Falsifying Marginal Treatment Effects: Supplemental Appendix

Minghai Mao* Pedro H. C. Sant'Anna† Xiaojun Song ‡

June 20, 2026

This supplemental appendix contains auxiliary lemmas, proofs of the main theorems, and additional results presented in the main text.

A Assumptions

This appendix presents the assumptions used to establish the asymptotic properties of the proposed tests. Throughout, $\rightsquigarrow$ denotes weak convergence in the sense of Hoffmann-Jorgensen. For a set $\mathbb{D}$, let $C(\mathbb{D})$ denote the space of continuous functions on $\mathbb{D}$ and let $l^\infty(\mathbb{D})$ denote the space of bounded functions on $\mathbb{D}$ equipped with the sup norm $\|f\|_\infty = \sup_{x\in\mathbb{D}} |f(x)|$. We use $\|\cdot\|_{2,\mu}$ for the $L_2(\mu)$ norm, write $\|\cdot\|_2$ when the measure is clear from the context, and let $|\cdot|$ denote the Euclidean norm.

To formalize the asymptotic arguments, we impose Assumptions A.1–A.5, which are slight modifications of Assumptions 1–6 in Escanciano, Jacho-Chávez and Lewbel (2014, EJL14 hereafter). Because Assumption A.1(iii) keeps the denominator in the kernel estimators bounded away from zero and the generated regressors in our setting are parametric, we do not require EJL14's Assumptions 7–11, which are mainly used to accommodate random trimming indicators and nonparametrically generated regressors. In addition, Assumption A.6 is added to control the effect of estimating the residuals and the propensity score quantile transformation in the copula based process.

Assumption A.1. *(i) The sample observations $\{(Y_i, Z_i, D_i)\}_{i=1}^n$ are a sequence of i.i.d. variables distributed as (Y, Z, D), where $Z_i = (X_i, Z_{0i})$. (ii) The parameter spaces C_α and C_{β_d} for $d \in \{0, 1\}$ are compact, and the true parameters $\alpha \in C_\alpha$ and $\beta_d \in C_{\beta_d}$, respectively. (iii) For all $\tilde\alpha \in C_\alpha$, $d \in \{0, 1\}$ and $p \in \mathcal{X}_P \subset (0, 1)$, the joint density of $(\tilde P, D)$, where $\tilde P = P(Z, \tilde\alpha)$, satisfies $f_{P,D}(p, d; \tilde\alpha) \geqslant \epsilon$ for some $\epsilon > 0$. (iv) The estimators $\hat\alpha$ and $\hat\beta_d$ satisfy the following asymptotic linear expansions:*

$$\sqrt{n}(\hat\alpha - \alpha) = \frac{1}{\sqrt{n}}\sum_{i=1}^n l(D_i, Z_i; \alpha) + o_p(1), \quad \sqrt{n}(\hat\beta_d - \beta_d) = \frac{1}{\sqrt{n}}\sum_{i=1}^n l_d(Y_i, Z_i; \beta_d, \alpha) + o_p(1), \tag{A.1}$$

where $l(\cdot)$ and $l_d(\cdot)$ denote the influence functions such that $E[l(D_i, Z_i; \alpha)] = 0$ and $E[l_d(Y_i, Z_i; \beta_d, \alpha)] = 0$, and the covariance matrices $E[l(D_i, Z_i; \alpha)l^\top(D_i, Z_i; \alpha)]$ and $E[l_d(Y_i, Z_i; \beta_d, \alpha)l_d^\top(Y_i, Z_i; \beta_d, \alpha)]$ exist and are

*Li Anmin Institute of Economic Research, Liaoning University
†Department of Economics, Emory University
‡Guanghua School of Management, Peking University

positive definite. There exist estimators $\hat{l}$ and $\hat{l}_d$ that satisfy

$$\frac{1}{n}\sum_{i=1}^{n}\|\hat{l}(D_i, Z_i; \hat{\alpha}) - l(D_i, Z_i; \alpha)\|^2 = o_p(1), \qquad \frac{1}{n}\sum_{i=1}^{n}\|\hat{l}_d(Y_i, Z_i; \hat{\beta}_d, \hat{\alpha}) - l_d(Y_i, Z_i; \beta_d, \alpha)\|^2 = o_p(1).$$

Also, the stochastic equicontinuity conditions for the nonparametric nuisance estimators used in the displayed influence function expansions are satisfied.

Assumption A.1(ii) is enough to restrict the class size of generated regressors. Assumption A.1(iii) implies $f_{P,D}(p, d; \tilde{\alpha})$ is bounded away from zero across the entire support range of the propensity score in both the treated and control groups for all $\tilde{\alpha} \in C_\alpha$. This requirement is inherently satisfied if the marginal treatment effect, $\text{MTE}(x, u)$, is identifiable for every u within the interval $\mathcal{X}_P$.

Assumption A.1(iv) is a high level asymptotic linearity condition for the first step estimators. It is satisfied for standard parametric propensity score estimators and for the Robinson estimator of the partially linear specification in (2.6), under the usual regularity, undersmoothing, and stochastic equicontinuity conditions for the nonparametric nuisance estimators. The discussion below sketches the verification for standard probit and Robinson first step estimators; the formal test validity only requires the high-level representations in Assumption A.1(iv).

For the propensity score, consider a probit specification $P(Z; \alpha) = \Phi(Z^\top\alpha)$ for example, where $\Phi(\cdot)$ denotes the standard normal distribution function. Suppose that $\hat{\alpha}$ is the maximum likelihood estimator with influence function

$$l(D_i, Z_i; \alpha) = I(\alpha)^{-1}\frac{\phi(Z_i^\top\alpha)\{D_i - \Phi(Z_i^\top\alpha)\}}{\Phi(Z_i^\top\alpha)\{1 - \Phi(Z_i^\top\alpha)\}}Z_i, \tag{A.2}$$

where $\phi(\cdot)$ is the standard normal density and $I(\alpha)$ is the Fisher information matrix.

With the parameter in propensity score function $\hat{\alpha}$ at hand, we next discuss the influence function for the estimator of the slope parameters $\beta = (\beta_0, \beta_1 - \beta_0)^\top$ in (2.6). Write $P_i(\hat{\alpha}) = P(Z_i; \hat{\alpha})$ and define the partialled-out variables $\mathring{Y}_i(\hat{\alpha}) = Y_i - E(Y_i \mid P_i(\hat{\alpha}))$, $\mathring{X}_i(\hat{\alpha}) = X_i - E(X_i \mid P_i(\hat{\alpha}))$, and $R_i(\hat{\alpha}) = (\mathring{X}_i(\hat{\alpha}), P_i(\hat{\alpha})\mathring{X}_i(\hat{\alpha}))^\top$. The Robinson estimator for β solves the sample analogue of

$$E\Big[R_i(\hat{\alpha})\left\{\mathring{Y}_i(\hat{\alpha}) - \mathring{X}_i(\hat{\alpha})^\top\beta_0 - P_i(\hat{\alpha})\mathring{X}_i(\hat{\alpha})^\top(\beta_1 - \beta_0)\right\}\Big] = 0.$$

Evaluated at the true parameter α, and treating the propensity score as known, the influence function of $\hat{\beta}$ is

$$l_\beta^{\text{Rob}}(Y_i, Z_i; \beta, \alpha) = S_{RR}^{-1}(\alpha)R_i(\alpha)\left\{\mathring{Y}_i(\alpha) - \mathring{X}_i(\alpha)^\top\beta_0 - P_i(\alpha)\mathring{X}_i(\alpha)^\top(\beta_1 - \beta_0)\right\}, \tag{A.3}$$

where $S_{RR}(\alpha) = E\left[R_i(\alpha)R_i(\alpha)^\top\right]$.

Because $P_i(\alpha) = P(Z_i; \alpha)$ is replaced by $P_i(\hat{\alpha}) = P(Z_i; \hat{\alpha})$ in practice, the influence function must also account for the first-step estimation of α. Hence

$$l_\beta(Y_i, Z_i; \beta, \alpha) = l_\beta^{\text{Rob}}(Y_i, Z_i; \beta, \alpha) + \Gamma_\alpha l(D_i, Z_i; \alpha), \tag{A.4}$$

where

$$\Gamma_\alpha = S_{RR}^{-1}(\alpha)E\Big[\partial_\alpha\left\{R_i(\alpha)\big(\mathring{Y}_i(\alpha) - \mathring{X}_i(\alpha)^\top\beta_0 - P_i(\alpha)\mathring{X}_i(\alpha)^\top(\beta_1 - \beta_0)\big)\right\}\Big]. \tag{A.5}$$

Equivalently, under the index sufficiency restrictions used in the partially linear MTE model, this adjustment can be written as

$$\Gamma_\alpha = -S_{RR}^{-1}(\alpha)E\big[R_i(\alpha)\,\nabla_\alpha P_i(\alpha)\,\{X_i^\top(\beta_1-\beta_0)+m'(P_i)\}\big]\,. \tag{A.6}$$

The scalar term $X_i^\top(\beta_1-\beta_0)+m'(P_i)$ is the marginal treatment effect implied by (2.6). Thus, the generated-regressor correction in (A.4) is exactly the effect of estimating the propensity score index inside the Robinson moment.

The stacked influence function for the Robinson estimator of $(\hat\beta_0,\,\hat\beta_1-\hat\beta_0)^\top$ is $l_\beta=(l_0,\,l_1-l_0)^\top$. The treatment-specific influence functions follow by the trivial identifications $l_0(Y_i,Z_i;\beta,\alpha)=l_0$ for $d=0$ and $l_1(Y_i,Z_i;\beta,\alpha)=l_0+[l_1-l_0]$ for $d=1$.

Remark A.1 (Role of index sufficiency in the first-step adjustment)**.** We explain how (A.5) simplifies to (A.6). Applying the product rule gives

$$E[\partial_\alpha\{R_i(\alpha)\,u_i(\alpha)\}]=E[(\partial_\alpha R_i(\alpha))\,u_i(\alpha)]+E[R_i(\alpha)\,\partial_\alpha u_i(\alpha)]\,.$$

where $u_i(\alpha)=\mathring{Y}_i(\alpha)-\mathring{X}_i(\alpha)^\top\beta_0-P_i(\alpha)\mathring{X}_i(\alpha)^\top(\beta_1-\beta_0)=Y_i-X_i^\top\beta_0-P_i(\alpha)X_i^\top(\beta_1-\beta_0)-m(P_i(\alpha))$.

For the *first term*: index sufficiency means the structural error $e_i=u_i(\alpha_0)$ satisfies $E(e_i\mid Z_i)=0$. Since $\partial_\alpha R_i(\alpha)$ depends on data only through Z_i, iterated expectations give $E[(\partial_\alpha R_i)\,e_i]=0$, so the first term vanishes. For the *second term*: the α-dependence of $u_i(\alpha)$ enters entirely through the propensity score index $P_i(\alpha)=P(Z_i,\alpha)$. By the chain rule, $\partial_\alpha u_i(\alpha)=(\partial_P u_i)\cdot\nabla_\alpha P_i$. Differentiating with respect to P yields:

$$\partial_P u_i=-\big\{X_i^\top(\beta_1-\beta_0)+m'(P_i)\big\},$$

where $-\partial_P u_i$ equals the MTE at (X_i,P_i) in (2.7). Substituting into the second term delivers (A.6).

If the empirical implementation uses a different first-step estimator, the same theory applies provided the conditions in Assumption A.1(iv) hold.

Assumption A.2. *The support $\mathcal{X}_P$ of the propensity score P is a compact interval $[\underline{p},\bar p]\subset[0,1]$. For $d\in\{0,1\}$, $\tilde\alpha\in C_\alpha$, $\tilde\beta_d\in C_{\beta_d}$, $v\in\mathcal{X}_V$, $p\in\mathcal{X}_P$, and $z\in\mathcal{X}_Z$, $f_{Z|P,D}(z\mid p,d;\tilde\alpha)$, $f_{P,D}(p,d;\tilde\alpha)$, and $F_{V|P,D}(v\mid p,d;\tilde\beta_d,\tilde\alpha)$ are r times continuously differentiable in p, and all these derivatives are uniformly bounded over the stated arguments.*

Assumption A.3. *The kernel function $K(t):\mathbf{R}\to\mathbf{R}$ is bounded, symmetric, continuously differentiable, and satisfies the following conditions: $\int K(t)dt=1$, $\int t^lK(t)dt=0$ for $l=1,\dots,r-1$, and $\int|t^rK(t)|dt<\infty$ for some $r\geqslant 2$; $|\partial K(t)/\partial t|\leqslant C$, and for some $\nu>1$, $|\partial K(t)/\partial t|\leqslant C|t|^{-\nu}$ for $|t|>L$, $0<L<\infty$.*

Assumption A.4. *The possibly data-dependent bandwidth h satisfies $\Pr(a_n\leqslant h\leqslant b_n)\to 1$ as $n\to\infty$, with a_n and b_n being deterministic sequences of positive numbers such that: (i) $b_n\to 0$ and $a_n n^{\frac12}/\log(n)\to\infty$; (ii) $nb_n^{2r}\to 0$.*

Assumptions A.2–A.4 provide the low-level regularity conditions used to control the bias and stochastic expansion of the kernel estimators. In particular, Assumption A.2 imposes the baseline smoothness in the index variable p, while Assumptions A.3 and A.4 specify the kernel and bandwidth restrictions. For example, the deterministic bandwidth $h=cn^{-\delta}$ with $\frac{1}{2r}<\delta<\frac12$ suffices for our purpose.

For the empirical-process arguments underlying the index statistic, however, these low-level smoothness conditions are not sufficient on their own. We therefore impose in Assumption A.5 a stronger high-level complexity condition on the relevant function classes, which strengthens the smoothness requirement in p and the tail behavior in v.

Assumption A.5. *Given $d \in \{0,1\}$, the class $\mathcal{F}^\eta$ consists of composite measurable maps on $\mathcal{X}_Z$ of the form $z \mapsto q_d(\tilde{P}(z) \,|\, v, \tilde{\beta}_d, \tilde{\alpha})$, where $q_d(\cdot \,|\, v, \tilde{\beta}_d, \tilde{\alpha}) : \mathcal{X}_P \to \mathbf{R}$ is indexed by $(v, \tilde{\beta}_d, \tilde{\alpha})$. The map $q_d(\cdot \,|\, v, \tilde{\beta}_d, \tilde{\alpha})$ satisfies the following three conditions:*

– *(a)* Lipschitz continuity: *There exists a universal constant C_L such that*

$$\left\| q_d(\cdot \,|\, v_1, \tilde{\beta}_{d1}, \tilde{\alpha}_1) - q_d(\cdot \,\big|\, v_2, \tilde{\beta}_{d2}, \tilde{\alpha}_2) \right\|_\infty \leqslant C_L \left|(v_1, \tilde{\beta}_{d1}, \tilde{\alpha}_1) - (v_2, \tilde{\beta}_{d2}, \tilde{\alpha}_2)\right|,$$

where the supremum is taken over $z \in \mathcal{X}_Z$.

– *(b)* Hölder smoothness: *For some $\eta \geqslant r$, for each $(v, \tilde{\beta}_d, \tilde{\alpha}) \in \mathcal{X}_V \times C_{\beta_d} \times C_\alpha$ we have*

$$q_d(\cdot \,|\, v, \tilde{\beta}_d, \tilde{\alpha}) \in C^\eta(\mathcal{X}_P),$$

where $C^\eta(\mathcal{X}_P)$ denotes the Hölder class of order η on $\mathcal{X}_P$ equipped with the norm

$$\|q_d(\cdot \,|\, v, \tilde{\beta}_d, \tilde{\alpha})\|_{\infty,\eta} := \max_{\kappa \leqslant \underline{\eta}} \sup_{p \in \mathcal{X}_P} \left|\partial^\kappa q_d(p \,|\, v, \tilde{\beta}_d, \tilde{\alpha})\right| + \max_{\kappa = \underline{\eta}} \sup_{p \neq p'} \frac{\left|\partial^\kappa q_d(p \,|\, v, \tilde{\beta}_d, \tilde{\alpha}) - \partial^\kappa q_d(p' \,|\, v, \tilde{\beta}_d, \tilde{\alpha})\right|}{\|p - p'\|^{\eta - \underline{\eta}}},$$

with $\underline{\eta}$ the largest integer strictly smaller than η. We assume that there exists $M > 0$ such that $\|q_d(\cdot \,|\, v, \tilde{\beta}_d, \tilde{\alpha})\|_{\infty,\eta} \leqslant M$.

– *(c)* Uniform tails. *There exist constants $C_0, c_0 > 0$ such that, for all $M > 0$,*

$$\sup_{(\tilde{\beta}_d, \tilde{\alpha}) \in C_{\beta_d} \times C_\alpha} \sup_{p \in \mathcal{X}_P} q_d(-M, p \,|\, \tilde{\beta}_d, \tilde{\alpha}) + \sup_{(\tilde{\beta}_d, \tilde{\alpha}) \in C_{\beta_d} \times C_\alpha} \sup_{p \in \mathcal{X}_P} \{1 - q_d(M, p \,|\, \tilde{\beta}_d, \tilde{\alpha})\} \leqslant C_0 e^{-c_0 M}.$$

Then, we have (i) $F_{V|P,D}(v|\cdot, d; \tilde{\beta}_d, \tilde{\alpha}) \in \mathcal{F}^\eta$; (ii) $\Pr(\hat{F}_{V|P,D}(v|\cdot, d; \tilde{\beta}_d, \tilde{\alpha}) \in \mathcal{F}^\eta) \to 1$. In addition, define

$$G_d(z \mid p; \tilde{\alpha}) := \Pr(Z \leqslant z \mid \tilde{P} = p, D = d),$$

and let $\mathcal{G}^\eta$ denote the corresponding class of composite maps with the same Lipschitz continuity in $(z, \tilde{\alpha})$ and the same Hölder smoothness in p as imposed above for $\mathcal{F}^\eta$. Assume that (i) $G_d(z \mid \cdot; \tilde{\alpha}) \in \mathcal{G}^\eta$; (ii) $\Pr(\hat{G}_d(z \mid \cdot; \tilde{\alpha}) \in \mathcal{G}^\eta) \to 1$.

