Covariate Distribution Balance via Propensity Scores: Supplemental Appendix

Pedro H. C. Sant'Anna* Xiaojun Song† Qi Xu‡

November 9, 2021

This supplemental appendix contains additional computational details, auxiliary lemmas, and proofs of the main theorems presented in the main text. Appendices S1 and S2 present the complete simulation results based on the stylized and empirically calibrated data generating process (DGP), respectively. Appendix S3 discusses closed-form representation of the integrated propensity score (IPS) objective function using the three families of weighting functions under Assumption 3 in the main text. Appendix S5 presents auxiliary lemmas. Appendix S6 collects all the proofs of the main results of the paper, while Appendix S7 collects the proofs of the results when treatment allocation is endogenous. Finally, Appendix S8 presents results for the case when one is interested in average, distributional and quantile treatment effects on the treated subpopulation.

*Microsoft and Vanderbilt University. E-mail: pedro.h.santanna@vanderbilt.edu.

†Peking University. E-mail: sxj@gsm.pku.edu.cn.

‡Vanderbilt University. E-mail: qi.xu.1@vanderbilt.edu.

S1 Simulation results for stylized DGP

This section display all the simulations results discussed in Section 5.1 of the main text.

Table S1.1: Monte Carlo study of the performance of IPW estimators for ATE and QTE based on different propensity score estimation methods. Sample size: $n = 500$.

	Correctly Specified Model						Misspecified Model					
	Bias	RMSE	relMSE	COV	ACIL	ARE	Bias	RMSE	relMSE	COV	ACIL	ARE
(a) **ATE**												
IPS_{exp}	0.14	3.68	0.89	0.95	14.07	1.22	1.93	4.03	0.78	0.92	13.76	1.36
IPS_{ind}	0.96	3.67	0.88	0.97	15.51	1.00	2.45	4.55	0.99	0.95	17.73	0.82
IPS_{proj}	0.15	3.61	0.85	0.94	13.82	1.26	0.50	3.40	0.55	0.96	15.04	1.14
$CBPS_{just}$	0.03	4.06	1.08	0.93	14.99	1.07	2.71	4.56	1.00	0.87	13.95	1.32
$CBPS_{over}$	0.05	3.91	1.00	0.95	15.53	1.00	2.56	4.57	1.00	0.93	16.05	1.00
MLE	-0.05	4.53	1.34	0.94	16.46	0.89	5.92	10.41	5.20	0.86	20.50	0.61
(b) **QTE(0.10)**												
IPS_{exp}	0.17	4.87	1.04	0.96	19.42	0.99	-3.96	6.34	1.32	0.87	19.80	0.99
IPS_{ind}	0.47	5.25	1.21	0.96	21.44	0.81	-2.86	5.96	1.17	0.94	22.81	0.75
IPS_{proj}	0.16	4.85	1.04	0.95	19.43	0.99	-2.15	5.58	1.02	0.99	29.85	0.44
$CBPS_{just}$	0.03	4.70	0.98	0.96	18.79	1.05	-3.16	5.81	1.11	0.89	19.39	1.03
$CBPS_{over}$	0.04	4.76	1.00	0.96	19.29	1.00	-2.79	5.52	1.00	0.92	19.72	1.00
MLE	0.01	4.76	1.00	0.96	18.90	1.04	-1.54	5.64	1.04	0.94	20.28	0.95
(c) **QTE(0.25)**												
IPS_{exp}	0.09	4.35	1.01	0.95	17.21	1.03	-2.08	4.79	1.23	0.93	17.01	1.07
IPS_{ind}	0.64	4.69	1.17	0.97	19.32	0.82	-1.10	4.60	1.14	0.97	19.55	0.81
IPS_{proj}	0.08	4.33	1.00	0.96	17.18	1.04	-1.32	4.63	1.15	0.98	22.91	0.59
$CBPS_{just}$	0.00	4.36	1.02	0.96	17.06	1.05	-1.25	4.46	1.07	0.95	16.99	1.07
$CBPS_{over}$	0.04	4.33	1.00	0.97	17.51	1.00	-1.10	4.32	1.00	0.95	17.58	1.00
MLE	-0.04	4.47	1.07	0.95	17.46	1.01	0.71	6.10	2.00	0.96	20.08	0.77
(d) **QTE(0.50)**												
IPS_{exp}	0.21	4.50	0.95	0.94	17.59	1.13	1.06	4.48	0.86	0.94	17.16	1.21
IPS_{ind}	0.94	4.65	1.01	0.96	19.15	0.96	1.73	4.86	1.01	0.96	19.81	0.91
IPS_{proj}	0.20	4.44	0.92	0.94	17.42	1.16	0.19	4.27	0.78	0.96	18.11	1.08
$CBPS_{just}$	0.08	4.79	1.07	0.94	18.42	1.03	1.94	4.94	1.04	0.93	17.56	1.15
$CBPS_{over}$	0.10	4.62	1.00	0.95	18.73	1.00	1.76	4.83	1.00	0.95	18.86	1.00
MLE	0.00	4.98	1.16	0.95	19.36	0.94	5.29	11.78	5.93	0.90	24.86	0.58
(e) **QTE(0.75)**												
IPS_{exp}	0.25	5.64	0.95	0.94	21.65	1.15	5.30	7.48	0.87	0.83	20.76	1.33
IPS_{ind}	1.34	5.47	0.90	0.95	22.06	1.11	5.48	7.76	0.94	0.89	24.83	0.93
IPS_{proj}	0.25	5.56	0.93	0.94	21.23	1.20	1.99	5.31	0.44	0.96	24.12	0.98
$CBPS_{just}$	0.17	6.17	1.14	0.93	23.19	1.00	6.16	8.29	1.07	0.79	21.34	1.25
$CBPS_{over}$	0.21	5.77	1.00	0.95	23.23	1.00	5.65	8.01	1.00	0.87	23.90	1.00
MLE	0.02	7.11	1.52	0.94	25.05	0.86	11.26	17.99	5.04	0.78	30.87	0.60
(f) **QTE(0.90)**												
IPS_{exp}	-0.06	8.29	1.00	0.93	30.81	1.06	10.12	13.23	1.01	0.76	30.58	1.21
IPS_{ind}	1.52	7.69	0.86	0.95	29.76	1.13	9.60	13.00	0.97	0.88	37.15	0.82
IPS_{proj}	-0.03	8.03	0.94	0.94	29.98	1.12	3.81	8.06	0.37	0.96	40.11	0.70
$CBPS_{just}$	-0.09	9.06	1.19	0.93	32.13	0.97	10.64	13.65	1.07	0.72	30.06	1.25
$CBPS_{over}$	-0.13	8.29	1.00	0.95	31.66	1.00	9.85	13.17	1.00	0.82	33.65	1.00
MLE	-0.30	10.92	1.74	0.92	34.93	0.82	15.81	22.69	2.97	0.66	34.75	0.94

Note: Simulations based on 1,000 Monte Carlo experiments. Bias, Monte Carlo Bias; RMSE, Monte Carlo root mean square error; relMSE, relative Monte Carlo mean square error; COV, Monte Carlo coverage of 95% normal confidence interval; ACIL, Monte Carlo average of 95% normal confidence interval length; ARE, asymptotic relative efficiency; ATE, average treatment effect; QTE(τ), quantile treatment effect at τ quantile. Both relMSE and ARE are expressed with respect to the IPW estimator based on the over-identified CBPS. The propensity score model is based on a logistic link function. IPS_{ind}, IPW estimator based on IPS estimator (2.10); IPS_{proj}, IPW estimator based on IPS estimator (2.11); IPS_{exp}, IPW estimator based on IPS estimator (2.12); $CBPS_{just}$, IPW estimator based on the (just-identified) CBPS estimator with moment equation (2.2), with $f(\mathbf{X}^*) = \mathbf{X}^*$; $CBPS_{over}$, IPW estimator based on the (over-identified) CBPS estimator with moment equation (2.2), with $f(\mathbf{X}^*) = \left(\mathbf{X}^{*,\prime}, \dot{p}(\mathbf{X}^*;\boldsymbol{\beta})^{\prime}\right)^{\prime}$, with $\dot{p}(\mathbf{X}^*;\boldsymbol{\beta})$ the derivative of the propensity score model with respect to $\boldsymbol{\beta}$; MLE , IPW estimator based on MLE. $\mathbf{X}^* = \mathbf{X}$ when the PS model is correctly specified, and $\mathbf{X}^* = \mathbf{W}$ when the PS is misspecified.

Table S1.2: Monte Carlo study of the performance of distributional covariate imbalances based on different propensity score estimation methods. Sample size: $n = 500$.

	Correctly Specified						Misspecified					
	KS_{bal}	$RCvM_{bal}$	KS_{bal1}	$RCvM_{bal1}$	KS_{bal0}	$RCvM_{bal0}$	KS_{bal}	$RCvM_{bal}$	KS_{bal1}	$RCvM_{bal1}$	KS_{bal0}	$RCvM_{bal0}$
IPS_{exp}	8.87	2.07	4.91	0.94	5.60	1.41	9.78	2.49	5.77	1.24	5.62	1.46
IPS_{ind}	8.16	1.77	4.65	0.84	4.96	1.13	8.65	2.13	5.06	1.07	4.99	1.22
IPS_{proj}	8.84	2.06	4.84	0.94	5.53	1.39	13.46	3.72	6.93	1.63	7.32	2.23
$CBPS_{just}$	9.22	2.15	5.12	0.98	5.85	1.47	10.28	2.63	6.05	1.32	5.74	1.52
$CBPS_{over}$	9.18	2.11	4.97	0.98	5.62	1.38	10.14	2.56	5.84	1.25	5.62	1.49
MLE	9.19	2.17	5.12	0.96	5.96	1.55	12.07	3.04	8.46	2.07	5.56	1.45

Note: Simulations based on 1,000 Monte Carlo experiments. $KS_{bal}, KS_{bal1}, KS_{bal0}, RCvM_{bal}, RCvM_{bal1}, RCvM_{bal0}$ as defined in 5.1. All imbalances are presented in percentage points. The propensity score model is based on a logistic link function. IPS_{ind}, IPW estimator based on IPS estimator (2.10); IPS_{proj}, IPW estimator based on IPS estimator (2.11); IPS_{exp}, IPW estimator based on IPS estimator (2.12); $CBPS_{just}$, IPW estimator based on the (just-identified) CBPS estimator with moment equation (2.2), with $f(\mathbf{X}^*) = \mathbf{X}^*$; $CBPS_{over}$, IPW estimator based on the (over-identified) CBPS estimator with moment equation (2.2), with $f(\mathbf{X}^*) = \left(\mathbf{X}^{*,\prime}, \dot{p}(\mathbf{X}^*;\boldsymbol{\beta})^{\prime}\right)^{\prime}$, with $\dot{p}(\mathbf{X}^*;\boldsymbol{\beta})$ the derivative of the propensity score model with respect to $\boldsymbol{\beta}$; MLE , IPW estimator based on MLE. $\mathbf{X}^* = \mathbf{X}$ when the PS model is correctly specified, and $\mathbf{X}^* = \mathbf{W}$ when the PS is misspecified.

Table S1.3: Monte Carlo study of the performance of IPW estimators for LATE and LQTE based on different instrument propensity score estimation methods. Sample size: $n = 500$.

	Correctly Specified Model						Misspecified Model					
	Bias	RMSE	relMSE	COV	ACIL	ARE	Bias	RMSE	relMSE	COV	ACIL	ARE
(a) **LATE**												
$LIPS_{exp}$	-0.24	4.39	0.74	0.95	17.76	1.76	5.11	6.62	0.38	0.94	21.45	1.30
$LIPS_{ind}$	-5.55	6.79	1.76	0.72	16.17	2.12	-0.78	4.79	0.20	0.96	20.22	1.47
$LIPS_{proj}$	-0.96	4.31	0.71	0.94	17.03	1.91	0.39	5.07	0.22	0.99	28.55	0.74
$CBPS_{just}$	-0.16	4.81	0.88	0.93	17.59	1.79	7.92	9.56	0.79	0.61	18.53	1.75
$CBPS_{over}$	1.37	5.12	1.00	0.98	23.54	1.00	9.03	10.74	1.00	0.74	24.51	1.00
MLE	-0.08	5.76	1.27	0.94	20.50	1.32	10.61	14.14	1.73	0.63	24.44	1.01
(b) $\mathbf{LQTE(0.10)}$												
$LIPS_{exp}$	-0.15	4.61	1.01	0.95	18.27	1.04	-1.67	4.91	1.18	0.92	18.20	1.02
$LIPS_{ind}$	-2.33	5.46	1.42	0.92	20.65	0.81	-3.97	6.24	1.92	0.87	19.74	0.87
$LIPS_{proj}$	-0.60	4.62	1.02	0.95	18.18	1.05	-0.92	5.54	1.51	0.96	24.38	0.57
$CBPS_{just}$	-0.11	4.56	0.99	0.94	17.56	1.13	-0.19	4.47	0.98	0.94	17.57	1.10
$CBPS_{over}$	0.45	4.58	1.00	0.95	18.63	1.00	0.37	4.51	1.00	0.95	18.42	1.00
MLE	-0.06	4.62	1.02	0.94	17.81	1.09	0.91	5.17	1.31	0.95	18.77	0.96
(c) $\mathbf{LQTE(0.25)}$												
$LIPS_{exp}$	-0.06	4.30	1.00	0.95	17.16	1.11	-0.42	4.22	0.83	0.96	17.40	1.06
$LIPS_{ind}$	-3.07	5.46	1.62	0.90	18.77	0.93	-3.22	5.49	1.40	0.90	18.39	0.95
$LIPS_{proj}$	-0.59	4.26	0.99	0.95	16.98	1.13	-0.86	4.78	1.06	0.96	20.57	0.76
$CBPS_{just}$	-0.01	4.17	0.94	0.95	16.59	1.19	1.26	4.37	0.89	0.94	16.57	1.17
$CBPS_{over}$	0.60	4.29	1.00	0.96	18.07	1.00	1.88	4.65	1.00	0.94	17.90	1.00
MLE	0.04	4.34	1.02	0.94	17.15	1.11	3.08	7.39	2.53	0.93	20.69	0.75
(d) $\mathbf{LQTE(0.50)}$												
$LIPS_{exp}$	-0.15	4.56	0.95	0.95	18.57	1.23	1.70	4.71	0.52	0.96	19.37	1.14
$LIPS_{ind}$	-4.92	6.60	1.99	0.85	18.68	1.22	-2.54	5.30	0.66	0.93	19.46	1.13
$LIPS_{proj}$	-0.88	4.52	0.93	0.95	18.12	1.29	-0.58	4.87	0.56	0.97	22.27	0.87
$CBPS_{just}$	-0.01	4.72	1.02	0.95	18.55	1.23	3.91	6.21	0.91	0.87	18.38	1.27
$CBPS_{over}$	0.69	4.68	1.00	0.98	20.61	1.00	4.34	6.52	1.00	0.91	20.71	1.00
MLE	0.02	5.01	1.15	0.96	19.78	1.09	7.32	13.58	4.34	0.85	25.71	0.65
(e) $\mathbf{LQTE(0.75)}$												
$LIPS_{exp}$	-0.55	6.12	0.95	0.94	23.71	1.33	4.98	7.47	0.50	0.93	25.44	1.23
$LIPS_{ind}$	-7.68	9.33	2.20	0.69	21.30	1.65	-1.64	6.02	0.33	0.95	25.17	1.25
$LIPS_{proj}$	-1.44	6.01	0.91	0.93	22.79	1.44	-0.23	6.05	0.33	0.98	33.36	0.71
$CBPS_{just}$	-0.34	6.57	1.09	0.93	24.47	1.25	7.64	10.16	0.93	0.76	24.34	1.34
$CBPS_{over}$	0.65	6.29	1.00	0.97	27.33	1.00	8.09	10.55	1.00	0.83	28.17	1.00
MLE	-0.35	7.78	1.53	0.94	26.83	1.04	12.79	19.96	3.58	0.76	32.98	0.73
(f) $\mathbf{LQTE(0.90)}$												
$LIPS_{exp}$	-1.17	9.65	1.00	0.94	39.29	1.16	10.14	13.89	0.61	0.86	37.37	1.15
$LIPS_{ind}$	-10.68	12.95	1.80	0.65	27.73	2.33	0.41	9.23	0.27	0.95	37.83	1.12
$LIPS_{proj}$	-2.11	9.30	0.93	0.93	34.13	1.54	0.88	9.42	0.28	0.99	54.96	0.53
$CBPS_{just}$	-1.14	11.01	1.30	0.92	39.31	1.16	13.04	17.22	0.94	0.70	34.93	1.32
$CBPS_{over}$	1.12	9.65	1.00	0.96	42.35	1.00	13.60	17.77	1.00	0.75	40.09	1.00
MLE	-1.84	15.04	2.43	0.92	48.55	0.76	16.79	23.64	1.77	0.63	36.97	1.18

Note: Simulations based on 1,000 Monte Carlo experiments. Bias, Monte Carlo Bias; RMSE, Monte Carlo root mean square error; relMSE, relative Monte Carlo mean square error; COV, Monte Carlo coverage of 95% normal confidence interval; ACIL, Monte Carlo average of 95% normal confidence interval length; ARE, asymptotic relative efficiency; LATE, local average treatment effect; LQTE(τ), local quantile treatment effect at τ quantile. Both relMSE and ARE are expressed with respect to the IPW estimator based on the over-identified CBPS. All instrument propensity scores is based on a logistic link function. $LIPS_{ind}$, $LIPS_{proj}$ and $LIPS_{exp}$ are the IPW estimators based on LIPS estimator (4.5) with the indicator, projection, and exponential weight function, respectively; $CBPS_{just}$, IPW estimator based on the (just-identified) CBPS estimator with moment equation (2.2), with Z in the place of D and $f(\mathbf{X}^*) = \mathbf{X}^*$; $CBPS_{over}$, IPW estimator based on the (over-identified) CBPS estimator with moment equation (2.2), with Z in the place of D and $f(\mathbf{X}^*) = \left(\mathbf{X}^{*\prime}, \dot{p}(\mathbf{X}^*; \boldsymbol{\beta})'\right)'$, with $\dot{p}(\mathbf{X}^*; \boldsymbol{\beta})$ the derivative of the instrument propensity score model with respect to $\boldsymbol{\beta}$; MLE, IPW estimator based on MLE. $\mathbf{X}^* = \mathbf{X}$ when the instrument PS model is correctly specified, and $\mathbf{X}^* = \mathbf{W}$ when the instrument PS is misspecified.

Table S1.4: Monte Carlo study of the performance of distributional covariate imbalances based on different instrument propensity score estimation methods. Sample size: $n = 500$.

	Correctly Specified						Misspecified					
	KS^{lte}_{bal}	$RCvM^{lte}_{bal}$	KS^{lte}_{bal1}	$RCvM^{lte}_{bal1}$	KS^{lte}_{bal0}	$RCvM^{lte}_{bal0}$	KS^{lte}_{bal}	$RCvM^{lte}_{bal}$	KS^{lte}_{bal1}	$RCvM^{lte}_{bal1}$	KS^{lte}_{bal0}	$RCvM^{lte}_{bal0}$
$LIPS_{exp}$	12.06	2.83	6.68	1.29	7.59	1.92	12.99	3.28	8.06	1.71	7.41	1.89
$LIPS_{ind}$	10.07	1.93	7.36	1.28	4.96	1.01	10.41	2.45	7.67	1.58	5.80	1.27
$LIPS_{proj}$	11.81	2.75	6.56	1.27	7.34	1.84	18.42	4.68	10.63	2.31	9.17	2.63
$CBPS_{just}$	12.58	2.94	7.01	1.34	8.00	2.01	14.37	3.68	8.45	1.84	8.05	2.14
$CBPS_{over}$	12.59	2.90	6.65	1.28	7.93	1.96	14.28	3.64	8.20	1.74	8.09	2.15
MLE	12.43	2.94	6.88	1.27	8.17	2.11	16.20	3.92	11.17	2.43	7.79	2.04

Note: Simulations based on 1,000 Monte Carlo experiments. $KS^{lte}_{bal}, KS^{lte}_{bal1}, KS^{lte}_{bal0}, RCvM^{lte}_{bal}, RCvM^{lte}_{bal1}, RCvM^{lte}_{bal0}$ as defined in 5.2. All imbalances are presented in percentage points. The instrument propensity score model is based on a logistic link function. $LIPS_{ind}$, $LIPS_{proj}$ and $LIPS_{exp}$ are the IPW estimators based on LIPS estimator (4.5) with the indicator, projection, and exponential weight function, respectively; $CBPS_{just}$, IPW estimator based on the (just-identified) CBPS estimator with moment equation (2.2), with Z in the place of D and $f(\mathbf{X}^*) = \mathbf{X}^*$; $CBPS_{over}$, IPW estimator based on the (over-identified) CBPS estimator with moment equation (2.2), with Z in the place of D and $f(\mathbf{X}^*) = \left(\mathbf{X}^{*,\prime}, \dot{p}(\mathbf{X}^*; \boldsymbol{\beta})'\right)'$, with $\dot{p}(\mathbf{X}^*; \boldsymbol{\beta})$ the derivative of the instrument propensity score model with respect to $\boldsymbol{\beta}$; MLE , IPW estimator based on MLE. $\mathbf{X}^* = \mathbf{X}$ when the instrument PS model is correctly specified, and $\mathbf{X}^* = \mathbf{W}$ when the instrument PS is misspecified.

S2 Simulation results for empirically calibrated DGP

This section display all the simulations results discussed in Section 5.2 of the main text.

Table S2.1: Empirically calibrated Monte Carlo study of the performance of IPW estimators for ITT parameters. Sample size: $n = 1,000$.

	Correctly Specified Model						Misspecified Model					
	Bias	RMSE	relMSE	COV	ACIL	ARE	Bias	RMSE	relMSE	COV	ACIL	ARE
(a) **ATE**												
IPS_{exp}	0.93	2.27	1.53	0.99	11.17	0.80	-5.07	6.68	6.85	0.76	12.72	1.03
IPS_{ind}	0.68	4.13	5.05	1.00	36.60	0.07	-5.13	7.57	8.80	0.99	26.40	0.24
IPS_{proj}	0.12	1.66	0.82	1.00	19.66	0.26	-0.89	2.46	0.93	0.95	7.91	2.67
$CBPS_{just}$	0.38	1.62	0.77	0.87	4.81	4.32	-3.68	5.04	3.90	0.65	8.74	2.19
$CBPS_{over}$	1.01	1.84	1.00	1.00	9.99	1.00	-0.46	2.55	1.00	1.00	12.93	1.00
MLE	0.11	1.56	0.72	0.93	5.45	3.36	-6.42	12.65	24.56	0.88	19.17	0.45
(b) **QTE(0.10)**												
IPS_{exp}	0.17	1.25	0.92	0.93	4.69	1.16	0.93	1.45	0.64	0.80	4.08	1.10
IPS_{ind}	0.52	1.37	1.10	0.92	5.12	0.97	0.69	1.52	0.70	0.85	4.91	0.76
IPS_{proj}	0.16	1.33	1.05	0.91	4.73	1.14	1.27	1.63	0.82	0.71	3.91	1.20
$CBPS_{just}$	0.20	1.45	1.24	0.89	4.66	1.17	1.17	1.59	0.78	0.75	4.08	1.10
$CBPS_{over}$	0.49	1.30	1.00	0.93	5.05	1.00	1.47	1.81	1.00	0.69	4.27	1.00
MLE	0.08	1.47	1.28	0.93	5.23	0.93	1.13	1.58	0.77	0.78	4.24	1.01
(b) **QTE(0.25)**												
IPS_{exp}	0.02	0.81	0.79	0.98	3.76	1.47	0.56	0.93	0.38	0.95	3.53	1.39
IPS_{ind}	0.26	0.95	1.08	0.99	4.59	0.98	0.30	0.92	0.37	0.98	4.48	0.86
IPS_{proj}	0.11	0.84	0.86	0.97	3.74	1.48	0.89	1.16	0.59	0.88	3.50	1.41
$CBPS_{just}$	0.22	1.01	1.23	0.94	3.81	1.43	0.80	1.11	0.54	0.90	3.60	1.33
$CBPS_{over}$	0.39	0.91	1.00	0.98	4.55	1.00	1.26	1.51	1.00	0.83	4.15	1.00
MLE	0.06	0.94	1.07	0.98	4.19	1.18	0.75	1.09	0.52	0.92	3.73	1.24
(b) **QTE(0.50)**												
IPS_{exp}	-0.12	0.99	0.75	0.96	4.09	1.80	0.31	0.97	0.23	0.96	3.77	1.92
IPS_{ind}	0.16	1.27	1.24	0.99	6.21	0.78	0.03	1.30	0.42	0.98	6.27	0.70
IPS_{proj}	0.01	1.00	0.77	0.96	4.31	1.62	0.95	1.34	0.45	0.87	3.81	1.89
$CBPS_{just}$	0.25	1.22	1.15	0.91	3.91	1.97	0.71	1.22	0.37	0.90	3.82	1.88
$CBPS_{over}$	0.50	1.14	1.00	0.98	5.48	1.00	1.64	1.99	1.00	0.82	5.24	1.00
MLE	0.05	1.17	1.05	0.95	4.48	1.49	0.55	1.15	0.33	0.92	3.97	1.74
(b) **QTE(0.75)**												
IPS_{exp}	-0.25	2.60	0.75	0.95	11.14	1.79	-1.32	2.83	0.45	0.92	10.51	2.01
IPS_{ind}	-0.08	4.66	2.41	0.98	22.92	0.42	-1.92	5.38	1.63	0.97	23.23	0.41
IPS_{proj}	-0.18	2.71	0.82	0.96	12.58	1.40	1.20	2.81	0.44	0.91	9.36	2.54
$CBPS_{just}$	0.21	2.94	0.96	0.84	8.53	3.05	-0.23	2.62	0.39	0.90	8.96	2.77
$CBPS_{over}$	1.37	3.00	1.00	0.98	14.90	1.00	3.12	4.22	1.00	0.94	14.92	1.00
MLE	0.08	3.13	1.09	0.88	9.10	2.68	-1.51	3.85	0.83	0.91	11.20	1.77
(b) **QTE(0.90)**												
IPS_{exp}	0.34	5.14	0.97	0.98	26.32	1.02	-8.59	10.61	4.07	0.78	26.41	1.17
IPS_{ind}	0.12	9.41	3.26	0.99	57.69	0.21	-9.13	13.67	6.76	0.96	53.04	0.29
IPS_{proj}	-0.74	5.19	0.99	1.00	35.70	0.56	-1.11	5.46	1.08	0.93	21.30	1.79
$CBPS_{just}$	-0.28	5.30	1.03	0.92	19.24	1.91	-6.19	8.55	2.64	0.80	22.11	1.66
$CBPS_{over}$	1.45	5.21	1.00	0.98	26.61	1.00	-0.06	5.26	1.00	0.99	28.51	1.00
MLE	-0.25	5.01	0.92	0.93	18.83	2.00	-12.17	41.76	63.09	0.82	35.39	0.65

Note: Simulations based on 1,000 Monte Carlo experiments. Bias, Monte Carlo Bias; RMSE, Monte Carlo root mean square error; relMSE, relative Monte Carlo mean square error; COV, Monte Carlo coverage of 95% normal confidence interval; ACIL, Monte Carlo average of 95% normal confidence interval length; ARE, asymptotic relative efficiency; ATE, average treatment effect; QTE(τ), quantile treatment effect at τ quantile. Both relMSE and ARE are expressed with respect to the IPW estimator based on the over-identified CBPS. The propensity score model is based on a logistic link function. IPS_{ind}, IPW estimator based on IPS estimator (2.10); IPS_{proj}, IPW estimator based on IPS estimator (2.11); IPS_{exp}, IPW estimator based on IPS estimator (2.12); $CBPS_{just}$, IPW estimator based on the (just-identified) CBPS estimator with moment equation (2.2), with $f(\mathbf{X}^*) = \mathbf{X}^*$; $CBPS_{over}$, IPW estimator based on the (over-identified) CBPS estimator with moment equation (2.2), with $f(\mathbf{X}^*) = \left(\mathbf{X}^{*\prime}, \dot{p}(\mathbf{X}^*;\boldsymbol{\beta})'\right)'$, with $\dot{p}(\mathbf{X}^*;\boldsymbol{\beta})$ the derivative of the propensity score model with respect to $\boldsymbol{\beta}$; MLE , IPW estimator based on MLE. $\mathbf{X}^*$ includes all relevant interaction terms when the PS model is correctly specified, and $\mathbf{X}^*$ does not include any interaction term when the PS is misspecified.

Table S2.2: Empirically calibrated Monte Carlo study of the performance of distributional covariate imbalances based on different propensity score estimation methods. Sample size: $n = 1,000$.

	Correctly Specified						Misspecified					
	KS_{bal}	$RCvM_{bal}$	KS_{bal1}	$RCvM_{bal1}$	KS_{bal0}	$RCvM_{bal0}$	KS_{bal}	$RCvM_{bal}$	KS_{bal1}	$RCvM_{bal1}$	KS_{bal0}	$RCvM_{bal0}$
IPS_{exp}	6.22	1.69	4.90	1.42	2.21	0.40	7.72	1.93	5.44	1.45	3.12	0.58
IPS_{ind}	5.73	1.35	4.23	1.07	2.59	0.39	7.39	1.59	5.02	1.21	3.56	0.49
IPS_{proj}	6.99	1.92	5.57	1.64	2.27	0.41	8.07	1.99	5.83	1.51	2.77	0.55
$CBPS_{just}$	7.00	1.97	5.77	1.74	2.40	0.43	7.75	1.97	5.94	1.61	3.19	0.50
$CBPS_{over}$	7.00	1.92	5.47	1.60	2.22	0.43	9.26	2.48	7.00	1.95	2.73	0.58
MLE	7.31	2.10	6.00	1.85	2.16	0.40	7.86	1.98	6.02	1.66	3.71	0.57

Note: Simulations based on 1,000 Monte Carlo experiments. $KS_{bal}, KS_{bal1}, KS_{bal0}, RCvM_{bal}, RCvM_{bal1}, RCvM_{bal0}$ as defined in 5.1. All imbalances are presented in percentage points. The propensity score model is based on a logistic link function. IPS_{ind}, IPW estimator based on IPS estimator (2.10); IPS_{proj}, IPW estimator based on IPS estimator (2.11); IPS_{exp}, IPW estimator based on IPS estimator (2.12); $CBPS_{just}$, IPW estimator based on the (just-identified) CBPS estimator with moment equation (2.2), with $f(\mathbf{X}^*) = \mathbf{X}^*$; $CBPS_{over}$, IPW estimator based on the (over-identified) CBPS estimator with moment equation (2.2), with $f(\mathbf{X}^*) = \left(\mathbf{X}^{*,\prime}, \dot{p}(\mathbf{X}^*; \boldsymbol{\beta})^\prime\right)^\prime$, with $\dot{p}(\mathbf{X}^*; \boldsymbol{\beta})$ the derivative of the propensity score model with respect to $\boldsymbol{\beta}$; MLE , IPW estimator based on MLE. $\mathbf{X}^*$ includes all relevant interaction terms when the PS model is correctly specified, and $\mathbf{X}^*$ does not include any interaction term when the PS is misspecified.

Table S2.3: Empirically calibrated Monte Carlo study of the performance of IPW estimators for local treatment effect parameters. Sample size: $n = 1,000$.

	Correctly Specified Model						Misspecified Model					
	Bias	RMSE	relMSE	COV	ACIL	ARE	Bias	RMSE	relMSE	COV	ACIL	ARE
(a) **LATE**												
$LIPS_{exp}$	0.92	2.65	1.22	1.00	16.20	0.72	-5.75	7.72	5.33	0.92	18.00	0.93
$LIPS_{ind}$	2.88	5.39	5.07	1.00	62.88	0.05	-3.00	5.68	2.88	1.00	30.40	0.33
$LIPS_{proj}$	0.03	2.20	0.85	1.00	28.32	0.23	-0.23	2.88	0.74	0.97	10.91	2.53
$CBPS_{just}$	0.55	2.21	0.85	0.88	6.70	4.18	-5.09	6.85	4.20	0.66	12.01	2.09
$CBPS_{over}$	1.25	2.39	1.00	1.00	13.70	1.00	-0.74	3.34	1.00	1.00	17.37	1.00
MLE	0.13	1.99	0.69	0.95	7.61	3.24	-8.37	16.01	22.92	0.86	24.28	0.51
(b) **LQTE(0.10)**												
$LIPS_{exp}$	0.22	1.62	0.98	0.92	5.99	1.07	1.22	1.89	0.83	0.81	5.42	0.96
$LIPS_{ind}$	0.68	1.69	1.07	0.92	6.47	0.92	1.21	1.87	0.81	0.83	5.77	0.84
$LIPS_{proj}$	0.20	1.71	1.09	0.91	6.03	1.06	1.47	1.97	0.90	0.74	5.01	1.12
$CBPS_{just}$	0.33	1.88	1.32	0.87	5.92	1.10	1.32	1.91	0.84	0.78	5.26	1.02
$CBPS_{over}$	0.56	1.64	1.00	0.91	6.20	1.00	1.60	2.07	1.00	0.73	5.31	1.00
MLE	0.15	1.90	1.34	0.91	6.42	0.93	1.27	1.90	0.83	0.80	5.40	0.97
(b) **LQTE(0.25)**												
$LIPS_{exp}$	-0.01	1.10	0.90	0.98	4.94	1.38	0.81	1.29	0.52	0.95	4.88	1.13
$LIPS_{ind}$	0.36	1.21	1.09	0.98	5.97	0.94	0.74	1.28	0.51	0.96	5.51	0.89
$LIPS_{proj}$	0.05	1.15	0.99	0.97	4.94	1.38	1.16	1.51	0.72	0.88	4.62	1.26
$CBPS_{just}$	0.21	1.30	1.26	0.94	5.03	1.33	0.94	1.37	0.58	0.92	4.74	1.20
$CBPS_{over}$	0.40	1.15	1.00	0.99	5.80	1.00	1.46	1.79	1.00	0.86	5.19	1.00
MLE	0.01	1.29	1.25	0.97	5.45	1.13	0.87	1.33	0.56	0.93	4.84	1.15
(b) **LQTE(0.50)**												
$LIPS_{exp}$	-0.10	1.31	0.81	0.96	5.60	1.65	0.76	1.52	0.36	0.93	5.39	1.54
$LIPS_{ind}$	0.49	1.62	1.22	0.99	8.43	0.73	0.76	1.69	0.44	0.98	7.96	0.71
$LIPS_{proj}$	-0.03	1.35	0.85	0.95	5.76	1.55	1.52	1.98	0.61	0.84	5.19	1.67
$CBPS_{just}$	0.30	1.59	1.18	0.91	5.29	1.84	0.99	1.64	0.42	0.90	5.17	1.68
$CBPS_{over}$	0.62	1.46	1.00	0.99	7.18	1.00	2.12	2.54	1.00	0.82	6.70	1.00
MLE	0.06	1.50	1.05	0.95	5.76	1.55	0.76	1.54	0.37	0.93	5.35	1.57
(b) **LQTE(0.75)**												
$LIPS_{exp}$	-0.40	3.46	0.79	0.96	15.62	1.67	-0.76	3.56	0.41	0.98	16.52	1.48
IPS_{ind}	1.53	5.41	1.92	0.99	33.42	0.37	0.22	5.40	0.95	1.00	32.02	0.39
IPS_{proj}	-0.56	3.56	0.83	0.96	16.95	1.42	2.67	4.27	0.59	0.88	13.24	2.30
$CBPS_{just}$	0.07	3.83	0.96	0.88	12.04	2.82	-0.19	3.53	0.40	0.93	13.03	2.38
$CBPS_{over}$	1.57	3.90	1.00	0.99	20.20	1.00	4.29	5.56	1.00	0.93	20.09	1.00
MLE	-0.20	3.66	0.88	0.89	14.14	2.04	-2.02	5.49	0.98	0.94	17.64	1.30
(b) **LQTE(0.90)**												
$LIPS_{exp}$	0.40	6.92	0.91	0.97	37.11	1.00	-11.40	15.14	4.01	0.84	54.44	0.67
$LIPS_{ind}$	4.37	10.89	2.25	0.99	86.01	0.19	-5.70	13.02	2.97	0.97	74.52	0.36
$LIPS_{proj}$	-0.90	7.01	0.93	0.99	49.40	0.56	0.52	7.55	1.00	0.93	33.27	1.78
$CBPS_{just}$	0.05	7.21	0.98	0.91	29.67	1.56	-10.05	13.70	3.28	0.78	37.60	1.40
$CBPS_{over}$	2.12	7.27	1.00	0.98	37.08	1.00	-0.31	7.56	1.00	0.98	44.45	1.00
MLE	0.08	6.94	0.91	0.92	26.36	1.98	-21.49	70.32	86.51	0.77	52.98	0.70

Note: Simulations based on 1,000 Monte Carlo experiments. Bias, Monte Carlo Bias; RMSE, Monte Carlo root mean square error; relMSE, relative Monte Carlo mean square error; COV, Monte Carlo coverage of 95% normal confidence interval; ACIL, Monte Carlo average of 95% normal confidence interval length; ARE, asymptotic relative efficiency; LATE, local average treatment effect; LQTE(τ), local quantile treatment effect at τ quantile. Both relMSE and ARE are expressed with respect to the IPW estimator based on the over-identified CBPS. All instrument propensity scores is based on a logistic link function. $LIPS_{ind}$, $LIPS_{proj}$ and $LIPS_{exp}$ are the IPW estimators based on LIPS estimator (4.5) with the indicator, projection, and exponential weight function, respectively; $CBPS_{just}$, IPW estimator based on the (just-identified) CBPS estimator with moment equation (2.2), with Z in the place of D and $f(\mathbf{X}^*) = \mathbf{X}^*$; $CBPS_{over}$, IPW estimator based on the (over-identified) CBPS estimator with moment equation (2.2), with Z in the place of D and $f(\mathbf{X}^*) = \left(\mathbf{X}^{*\prime}, \dot{p}(\mathbf{X}^*;\boldsymbol{\beta})'\right)'$, with $\dot{p}(\mathbf{X}^*;\boldsymbol{\beta})$ the derivative of the instrument propensity score model with respect to $\boldsymbol{\beta}$; MLE , IPW estimator based on MLE. $\mathbf{X}^*$ includes all relevant interaction terms when the PS model is correctly specified, and $\mathbf{X}^*$ does not include any interaction term when the PS is misspecified.

Table S2.4: Empirically calibrated Monte Carlo study of the performance of distributional covariate imbalances based on different instrument propensity score estimation methods. Sample size: $n = 1,000$.

	Correctly Specified						Misspecified					
	KS_{bal}^{lte}	$RCvM_{bal}^{lte}$	KS_{bal1}^{lte}	$RCvM_{bal1}^{lte}$	KS_{bal0}^{lte}	$RCvM_{bal0}^{lte}$	KS_{bal}^{lte}	$RCvM_{bal}^{lte}$	KS_{bal1}^{lte}	$RCvM_{bal1}^{lte}$	KS_{bal0}^{lte}	$RCvM_{bal0}^{lte}$
$LIPS_{exp}$	8.85	2.41	7.03	2.05	3.15	0.57	11.48	2.89	8.15	2.20	4.72	0.80
$LIPS_{ind}$	7.85	1.88	6.05	1.54	3.59	0.57	10.46	2.27	7.51	1.81	4.95	0.73
$LIPS_{proj}$	10.19	2.79	8.08	2.39	3.25	0.59	11.54	2.89	8.52	2.24	3.91	0.74
$CBPS_{just}$	9.98	2.84	8.25	2.54	3.43	0.61	10.90	2.80	8.32	2.30	4.54	0.70
$CBPS_{over}$	9.69	2.66	7.61	2.22	3.09	0.59	12.51	3.43	9.51	2.72	3.71	0.78
MLE	10.32	2.98	8.48	2.62	3.08	0.57	10.90	2.78	8.35	2.35	5.12	0.79

Note: Simulations based on 1,000 Monte Carlo experiments. $KS_{bal}^{lte}, KS_{bal1}^{lte}, KS_{bal0}^{lte}, RCvM_{bal}^{lte}, RCvM_{bal1}^{lte}, RCvM_{bal0}^{lte}$ as defined in 5.2. All imbalances are presented in percentage points. The instrument propensity score model is based on a logistic link function. $LIPS_{ind}$, $LIPS_{proj}$ and $LIPS_{exp}$ are the IPW estimators based on LIPS estimator (4.5) with the indicator, projection, and exponential weight function, respectively; $CBPS_{just}$, IPW estimator based on the (just-identified) CBPS estimator with moment equation (2.2), with Z in the place of D and $f(\mathbf{X}^*) = \mathbf{X}^*$; $CBPS_{over}$, IPW estimator based on the (over-identified) CBPS estimator with moment equation (2.2), with Z in the place of D and $f(\mathbf{X}^*) = (\mathbf{X}^{*,\prime}, \dot{p}(\mathbf{X}^*;\boldsymbol{\beta})')'$, with $\dot{p}(\mathbf{X}^*;\boldsymbol{\beta})$ the derivative of the instrument propensity score model with respect to $\boldsymbol{\beta}$; MLE, IPW estimator based on MLE. $\mathbf{X}^*$ includes all relevant interaction terms when the PS model is correctly specified, and $\mathbf{X}^*$ does not include any interaction term when the PS is misspecified.

S3 Closed-form representation of the IPS objective functions

In this section, we derive the closed-form representation of the IPS objective function $Q_{n,w}(\boldsymbol{\beta})$ using the three families of weighting functions under Assumption 3 in the main text.

First, recall from Section 2.2 that

$$\widehat{\boldsymbol{\beta}}_{n,w}^{ips} = \arg\min_{\boldsymbol{\beta}\in\Theta} Q_{n,w}(\boldsymbol{\beta}),$$

where $Q_{n,w}(\boldsymbol{\beta}) = \int \|\mathbf{H}_{n,w}(\boldsymbol{\beta},\mathbf{u})\|^2 \Psi_n(d\mathbf{u})$, Ψ_n is an uniformly consistent estimator of Ψ, $\mathbf{H}_{n,w}(\boldsymbol{\beta},\mathbf{u}) = \mathbb{E}_n[\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta})\, w(\mathbf{X};\mathbf{u})]$, with $\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta}) = (h_{n,1}(D,\mathbf{X};\boldsymbol{\beta}), h_{n,0}(D,\mathbf{X};\boldsymbol{\beta}))'$, $h_{n,d}(D,\mathbf{X};\boldsymbol{\beta}) = \varpi_{n,d}^{ps}(D,\mathbf{X};\boldsymbol{\beta}) - 1$, $d\in\{0,1\}$, and

$$\varpi_{n,1}^{ps}(D,\mathbf{X};\boldsymbol{\beta}) = \frac{D}{p(\mathbf{X};\boldsymbol{\beta})}\Bigg/\mathbb{E}_n\left[\frac{D}{p(\mathbf{X};\boldsymbol{\beta})}\right],$$

$$\varpi_{n,0}^{ps}(D,\mathbf{X};\boldsymbol{\beta}) = \frac{1-D}{1-p(\mathbf{X};\boldsymbol{\beta})}\Bigg/\mathbb{E}_n\left[\frac{1-D}{1-p(\mathbf{X};\boldsymbol{\beta})}\right].$$

In the following, we derive the closed-form representation of the IPS objective function of the estimators in (2.10)-(2.12). In light of Remark 3.2, we emphasize the role played by $h_{n,1}$ and $h_{n,0}$ in the computation of the objective functions.

Case 1: Indicator Weights

As introduced in (2.10), in this case we have $\Psi_n(\mathbf{u}) = F_{n,\mathbf{X}}(\mathbf{u})$ and $w(\mathbf{X};\mathbf{u}) = 1\{\mathbf{X}\le\mathbf{u}\}$. For any $\mathbf{g}(\mathbf{x})$, $\int \mathbf{g}(\mathbf{u}) F_{n,\mathbf{X}}(d\mathbf{u}) = n^{-1}\sum_{j=1}^n \mathbf{g}(\mathbf{X}_j)$. Since $1\{\mathbf{X}\le\mathbf{u}\}$ is real-valued, conjugate transpose reduces to direct transpose. Hence,

$$
\begin{aligned}
\mathbf{H}_{n,ind}\left(\boldsymbol{\beta},\mathbf{u}\right) &= \frac{}{1} n \sum_{j=1}^{n} \mathbf{h}_n\left(D_j,\mathbf{X}_j;\boldsymbol{\beta}\right) 1\left(\mathbf{X}_j \leq \mathbf{u}\right) \\
&= \frac{1}{n} \sum_{j=1}^{n} \left(h_{n,1}\left(D_j,\mathbf{X}_j;\boldsymbol{\beta}\right), h_{n,0}\left(D_j,\mathbf{X}_j;\boldsymbol{\beta}\right)\right)' 1\left(\mathbf{X}_j \leq \mathbf{u}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
Q_{n,ind}\left(\boldsymbol{\beta}\right) &= \int_{[-\infty,\infty]^k} \left\|\mathbb{E}_n\left[\mathbf{h}_n\left(D,\mathbf{X};\boldsymbol{\beta}\right) 1(\mathbf{X} \leq \mathbf{u})\right]\right\|^2 \, F_{n,\mathbf{X}}(d\mathbf{u}); \\
&= \frac{1}{n^3} \sum_{l=1}^{n} \left(\sum_{j=1}^{n} \mathbf{h}_n\left(D_j,\mathbf{X}_j;\boldsymbol{\beta}\right) 1\left(\mathbf{X}_j \leq \mathbf{X}_l\right) \right)' \left(\sum_{j=1}^{n} \mathbf{h}_n\left(D_j,\mathbf{X}_j;\boldsymbol{\beta}\right) 1\left(\mathbf{X}_j \leq \mathbf{X}_l\right) \right) \\
&= \frac{1}{n^3} \sum_{l=1}^{n} \left(\sum_{j=1}^{n} h_{n,1}\left(D_j,\mathbf{X}_j;\boldsymbol{\beta}\right) 1\left(\mathbf{X}_j \leq \mathbf{X}_l\right) \right)^2 \\
&+ \frac{1}{n^3} \sum_{l=1}^{n} \left(\sum_{j=1}^{n} h_{n,0}\left(D_j,\mathbf{X}_j;\boldsymbol{\beta}\right) 1\left(\mathbf{X}_j \leq \mathbf{X}_l\right) \right)^2 .
\end{aligned}
$$

Case 2: Projection Weights

As introduced in (2.11), in this case we have $\Psi_n\left(\mathbf{u}\right) = n^{-1} \sum_{i=1}^{n} 1\left(\boldsymbol{\gamma}'\mathbf{X}_j \leq u\right) \times \boldsymbol{\gamma}$ and $w\left(\mathbf{X};\mathbf{u}\right) = 1\{\boldsymbol{\gamma}'\mathbf{X} \leq u\}$. Before proceeding to the detailed derivations, from Appendix B of Escanciano (2006), we have that

$$
\begin{aligned}
&\int_{[-\infty,\infty]\times\mathbb{S}_k} 1\left(\boldsymbol{\gamma}'\mathbf{X}_j \leq u\right) 1\left(\boldsymbol{\gamma}'\mathbf{X}_s \leq u\right) F_{n,\boldsymbol{\gamma}'\mathbf{X}}\left(du\right) d\boldsymbol{\gamma} \\
&= \frac{1}{n} \sum_{r=1}^{n} \int_{\mathbb{S}_k} 1\left(\boldsymbol{\gamma}'\mathbf{X}_j \leq \boldsymbol{\gamma}'\mathbf{X}_r\right) 1\left(\boldsymbol{\gamma}'\mathbf{X}_s \leq \boldsymbol{\gamma}'\mathbf{X}_r\right) d\boldsymbol{\gamma} \\
&= \frac{1}{n} \sum_{r=1}^{n} A_{jsr} \\
&\equiv A_{js},
\end{aligned}
$$

where A_{jsr} is proportional to the volume of a spherical wedge and can be computed as

$$
A_{jsr} \equiv A_{jsr}^{(0)} \frac{\pi^{(k/2)-1}}{\Gamma\left(\dfrac{k}{2}\right)},
$$

where $\Gamma(\cdot)$ is the gamma function and

$$A^{(0)}_{jsr} = \begin{cases} 2\pi & if\ X_i = X_r = X_j, \\ \pi & if\ X_i = X_j,\ X_i = X_r\ or\ X_j = X_r, \\ \left| \pi - \arccos\left(\frac{(X_i - X_r)'(X_j - X_r)}{\|X_i - X_r\| \|X_j - X_r\|} \right) \right| & otherwise. \end{cases}$$

With this result in hand, and the fact that $1(\boldsymbol{\gamma}'\mathbf{X} \leq u)$ is real-valued, the objective function $Q_{n,proj}(\boldsymbol{\beta})$ can be written as

$$\begin{aligned} Q_{n,proj}(\boldsymbol{\beta}) &= \int_{[-\infty,\infty]\times\mathbb{S}_k} \left\| \mathbb{E}_n \left[\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta}) 1(\boldsymbol{\gamma}'\mathbf{X} \leq u) \right] \right\|^2 F_{n,\boldsymbol{\gamma}}(du) d\boldsymbol{\gamma} \\ &= \frac{1}{n^2} \int_{\mathbb{R}\times\mathbb{S}_k} \left(\sum_{j=1}^n h_{n,1}(D_j,\mathbf{X}_j;\boldsymbol{\beta}) 1\left(\boldsymbol{\gamma}'\mathbf{X}_j \leq u\right) \right)^2 F_{n,\boldsymbol{\gamma}'\mathbf{X}}(du)\, d\boldsymbol{\gamma} \\ &+ \frac{1}{n^2} \int_{\mathbb{R}\times\mathbb{S}_k} \left(\sum_{j=1}^n h_{n,0}(D_j,\mathbf{X}_j;\boldsymbol{\beta}) 1\left(\boldsymbol{\gamma}'\mathbf{X}_j \leq u\right) \right)^2 F_{n,\boldsymbol{\gamma}'\mathbf{X}}(du)\, d\boldsymbol{\gamma} \\ &= \frac{1}{n^2} \sum_{j=1}^n \sum_{s=1}^n h_{n,1}(D_j,\mathbf{X}_j;\boldsymbol{\beta}) h_{n,1}(D_s,\mathbf{X}_s;\boldsymbol{\beta}) A_{js} \\ &+ \frac{1}{n^2} \sum_{j=1}^n \sum_{s=1}^n h_{n,0}(D_j,\mathbf{X}_j;\boldsymbol{\beta}) h_{n,0}(D_s,\mathbf{X}_s;\boldsymbol{\beta}) A_{js}. \end{aligned}$$

Case 3: Exponential Weights

Finally, as introduced in (2.12), now we have $w(\mathbf{X};\mathbf{u}) = exp(i\mathbf{u}'\Phi(\mathbf{X}))$ and

$$\Psi_n(d\mathbf{u}) = \Psi(d\mathbf{u}) = \frac{\exp(-\mathbf{u}'\mathbf{u}/2)}{(2\pi)^{k/2}} d\mathbf{u}.$$

For notational convenience, let $\mathbf{h}_{n,j}(\boldsymbol{\beta}) \equiv \mathbf{h}_n(D_j,\mathbf{X}_j;\boldsymbol{\beta})$. Hence

$$\begin{aligned} &Q_{n,exp}(\boldsymbol{\beta}) \\ &= \int_{\mathbb{R}^k} \left\| \mathbb{E}_n \left[\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta}) \exp(i\mathbf{u}'\Phi(\mathbf{X})) \right] \right\|^2 \frac{\exp\left(-\frac{1}{2}\mathbf{u}'\mathbf{u}\right)}{(2\pi)^{k/2}} d\mathbf{u} \\ &= \frac{1}{n^2} \int_{\mathbb{R}^k} \left(\sum_{j=1}^n \mathbf{h}_{n,j}(\boldsymbol{\beta}) \exp\left(-i\mathbf{u}'\Phi(\mathbf{X}_j)\right) \right)' \left(\sum_{s=1}^n \mathbf{h}_{n,s}(\boldsymbol{\beta}) \exp\left(i\mathbf{u}'\Phi(\mathbf{X}_j)\right) \right) \Psi(d\mathbf{u}) \\ &= \frac{1}{n^2} \sum_{j=1}^n \sum_{s=1}^n \left\{ \mathbf{h}'_{n,j}(\boldsymbol{\beta}) \mathbf{h}_{n,s}(\boldsymbol{\beta}) \int_{\mathbb{R}^k} \exp\left(i\mathbf{u}'\left(\Phi(\mathbf{X}_j) - \Phi(\mathbf{X}_s)\right)\right) \frac{\exp(-\mathbf{u}'\mathbf{u}/2)}{(2\pi)^{k/2}} d\mathbf{u} \right\} \end{aligned}$$

$$= \frac{1}{n^2} \sum_{j=1}^{n} \sum_{s=1}^{n} \mathbf{h}'_{n,j}(\boldsymbol{\beta}) \mathbf{h}_{n,s}(\boldsymbol{\beta}) \exp\left\{-\frac{1}{2} ||\Phi(\mathbf{X}_j) - \Phi(\mathbf{X}_s)||^2\right\}. \tag{S3.1}$$

To get the last equality, we exploit that

$$\int_{\mathbb{R}^k} \exp\left(i\mathbf{u}'\mathbf{t}\right) \cdot \frac{\exp\left(-\mathbf{u}'\mathbf{u}/2\right)}{(2\pi)^{k/2}} d\mathbf{u} = \mathbb{E}_{\mathbf{U}}[\exp\left(i\mathbf{U}'\mathbf{t}\right)]$$
$$= \exp\left\{-\mathbf{t}'\mathbf{t}/2\right\},$$

where we use the definition of characteristic function for the random variable $\mathbf{U}$, and exploits that $\mathbf{U}$ follows a standard k-variate normal distribution. Letting $\mathbf{t} = \Phi(\mathbf{X}_j) - \Phi(\mathbf{X}_s)$, (S3.1) follows immediately.