Assumption A.5 is a high level complexity condition that delivers the entropy and stochastic equicontinuity required for the uniform empirical process approximation in Lemma E.1; see, e.g., Assumption 6 in Escanciano et al. (2014) and Assumption 3.4 in Ichimura and Lee (2010). Assumption A.2 imposes the baseline smoothness in p used mainly for kernel bias control. To handle the parametric perturbations induced by $(\tilde{\alpha}, \tilde{\beta}_d)$, we further impose Assumption A.6, which requires continuous differentiability of $f_{P,D}$ and $F_{V|P,D}$ with respect to $(p, \tilde{\alpha})$ and $(p, \tilde{\beta}_d, \tilde{\alpha})$. These differentiability conditions are central for the numerical derivative estimation of the correction terms and, at the same time, provide the local stability and stochastic equicontinuity used in Lemmas E.1 and E.2.

Assumption A.6. *For each $d \in \{0,1\}$, the following conditions hold.*

(i) The joint density $f_{P,D}(p,d;\tilde{\alpha})$ is continuously differentiable in $(p,\tilde{\alpha})$ on $\mathcal{X}_P \times C_\alpha$, with uniformly continuous and bounded derivatives.

(ii) The conditional distribution $F_{V|P,D}(v \mid p,d;\tilde{\beta}_d,\tilde{\alpha})$ is continuously differentiable in $(p,\tilde{\beta}_d,\tilde{\alpha})$, with uniformly continuous and bounded derivatives over

$$(v,p,\tilde{\beta}_d,\tilde{\alpha}) \in \mathcal{X}_V \times \mathcal{X}_P \times C_{\beta_d} \times C_\alpha.$$

In addition, $F_{V|P,D}(\cdot \mid p,d;\tilde{\beta}_d,\tilde{\alpha})$ admits a conditional density in v that is uniformly bounded over the same set.

The implications of Assumption A.6 can be organized as follows. First, it gives boundary mass controls for both directions of generated regressor perturbation. By Assumption A.6(i), uniformly in $\tilde{\alpha}$,

$$\sup_{p\in\mathcal{X}_P} \Pr\{|P(Z;\tilde{\alpha}) - p| \leqslant \delta\} \leqslant C\delta.$$

This controls perturbations of the propensity score density function through $\tilde{\alpha}$.

By Assumption A.6(ii), uniformly in $(\tilde{\alpha},\tilde{\beta}_d)$,

$$\sup_{p\in\mathcal{X}_P,\, v\in\mathcal{X}_V} \Pr\{|Y - X^\top\tilde{\beta}_d - v| \leqslant \delta \mid D = d, P(Z;\tilde{\alpha}) = p\} \leqslant C\delta,$$

and, for $|p - p'| \leqslant \delta$,

$$\sup_{p\in\mathcal{X}_P,\, v\in\mathcal{X}_V} \left|\Pr\{Y - X^\top\tilde{\beta}_d \leqslant v \mid D = d, P(Z;\tilde{\alpha}) = p\} - \Pr\{Y - X^\top\tilde{\beta}_d \leqslant v \mid D = d, P(Z;\tilde{\alpha}) = p'\}\right| \leqslant C\delta.$$

The first bound controls local perturbations in the residual direction, while the second controls local perturbations in the propensity-score direction. In particular, changes in $\tilde{\beta}_d$ shift the residual threshold by $X^\top(\tilde{\beta}_d - \beta_d)$, whereas changes in $\tilde{\alpha}$ shift the propensity score $P(Z;\tilde{\alpha})$. Hence, the continuous differentiability, together with the uniform continuity and boundedness of the derivatives imposed in Assumption A.6, provides the local stability needed to control perturbations induced by $\tilde{\beta}_d$ and $\tilde{\alpha}$.

Second, by Assumptions A.1(iii), A.2, and A.6(i), the marginal density $f_P(p;\tilde{\alpha}) = \sum_{d=0}^{1} f_{P,D}(p,d;\tilde{\alpha})$ is bounded away from zero and continuously differentiable on its full support. Hence the quantile map $F_P^{-1}(u;\tilde{\alpha})$ is well defined for all $u \in [0,1]$. Moreover, the same boundary-mass control applies uniformly at the induced quantile cutoffs:

$$\sup_{u\in[0,1]} \Pr\left(\left|P(Z;\tilde{\alpha}) - F_P^{-1}(u;\tilde{\alpha})\right| \leqslant \delta\right) \leqslant C\delta.$$

Third, using $f_{P,D}$ and $F_{V|P,D}$, define $F_{V,D|P}(v,d \mid p;\tilde{\beta}_d,\tilde{\alpha})$. Then the copula process can be written as

$$C_d(\lambda;\tilde{\alpha},\tilde{\beta}_d) = \int_0^u \Big[F_{V,D|P}\Big(v_2,d \mid F_P^{-1}(\bar{u};\tilde{\alpha});\tilde{\beta}_d,\tilde{\alpha}\Big) - F_{V,D|P}\Big(v_1,d \mid F_P^{-1}(\bar{u};\tilde{\alpha});\tilde{\beta}_d,\tilde{\alpha}\Big)\Big]\,d\bar{u}.$$

Therefore, the smoothness of $f_{P,D}$ and $F_{V|P,D}$, together with the full-domain stability of the quantile map established above, implies that $C_d(\lambda;\tilde{\alpha},\tilde{\beta}_d)$ is continuously differentiable in $(u,\tilde{\alpha},\tilde{\beta}_d)$ for $u \in [0,1]$. Thus, under our maintained conditions, the copula process and the partial LCM operator can be analyzed on the full rank set $[0,1]$, consistent with the formulation in Seo (2018); no interior trimming is required.[1]

Remark A.2 (Numerical derivative estimation)**.** Let $a_n \to 0$ satisfy $na_n/\log n \to \infty$. In the implementation, the terms $\hat{A}_d$, $\hat{B}_d$, $\partial_u\hat{C}_d$, $\partial_{\beta_d}\hat{C}_d$, and $\partial_\alpha\hat{C}_d$, whose explicit forms are given in Section B, are computed by symmetric finite differences with perturbation step a_n. In the implementation we set the order of a_n as $n^{-1/2}$, which satisfies the required rate condition. The uniform consistency of these numerical derivatives is established in the proof of Theorem 3.1 under Assumptions A.2 and A.6.

B Implementation of the Test

In this section, we describe how to compute the test statistics $\hat{T}_I$ and $\hat{T}_M$ and implement the bootstrap inference procedure described in **Algorithm**.

Let $\{(Y_i,Z_i,D_i)\}_{i=1}^n$ denote a sample of size n from the joint distribution of (Y,Z,D). Throughout this section and the proofs below, we use P_i, $\hat{P}_i$, and $\tilde{P}_i$ as shorthand for $P_i(\alpha)$, $P_i(\hat{\alpha})$, and $P_i(\tilde{\alpha})$, respectively. More generally, we suppress the dependence of estimation objects on $(\hat{\alpha},\hat{\beta}_d)$ and of population quantities on (α,β_d) when no confusion arises; parameter arguments are made explicit when taking derivatives or considering perturbations of first step estimators.

Using this notation, we proceed as follows. In the following subsections, we construct the empirical processes $\hat{U}_d(\omega)$ and $\hat{C}_d(\lambda)$ underlying the test statistics $\hat{T}_I$ and $\hat{T}_M$, respectively. For the asymptotic analysis, we derive the corresponding influence functions ζ_d^I and ζ_d^M and use their sample analogues, $\hat{\zeta}_d^I$ and $\hat{\zeta}_d^M$, to implement the multiplier bootstrap. For the monotonicity statistic $\hat{T}_M$, we also describe the numerical approximation to the functional derivative of the least concave majorant (LCM) operator required for computation.

B.1 The Index Statistics and Its Influence Function

The sample analogue of $\sqrt{n}U_d(\omega)$ in (3.1) is

$$\sqrt{n}\hat{U}_d(\omega) = \frac{1}{\sqrt{n}}\sum_{i=1}^n[\mathbf{1}(\hat{V}_{di} \leqslant v) - \hat{F}_{V|P,D}(v|\hat{P}_i,d;\hat{\beta}_d,\hat{\alpha})]\psi_i(p,z,d;\hat{\alpha}), \tag{B.1}$$

where $\psi_i(p,z,d;\hat{\alpha}) = \mathbf{1}(\hat{P}_i \leqslant p, Z_i \leqslant z, D_i = d)$ is the weighting function.

[1] The standard Bahadur representation for empirical quantiles is used only on compact subsets of $(0,1)$. The endpoint cases $u=0$ and $u=1$ are treated separately: the corresponding boundary terms are either zero or reduce to empirical processes indexed only by (v_1,v_2), whose limits follow from the same VC-type empirical process arguments. Hence the full domain statement on $[0,1]$ is obtained by combining the interior expansion with stochastic equicontinuity and continuous extension.

We estimate the conditional CDF, $F_{V|P,D}(v|P_i, d; \beta_d, \alpha) = \Pr(V_i \leqslant v \mid P_i, D_i = d)$, by the Nadaraya–Watson estimator $\hat{F}_{V|P,D}(v|\hat{P}_i, d; \hat{\beta}_d, \hat{\alpha})$ constructed from the generated regressors $\hat{P}_i$ and $\hat{V}_{di}$:

$$\hat{F}_{V|P,D}(v|\hat{P}_i, d; \hat{\beta}_d, \hat{\alpha}) = \frac{\sum_{j=1}^n \mathbf{1}(\hat{V}_{dj} \leqslant v, D_j = d) K_h(\hat{P}_j - \hat{P}_i)}{\sum_{j=1}^n \mathbf{1}(D_j = d) K_h(\hat{P}_j - \hat{P}_i)}, \tag{B.2}$$

with $K_h(t) = \frac{1}{h} K(\frac{t}{h})$, where $K(\cdot)$ is the kernel function and $h = h_n$ is the bandwidth. The notation $\hat{F}_{V|P,D}(v \mid p, d; \tilde{\beta}_d, \tilde{\alpha})$ and $F_{V|P,D}(v \mid p, d; \tilde{\beta}_d, \tilde{\alpha})$ that places $\tilde{\beta}_d$ before $\tilde{\alpha}$ highlights $\tilde{\beta}_d$ governs the residual $V(\tilde{\beta}_d)$, whereas $\tilde{\alpha}$ governs the conditional index $P(\tilde{\alpha})$.

Lemma E.1 shows that $\hat{U}_d(\omega)$ admits an asymptotic linear representation with influence function ζ_d^I by

$$\begin{aligned} \zeta_d^I(Y_i, D_i, Z_i; \omega, \alpha, \beta_d) = &\left[\mathbf{1}(V_{di} \leqslant v) - F_{V|P,D}(v|P_i, d; \beta_d, \alpha)\right] \psi_i^{\perp}(p, z, d; \alpha) \\ &+ A_d^{\top} l(D_i, Z_i; \alpha) + B_d^{\top} l_d(Y_i, Z_i; \beta_d, \alpha), \end{aligned} \tag{B.3}$$

where $\psi_i^{\perp}(p, z, d; \alpha) = \psi_i(p, z, d; \alpha) - \iota_d(P_i)\mathbf{1}(D_i = d)$ denotes the orthogonalized weighting function with the projection $\iota_d(P_i) = E[\psi_i(p, z, d; \alpha) \mid P = P_i, D_i = d] = \mathbf{1}(P_i \leqslant p) G_d(z \mid P_i)$.

The first term captures the nonparametric estimation effect arising from kernel estimation of $F_{V|P,D}(\cdot)$, while $A_d^{\top} l(D_i, Z_i; \alpha)$ and $B_d^{\top} l_d(Y_i, Z_i; \beta_d, \alpha)$ account for the parametric estimation effects of $\hat{\alpha}$ and $\hat{\beta}_d$, respectively. The terms A_d and B_d are respectively given by

$$\begin{aligned} A_d &= -E[\partial_\alpha F_{V|P,D}(v|P_i, d; \beta_d, \alpha) \psi_i^{\perp}(p, z, d; \alpha)] \\ B_d &= \partial_{\beta_d} F_{V,P,Z,D}(v, p, z, d; \beta_d, \alpha) - E[\partial_{\beta_d} F_{V|P,D}(v|P_i, d; \beta_d, \alpha) \psi_i(p, z, d; \alpha)], \end{aligned} \tag{B.4}$$

where $F_{V,P,Z,D}(v, p, z, d; \beta_d, \alpha) = \Pr(V_i(\beta_d) \leqslant v, P_i(\alpha) \leqslant p, Z_i \leqslant z, D_i = d)$.

To implement the multiplier bootstrap, we replace the unknown quantities in ζ_d^I by sample analogues and define the plug-in estimator

$$\begin{aligned} \hat{\zeta}_d^I(Y_i, D_i, Z_i; \omega, \hat{\alpha}, \hat{\beta}_d) = &[\mathbf{1}(\hat{V}_{di} \leqslant v) - \hat{F}_{V|P,D}(v|\hat{P}_i, d; \hat{\beta}_d, \hat{\alpha})] \hat{\psi}_i^{\perp}(p, z, d; \hat{\alpha}) \\ &+ \hat{A}_d^{\top} \hat{l}(D_i, Z_i; \hat{\alpha}) + \hat{B}_d^{\top} \hat{l}_d(Y_i, Z_i; \hat{\beta}_d, \hat{\alpha}), \end{aligned} \tag{B.5}$$

Let $\hat{A}_d$, $\hat{B}_d$, and $\hat{\psi}_i^{\perp}(p, z, d; \hat{\alpha}) = \psi_i(p, z, d; \hat{\alpha}) - \hat{\iota}_d(\hat{P}_i; \hat{\alpha}) \mathbf{1}(D_i = d)$ denote the corresponding sample analogues of A_d, B_d. Here, $\hat{\iota}_d(\hat{P}_i; \hat{\alpha}) = \hat{E}[\psi_i(p, z, d; \hat{\alpha}) \mid P = \hat{P}_i, D = d]$ is computed by the NW estimator.

In practice, we evaluate the correction term

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \left[\hat{A}_d^{\top} \hat{l}(D_i, Z_i; \hat{\alpha}) + \hat{B}_d^{\top} \hat{l}_d(Y_i, Z_i; \hat{\beta}_d, \hat{\alpha})\right]$$

using a numerical perturbation scheme. Specifically, we perturb $\hat{\alpha}$ and $\hat{\beta}_d$ by their bootstrap counterparts and record the induced change in the estimated distributional objects of $F_{V|P,D}(v|P_i, d; \beta_d, \alpha) \psi_i^{\perp}(p, z, d; \alpha)$, $F_{V,P,Z,D}(v, p, z, d; \beta_d, \alpha)$ and $F_{V|P,D}(v|P_i, d; \beta_d, \alpha) \psi_i(p, z, d; \alpha)$. The only additional regularization needed in this step concerns the estimation of $\partial_{\beta_d} F_{V,P,Z,D}(v, p, z, d; \beta_d, \alpha)$. The empirical CDF based on $\mathbf{1}(\hat{V}_{di} \leqslant v, \hat{P}_i \leqslant p, Z_i \leqslant z, D_i = d)$ is discontinuous in $\hat{V}_{di}$ and therefore yields unstable numerical derivatives. To overcome this issue, we replace the indicator $\mathbf{1}(\hat{V}_{di} \leqslant v)$ by the smoothed version

$$\int_{-\infty}^{v} K_\ell(\hat{V}_{di} - \bar{v})\, d\bar{v},$$

where $K_\ell(t) = \frac{1}{\ell}K(\frac{t}{\ell})$ and $\ell = \ell_n$ is a smoothing bandwidth. This regularization stabilizes the numerical derivative while preserving the asymptotic properties of the estimator.

B.2 The Monotonicity Statistics and its Influence Function

The feasible estimator for the estimated copula process is $\hat{C}_d(\lambda) = \hat{C}_d(\lambda; \hat{\alpha}, \hat{\beta}_d)$, so that

$$\hat{C}_d(\lambda) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(v_1 \leqslant \hat{V}_{di} \leqslant v_2, D_i = d, \hat{F}_P(\hat{P}_i) \leqslant u). \tag{B.6}$$

where $\hat{F}_P(p) = n^{-1}\sum_{i=1}^{n} \mathbf{1}(\hat{P}_i \leqslant p)$.

Since $\hat{V}_{di}$ and $\hat{P}_i$ are generated regressors, the asymptotic behaviour of $\hat{C}_d(\lambda)$ must account for first-step estimation effects. Under regularity conditions, the empirical copula process $\sqrt{n}\{\hat{C}_d(\lambda) - C_d(\lambda)\}$ admits an asymptotic linear representation with influence function ζ_d^M given by

$$\begin{aligned}
&\zeta_d^M(Y_i, D_i, Z_i, \lambda, \alpha, \beta_d) \\
=&\mathbf{1}(v_1 \leqslant V_{di} \leqslant v_2, D_i = d, F_P(P_i) \leqslant u) - C_d(\lambda) - \partial_u C_d(\lambda)\{\mathbf{1}(F_P(P_i) \leqslant u) - u\} \\
&+ l^\top(D_i, Z_i, \alpha)\partial_\alpha C_d(\lambda) + l_d^\top(Y_i, Z_i, \beta_d)\partial_{\beta_d} C_d(\lambda).
\end{aligned} \tag{B.7}$$

The first three terms on the right-hand side coincide with the influence function of the classical empirical copula process; see Gänssler and Stute (1987) and Fermanian, Radulovic and Wegkamp (2004). The last two terms account for the estimation effects of $\hat{\alpha}$ and $\hat{\beta}_d$ induced by the plug-in procedure; see Van der Vaart (2000). Here, $\partial_\alpha C_d(\lambda)$ denotes the total derivative. By the chain rule,

$$\partial_\alpha C_d(\lambda) = \partial_\alpha H_d(v_1, v_2, F_P^{-1}(u); \alpha, \beta_d) + \partial_u C_d(\lambda)\, f_P(F_P^{-1}(u))\, \partial_\alpha F_P^{-1}(u; \alpha),$$

where $H_d(v_1, v_2, t; \alpha, \beta_d) = \Pr(v_1 \leqslant V_{di} \leqslant v_2, D_i = d, P_i \leqslant t)$. Equivalently, using $\partial_\alpha F_P^{-1} = -f_P^{-1}\partial_\alpha F_P$, $\partial_\alpha C_d = \partial_\alpha H_d - \partial_u C_d\, \partial_\alpha F_P$. Thus, $\partial_\alpha C_d(\lambda)$ incorporates both the direct effect of α on H_d and the indirect effect through the rank transformation. The second term is the derivative of the quantile map. Thus, when $\partial_\alpha \hat{C}_d$ is computed by perturbing α, the propensity scores, empirical distribution, quantiles, and ranks must be recomputed under the perturbed value of α.

For implementation, we replace the unknown quantities by their sample analogues and define the plug-in estimator:

$$\begin{aligned}
&\hat{\zeta}_d^M(Y_i, D_i, Z_i; \lambda, \hat{\alpha}, \hat{\beta}_d) \\
=&\mathbf{1}(v_1 \leqslant \hat{V}_{di} \leqslant v_2, D_i = d, \hat{F}_P(\hat{P}_i) \leqslant u) - \hat{C}_d(\lambda) - \partial_u \hat{C}_d(\lambda)\{\mathbf{1}(\hat{F}_P(\hat{P}_i) \leqslant u) - u\} \\
&+ \hat{l}_d^\top(Y_i, Z_i; \hat{\beta}_d, \hat{\alpha})\partial_{\beta_d}\hat{C}_d(\lambda) + \hat{l}^\top(D_i, Z_i; \hat{\alpha})\partial_\alpha \hat{C}_d(\lambda),
\end{aligned}$$

where $\partial_u \hat{C}_d(\lambda)$ and the remaining derivative terms are consistent estimators of their population counterparts.

In practice, no fully data-driven optimal choice is currently available for estimating $\partial_u C_d(\lambda)$, and rule-of-thumb smoothing such as $h = n^{-1/2}$ is commonly adopted; see Bücher and Dette (2010). In addition, the aggregate contribution

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\big[\hat{l}_d^{\top}(Y_i, Z_i; \hat{\beta}_d, \hat{\alpha})\partial_{\beta_d}\hat{C}_d(\lambda) + \hat{l}^{\top}(D_i, Z_i; \hat{\alpha})\partial_{\alpha}\hat{C}_d(\lambda)\big]$$

is computed using the same numerical perturbation scheme described above, perturbing $\hat{\alpha}$ and $\hat{\beta}_d$ by their bootstrap counterparts and recording the induced change in $\hat{C}_d(\lambda)$. When perturbing $\hat{\alpha}$, the resulting empirical ranks are recomputed as described above, so that the numerical derivative estimates the total derivative $\partial_\alpha C_d(\lambda)$.

To approximate the limiting copula process $\sqrt{n}\{\hat{C}_d(\lambda) - C_d(\lambda)\}$, we employ a multiplier bootstrap procedure, which is justified under conditional weak convergence in probability; see Theorem 2.9.6 of Van der Vaart and Wellner (1996).

B.3 The LCM Operator and Its Functional Derivative

We first recall the definition of the least concave majorant (LCM) operator on an interval $[a, b]$. For a bounded function $f : [a, b] \to \mathbb{R}$, define

$$\theta_{[a,b]}(f(\cdot))(u) \equiv \inf\{g(\cdot) : g \text{ is concave in } t \text{ and } g(t) \geqslant f(t) \text{ for all } t \in [a, b]\}, \tag{B.8}$$

and we write θ as shorthand for $\theta_{[0,1]}$.

We next introduce the partial LCM operator acting on a function f defined on $\Lambda = \{(v_1, v_2, u) : v_1 \leqslant v_2,\ u \in [0, 1]\}$. The operator applies the LCM in the u direction for each fixed (v_1, v_2). Formally, the partial LCM operator $\tilde{\theta}$ is defined by

$$\tilde{\theta}(f)(v_1, v_2, u) \equiv \theta\big(f(v_1, v_2, \cdot)\big)(u), \tag{B.9}$$

that is,

$$\tilde{\theta}(f)(v_1, v_2, u) \equiv \inf\{g_{v_1,v_2}(u) : g_{v_1,v_2} \text{ is concave in } u \text{ and } g_{v_1,v_2}(u) \geqslant f(v_1, v_2, u) \text{ for all } u \in [0, 1]\}.$$

The partial LCM operator $\tilde{\theta} : l^{\infty}(\Lambda) \to l^{\infty}(\Lambda)$ is Hadamard directionally differentiable tangentially to $C(\Lambda)$; see Shapiro (1990) and Fang and Santos (2018). This property allows us to apply the functional delta method to derive the asymptotic distribution of $\hat{T}_M$.

For $\lambda = (v_1, v_2, u) \in \Lambda$ and $f \in l^{\infty}(\Lambda)$, the directional derivatives of the partial LCM operator $\tilde{\theta}$ at the relevant concave functions $C_0(v_1, v_2, \cdot)$ and $-C_1(v_1, v_2, \cdot)$ are obtained by applying the LCM to the direction f over the maximal intervals of affinity of $C_0(v_1, v_2, \cdot)$ and $-C_1(v_1, v_2, \cdot)$:

$$\tilde{\theta}'_{C_0} f(\lambda) = \tilde{\theta}_{\mathcal{I}_{0,\lambda}}(f)(\lambda), \qquad \tilde{\theta}'_{-C_1} f(\lambda) = \tilde{\theta}_{\mathcal{I}_{1,\lambda}}\big(f\big)(\lambda), \tag{B.10}$$

where $\mathcal{I}_{0,\lambda}$ and $\mathcal{I}_{1,\lambda}$ are the maximal intervals containing u on which the concave functions $C_0(v_1, v_2, \cdot)$ and $-C_1(v_1, v_2, \cdot)$ are affine in u, respectively.

Consequently, the directional derivatives of the gap operators are

$$\dot{\mathcal{T}}_0(f)(\lambda) = \tilde{\theta}_{\mathcal{I}_{0,\lambda}}(f)(\lambda) - f(\lambda), \qquad \dot{\mathcal{T}}_1(f)(\lambda) = \tilde{\theta}_{\mathcal{I}_{1,\lambda}}(-f)(\lambda) + f(\lambda). \tag{B.11}$$

Note that for the gap operator $\mathcal{T}_1(f) = \tilde{\theta}(-f) + f$ when $d = 1$, the chain rule introduces a sign change inside the LCM directional derivative of $\dot{\mathcal{T}}_1$. Moreover, if $C_0(v_1, v_2, \cdot)$ is strictly concave or $C_1(v_1, v_2, \cdot)$ is strictly convex at u, then $\mathcal{I}_{0,\lambda} = u$ and $\mathcal{I}_{1,\lambda} = u$, respectively, so that the corresponding derivatives reduce to the zero map.

The contact set $B_C = B_{C_1} \cup B_{C_0}$ is defined as

$$B_{C_d} = \{\lambda \in \Lambda : C_d(v_1, v_2, \cdot) \text{ is locally affine at } u\}. \tag{B.12}$$

A non-trivial contact set implies that $C_d(v_1, v_2, \cdot)$ is affine in u on a set of positive measure for some $v_1 < v_2$, so that the monotonicity inequalities bind with positive measure.

The LCM operator itself can be computed using standard algorithms as in Delgado and Escanciano (2012); Seo (2018). For its directional derivative, one may either use the numerical approximation in (3.9) or estimate the contact set B_C explicitly following Beare and Moon (2015) and Seo (2018). Because the contact set based approach requires additional tuning parameters and is computationally more burdensome, we adopt the numerical estimator for stability and ease of implementation.

C Proof for Section 2

Proposition 2.1. Parts (i) and (ii) are the observable implications of Assumptions 1–3 established by Heckman and Vytlacil (2005). Part (i) follows by taking $G(Y) = \mathbf{1}(Y \leqslant y)$ in their characterization, and part (ii) follows by taking $G(Y) = \mathbf{1}(y_1 \leqslant Y \leqslant y_2)$.

It remains to prove the sharpness claim in part (iii). Throughout this proof all probabilities are conditional on $X = x$, for a fixed $x \in \mathcal{X}_X$. Let $\mathcal{P}_x$ denote the support of $P(Z)$ conditional on $X = x$. By the index sufficiency restriction, for any z, z' with $P(z) = P(z') = p \in \mathcal{P}_x$,

$$\Pr(Y \in B, D = d \mid Z = z) = \Pr(Y \in B, D = d \mid Z = z')$$

for every Borel set $B \subseteq \mathcal{X}_Y$ and $d \in \{0, 1\}$; the equality first holds for half-lines by (2.4) and then extends to Borel sets by the monotone class theorem. Hence the following subdistributions are well defined:

$$Q_{1,x}(B, p) = \Pr(Y \in B, D = 1 \mid P = p, X = x), \qquad Q_{0,x}(B, p) = \Pr(Y \in B, D = 0 \mid P = p, X = x).$$

They satisfy

$$Q_{1,x}(\mathcal{X}_Y, p) = p, \qquad Q_{0,x}(\mathcal{X}_Y, p) = 1 - p.$$

Moreover, (2.5) implies that, for every Borel set $B \subseteq \mathcal{X}_Y$ and any $p' \leqslant p$ in $\mathcal{P}_x$,

$$Q_{1,x}(B, p) - Q_{1,x}(B, p') \geqslant 0, \qquad Q_{0,x}(B, p') - Q_{0,x}(B, p) \geqslant 0.$$

Again, the extension from intervals to Borel sets follows by a monotone class argument.

We now construct the latent distribution explicitly. Extend $Q_{1,x}(B,\cdot)$ and $Q_{0,x}(B,\cdot)$ from $\mathcal{P}_x$ to $[0,1]$ as follows. For $u \in \mathcal{P}_x$, set

$$\bar{Q}_{1,x}(B,u) = Q_{1,x}(B,u), \qquad \bar{Q}_{0,x}(B,u) = Q_{0,x}(B,u).$$

At the endpoint $u = 0$, if $0 \in \mathcal{P}_x$, then $\bar{Q}_{1,x}(\cdot,0)$ and $\bar{Q}_{0,x}(\cdot,0)$ are already determined by the observed subdistributions. If $0 \notin \mathcal{P}_x$, let $\underline{p}_x = \inf \mathcal{P}_x$, set

$$\bar{Q}_{1,x}(\cdot,0) = 0,$$

and define

$$\bar{Q}_{0,x}(\cdot,0) = Q_{0,x}(\cdot,\underline{p}_x) + \underline{p}_x R_{0,x}(\cdot),$$

where $R_{0,x}$ is an arbitrary probability measure on $\mathcal{X}_Y$; this yields $\bar{Q}_{0,x}(\mathcal{X}_Y,0) = 1$.

At the endpoint $u = 1$, if $1 \in \mathcal{P}_x$, then $\bar{Q}_{1,x}(\cdot,1)$ and $\bar{Q}_{0,x}(\cdot,1)$ are already determined by the observed subdistributions. If $1 \notin \mathcal{P}_x$, let $\bar{p}_x = \sup \mathcal{P}_x$, set

$$\bar{Q}_{0,x}(\cdot,1) = 0,$$

and define

$$\bar{Q}_{1,x}(\cdot,1) = Q_{1,x}(\cdot,\bar{p}_x) + (1-\bar{p}_x)R_{1,x}(\cdot),$$

where $R_{1,x}$ is an arbitrary probability measure on $\mathcal{X}_Y$; this yields $\bar{Q}_{1,x}(\mathcal{X}_Y,1) = 1$.

Then, on each connected component (a,b) of $[0,1]\backslash(\mathcal{P}_x \cup \{0,1\})$, define for $u \in (a,b)$

$$\bar{Q}_{d,x}(B,u) = \frac{b-u}{b-a}\bar{Q}_{d,x}(B,a) + \frac{u-a}{b-a}\bar{Q}_{d,x}(B,b), \qquad d \in \{0,1\}.$$

By construction, $u \mapsto \bar{Q}_{1,x}(B,u)$ is nondecreasing and $u \mapsto \bar{Q}_{0,x}(B,u)$ is nonincreasing for every Borel set $B \subseteq \mathcal{X}_Y$, and

$$\bar{Q}_{1,x}(\mathcal{X}_Y,u) = u, \qquad \bar{Q}_{0,x}(\mathcal{X}_Y,u) = 1-u, \qquad u \in [0,1].$$

Define set functions on rectangles by $\mathcal{X}_Y \times [0,1]$ by

$$\mu_{1,x}(B \times (s,t]) = \bar{Q}_{1,x}(B,t) - \bar{Q}_{1,x}(B,s), \qquad 0 \leqslant s < t \leqslant 1,$$

and

$$\mu_{0,x}(B \times (s,t]) = \bar{Q}_{0,x}(B,s) - \bar{Q}_{0,x}(B,t), \qquad 0 \leqslant s < t \leqslant 1.$$

The monotonicity just established implies that the rectangle increments above are nonnegative. For fixed u, $\bar{Q}_{d,x}(\cdot,u)$ is countably additive in the outcome set B, while in the u-coordinate the increments telescope over adjacent intervals, i.e. $\mu_{d,x}(B \times (s,t]) = \mu_{d,x}(B \times (s,r]) + \mu_{d,x}(B \times (r,t])$ for $s < r < t$. Hence, these formulas define nonnegative finitely additive set functions on finite disjoint unions of rectangles $B \times (s,t]$. Moreover, countable additivity on this ring follows from the countable additivity of $\bar{Q}_{d,x}(\cdot,u)$ in B together with continuity from above of the monotone increment maps in the interval coordinate.

Their second marginals are Lebesgue measure. Indeed,

$$\mu_{1,x}(\mathcal{X}_Y \times (s,t]) = t - s, \qquad \mu_{0,x}(\mathcal{X}_Y \times (s,t]) = t - s,$$

and the intervals $(s,t]$ generate $\mathcal{B}([0,1])$. Thus the second marginal of each $\mu_{d,x}$ is du.

Since $\mathcal{X}_Y$ and $[0,1]$ are Borel spaces, Kallenberg's conditional distribution theorem (Kallenberg, 1997, Theorem 5.3) yields a well defined probability kernel $K_{d,x}(\cdot \mid u)$ such that

$$\mu_{d,x}(B \times A) = \int_A K_{d,x}(B \mid u)\, du, \qquad d \in \{0,1\}.$$

The kernel representation recovers the extended subdistributions. For every $p \in [0,1]$,

$$\bar{Q}_{1,x}(B,p) = \mu_{1,x}(B \times (0,p]) = \int_0^p K_{1,x}(B \mid u)\, du,$$

and

$$\bar{Q}_{0,x}(B,p) = \mu_{0,x}(B \times (p,1]) = \int_p^1 K_{0,x}(B \mid u)\, du.$$

We now construct a latent model that reproduces the observable distribution. Let $\tilde{U}_D$ satisfy

$$\Pr(\tilde{U}_D \leqslant u \mid X = x, Z_0 = z_0) = u, \qquad u \in [0,1],$$

so that $\tilde{U}_D \mid X = x \sim U[0,1]$ and $\tilde{U}_D \perp Z_0 \mid X = x$. Define the treatment rule

$$\tilde{D} = \mathbf{1}\{P(Z) \geqslant \tilde{U}_D\}.$$

Next, for each $d \in \{0,1\}$, define the marginal conditional law of $\tilde{Y}(d) \mid \tilde{U}_D = u, X = x, Z_0 = z_0$ by

$$\Pr(\tilde{Y}(d) \in B \mid \tilde{U}_D = u, X = x, Z_0 = z_0) = K_{d,x}(B \mid u), \qquad B \in \mathcal{B}(\mathcal{X}_Y).$$

Then, define the joint law of the two potential outcomes by choosing any measurable coupling of the two marginals $K_{1,x}(\cdot \mid u)$ and $K_{0,x}(\cdot \mid u)$; for example, take their product coupling. Since only one potential outcome is observed in each treatment state, any coupling yields the same observable distribution.

Since these conditional laws depend on (u,x) but not on z_0, we have

$$(\tilde{Y}(1), \tilde{Y}(0), \tilde{U}_D) \perp Z_0 \mid X = x.$$

Because the construction is carried out separately for each x, it follows that

$$(\tilde{Y}(1), \tilde{Y}(0), \tilde{U}_D) \perp Z_0 \mid X,$$

and Assumptions 1–3 are satisfied: exclusion holds because $\tilde{Y}(d)$ does not depend on z_0, random assignment holds by the display above, and monotonicity holds by the threshold rule $\tilde{D} = \mathbf{1}\{P(Z) \geqslant \tilde{U}_D\}$.

It remains to show observational equivalence. It is enough to consider $z = (x, z_0)$ in the conditional support of Z given $X = x$. Then $p = P(z) \in \mathcal{P}_x$, so the extended subdistributions coincide with the observed subdistributions at this value of p. For any Borel set $B \subseteq \mathcal{X}_Y$,

$$\begin{aligned}\Pr(\tilde{Y} \in B, \tilde{D} = 1 \mid Z = z) &= \Pr(\tilde{Y}(1) \in B, \tilde{U}_D \leqslant p \mid X = x, Z_0 = z_0) \\ &= \int_0^p K_{1,x}(B \mid u)\, du \\ &= \bar{Q}_{1,x}(B, p) \\ &= Q_{1,x}(B, p) \\ &= \Pr(Y \in B, D = 1 \mid P = p, X = x) \\ &= \Pr(Y \in B, D = 1 \mid Z = z),\end{aligned}$$

where the fourth equality uses $p \in \mathcal{P}_x$ and the last equality uses index sufficiency. Similarly,

$$\begin{aligned}\Pr(\tilde{Y} \in B, \tilde{D} = 0 \mid Z = z) &= \Pr(\tilde{Y}(0) \in B, \tilde{U}_D > p \mid X = x, Z_0 = z_0) \\ &= \int_p^1 K_{0,x}(B \mid u)\, du \\ &= \bar{Q}_{0,x}(B, p) \\ &= Q_{0,x}(B, p) \\ &= \Pr(Y \in B, D = 0 \mid P = p, X = x) \\ &= \Pr(Y \in B, D = 0 \mid Z = z).\end{aligned}$$

Thus, conditional on every $Z = z$, the joint distribution of $(\tilde{Y}, \tilde{D})$ coincides with that of (Y, D). Therefore $(\tilde{Y}, \tilde{D}, Z)$ and (Y, D, Z) have the same observable distribution. This proves the sharpness claim in part (iii). □

Proof of Proposition 2.2. We first prove parts (i) and (ii). Under Assumptions 1 and 4, write $Y(d) = X^\top \beta_d + V(d)$. Under Assumption 5, $(V(d), U_D) \perp Z$ for $d \in \{0, 1\}$. After the usual normalization, Assumption 3 gives $D = \mathbf{1}\{P(Z) \geqslant U_D\}$. For $d = 1$ and $p = P(z)$,

$$\Pr(V \leqslant v \mid Z = z, D = 1) = \Pr(V(1) \leqslant v \mid Z = z, U_D \leqslant p) = \Pr(V(1) \leqslant v \mid U_D \leqslant p),$$

where the last equality uses $(V(1), U_D) \perp Z$. The right-hand side depends on z only through p, and hence equals $\Pr(V \leqslant v \mid P = p, D = 1)$. The proof for $d = 0$ is identical with $U_D > p$. This proves (2.8). For any Borel set $B \subseteq \mathcal{X}_V$,

$$\Pr(V \in B, D = 1 \mid P = p) = \Pr(V(1) \in B, U_D \leqslant p) = \int_0^p \Pr(V(1) \in B \mid U_D = u)\, du,$$

which is nondecreasing in p, while

$$\Pr(V \in B, D = 0 \mid P = p) = \Pr(V(0) \in B, U_D > p) = \int_p^1 \Pr(V(0) \in B \mid U_D = u)\, du,$$

which is nonincreasing in p. Taking $B = [v_1, v_2]$ gives (2.9).