Thus, from (S3.1) and the definition of $\mathbf{h}_{n,j}(\boldsymbol{\beta})$, we have

$$Q_{n,exp}(\boldsymbol{\beta}) = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{s=1}^{n} h_{n,1}(D_j, \mathbf{X}_j; \boldsymbol{\beta}) h_{n,1}(D_s, \mathbf{X}_s; \boldsymbol{\beta}) \exp\left\{-\frac{1}{2} ||\Phi(\mathbf{X}_j) - \Phi(\mathbf{X}_s)||^2\right\}$$
$$+ \frac{1}{n^2} \sum_{j=1}^{n} \sum_{s=1}^{n} h_{n,0}(D_j, \mathbf{X}_j; \boldsymbol{\beta}) h_{n,0}(D_s, \mathbf{X}_s; \boldsymbol{\beta}) \exp\left\{-\frac{1}{2} ||\Phi(\mathbf{X}_j) - \Phi(\mathbf{X}_s)||^2\right\}.$$

S4 Additional Monte Carlo results

In this Section, we report Monte Carlo results under our stylized setup with sample size $n = 200$ and $n = 1,000$. Like in Section 5.1 of the main text, we consider an unconfounded and a local treatment effect setup. For additional details, see Section 5.1 of the main text. Since the results are qualitatively similar to the results with $n = 500$ discussed in the main text, we avoid introducing additional redundant discussions.

Table S4.1: Monte Carlo study of the performance of IPW estimators for ATE and QTE based on different propensity score estimation methods. DGP describe in Section 5.1. Sample size: $n = 200$.

	Correctly Specified Model						Misspecified Model					
	Bias	RMSE	relMSE	COV	ACIL	ARE	Bias	RMSE	relMSE	COV	ACIL	ARE
(a) **ATE**												
IPS_{exp}	0.17	5.74	0.91	0.94	21.28	1.21	1.45	5.82	0.92	0.92	20.78	1.25
IPS_{ind}	1.85	5.88	0.96	0.94	23.26	1.01	2.27	6.38	1.10	0.95	24.98	0.87
IPS_{proj}	0.18	5.62	0.87	0.93	20.84	1.26	0.45	5.38	0.78	0.96	22.56	1.06
$CBPS_{just}$	0.20	6.44	1.14	0.92	22.16	1.11	2.27	6.47	1.13	0.90	21.34	1.19
$CBPS_{over}$	0.19	6.02	1.00	0.94	23.37	1.00	1.61	6.07	1.00	0.94	23.26	1.00
MLE	0.26	7.33	1.48	0.93	24.42	0.92	4.62	11.37	3.51	0.90	26.73	0.76
(b) **QTE(0.10)**												
IPS_{exp}	0.25	8.25	1.09	0.95	31.09	0.99	-3.95	9.13	1.12	0.92	31.74	1.00
IPS_{ind}	1.09	8.47	1.15	0.95	33.42	0.86	-2.38	9.06	1.10	0.95	36.20	0.77
IPS_{proj}	0.26	8.18	1.07	0.95	30.98	1.00	-1.53	8.77	1.03	0.98	42.61	0.56
$CBPS_{just}$	0.30	7.94	1.01	0.94	29.92	1.07	-3.41	9.09	1.11	0.91	31.22	1.04
$CBPS_{over}$	0.24	7.91	1.00	0.96	30.94	1.00	-3.04	8.62	1.00	0.95	31.81	1.00
MLE	0.36	7.98	1.02	0.95	30.03	1.06	-1.72	10.12	1.38	0.93	33.08	0.92
(b) **QTE(0.25)**												
IPS_{exp}	0.08	6.87	1.03	0.95	27.55	1.03	-2.10	7.16	1.11	0.94	27.23	1.06
IPS_{ind}	1.35	7.11	1.11	0.96	29.91	0.87	-0.68	6.83	1.01	0.96	30.44	0.84
IPS_{proj}	0.06	6.83	1.02	0.95	27.30	1.05	-0.83	6.96	1.05	0.98	33.69	0.69
$CBPS_{just}$	-0.02	6.89	1.04	0.95	27.15	1.06	-1.57	7.13	1.10	0.94	27.39	1.04
$CBPS_{over}$	0.08	6.77	1.00	0.96	27.94	1.00	-1.45	6.79	1.00	0.95	27.98	1.00
MLE	-0.06	7.19	1.13	0.95	27.75	1.01	0.36	10.39	2.34	0.94	30.84	0.82
(b) **QTE(0.50)**												
IPS_{exp}	0.24	6.82	0.94	0.95	27.64	1.13	1.01	6.73	0.95	0.96	27.09	1.15
IPS_{ind}	1.93	6.97	0.98	0.95	29.34	1.01	1.99	7.21	1.09	0.96	29.93	0.95
IPS_{proj}	0.28	6.70	0.91	0.95	27.20	1.17	0.24	6.46	0.87	0.97	28.75	1.02
$CBPS_{just}$	0.34	7.45	1.12	0.93	28.76	1.05	2.03	7.45	1.16	0.94	28.02	1.08
$CBPS_{over}$	0.25	7.03	1.00	0.96	29.43	1.00	1.24	6.92	1.00	0.96	29.11	1.00
MLE	0.52	8.59	1.49	0.94	30.42	0.94	4.73	13.78	3.97	0.93	34.67	0.70
(b) **QTE(0.75)**												
IPS_{exp}	0.31	8.88	0.97	0.92	32.58	1.19	4.74	9.70	0.99	0.89	31.42	1.24
IPS_{ind}	2.47	8.70	0.93	0.94	33.29	1.14	4.98	9.85	1.02	0.93	35.88	0.95
IPS_{proj}	0.26	8.58	0.91	0.93	31.70	1.25	1.54	8.11	0.69	0.95	35.04	1.00
$CBPS_{just}$	0.29	9.90	1.20	0.92	34.89	1.03	5.93	10.92	1.25	0.87	32.98	1.13
$CBPS_{over}$	0.23	9.02	1.00	0.95	35.47	1.00	4.53	9.76	1.00	0.93	35.00	1.00
MLE	0.54	11.25	1.56	0.92	38.33	0.86	9.47	17.88	3.35	0.85	39.96	0.77
(b) **QTE(0.90)**												
IPS_{exp}	-0.02	14.20	1.10	0.85	41.68	1.13	8.97	16.32	1.07	0.79	40.41	1.14
IPS_{ind}	2.91	13.10	0.94	0.88	41.67	1.13	8.66	16.67	1.12	0.85	46.67	0.86
IPS_{proj}	0.03	13.35	0.98	0.87	40.75	1.19	3.19	12.70	0.65	0.93	51.37	0.71
$CBPS_{just}$	0.16	15.52	1.32	0.82	42.26	1.10	9.79	17.03	1.16	0.76	39.93	1.17
$CBPS_{over}$	0.22	13.52	1.00	0.89	44.39	1.00	7.93	15.78	1.00	0.84	43.19	1.00
MLE	-0.09	17.51	1.68	0.81	44.44	1.00	12.29	21.74	1.90	0.69	38.76	1.24

Note: Simulations based on 1,000 Monte Carlo experiments. Bias, Monte Carlo Bias; RMSE, Monte Carlo root mean square error; relMSE, relative Monte Carlo mean square error; COV, Monte Carlo coverage of 95% normal confidence interval; ACIL, Monte Carlo average of 95% normal confidence interval length; ARE, asymptotic relative efficiency; ATE, average treatment effect; QTE(τ), quantile treatment effect at τ quantile. Both relMSE and ARE are expressed with respect to the IPW estimator based on the overidentified CBPS. The propensity score model is based on a logistic link function. IPS_{ind}, IPW estimator based on IPS estimator (2.10); IPS_{proj}, IPW estimator based on IPS estimator (2.11); IPS_{exp}, IPW estimator based on IPS estimator (2.12); $CBPS_{just}$, IPW estimator based on the (just-identified) CBPS estimator with moment equation (2.2), with $f(\mathbf{X}^*) = \mathbf{X}^*$; $CBPS_{over}$, IPW estimator based on the (overidentified) CBPS estimator with moment equation (2.2), with $f(\mathbf{X}^*) = \left(\mathbf{X}^{*\prime}, \dot{p}(\mathbf{X}^*;\boldsymbol{\beta})\right)'$, with $\dot{p}(\mathbf{X}^*;\boldsymbol{\beta})$ the derivative of the propensity score model with respect to $\boldsymbol{\beta}$; MLE , IPW estimator based on MLE. $\mathbf{X}^* = \mathbf{X}$ when the PS model is correctly specified, and $\mathbf{X}^* = \mathbf{W}$ when the PS is misspecified.

Table S4.2: Monte Carlo study of the performance of IPW estimators for ATE and QTE based on different propensity score estimation methods. DGP describe in Section 5.1. Sample size: $n = 1,000$.

	Correctly Specified Model						Misspecified Model					
	Bias	RMSE	relMSE	COV	ACIL	ARE	Bias	RMSE	relMSE	COV	ACIL	ARE
(a) **ATE**												
IPS_{exp}	0.01	2.66	0.85	0.94	10.14	1.25	2.04	3.30	0.63	0.87	10.05	1.47
IPS_{ind}	0.48	2.63	0.84	0.97	11.26	1.01	2.37	3.68	0.78	0.94	13.39	0.83
IPS_{proj}	0.02	2.62	0.83	0.94	9.98	1.29	0.48	2.55	0.38	0.96	11.14	1.20
$CBPS_{just}$	0.01	2.91	1.02	0.94	10.90	1.08	2.80	3.84	0.85	0.80	10.00	1.49
$CBPS_{over}$	0.02	2.88	1.00	0.95	11.33	1.00	3.09	4.15	1.00	0.88	12.19	1.00
MLE	-0.01	3.10	1.16	0.94	11.78	0.92	6.51	10.28	6.12	0.76	16.72	0.53
(b) **QTE(0.10)**												
IPS_{exp}	0.02	3.57	1.06	0.95	13.81	0.98	-4.06	5.45	1.52	0.79	14.07	0.99
IPS_{ind}	0.23	3.93	1.29	0.95	15.49	0.78	-3.19	4.97	1.27	0.91	16.48	0.72
IPS_{proj}	0.03	3.58	1.07	0.95	13.84	0.98	-2.66	4.67	1.12	0.98	24.85	0.32
$CBPS_{just}$	0.08	3.43	0.98	0.95	13.38	1.05	-3.17	4.78	1.17	0.84	13.74	1.04
$CBPS_{over}$	0.05	3.47	1.00	0.96	13.69	1.00	-2.71	4.42	1.00	0.89	14.00	1.00
MLE	0.04	3.47	1.00	0.95	13.46	1.03	-1.37	4.71	1.14	0.92	14.90	0.88
(b) **QTE(0.25)**												
IPS_{exp}	0.03	3.05	1.02	0.95	12.13	1.05	-2.21	3.74	1.45	0.89	12.02	1.09
IPS_{ind}	0.30	3.27	1.18	0.96	13.95	0.79	-1.45	3.46	1.24	0.96	14.16	0.78
IPS_{proj}	0.06	3.05	1.03	0.94	12.14	1.05	-1.77	3.68	1.40	0.98	18.54	0.46
$CBPS_{just}$	0.04	3.04	1.02	0.95	12.05	1.06	-1.31	3.28	1.12	0.93	11.96	1.10
$CBPS_{over}$	0.00	3.01	1.00	0.96	12.41	1.00	-0.96	3.11	1.00	0.95	12.54	1.00
MLE	0.00	3.09	1.05	0.95	12.32	1.02	1.08	6.42	4.28	0.95	15.01	0.70
(b) **QTE(0.50)**												
IPS_{exp}	0.07	3.06	0.88	0.95	12.40	1.15	1.00	3.18	0.72	0.94	12.13	1.27
IPS_{ind}	0.48	3.18	0.95	0.97	13.78	0.93	1.52	3.49	0.86	0.96	14.37	0.90
IPS_{proj}	0.06	3.05	0.88	0.95	12.31	1.17	0.00	3.03	0.65	0.97	13.15	1.08
$CBPS_{just}$	0.07	3.23	0.98	0.96	13.03	1.05	1.88	3.62	0.93	0.90	12.31	1.23
$CBPS_{over}$	0.10	3.26	1.00	0.95	13.32	1.00	2.11	3.76	1.00	0.93	13.64	1.00
MLE	0.06	3.35	1.06	0.95	13.59	0.96	5.52	11.00	8.55	0.86	19.80	0.47
(b) **QTE(0.75)**												
IPS_{exp}	0.04	4.14	0.93	0.94	15.60	1.17	5.44	6.55	0.75	0.72	15.11	1.45
IPS_{ind}	0.67	3.96	0.85	0.96	15.94	1.12	5.45	6.66	0.78	0.85	18.45	0.97
IPS_{proj}	0.05	4.09	0.91	0.93	15.36	1.20	2.32	4.19	0.31	0.96	18.15	1.01
$CBPS_{just}$	0.04	4.45	1.07	0.93	16.73	1.01	6.24	7.33	0.94	0.63	15.19	1.43
$CBPS_{over}$	0.05	4.30	1.00	0.95	16.84	1.00	6.43	7.55	1.00	0.76	18.20	1.00
MLE	0.03	4.60	1.14	0.94	17.73	0.90	12.15	17.86	5.60	0.63	24.85	0.54
(b) **QTE(0.90)**												
IPS_{exp}	-0.14	6.16	0.97	0.93	22.97	1.07	10.57	12.34	0.91	0.62	24.25	1.26
IPS_{ind}	0.75	5.70	0.83	0.95	22.02	1.17	9.84	11.64	0.81	0.78	28.92	0.88
IPS_{proj}	-0.15	6.04	0.93	0.93	22.52	1.11	4.51	6.82	0.28	0.96	32.73	0.69
$CBPS_{just}$	-0.19	6.52	1.09	0.93	23.89	0.99	10.84	12.40	0.92	0.52	22.47	1.46
$CBPS_{over}$	-0.10	6.26	1.00	0.94	23.77	1.00	11.11	12.93	1.00	0.67	27.19	1.00
MLE	-0.31	7.09	1.29	0.94	25.71	0.86	17.29	22.69	3.08	0.52	30.01	0.82

Note: Simulations based on 1,000 Monte Carlo experiments. Bias, Monte Carlo Bias; RMSE, Monte Carlo root mean square error; relMSE, relative Monte Carlo mean square error; COV, Monte Carlo coverage of 95% normal confidence interval; ACIL, Monte Carlo average of 95% normal confidence interval length; ARE, asymptotic relative efficiency; ATE, average treatment effect; QTE(τ), quantile treatment effect at τ quantile. Both relMSE and ARE are expressed with respect to the IPW estimator based on the overidentified CBPS. The propensity score model is based on a logistic link function. IPS_{ind}, IPW estimator based on IPS estimator (2.10); IPS_{proj}, IPW estimator based on IPS estimator (2.11); IPS_{exp}, IPW estimator based on IPS estimator (2.12); $CBPS_{just}$, IPW estimator based on the (just-identified) CBPS estimator with moment equation (2.2), with $f(\mathbf{X}^*) = \mathbf{X}^*$; $CBPS_{over}$, IPW estimator based on the (overidentified) CBPS estimator with moment equation (2.2), with $f(\mathbf{X}^*) = \left(\mathbf{X}^{*\prime}, \dot{p}(\mathbf{X}^*;\boldsymbol{\beta})'\right)'$, with $\dot{p}(\mathbf{X}^*;\boldsymbol{\beta})$ the derivative of the propensity score model with respect to $\boldsymbol{\beta}$; MLE , IPW estimator based on MLE. $\mathbf{X}^* = \mathbf{X}$ when the PS model is correctly specified, and $\mathbf{X}^* = \mathbf{W}$ when the PS is misspecified.

Table S4.3: Monte Carlo study of the performance of IPW estimators for LATE and LQTE based on different instrument propensity score estimation methods. DGP describe in Section 5.1. Sample size: $n = 200$.

	Correctly Specified Model						Misspecified Model					
	Bias	RMSE	relMSE	COV	ACIL	ARE	Bias	RMSE	relMSE	COV	ACIL	ARE
(a) **LATE**												
$LIPS_{exp}$	-1.03	6.76	0.72	0.94	26.67	1.87	4.04	7.56	0.40	0.97	32.08	1.35
$LIPS_{ind}$	-5.60	8.24	1.07	0.90	26.58	1.89	-3.31	7.96	0.44	0.93	30.25	1.52
$LIPS_{proj}$	-2.55	6.75	0.72	0.93	25.15	2.11	-1.15	6.94	0.34	0.98	37.21	1.00
$CBPS_{just}$	-0.37	7.60	0.91	0.90	25.79	2.00	6.95	10.59	0.78	0.82	27.16	1.88
$CBPS_{over}$	2.10	7.98	1.00	0.98	36.51	1.00	8.28	11.96	1.00	0.92	37.24	1.00
MLE	0.19	9.04	1.29	0.92	30.21	1.46	9.15	14.74	1.52	0.82	32.37	1.32
(b) **LQTE(0.10)**												
$LIPS_{exp}$	-0.54	7.34	1.06	0.94	29.37	1.05	-1.67	7.37	1.10	0.94	29.22	1.03
$LIPS_{ind}$	-0.55	8.39	1.39	0.95	38.13	0.62	-3.71	8.90	1.60	0.91	35.28	0.70
$LIPS_{proj}$	-1.45	7.31	1.06	0.93	29.04	1.07	-0.66	7.98	1.29	0.96	40.35	0.54
$CBPS_{just}$	-0.21	7.08	0.99	0.94	28.21	1.13	-0.43	7.10	1.02	0.94	28.41	1.09
$CBPS_{over}$	0.68	7.11	1.00	0.96	30.04	1.00	0.27	7.04	1.00	0.96	29.62	1.00
MLE	0.00	7.15	1.01	0.94	28.64	1.10	0.77	9.93	1.99	0.93	30.20	0.96
(b) **LQTE(0.25)**												
$LIPS_{exp}$	-0.51	6.53	1.01	0.95	27.36	1.13	-0.44	6.53	0.91	0.96	27.78	1.06
$LIPS_{ind}$	-1.87	7.31	1.26	0.95	32.58	0.79	-3.41	7.80	1.30	0.92	31.06	0.84
$LIPS_{proj}$	-1.59	6.54	1.01	0.94	26.80	1.17	-0.87	7.06	1.06	0.97	35.20	0.66
$CBPS_{just}$	-0.05	6.48	0.99	0.95	26.50	1.20	1.15	6.88	1.01	0.94	26.72	1.14
$CBPS_{over}$	1.04	6.51	1.00	0.97	29.03	1.00	1.74	6.84	1.00	0.96	28.53	1.00
MLE	0.25	6.83	1.10	0.95	27.55	1.11	2.91	11.41	2.78	0.94	30.29	0.89
(b) **LQTE(0.50)**												
$LIPS_{exp}$	-1.01	7.37	1.01	0.95	29.03	1.26	1.10	6.92	0.72	0.97	30.22	1.13
$LIPS_{ind}$	-4.56	8.26	1.27	0.93	30.68	1.13	-3.82	8.14	0.99	0.94	31.05	1.07
$LIPS_{proj}$	-2.44	7.34	1.00	0.94	27.98	1.36	-1.68	7.14	0.76	0.98	34.44	0.87
$CBPS_{just}$	-0.24	7.60	1.08	0.94	29.24	1.25	3.31	8.24	1.02	0.92	29.18	1.21
$CBPS_{over}$	0.81	7.33	1.00	0.97	32.64	1.00	3.46	8.18	1.00	0.95	32.15	1.00
MLE	0.09	8.79	1.44	0.94	31.46	1.08	6.06	14.53	3.16	0.91	35.42	0.82
(b) **LQTE(0.75)**												
$LIPS_{exp}$	-1.64	9.29	0.90	0.93	36.20	1.40	4.02	9.51	0.58	0.95	38.25	1.18
$LIPS_{ind}$	-8.46	11.62	1.40	0.84	33.42	1.64	-4.51	9.97	0.64	0.92	36.60	1.29
$LIPS_{proj}$	-3.37	9.30	0.90	0.92	34.18	1.57	-1.97	9.09	0.53	0.97	47.00	0.78
$CBPS_{just}$	-0.43	10.74	1.20	0.92	37.71	1.29	6.92	12.52	1.00	0.87	36.94	1.26
$CBPS_{over}$	1.07	9.82	1.00	0.97	42.86	1.00	6.95	12.49	1.00	0.92	41.51	1.00
MLE	0.04	12.87	1.72	0.92	43.82	0.96	10.50	18.97	2.31	0.84	42.80	0.94
(b) **LQTE(0.90)**												
$LIPS_{exp}$	-3.84	16.31	0.98	0.90	62.09	1.22	7.89	16.62	0.73	0.87	48.46	1.09
$LIPS_{ind}$	-13.00	16.83	1.04	0.74	42.01	2.66	-5.13	14.81	0.58	0.89	49.29	1.05
$LIPS_{proj}$	-5.96	15.18	0.85	0.88	53.83	1.62	-2.53	14.84	0.58	0.94	71.47	0.50
$CBPS_{just}$	-3.41	19.14	1.35	0.85	65.17	1.10	10.81	20.07	1.06	0.74	43.89	1.33
$CBPS_{over}$	0.62	16.46	1.00	0.94	68.48	1.00	11.19	19.51	1.00	0.83	50.60	1.00
MLE	-3.94	21.81	1.76	0.85	74.54	0.84	12.03	21.72	1.24	0.72	43.80	1.33

Note: Simulations based on 1,000 Monte Carlo experiments. Bias, Monte Carlo Bias; RMSE, Monte Carlo root mean square error; relMSE, relative Monte Carlo mean square error; COV, Monte Carlo coverage of 95% normal confidence interval; ACIL, Monte Carlo average of 95% normal confidence interval length; ARE, asymptotic relative efficiency; LATE, local average treatment effect; LQTE(τ), local quantile treatment effect at τ quantile. Both relMSE and ARE are expressed with respect to the IPW estimator based on the overidentified CBPS. All instrument propensity scores is based on a logistic link function. $LIPS_{ind}$, $LIPS_{proj}$ and $LIPS_{exp}$ are the IPW estimators based on LIPS estimator (4.5) with the indicator, projection, and exponential weight function, respectively; $CBPS_{just}$, IPW estimator based on the (just-identified) CBPS estimator with moment equation (2.2), with Z in the place of D and $f(\mathbf{X}^*) = \mathbf{X}^*$; $CBPS_{over}$, IPW estimator based on the (overidentified) CBPS estimator with moment equation (2.2), with Z in the place of D and $f(\mathbf{X}^*) = \left(\mathbf{X}^{*\prime}, \dot{p}(\mathbf{X}^*;\boldsymbol{\beta})'\right)'$, with $\dot{p}(\mathbf{X}^*;\boldsymbol{\beta})$ the derivative of the instrument propensity score model with respect to $\boldsymbol{\beta}$; MLE , IPW estimator based on MLE. $\mathbf{X}^* = \mathbf{X}$ when the instrument PS model is correctly specified, and $\mathbf{X}^* = \mathbf{W}$ when the instrument PS is misspecified.

Table S4.4: Monte Carlo study of the performance of IPW estimators for LATE and LQTE based on different instrument propensity score estimation methods. DGP describe in Section 5.1. Sample size: $n = 1,000$.