It remains to prove sharpness (iii). Let $\mathcal{P} = \operatorname{supp}\{P(Z)\} \subseteq [0,1]$. For $p \in \mathcal{P}$ and Borel $B \subseteq \mathcal{X}_V$, define

$$Q_1(B,p) = \Pr(V \in B, D = 1 \mid P = p), \qquad Q_0(B,p) = \Pr(V \in B, D = 0 \mid P = p).$$

Residual index sufficiency implies that, if $P(z) = p$, then $Q_d(B,p) = \Pr(V \in B, D = d \mid Z = z)$ for $d \in \{0,1\}$. Moreover,

$$Q_1(\mathcal{X}_V, p) = p, \qquad Q_0(\mathcal{X}_V, p) = 1 - p,$$

and the residual nesting inequalities imply that $Q_1(B,\cdot)$ is nondecreasing and $Q_0(B,\cdot)$ is nonincreasing on $\mathcal{P}$.

Extend Q_1 and Q_0 from $\mathcal{P}$ to $[0,1]$ by the same concrete linear interpolation construction used in the proof of Proposition 2.1, now with $\mathcal{X}_Y$ replaced by $\mathcal{X}_V$. Denote the extensions by $\bar{Q}_1(B,u)$ and $\bar{Q}_0(B,u)$, with $\bar{Q}_1(\mathcal{X}_V,u) = u$ and $\bar{Q}_0(\mathcal{X}_V,u) = 1-u$.

Exactly as in the proof of Proposition 2.1, there exist well defined probability kernels $K_1(\cdot \mid u)$ and $K_0(\cdot \mid u)$ such that

$$\mu_d(B \times A) = \int_A K_d(B \mid u)\, du, \qquad d \in \{0,1\}.$$

Therefore,

$$\bar{Q}_1(B,p) = \int_0^p K_1(B \mid u)\, du, \qquad \bar{Q}_0(B,p) = \int_p^1 K_0(B \mid u)\, du,$$

and hence, for every observable $p \in \mathcal{P}$,

$$Q_1(B,p) = \int_0^p K_1(B \mid u)\, du, \qquad Q_0(B,p) = \int_p^1 K_0(B \mid u)\, du.$$

Now let $\tilde{U}_D \sim U[0,1]$ be independent of Z and set $\tilde{D} = \mathbf{1}\{P(Z) \geqslant \tilde{U}_D\}$. Conditional on $\tilde{U}_D = u$, choose any measurable coupling of the two marginal kernels $K_1(\cdot \mid u)$ and $K_0(\cdot \mid u)$; for example, draw $\tilde{V}(1)$ and $\tilde{V}(0)$ independently conditional on u, with marginals $K_1(\cdot \mid u)$ and $K_0(\cdot \mid u)$. The conditional laws do not depend on Z, so $(\tilde{V}(1), \tilde{V}(0), \tilde{U}_D) \perp Z$.

For any z in the support of Z, let $p = P(z) \in \mathcal{P}$. Then, for every Borel $B \subseteq \mathcal{X}_V$,

$$\Pr(\tilde{V} \in B, \tilde{D} = 1 \mid Z = z) = \int_0^p K_1(B \mid u)\, du = \bar{Q}_1(B,p) = Q_1(B,p) = \Pr(V \in B, D = 1 \mid Z = z),$$

and similarly

$$\Pr(\tilde{V} \in B, \tilde{D} = 0 \mid Z = z) = \int_p^1 K_0(B \mid u)\, du = \bar{Q}_0(B,p) = Q_0(B,p) = \Pr(V \in B, D = 0 \mid Z = z).$$

Thus $(\tilde{V}, \tilde{D}, Z)$ and (V, D, Z) have the same distribution.

Finally define

$$\tilde{Y}(d, z_0) = X^\top \beta_d + \tilde{V}(d), \qquad \tilde{Y} = \tilde{D}\, \tilde{Y}(1, Z_0) + (1 - \tilde{D})\, \tilde{Y}(0, Z_0).$$

Then Assumptions 1, 3, 4, and 5 hold, and Assumption 2 follows from Assumption 5. Since conditioning on

$Z = z$ fixes $X = x$,

$$\Pr(\tilde{Y} \leqslant y, \tilde{D} = d \mid Z = z) = \Pr(\tilde{V} \leqslant y - x^\top \beta_d, \tilde{D} = d \mid Z = z) = \Pr(Y \leqslant y, D = d \mid Z = z),$$

so $(\tilde{Y}, \tilde{D}, Z)$ and (Y, D, Z) have the same distribution. □

D Proof of Theorem 3.1

Proof of Theorem 3.1. The proof of part (i) is based on the conditional multiplier central limit theorem, namely, Theorem 2.9.6 in Van der Vaart and Wellner (1996), together with the weak convergence result established above. For later use, define

$$\begin{aligned} \hat{F}_{V|P,D}(v|\hat{P}_i, d; \hat{\beta}_d, \hat{\alpha}) &= \frac{\sum_{j=1}^n \mathbf{1}(D_j = d)\mathbf{1}(\hat{V}_{dj} \leqslant v) K_h(\hat{P}_j - \hat{P}_i)}{\sum_{j=1}^n \mathbf{1}(D_j = d) K_h(\hat{P}_j - \hat{P}_i)}, \\ \hat{f}_{P,D}(\hat{P}_i, d) &= \frac{1}{n}\sum_{j=1}^n \mathbf{1}(D_j = d) K_h(\hat{P}_j - \hat{P}_i). \end{aligned} \tag{D.1}$$

Consider the infeasible multiplier process

$$\sqrt{n} U_d^\star(\omega) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i \zeta_d^I(Y_i, D_i, Z_i; \omega, \alpha, \beta_d). \tag{D.2}$$

By Theorem 2.9.6 in Van der Vaart and Wellner (1996), conditional on the original sample $\{(Y_i, D_i, Z_i)\}_{i=1}^n$,

$$\sqrt{n} U_d^\star(\omega) \underset{\xi}{\rightsquigarrow} \mathbb{G}_{U_d}(\omega) \quad \text{in } l^\infty(\mathscr{W}), \tag{D.3}$$

where $\underset{\xi}{\rightsquigarrow}$ means that

$$\sup_{f \in \mathrm{BL}_1} \left| E_\xi\big[f(\sqrt{n} U_d^\star)\big] - E\big[f(\mathbb{G}_{U_d})\big]\right| \to 0$$

and BL_1 denotes the class of all Lipschitz functions bounded by one.

It remains to show that the feasible bootstrap process $\hat{U}_d^\star(\omega)$ in (3.7) is asymptotically equivalent to $U_d^\star(\omega)$ in (D.2). As both $\hat{\zeta}_d^I(Y_i, D_i, Z_i; \omega, \tilde{\alpha}, \tilde{\beta}_d)$ and $\zeta_d^I(Y_i, D_i, Z_i; \omega, \tilde{\alpha}, \tilde{\beta}_d)$ are Donsker classes by Lemma E.1, it suffices to show that

$$\begin{aligned} \sup_{\omega \in \mathscr{W}} \sqrt{n} \left| \hat{U}_d^\star(\omega; \alpha, \beta_d) - U_d^\star(\omega; \alpha, \beta_d) \right| &= o_p(1), \\ \sup_{\omega \in \mathscr{W}} \sqrt{n} \left| U_d^\star(\omega; \hat{\alpha}, \hat{\beta}_d) - U_d^\star(\omega; \alpha, \beta_d) \right| &= o_p(1). \end{aligned} \tag{D.4}$$

To show the first line above, we decompose the difference $\sqrt{n}(\hat{U}_d^\star(\omega; \alpha, \beta_d) - U_d^\star(\omega; \alpha, \beta_d))$ into three terms that capture the nonparametric kernel estimation error and a correction term that captures the parametric first step estimation effect. Specifically, we keep the parameters fixed at their true values and suppress (α, β_d) from the notation in the subsequent decomposition.

$$\sqrt{n}(\hat{U}_d^\star(\omega; \alpha, \beta_d) - U_d^\star(\omega; \alpha, \beta_d))$$

$$
\begin{aligned}
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\left[\hat{\zeta}_d^I(Y_i,D_i,Z_i;\omega,\alpha,\beta_d)-\zeta_d^I(Y_i,D_i,Z_i;\omega,\alpha,\beta_d)\right]\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\left[\mathbf{1}(V_{di}\leqslant v)-\hat{F}_{V|P,D}(v|P_i,d)\right]\hat{\psi}_i^{\perp}(p,z,d)\\
-&\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\left[\mathbf{1}(V_{di}\leqslant v)-F_{V|P,D}(v|P_i,d)\right]\psi_i^{\perp}(p,z,d)\\
&\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\left[\hat{A}_d^\top l(D_i,Z_i;\alpha)+\hat{B}_d^\top l_d(Y_i,Z_i;\beta_d,\alpha)\right]-\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\left[A_d^\top l(D_i,Z_i;\alpha)+B_d^\top l_d(Y_i,Z_i;\beta_d,\alpha)\right]\\
=&-B_{1n}(\omega)-B_{2n}(\omega)-B_{3n}(\omega)+C_n(\omega),
\end{aligned}
\tag{D.5}
$$

where

$$
B_{1n}(\omega)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\big(\mathbf{1}(V_i\leqslant v)-F_{V|P,D}(v|P_i,d)\big)\big(\hat{\iota}_d(P_i)-\iota_d(P_i)\big)\mathbf{1}(D_i=d),\tag{D.6}
$$

$$
B_{2n}(\omega)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\big(\hat{F}_{V|P,D}(v|P_i,d)-F_{V|P,D}(v|P_i,d)\big)\psi_i^{\perp}(p,z,d),\tag{D.7}
$$

$$
B_{3n}(\omega)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\big(\hat{F}_{V|P,D}(v|P_i,d)-F_{V|P,D}(v|P_i,d)\big)\big(\hat{\iota}_d(P_i)-\iota_d(P_i)\big)\mathbf{1}(D_i=d).\tag{D.8}
$$

$$
C_n(\omega)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\left[(\hat{A}_d^\top-A_d^\top)l(D_i,Z_i;\alpha)+(\hat{B}_d^\top-B_d^\top)l_d(Y_i,Z_i;\beta_d,\alpha)\right].\tag{D.9}
$$

where $\iota_d(P_i)=E[\psi_i(p,z,d)\mid P=P_i,D_i=d]$ and let $\hat{\iota}_d(P_i)$ denote its NW estimator:

$$
\hat{\iota}_d(P_i)=\frac{\sum_{j=1}^n K_h(P_i-P_j)\psi_j(p,z,d)}{\sum_{j=1}^n K_h(P_i-P_j)\mathbf{1}(D_j=d)}
$$

As both processes $\sqrt{n}U_d(\omega;\tilde{\alpha},\tilde{\beta}_d)$ and $\sqrt{n}\hat{U}_d(\omega;\tilde{\alpha},\tilde{\beta}_d)$ form Donsker classes by Lemma E.1, it suffices to show that the conditional variances of $B_{1n}(\omega)$, $B_{2n}(\omega)$, and $B_{3n}(\omega)$ given the original sample are all $o_p(1)$. Define

$$
\Delta_{Fi}(\omega):=\hat{F}_{V|P,D}(v\mid P_i,d)-F_{V|P,D}(v\mid P_i,d),\qquad \Delta_{\iota i}(\omega):=\hat{\iota}_d(P_i)-\iota_d(P_i).
$$

By Assumptions A.2–A.4, both $\Delta_{Fi}(\omega)$ and $\Delta_{\iota i}(\omega)$ converge to zero in L^2 norm and are uniformly bounded in absolute value by a constant (say, two).

The conditional variance of $B_{1n}(\omega)$, $B_{2n}(\omega)$ and $B_{3n}(\omega)$ on the original sample can be written as follows:

$$
\begin{aligned}
\mathrm{Var}_\xi\big(B_{1n}(\omega)\mid\mathcal{Z}_n\big)&=\frac{1}{n}\sum_{i=1}^n(\mathbf{1}(V_i\leqslant v)-F_{V|P,D}(v|P_i,d))^2\Delta_{\iota i}(\omega)^2\mathbf{1}(D_i=d)\\
\mathrm{Var}_\xi\big(B_{2n}(\omega)\mid\mathcal{Z}_n\big)&=\frac{1}{n}\sum_{i=1}^n\psi_i^{\perp}(p,z,d)^2\Delta_{Fi}(\omega)^2\mathbf{1}(D_i=d)\\
\mathrm{Var}_\xi\big(B_{3n}(\omega)\mid\mathcal{Z}_n\big)&=\frac{1}{n}\sum_{i=1}^n\Delta_{Fi}(\omega)^2\Delta_{\iota i}(\omega)^2\mathbf{1}(D_i=d).
\end{aligned}
\tag{D.10}
$$

Notice that $(\mathbf{1}(V_i\leqslant v)-F_{V|P,D}(v|P_i,d))^2$ and $\psi_i^{\perp}(p,z,d)^2$ are uniformly bounded by a constant one.

Therefore,

$$
\begin{aligned}
&\sup_{\omega\in\mathscr{W}} \operatorname{Var}_\xi\big(B_{1n}(\omega)\mid \mathcal{Z}_n\big) \leqslant C \sup_{\omega\in\mathscr{W}} \frac{1}{n}\sum_{i=1}^n \Delta_{\iota i}(\omega)^2 = o_p(1).\\
&\sup_{\omega\in\mathscr{W}} \operatorname{Var}_\xi\big(B_{2n}(\omega)\mid \mathcal{Z}_n\big) \leqslant C \sup_{\omega\in\mathscr{W}} \frac{1}{n}\sum_{i=1}^n \Delta_{Fi}(\omega)^2 = o_p(1).\\
&\sup_{\omega\in\mathscr{W}} \operatorname{Var}_\xi\big(B_{3n}(\omega)\mid \mathcal{Z}_n\big) \leqslant C \sup_{\omega\in\mathscr{W}} \frac{1}{n}\sum_{i=1}^n \Delta_{\iota i}(\omega)^2 = o_p(1).
\end{aligned}
\tag{D.11}
$$

Finally, the correction term $C_n(\omega)$ in (D.9) is $o_p(1)$ uniformly in ω once $\hat{A}_d$ and $\hat{B}_d$ are uniformly consistent for A_d and B_d, respectively. Let r denote the size of a local joint perturbation of (α, β_d). Inspecting (B.4), the numerical coefficients are finite-difference approximations to derivatives of the conditional CDF $F_{V|P,D}$ and of the joint distribution $F_{V,P,Z,D}$. The relevant local stability condition is

$$
\sup_{p\in\mathcal{X}_P,\, v\in\mathcal{X}_V} \sup_{\|(\tilde{\alpha},\tilde{\beta}_d)-(\alpha,\beta_d)\|\leqslant r} \Big| F_{V|P,D}(v\mid p,d;\tilde{\beta}_d,\tilde{\alpha}) - F_{V|P,D}(v\mid p,d;\beta_d,\alpha)\Big| \leqslant Cr.
$$

Both the β_d-perturbation and α-perturbation are controlled by the continuous differentiability of $F_{V|P,D}$ imposed in Assumption A.6(ii).

For the joint distribution term, use the representation

$$
F_{V,P,Z,D}(v,p,z,d;\beta_d,\alpha) = \int_{\mathcal{X}_P\cap(-\infty,p]}\int_{\mathcal{X}_Z\cap(-\infty,z]} F_{V|P,D}(v\mid \bar{p},d;\beta_d,\alpha)\, f_{Z|P,D}(\bar{z}\mid \bar{p},d;\alpha) f_{P,D}(\bar{p},d;\alpha)\, d\bar{z}\, d\bar{p}.
$$

The preceding local stability bound, together with the bounded and smooth density terms in Assumptions A.2 and A.6, provides an integrable envelope, so dominated convergence justifies passing the local derivatives through the integral. Hence the population finite differences converge uniformly to the derivatives defining A_d and B_d. For the sample analogues, uniform consistency of $\hat{F}_{V|P,D}$ and of the empirical (smoothed) joint distribution estimator implies that each perturbed distributional object is uniformly close to its population counterpart over (v,p,z,α,β_d). Moreover, the associated empirical and kernel processes are stochastically equicontinuous in (α,β_d) because of the boundary mass control by Assumption A.6. Therefore the sample finite differences are uniformly close to the corresponding population finite differences.

Combining the bounds for $B_{1n}(\omega)$, $B_{2n}(\omega)$, $B_{3n}(\omega)$, and $C_n(\omega)$ yields

$$
\sup_{\omega\in\mathscr{W}} \sqrt{n}\Big|\hat{U}_d^\star(\omega;\alpha,\beta_d) - U_d^\star(\omega;\alpha,\beta_d)\Big| = o_p(1). \tag{D.12}
$$

Next, since the class $\{\zeta_d^I(Y_i,D_i,Z_i;\omega,\alpha,\beta_d) : \omega\in\mathscr{W}, (\alpha,\beta_d)\in C_\alpha\times C_{\beta_d}\}$ is Donsker by Lemma E.1, the associated multiplier empirical process is asymptotically equicontinuous in (α,β_d). Together with the consistency of $(\hat{\alpha},\hat{\beta}_d)$, this implies

$$
\sup_{\omega\in\mathscr{W}} \sqrt{n}\Big|U_d^\star(\omega;\hat{\alpha},\hat{\beta}_d) - U_d^\star(\omega;\alpha,\beta_d)\Big| = o_p(1). \tag{D.13}
$$

Therefore,

$$\sup_{\omega \in \mathscr{W}} \sqrt{n} \left| \hat{U}_d^{\star}(\omega) - U_d^{\star}(\omega) \right| = o_p(1). \tag{D.14}$$

It follows that the feasible multiplier process $\sqrt{n}\hat{U}_d^{\star}$ and the infeasible multiplier process $\sqrt{n}U_d^{\star}$ have the same conditional weak limit. Hence,

$$\sqrt{n}\hat{U}_d^{\star}(\omega) \underset{\xi}{\rightsquigarrow} \mathbb{G}_{U_d}(\omega) \quad \text{in probability.} \tag{D.15}$$

This completes the proof of part (1).

The proof of part (ii) consists of two main steps: (1) establishing the bootstrap consistency for the copula process, i.e., for each $d \in \{0, 1\}$,

$$\hat{\mathbb{G}}_{C_d}^{\star}(\lambda) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \hat{\zeta}_d^M(Y_i, D_i, Z_i; \lambda, \hat{\alpha}, \hat{\beta}_d) \underset{\xi}{\rightsquigarrow} \mathbb{G}_{C_d}(\lambda) \quad \text{in probability in } \ell^{\infty}(\Lambda). \tag{D.16}$$

and (2) verifying the consistency of the numerical LCM operator, so that the numerical delta bootstrap method for directionally differentiable maps can be applied to the LCM gap operator $\mathcal{T}_d$.

Step 1: Bootstrap consistency for the copula process.

Lemma E.2 establishes two facts for each $d \in \{0, 1\}$: (i) the class $\{\zeta_d^M(\cdot; \lambda) : \lambda \in \Lambda\}$ is P-Donsker, and (ii) the uniform asymptotic linear representation $\sqrt{n}(\hat{C}_d - C_d)(\lambda) = n^{-1/2} \sum_{i=1}^{n} \zeta_d^M(Y_i, D_i, Z_i; \lambda, \alpha, \beta_d) + o_p(1)$ holds in $\ell^{\infty}(\Lambda)$. By the multiplier central limit theorem (Theorem 2.9.6 of Van der Vaart and Wellner (1996)), the infeasible multiplier process therefore converges conditionally:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \zeta_d^M(Y_i, D_i, Z_i; \lambda, \alpha, \beta_d) \underset{\xi}{\rightsquigarrow} \mathbb{G}_{C_d}(\lambda) \quad \text{in probability in } \ell^{\infty}(\Lambda).$$

To replace ζ_d^M by the feasible version $\hat{\zeta}_d^M$, it suffices to establish the $L_2(P_n)$ plug-in consistency

$$\sup_{\lambda \in \Lambda} P_n \{\hat{\zeta}_d^M(Y, D, Z; \lambda, \hat{\alpha}, \hat{\beta}_d) - \zeta_d^M(Y, D, Z; \lambda, \alpha, \beta_d)\}^2 = o_p(1), \tag{D.17}$$

since the multiplier maximal inequality applied conditionally on the data then gives $\sup_{\lambda} |\hat{\mathbb{G}}_{C_d}^{\star}(\lambda) - n^{-1/2} \sum \xi_i \zeta_d^M(\lambda)| = o_p(1)$.

To verify (D.17), first separate the influence function into a nonparametric copula component and a parametric nuisance component:

$$\begin{aligned} \zeta_{d0}^M(Y_i, D_i, Z_i; \lambda, \alpha, \beta_d) &= \mathbf{1}(v_1 \leqslant V_{di} \leqslant v_2, D_i = d, F_P(P_i) \leqslant u) - C_d(\lambda) \\ &\quad - \partial_u C_d(\lambda)\{\mathbf{1}(F_P(P_i) \leqslant u) - u\}, \\ \zeta_{d1}^M(Y_i, D_i, Z_i; \lambda, \alpha, \beta_d) &= l_d^{\top}(Y_i, Z_i; \beta_d, \alpha)\, \partial_{\beta_d} C_d(\lambda) + l^{\top}(D_i, Z_i; \alpha)\, \partial_{\alpha} C_d(\lambda). \end{aligned} \tag{D.18}$$

Consider first ζ_{d0}^M. Let $\tilde{\zeta}_{d0}^M$ denote the intermediate quantity obtained by replacing the nonparametric ingredients (V_{di}, P_i, C_d, F_P) with their estimates $(\hat{V}_{di}, \hat{P}_i, \hat{C}_d, \hat{F}_P)$, while keeping the derivative coefficient

$\partial_u C_d(\lambda)$ at its population value. The same argument used in the proof of Lemma E.2 implies

$$\sup_{\lambda\in\Lambda} P_n\{\tilde{\zeta}_{d0}^M(\lambda) - \zeta_{d0}^M(\lambda)\}^2 = o_p(1).$$

The fully feasible version $\hat{\zeta}_{d0}^M$ differs from $\tilde{\zeta}_{d0}^M$ only by replacing $\partial_u C_d(\lambda)$ with the numerical estimator $\partial_u \hat{C}_d(\lambda)$. Hence

$$\hat{\zeta}_{d0}^M(\lambda) - \tilde{\zeta}_{d0}^M(\lambda) = \{\partial_u C_d(\lambda) - \partial_u \hat{C}_d(\lambda)\}\{\mathbf{1}(\hat{F}_P(\hat{P}_i) \leqslant u) - u\},$$

whose $L_2(P_n)$ norm is bounded by $|\partial_u \hat{C}_d(\lambda) - \partial_u C_d(\lambda)|$.

For ζ_{d1}^M, the feasible version $\hat{\zeta}_{d1}^M$ replaces (α, β_d) by $(\hat{\alpha}, \hat{\beta}_d)$, the derivatives $\partial_{\beta_d} C_d$ and $\partial_\alpha C_d$ by their numerical perturbation estimators, and the influence functions l, l_d by $\hat{l}, \hat{l}_d$. The $L_2(P_n)$ consistency of $\hat{l}$ and $\hat{l}_d$ is guaranteed by Assumption A.1(iv), and the population derivatives $\partial_\alpha C_d$ and $\partial_{\beta_d} C_d$ are uniformly bounded over Λ by Assumption A.6. Hence the $L_2(P_n)$-consistency of $\hat{\zeta}_d^M$ for ζ_d^M follows once we establish

$$\begin{aligned} \sup_{\lambda\in\Lambda} |\partial_u \hat{C}_d(\lambda) - \partial_u C_d(\lambda)| &= o_p(1), \\ \sup_{\lambda\in\Lambda} \|\partial_{\beta_d} \hat{C}_d(\lambda) - \partial_{\beta_d} C_d(\lambda)\| &= o_p(1), \\ \sup_{\lambda\in\Lambda} \|\partial_\alpha \hat{C}_d(\lambda) - \partial_\alpha C_d(\lambda)\| &= o_p(1). \end{aligned} \tag{D.19}$$

All three statements in (D.19) are therefore the only remaining verification; they are proved by the same numerical perturbation argument.

(a) Uniform consistency of $\partial_u \hat{C}_d$. Write $\lambda = (v_1, v_2, u)$ and estimate $\partial_u \hat{C}_d$ by the symmetric difference quotient

$$\partial_u \hat{C}_d(\lambda) = \frac{\sqrt{n}}{2}\Big[\hat{C}_d(v_1, v_2, u + n^{-1/2}) - \hat{C}_d(v_1, v_2, u - n^{-1/2})\Big],$$

with the obvious one-sided modification at the boundary. Then

$$\sup_{\lambda\in\Lambda} \Big|\partial_u \hat{C}_d(\lambda) - \partial_u C_d(\lambda)\Big| \leqslant A_n + B_n,$$

where

$$A_n = \sup_{\lambda\in\Lambda} \left|\frac{\sqrt{n}}{2}\Big[C_d(v_1, v_2, u + n^{-1/2}) - C_d(v_1, v_2, u - n^{-1/2})\Big] - \partial_u C_d(\lambda)\right|,$$

and

$$B_n = \frac{1}{2}\sup_{\lambda\in\Lambda} \Big|\sqrt{n}(\hat{C}_d - C_d)(v_1, v_2, u + n^{-1/2}) - \sqrt{n}(\hat{C}_d - C_d)(v_1, v_2, u - n^{-1/2})\Big|.$$

Under Assumptions A.2 and A.6, the copula function $C_d(\lambda)$ is continuously differentiable in u, with a uniformly continuous derivative $\partial_u C_d(\lambda)$ over Λ; hence $A_n = o(1)$ by the mean-value theorem and the uniform continuity of the derivative. Lemma E.2 implies that the empirical copula process $\sqrt{n}(\hat{C}_d - C_d)(\lambda)$ is asymptotically equicontinuous on Λ, so $B_n = o_p(1)$. Therefore the first line of (D.19) holds.

(b) Uniform consistency of $\partial_{\beta_d} \hat{C}_d$ and $\partial_\alpha \hat{C}_d$. We give the argument for $\partial_{\beta_d} \hat{C}_d$; the proof for $\partial_\alpha \hat{C}_d$ is analogous, with the additional ingredient that α-perturbations affect both the propensity score threshold and the rank transform, whose stability is guaranteed by the quantile boundary-mass control implied by Assumption A.6.

For the k-th component of β_d, set $a_n = n^{-1/2}$ and estimate the derivative by the symmetric difference quotient

$$\partial_{\beta_{d,k}} \hat{C}_d(\lambda) = \frac{1}{2a_n}\Big[\hat{C}_d(\lambda; \hat{\alpha}, \hat{\beta}_d + a_n e_k) - \hat{C}_d(\lambda; \hat{\alpha}, \hat{\beta}_d - a_n e_k)\Big].$$

Define the population counterpart centered at the estimated parameter values:

$$\partial_{\beta_{d,k}} \tilde{C}_d(\lambda) = \frac{1}{2a_n}\Big[C_d(\lambda; \hat{\alpha}, \hat{\beta}_d + a_n e_k) - C_d(\lambda; \hat{\alpha}, \hat{\beta}_d - a_n e_k)\Big].$$

Then

$$\begin{aligned}
&\big|\partial_{\beta_{d,k}} \hat{C}_d(\lambda) - \partial_{\beta_{d,k}} C_d(\lambda; \alpha, \beta_d)\big| \\
&\leqslant \big|\partial_{\beta_{d,k}} \hat{C}_d(\lambda) - \partial_{\beta_{d,k}} \tilde{C}_d(\lambda)\big| + \big|\partial_{\beta_{d,k}} \tilde{C}_d(\lambda) - \partial_{\beta_{d,k}} C_d(\lambda; \hat{\alpha}, \hat{\beta}_d)\big| \\
&\quad + \big|\partial_{\beta_{d,k}} C_d(\lambda; \hat{\alpha}, \hat{\beta}_d) - \partial_{\beta_{d,k}} C_d(\lambda; \alpha, \beta_d)\big|.
\end{aligned}$$

The first term is a stochastic equicontinuity term $B_{n,k}$: it involves the difference of the empirical copula process $\sqrt{n}(\hat{C}_d - C_d)$ evaluated at perturbed residuals, and is $o_p(1)$ uniformly in λ by the asymptotic equicontinuity established in Lemma E.2.

The second term is a deterministic bias term $A_{n,k}$. Perturbing $\hat{\beta}_d$ by $\pm a_n e_k$ shifts the residual cutoffs by $\pm a_n X_{ik}$. Because $C_d(\lambda; a, b)$ is continuously differentiable in b with uniformly continuous and bounded derivatives (Assumptions A.2 and A.6) and the perturbation step $a_n \to 0$, the mean-value theorem gives $\sup_\lambda |A_{n,k}| = o(1)$.

The third term is $o_p(1)$ uniformly in λ by the consistency of $(\hat{\alpha}, \hat{\beta}_d)$ and the uniform continuity of $\partial_{\beta_d} C_d(\lambda; \tilde{\alpha}, \tilde{\beta}_d)$ in its parameter arguments. Hence

$$\sup_{\lambda \in \Lambda} \|\partial_{\beta_d} \hat{C}_d(\lambda) - \partial_{\beta_d} C_d(\lambda)\| = o_p(1),$$

and the same argument, incorporating the quantile boundary mass control, yields the analogous statement for $\partial_\alpha \hat{C}_d$. This verifies the last two lines of (D.19).

With (D.19) established, the discussion above shows that $\hat{\zeta}_d^M$ is $L_2(P_n)$-consistent for its population counterpart. Hence the full plug-in consistency (D.17) holds, and the conditional weak convergence (D.16) follows.

Step 2: Consistency of the numerical LCM operator

Lemma E.2 gives $\sqrt{n}(\hat{C}_d - C_d) \rightsquigarrow \mathbb{G}_{C_d}$ in $\ell^\infty(\Lambda)$. Under H_0^M, $\mathcal{T}_d C_d = 0$. The partial LCM gap operators $\mathcal{T}_d$ are Hadamard directionally differentiable tangentially to $C(\Lambda)$; see Shapiro (1990) and Fang and Santos (2018). Therefore the extended delta method for directionally differentiable maps (Proposition 2.1 of Fang and Santos (2018)) gives the limit distribution of $\hat{T}_M$ stated in the theorem.

For the bootstrap validity, we apply Theorem 3.2 of Fang and Santos (2018), which requires four conditions: (i) the domain and range are Banach spaces (here $\ell^\infty(\Lambda)$ and $\mathbb{R}$); (ii) the original estimator satisfies a weak convergence statement (Lemma E.2); (iii) the bootstrap process consistently estimates the law of the limit (established in Step 1 above); and (iv) the numerical derivative estimator $\hat{\dot{\mathcal{T}}}_d$ in (3.9) consistently estimates the Hadamard directional derivative $\dot{\mathcal{T}}_d$ of $\mathcal{T}_d$ such that, for each $h \in C(\Lambda)$,

$$\dot{\mathcal{T}}_0(h) = \tilde{\theta}_{C_0}(h) - h, \qquad \dot{\mathcal{T}}_1(h) = \tilde{\theta}_{-C_1}(-h) + h.$$

Then, condition (iv) is verified by the Lemma S.3.8 of Fang and Santos (2018) as the tuning parameter κ in (3.9) satisfies $\kappa \to 0$ and $\sqrt{n}\kappa \to \infty$.

When the contact set is empty, $\tilde{\theta}_{C_0}$ and $\tilde{\theta}_{-C_1}$ reduce to the identity map, so that $\tilde{\theta}_{C_0}(h) = h$ and $\tilde{\theta}_{-C_1}(-h) = -h$ for every $h \in C(\Lambda)$. Hence $\dot{\mathcal{T}}_0(h) = \tilde{\theta}_{C_0}(h) - h = 0$ and $\dot{\mathcal{T}}_1(h) = \tilde{\theta}_{-C_1}(-h) + h = 0$, and the delta method gives $\hat{T}_M \to_p 0$. For any fixed $\eta > 0$, the regularized bootstrap critical value $\check{c}^{\star}_{M,\alpha}$ is asymptotically bounded below by η, so $\Pr(\hat{T}_M \geqslant \check{c}^{\star}_{M,\alpha}) \to 0$.

(3) The conclusion of Theorem 3.1(iii) is the direct result of parts (1) and (2). □

E Lemmas and Intermediary Results

E.1 Asymptotic Null Distribution of Index Statistics

This subsection derives the asymptotic null distribution of $\hat{T}_I$, which depends on the influence function of the empirical process $\hat{U}_d(\omega)$. To characterize the influence function ζ^I_d, we introduce the following empirical process for a generic parameter value $(\tilde{\alpha}, \tilde{\beta}_d) \in C_\alpha \times C_{\beta_d}$:

$$\hat{U}_d(\omega; \tilde{\alpha}, \tilde{\beta}_d) = \frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{1}(\tilde{V}_{di} \leqslant v) - \hat{F}_{V|P,D}(v|\tilde{P}_i, d; \tilde{\beta}_d, \tilde{\alpha})\right]\psi_i(p, z, d; \tilde{\alpha}), \tag{E.1}$$

where $\tilde{V}_{di} = Y_i - X_i^\top \tilde{\beta}_d$, $\tilde{P}_i = P(Z_i, \tilde{\alpha})$, and $\psi_i(p, z, d; \tilde{\alpha}) = \mathbf{1}(\tilde{P}_i \leqslant p, Z_i \leqslant z, D_i = d)$.

The next lemma formalizes the application of EJL14 to our process $\hat{U}_d(\omega)$ in (B.1).

Lemma E.1. *Suppose H^I_0 in (3.2) holds, and let Assumptions A.1–A.5 be satisfied. Then, for each $d \in \{0, 1\}$, the following statements hold.*
(i) ***Uniform approximation.*** *We have*

$$\sup_{\omega \in \mathscr{W}} \sup_{(\tilde{\alpha}, \tilde{\beta}_d) \in C_\alpha \times C_{\beta_d}} \sqrt{n}|\hat{U}_d(\omega; \tilde{\alpha}, \tilde{\beta}_d) - U^{\perp}_d(\omega; \tilde{\alpha}, \tilde{\beta}_d)| = o_p(1). \tag{E.2}$$

where the weighted empirical process $U^{\perp}_d(\omega; \tilde{\alpha}, \tilde{\beta}_d)$ is expressed as

$$U^{\perp}_d(\omega; \tilde{\alpha}, \tilde{\beta}_d) = \frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{1}(\tilde{V}_{di} \leqslant v) - F_{V|P,D}(v|\tilde{P}_i, d; \tilde{\beta}_d, \tilde{\alpha})\right]\psi^{\perp}_i(p, z, d, \tilde{\alpha}),$$

(ii) ***Asymptotic linear representation.*** *The process $\sqrt{n}\hat{U}_d(\omega)$ admits the uniform expansion*

$$\sup_{\omega \in \mathscr{W}} \sqrt{n}\left|\hat{U}_d(\omega) - \frac{1}{n}\sum_{i=1}^{n}\zeta^I_d(Y_i, D_i, Z_i; \omega, \alpha, \beta_d)\right| = o_p(1), \tag{E.3}$$

and hence

$$\sqrt{n}\hat{U}_d(\omega) \rightsquigarrow \mathbb{G}_{U_d}(\omega) \text{ in } l^\infty(\mathscr{W}), \tag{E.4}$$

where $\mathbb{G}_{U_d}(\omega)$ *is a mean-zero Gaussian process with covariance function*

$$E\left[\zeta_d^I(Y_i, D_i, Z_i; \omega_1, \alpha, \beta_d)\zeta_d^I(Y_i, D_i, Z_i; \omega_2, \alpha, \beta_d)\right].$$

Consequently,

$$\hat{T}_I \rightsquigarrow \max_{d\in\{0,1\}} \|\mathbb{G}_{U_d}(\omega)\|_\infty. \tag{E.5}$$

Proof of Lemma E.1. The proof follows from Theorems 3.1 and 3.2 of Escanciano et al. (2014), which imply parts (i) and (ii) of Lemma E.1, respectively. The only difference from their setting is that our empirical process $U_d(\omega; \tilde{\alpha}, \tilde{\beta}_d)$ contains parametric generated regressors, while no trimming indicator is needed here.