	Correctly Specified Model						Misspecified Model					
	Bias	RMSE	relMSE	COV	ACIL	ARE	Bias	RMSE	relMSE	COV	ACIL	ARE
(*a*) **LATE**												
$LIPS_{exp}$	-0.29	3.17	0.74	0.95	12.78	1.74	5.34	6.12	0.35	0.85	15.81	1.30
$LIPS_{ind}$	-4.55	5.39	2.14	0.64	11.27	2.23	1.02	3.54	0.12	0.98	15.29	1.39
$LIPS_{proj}$	-0.64	3.14	0.72	0.94	12.33	1.86	1.38	4.10	0.16	0.98	22.80	0.62
$CBPS_{just}$	-0.16	3.42	0.86	0.94	12.80	1.73	8.23	9.02	0.76	0.35	13.58	1.76
$CBPS_{over}$	0.77	3.68	1.00	0.98	16.84	1.00	9.46	10.37	1.00	0.46	18.00	1.00
MLE	-0.08	3.87	1.11	0.95	14.68	1.32	11.16	13.93	1.80	0.38	19.54	0.85
(*b*) **LQTE(0.10)**												
$LIPS_{exp}$	0.02	3.33	1.05	0.95	12.90	1.05	-1.71	3.75	1.32	0.91	12.92	1.02
$LIPS_{ind}$	-2.23	4.22	1.70	0.89	13.97	0.89	-3.43	4.95	2.30	0.82	13.90	0.89
$LIPS_{proj}$	-0.20	3.33	1.06	0.95	12.91	1.05	-0.82	4.23	1.69	0.96	17.51	0.56
$CBPS_{just}$	0.08	3.17	0.96	0.95	12.42	1.13	-0.05	3.18	0.95	0.95	12.46	1.10
$CBPS_{over}$	0.41	3.24	1.00	0.96	13.20	1.00	0.55	3.26	1.00	0.95	13.08	1.00
MLE	0.14	3.22	0.99	0.95	12.61	1.10	1.15	4.21	1.67	0.95	13.78	0.90
(*b*) **LQTE(0.25)**												
$LIPS_{exp}$	-0.03	2.92	1.00	0.96	12.06	1.12	-0.45	2.89	0.67	0.96	12.31	1.07
$LIPS_{ind}$	-2.82	4.14	2.00	0.87	12.83	0.99	-2.55	3.97	1.25	0.89	13.08	0.94
$LIPS_{proj}$	-0.30	2.93	1.00	0.95	12.01	1.13	-0.61	3.32	0.88	0.97	14.29	0.79
$CBPS_{just}$	0.06	2.86	0.96	0.96	11.65	1.20	1.38	3.13	0.78	0.94	11.63	1.19
$CBPS_{over}$	0.43	2.93	1.00	0.97	12.76	1.00	2.01	3.54	1.00	0.93	12.71	1.00
MLE	0.10	2.96	1.02	0.96	12.03	1.13	3.30	7.36	4.31	0.91	14.88	0.73
(*b*) **LQTE(0.50)**												
$LIPS_{exp}$	-0.20	3.28	0.96	0.96	13.11	1.24	1.71	3.47	0.37	0.96	13.80	1.18
$LIPS_{ind}$	-4.27	5.33	2.53	0.77	13.05	1.25	-1.44	3.57	0.39	0.94	14.19	1.11
$LIPS_{proj}$	-0.56	3.27	0.95	0.95	12.90	1.28	-0.25	3.42	0.35	0.98	15.62	0.92
$CBPS_{just}$	-0.05	3.30	0.97	0.95	13.10	1.24	3.94	5.05	0.77	0.80	12.95	1.34
$CBPS_{over}$	0.38	3.35	1.00	0.97	14.61	1.00	4.64	5.74	1.00	0.82	14.98	1.00
MLE	0.00	3.47	1.07	0.95	13.85	1.11	7.60	12.81	4.97	0.76	20.28	0.55
(*b*) **LQTE(0.75)**												
$LIPS_{exp}$	-0.45	4.35	0.93	0.95	16.98	1.34	5.34	6.57	0.42	0.86	18.75	1.27
$LIPS_{ind}$	-6.10	7.25	2.58	0.64	15.48	1.62	0.34	4.16	0.17	0.98	19.17	1.22
$LIPS_{proj}$	-0.92	4.28	0.90	0.95	16.45	1.43	0.73	4.73	0.22	0.98	26.04	0.66
$CBPS_{just}$	-0.24	4.52	1.00	0.95	17.57	1.25	8.07	9.25	0.82	0.56	17.46	1.47
$CBPS_{over}$	0.37	4.52	1.00	0.97	19.67	1.00	8.94	10.19	1.00	0.64	21.17	1.00
MLE	-0.19	4.79	1.12	0.95	18.97	1.08	13.58	18.85	3.42	0.63	27.58	0.59
(*b*) **LQTE(0.90)**												
$LIPS_{exp}$	-0.74	6.67	0.96	0.95	27.99	1.15	10.96	13.24	0.57	0.74	29.51	1.18
$LIPS_{ind}$	-7.98	9.80	2.07	0.64	21.08	2.03	3.68	7.48	0.18	0.96	29.56	1.18
$LIPS_{proj}$	-1.24	6.43	0.89	0.94	24.84	1.46	2.55	7.66	0.19	0.99	44.40	0.52
$CBPS_{just}$	-0.61	7.23	1.12	0.95	26.93	1.24	13.74	15.73	0.81	0.50	26.72	1.44
$CBPS_{over}$	0.57	6.82	1.00	0.97	30.04	1.00	15.17	17.52	1.00	0.55	32.11	1.00
MLE	-0.68	8.20	1.45	0.96	31.90	0.89	18.91	24.06	1.89	0.48	30.30	1.12

Note: Simulations based on 1,000 Monte Carlo experiments. Bias, Monte Carlo Bias; RMSE, Monte Carlo root mean square error; relMSE, relative Monte Carlo mean square error; COV, Monte Carlo coverage of 95% normal confidence interval; ACIL, Monte Carlo average of 95% normal confidence interval length; ARE, asymptotic relative efficiency; LATE, local average treatment effect; LQTE(τ), local quantile treatment effect at τ quantile. Both relMSE and ARE are expressed with respect to the IPW estimator based on the overidentified CBPS. All instrument propensity scores is based on a logistic link function. $LIPS_{ind}$, $LIPS_{proj}$ and $LIPS_{exp}$ are the IPW estimators based on LIPS estimator (4.5) with the indicator, projection, and exponential weight function, respectively; $CBPS_{just}$, IPW estimator based on the (just-identified) CBPS estimator with moment equation (2.2), with Z in the place of D and $f(\mathbf{X}^*) = \mathbf{X}^*$; $CBPS_{over}$, IPW estimator based on the (overidentified) CBPS estimator with moment equation (2.2), with Z in the place of D and $f(\mathbf{X}^*) = \left(\mathbf{X}^{*\prime}, \dot{p}(\mathbf{X}^*;\boldsymbol{\beta})'\right)'$, with $\dot{p}(\mathbf{X}^*;\boldsymbol{\beta})$ the derivative of the instrument propensity score model with respect to $\boldsymbol{\beta}$; MLE , IPW estimator based on MLE. $\mathbf{X}^* = \mathbf{X}$ when the instrument PS model is correctly specified, and $\mathbf{X}^* = \mathbf{W}$ when the instrument PS is misspecified.

S5 Auxiliary Lemmas

In this Section, we present and prove some auxiliary lemmas that help proving the main results of the paper.

Lemma S5.1 *Let Π be a compact, convex subset of $\mathbb{R}^k$ with a non-empty interior. Then*

$$\mathcal{W}_{ind}=\left\{\mathbf{x}\in\mathcal{X}\mapsto 1(\mathbf{x}\leq\mathbf{u}):\mathbf{u}\in[-\infty,\infty]^k\right\},$$

$$\mathcal{W}_{proj}=\left\{\mathbf{x}\in\mathcal{X}\mapsto 1\left\{\boldsymbol{\gamma}'\mathbf{x}\leq u\right\}:(\boldsymbol{\gamma},u)\in\mathbb{S}_k\times[-\infty,\infty]\right\},$$

$$\mathcal{W}_{exp}=\left\{\mathbf{x}\in\mathcal{X}\mapsto \exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right)):\mathbf{u}\in\Pi\right\},$$

are uniformly bounded Donsker classes of functions.

Proof of Lemma S5.1: The uniform boundedness property follows from the fact that $1(\boldsymbol{x}\leq\boldsymbol{u})\leq 1$, $1\{\boldsymbol{\gamma}'\mathbf{x}\leq u\}\leq 1$ and $|\exp(i\mathbf{u}'\Phi(\mathbf{x}))|=|\cos(\mathbf{u}'\Phi(\mathbf{x}))+i\sin(\mathbf{u}'\Phi(\mathbf{x}))|\leq 1$. From Example 2.5.4 in van der Vaart and Wellner (1996), $\mathcal{W}_{ind}$ is Donsker. From Theorems 2.5.2, 2.6.7 and Problem 14 on page 152 in van der Vaart and Wellner (1996), $\mathcal{W}_{\text{proj}}$ is Donsker. Finally, since $\exp(i\mathbf{u}'\Phi(\mathbf{x}))$ is infinitely differentiable with respect to $\mathbf{u}$, and all derivatives are uniformly bounded on Π, the Donsker property of $\mathcal{W}_{\exp}$ follows from Theorem 2.5.6 and Corollary 2.7.2 in van der Vaart and Wellner (1996) .■

Lemma S5.2 *Under Assumption 2$(i)-(iii)$, the classes of functions*

$$\mathcal{F}_1\equiv\{(d,\mathbf{x})\in\{0,1\}\times\mathcal{X}\mapsto d/p\left(\mathbf{x};\boldsymbol{\beta}\right):\boldsymbol{\beta}\in\Theta\},$$

$$\mathcal{F}_2\equiv\{(d,\mathbf{x})\in\{0,1\}\times\mathcal{X}\mapsto(1-d)/\left(1-p\left(\mathbf{x};\boldsymbol{\beta}\right)\right):\boldsymbol{\beta}\in\Theta\},$$

$$\mathcal{F}_3\equiv\mathcal{F}_1\cdot\mathcal{W},$$

$$\mathcal{F}_4\equiv\mathcal{F}_2\cdot\mathcal{W},$$

where $\mathcal{W}$ is either equal to $\mathcal{W}_{ind}$, $\mathcal{W}_{proj}$ or $\mathcal{W}_{exp}$, are Glivenko-Cantelli.

Proof of Lemma S5.2: By Example 19.8 in van der Vaart (1998), $\mathcal{F}_1$ and $\mathcal{F}_2$ are Glivenko-Cantelli (GC) classes under Assumption 2$(i)-(iii)$. By Lemma S5.1, $\mathcal{W}_{ind}$, $\mathcal{W}_{\text{proj}}$ and $\mathcal{W}_{\text{exp}}$ are uniformly bounded Donsker classes of functions, and therefore they are also GC, see, e.g., Lemma 8.17 in Kosorok (2008). Finally, by Corollary 9.26 in Kosorok (2008), $\mathcal{F}_3$ and $\mathcal{F}_4$ are GC. ■

Let

$$\widehat{C}_{ind,F_{n,\mathbf{X}}}=2\int_{[-\infty,\infty]^k}\dot{\mathbf{H}}'_{n,ind}(\widehat{\boldsymbol{\beta}}^{ips}_{n,ind},\mathbf{u})\dot{\mathbf{H}}_{n,ind}(\widetilde{\boldsymbol{\beta}},\mathbf{u})F_{n,\mathbf{X}}(d\mathbf{u}),$$

$$\widehat{C}_{proj,F_{n,\gamma}} = 2\int_{[-\infty,\infty]\times\mathbb{S}_k} \dot{\mathbf{H}}'_{n,proj}(\widehat{\boldsymbol{\beta}}^{ips}_{n,proj},\mathbf{u})\dot{\mathbf{H}}_{n,proj}(\widetilde{\boldsymbol{\beta}},\mathbf{u})F_{n,\boldsymbol{\gamma}}(du)d\boldsymbol{\gamma},$$

and

$$\widehat{C}_{\exp,\Phi} = \int_{\mathbb{R}^k} \dot{\mathbf{H}}^c_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\mathbf{u})\dot{\mathbf{H}}_{n,\exp}(\widetilde{\boldsymbol{\beta}},\mathbf{u})\phi\left(\mathbf{u}\right)d\mathbf{u} + \int_{\mathbb{R}^k} \dot{\mathbf{H}}'_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\mathbf{u})\left(\dot{\mathbf{H}}'_{n,\exp}(\widetilde{\boldsymbol{\beta}},\mathbf{u})\right)^c\phi\left(\mathbf{u}\right)d\mathbf{u},$$

where $\phi\left(\mathbf{u}\right)$ is the standard k-variate normal density function and $\widetilde{\boldsymbol{\beta}}$ satisfies $\left\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\| \leq \left\|\widehat{\boldsymbol{\beta}}^{ips}_{n,w}-\boldsymbol{\beta}_0\right\|$. Furthermore, write

$$C_{ind,F_{\mathbf{X}}} = 2\int_{[-\infty,\infty]^k} \dot{\mathbf{H}}'_{ind}(\boldsymbol{\beta}_0,\mathbf{u})\dot{\mathbf{H}}_{ind}(\boldsymbol{\beta}_0,\mathbf{u})F_{\mathbf{X}}(d\mathbf{u}),$$

$$C_{proj,F_{\boldsymbol{\gamma}}} = 2\int_{[-\infty,\infty]\times\mathbb{S}_k} \dot{\mathbf{H}}'_{proj}(\boldsymbol{\beta}_0,\mathbf{u})\dot{\mathbf{H}}_{proj}(\boldsymbol{\beta}_0,\mathbf{u})F_{\boldsymbol{\gamma}}(du)d\boldsymbol{\gamma},$$

$$C_{\exp,\Phi} = \int_{\mathbb{R}^k}\left(\dot{\mathbf{H}}^c_{\exp}(\boldsymbol{\beta}_0,\mathbf{u})\dot{\mathbf{H}}_w(\boldsymbol{\beta}_0,\mathbf{u}) + \dot{\mathbf{H}}'_{\exp}(\boldsymbol{\beta}_0,\mathbf{u})\left(\dot{\mathbf{H}}'_{\exp}(\boldsymbol{\beta}_0,\mathbf{u})\right)^c\right)\phi\left(\mathbf{u}\right)d\mathbf{u}.$$

Lemma S5.3 *Let $\mathcal{W}$ be equal to either $\mathcal{W}_{ind}$, $\mathcal{W}_{proj}$ or $\mathcal{W}_{exp}$. Then, under Assumption 2,*

$$\mathcal{F}_5 \equiv \left\{(d,\mathbf{x})\in\{0,1\}\times\mathcal{X}\mapsto \frac{d}{p\left(\mathbf{x};\boldsymbol{\beta}\right)^2}\dot{p}\left(\mathbf{x};\boldsymbol{\beta}\right),\boldsymbol{\beta}\in\Theta_0\right\},$$

$$\mathcal{F}_6 \equiv \left\{(d,\mathbf{x})\in\{0,1\}\times\mathcal{X}\mapsto \frac{1-d}{\left(1-p\left(\mathbf{x};\boldsymbol{\beta}\right)\right)^2}\dot{p}\left(\mathbf{x};\boldsymbol{\beta}\right),\boldsymbol{\beta}\in\Theta_0\right\},$$

$$\mathcal{F}_7 \equiv \mathcal{F}_5\cdot\mathcal{W},$$

$$\mathcal{F}_8 \equiv \mathcal{F}_6\cdot\mathcal{W},$$

are Glivenko-Cantelli classes of functions. Furthermore,

$$\widehat{C}_{ind,F_{n,\mathbf{X}}} - C_{ind,F_{\mathbf{X}}} = o_p\left(1\right),$$

$$\widehat{C}_{proj,F_{n,\boldsymbol{\gamma}}} - C_{proj,F_{\boldsymbol{\gamma}}} = o_p\left(1\right),$$

$$\widehat{C}_{\exp,\Phi} - C_{\exp,\Phi} = o_p\left(1\right).$$

Proof of Lemma S5.3: By Example 19.8 in van der Vaart (1998), $\mathcal{F}_5$ and $\mathcal{F}_6$ are Glivenko-Cantelli (GC) classes under Assumption 2. By Lemma S5.1, $\mathcal{W}_{ind}$, $\mathcal{W}_{\text{proj}}$ and $\mathcal{W}_{\text{exp}}$ are uniformly bounded Donsker classes of functions, and therefore they are also GC. Finally, by Corollary 9.27 in Kosorok (2008), $\mathcal{F}_3$ and $\mathcal{F}_4$ are GC.

Next, from the first part of Theorem 3.1, we have that $\widehat{\boldsymbol{\beta}}^{ips}_{n,w} \xrightarrow{p.} \boldsymbol{\beta}_0$, which in turn implies that

$\widehat{\boldsymbol{\beta}}^{ips}_{n,w}, \widetilde{\boldsymbol{\beta}} \in \Theta_0$ with probability approaching one. Thus, from Lemma S5.1 and a direct application of the CMT, we conclude that

$$\widehat{C}_{ind,F_{n,\mathbf{X}}} - C_{ind,F_{\mathbf{X}}} = o_p(1),$$

$$\widehat{C}_{proj,F_{n,\boldsymbol{\gamma}}} - C_{proj,F_{\boldsymbol{\gamma}}} = o_p(1).$$

To conclude the proof of this lemma, we need to show that

$$\widehat{C}_{\exp,\Phi} - C_{\exp,\Phi} = o_p(1).$$

Toward this end, as in the consistency proof of Theorem 3.1, fix an arbitrarily small $\epsilon > 0$ and choose a compact and convex set Π such that

$$\left| \int_{\mathbb{R}^k \setminus \Pi} \phi(\mathbf{u})\, d\mathbf{u} \right| \leq \epsilon. \tag{S5.1}$$

Then, write

$$\begin{aligned}
&\int_{\mathbb{R}^k} A_{n,1}\left(\mathbf{u}; \widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \widetilde{\boldsymbol{\beta}}\right) \phi(\mathbf{u})\, d\mathbf{u} \\
&= \int_{\Pi} A_{n,1}\left(\mathbf{u}; \widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \widetilde{\boldsymbol{\beta}}\right) \phi(\mathbf{u})\, d\mathbf{u} + \int_{\mathbb{R}^k \setminus \Pi} A_{n,1}\left(\mathbf{u}; \widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \widetilde{\boldsymbol{\beta}}\right) \phi(\mathbf{u})\, d\mathbf{u}, \\
&\equiv J_{3n} + J_{4n}.
\end{aligned}$$

with

$$A_{n,1}\left(\mathbf{u}; \widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \widetilde{\boldsymbol{\beta}}\right) \equiv \dot{\mathbf{H}}^c_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \mathbf{u}) \dot{\mathbf{H}}_{n,\exp}(\widetilde{\boldsymbol{\beta}}, \mathbf{u}) + \dot{\mathbf{H}}'_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \mathbf{u}) \left(\dot{\mathbf{H}}'_{n,\exp}(\widetilde{\boldsymbol{\beta}}, \mathbf{u})\right)^c.$$

Let

$$A_1(\mathbf{u}; \boldsymbol{\beta}_0, \boldsymbol{\beta}) \equiv \dot{\mathbf{H}}^c_{\exp}(\boldsymbol{\beta}_0, \mathbf{u}) \dot{\mathbf{H}}_{\exp}(\boldsymbol{\beta}, \mathbf{u}) + \dot{\mathbf{H}}'_{\exp}(\boldsymbol{\beta}_0, \mathbf{u}) \left(\dot{\mathbf{H}}'_{\exp}(\boldsymbol{\beta}, \mathbf{u})\right)^c.$$

From the GC results above and the CMT, we have that

$$\sup_{\mathbf{u} \in \Pi} \left\| A_{n,1}\left(\mathbf{u}; \widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \widetilde{\boldsymbol{\beta}}\right) - A_1(\mathbf{u}; \boldsymbol{\beta}_0, \boldsymbol{\beta}_0) \right\| \xrightarrow{p} 0.$$

Thus, by another application of the CMT, it follows that

$$J_{3n} = \int_{\Pi} A_1(\mathbf{u}; \boldsymbol{\beta}_0, \boldsymbol{\beta}_0) \phi(\mathbf{u})\, d\mathbf{u} + o_p(1).$$

For J_{4n}, since $|\exp(i\mathbf{u}'\Phi(\mathbf{x})| \leq 1$, we have that under Assumption 2, for all $\boldsymbol{\beta} \in \Theta_0$,

$$\left\| \dot{\mathbf{H}}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u}) \right\| \leq \mathbb{E}_n[b(\mathbf{X})] = O_p(1)$$

for some integrable function $b(\mathbf{X})$. Hence, by (S5.1) we have that $J_{4n} = O_p(\epsilon)$. Since $\epsilon > 0$ is arbitrary, this concludes the proof. ■

Lemma S5.4 *Let Π be a compact, convex subset of $\mathbb{R}^k$ with a non-empty interior. Then, under Assumption 2,*

$$\mathcal{F}_{ind} \equiv \left\{(d,\mathbf{x}) \in \{0,1\} \times \mathcal{X} \mapsto \mathbf{h}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right) 1(\mathbf{x} \leq \mathbf{u}) : \mathbf{u} \in [-\infty,\infty]^k\right\},$$

$$\mathcal{F}_{proj} \equiv \left\{(d,\mathbf{x}) \in \{0,1\} \times \mathcal{X} \mapsto \mathbf{h}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right) 1\left\{\boldsymbol{\gamma}'\mathbf{x} \leq u\right\} : (\boldsymbol{\gamma},u) \in \mathbb{S}_k \times [-\infty,\infty]\right\}$$

$$\mathcal{F}_{\exp} \equiv \left\{(d,\mathbf{x}) \in \{0,1\} \times \mathcal{X} \mapsto \mathbf{h}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right) \exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right)) : \mathbf{u} \in \Pi\right\},$$

are Donsker classes of functions.

Proof of Lemma S5.4: The Donsker properties follow directly from Lemma S5.1, Assumption 2(ii), and Corollary 9.32 in Kosorok (2008). ■

Define

$$A_{n,2,ind}\left(\mathbf{x}\right) = 2 \cdot \int_{[-\infty,\infty]^k} \dot{\mathbf{H}}_{n,ind}'(\widehat{\boldsymbol{\beta}}_{n,ind}^{ips},\mathbf{u}) 1(\mathbf{x} \leq \mathbf{u}) F_{n,\mathbf{X}}(d\mathbf{u}),$$

$$A_{n,2,proj}\left(\mathbf{x}\right) = 2 \cdot \int_{[-\infty,\infty]\times\mathbb{S}_k} \dot{\mathbf{H}}_{n,proj}'(\widehat{\boldsymbol{\beta}}_{n,proj}^{ips},(u,\boldsymbol{\gamma})) 1\left\{\boldsymbol{\gamma}'\mathbf{x} \leq u\right\} F_{n,\boldsymbol{\gamma}}(du)d\boldsymbol{\gamma},$$

$$A_{n,2,\exp}\left(\mathbf{x}\right) = \int_{\mathbb{R}^k} \left(\dot{\mathbf{H}}_{n,\exp}^{c}(\widehat{\boldsymbol{\beta}}_{n,\exp}^{ips},\mathbf{u}) \exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right)) + \dot{\mathbf{H}}_{n,\exp}'(\widehat{\boldsymbol{\beta}}_{n,\exp}^{ips},\mathbf{u}) \exp(-i\mathbf{u}'\Phi\left(\mathbf{x}\right))\right) \phi\left(\mathbf{u}\right) d\mathbf{u},$$

and let

$$A_{2,ind}\left(\mathbf{x}\right) = 2 \cdot \int_{[-\infty,\infty]^k} \left(\dot{\mathbf{H}}_{ind}'(\boldsymbol{\beta}_0,\mathbf{u}) 1(\mathbf{x} \leq \mathbf{u})\right) F_{\mathbf{X}}(d\mathbf{u}),$$

$$A_{2,proj}\left(\mathbf{x}\right) = 2 \cdot \int_{[-\infty,\infty]\times\mathbb{S}_k} \dot{\mathbf{H}}_{proj}'(\boldsymbol{\beta}_0,(u,\boldsymbol{\gamma})) 1\left\{\boldsymbol{\gamma}'\mathbf{x} \leq u\right\} F_{\boldsymbol{\gamma}}(du)d\boldsymbol{\gamma},$$

$$A_{2,\exp}\left(\mathbf{x}\right) = \int_{\mathbb{R}^k} \left(\dot{\mathbf{H}}_{\exp}^{c}(\boldsymbol{\beta}_0,\mathbf{u}) \exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right)) + \dot{\mathbf{H}}_{\exp}'(\boldsymbol{\beta}_0,\mathbf{u}) \exp(-i\mathbf{u}'\Phi\left(\mathbf{x}\right))\right) \phi\left(\mathbf{u}\right) d\mathbf{u}.$$

Lemma S5.5 *Under Assumption 2,*

$$\mathbb{E}_n\left[A_{n,2,ind}\left(\mathbf{X}\right) \cdot \mathbf{h}_n\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right] = \mathbb{E}_n\left[A_{2,ind}\left(\mathbf{X}\right) \cdot \mathbf{h}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right] + o_p\left(n^{-1/2}\right). \tag{S5.2}$$

$$\mathbb{E}_n\left[A_{n,2,proj}\left(\mathbf{X}\right) \cdot \mathbf{h}_n\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right] = \mathbb{E}_n\left[A_{2,proj}\left(\mathbf{X}\right) \cdot \mathbf{h}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right] + o_p\left(n^{-1/2}\right), \tag{S5.3}$$

$$\mathbb{E}_n\left[A_{n,2,\exp}\left(\mathbf{X}\right) \cdot \mathbf{h}_n\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right] = \mathbb{E}_n\left[A_{2,\exp}\left(\mathbf{X}\right) \cdot \mathbf{h}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right] + o_p\left(n^{-1/2}\right). \tag{S5.4}$$

Proof of Lemma S5.5: First note that

$$\sqrt{n}\mathbb{E}_n\left[A_{n,2,ind}\left(\mathbf{X}\right) \cdot \mathbf{h}_n\left(D,X;\boldsymbol{\beta}_0\right)\right] = 2\int \dot{\mathbf{H}}_{n,ind}\left(\widehat{\boldsymbol{\beta}}_{n,ind}^{ips},\mathbf{u}\right) \sqrt{n}\mathbf{H}_{n,ind}\left(\boldsymbol{\beta}_0,\mathbf{u}\right) F_{n,\mathbf{X}}(d\mathbf{u}).$$

Then, from Lemma S5.3, Lemma S5.4, the CMT, and the fact that $\mathbf{H}_{ind}(\boldsymbol{\beta}_0, \mathbf{u}) = 0$ $a.e$, it follows that, uniformly in $\mathbf{u} \in [-\infty, \infty]^k$,

$$\dot{\mathbf{H}}'_{n,ind}\left(\widehat{\boldsymbol{\beta}}^{ips}_{n,ind}, \mathbf{u}\right) \cdot \sqrt{n}\mathbf{H}_{n,ind}(\boldsymbol{\beta}_0, \mathbf{u}) = \dot{\mathbf{H}}'_{ind}(\boldsymbol{\beta}_0, \mathbf{u}) \cdot \sqrt{n}\mathbb{E}_n\left[\mathbf{h}(D, \mathbf{X}; \boldsymbol{\beta}_0) 1(\mathbf{X} \leq \mathbf{u})\right] + o_p(1).$$

Furthermore, the process

$$\dot{\mathbf{H}}'_{ind}(\boldsymbol{\beta}_0, \mathbf{u}) \cdot \sqrt{n}\mathbb{E}_n\left[\mathbf{h}(D, \mathbf{X}; \boldsymbol{\beta}_0) 1(\mathbf{X} \leq \mathbf{u})\right]$$

is asymptotically tight in $\ell^\infty\left([-\infty, \infty]^k\right)$. Given Lemma S5.1 and the Glivenko-Cantelli theorem, we apply Lemma 3.1 of Chang (1990) to conclude (S5.2).