(1) For $(\tilde{\alpha}, \tilde{\beta}_d) \in C_\alpha \times C_{\beta_d}$, we first rewrite $\hat{U}_d(\omega; \tilde{\alpha}, \tilde{\beta}_d)$ by adding and subtracting the population counterpart $F_{V|P,D}(v|\tilde{P}_i, d; \tilde{\beta}_d, \tilde{\alpha})$. This algebraic decomposition separates the empirical process term from the kernel estimation term. Theorem 2.11.9 of Van der Vaart and Wellner (1996) will then be used to verify stochastic equicontinuity of the former.

To verify the stochastic equicontinuity conditions, define

$$Z_{ni}(\upsilon) = \frac{1}{\sqrt{n}}\big[g_i(v, \tilde{\beta}_d) - q_d(P(Z_i, \tilde{\alpha}) \mid v, \tilde{\beta}_d, \tilde{\alpha})\big]\psi_i(p, z, d; \tilde{\alpha}),$$

where $\upsilon = (g, q_d, \psi) \in \Upsilon = \mathcal{I}_d \times \mathcal{F}^\eta \times \Psi$, with

$$\mathcal{I}_d = \left\{g_i(v, \tilde{\beta}_d) = \mathbf{1}(Y_i - X_i^\top \tilde{\beta}_d \leqslant v) : (v, \tilde{\beta}_d) \in \mathcal{X}_V \times C_{\beta_d}\right\},$$
$$\Psi = \left\{\psi_i(p, z, d; \tilde{\alpha}) = \mathbf{1}(\tilde{P}_i \leqslant p, Z_i \leqslant z, D_i = d) : (p, z, \tilde{\alpha}) \in \mathcal{X}_P \times \mathcal{X}_Z \times C_\alpha\right\}.$$

Here $\mathcal{F}^\eta$ is exactly the function class introduced in Assumption A.5; we do not restate its definition. In particular, for each $(v, \tilde{\beta}_d, \tilde{\alpha})$, the map $z \mapsto q_d(P(z, \tilde{\alpha}) \mid v, \tilde{\beta}_d, \tilde{\alpha})$ belongs to $\mathcal{F}^\eta$, and its estimator $\hat{q}_d$ also belongs to $\mathcal{F}^\eta$ with probability approaching one by Assumption A.5.

With this convention,

$$\upsilon = (g, q_d, \psi), \qquad q_d(\cdot) := F_{V|P,D}\big(v \mid \cdot, d; \tilde{\beta}_d, \tilde{\alpha}\big),$$

corresponds to the population conditional CDF along the composite map, and

$$\hat{\upsilon} = (g, \hat{q}_d, \psi), \qquad \hat{q}_d(\cdot) := \hat{F}_{V|P,D}\big(v \mid \cdot, d; \tilde{\beta}_d, \tilde{\alpha}\big),$$

to the kernel estimator.

Let $\upsilon_1 = (g_1, q_{d1}, \psi_1)$ and $\upsilon_2 = (g_2, q_{d2}, \psi_2)$ and define $\rho(\upsilon_1, \upsilon_2)$ be the semimetric

$$\rho^2(\upsilon_1, \upsilon_2) := \|g_1 - g_2\|_2^2 + \|q_{d1} - q_{d2}\|_\infty^2 + \|\psi_1 - \psi_2\|_2^2$$

To apply Theorem 2.11.9, we verify the following conditions.

(a) Envelope and Lindeberg condition. Since $\sqrt{n}Z_{ni}(\upsilon)$ is bounded by a constant one as both the indicator

and the weighting function are bounded by one, we have

$$\sum_{i=1}^{n} E[\sup_{\upsilon} |Z_{ni}(\upsilon)| \mathbf{1}(\sup_{\upsilon} |Z_{ni}(\upsilon)| \geqslant \delta)] \leqslant 0$$

This yields the required Lindeberg-type condition for the triangular array.

(b) Local oscillation bound. For a fixed $\upsilon_1 = (g_1, q_{d1}, \psi_1)$ and $\upsilon_2 = (g_2, q_{d2}, \psi_2)$ such that $\rho(\upsilon_1, \upsilon_2) \leqslant \delta$, the q_d component is $O(\delta)$ by the Lipschitz property in Assumption A.5. For the indicator components, perturbations in $\tilde{\beta}_d$ or $(p, \tilde{\alpha})$ change g_i or ψ_i only when an observation lies within an $O(\delta)$ neighborhood of the relevant threshold; the boundary-mass bounds implied by Assumption A.6 therefore yield the same L_2 order. Hence,

$$\sum_{i=1}^{n} E\Big[\sup_{\rho(\upsilon_1, \upsilon_2) \leqslant \delta} |Z_{ni}(\upsilon_1) - Z_{ni}(\upsilon_2)|^2 \Big] \leqslant C\delta^2.$$

This is the asymptotic equicontinuity modulus required by the theorem.

(c) Entropy bound. It remains to verify that $N_{[]}(\epsilon, \Upsilon, \|.\|_2)$ is bounded by three components: the indicator class in $\mathcal{I}_d$, the smooth class $\mathcal{F}^{\eta}$, and the class of weighting functions Ψ indexed by $(p, z, \tilde{\alpha})$ as following:

$$N_{[]}(\epsilon, \Upsilon, \rho) \leqslant N_{[]}(\epsilon, \mathcal{I}_d, \|.\|_2) \times N_{[]}(\epsilon C, \mathcal{F}^{\eta}, \|.\|_{\infty}) \times N_{[]}(\epsilon, \Psi, \|.\|_2)$$

where $\mathcal{I}_d$ and Ψ are the indicator classes defined above. Each class is obtained by composing a threshold indicator function with a finite-dimensional parametric index: $Y_i - X_i^{\top} \tilde{\beta}_d$ for $\mathcal{I}_d$ and $P(Z_i, \tilde{\alpha})$ for Ψ. Therefore, both bracketing numbers $N_{[]}(\epsilon, \mathcal{I}_d, \| \cdot \|_2)$ and $N_{[]}(\epsilon, \Psi, \| \cdot \|_2)$ are of polynomial order in ϵ^{-1}.

To control $N_{[]}(\epsilon C, \mathcal{F}^{\eta}, \| \cdot \|_{\infty})$, note that under Assumption A.5(c), the v-dimension can be truncated to $[-M_{\epsilon}, M_{\epsilon}]$ with uniformly negligible error. Since in Assumption A.5 the Hölder smoothness is imposed only in p, uniformly over $(v, \tilde{\beta}_d, \tilde{\alpha})$, we obtain

$$N_{[]}(\epsilon C, \mathcal{F}^{\eta}, \| \cdot \|_{\infty}) \leqslant N\big(\epsilon, C_{\alpha} \times C_{\beta_d} \times [-M_{\epsilon}, M_{\epsilon}], | \cdot |\big) \times N\big(\epsilon, C^{\eta}(\mathcal{X}_P), \| \cdot \|_{\infty}\big).$$

The first term is of polynomial order in ϵ^{-1}. Moreover, by the uniform exponential tail condition in Assumption A.5(c), the truncation level M_{ϵ} need only grow at a logarithmic rate in ϵ^{-1}, so its contribution is asymptotically negligible relative to the polynomial entropy bound. Moreover, it is well known that for the Hölder ball $C^{\eta}(\mathcal{X}_P)$, $\log N\big(\epsilon, C^{\eta}(\mathcal{X}_P), \| \cdot \|_{\infty}\big) \lesssim \big(1/\epsilon\big)^{1/\eta}$. Therefore,

$$\int_0^1 \sqrt{\log N_{[]}(\epsilon, \Upsilon, \rho)}\, d\epsilon < \infty.$$

(d) Consequence. It follows from Theorem 2.11.9 of Van der Vaart and Wellner (1996) that the empirical process

$$\mathbb{G}_n(\upsilon) := \sum_{i=1}^{n} \Big(Z_{ni}(\upsilon) - E[Z_{ni}(\upsilon)] \Big)$$

is asymptotically stochastically equicontinuous on (Υ, ρ). In particular, consider $\hat{\upsilon} = (g, \hat{q}_d, \psi)$ and $\upsilon = (g, q_d, \psi)$ such that $\rho(\hat{\upsilon}, \upsilon) \to 0$,

$$\mathbb{G}_n(\hat{\upsilon}) - \mathbb{G}_n(\upsilon) = o_p(1). \tag{E.6}$$

Hence, we can decompose $\hat{U}_d(\omega; \tilde{\beta}_d, \tilde{\alpha})$ as the sum of an empirical process term and a term capturing the difference between the kernel estimator $\hat{F}_{V|P,D}$ and the true function $F_{V|P,D}$ as following:

$$
\begin{aligned}
\sqrt{n}\hat{U}_d(\omega; \tilde{\beta}_d, \tilde{\alpha}) = & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[\mathbf{1}(\tilde{V}_{di} \leqslant v) - F_{V|P,D}(v|\tilde{P}_i, d; \tilde{\beta}_d, \tilde{\alpha}) \Big] \psi_i(p, z, d; \tilde{\alpha}) \\
& - \sqrt{n}\, E_{\mathcal{Z}_n} \Big[\big(\hat{F}_{V|P,D}(v|\tilde{P}_i, d; \tilde{\beta}_d, \tilde{\alpha}) - F_{V|P,D}(v|\tilde{P}_i, d; \tilde{\beta}_d, \tilde{\alpha}) \big) \psi_i(p, z, d; \tilde{\alpha}) \Big] + o_p(1).
\end{aligned} \tag{E.7}
$$

where $E_{\mathcal{Z}_n}$ denotes the conditional mean on the original sample $\mathcal{Z}_n$, henceforth, $E_{\mathcal{Z}_n}[\hat{g}(X)] = E[\hat{g}(X)|\mathcal{Z}_n]$, and the remaining task is to linearize the second term involving the kernel estimator $\hat{F}_{V|P,D}$ in further steps.

For all $\tilde{\beta}_d \in C_{\beta_d}$ and $\tilde{\alpha} \in C_\alpha$, we want to show that

$$
\begin{aligned}
& \sqrt{n} E_{\mathcal{Z}_n} \left[(\hat{F}_{V|P,D}(v|P_i, d; \tilde{\beta}_d, \tilde{\alpha}) - F_{V|P,D}(v|P_i, d; \tilde{\beta}_d, \tilde{\alpha})) \psi_i(p, z, d; \tilde{\alpha}) \right] \\
= & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \mathbf{1}(\tilde{V}_{di} \leqslant v) - F_{V|P,D}(v|P_i, d; \tilde{\beta}_d, \tilde{\alpha}) \right\} \iota_d(\tilde{P}_i, \tilde{\alpha}) \mathbf{1}(D_i = d) + o_p(1),
\end{aligned} \tag{E.8}
$$

where $\iota_d(\tilde{P}_i, \tilde{\alpha})$ denotes $E[\psi_i(p, z, d; \tilde{\alpha}) \mid \tilde{P}_i, D = d]$.

The above projection step, which converts the nonparametric estimation effect into an influence-function representation, is a general result for nonparametric conditional mean estimators (see, e.g., Newey (1994); Escanciano et al. (2014)). The argument below for establishing (E.8) follows the same kernel linearization steps as in the verification of A.20 in Escanciano et al. (2014). The key ingredients are Assumptions A.4 and A.5, which control the bias of the kernel estimators so that

$$
\begin{aligned}
& \sqrt{n} \big(\hat{F}_{V|P,D}(v|P, d; \tilde{\beta}_d, \tilde{\alpha}) - F_{V|P,D}(v|P, d; \tilde{\beta}_d, \tilde{\alpha}) \big) \\
= & f^{-1}(P, d; \tilde{\alpha}) \Big[\hat{T}_{V|P,D}(v|P, d; \tilde{\beta}_d, \tilde{\alpha}) - T_{V|P,D}(v|P, d; \tilde{\beta}_d, \tilde{\alpha}) \\
& \qquad - F_{V|P,D}(v|P, d; \tilde{\beta}_d, \tilde{\alpha}) \big(\hat{f}(P, d; \tilde{\alpha}) - f(P, d; \tilde{\alpha}) \big) \Big] + o_p(1),
\end{aligned} \tag{E.9}
$$

where

$$
\begin{aligned}
T_{V|P,D}(v|P, d; \tilde{\beta}_d, \tilde{\alpha}) &= F_{V|P,D}(v|P, d; \tilde{\beta}_d, \tilde{\alpha}) f(P, d; \tilde{\alpha}), \\
\hat{T}_{V|P,D}(v|P, d; \tilde{\beta}_d, \tilde{\alpha}) &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(\tilde{V}_{di} \leqslant v, D_i = d) K_h(P - \tilde{P}_i), \\
\hat{f}(P, d; \tilde{\alpha}) &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(D_i = d) K_h(P - \tilde{P}_i).
\end{aligned}
$$

Using (E.9), we can rewrite (E.8) as

$$
\begin{aligned}
& \sqrt{n} E_{\mathcal{Z}_n} \left[(\hat{F}_{V|P,D}(v|P, d; \tilde{\beta}_d, \tilde{\alpha}) - F_{V|P,D}(v|P, d; \tilde{\beta}_d, \tilde{\alpha})) \psi_i(p, z, d; \tilde{\alpha}) \right] \\
= & \sqrt{n} \int \big[\hat{T}_{V|P,D}(v|\bar{p}, d; \tilde{\beta}_d, \tilde{\alpha}) - T_{V|P,D}(v|\bar{p}, d; \tilde{\beta}_d, \tilde{\alpha}) \big] \iota_d(\bar{p}; \tilde{\alpha})\, d\bar{p} && \text{(E.10)} \\
& - \sqrt{n} \int F_{V|P,D}(v|\bar{p}, d; \tilde{\beta}_d, \tilde{\alpha}) \big[\hat{f}(\bar{p}, d; \tilde{\alpha}) - f(\bar{p}, d; \tilde{\alpha}) \big] \iota_d(\bar{p}; \tilde{\alpha})\, d\bar{p} + o_p(1). && \text{(E.11)}
\end{aligned}
$$

where, with a slight abuse of notation, we write $\iota_d(\bar{p}; \tilde{\alpha}) = E[\psi_i(p, z, d; \tilde{\alpha}) \mid \tilde{P}_i = \bar{p}, D = d]$.

By Assumptions A.4 and A.5 on the bandwidth and kernel function, the bias term between $T_{V|P,D}(v|\bar{p},d;\tilde{\beta}_d,\tilde{\alpha})$ and $E[\hat{T}_{V|P,D}(v|\bar{p},d;\tilde{\beta}_d,\tilde{\alpha})]$ is $o(n^{-1/2})$ uniformly over $(\tilde{\beta}_d,\tilde{\alpha})$. Hence,

$$\begin{aligned}
&\sqrt{n}\int\big[\hat{T}_{V|P,D}(v|\bar{p},d;\tilde{\beta}_d,\tilde{\alpha})-T_{V|P,D}(v|\bar{p},d;\tilde{\beta}_d,\tilde{\alpha})\big]\iota_d(\bar{p};\tilde{\alpha})\,d\bar{p}\\
=&\sqrt{n}\int\big[\hat{T}_{V|P,D}(v|\bar{p},d;\tilde{\beta}_d,\tilde{\alpha})-E\hat{T}_{V|P,D}(v|\bar{p},d;\tilde{\beta}_d,\tilde{\alpha})\big]\iota_d(\bar{p};\tilde{\alpha})\,d\bar{p}+o_p(1)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\iota_d(\tilde{P}_i)\mathbf{1}(\tilde{V}_{di}\leqslant v,D_i=d)-E\big[\iota_d(\tilde{P}_i)\mathbf{1}(D_i=d)F_{V|P,D}(v|P_i,d;\tilde{\beta}_d,\tilde{\alpha})\big]\Big\}+o_p(1).
\end{aligned}\tag{E.12}$$

Similarly,

$$\begin{aligned}
&\sqrt{n}\int F_{V|P,D}(v|\bar{p},d;\tilde{\beta}_d,\tilde{\alpha})\big(\hat{f}(\bar{p},d;\tilde{\alpha})-f(\bar{p},d;\tilde{\alpha})\big)\iota_d(\bar{p};\tilde{\alpha})\,d\bar{p}\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\iota_d(\tilde{P}_i)\mathbf{1}(D_i=d)F_{V|P,D}(v|P_i,d;\tilde{\beta}_d,\tilde{\alpha})-E\big[\iota_d(\tilde{P}_i)\mathbf{1}(D_i=d)F_{V|P,D}(v|P_i,d;\tilde{\beta}_d,\tilde{\alpha})\big]\Big\}\\
&+o_p(1).
\end{aligned}\tag{E.13}$$

Combining the two displays yields (E.8) holds. This establishes the desired claim.

(2) To establish part (ii), set $(\hat{\alpha},\hat{\beta}_d)$ for specific values of $(\tilde{\alpha},\tilde{\beta}_d)$. By part (i),

$$\begin{aligned}
\sqrt{n}\hat{U}_d(\omega)=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\mathbf{1}(\hat{V}_{di}\leqslant v)-\hat{F}_{V|P,D}(v|P_i,d;\hat{\beta}_d,\hat{\alpha})\Big\}\,\psi_i(p,z,d;\hat{\alpha})\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\mathbf{1}(\hat{V}_{di}\leqslant v)-F_{V|P,D}(v|P_i,d;\hat{\beta}_d,\hat{\alpha})\Big\}\,\psi_i^{\perp}(p,z,d;\hat{\alpha})+o_p(1).
\end{aligned}\tag{E.14}$$

To further decompose this empirical process, we follow equation (4) in Escanciano et al. (2014) and write

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\mathbf{1}(\hat{V}_{di}\leqslant v)-F_{V|P,D}(v|\hat{P}_i,d;\hat{\beta}_d,\hat{\alpha})\Big\}\,\psi_i^{\perp}(p,z,d;\hat{\alpha})\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big\{\mathbf{1}(V_{di}\leqslant v)-F_{V|P,D}(v|P_i,d;\beta_d,\alpha)\big\}\,\psi_i^{\perp}(p,z,d;\alpha)\\
&+\sqrt{n}E_{\mathcal{Z}_n}\Big[\Big\{\mathbf{1}(\hat{V}_{di}\leqslant v)-F_{V|P,D}(v|\hat{P}_i,d;\hat{\beta}_d,\hat{\alpha})\Big\}\psi_i^{\perp}(p,z,d;\hat{\alpha})\\
&\qquad-\Big\{\mathbf{1}(V_{di}\leqslant v)-F_{V|P,D}(v|P_i,d;\beta_d,\alpha)\Big\}\psi_i^{\perp}(p,z,d;\alpha)\Big].
\end{aligned}\tag{E.15}$$

This decomposition separates the leading empirical process from the contribution of the generated regressors. In our setting, the latter only depends on parametric first-step estimators $(\hat{\alpha},\hat{\beta}_d)$.

Let the second term on the right-hand side of (E.15) be decomposed as

$$A_{1n}(\omega)=\sqrt{n}E_{\mathcal{Z}_n}\Big[\Big\{\mathbf{1}(\hat{V}_{di}\leqslant v)-\mathbf{1}(V_{di}\leqslant v)\Big\}\,\psi_i^{\perp}(p,z,d;\hat{\alpha})\Big],\tag{E.16}$$

$$A_{2n}(\omega)=-\sqrt{n}E_{\mathcal{Z}_n}\Big[\Big\{F_{V|P,D}(v|\hat{P}_i,d;\hat{\beta}_d,\hat{\alpha})-F_{V|P,D}(v|P_i,d;\beta_d,\alpha)\Big\}\,\psi_i^{\perp}(p,z,d;\hat{\alpha})\Big],\tag{E.17}$$

$$A_{3n}(\omega)=-\sqrt{n}E_{\mathcal{Z}_n}\big[\big\{\mathbf{1}(V_{di}\leqslant v)-F_{V|P,D}(v|P_i,d;\beta_d,\alpha)\big\}$$

$$\times \left\{\psi_i^{\perp}(p,z,d;\hat{\alpha}) - \psi_i^{\perp}(p,z,d;\alpha)\right\}\Big]. \tag{E.18}$$

For the term in (E.16), consider $\psi_i^{\perp}(p,z,d;\hat{\alpha}) = \psi_i(p,z,d;\hat{\alpha}) - \iota_d(\hat{P}_i;\hat{\alpha})\mathbf{1}(D_i = d)$, a first-order expansion yields

$$\begin{aligned}
A_{1n}(\omega) =& \sqrt{n} E_{\mathcal{Z}_n}\left[\left\{\mathbf{1}(\hat{V}_{di} \leqslant v) - \mathbf{1}(V_{di} \leqslant v)\right\}\psi_i(p,z,d;\hat{\alpha})\right] \\
&- \sqrt{n} E_{\mathcal{Z}_n}\left[\left\{\mathbf{1}(\hat{V}_{di} \leqslant v) - \mathbf{1}(V_{di} \leqslant v)\right\}\iota_d(\hat{P}_i;\hat{\alpha})\mathbf{1}(D_i = d)\right] \\
=& \sqrt{n}(\hat{\beta}_d - \beta_d)^{\top}\Big[\partial_{\beta_d} F_{V,P,Z,D}(v,p,z,d;\beta_d,\alpha) \\
&\qquad - E\Big\{\partial_{\beta_d} F_{V|P,D}(v|P_i,d;\beta_d,\alpha)\iota_d(P_i;\alpha)\mathbf{1}(D_i = d)\Big\}\Big] + o_p(1).
\end{aligned} \tag{E.19}$$

For the second term on the right-hand side of (E.19), the law of iterated expectations gives:

$$\begin{aligned}
&\sqrt{n} E_{\mathcal{Z}_n}\left[\left\{\mathbf{1}(\hat{V}_{di} \leqslant v) - \mathbf{1}(V_{di} \leqslant v)\right\}\iota_d(\hat{P}_i;\hat{\alpha})\mathbf{1}(D_i = d)\right] \\
=& \sqrt{n} E_{\mathcal{Z}_n}\left[E\left(\mathbf{1}(\hat{V}_{di} \leqslant v) - \mathbf{1}(V_{di} \leqslant v) \mid \hat{P}_i, D_i = d\right)\iota_d(\hat{P}_i;\hat{\alpha})\mathbf{1}(D_i = d)\right] \\
=& \sqrt{n}(\hat{\beta}_d - \beta_d)^{\top} E_{\mathcal{Z}_n}\left[\partial_{\beta_d} F_{V|P,D}(v|\hat{P}_i,d;\hat{\beta}_d,\hat{\alpha})\iota_d(\hat{P}_i;\hat{\alpha})\mathbf{1}(D_i = d)\right] + o_p(1).
\end{aligned}$$

The second equality in the display above follows from a first-order Taylor expansion of $F_{V|P,D}(v|\hat{P}_i,d;\hat{\beta}_d,\hat{\alpha})$ around (P_i,β_d,α) and the continuous differentiability of $F_{V|P,D}(v|P,d;\beta_d,\alpha)$ under Assumption A.2 and A.6.

Next, the term in (E.17) reflects the effect of estimating both α and β_d inside $F_{V|P,D}(v|P,d;\beta_d,\alpha)$. A first-order expansion gives

$$\begin{aligned}
A_{2n}(\omega) =& -\sqrt{n} E_{\mathcal{Z}_n}\left[\left\{F_{V|P,D}(v|\hat{P}_i,d;\hat{\beta}_d,\hat{\alpha}) - F_{V|P,D}(v|P_i,d;\beta_d,\alpha)\right\}\psi_i^{\perp}(p,z,d;\hat{\alpha})\right] \\
=& -\sqrt{n}(\hat{\beta}_d - \beta_d)^{\top} E\left\{\partial_{\beta_d} F_{V|P,D}(v|P_i,d;\beta_d,\alpha)\psi_i^{\perp}(p,z,d;\alpha)\right\} \\
&- \sqrt{n}(\hat{\alpha} - \alpha)^{\top} E\left\{\partial_{\alpha} F_{V|P,D}(v|P_i,d;\beta_d,\alpha)\psi_i^{\perp}(p,z,d;\alpha)\right\} + o_p(1).
\end{aligned} \tag{E.20}$$

Finally, the term $A_{3n}(\omega)$ in (E.18) is $o_p(1)$ because

$$E\left[\mathbf{1}(V_{di} \leqslant v) - F_{V|P,D}(v|P_i,d;\beta_d,\alpha) \mid P_i, Z_i, D_i = d\right] = 0,$$

Combining (E.16)–(E.18) yields

$$\begin{aligned}
&A_{1n}(\omega) + A_{2n}(\omega) + A_{3n}(\omega) \\
=& \sqrt{n}(\hat{\beta}_d - \beta_d)^{\top}\Big[\partial_{\beta_d} F_{V,P,Z,D}(v,p,z,d;\beta_d,\alpha) - E\Big\{\partial_{\beta_d} F_{V|P,D}(v|P_i,d;\beta_d,\alpha)\psi_i(p,z,d;\alpha)\Big\}\Big] \\
&- \sqrt{n}(\hat{\alpha} - \alpha)^{\top} E\left\{\partial_{\alpha} F_{V|P,D}(v|P_i,d;\beta_d,\alpha)\psi_i^{\perp}(p,z,d;\alpha)\right\} + o_p(1) \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\partial_{\beta_d}^{\top} F_{V,P,Z,D}(v,p,z,d;\beta_d,\alpha) - E\Big\{\partial_{\beta_d}^{\top} F_{V|P,D}(v|P_i,d;\beta_d,\alpha)\psi_i(p,z,d;\alpha)\Big\}\Big] l_d(Y_i,Z_i;\beta_d) \\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^{n} E\left\{\partial_{\alpha}^{\top} F_{V|P,D}(v|P_i,d;\beta_d,\alpha)\psi_i^{\perp}(p,z,d;\alpha)\right\} l(D_i,Z_i;\alpha) + o_p(1).
\end{aligned} \tag{E.21}$$

where the second equality follows from the asymptotic linearity of the first-step estimators in Assumption A.1(iv).

It remains to verify that the influence-function class

$$\mathcal{Z}_d^I = \left\{\zeta_d^I(\cdot;\omega,\alpha,\beta_d) : \omega \in \mathscr{W}\right\}$$

is P-Donsker. Write $\mathcal{Z}_d^I$ as the sum of three classes,

$$\mathcal{Z}_{d,0}^I = \left\{\left[\mathbf{1}(V_d \leqslant v) - F_{V|P,D}(v|P,d;\beta_d,\alpha)\right]\psi^{\perp}(p,z,d;\alpha) : \omega \in \mathscr{W}\right\},$$

$$\mathcal{Z}_{d,\alpha}^I = \left\{A_d(\omega)^{\top} l(D,Z;\alpha) : \omega \in \mathscr{W}\right\}, \qquad \mathcal{Z}_{d,\beta}^I = \left\{B_d(\omega)^{\top} l_d(Y,Z;\beta_d,\alpha) : \omega \in \mathscr{W}\right\}.$$

Under Assumption A.1(iv), the first step estimators admit fixed asymptotic linear representations with influence functions $l(D,Z;\alpha)$ and $l_d(Y,Z;\beta_d,\alpha)$ that are square integrable. This condition already incorporates the regularity of any nonparametric nuisance estimators used to obtain the Robinson first-step estimator. Hence, in the present Donsker verification, l and l_d are treated as fixed population influence functions, while the indexing over ω enters only through the scalar coefficient functions $A_d(\omega)$ and $B_d(\omega)$. Because α and β_d are finite-dimensional and $A_d(\omega)$ and $B_d(\omega)$ are bounded and uniformly continuous over ω, the classes $\mathcal{Z}_{d,\alpha}^I$ and $\mathcal{Z}_{d,\beta}^I$ are finite-dimensional linear spans of fixed $L_2(P)$ functions and are therefore P-Donsker.

Indeed, the first class $\mathcal{Z}_{d,0}^I$ is a bounded product class of the residual component

$$\mathcal{H}_{d,1} = \left\{\mathbf{1}(V_d \leqslant v) - F_{V|P,D}(v|P,d;\beta_d,\alpha) : v \in \mathcal{X}_V\right\}$$

and the weighting component

$$\mathcal{H}_{d,2} = \left\{\psi^{\perp}(p,z,d;\alpha) : (p,z) \in \mathcal{X}_P \times \mathcal{X}_Z\right\}.$$

For $\mathcal{H}_{d,1}$, the integrability of the bracketing entropy follows directly from the entropy bound for $\mathcal{F}^{\eta}$ established in part (1) and the fact that $F_{V|P,D}(v|\cdot,d;\beta_d,\alpha) \in \mathcal{F}^{\eta}$ by Assumption A.5. Also, for the weighting class $\mathcal{H}_{d,2}$

$$\{\psi^{\perp}(p,z,d;\alpha) : (p,z) \in \mathcal{X}_P \times \mathcal{X}_Z\}$$

has the same entropy order as the VC class Ψ, up to the projected component $\iota_d(P)$. By definition,

$$\iota_d(P) = \mathbf{1}(P \leqslant p)\, G(z \mid P;\alpha), \qquad G(z \mid p;\alpha) := \Pr(Z \leqslant z \mid P = p, D = d),$$

and Assumption A.5 implies that $G(z \mid \cdot;\alpha) \in \mathcal{G}^{\eta}$. Hence the projection term is obtained from a bounded indicator class multiplied by a class with the same entropy order as $\mathcal{G}^{\eta}$, so it does not alter the integrability of the bracketing entropy.

Since $\mathcal{Z}_{d,0}^I = \mathcal{H}_{d,1} \cdot \mathcal{H}_{d,2}$ is a bounded product of two classes with integrable bracketing entropy, its bracketing entropy is again integrable. Therefore,

$$\int_0^1 \sqrt{\log N_{[]}(\epsilon, \mathcal{Z}_{d,0}^I, L_2(P))}\, d\epsilon < \infty.$$

This proves part (ii). In addition, the same entropy argument also applies to the feasible analogue obtained by replacing $F_{V|P,D}$ and G with $\hat{F}_{V|P,D}$ and $\hat{G}$, since these estimators belong to $\mathcal{F}^\eta$ and $\mathcal{G}^\eta$ with probability approaching one. $\square$

E.2 Asymptotic Null Distribution of Monotonicity Statistics

To derive the asymptotic null distribution of $\hat{T}_M$, we first establish the weak convergence of the empirical copula process $\sqrt{n}(\hat{C}_d(\lambda) - C_d(\lambda))$ as follows:

Lemma E.2. *Under Assumptions A.1, A.2, and A.6, for $d \in \{0,1\}$,*

$$\sup_{\lambda\in\Lambda} |\sqrt{n}(\hat{C}_d(\lambda) - C_d(\lambda)) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_d^M(Y_i, D_i, Z_i, \lambda, \alpha, \beta_d)| = o_p(1), \tag{E.22}$$

where

$$\begin{aligned}\zeta_d^M(Y_i, D_i, Z_i; \lambda, \alpha, \beta_d) = {}& \mathbf{1}\{v_1 \leqslant V_{di} \leqslant v_2, D_i = d, P_i \leqslant F_P^{-1}(u)\} - C_d(\lambda) \\ & - \partial_u C_d(\lambda)\{\mathbf{1}(P_i \leqslant F_P^{-1}(u)) - u\} \\ & + l(D_i, Z_i; \alpha)^\top \partial_\alpha C_d(\lambda) + l_d(Y_i, Z_i; \beta_d, \alpha)^\top \partial_{\beta_d} C_d(\lambda).\end{aligned}$$

Furthermore,

$$\sqrt{n}(\hat{C}_d(\lambda) - C_d(\lambda)) \rightsquigarrow \mathbb{G}_{C_d}(\lambda) \text{ in } l^\infty(\Lambda), \tag{E.23}$$

where $\mathbb{G}_{C_d}(\lambda)$ denotes a zero mean Gaussian process with covariance structure

$$E[\zeta_d^M(Y_i, D_i, Z_i, \lambda_1, \alpha, \beta_d)\zeta_d^M(Y_i, D_i, Z_i, \lambda_2, \alpha, \beta_d)]$$

.

Proof. Let P_n denote the empirical measure and let $\mathbb{G}_n = \sqrt{n}(P_n - P)$. For $\lambda = (v_1, v_2, u)$, define

$$f_\lambda(Y, D, Z; a, b, t) = \mathbf{1}\{v_1 \leqslant Y - X^\top b \leqslant v_2,\ D = d,\ P(Z; a) \leqslant t\},$$

where the rank u and the propensity threshold t are linked by $u = F_P(t; a)$ (equivalently $t = F_P^{-1}(u; a)$). In Steps 1–2 below, t varies freely and u is understood through this relation; in Steps 3–5, u is the primary variable and t is set to the corresponding quantile.

Step 1: Empirical process regularity. Consider the two classes

$$\mathcal{G}_P = \{\mathbf{1}(P(Z; a) \leqslant t) : (a, t) \in C_\alpha \times \mathcal{X}_P\},$$

and

$$\mathcal{F}_d = \{f_\lambda(Y, D, Z; a, b, t) : \lambda \in \Lambda,\ (a, b, t) \in C_\alpha \times C_{\beta_d} \times \mathcal{X}_P\}.$$

By Lemmas 2.6.15, 2.6.16, and 2.6.18 of Van der Vaart and Wellner (1996), the propensity-score threshold class $\mathcal{G}_P$ is VC-subgraph. Indeed, $P(Z; a)$ is a finite-dimensional parametric transformation of Z, and the sets $\{z : P(z; a) \leqslant t\}$ are generated by thresholding this parametric index over $(a, t) \in C_\alpha \times \mathcal{X}_P$. The same argument applies to the residual threshold classes $\{\mathbf{1}(Y - X^\top b \leqslant v) : (v, b) \in \mathcal{X}_V \times C_{\beta_d}\}$. Therefore,

$\mathbf{1}\{v_1 \leqslant V_d(b) \leqslant v_2\}$ is the intersection of two such threshold classes and is VC-type. Since $D = d$ is a fixed treatment indicator, the class $\mathcal{F}_d$ is formed by finite intersections of the treatment indicator, the residual threshold sets, and the propensity score threshold set. Finite intersections of VC-subgraph classes are VC-type. Hence both $\mathcal{G}_P$ and $\mathcal{F}_d$ are pointwise measurable, uniformly bounded, and P-Donsker.

The boundary mass bounds implied by Assumption A.6 give $L_2(P)$-continuity of these classes in their finite-dimensional indices. For example, perturbing v_1, v_2, b, a, or t only changes the indicator on sets where either $Y - X^\top b$ is close to a residual boundary or $P(Z; a)$ is close to a propensity-score boundary. These sets have probability of order equal to the perturbation size. Therefore, the Donsker processes indexed by $\mathcal{G}_P$ and $\mathcal{F}_d$ are stochastically equicontinuous in the relevant parameters.