The proof of (S5.3) follows exactly the same steps and is therefore omitted. To prove (S5.4), fix an arbitrarily small $\epsilon > 0$ and choose a compact and convex set Π such that

$$\left|\int_{\mathbb{R}^k \setminus \Pi} \phi(\mathbf{u})\, d\mathbf{u}\right| \leq \epsilon. \tag{S5.5}$$

Then write

$$\begin{aligned}
&\sqrt{n}\mathbb{E}_n\left[A_{n,2,\exp}(\mathbf{X}) \cdot \mathbf{h}_n(D, X; \boldsymbol{\beta}_0)\right] \\
&= \int_{\mathbb{R}^k} \dot{\mathbf{H}}^c_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \mathbf{u})\sqrt{n}\mathbf{H}_{n,\exp}(\boldsymbol{\beta}_0, \mathbf{u})\, \phi(\mathbf{u})\, d\mathbf{u} \\
&+ \int_{\mathbb{R}^k} \dot{\mathbf{H}}'_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \mathbf{u})\sqrt{n}\left(\mathbf{H}'_{n,\exp}(\boldsymbol{\beta}_0, \mathbf{u})\right)^c \phi(\mathbf{u})\, d\mathbf{u} \\
&= \int_{\Pi} \widehat{A}_{n,3}\left(\mathbf{u}; \widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \boldsymbol{\beta}_0\right) \phi(\mathbf{u})\, d\mathbf{u} + \int_{\mathbb{R}^k \setminus \Pi} \widehat{A}_{n,3}\left(\mathbf{u}; \widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \boldsymbol{\beta}_0\right) \phi(\mathbf{u})\, d\mathbf{u}, \\
&\equiv J_{5n} + J_{6n}.
\end{aligned}$$

with

$$\begin{aligned}
\widehat{A}_{n,3}\left(\mathbf{u}; \widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \boldsymbol{\beta}_0\right) \equiv\ & \dot{\mathbf{H}}^c_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \mathbf{u})\sqrt{n}\mathbf{H}_{n,\exp}(\boldsymbol{\beta}_0, \mathbf{u}) \\
&+ \dot{\mathbf{H}}'_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \mathbf{u})\sqrt{n}\left(\mathbf{H}'_{n,\exp}(\boldsymbol{\beta}_0, \mathbf{u})\right)^c.
\end{aligned}$$

Let

$$A_{n,3}(\mathbf{u}; \boldsymbol{\beta}_0, \boldsymbol{\beta}_0) \equiv \dot{\mathbf{H}}^c_{\exp}(\boldsymbol{\beta}_0, \mathbf{u})\sqrt{n}\mathbf{H}_{n,\exp}(\boldsymbol{\beta}_0, \mathbf{u}) + \dot{\mathbf{H}}'_{\exp}(\boldsymbol{\beta}_0, \mathbf{u})\sqrt{n}\left(\mathbf{H}'_{n,\exp}(\boldsymbol{\beta}_0, \mathbf{u})\right)^c.$$

From Lemma S5.3, Lemma S5.4, the CMT, and the fact that $\mathbf{H}_{\exp}(\boldsymbol{\beta}_0, \mathbf{u}) = 0$ $a.e$, it follows that, uniformly in $\mathbf{u} \in \Pi$,

$$\widehat{A}_{n,3}\left(\mathbf{u}; \widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \boldsymbol{\beta}_0\right) = A_{n,3}(\mathbf{u}; \boldsymbol{\beta}_0, \boldsymbol{\beta}_0) + o_p(1).$$

Furthermore, the process $A_{n,3}(\mathbf{u};\boldsymbol{\beta}_0,\boldsymbol{\beta})$ is asymptotically tight in $\ell^\infty(\Pi)$. Given that $\int(\cdot)\phi(\mathbf{u})\,d\mathbf{u}$ is a nonrandom continuous functional, by the CMT, we conclude that

$$J_{5n}=\int_\Pi A_{n,3}(\mathbf{u};\boldsymbol{\beta}_0,\boldsymbol{\beta}_0)\,\phi(\mathbf{u})\,d\mathbf{u}+o_p(1).$$

For J_{6n}, since $|\exp(i\mathbf{u}'\Phi(\mathbf{x})|\le 1$, we have that under Assumption 2,

$$\left\|\dot{\mathbf{H}}_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\mathbf{u})\right\|\le C\cdot\mathbb{E}_n\left[b(\mathbf{X})\right]=O_p(1),$$

for some $C<\infty$ and some integrable function $b(\mathbf{X})$. Furthermore, $\mathbb{E}\left[\|\mathbf{h}(D,\mathbf{X};\boldsymbol{\beta}_0)\|\right]<\infty$,

$$\sqrt{n}\left\|\mathbf{H}_{n,\exp}(\boldsymbol{\beta}_0,\mathbf{u})\right\|\le C\cdot\sqrt{n}\mathbb{E}_n\left[\|\mathbf{h}(D,\mathbf{X};\boldsymbol{\beta}_0)\|\right]=O_p\left(n^{1/2}\right).$$

Hence, by (S5.5) we have that $J_{6n}=O_p(\epsilon\cdot n^{1/2})$. Since $\epsilon>0$ is arbitrary, we can pick ϵ such that, for some $\delta>0$, $\epsilon=o\left(n^{-1/2-\delta}\right)$, which concludes the proof. ■

S6 Proofs of Main Results

Proof of Lemma 2.2: The first part, $Q_w(\boldsymbol{\beta})\ge 0$, follows directly from the definition. Next, as discussed in Section 2.2, the covariate balancing condition (2.1) is equivalent to (2.4), implying that $Q_w(\boldsymbol{\beta}_0)=0$.

To complete the proof we then need to show that if $Q_w(\boldsymbol{\beta})=0$, then $\boldsymbol{\beta}=\boldsymbol{\beta}_0$. Towards this end, recall that if $Q_w(\boldsymbol{\beta})=0$, it must be that $\mathbf{H}_w(\boldsymbol{\beta},\boldsymbol{u})=0$ *a.e.* on Π, because $||\cdot||\ge 0$ and the integrating probability measure Ψ is absolutely continuous with respect to a dominating measure on Π. However, $\mathbf{H}_w(\boldsymbol{\beta},\boldsymbol{u})=0$ *a.e.* on Π if and only if $\mathbb{E}\left[\mathbf{h}(D,\mathbf{X};\boldsymbol{\beta})\,|\mathbf{X}\right]=0$ *a.s.*, which is equivalent to

$$\mathbb{E}\left[\frac{D}{p(\mathbf{X};\boldsymbol{\beta})}\middle|\mathbf{X}\right]=\mathbb{E}\left[\frac{D}{p(\mathbf{X};\boldsymbol{\beta})}\right]\ \ a.s.\ \text{ and }\ \mathbb{E}\left[\frac{1-D}{1-p(\mathbf{X};\boldsymbol{\beta})}\middle|\mathbf{X}\right]=\mathbb{E}\left[\frac{1-D}{1-p(\mathbf{X};\boldsymbol{\beta})}\right]a.s..$$

The above is further equivalent to

$$\frac{p(\mathbf{X};\boldsymbol{\beta}_0)}{p(\mathbf{X};\boldsymbol{\beta})}=\mathbb{E}\left[\frac{p(\mathbf{X};\boldsymbol{\beta}_0)}{p(\mathbf{X};\boldsymbol{\beta})}\right]\ \ a.s.\ \text{ and }\ \frac{1-p(\mathbf{X};\boldsymbol{\beta}_0)}{1-p(\mathbf{X};\boldsymbol{\beta})}=\mathbb{E}\left[\frac{1-p(\mathbf{X};\boldsymbol{\beta}_0)}{1-p(\mathbf{X};\boldsymbol{\beta})}\right]a.s..$$

That is, there exist some constants c_1 and c_0, potentially depending on β, such that $p(\mathbf{X};\boldsymbol{\beta}_0)=c_1p(\mathbf{X};\boldsymbol{\beta})$ *a.s.* and $1-p(\mathbf{X};\boldsymbol{\beta}_0)=c_0(1-p(\mathbf{X};\boldsymbol{\beta}))$ *a.s.*, which imply

$$(c_1-c_0)\,p(\mathbf{X};\boldsymbol{\beta})=1-c_0\quad a.s. \tag{S6.1}$$

If $c_1\ne c_0$, then $p(\mathbf{X};\boldsymbol{\beta})=(1-c_0)/(c_1-c_0)$ *a.s.* so that $p(\mathbf{X};\boldsymbol{\beta})$ degenerates to a constant. This leads to a contradiction. In light of (S6.1), we then conclude that $c_1=c_0=1$ and that $\mathbf{H}_w(\boldsymbol{\beta},\boldsymbol{u})=0$ *a.e.* on Π is equivalent to $p(\mathbf{X};\boldsymbol{\beta})=p(\mathbf{X};\boldsymbol{\beta}_0)$ *a.s.*. Given that $p(\mathbf{X};\boldsymbol{\beta})=p(\mathbf{X};\boldsymbol{\beta}_0)$ *a.s.* for a unique $\boldsymbol{\beta}_0$, we must have $\boldsymbol{\beta}=\boldsymbol{\beta}_0$. This concludes the proof. ■

Proof of Theorem 3.1: We first show that $\widehat{\boldsymbol{\beta}}_{n,w}^{ips} - \boldsymbol{\beta}_0 = o_p(1)$ using M-estimator theory, see e.g. Theorem 5.7 in van der Vaart (1998). Since Lemma 2.2 already established that $Q_w(\boldsymbol{\beta})$ achieves the unique minimum value at $\boldsymbol{\beta}_0$, and by Assumption 2 we have that $\mathbf{H}_w(\boldsymbol{\beta}, \mathbf{u})$ is continuous at each $\boldsymbol{\beta} \in \Theta$, Θ is compact, we have that by Exercise 5.27 in van der Vaart (1998) for every $\varepsilon > 0$

$$\inf_{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\geq\varepsilon} Q_w(\boldsymbol{\beta}) > Q_w(\boldsymbol{\beta}_0).$$

Thus, to establish consistency of $\widehat{\boldsymbol{\beta}}_{n,w}^{ips}$ it suffices to show that, as $n \to \infty$,

$$\sup_{\boldsymbol{\beta}\in\Theta} |Q_{n,w}(\boldsymbol{\beta}) - Q_w(\boldsymbol{\beta})| \xrightarrow{p} 0.$$

From Lemma S5.1 we have that $F_{n,\mathbf{X}}$ and $F_{n,\boldsymbol{\gamma}}$ are uniformly consistent for $F_{\mathbf{X}}$ and $F_{\boldsymbol{\gamma}}$, respectively, whereas by Lemma S5.2 and the continuous mapping theorem (CMT), we have that

$$\sup_{(\boldsymbol{\beta},\mathbf{u})\in\Theta\times[-\infty,\infty]^k} \|\mathbf{H}_{n,ind}(\boldsymbol{\beta}, \mathbf{u}) - \mathbf{H}_{ind}(\boldsymbol{\beta}, \mathbf{u})\| \xrightarrow{p} 0,$$

$$\sup_{(\boldsymbol{\beta},\mathbf{u})\in\Theta\times([-\infty,\infty]\times\mathbb{S}_k)} \|\mathbf{H}_{n,proj}(\boldsymbol{\beta}, \mathbf{u}) - \mathbf{H}_{proj}(\boldsymbol{\beta}, \mathbf{u})\| \xrightarrow{p} 0.$$

Given that integration is a linear functional, by the CMT we have that

$$\sup_{\boldsymbol{\beta}\in\Theta} |Q_{n,ind}(\boldsymbol{\beta}) - Q_{ind}(\boldsymbol{\beta})| \xrightarrow{p} 0,$$

$$\sup_{\boldsymbol{\beta}\in\Theta} |Q_{n,proj}(\boldsymbol{\beta}) - Q_{proj}(\boldsymbol{\beta})| \xrightarrow{p} 0.$$

To complete the consistency proof, we need to show that

$$\sup_{\boldsymbol{\beta}\in\Theta} \left| \int_{\mathbb{R}^k} \left(||\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u})||^2 - \|\mathbf{H}_{\exp}(\boldsymbol{\beta}, \mathbf{u})\|^2 \right) \phi(\mathbf{u})\, d\mathbf{u} \right| \xrightarrow{p} 0,$$

where $\phi(\mathbf{u})$ is the standard k-variate normal density function. Fix an arbitrarily small $\epsilon > 0$ and choose a compact and convex set Π such that

$$\left| \int_{\mathbb{R}^k\setminus\Pi} \phi(\mathbf{u})\, d\mathbf{u} \right| \leq \epsilon. \tag{S6.2}$$

Then, write

$$\begin{aligned}
&\int_{\mathbb{R}^k} ||\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u})||^2 \phi(\mathbf{u})\, d\mathbf{u} \\
&= \int_{\Pi} ||\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u})||^2 \phi(\mathbf{u})\, d\mathbf{u} + \int_{\mathbb{R}^k\setminus\Pi} ||\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u})||^2 \phi(\mathbf{u})\, d\mathbf{u} \\
&\equiv J_{1n} + J_{2n}.
\end{aligned}$$

From Lemma S5.2 and the CMT, we have that

$$\sup_{(\boldsymbol{\beta},\mathbf{u})\in\Theta\times\Pi}\|\mathbf{H}_{n,\exp}(\boldsymbol{\beta},\mathbf{u})-\mathbf{H}_{\exp}(\boldsymbol{\beta},\mathbf{u})\|\xrightarrow{p}0.$$

Thus, by another application of the CMT, it follows that

$$J_{1n}=\int_{\Pi}\|\mathbf{H}_{\exp}(\boldsymbol{\beta},\mathbf{u})\|^{2}\,\phi\left(\mathbf{u}\right)d\mathbf{u}+o_{p}\left(1\right)$$

uniformly in $\boldsymbol{\beta}\in\Theta$. For J_{2n}, since $|\exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right))|\leq 1$, we have that under Assumption 2(ii)

$$\|\mathbf{H}_{n,\exp}(\boldsymbol{\beta},\mathbf{u})\|\leq C$$

for some $C<\infty$. Hence, by (S6.2) we have that $J_{2n}=O_p(\epsilon)$. Since $\epsilon>0$ is arbitrary, we conclude that

$$\sup_{\boldsymbol{\beta}\in\Theta}|Q_{n,\exp}(\boldsymbol{\beta})-Q_{\exp}(\boldsymbol{\beta})|\xrightarrow{p}0.$$

Next, we derive the asymptotic linear representation of $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}-\boldsymbol{\beta}_0\right)$. Towards this end, note that the first order condition of $\min_{\beta}Q_{n,w}(\boldsymbol{\beta})$ is given as follows

$$\int\left\{\dot{\mathbf{H}}^{c}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\mathbf{H}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})+\left(\mathbf{H}^{c}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\dot{\mathbf{H}}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\right)'\right\}\Psi_n(d\mathbf{u})=0. \tag{S6.3}$$

By the mean value theorem (after properly extending to the case of complex-valued functions of real variables), and Assumption 2 (i), we have that

$$\mathbf{H}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})=\mathbf{H}_{n,w}(\boldsymbol{\beta}_0,\mathbf{u})+\dot{\mathbf{H}}_{n,w}(\widetilde{\boldsymbol{\beta}},\mathbf{u})(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}-\boldsymbol{\beta}_0), \tag{S6.4}$$

where $\widetilde{\boldsymbol{\beta}}$ satisfies $\left\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|\leq\left\|\widehat{\boldsymbol{\beta}}^{ips}_{n,w}-\boldsymbol{\beta}_0\right\|$. Plugging (S6.4) into (S6.3), we can write

$$\int\left(\dot{\mathbf{H}}^{c}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\mathbf{H}_{n,w}(\boldsymbol{\beta}_0,\mathbf{u})+\dot{\mathbf{H}}'_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\left(\mathbf{H}'_{n,w}(\boldsymbol{\beta}_0,\mathbf{u})\right)^{c}\right)\Psi_n(d\mathbf{u})$$
$$+\widehat{C}_{w,\Psi_n}\left(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}-\boldsymbol{\beta}_0\right)=0$$

where

$$\widehat{C}_{w,\Psi_n}=\int\left(\dot{\mathbf{H}}^{c}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\dot{\mathbf{H}}_{n,w}(\widetilde{\boldsymbol{\beta}},\mathbf{u})+\dot{\mathbf{H}}'_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\left(\dot{\mathbf{H}}'_{n,w}(\widetilde{\boldsymbol{\beta}},\mathbf{u})\right)^{c}\right)\Psi_n(d\mathbf{u}).$$

Therefore

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}-\boldsymbol{\beta}_0\right)$$

$$= -\widehat{C}_{w,\Psi_n}^{-1} \cdot \sqrt{n} \int \left(\dot{\mathbf{H}}_{n,w}^{c}(\widehat{\boldsymbol{\beta}}_{n,w}^{ips}, \mathbf{u}) \mathbf{H}_{n,w}(\boldsymbol{\beta}_0, \mathbf{u}) + \dot{\mathbf{H}}_{n,w}^{\prime}(\widehat{\boldsymbol{\beta}}_{n,w}^{ips}, \mathbf{u}) \left(\mathbf{H}_{n,w}^{\prime}(\boldsymbol{\beta}_0, \mathbf{u}) \right)^{c} \right) \Psi_n(d\mathbf{u}). \quad \text{(S6.5)}$$

By exploiting that $\mathbf{H}_{n,w}(\boldsymbol{\beta}_0, \mathbf{u}) = \mathbb{E}_n \left[\mathbf{h}_n (D, \mathbf{X}; \boldsymbol{\beta}_0) \, w(\mathbf{X}; \mathbf{u}) \right]$, we can express (S6.5) as

$$\sqrt{n} \left(\widehat{\boldsymbol{\beta}}_{n,w}^{ips} - \boldsymbol{\beta}_0 \right)$$
$$= -\widehat{C}_{w,\Psi_n}^{-1} \cdot \sqrt{n} \mathbb{E}_n \left[\int \left(\dot{\mathbf{H}}_{n,w}^{c}(\widehat{\boldsymbol{\beta}}_{n,w}^{ips}, \mathbf{u}) w(\mathbf{X}; \mathbf{u}) + \dot{\mathbf{H}}_{n,w}^{\prime}(\widehat{\boldsymbol{\beta}}_{n,w}^{ips}, \mathbf{u}) w^{c}(\mathbf{X}; \mathbf{u}) \right) \Psi_n(d\mathbf{u}) \right.$$
$$\left. \cdot \mathbf{h}_n (D, \mathbf{X}; \boldsymbol{\beta}_0) \right] \quad \text{(S6.6)}$$

From Lemma S5.3 we have that

$$\widehat{C}_{w,\Psi_n} = C_{w,\Psi} + o_p(1), \quad \text{(S6.7)}$$

whereas by Lemma S5.5 we have that

$$\sqrt{n} \mathbb{E}_n \left[\int \left(\dot{\mathbf{H}}_{w}^{c}(\widehat{\boldsymbol{\beta}}_{n,w}^{ips}, \mathbf{u}) w(\mathbf{X}; \mathbf{u}) + \dot{\mathbf{H}}_{n,w}^{\prime}(\widehat{\boldsymbol{\beta}}_{n,w}^{ips}, \mathbf{u}) w^{c}(\mathbf{X}; \mathbf{u}) \right) \Psi_n(d\mathbf{u}) \cdot \mathbf{h}_n (D, \mathbf{X}; \boldsymbol{\beta}_0) \right]$$
$$= \sqrt{n} \mathbb{E}_n \left[\int \left(\dot{\mathbf{H}}_{w}^{c}(\boldsymbol{\beta}_{\mathbf{0}}, \mathbf{u}) w(\mathbf{X}; \mathbf{u}) + \dot{\mathbf{H}}_{w}^{\prime}(\boldsymbol{\beta}_{\mathbf{0}}, \mathbf{u}) w^{c}(\mathbf{X}, \mathbf{u}) \right) \Psi(d\mathbf{u}) \cdot \mathbf{h} (D, \mathbf{X}; \boldsymbol{\beta}_0) \right] + o_p(1). \quad \text{(S6.8)}$$

Thus, from (S6.6)-(S6.8), we conclude that

$$\sqrt{n} \left(\widehat{\boldsymbol{\beta}}_{n,w}^{ips} - \boldsymbol{\beta}_0 \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l_{w,\Psi} (D_i, \mathbf{X}_i; \boldsymbol{\beta}_0) + o_p(1),$$

with

$$l_{w,\Psi} (D, \mathbf{X}; \boldsymbol{\beta}_0) = -C_{w,\Psi}^{-1} \cdot \int \left(\dot{\mathbf{H}}_{w}^{c}(\boldsymbol{\beta}_0, \mathbf{u}) w(\mathbf{X}; \mathbf{u}) + \dot{\mathbf{H}}_{w}^{\prime}(\boldsymbol{\beta}_0, \mathbf{u}) w^{c}(\mathbf{X}; \mathbf{u}) \right) \Psi(d\mathbf{u}) \cdot \mathbf{h} (D, \mathbf{X}; \boldsymbol{\beta}_0).$$

Since $\dot{\mathbf{H}}_{w}^{c}(\boldsymbol{\beta}_0, \mathbf{u}) w(\mathbf{X}; \mathbf{u}) + \dot{\mathbf{H}}_{w}^{\prime}(\boldsymbol{\beta}_0, \mathbf{u}) w^{c}(\mathbf{X}; \mathbf{u})$ is real-valued, under Assumptions 2-3, as long as $\mathbb{E}[||l_{w,\Psi}(D, X; \beta_0)||^2] < \infty$, the asymptotic normality result follows directly from the application of the standard central limit theorem. Next, we show that $l_{w,\Psi}(D, \mathbf{X}; \boldsymbol{\beta}_0)$ is indeed square integrable when $C_{w,\Psi}$ is positive definite. Here, it suffices to show that $\mathbb{E} \left[||s(D, \mathbf{X}; \boldsymbol{\beta}_{\mathbf{0}})||^2 \right] < \infty$ where

$$s(D, \mathbf{X}; \boldsymbol{\beta}_{\mathbf{0}}) \equiv \int \left\{ \dot{\mathbf{H}}_{w}^{c}(\boldsymbol{\beta}_0, \mathbf{u}) w(\mathbf{X}; \mathbf{u}) + \dot{\mathbf{H}}_{w}^{\prime}(\boldsymbol{\beta}_0, \mathbf{u}) w^{c}(\mathbf{X}; \mathbf{u}) \right\} \Psi(d\mathbf{u}) \cdot \mathbf{h} (D, \mathbf{X}; \boldsymbol{\beta}_0).$$

Let $K(\mathbf{x}, \mathbf{u}; \boldsymbol{\beta}_{\mathbf{0}}) \equiv \dot{\mathbf{H}}_{w}^{c}(\boldsymbol{\beta}_0, \mathbf{u}) w(\mathbf{x}; \mathbf{u}) + \dot{\mathbf{H}}_{w}^{\prime}(\boldsymbol{\beta}_0, \mathbf{u}) w^{c}(\mathbf{x}; \mathbf{u})$. Then, for some $C < \infty$,

$$||s(d, \mathbf{x}; \boldsymbol{\beta}_{\mathbf{0}})||^2 \leq \int ||K(\mathbf{x}, \mathbf{u}; \boldsymbol{\beta}_{\mathbf{0}}) \mathbf{h} (d, \mathbf{x}; \boldsymbol{\beta}_0) ||^2 \Psi(d\mathbf{u})$$

$$\leq \int ||K(\mathbf{x},\mathbf{u};\boldsymbol{\beta_0})||^2 \cdot ||\mathbf{h}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right)||^2 \Psi(d\mathbf{u})$$

$$\leq C \cdot \int ||K(\mathbf{x},\mathbf{u};\boldsymbol{\beta_0})||^2 \Psi(d\mathbf{u}),$$

where the first inequality follows from Jensen's inequality, the second inequality follows from the Cauchy-Schwarz inequality, and the third from Assumption 2(ii). Finally, from Lemma S5.1 we have $w(\mathbf{X};\mathbf{u})$ is uniformly bounded, and therefore,

$$\begin{aligned}\mathbb{E}\left[||K(\mathbf{X},\mathbf{u};\boldsymbol{\beta_0})||^2\right] &\leq C \cdot ||\dot{\mathbf{H}}_w(\boldsymbol{\beta}_0,\mathbf{u})||^2 \\ &\leq C \cdot \left(\mathbb{E}\left[\left\|\dot{\mathbf{h}}\left(\mathbf{X};\boldsymbol{\beta}_0\right)\right\|\right]\right)^2 \\ &< \infty,\end{aligned}$$

where the last inequality follows from Assumption 2(iv). This concludes our proof.■

Proof of Theorem 3.2: The proof is divided into three parts.

Part 1: Asymptotic Properties of the Average Treatment Effect.

It suffices to show that

$$\sqrt{n}\left(\widehat{ATE}_n^{ips} - ATE\right) = \sqrt{n}\mathbb{E}_n\left[\psi_{w,\Psi}^{ate}\left(Y,D,\mathbf{X}\right)\right] + o_p\left(1\right) \quad \text{(S6.9)}$$

where $\psi_{w,\Psi}^{ate}\left(Y,D,\mathbf{X}\right) = g^{ate}(Y,X,D) - l_{w,\Psi}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)' \mathbf{G}_{\boldsymbol{\beta}}^{ate}$,

$$\mathbb{E}\left[\psi_{w,\Psi}^{ate}\left(Y,D,\mathbf{X}\right)\right] = 0$$

and

$$\mathbb{E}\left[\psi_{w,\Psi}^{ate}\left(Y,D,\mathbf{X}\right)^2\right] < \infty,$$

i.e., that $\sqrt{n}\left(\widehat{ATE}_n^{ips} - ATE\right)$ admits an asymptotically linear representation. We show this by combining the mean value theorem, continuos mapping theorem, and the results in Theorem 3.1.

We first show that

$$\begin{aligned}&\mathbb{E}_n[\varpi_{n,1}^{ps}(D,\mathbf{X};\widehat{\boldsymbol{\beta}}_{n,w}^{ips})Y - \mathbb{E}\left[Y\left(1\right)\right]] \\ &\quad = \mathbb{E}_n\left[\varpi_1^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\cdot\left(Y - \mathbb{E}\left[Y\left(1\right)\right]\right) - l_{w,\Psi}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)' \mathbf{G}_{\boldsymbol{\beta},1}^{ate}\right] + o_p\left(n^{-1/2}\right),\end{aligned} \quad \text{(S6.10)}$$

where

$$\mathbf{G}_{\boldsymbol{\beta},1}^{ate} = \mathbb{E}\left[\frac{g_1^{ate}}{p(\mathbf{X};\boldsymbol{\beta_0})}\cdot\dot{p}\left(\mathbf{X};\boldsymbol{\beta_0}\right)\right],$$

and $g_1^{ate}\left(Y,D,\mathbf{X}\right) = \varpi_1^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\cdot\left(Y - \mathbb{E}\left[Y\left(1\right)\right]\right)$. By the mean value theorem and some manipulations, we have that

$$\begin{aligned}
&\mathbb{E}_n[\varpi_{n,1}^{ps}(D,\mathbf{X};\widehat{\boldsymbol{\beta}}_{n,w}^{ips})Y]\\
&=\mathbb{E}_n\left[\varpi_{n,1}^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)Y\right]\\
&-\mathbb{E}_n\left[\frac{\varpi_{n,1}^{ps}\left(D,\mathbf{X};\widetilde{\boldsymbol{\beta}}\right)\left(Y-\mathbb{E}_n\left[\varpi_{n,1}^{ps}\left(D,\mathbf{X};\widetilde{\boldsymbol{\beta}}\right)Y\right]\right)\cdot\dot{p}\left(\mathbf{X};\widetilde{\boldsymbol{\beta}}\right)'}{p\left(\mathbf{X};\widetilde{\boldsymbol{\beta}}\right)}\right]\left(\widehat{\boldsymbol{\beta}}_{n,w}^{ips}-\boldsymbol{\beta}_0\right),
\end{aligned}$$

where $\widetilde{\boldsymbol{\beta}}$ satisfies $\left\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\|\leq\left\|\widehat{\boldsymbol{\beta}}_{n,w}^{ips}-\boldsymbol{\beta}_0\right\|$. From Theorem 3.1, we have that

$$\begin{aligned}
\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{n,w}^{ips}-\boldsymbol{\beta}_0\right)&=\sqrt{n}\mathbb{E}_n\left[l_{w,\Psi}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right]+o_p\left(1\right)\\
&=O_p\left(1\right),
\end{aligned}$$

and therefore, by the CMT, under Assumptions 2-3,

$$\begin{aligned}
&\mathbb{E}_n[\varpi_{n,1}^{ps}(D,\mathbf{X};\widehat{\boldsymbol{\beta}}_{n,w}^{ips})Y]\\
&\qquad=\mathbb{E}_n\left[\varpi_{n,1}^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)Y\right]-\mathbb{E}_n\left[l_{w,\Psi}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)'\cdot\mathbf{G}_{\boldsymbol{\beta},1}^{ate}\right]+o_p\left(n^{-1/2}\right). \quad \text{(S6.11)}
\end{aligned}$$

Given that, under Assumption 2,

$$\mathbb{E}_n\left[\frac{D}{p\left(\mathbf{X};\boldsymbol{\beta}_0\right)}\right]-\mathbb{E}\left[\frac{D}{p\left(\mathbf{X};\boldsymbol{\beta}_0\right)}\right]=O_p\left(n^{-1/2}\right),$$

we have that, by the CMT,

$$\begin{aligned}
\mathbb{E}_n\left[\varpi_{n,1}^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)Y\right]&=\mathbb{E}_n\left[\varpi_1^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\left(Y-\mathbb{E}\left[\varpi_1^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)Y\right]\right)\right]\\
&\quad+\mathbb{E}\left[\varpi_1^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)Y\right]+o_p\left(n^{-1/2}\right),\\
&=\mathbb{E}_n\left[\varpi_1^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\left(Y-\mathbb{E}\left[\varpi_1^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)Y\right]\right)\right]\\
&\quad+\mathbb{E}\left[Y\left(1\right)\right]+o_p\left(n^{-1/2}\right), \qquad \text{(S6.12)}
\end{aligned}$$

where the last step follows from Assumption 1. Hence, from (S6.11) and (S6.12), we conclude the proof of (S6.10).

By symmetry, we have that

$$\begin{aligned}
&\mathbb{E}_n[\varpi_{n,0}^{ps}(D,\mathbf{X};\widehat{\boldsymbol{\beta}}_{n,w}^{ips})Y-\mathbb{E}\left[Y\left(0\right)\right]]\\
&\quad=\mathbb{E}_n\left[\varpi_0^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\cdot\left(Y-\mathbb{E}\left[Y\left(0\right)\right]\right)+l_{w,\Psi}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)'\mathbf{G}_{\boldsymbol{\beta},0}^{ate}\right]+o_p\left(n^{-1/2}\right), \quad \text{(S6.13)}
\end{aligned}$$

where

$$\mathbf{G}_{\boldsymbol{\beta},0}^{ate} = \mathbb{E}\left[\frac{g_0^{ate}}{1-p(\mathbf{X};\boldsymbol{\beta}_0)} \cdot \dot{p}\left(\mathbf{X};\boldsymbol{\beta}_0\right)\right],$$

and $g_0^{ate}\left(Y,D,\mathbf{X}\right) = \varpi_0^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right) \cdot \left(Y - \mathbb{E}\left[Y\left(0\right)\right]\right)$.

By combining (S6.10) with (S6.13), we have that

$$\sqrt{n}\left(\widehat{ATE}_n^{ips} - ATE\right) = \mathbb{E}_n\left[\psi_{w,\Psi}^{ate}\left(Y,D,\mathbf{X}\right)\right] + o_p\left(1\right),$$

where $\mathbb{E}\left[\psi_{w,\Psi}^{ate}\left(Y,D,\mathbf{X}\right)\right] = 0$ follows from the law of iterated expectations and Assumption 1. Next, note that

$$\begin{aligned}
\mathbb{E}[\psi_{w,\Psi}^{ate}(Y,D,\mathbf{X})^2] &= \mathbb{E}[(g^{ate}\left(Y,D,\mathbf{X}\right) - l_{w,\Psi}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)'\,\mathbf{G}_{\boldsymbol{\beta}}^{ate})^2] \\
&= \mathbb{E}\left[g^{ate2} - 2g^{ate}l_{w,\Psi}'\mathbf{G}_{\boldsymbol{\beta}}^{ate} + (l_{w,\Psi}'\mathbf{G}_{\boldsymbol{\beta}}^{ate})^2\right] \\
&\leq \mathbb{E}[g^{ate2}] + \mathbb{E}[(l_{w,\Psi}'\mathbf{G}_{\boldsymbol{\beta}}^{ate})^2] + 2\mathbb{E}[|g^{ate}l_{w,\Psi}'\mathbf{G}_{\boldsymbol{\beta}}^{ate}|] \qquad \text{(S6.14)}
\end{aligned}$$

Let $C_1 \equiv \sup_{d,\mathbf{x}} \varpi_1^{ps}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right)^2$ and $C_2 \equiv \sup_{d,\mathbf{x}} \varpi_0^{ps}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right)^2$, and note that, under Assumption 2(ii), $1 \leq C_1, C_2 < \infty$. Then

$$\begin{aligned}
\mathbb{E}[g^{ate2}] &= \mathbb{E}[\varpi_1^{ps2}(Y - \mathbb{E}[Y(1)])^2 + \varpi_0^{ps2}(Y - \mathbb{E}[Y(0)])^2] \\
&\leq C_1\mathbb{E}[(Y\left(1\right) - \mathbb{E}[Y(1)])^2] + C_2\mathbb{E}[(Y\left(0\right) - \mathbb{E}[Y(0)])^2] \\
&< \infty, \qquad \text{(S6.15)}
\end{aligned}$$

where the first equality follows from $\varpi_1^{ps} \cdot \varpi_0^{ps} = 0$ $a.s.$, the first inequality follows from Assumption (1) and Assumption 2(ii), whereas the last inequality follows from Assumption 4(i).

Next, by Cauchy-Schwarz inequality, Theorem 3.1, and Assumption 4(ii),

$$\begin{aligned}
\mathbb{E}[(l_{w,\Psi}'\mathbf{G}_{\boldsymbol{\beta}}^{ate})^2] &\leq ||\mathbf{G}_{\boldsymbol{\beta}}^{ate}||^2 \cdot \mathbb{E}[||l_{w,\Psi}||^2] \\
&< \infty, \qquad \text{(S6.16)}
\end{aligned}$$

whereas, by Cauchy-Schwarz inequality, (S6.15) and (S6.16),

$$\begin{aligned}
\mathbb{E}[|g^{ate}l_{w,\Psi}'\mathbf{G}_{\boldsymbol{\beta}}^{ate}|] &\leq \mathbb{E}[|g^{ate}|^2]^{1/2} \cdot \mathbb{E}[|l_{w,\Psi}'\mathbf{G}_{\boldsymbol{\beta}}^{ate}|^2]^{1/2} \\
&< \infty. \qquad \text{(S6.17)}
\end{aligned}$$

Hence, $\mathbb{E}[\psi_{w,\Psi}^{ate}(Y,D,\mathbf{X})^2] < \infty$ follows from (S6.14)-(S6.17), which concludes the proof of (S6.9).

Part 2: Asymptotic Properties of the Distribution Treatment Effects.

The (uniform) asymptotic linear representation for the Distribution Treatment Effect parameter follows from exactly the same steps as in Part 1 and is therefore omitted. Next, we show that the classes of functions

$$\mathcal{F}_{1,dte} \equiv \left\{ (z,d,\mathbf{x}) \in \mathbb{R}\times \{0,1\} \times \mathcal{X} \mapsto \psi^{dte}_{1,w,\Psi}(z,d,\mathbf{x};y) : y \in [-\infty,\infty] \right\},$$

$$\mathcal{F}_{0,dte} \equiv \left\{ (z,d,\mathbf{x}) \in \mathbb{R}\times \{0,1\} \times \mathcal{X} \mapsto \psi^{dte}_{0,w,\Psi}(z,d,\mathbf{x};y) : y \in [-\infty,\infty] \right\}$$

are Donsker, where

$$\psi^{dte}_{1,w,\Psi}(z,d,\mathbf{x};y) = g^{dte}_{1}(z,d,\mathbf{x};y) - l_{w,\Psi}(d,\mathbf{x};\boldsymbol{\beta}_0)' \cdot \mathbf{G}^{dte}_{1,\boldsymbol{\beta}}(y),$$

$$\psi^{dte}_{0,w,\Psi}(z,d,\mathbf{x};y) = g^{dte}_{0}(z,d,\mathbf{x};y) + l_{w,\Psi}(d,\mathbf{x};\boldsymbol{\beta}_0)' \cdot \mathbf{G}^{dte}_{0,\boldsymbol{\beta}}(y),$$

and

$$\mathbf{G}^{dte}_{1,\boldsymbol{\beta}}(y) = \mathbb{E}\left[\frac{g^{dte}_{1}(Y,D,\mathbf{X};y)}{p(\mathbf{X};\boldsymbol{\beta}_0)} \cdot \dot{p}(\mathbf{X};\boldsymbol{\beta}_0) \right],$$

$$\mathbf{G}^{dte}_{0,\boldsymbol{\beta}}(y) = \mathbb{E}\left[\frac{g^{dte}_{0}(Y,D,\mathbf{X};y)}{1-p(\mathbf{X};\boldsymbol{\beta}_0)} \cdot \dot{p}(\mathbf{X};\boldsymbol{\beta}_0) \right].$$

Toward this end, note that the classes of functions $\left\{ l_{w,\Psi}(d,\mathbf{x};\boldsymbol{\beta}_0)' \cdot \mathbf{G}^{dte}_{1,\boldsymbol{\beta}}(y) : y \in [-\infty,\infty] \right\}$ and $\left\{ l_{w,\Psi}(d,\mathbf{x};\boldsymbol{\beta}_0)' \cdot \mathbf{G}^{dte}_{0,\boldsymbol{\beta}}(y) : y \in [-\infty,\infty] \right\}$are Donsker since they depend on y in a deterministic manner, $\mathbf{G}^{dte}_{d,\boldsymbol{\beta}}(y) < \infty$, $d \in \{0,1\}$, and, by Theorem 3.1, $\mathbb{E}[||l_{w,\Psi}||^2] < \infty$. The Donsker property of $\left\{ g^{dte}_{1}(z,d,\mathbf{x};y) : y \in [-\infty,\infty] \right\}$ and $\left\{ g^{dte}_{0}(z,d,\mathbf{x};y) : y \in [-\infty,\infty] \right\}$ follows from Lemma S5.1, Assumption 2(ii), and Corollary 9.32 in Kosorok (2008). Thus, from Corollary 9.32 in Kosorok (2008), we conclude that $\mathcal{F}_{1,dte}$ and $\mathcal{F}_{0,dte}$ are Donsker.

Let $\boldsymbol{\lambda}^{dte}_{w,\Psi}(z,d,\mathbf{x};\cdot) = \left(\psi^{dte}_{1,w,\Psi}(z,d,\mathbf{x};\cdot), \psi^{dte}_{0,w,\Psi}(z,d,\mathbf{x};\cdot) \right)'$, and denote

$$\mathbb{G}^{dte,(1,0)}_{n,w,\Psi}(\cdot) = \sqrt{n}\mathbb{E}_n \left[\boldsymbol{\lambda}^{dte}_{w,\Psi}(Y,D,\mathbf{X};\cdot) \right].$$

Thus, under Assumptions 1-4,

$$\mathbb{G}^{dte,(1,0)}_{n,w,\Psi}(\cdot) \Rightarrow \mathbb{G}^{dte,(1,0)}_{\infty,w,\Psi}(\cdot) \text{ in } \ell^{\infty}([-\infty,\infty]) \times \ell^{\infty}([-\infty,\infty]), \tag{S6.18}$$

where $\Rightarrow$ denotes weak convergence in the sense of J. Hoffman-Jørgensen, see e.g. van der Vaart and Wellner (1996), $\ell^{\infty}(T)$ is the collection of all bounded functions $f : T \mapsto \mathbb{R}$, and $\mathbb{G}^{dte,(1,0)}_{\infty,w,\Psi}(\cdot)$ is a tight, two-dimensional mean zero Gaussian process with covariance kernel

$$\Gamma(\mathbf{y}_1,\mathbf{y}_2) = \mathbb{E}\left[\boldsymbol{\lambda}^{dte}_{w,\Psi}(Y,D,\mathbf{X};\mathbf{y}_1)\, \boldsymbol{\lambda}^{dte}_{w,\Psi}(Y,D,\mathbf{X};\mathbf{y}_2)' \right],$$

in which

$$\boldsymbol{\lambda}^{dte}_{w,\Psi}(z,d,\mathbf{x};\mathbf{y}) = \left(\psi^{dte}_{1,w,\Psi}(z,d,\mathbf{x};y_1), \psi^{dte}_{0,w,\Psi}(z,d,\mathbf{x};y_2) \right)'.$$

By the CMT, we have that

$$\sqrt{n}\left(\widehat{DTE}_n^{ips} - DTE\right)(\cdot) = (1,-1)\,\mathbb{G}_{n,w,\Psi}^{dte,(1,0)}(\cdot) + o_p(1),$$

$$\Rightarrow \mathbb{G}_{\infty,w,\Psi}^{dte}(\cdot) \text{ in } \ell^{\infty}([-\infty,\infty])$$

where $\mathbb{G}_{\infty,w,\Psi}^{dte}(\cdot)$ is a tight, univariate mean zero Gaussian process with covariance kernel

$$\Gamma_{dte}(y_1,y_2) = \mathbb{E}\left[\psi_{w,\Psi}^{dte}(Y,D,\mathbf{X};y_1)\,\psi_{w,\Psi}^{dte}(Y,D,\mathbf{X};y_2)\right].$$

The asymptotic normality result now follows by simply fixing y.

Part 3: Asymptotic Properties of the Quantile Treatment Effects.

Define $\widehat{\mathbf{q}}_n^{ips}(\boldsymbol{\tau}) = \left(\widehat{q}_{n,Y(1)}^{ips}(\tau_1), \widehat{q}_{n,Y(0)}^{ips}(\tau_2)\right)'$, $\mathbf{q}(\boldsymbol{\tau}) = \left(q_{Y(1)}(\tau_1), q_{Y(0)}(\tau_2)\right)'$, and also $\mathbf{f}^{-1}(\tau) = \left(f_{Y(1)}^{-1}\left(q_{Y(1)}(\tau_1)\right), f_{Y(0)}^{-1}\left(q_{Y(0)}(\tau_2)\right)\right)'$ and $\boldsymbol{\tau} = (\tau_1,\tau_2) \in [a_1,a_2]^2$, where a_1 and a_2 satisfy $0 < a_1 < a_2 < 1$. Under Assumptions 1-4, we have that, from (S6.18), Lemma 21.4 in van der Vaart (1998), and the functional delta method, see e.g. Theorem 20.8 in van der Vaart (1998),

$$\sqrt{n}\left(\widehat{\mathbf{q}}_n^{ips} - \mathbf{q}\right)(\cdot) = -\mathbf{f}^{-1}(\cdot)' \cdot \mathbb{G}_{n,w,\Psi}^{dte,(1,0)}(\mathbf{q}(\cdot)) + o_p(1)$$

$$\Rightarrow -\mathbf{f}^{-1}(\cdot)' \cdot \mathbb{G}_{\infty,w,\Psi}^{dte,(1,0)}(\mathbf{q}(\cdot)) \text{ in } \ell^{\infty}([a_1,a_2]) \times \ell^{\infty}([a_1,a_2]).$$

Then, by the CMT,

$$\sqrt{n}\left(\widehat{QTE}_n^{ips} - QTE\right)(\cdot) = (1,-1)\cdot\left(-\mathbf{f}^{-1}(\cdot)' \cdot \mathbb{G}_{n,w,\Psi}^{dte,(1,0)}(\mathbf{q}(\cdot))\right) + o_p(1)$$

$$\Rightarrow \mathbb{G}_{\infty,w,\Psi}^{qte}(\cdot) \text{ in } \ell^{\infty}[a_1,a_2]$$

where $\mathbb{G}_{\infty,w,\Psi}^{qte}(\cdot)$ is a tight, mean zero Gaussian process with covariance kernel

$$\Gamma_{qte}(\tau_1,\tau_2) = \mathbb{E}\left[\psi_{w,\Psi}^{qte}(Y,D,\mathbf{X};\tau_1)\,\psi_{w,\Psi}^{qte}(Y,D,\mathbf{X};\tau_2)\right].$$

The asymptotic normality result now follows by simply fixing τ.■

S7 Proofs of the Results under Endogeneity

In this section, we will introduce additional notations for and present the proof of Theorem 4.1 and 4.2. First of all, it is straightforward to show that the LIPS objective function $Q_{n,w}^{lte}(\boldsymbol{\beta}) = \int_{\Pi}\left\|\mathbf{H}_{n,w}^{lte}(\boldsymbol{\beta},\mathbf{u})\right\|^2 \Psi_n(d\mathbf{u})$, has closed-form representation analogous to the IPS case. All the proofs in Appendix S3 follow through by replacing $\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta})$ with $\mathbf{h}_n^{lte}(D,Z,\mathbf{X};\boldsymbol{\beta})$.

S7.I Lemmas for Section 4

Before proving Theorem 4.1 and 4.2, we point out that results similar to Lemma 2.2 continue to hold and so do the auxiliary lemmas analogous to those in Appendix S5.

Lemma S7.1 *(Lemma 2.2) Let $\Theta \subset \mathbb{R}^k$ be the parameter space, and assume that (4.3) is satisfied for a unique $\boldsymbol{\beta}_0^{lte} \in \Theta$. Then $Q_w^{lte}(\boldsymbol{\beta}) \geq 0$, $\forall \boldsymbol{\beta} \in \Theta$, and $Q_w^{lte}(\boldsymbol{\beta}_0^{lte}) = 0$ if and only if the the covariate balancing condition (4.3) holds.*

Proof of Lemma S7.1: Note that $Q_w^{lte}(\boldsymbol{\beta}) \geq 0$ follows trivially from the definition. Next, analogous to the discussion in Section 2.2, the covariate balancing condition among compliers (4.3) is equivalent to (4.4), implying that $Q_w^{lte}(\boldsymbol{\beta}_0^{lte}) = 0$.

To complete the proof we then need to show that if $Q_w^{lte}(\boldsymbol{\beta}) = 0$, then $\boldsymbol{\beta} = \boldsymbol{\beta}_0^{lte}$. Towards this end, recall that if $Q_w^{lte}(\boldsymbol{\beta}) = 0$, it must be that $\mathbf{H}_w^{lte}(\boldsymbol{\beta}, \boldsymbol{u}) = 0$ *a.e.* on Π, because $|| \cdot || \geq 0$ and the integrating probability measure Ψ is absolutely continuous with respect to a dominating measure on Π. However, $\mathbf{H}_w^{lte}(\boldsymbol{\beta}, \boldsymbol{u}) = 0$ *a.e.* on Π if and only if $\mathbb{E}\left[\mathbf{h}^{lte}(D, Z, \mathbf{X}; \boldsymbol{\beta}) | \mathbf{X}\right] = 0$ *a.s.*, which is equivalent to

$$\frac{1}{\kappa_1(\boldsymbol{\beta})}\mathbb{E}\left[\frac{DZ}{q(\mathbf{X};\boldsymbol{\beta})} - \frac{D(1-Z)}{1-q(\mathbf{X};\boldsymbol{\beta})}\middle| \mathbf{X}\right] = \frac{1}{\kappa(\boldsymbol{\beta})}\mathbb{E}\left[1 - \frac{(1-D)Z}{q(\mathbf{X};\boldsymbol{\beta})} - \frac{D(1-Z)}{1-q(\mathbf{X};\boldsymbol{\beta})}\middle| \mathbf{X}\right] \quad a.s.,$$

and

$$\frac{1}{\kappa_0(\boldsymbol{\beta})}\mathbb{E}\left[\frac{(1-D)Z}{q(\mathbf{X};\boldsymbol{\beta})} - \frac{(1-D)(1-Z)}{1-q(\mathbf{X};\boldsymbol{\beta})}\middle| \mathbf{X}\right] = \frac{1}{\kappa(\boldsymbol{\beta})}\mathbb{E}\left[1 - \frac{(1-D)Z}{q(\mathbf{X};\boldsymbol{\beta})} - \frac{D(1-Z)}{1-q(\mathbf{X};\boldsymbol{\beta})}\middle| \mathbf{X}\right] \quad a.s..$$

Let $p_{dz}(\mathbf{X}) = \mathbb{P}(D = d | \mathbf{X}, Z = z)$ for $d, z \in \{0, 1\}$. By straightforward calculation, the two equations given above are further equivalent to

$$\begin{aligned}\frac{1}{\kappa_1(\boldsymbol{\beta})}&\left[p_{11}(\mathbf{X})\frac{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{q(\mathbf{X};\boldsymbol{\beta})} - p_{10}(\mathbf{X})\frac{1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{1-q(\mathbf{X};\boldsymbol{\beta})}\right] \\ &= \frac{1}{\kappa(\boldsymbol{\beta})}\left[1 - p_{01}(\mathbf{X})\frac{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{q(\mathbf{X};\boldsymbol{\beta})} - p_{10}(\mathbf{X})\frac{1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{1-q(\mathbf{X};\boldsymbol{\beta})}\right] \quad a.s., \end{aligned} \tag{S7.1}$$

and

$$\begin{aligned}\frac{1}{\kappa_0(\boldsymbol{\beta})}&\left[p_{01}(\mathbf{X})\frac{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{q(\mathbf{X};\boldsymbol{\beta})} - p_{00}(\mathbf{X})\frac{1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{1-q(\mathbf{X};\boldsymbol{\beta})}\right] \\ &= \frac{1}{\kappa(\boldsymbol{\beta})}\left[1 - p_{01}(\mathbf{X})\frac{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{q(\mathbf{X};\boldsymbol{\beta})} - p_{10}(\mathbf{X})\frac{1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{1-q(\mathbf{X};\boldsymbol{\beta})}\right] \quad a.s.. \end{aligned} \tag{S7.2}$$

Denote $c_1(\mathbf{X}) = q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)/q(\mathbf{X};\boldsymbol{\beta})$ and $c_0(\mathbf{X}) = \left(1 - q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right)/(1-q(\mathbf{X};\boldsymbol{\beta}))$. Note that

$p_{1z}(\mathbf{X}) + p_{0z}(\mathbf{X}) = 1$ $a.s.$ for $z \in \{0,1\}$. Then dividing (S7.1) by (S7.2) and rearranging yield

$$\begin{aligned}\frac{\kappa_0(\boldsymbol{\beta})}{\kappa_1(\boldsymbol{\beta})} =& \frac{(1-p_{11}(\mathbf{X}))\,c_1(\mathbf{X}) - (1-p_{10}(\mathbf{X}))\,c_0(\mathbf{X})}{p_{11}(\mathbf{X})\,c_1(\mathbf{X}) - p_{10}(\mathbf{X})\,c_0(\mathbf{X})} \\ =& \frac{c_1(\mathbf{X}) - c_0(\mathbf{X})}{p_{11}(\mathbf{X})\,c_1(\mathbf{X}) - p_{10}(\mathbf{X})\,c_0(\mathbf{X})} - 1 \;\; a.s.,\end{aligned}$$

which implies $c_1(\mathbf{X}) = c_0(\mathbf{X})$ $a.s.$ by noting that $\kappa_0(\boldsymbol{\beta})/\kappa_1(\boldsymbol{\beta})$ is a constant. Otherwise, it will lead to a contradiction since the right-hand side of the above equation is a non-degenerate measurable function of $\mathbf{X}$. With the help of $c_1(\mathbf{X}) = c_0(\mathbf{X})$ $a.s.$, (S7.1) is equivalent to

$$\begin{aligned}\frac{\kappa(\boldsymbol{\beta})}{\kappa_1(\boldsymbol{\beta})} =& \frac{1-(1-p_{11}(\mathbf{X}))\,c_1(\mathbf{X}) - p_{10}(\mathbf{X})\,c_1(\mathbf{X})}{(p_{11}(\mathbf{X}) - p_{10}(\mathbf{X}))\,c_1(\mathbf{X})} \\ =& \frac{1 - c_1(\mathbf{X})}{(p_{11}(\mathbf{X}) - p_{10}(\mathbf{X}))\,c_1(\mathbf{X})} + 1 \;\; a.s.,\end{aligned}$$

which implies $c_1(\mathbf{X}) = 1$ $a.s.$ by noting that $\kappa(\boldsymbol{\beta})/\kappa_1(\boldsymbol{\beta})$ is a constant. Otherwise, it will lead to a contradiction. Combing the previous results, we have $c_1(\mathbf{X}) = c_0(\mathbf{X}) = 1$ $a.s.$. We then conclude that $\mathbf{H}_w^{lte}(\boldsymbol{\beta}, \boldsymbol{u}) = 0$ $a.e.$ on Π is equivalent to $q(\mathbf{X};\boldsymbol{\beta}) = q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)$ $a.s.$. Given that $q(\mathbf{X};\boldsymbol{\beta}) = q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)$ $a.s.$ for a unique $\boldsymbol{\beta}_0^{lte}$, we must have $\boldsymbol{\beta} = \boldsymbol{\beta}_0^{lte}$. This concludes the proof. ■

Lemma S7.2 *(Lemma S5.2) Under Assumption 6 $(i)-(iii)$, the classes of functions*

$$\mathcal{F}_1^{lte} \equiv \{(d,z,\mathbf{x}) \in \{0,1\}\times\{0,1\}\times\mathcal{X} \mapsto d\left(\frac{z}{q(\mathbf{x};\boldsymbol{\beta})} - \frac{(1-z)}{1-q(\mathbf{x};\boldsymbol{\beta})}\right) : \boldsymbol{\beta}\in\Theta\},$$

$$\mathcal{F}_2^{lte} \equiv \{(d,z,\mathbf{x}) \in \{0,1\}\times\{0,1\}\times\mathcal{X} \mapsto \left(1 - \frac{(1-d)z}{q(\mathbf{x};\boldsymbol{\beta})} - \frac{d(1-z)}{1-q(\mathbf{x};\boldsymbol{\beta})}\right) : \boldsymbol{\beta}\in\Theta\},$$

$$\mathcal{F}_3^{lte} \equiv \mathcal{F}_1^{lte}\cdot\mathcal{W},$$

$$\mathcal{F}_4^{lte} \equiv \mathcal{F}_2^{lte}\cdot\mathcal{W},$$

where $\mathcal{W}$ is either equal to $\mathcal{W}_{ind}$, $\mathcal{W}_{proj}$ or $\mathcal{W}_{exp}$, are Glivenko-Cantelli.