Step 2: Quantile expansion. By the Donsker property and stochastic equicontinuity of $\mathcal{G}_P$, together with the first step expansion for $\hat{\alpha}$,

$$\sup_{t \in \mathcal{X}_P} \left| \sqrt{n}\{\hat{F}_P(t; \hat{\alpha}) - F_P(t; \alpha)\} - \mathbb{G}_n\{\mathbf{1}(P(Z; \alpha) \leqslant t)\} - \partial_\alpha F_P(t; \alpha)^\top \sqrt{n}(\hat{\alpha} - \alpha) \right| = o_p(1).$$

By Assumptions A.1(iii), A.2, and A.6, $f_P(\cdot; \alpha)$ is bounded away from zero on $\mathcal{X}_P$. The standard Bahadur representation for empirical quantiles (Lemma 3.9.23 of Van der Vaart and Wellner (1996)) applies on compact subsets of $(0, 1)$. At the endpoints $u = 0$ and $u = 1$, the empirical copula process and the rank-correction term are handled directly by their boundary values. Stochastic equicontinuity on Λ (Step 1) then extends the interior representation to the closed interval $[0, 1]$.

$$\sup_{u \in [0,1]} \Bigg| \sqrt{n}\big(\hat{F}_P^{-1}(u; \hat{\alpha}) - F_P^{-1}(u; \alpha)\big) + f_P^{-1}\big(F_P^{-1}(u; \alpha)\big) \times \Big[\mathbb{G}_n\{\mathbf{1}(F_P(P; \alpha) \leqslant u) - u\} + \partial_\alpha F_P\big(F_P^{-1}(u; \alpha); \alpha\big)^\top \sqrt{n}(\hat{\alpha} - \alpha)\Big] \Bigg| = o_p(1).$$

Step 3: Decomposition of the copula process. For each $\lambda = (v_1, v_2, u)$,

$$\begin{aligned}\sqrt{n}\{\hat{C}_d(\lambda) - C_d(\lambda)\} = {} & \mathbb{G}_n f_\lambda(Y, D, Z; \alpha, \beta_d, F_P^{-1}(u; \alpha)) \\ & + \sqrt{n}\{P f_\lambda(Y, D, Z; \hat{\alpha}, \hat{\beta}_d, \hat{F}_P^{-1}(u; \hat{\alpha})) - P f_\lambda(Y, D, Z; \alpha, \beta_d, F_P^{-1}(u; \alpha))\} + r_n(\lambda),\end{aligned}$$

where

$$r_n(\lambda) = \mathbb{G}_n\{f_\lambda(Y, D, Z; \hat{\alpha}, \hat{\beta}_d, \hat{F}_P^{-1}(u; \hat{\alpha})) - f_\lambda(Y, D, Z; \alpha, \beta_d, F_P^{-1}(u; \alpha))\}.$$

The stochastic equicontinuity established in Step 1, together with $(\hat{\alpha}, \hat{\beta}_d, \hat{F}_P^{-1}(\cdot; \hat{\alpha})) \to_p (\alpha, \beta_d, F_P^{-1}(\cdot; \alpha))$, gives

$$\sup_{\lambda \in \Lambda} |r_n(\lambda)| = o_p(1).$$

Step 4: Drift expansion. Since

$$P f_\lambda(Y, D, Z; a, b, t) = H_d(v_1, v_2, t; a, b) = \int_0^t \big[F_{V|P,D}(v_2 \mid p, d; b, a) - F_{V|P,D}(v_1 \mid p, d; b, a)\big] f_{P,D}(p, d; a)\, dp,$$

the differentiability of H_d implied by Assumption A.6 gives the uniform first-order expansion

$$\begin{aligned}&\sqrt{n}\{Pf_\lambda(Y,D,Z;\hat\alpha,\hat\beta_d,\hat F_P^{-1}(u;\hat\alpha))-Pf_\lambda(Y,D,Z;\alpha,\beta_d,F_P^{-1}(u;\alpha))\}\\&=\partial_\alpha H_d(v_1,v_2,F_P^{-1}(u;\alpha);\alpha,\beta_d)^\top\sqrt{n}(\hat\alpha-\alpha)+\partial_{\beta_d}H_d(v_1,v_2,F_P^{-1}(u;\alpha);\alpha,\beta_d)^\top\sqrt{n}(\hat\beta_d-\beta_d)\\&\quad+\partial_t H_d(v_1,v_2,F_P^{-1}(u;\alpha);\alpha,\beta_d)\sqrt{n}\{\hat F_P^{-1}(u;\hat\alpha)-F_P^{-1}(u;\alpha)\}+o_p(1),\end{aligned}$$

uniformly in $\lambda\in\Lambda$.

Substituting the quantile expansion from Step 2 and using

$$\partial_t H_d(v_1,v_2,F_P^{-1}(u;\alpha);\alpha,\beta_d)=f_P(F_P^{-1}(u;\alpha);\alpha)\partial_u C_d(\lambda),$$

we obtain

$$\begin{aligned}&\sqrt{n}\{Pf_\lambda(Y,D,Z;\hat\alpha,\hat\beta_d,\hat F_P^{-1}(u;\hat\alpha))-Pf_\lambda(Y,D,Z;\alpha,\beta_d,F_P^{-1}(u;\alpha))\}\\&=\partial_{\beta_d}C_d(\lambda)^\top\sqrt{n}(\hat\beta_d-\beta_d)+\partial_\alpha C_d(\lambda)^\top\sqrt{n}(\hat\alpha-\alpha)\\&\quad-\partial_u C_d(\lambda)\mathbb{G}_n\{\mathbf{1}(P(Z;\alpha)\leqslant F_P^{-1}(u;\alpha))-u\}+o_p(1),\end{aligned}$$

where $\partial_\alpha C_d$ is the total derivative defined above.

Step 5: Influence function representation. Substituting the asymptotic linear representations of $\hat\alpha$ and $\hat\beta_d$ into the display in Step 4 and combining with the empirical-process term from Step 3 gives

$$\sqrt{n}\{\hat C_d(\lambda)-C_d(\lambda)\}=\frac{1}{\sqrt{n}}\sum_{i=1}^n\zeta_d^M(Y_i,D_i,Z_i;\lambda,\alpha,\beta_d)+o_p(1)$$

uniformly in $\lambda\in\Lambda$.

Finally, the influence-function class is the sum of a VC-type centered indicator class, the rank correction class $\{\partial_u C_d(\lambda)(\mathbf{1}(P_i\leqslant F_P^{-1}(u;\alpha))-u):\lambda\in\Lambda\}$, and finite dimensional linear spans of the first step influence functions multiplied by bounded smooth coefficients. These classes are Donsker under the conditions above and the moment conditions implicit in the first step linear representations. Hence the uniform expansion implies weak convergence in $\ell^\infty(\Lambda)$ to the Gaussian process stated in the lemma.

□

F Simulation Results: All Significance Levels

This appendix reports the full simulation results at the 1%, 5%, and 10% significance levels. Tables 1, 2, and 3 correspond to Tables 1, 2, and 3 in the main text, respectively.

Table 1: Simulation Results for Size (All Significance Levels)

		$n = 200$			$n = 500$			$n = 1000$			$n = 2000$		
DGP	Test	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
N1	$\hat{T}_I$	0.008	0.044	0.091	0.005	0.043	0.078	0.003	0.041	0.077	0.013	0.059	0.113
	$\hat{T}_M$	0.000	0.002	0.008	0.000	0.002	0.009	0.000	0.002	0.005	0.000	0.002	0.013
	$\hat{T}_{\text{Max}}$	0.000	0.002	0.009	0.001	0.007	0.016	0.000	0.013	0.026	0.007	0.039	0.077
	$\hat{T}_{\text{Sum}}$	0.000	0.006	0.020	0.001	0.004	0.016	0.001	0.005	0.014	0.001	0.008	0.024
	$\hat{T}_{\text{BC}}$	0.005	0.024	0.046	0.003	0.020	0.044	0.001	0.019	0.043	0.009	0.032	0.061
N2	$\hat{T}_I$	0.014	0.056	0.112	0.008	0.040	0.084	0.008	0.054	0.099	0.010	0.043	0.076
	$\hat{T}_M$	0.000	0.002	0.010	0.000	0.002	0.009	0.000	0.004	0.012	0.000	0.002	0.009
	$\hat{T}_{\text{Max}}$	0.000	0.004	0.014	0.001	0.004	0.018	0.002	0.012	0.046	0.006	0.027	0.050
	$\hat{T}_{\text{Sum}}$	0.001	0.008	0.021	0.002	0.013	0.025	0.002	0.010	0.027	0.003	0.013	0.024
	$\hat{T}_{\text{BC}}$	0.009	0.026	0.058	0.005	0.025	0.042	0.005	0.033	0.058	0.006	0.022	0.045
N3	$\hat{T}_I$	0.002	0.046	0.092	0.008	0.050	0.094	0.014	0.058	0.109	0.008	0.046	0.090
	$\hat{T}_M$	0.000	0.003	0.022	0.000	0.005	0.026	0.001	0.016	0.029	0.002	0.016	0.033
	$\hat{T}_{\text{Max}}$	0.000	0.003	0.021	0.000	0.005	0.028	0.002	0.019	0.044	0.001	0.021	0.047
	$\hat{T}_{\text{Sum}}$	0.002	0.008	0.025	0.001	0.011	0.039	0.005	0.026	0.055	0.003	0.017	0.042
	$\hat{T}_{\text{BC}}$	0.001	0.018	0.048	0.006	0.026	0.055	0.011	0.040	0.074	0.007	0.027	0.061
N4	$\hat{T}_I$	0.005	0.039	0.081	0.004	0.035	0.083	0.011	0.049	0.091	0.014	0.056	0.099
	$\hat{T}_M$	0.000	0.011	0.040	0.000	0.023	0.060	0.004	0.032	0.076	0.009	0.050	0.102
	$\hat{T}_{\text{Max}}$	0.000	0.011	0.042	0.000	0.023	0.057	0.004	0.032	0.077	0.010	0.048	0.099
	$\hat{T}_{\text{Sum}}$	0.002	0.015	0.039	0.001	0.021	0.056	0.004	0.034	0.087	0.009	0.058	0.114
	$\hat{T}_{\text{BC}}$	0.004	0.023	0.049	0.002	0.023	0.057	0.014	0.047	0.081	0.017	0.057	0.102

Table 2: Power Performance under Violations of H_0^I (All Significance Levels)

		$n = 200$			$n = 500$			$n = 1000$			$n = 2000$		
DGP	Test	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
ALT1(I)	$\hat{T}_I$	0.068	0.178	0.285	0.282	0.535	0.663	0.686	0.876	0.936	0.991	0.998	1.000
	$\hat{T}_M$	0.000	0.005	0.013	0.000	0.000	0.007	0.001	0.006	0.011	0.005	0.025	0.041
	$\hat{T}_{\text{Max}}$	0.002	0.020	0.037	0.069	0.205	0.312	0.505	0.740	0.831	0.980	0.998	0.999
	$\hat{T}_{\text{Sum}}$	0.004	0.041	0.093	0.059	0.212	0.315	0.315	0.613	0.738	0.872	0.971	0.984
	$\hat{T}_{\text{BC}}$	0.046	0.125	0.181	0.230	0.428	0.535	0.624	0.807	0.876	0.982	0.998	0.998
ALT2(I)	$\hat{T}_I$	0.025	0.106	0.202	0.067	0.227	0.364	0.290	0.646	0.788	0.892	0.991	0.997
	$\hat{T}_M$	0.000	0.002	0.007	0.000	0.005	0.009	0.000	0.006	0.015	0.001	0.010	0.021
	$\hat{T}_{\text{Max}}$	0.001	0.011	0.020	0.011	0.044	0.101	0.152	0.420	0.587	0.849	0.982	0.994
	$\hat{T}_{\text{Sum}}$	0.003	0.022	0.047	0.008	0.049	0.111	0.064	0.232	0.386	0.541	0.843	0.934
	$\hat{T}_{\text{BC}}$	0.018	0.060	0.108	0.046	0.135	0.229	0.229	0.498	0.649	0.835	0.974	0.991
ALT3(I)	$\hat{T}_I$	0.015	0.062	0.134	0.019	0.102	0.201	0.089	0.313	0.480	0.452	0.802	0.915
	$\hat{T}_M$	0.000	0.000	0.010	0.000	0.005	0.017	0.000	0.005	0.023	0.000	0.012	0.030
	$\hat{T}_{\text{Max}}$	0.000	0.002	0.013	0.000	0.015	0.034	0.026	0.107	0.207	0.340	0.690	0.815
	$\hat{T}_{\text{Sum}}$	0.001	0.014	0.036	0.003	0.026	0.061	0.012	0.107	0.194	0.158	0.467	0.639
	$\hat{T}_{\text{BC}}$	0.011	0.038	0.062	0.014	0.055	0.107	0.064	0.186	0.316	0.386	0.692	0.807
ALT4(I)	$\hat{T}_I$	0.007	0.047	0.085	0.007	0.051	0.105	0.030	0.122	0.223	0.176	0.457	0.620
	$\hat{T}_M$	0.000	0.002	0.008	0.000	0.004	0.008	0.000	0.003	0.006	0.000	0.000	0.008
	$\hat{T}_{\text{Max}}$	0.000	0.003	0.011	0.001	0.010	0.025	0.011	0.066	0.113	0.145	0.390	0.539
	$\hat{T}_{\text{Sum}}$	0.000	0.005	0.015	0.002	0.007	0.022	0.004	0.032	0.061	0.027	0.136	0.256
	$\hat{T}_{\text{BC}}$	0.003	0.029	0.049	0.006	0.025	0.055	0.020	0.081	0.124	0.133	0.319	0.457

Table 3: Power Performance under Violations of H_0^M (All Significance Levels)

		$n=200$			$n=500$			$n=1000$			$n=2000$		
DGP	Test	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
ALT1(M)	$\hat{T}_I$	0.014	0.062	0.107	0.010	0.053	0.106	0.005	0.041	0.099	0.009	0.057	0.118
	$\hat{T}_M$	0.256	0.666	0.825	0.973	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$\hat{T}_{\text{Max}}$	0.256	0.665	0.826	0.973	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$\hat{T}_{\text{Sum}}$	0.184	0.532	0.715	0.921	0.991	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	$\hat{T}_{\text{BC}}$	0.189	0.528	0.687	0.955	0.998	0.998	1.000	1.000	1.000	1.000	1.000	1.000
ALT2(M)	$\hat{T}_I$	0.010	0.060	0.113	0.007	0.056	0.102	0.013	0.046	0.098	0.014	0.054	0.100
	$\hat{T}_M$	0.001	0.016	0.044	0.047	0.223	0.383	0.540	0.926	0.980	0.997	1.000	1.000
	$\hat{T}_{\text{Max}}$	0.001	0.017	0.049	0.045	0.201	0.357	0.474	0.883	0.960	0.994	1.000	1.000
	$\hat{T}_{\text{Sum}}$	0.005	0.029	0.082	0.028	0.155	0.284	0.239	0.687	0.848	0.890	0.998	0.999
	$\hat{T}_{\text{BC}}$	0.009	0.041	0.076	0.030	0.152	0.272	0.416	0.818	0.928	0.993	1.000	1.000
ALT3(M)	$\hat{T}_I$	0.008	0.060	0.111	0.010	0.055	0.102	0.016	0.042	0.094	0.011	0.067	0.105
	$\hat{T}_M$	0.000	0.010	0.046	0.013	0.084	0.164	0.045	0.207	0.362	0.216	0.491	0.658
	$\hat{T}_{\text{Max}}$	0.000	0.010	0.046	0.013	0.081	0.161	0.045	0.211	0.352	0.210	0.474	0.635
	$\hat{T}_{\text{Sum}}$	0.001	0.024	0.068	0.006	0.056	0.141	0.026	0.157	0.281	0.118	0.368	0.543
	$\hat{T}_{\text{BC}}$	0.005	0.027	0.070	0.013	0.071	0.137	0.044	0.151	0.242	0.176	0.384	0.530
ALT4(M)	$\hat{T}_I$	0.009	0.043	0.077	0.007	0.042	0.091	0.009	0.056	0.100	0.011	0.055	0.105
	$\hat{T}_M$	0.000	0.010	0.033	0.004	0.027	0.067	0.022	0.100	0.207	0.124	0.460	0.672
	$\hat{T}_{\text{Max}}$	0.000	0.011	0.034	0.003	0.029	0.068	0.020	0.096	0.195	0.099	0.390	0.606
	$\hat{T}_{\text{Sum}}$	0.001	0.013	0.036	0.004	0.025	0.068	0.012	0.076	0.161	0.049	0.232	0.418
	$\hat{T}_{\text{BC}}$	0.006	0.028	0.052	0.007	0.034	0.069	0.025	0.067	0.152	0.097	0.318	0.504

References

Beare, Brendan K. and Jong-Myun Moon, "NONPARAMETRIC TESTS OF DENSITY RATIO ORDERING," *Econometric Theory*, 6 2015, *31* (03), 471–492.

Bücher, Axel and Holger Dette, "A note on bootstrap approximations for the empirical copula process," *Statistics & Probability Letters*, 12 2010, *80* (23-24), 1925–1932.

Delgado, Miguel A. and Juan Carlos Escanciano, "Distribution-free tests of stochastic monotonicity," *Journal of Econometrics*, 9 2012, *170* (1), 68–75.

der Vaart, Aad W. Van and Jon A. Wellner, *Weak Convergence and Empirical Processes* Springer Series in Statistics, New York, NY: Springer New York, 1996.

Escanciano, Juan Carlos, David T. Jacho-Chávez, and Arthur Lewbel, "Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing," *Journal of Econometrics*, 2014, *178* (PART 3), 426–443.

Fang, Zheng and Andres Santos, "Inference on Directionally Differentiable Functions," *The Review of Economic Studies*, 9 2018, *86* (1), 377–412.

Fermanian, Jean-David, Dragan Radulovic, and Marten Wegkamp, "Weak convergence of empirical copula processes," *Bernoulli*, 10 2004, *10* (5), 847–860.

Gänssler, Peter and W Stute, *Seminar on Empirical Processes*, Vol. 9 of *DMV Seminar*, Basel: Birkhäuser Basel, 1987.

Heckman, James J. and Edward Vytlacil, "Structural equations, treatment effects, and econometric policy evaluation," *Econometrica*, 5 2005, *73* (3), 669–738.

Ichimura, Hidehiko and Sokbae Lee, "Characterization of the asymptotic distribution of semiparametric M-estimators," *Journal of Econometrics*, 2010, *159* (2), 252–266.

Kallenberg, Olav, *Foundations of modern probability*, Springer, 1997.

Newey, Whitney K., "The asymptotic variance of semiparametric estimators," *Econometrica*, 1994.

Seo, Juwon, "Tests of stochastic monotonicity with improved power," *Journal of Econometrics*, 11 2018, *207* (1), 53–70.

Shapiro, A., "On concepts of directional differentiability," *Journal of Optimization Theory and Applications*, 9 1990, *66* (3), 477–487.

Van der Vaart, *Asymptotic statistics*, Cambridge: Cambridge University Press, 2000.