Proof of Lemma S7.2 The Glivenko-Cantelli property of $\mathcal{F}_1^{lte}$ and $\mathcal{F}_2^{lte}$ follows from Example 19.8 in van der Vaart (1998) under Assumption $6(i)-(iii)$. ■

The family of functions associated with $\mathbf{h}_0^{lte}$,

$$\mathcal{F}_0^{lte} \equiv \{(d,z,\mathbf{x}) \in \{0,1\}\times\{0,1\}\times\mathcal{X} \mapsto (1-d)\left(z/q(\mathbf{x};\boldsymbol{\beta}) - (1-z)/(1-q(\mathbf{x};\boldsymbol{\beta}))\right) : \boldsymbol{\beta}\in\Theta\}.$$

is also Glivenko-Cantelli as a result of Lemma S7.2.

Let

$$\widehat{C}^{lte}_{ind,F_{n,\mathbf{X}}} = 2\int_{[-\infty,\infty]^k} \dot{\mathbf{H}}^{lte}_{n,ind}(\widehat{\boldsymbol{\beta}}^{lips}_{n,ind},\mathbf{u})'\dot{\mathbf{H}}^{lte}_{n,ind}(\widetilde{\boldsymbol{\beta}},\mathbf{u})F_{n,\mathbf{X}}(d\mathbf{u}),$$

$$\widehat{C}^{lte}_{proj,F_{n,\gamma}} = 2\int_{[-\infty,\infty]\times\mathbb{S}_k} \dot{\mathbf{H}}^{lte}_{n,proj}(\widehat{\boldsymbol{\beta}}^{lips}_{n,proj},\mathbf{u})'\dot{\mathbf{H}}^{lte}_{n,proj}(\widetilde{\boldsymbol{\beta}},\mathbf{u})F_{n,\boldsymbol{\gamma}}(du)d\boldsymbol{\gamma},$$

and

$$\widehat{C}^{lte}_{\exp,\Phi} = \int_{\mathbb{R}^k} \dot{\mathbf{H}}^{lte}_{n,\exp}(\widehat{\boldsymbol{\beta}}^{lips}_{n,\exp},\mathbf{u})^c\dot{\mathbf{H}}^{lte}_{n,\exp}(\widetilde{\boldsymbol{\beta}},\mathbf{u})\phi\left(\mathbf{u}\right)d\mathbf{u} \\ + \int_{\mathbb{R}^k} \dot{\mathbf{H}}^{lte}_{n,\exp}(\widehat{\boldsymbol{\beta}}^{lips}_{n,\exp},\mathbf{u})'\left(\dot{\mathbf{H}}^{lte}_{n,\exp}(\widetilde{\boldsymbol{\beta}},\mathbf{u})'\right)^c\phi\left(\mathbf{u}\right)d\mathbf{u},$$

where $\phi\left(\mathbf{u}\right)$ is the standard k-variate normal density function and $\widetilde{\boldsymbol{\beta}}$ satisfies $\left\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\| \leq \left\|\widehat{\boldsymbol{\beta}}^{lips}_{n,w}-\boldsymbol{\beta}^{lte}_0\right\|$. Furthermore, write

$$C^{lte}_{ind,F_{\mathbf{X}}} = 2\int_{[-\infty,\infty]^k} \dot{\mathbf{H}}^{lte}_{ind}(\boldsymbol{\beta}^{lte}_0,\mathbf{u})'\dot{\mathbf{H}}^{lte}_{ind}(\boldsymbol{\beta}^{lte}_0,\mathbf{u})F_{\mathbf{X}}(d\mathbf{u}),$$

$$C^{lte}_{proj,F_{\boldsymbol{\gamma}}} = 2\int_{[-\infty,\infty]\times\mathbb{S}_k} \dot{\mathbf{H}}^{lte}_{proj}(\boldsymbol{\beta}^{lte}_0,\mathbf{u})'\dot{\mathbf{H}}^{lte}_{proj}(\boldsymbol{\beta}^{lte}_0,\mathbf{u})F_{\boldsymbol{\gamma}}(du)d\boldsymbol{\gamma},$$

$$C^{lte}_{\exp,\Phi} = \int_{\mathbb{R}^k}\left(\dot{\mathbf{H}}^{lte}_{\exp}(\boldsymbol{\beta}^{lte}_0,\mathbf{u})^c\dot{\mathbf{H}}^{lte}_{exp}(\boldsymbol{\beta}^{lte}_0,\mathbf{u}) + \dot{\mathbf{H}}^{lte}_{\exp}(\boldsymbol{\beta}^{lte}_0,\mathbf{u})'\left(\dot{\mathbf{H}}^{lte}_{\exp}(\boldsymbol{\beta}^{lte}_0,\mathbf{u})'\right)^c\right)\phi\left(\mathbf{u}\right)d\mathbf{u}.$$

Lemma S7.3 *(Lemma S5.3) Let $\mathcal{W}$ be equal to either $\mathcal{W}_{ind}$, $\mathcal{W}_{proj}$ or $\mathcal{W}_{exp}$. Then, under Assumption 6,*

$$\mathcal{F}^{lte}_5 \equiv \left\{(d,z,\mathbf{x})\in\{0,1\}\times\{0,1\}\times\mathcal{X}\mapsto d\left(\frac{z}{q\left(\mathbf{x};\boldsymbol{\beta}\right)^2}+\frac{1-z}{(1-q\left(\mathbf{x};\boldsymbol{\beta}\right))^2}\right)\dot{q}\left(\mathbf{x};\boldsymbol{\beta}\right),\boldsymbol{\beta}\in\Theta^{lte}_0\right\},$$

$$\mathcal{F}^{lte}_6 \equiv \left\{(d,z,\mathbf{x})\in\{0,1\}\times\{0,1\}\times\mathcal{X}\mapsto \left(\frac{(1-d)z}{q\left(\mathbf{x};\boldsymbol{\beta}\right)^2}-\frac{d(1-z)}{(1-q\left(\mathbf{x};\boldsymbol{\beta}\right))^2}\right)\dot{q}\left(\mathbf{x};\boldsymbol{\beta}\right),\boldsymbol{\beta}\in\Theta^{lte}_0\right\},$$

$$\mathcal{F}^{lte}_7 \equiv \mathcal{F}^{lte}_5\cdot\mathcal{W},$$

$$\mathcal{F}^{lte}_8 \equiv \mathcal{F}^{lte}_6\cdot\mathcal{W},$$

are Glivenko-Cantelli classes of functions. Furthermore,

$$\widehat{C}^{lte}_{ind,F_{n,\mathbf{X}}} - C^{lte}_{ind,F_{\mathbf{X}}} = o_p\left(1\right),$$

$$\widehat{C}^{lte}_{proj,F_{n,\boldsymbol{\gamma}}} - C^{lte}_{proj,F_{\boldsymbol{\gamma}}} = o_p\left(1\right),$$

$$\widehat{C}^{lte}_{\exp,\Phi} - C^{lte}_{\exp,\Phi} = o_p\left(1\right).$$

Proof of Lemma S7.3: The follows from the same steps as in the proof of Lemma S5.3 and is therefore omitted. ■

As in Lemma S7.2, it follows trivially from Lemma S7.3 that

$$\mathcal{F}_9^{lte} \equiv \left\{(d,z,\mathbf{x}) \in \{0,1\} \times \{0,1\} \times \mathcal{X} \mapsto (1-d) \cdot \left(\frac{z}{q\left(\mathbf{x};\boldsymbol{\beta}\right)^2} + \frac{1-z}{(1-q\left(\mathbf{x};\boldsymbol{\beta}\right))^2}\right) \dot{q}\left(\mathbf{x};\boldsymbol{\beta}\right), \boldsymbol{\beta} \in \Theta_0^{lte}\right\},$$

is Glivenko-Cantelli.

Lemma S7.4 *(Lemma S5.4) Let* Π *be a compact, convex subset of* $\mathbb{R}^k$ *with a non-empty interior. Then, under Assumption 6,*

$$\mathcal{F}_{ind}^{lte} \equiv \left\{(d,z,\mathbf{x}) \in \{0,1\} \times \{0,1\} \times \mathcal{X} \mapsto \mathbf{h}\left(d,z,\mathbf{x};\boldsymbol{\beta}_0^{lte}\right) 1(\mathbf{x} \leq \mathbf{u}) : \mathbf{u} \in [-\infty,\infty]^k\right\},$$

$$\mathcal{F}_{proj}^{lte} \equiv \left\{(d,z,\mathbf{x}) \in \{0,1\} \times \{0,1\} \times \mathcal{X} \mapsto \mathbf{h}\left(d,z,\mathbf{x};\boldsymbol{\beta}_0^{lte}\right) 1\left\{\boldsymbol{\gamma}'\mathbf{x} \leq u\right\} : (\boldsymbol{\gamma},u) \in \mathbb{S}_k \times [-\infty,\infty]\right\},$$

$$\mathcal{F}_{\exp}^{lte} \equiv \left\{(d,z,\mathbf{x}) \in \{0,1\} \times \{0,1\} \times \mathcal{X} \mapsto \mathbf{h}\left(d,z,\mathbf{x};\boldsymbol{\beta}_0^{lte}\right) \exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right)) : \mathbf{u} \in \Pi\right\},$$

are Donsker classes of functions.

Proof of Lemma S7.4: The follows from the same steps as in the proof of Lemma S5.4 and is therefore omitted. ■

Next, define

$$A_{2,ind}^{lte}\left(\mathbf{x}\right) = 2 \cdot \int_{[-\infty,\infty]^k} \left(\dot{\mathbf{H}}_{ind}^{lte}(\boldsymbol{\beta}_0^{lte},\mathbf{u})'1(\mathbf{x} \leq \mathbf{u})\right) F_{\mathbf{X}}(d\mathbf{u}),$$

$$A_{2,proj}^{lte}\left(\mathbf{x}\right) = 2 \cdot \int_{[-\infty,\infty]\times\mathbb{S}_k} \dot{\mathbf{H}}_{proj}^{lte}(\boldsymbol{\beta}_0^{lte},(u,\boldsymbol{\gamma}))'1\left\{\boldsymbol{\gamma}'\mathbf{x} \leq u\right\} F_{\boldsymbol{\gamma}}(du)d\boldsymbol{\gamma},$$

$$A_{2,\exp}^{lte}\left(\mathbf{x}\right) = \int_{\mathbb{R}^k} \left(\dot{\mathbf{H}}_{\exp}^{lte}(\boldsymbol{\beta}_0^{lte},\mathbf{u})^c \exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right)) + \dot{\mathbf{H}}_{\exp}^{lte}(\boldsymbol{\beta}_0^{lte},\mathbf{u})' \exp(-i\mathbf{u}'\Phi\left(\mathbf{x}\right))\right) \phi\left(\mathbf{u}\right) d\mathbf{u},$$

and let $A_{n,2,ind}^{lte}\left(\mathbf{x}\right), A_{n,2,proj}^{lte}\left(\mathbf{x}\right)$ and $A_{n,2,exp}^{lte}\left(\mathbf{x}\right)$ be the counterparts in the sample.

Lemma S7.5 *(Lemma S5.5) Under Assumption 6,*

$$\mathbb{E}_n\left[A_{n,2,ind}^{lte}\left(\mathbf{X}\right) \cdot \mathbf{h}_n\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right] = \mathbb{E}_n\left[A_{2,ind}^{lte}\left(\mathbf{X}\right) \cdot \mathbf{h}\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right] + o_p\left(n^{-1/2}\right), \quad \text{(S7.3)}$$

$$\mathbb{E}_n\left[A_{n,2,proj}^{lte}\left(\mathbf{X}\right) \cdot \mathbf{h}_n\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right] = \mathbb{E}_n\left[A_{2,proj}^{lte}\left(\mathbf{X}\right) \cdot \mathbf{h}\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right] + o_p\left(n^{-1/2}\right), \quad \text{(S7.4)}$$

$$\mathbb{E}_n\left[A_{n,2,\exp}^{lte}\left(\mathbf{X}\right) \cdot \mathbf{h}_n\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right] = \mathbb{E}_n\left[A_{2,\exp}^{lte}\left(\mathbf{X}\right) \cdot \mathbf{h}\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right] + o_p\left(n^{-1/2}\right). \quad \text{(S7.5)}$$

Proof of Lemma S7.5: The follows from the same steps as in the proof of Lemma S5.5 and is therefore omitted. ■

S7.II Proof of Main Results

Proof of Theorem 4.1: We first show consistency of $\widehat{\boldsymbol{\beta}}^{lips}_{n,w}$. From Lemma S7.1 we know that $Q^{lte}_w(\boldsymbol{\beta})$ is uniquely minimized at $\boldsymbol{\beta}^{lte}_0$, and that, under Assumption 6, $\mathbf{H}^{lte}_w(\boldsymbol{\beta},\mathbf{u})$ is continuous at each $\boldsymbol{\beta}\in\Theta$, Θ is compact, we have that by Exercise 5.27 in van der Vaart (1998) for every $\varepsilon>0$

$$\inf_{\boldsymbol{\beta}:\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^{lte}_0\right\|\geq\varepsilon} Q^{lte}_w(\boldsymbol{\beta})>Q^{lte}_w(\boldsymbol{\beta}^{lte}_0).$$

Therefore, the consistency of $\widehat{\boldsymbol{\beta}}^{lips}_{n,w}$ follows immediately from the uniform convergence of $Q^{lte}_{n,w}(\boldsymbol{\beta})$ over Θ as $n\to\infty$. Lemma S5.1, S7.2 and CMT ensure that

$$\sup_{(\boldsymbol{\beta},\mathbf{u})\in\Theta\times\Pi}\left\|\mathbf{H}^{lte}_{n,w}(\boldsymbol{\beta},\mathbf{u})-\mathbf{H}^{lte}_w(\boldsymbol{\beta},\mathbf{u})\right\|\xrightarrow{p}0,\ w\in\{ind,proj\}\,.$$

and for $w=exp$, the same arguments as in the proof of Theorem 4.1 can be applied to show the uniform convergence.

To derive the asymptotic linear representation of $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{lips}_{n,w}-\boldsymbol{\beta}^{lte}_0\right)$, we first apply Taylor expansion to the first order condition of $Q^{lte}_{n,w}(\boldsymbol{\beta})$, which gives

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{lips}_{n,w}-\boldsymbol{\beta}^{lte}_0\right)\\
&=-(\widehat{C}^{lte}_{w,\Psi_n})^{-1}\cdot\sqrt{n}\int\left(\dot{\mathbf{H}}^{lte}_{n,w}(\widehat{\boldsymbol{\beta}}^{lips}_{n,w},\mathbf{u})^c\mathbf{H}^{lte}_{n,w}(\boldsymbol{\beta}^{lte}_0,\mathbf{u})+\dot{\mathbf{H}}^{lte}_{n,w}(\widehat{\boldsymbol{\beta}}^{lips}_{n,w},\mathbf{u})'\left(\mathbf{H}^{lte}_{n,w}(\boldsymbol{\beta}^{lte}_0,\mathbf{u})'\right)^c\right)\Psi_n(d\mathbf{u})\\
&\quad=-(\widehat{C}^{lte}_{w,\Psi_n})^{-1}\cdot\sqrt{n}\mathbb{E}_n\left[\int\left(\dot{\mathbf{H}}^{lte}_{n,w}(\widehat{\boldsymbol{\beta}}^{lips}_{n,w},\mathbf{u})^c w(\mathbf{X};\mathbf{u})+\dot{\mathbf{H}}^{lte}_{n,w}(\widehat{\boldsymbol{\beta}}^{lips}_{n,w},\mathbf{u})'w^c(\mathbf{X},\mathbf{u})\right)\Psi_n(d\mathbf{u})\right.\\
&\qquad\left.\cdot\mathbf{h}_n\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)\right]
\end{aligned}\tag{S7.6}$$

From Lemma S7.3 and Lemma S7.5 we have that

$$\widehat{C}^{lte}_{w,\Psi_n}=C^{lte}_{w,\Psi}+o_p(1)\,,\tag{S7.7}$$

$$\mathbb{E}_n\left[A^{lte}_{n,2,ind}(\mathbf{X})\cdot\mathbf{h}_n\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)\right]=\mathbb{E}_n\left[A^{lte}_{2,ind}(\mathbf{X})\cdot\mathbf{h}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)\right]+o_p(1)\,.\tag{S7.8}$$

Consistency of $\widehat{\boldsymbol{\beta}}^{lips}_{n,w}$ follows from (S7.7) and (S7.8) with $l^{lte}_{w,\Psi}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)$ given by (4.7). Asymptotic normality results from the square integrability of $l^{lte}_{w,\Psi}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)$, which is further guaranteed by the uniform boundedness of $w(\mathbf{X};\mathbf{u})$ and Assumption 6(ii). ■

Before presenting the proof of Theorem 4.2, we define some quantities related to the influence function of $\psi^j_{w,\Psi}$, for $w\in\{ind,exp,proj\}$. Let

$$\psi^{late}_{w,\Psi}(Y,D,Z,\mathbf{X})=g^{late}(Y,D,Z,\mathbf{X})-l^{lte}_{w,\Psi}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)'\cdot\mathbf{G}^{late}_{\boldsymbol{\beta}},\tag{S7.9}$$

$$\psi_{w,\Psi}^{ldte}\left(Y,D,Z,\mathbf{X};y\right)=g^{ldte}\left(Y,D,Z,\mathbf{X};y\right)-l_{w,\Psi}^{lte}\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)'\cdot\mathbf{G}_{\boldsymbol{\beta}}^{ldte}\left(y\right),\tag{S7.10}$$

$$\psi_{w,\Psi}^{lqte}\left(Y,D,Z,\mathbf{X};\tau\right)=-\left(g^{lqte}\left(Y,D,Z,\mathbf{X};\tau\right)-l_{w,\Psi}^{lte}\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)'\cdot\mathbf{G}_{\boldsymbol{\beta}^{lte}}^{lqte}\left(\tau\right)\right),\tag{S7.11}$$

where, for $j\in\{late,ldte,lqte\}$, $g^j\left(Y,D,Z,\mathbf{X}\right)=g_1^j\left(Y,D,Z,\mathbf{X}\right)-g_0^j\left(Y,D,Z,\mathbf{X}\right)$, with

$$g_d^{late}\left(Y,D,Z,\mathbf{X}\right)=\varpi_d^{lte}\left(D,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\cdot\left(Y-\mathbb{E}\left[Y\left(d\right)|\mathcal{C}\right]\right),$$

$$g_d^{ldte}\left(Y,D,Z,\mathbf{X};y\right)=\varpi_d^{lte}\left(D,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\cdot\left(1\left\{Y\leq y\right\}-F_{Y(d)|\mathcal{C}}\left(y\right)\right),$$

$$g_d^{lqte}\left(Y,D,Z,\mathbf{X};\tau\right)=\frac{\varpi_d^{lte}\left(D,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\cdot\left(1\left\{Y\leq q_{Y(d)|\mathcal{C}}\left(\tau\right)\right\}-\tau\right)}{f_{Y(d)|\mathcal{C}}\left(q_{Y(d)|\mathcal{C}}\left(\tau\right)\right)},$$

and

$$\mathbf{G}_{\boldsymbol{\beta}}^{late}=\mathbb{E}\left[\left(\sum_{d=0,1}\frac{1\{D=d\}(Y-\mathbb{E}[Y(d)|\mathcal{C}])}{(-1)^{d+1}\kappa_d(\boldsymbol{\beta}_0^{lte})}\right)\cdot\left(\frac{Z}{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)^2}+\frac{1-Z}{\left(1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right)^2}\right)\dot{q}\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right],$$

$$\mathbf{G}_{\boldsymbol{\beta}}^{ldte}\left(y\right)=\mathbb{E}\left[\left(\sum_{d=0,1}\frac{1\{D=d\}(1\{Y\leq y\}-F_{Y(d)|\mathcal{C}}(y))}{(-1)^{d+1}\kappa_d(\boldsymbol{\beta}_0^{lte})}\right)\cdot\left(\frac{Z}{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)^2}+\frac{1-Z}{(1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right))^2}\right)\dot{q}\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right],$$

$$\mathbf{G}_{\boldsymbol{\beta}}^{lqte}\left(\tau\right)=\mathbb{E}\left[\left(\sum_{d=0,1}\frac{1\{D=d\}(1\{Y\leq q_{Y(d)|\mathcal{C}}(\tau)\}-\tau)}{(-1)^{d+1}\kappa_d(\boldsymbol{\beta}_0^{lte})}\right)\cdot\left(\frac{Z}{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)^2}+\frac{1-Z}{(1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right))^2}\right)\dot{q}\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right].$$

The functions g^{late}, g^{ldte} and g^{lqte} are the influence functions of the LATE, LDTE and LQTE estimators, respectively, when the instrumental propensity score parameters $\boldsymbol{\beta}_0^{lte}$ are known. We denote $\Omega_{w,\Psi}^{late}=\mathbb{E}\left[\psi_{w,\Psi}^{late}\left(Y,D,Z,\mathbf{X}\right)^2\right]$, $\Omega_{w,\Psi,y}^{ldte}=\mathbb{E}\left[\psi_{w,\Psi}^{ldte}\left(Y,D,Z,\mathbf{X};y\right)^2\right]$, and $\Omega_{w,\Psi,\tau}^{lqte}=\mathbb{E}\left[\psi_{w,\Psi}^{lqte}\left(Y,D,Z,\mathbf{X};\tau\right)^2\right]$.

Part 1: Asymptotic Properties of the Local Average Treatment Effect.

As in the proof of Theorem 3.2, we show that

$$\sqrt{n}\left(\widehat{LATE}_n^{lips}-LATE\right)=\sqrt{n}\mathbb{E}_n\left[\psi_{w,\Psi}^{late}\left(Y,D,Z,\mathbf{X}\right)\right]+o_p\left(1\right)\tag{S7.12}$$

where $\mathbb{E}\left[\psi_{w,\Psi}^{late}\left(Y,D,Z,\mathbf{X}\right)\right]=0$ and $\mathbb{E}\left[\psi_{w,\Psi}^{late}\left(Y,D,Z,\mathbf{X}\right)^2\right]<\infty$.

The key is to show that

$$\begin{aligned}&\mathbb{E}_n[\varpi_{n,d}^{lte}(D,Z,\mathbf{X};\widehat{\boldsymbol{\beta}}_{n,w}^{lips})Y]-\mathbb{E}\left[Y\left(d\right)|\mathcal{C}\right]\\&\quad=\mathbb{E}_n\left[\varpi_d^{lte}\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\cdot\left(Y-\mathbb{E}\left[Y\left(d\right)|\mathcal{C}\right]\right)-l_{w,\Psi}^{lte}\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)'\mathbf{G}_{\boldsymbol{\beta},d}^{late}\right]+o_p\left(n^{-1/2}\right),\end{aligned}\tag{S7.13}$$

where $\mathbf{G}^{late}_{\boldsymbol{\beta},d} = \mathbb{E}\left[\frac{1\{D=d\}(Y-\mathbb{E}[Y(d)|\mathcal{C}])}{\kappa_d(\boldsymbol{\beta}^{lte}_0)} \cdot \left(\frac{Z}{q(\mathbf{X};\boldsymbol{\beta}^{lte}_0)^2} + \frac{1-Z}{(1-q(\mathbf{X};\boldsymbol{\beta}^{lte}_0))^2}\right) \dot{q}\left(\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)\right]$

Without loss of generality, we focus on $d=1$. By Taylor expanding $\mathbb{E}_n[\varpi^{lte}_{n,d}(D,Z,\mathbf{X};\widehat{\boldsymbol{\beta}}^{lips}_{n,w})Y]$ around $\boldsymbol{\beta}^{lte}_0$ and some algebraic manipulations, we have that

$$
\begin{aligned}
&\mathbb{E}_n[\varpi^{lte}_{n,1}(D,Z,\mathbf{X};\widehat{\boldsymbol{\beta}}^{lips}_{n,w})Y] \\
&= \mathbb{E}_n\left[\varpi^{lte}_{n,1}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)Y\right] \\
&- \mathbb{E}_n\left[\left(\frac{Z}{q\left(\mathbf{X};\widetilde{\boldsymbol{\beta}}\right)^2} + \frac{1-Z}{\left(1-q\left(\mathbf{X};\widetilde{\boldsymbol{\beta}}\right)\right)^2}\right)\left(Y - \mathbb{E}_n\left[\varpi^{lte}_{n,1}\left(D,Z,\mathbf{X};\widetilde{\boldsymbol{\beta}}\right)Y\right]\right)\cdot \dot{q}\left(\mathbf{X};\widetilde{\boldsymbol{\beta}}\right)'\right]\left(\widehat{\boldsymbol{\beta}}^{lips}_{n,w} - \boldsymbol{\beta}^{lte}_0\right),
\end{aligned}
$$

where $\widetilde{\boldsymbol{\beta}}$ satisfies $\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^{lte}_0\right\| \le \left\|\widehat{\boldsymbol{\beta}}^{lips}_{n,w} - \boldsymbol{\beta}^{lte}_0\right\|$. From Theorem 4.1, we have that

$$
\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{lips}_{n,w} - \boldsymbol{\beta}_0\right) = \sqrt{n}\mathbb{E}_n\left[l^{lte}_{w,\Psi}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)\right] + o_p(1)
$$

and therefore by CMT,

$$
\begin{aligned}
&\mathbb{E}_n[\varpi^{lte}_{n,1}(D,Z,\mathbf{X};\widehat{\boldsymbol{\beta}}^{lips}_{n,w})Y] \\
&\quad = \mathbb{E}_n\left[\varpi^{lte}_{n,1}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)Y\right] - \mathbb{E}_n\left[l^{lte}_{w,\Psi}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)'\cdot \mathbf{G}^{late}_{\boldsymbol{\beta},1}\right] + o_p\left(n^{-1/2}\right). \quad \text{(S7.14)}
\end{aligned}
$$

To get the influence function of the first term, we use the fact that $\varpi^{lte}_{n,1}$ is normalized with mean equal to 1 and that $\varpi^{lte}_{n,1} \xrightarrow{a.s.} \varpi^{lte}_1$ by Lemma S7.2,

$$
\begin{aligned}
\mathbb{E}_n\left[\varpi^{lte}_{n,1}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)Y\right] - \mathbb{E}[Y(1)|\mathcal{C}] &= \mathbb{E}_n\left[\varpi^{lte}_{n,1}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)(Y-\mathbb{E}[Y(1)|\mathcal{C}])\right] \quad \text{(S7.15)} \\
&= \mathbb{E}_n\left[\varpi^{lte}_{1}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)(Y-\mathbb{E}[Y(1)|\mathcal{C}])\right] + o_p(n^{-1/2}) \quad \text{(S7.16)}
\end{aligned}
$$

We can also show $\mathbb{E}[\varpi^{lte}_{n,0}Y]$ admits an asymptotic linear representation and (S7.12) follows by CMT and the orthogonality between $(g^{late}_1, \mathbf{G}^{late}_{\boldsymbol{\beta},1})$ and $(g^{late}_0, \mathbf{G}^{late}_{\boldsymbol{\beta},0})$.

Lastly, we need to show $\psi^{late}_{w,\Psi}$ is square integrable. By Assumption 6 (uniform boundedness of ϖ^{lte}_d), Assumption 7 (conditional square integrability of $Y(d)$) and Theorem 4.1 (square integrability of $l^{lte}_{w,\Psi}$), a standard application of Cauchy-Schwartz inequality would lead to the desired result. ■

Part 2: Asymptotic Properties of the Local Distribution Treatment Effects.

The (uniform) asymptotic linear representation for the Local Distribution Treatment Effect parameter without rearrangement can be derived as in Part 1.

We define the rearranging operator of CDF as $\mathcal{R}_{\mathcal{F}d}: F_{Y(d)|\mathcal{C}}(\cdot) \mapsto F^r_{Y(d)|\mathcal{C}}(\cdot) \equiv \int 1\{F^{-1}_{Y(d)|\mathcal{C}}(\tau) \le \cdot\}d\tau$, for $d=0,1$. Then, by the monotonicity of $F_{Y(d)|\mathcal{C}}(\cdot)$ and Corollary 3 of Chernozhukov et al.

(2010),

$$\sqrt{n}(\widehat{F^r}_{n,\varpi_d^{lte}\cdot Y}(\cdot) - F_{Y(d)|\mathcal{C}}(\cdot)) = \sqrt{n}(\widehat{F}_{n,\varpi_d^{lte}\cdot Y}(\cdot) - F_{Y(d)|\mathcal{C}}(\cdot)) + o_p(1), \tag{S7.17}$$

uniformly over $[q_{Y(d)|\mathcal{C}}(a_1) - \epsilon, q_{Y(d)|\mathcal{C}}(a_2) + \epsilon]$, for $d = 0, 1$.

Therefore, as in the proof of Theorem 3.2 we only need to show that the classes of functions

$$\mathcal{F}_{1,ldte} \equiv \left\{(v, d, z, \mathbf{x}) \in \{0,1\} \times \{0,1\} \times \mathcal{X} \mapsto \psi_{1,w,\Psi}^{ldte}(v, d, z, \mathbf{x};y) : y \in [-\infty, \infty]\right\},$$

$$\mathcal{F}_{0,ldte} \equiv \left\{(v, d, z, \mathbf{x}) \in \{0,1\} \times \{0,1\} \times \mathcal{X} \mapsto \psi_{0,w,\Psi}^{ldte}(v, d, z, \mathbf{x};y) : y \in [-\infty, \infty]\right\},$$

are Donsker, where

$$\psi_{1,w,\Psi}^{ldte}(v, d, z, \mathbf{x};y) = g_1^{ldte}(v, d, z, \mathbf{x}; y) - l_{w,\Psi}^{lte}(d, z, \mathbf{x}; \boldsymbol{\beta}_0)' \cdot \mathbf{G}_{1,\boldsymbol{\beta}}^{ldte}(y),$$

$$\psi_{0,w,\Psi}^{ldte}(v, d, z, \mathbf{x};y) = g_0^{ldte}(v, d, z, \mathbf{x}; y) - l_{w,\Psi}^{lte}(d, z, \mathbf{x}; \boldsymbol{\beta}_0)' \cdot \mathbf{G}_{0,\boldsymbol{\beta}}^{ldte}(y),$$

and

$$\mathbf{G}_{\boldsymbol{\beta},d}^{ldte} = \mathbb{E}\left[\frac{1\{D = d\}(1\{Y \leq y\} - F_{Y(d)|\mathcal{C}}(y))}{\kappa_d(\boldsymbol{\beta}_0^{lte})} \cdot \left(\frac{Z}{q\left(\mathbf{X}; \boldsymbol{\beta}_0^{lte}\right)^2} + \frac{1-Z}{(1 - q\left(\mathbf{X}; \boldsymbol{\beta}_0^{lte}\right))^2}\right) \dot{q}\left(\mathbf{X}; \boldsymbol{\beta}_0^{lte}\right)\right].$$

First note that $\left\{l_{w,\Psi}^{lte}\left(d, z, \mathbf{x}; \boldsymbol{\beta}_0^{lte}\right)' \cdot \mathbf{G}_{d,\boldsymbol{\beta}}^{ldte}(y) : y \in [-\infty, \infty]\right\}$ is Donsker since they are deterministic functions of y, $\mathbf{G}_{d,\boldsymbol{\beta}}^{ldte}(y) < \infty$, $d \in \{0,1\}$, and, by Theorem 4.1, the square integrability of $l_{w,\Psi}^{lte}$. The Donsker property of $\left\{g_d^{ldte}(v, d, z, \mathbf{x}; y) : y \in [-\infty, \infty]\right\}$ follows from Lemma S5.1, Assumption 6, and Corollary 9.32 in Kosorok (2008). Thus, from Corollary 9.32 in Kosorok (2008), we conclude that $\mathcal{F}_{1,ldte}$ and $\mathcal{F}_{0,ldte}$ are Donsker.

Hence, under Assumptions 3, 5-7 ,

$$\mathbb{G}_{n,w,\Psi}^{ldte,(1,0)}(\cdot) \equiv \sqrt{n}\mathbb{E}_n\left[\left(\psi_{1,w,\Psi}^{ldte}(v, d, z, \mathbf{x};\cdot), \psi_{0,w,\Psi}^{ldte}(v, d, z, \mathbf{x};\cdot)\right)'\right] \tag{S7.18}$$

$$\Rightarrow \mathbb{G}_{\infty,w,\Psi}^{ldte,(1,0)}(\cdot) \text{ in } \ell^\infty([-\infty, \infty]) \times \ell^\infty([-\infty, \infty]), \tag{S7.19}$$

where $\mathbb{G}_{\infty,w,\Psi}^{ldte,(1,0)}(\cdot)$ is a tight, two-dimensional mean zero Gaussian process with covariance kernel $\Gamma(y_1, y_2) = \mathbb{E}[(\psi_{1,w,\Psi}^{ldte}(y_1), \psi_{0,w,\Psi}^{ldte}(y_1))(\psi_{1,w,\Psi}^{ldte}(y_2), \psi_{0,w,\Psi}^{ldte}(y_2))']$, Applying CMT, it follows that

$$\sqrt{n}\left(\widehat{LDTE}_n^{lips} - LDTE\right)(\cdot) = (1, -1)\,\mathbb{G}_{n,w,\Psi}^{ldte,(1,0)}(\cdot) + o_p(1),$$

$$\Rightarrow \mathbb{G}_{\infty,w,\Psi}^{ldte}(\cdot) \text{ in } \ell^\infty([-\infty, \infty])$$

where $\mathbb{G}_{\infty,w,\Psi}^{ldte}(\cdot)$ is a tight, univariate mean zero Gaussian process with covariance kernel

$$\Gamma_{ldte}(y_1, y_2) = \mathbb{E}\left[\psi_{w,\Psi}^{ldte}(Y, D, Z, \mathbf{X}; y_1)\,\psi_{w,\Psi}^{ldte}(Y, D, Z, \mathbf{X}; y_2)\right].$$

Fixing y leads to asymptotic normality result of $\widehat{LDTE}^{lips}$ in Theorem 4.2.

Part 3: Asymptotic Properties of the Local Quantile Treatment Effects.

Likewise, by Corollary 3 of Chernozhukov et al. (2010) and the monotonicity of $F^{-1}_{Y(d)|\mathcal{C}}(\cdot)$, the rearranged quantile estimator $\widehat{Fr}^{-1}_{n,\varpi^{lte}_d \cdot Y}(\cdot)$ have the same first order asymptotic distribution as $\widehat{F}^{-1}_{n,\varpi^{lte}_d \cdot Y}(\cdot)$, for $d = 0, 1$. Therefore, we can focus on deriving the asymptotic property of the original quantile estimator.

Under Assumptions 3, 5-7, we can use Lemma 21.4 in van der Vaart (1998), and the functional delta method to show that (S7.19) leads to

$$\sqrt{n}\left(\widehat{\mathbf{q}}^{lips}_n - \mathbf{q}^{lte}\right)(\cdot) = -\mathbf{f}^{-1}_{lte}(\cdot)' \cdot \mathbb{G}^{ldte,(1,0)}_{n,w,\Psi}\left(\mathbf{q}^{lte}(\cdot)\right) + o_p(1)$$

$$\Rightarrow -\mathbf{f}^{-1}_{lte}(\cdot)' \cdot \mathbb{G}^{ldte,(1,0)}_{\infty,w,\Psi}\left(\mathbf{q}^{lte}(\cdot)\right) \text{ in } \ell^{\infty}([a_1, a_2]) \times \ell^{\infty}([a_1, a_2]),$$

where $\widehat{\mathbf{q}}^{lips}_n(\boldsymbol{\tau}) = \left(\widehat{q}^{lips}_{n,Y(1)|\mathcal{C}}(\tau_1), \widehat{q}^{lte}_{n,Y(0)|\mathcal{C}}(\tau_2)\right)'$, $\mathbf{q}^{lte}(\boldsymbol{\tau}) = \left(q_{Y(1)|\mathcal{C}}(\tau_1), q_{Y(0)|\mathcal{C}}(\tau_2)\right)'$, $\mathbf{f}^{-1}_{lte}(\tau) = \left(f^{-1}_{Y(1)|\mathcal{C}}\left(q_{Y(1)|\mathcal{C}}(\tau_1)\right), f^{-1}_{Y(0)|\mathcal{C}}\left(q_{Y(0)|\mathcal{C}}(\tau_2)\right)\right)'$, $\boldsymbol{\tau} = (\tau_1, \tau_2) \in [a_1, a_2]^2$, and for all $0 < a_1 < a_2 < 1$.

Applying CMT again yields

$$\sqrt{n}\left(\widehat{LQTE}^{lips}_n - LQTE\right)(\cdot) = (1, -1) \cdot \left(-\mathbf{f}^{-1}_{lte}(\cdot)' \cdot \mathbb{G}^{ldte,(1,0)}_{n,w,\Psi}\left(\mathbf{q}^{lte}(\cdot)\right)\right) + o_p(1)$$

$$\Rightarrow \mathbb{G}^{lqte}_{\infty,w,\Psi}(\cdot) \text{ in } \ell^{\infty}[a_1, a_2],$$

where $\mathbb{G}^{lqte}_{\infty,w,\Psi}(\cdot)$ is a tight, mean zero Gaussian process with covariance kernel

$$\Gamma_{lqte}(\tau_1, \tau_2) = \mathbb{E}\left[\psi^{lqte}_{w,\Psi}(Y, D, Z, \mathbf{X}; \tau_1)\, \psi^{lqte}_{w,\Psi}(Y, D, Z, \mathbf{X}; \tau_2)\right].$$

Now, fixing τ leads to asymptotic normality result of $\widehat{LQTE}^{lips}$ in Theorem 4.2. ■

S8 Estimating Treatment Effects on the Treated

In this section we focus on treatment effect parameters for the treated subpopulation instead of the overall population. Heckman et al. (1997) argue that analyzing treatment effects on the treated instead of overall treatment effects is more interesting when the policy intervention is directed at individuals with certain characteristics. For instance, if an employment program (or a clinical treatment) is directed at individuals who face barriers to employment (or who have some specific symptoms), perhaps there is little interest in analyzing the effect of this intervention on individuals with strong labor market attachment (or on individual who does not have these symptoms). Another potential advantage of focusing on the treated subpopulation is that one can weaken the overlap condition in Assumption 2(ii) by allowing the PS to be close or even exactly equal to zero. This is particularly important in one of our applications in Section 6.

Analogous to the discussion in the previous section, here the goal is to make inference about the average, distributional and quantile treatment effect on the treated, defined as $ATT = \mathbb{E}[Y(1)|D=1] - \mathbb{E}[Y(0)|D=1]$, $DTT(y) = F_{Y(1)|D=1}(y) - F_{Y(0)|D=1}(y)$, and $QTT(\tau) = q_{Y(1)|D=1}(\tau) - q_{Y(0)|D=1}(\tau)$, respectively, where, for $d \in \{0,1\}$, $F_{Y(d)|D=1}(y) = \mathbb{E}[1\{Y(d) \le y\}|D=1]$, $y \in \mathbb{R}$, and $q_{Y(d)|D=1}(\tau) = \inf\left\{y : F_{Y(d)|D=1}(y) \ge \tau\right\}$, $\tau \in (0,1)$.

Let $w_1^{tt,ps}(D,\mathbf{X}) = D/\,\mathbb{E}[D]$ and

$$w_0^{tt,ps}(D,\mathbf{X}) = \frac{(1-D)\,p(\mathbf{X})}{1-p(\mathbf{X})} \Bigg/ \mathbb{E}\left[\frac{(1-D)\,p(\mathbf{X})}{1-p(\mathbf{X})}\right].$$

Note that functionals of the distribution of $Y(1)$ for the treated subpopulation can be directly estimated from the data using the analogy principle. Thus, when analyzing treatment effects on the treated, the main challenge faced is to identify and make inference about functionals of the distribution of $Y(0)$ for the treated subpopulation. Towards this end, we make the following assumptions.

Assumption S8.1 *(a) Given $\mathbf{X}$, $Y(0)$ is independent from D; and (b) for all $\mathbf{x} \in \mathcal{X}$, $p(\mathbf{x})$ is uniformly bounded away from one.*

Assumption S8.2 *For $d \in \{0,1\}$, (i) $\mathbb{E}\left[Y(d)^2|D=1\right] < M$ for some $0 < M < \infty$, (ii)*

$$\mathbb{E}\left[\sup_{\beta \in \Theta_0}\left\|\frac{w_0^{tt,ps}(D,\mathbf{X};\boldsymbol{\beta})\,(Y - \mathbb{E}[Y(0)|D=1])}{p(\mathbf{X};\boldsymbol{\beta})\,(1-p(\mathbf{X};\boldsymbol{\beta}))}\cdot \dot{p}(\mathbf{X};\boldsymbol{\beta})\right\|\right] < \infty,$$

and (iii) for some $\varepsilon > 0$, $0 < a_1 < a_2 < 1$, $F_{Y(d)|D=1}$ is continuously differentiable on $\left[q_{Y(d)|D=1}(a_1) - \varepsilon, q_{Y(d)|D=1}(a_2) + \varepsilon\right]$ with strictly positive derivative $f_{Y(d)|D=1}$.

Assumption S8.1 is a weaker version of Assumption 1, where we do not impose any lower bound on the PS, nor make any assumption about the relationship of $Y(1)$, D, and $\mathbf{X}$. Assumption S8.2 is the analogue of Assumption 4.

As shown by Heckman et al. (1997), under Assumptions S8.1 - S8.2, the ATT is identified by

$$ATT = \mathbb{E}\left[\left(w_1^{tt,ps}(D,\mathbf{X}) - w_0^{tt,ps}(D,\mathbf{X})\right)Y\right].$$

Analogously, $F_{Y(0)|D=1}(y)$ is identified by

$$F_{Y(0)|D=1}(y) = \mathbb{E}\left[w_0^{tt,ps}(D,\mathbf{X})\,1\{Y \le y\}\right],$$

implying that both $DTT(y)$ and $QTT(\tau)$ can also be written as functionals of the observed data; see e.g. Firpo (2007) . Such identification results suggest that we can estimate the ATT, $DTT(y)$ and $QTT(\tau)$ by

$$\widehat{ATT}_n^{ips} = \mathbb{E}_n\left[\left(w_{n,1}^{tt,ps}(D,\mathbf{X}) - w_{n,0}^{tt,ps}\left(D,\mathbf{X};\widehat{\boldsymbol{\beta}}_{n,w}^{ips}\right)\right)Y\right],$$

$$\widehat{DTT}_n^{ips}(y) = \mathbb{E}_n\left[\left(w_{n,1}^{tt,ps}(D,\mathbf{X}) - w_{n,0}^{tt,ps}\left(D,\mathbf{X};\widehat{\boldsymbol{\beta}}_{n,w}^{ips}\right)\right)1\{Y \le y\}\right],$$

$$\widehat{QTT}_n^{ips}(\tau) = \widehat{q}_{n,Y(1)|D=1}(\tau) - \widehat{q}_{n,Y(0)|D=1}^{ips}(\tau),$$

where

$$\widehat{q}_{n,Y(1)|D=1} = \arg\min_{q\in\mathbb{R}} \mathbb{E}_n\left[w_{n,1}^{tt,ps}(D,\mathbf{X}) \cdot \rho_\tau(Y-q)\right],$$

$$\widehat{q}_{n,Y(0)|D=1}^{ips} = \arg\min_{q\in\mathbb{R}} \mathbb{E}_n\left[w_{n,0}^{tt,ps}\left(D,\mathbf{X};\widehat{\boldsymbol{\beta}}_{n,w}^{ips}\right) \cdot \rho_\tau(Y-q)\right],$$

$w_{n,1}^{tt,ps}(D,\mathbf{X}) = D/\,\mathbb{E}_n[D]$, and

$$w_{n,0}^{tt,ps}(D,\mathbf{X};\boldsymbol{\beta}) = \frac{(1-D)\,p(\mathbf{X};\boldsymbol{\beta})}{1-p(\mathbf{X};\boldsymbol{\beta})} \Big/ \mathbb{E}_n\left[\frac{(1-D)\,p(\mathbf{X};\boldsymbol{\beta})}{1-p(\mathbf{X};\boldsymbol{\beta})}\right].$$

The next theorem summarizes the asymptotic properties of the IPW estimators for the treatment effect on the treated based on the IPS. For $j \in \{att, dtt, qtt\}$, let $g^j(Y,D,\mathbf{X}) = g_1^j(Y,D,\mathbf{X}) - g_0^j(Y,D,\mathbf{X})$, with, for $d\in\{0,1\}$,

$$g_d^{att}(Y,D,\mathbf{X}) = w_d^{tt,ps}(D,\mathbf{X};\boldsymbol{\beta}_0)\cdot(Y-\mathbb{E}[Y(d)|D=1]),$$

$$g_d^{dtt}(Y,D,\mathbf{X};y) = w_d^{tt,ps}(D,\mathbf{X};\boldsymbol{\beta}_0)\cdot\left(1\{Y\le y\} - F_{Y(d)|D=1}(y)\right),$$

$$g_d^{qtt}(Y,D,\mathbf{X};\tau) = \frac{w_d^{tt,ps}(D,\mathbf{X};\boldsymbol{\beta}_0)\cdot\left(1\left\{Y\le q_{Y(d)|D=1}(\tau)\right\}-\tau\right)}{f_{Y(d)|D=1}\left(q_{Y(d)|D=1}(\tau)\right)}.$$

Finally, let $\Omega_{w,\Psi}^{att} = \mathbb{E}\left[\psi_{w,\Psi}^{att}(Y,D,\mathbf{X})^2\right]$, $\Omega_{w,\Psi,y}^{dtt} = \mathbb{E}\left[\psi_{w,\Psi}^{dtt}(Y,D,\mathbf{X};y)^2\right]$, and $\Omega_{w,\Psi,\tau}^{qtt} = \mathbb{E}\left[\psi_{w,\Psi}^{qtt}(Y,D,\mathbf{X};\tau)^2\right]$, where $\psi_{w,\Psi}^{att}$, $\psi_{w,\Psi}^{dtt}$, and $\psi_{w,\Psi}^{qtt}$ are defined analogously to (3.5)-(3.7), but with g^{att}, g^{dtt}, g^{qtt} playing the role of g^{ate}, g^{dte}, g^{qte}, respectively, and

$$\mathbf{G}_{\boldsymbol{\beta}}^{att} = \mathbb{E}\left[\frac{g_0^{att}(Y,D,\mathbf{X})}{p(\mathbf{X};\boldsymbol{\beta}_0)(1-p(\mathbf{X};\boldsymbol{\beta}_0))}\cdot\dot{p}(\mathbf{X};\boldsymbol{\beta}_0)\right],$$

$$\mathbf{G}_{\boldsymbol{\beta}}^{dtt}(y) = \mathbb{E}\left[\frac{g_0^{dtt}(Y,D,\mathbf{X};y)}{p(\mathbf{X};\boldsymbol{\beta}_0)(1-p(\mathbf{X};\boldsymbol{\beta}_0))}\cdot\dot{p}(\mathbf{X};\boldsymbol{\beta}_0)\right],$$

$$\mathbf{G}_{\boldsymbol{\beta}}^{qtt}(\tau) = \mathbb{E}\left[\frac{g_0^{qtt}(Y,D,\mathbf{X};\tau)}{p(\mathbf{X};\boldsymbol{\beta}_0)(1-p(\mathbf{X};\boldsymbol{\beta}_0))}\cdot\dot{p}(\mathbf{X};\boldsymbol{\beta}_0)\right],$$

playing the role of $\mathbf{G}_{\boldsymbol{\beta}}^{ate}$, $\mathbf{G}_{\boldsymbol{\beta}}^{dte}$, and $\mathbf{G}_{\boldsymbol{\beta}}^{qte}$, respectively.

Theorem S8.1 *Under Assumptions 2, 3, S8.1, and S8.2, for each $y\in\mathbb{R}$, $\tau\in[\varepsilon,1-\varepsilon]$, we have that, as $n\to\infty$,*

$$\sqrt{n}\left(\widehat{ATT}_n^{ips} - ATT\right) \xrightarrow{d} N\left(0,\Omega_{w,\Psi}^{att}\right),$$

$$\sqrt{n}\left(\widehat{DTT}_n^{ips} - DTT\right)(y) \xrightarrow{d} N\left(0,\Omega_{w,\Psi,y}^{dtt}\right),$$

$$\sqrt{n}\left(\widehat{QTT}_n^{ips} - QTT\right)(\tau) \xrightarrow{d} N\left(0, \Omega^{qtt}_{w,\Psi,\tau}\right).$$

Remark S8.1 When average, distributional and quantile treatment effect on the treated are the main parameters of interest, instead of using (2.7), one may wish to estimate $\boldsymbol{\beta}_0$ such that, for every measurable, integrable function $f(\mathbf{X})$ of the covariates,

$$\mathbb{E}\left[\left(\left(\frac{(1-D)\,p(\mathbf{X};\boldsymbol{\beta}_0)}{1-p(\mathbf{X};\boldsymbol{\beta}_0)}\Bigg/\mathbb{E}\left[\frac{(1-D)\,p(\mathbf{X};\boldsymbol{\beta}_0)}{1-p(\mathbf{X};\boldsymbol{\beta}_0)}\right]\right) - \frac{D}{\mathbb{E}[D]}\right) f(\mathbf{X})\right] = \mathbf{0}. \tag{S8.1}$$

From the discussion in Section 2, and the fact that

$$\frac{(1-D)\,p(\mathbf{X};\boldsymbol{\beta}_0)}{1-p(\mathbf{X};\boldsymbol{\beta}_0)} - D = \frac{(1-D)}{1-p(\mathbf{X};\boldsymbol{\beta}_0)} - 1,$$

and $\mathbb{E}\left[(1-D)\,p(\mathbf{X};\boldsymbol{\beta}_0)/\left(1-p(\mathbf{X};\boldsymbol{\beta}_0)\right)\right] = \mathbb{E}[D]$, we can conclude that one can use

$$H_{0,w}(\boldsymbol{\beta},\mathbf{u}) = \mathbb{E}\left[\left(\left(\frac{(1-D)}{1-p(\mathbf{X};\boldsymbol{\beta})}\Bigg/\mathbb{E}\left[\frac{(1-D)}{1-p(\mathbf{X};\boldsymbol{\beta})}\right]\right) - 1\right) w(\mathbf{X};\mathbf{u})\right]$$

to construct a minimum distance estimator for $\boldsymbol{\beta}_0$ analogous to (2.5). In order to avoid additional cumbersome notation, the results stated in Theorem S8.1 do not use this alternative IPS estimator, though such a modification is straightforward.

References

Chang, M. N. (1990), "Weak Convergence of a Self-Consistent Estimator of the Survival Function with Doubly Censored Data," *The Annals of Statistics*, 18, 391–404.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010), "Quantile and Probability Curves Without Crossing," *Econometrica*, 78(3), 1093–1125.

Escanciano, J. C. (2006), "A consistent diagnostic test for regression models using projections," *Econometric Theory*, 22, 1030–1051.

Firpo, S. (2007), "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75(1), 259–276.

Heckman, J. J., Ichimura, H., and Todd, P. (1997), "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme," *The Review of Economic Studies*, 64(4), 605–654.

Kosorok, M. R. (2008), *Introduction to empirical processes and semiparametric inference*, New York, NY: Springer.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge: Cambridge University Press.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.