Online Appendix

Double Robust, Flexible Adjustment Methods for Causal Inference: An Overview and an Evaluation

Nathan I. Hoffmann, Departments of Sociology and Statistics, UCLA

July 26, 2024

Contents

A LaLonde NSW Data

As another evaluation of these methods, I use data from LaLonde's (1986) study of the National Supported Work Demonstration (NSW), as provided by Dehejia & Wahba (1999). Between March 1975 and July 1977, the NSW randomly provided training to disadvantaged workers. LaLonde used earnings in 1978 as the outcome of interest; comparing earnings in this year for treated and untreated workers allows an experimental estimate of the effect of the intervention. Restricting the sample to men, this study had 297 treated and 425 control participants. Covariates include age, education in years of schooling, earnings in 1975, and dichotomous variables for Black and Hispanic race, married, and not having a high school degree. Following Dehejia & Wahba (1999), I add a variable indicating whether each respondent's earnings in 1975 was $0 – i.e., they were unemployed.

LaLonde compared these experimental estimates to control samples drawn from the Panel Study of Income Dynamics (PSID) and Westat's Matched Current Population Survey-Social Security Administration File (CPS). The PSID-1 sample (n = 2,490) contains all male household heads under 55 who did not classify themselves as retired in 1975, and the PSID-3 sample (n = 128) further restricts this to men who were not working in the spring of 1976 or 1975. The CPS-1 sample (n = 15,992) includes all CPS males under 55, and CPS-3 (n = 429) restricts this two those who were not working in March 1976 whose earnings in 1975 were below the poverty level. Restricting these observational samples gets closer to the group eligible for the NSW.

Following Dehejia & Wahba (1999), I present results for the original samples analyzed by LaLonde (1986), but I also include results using a subsample of the experimental group that has 1974 earnings data available (185 treated and 260 control participants) and include this additional covariate, along with an indicator variable for no earnings in 1974.

Results are presented in Figure A.1, with a table in the Appendix (Table A.1). Standard errors are based on 100 bootstrap samples. We first focus on the original LaLonde dataset, which did not include 1974 earnings. The "experimental" estimates provide a baseline for the comparison, suggesting that the program resulted in an earnings gain of about $800. Some methods calculate widely different results for the experimental estimates, highlighting their instability. Echoing results from the simulations, methods that include logit models are particularly unstable.

If selection on observables holds, then we should be able to recover experimental estimates from the non-experimental control groups. Most of the methods do not perform very well, estimating treatment effects with the wrong sign. The exception is in the PSID-3 sample, where 12 of the 17 methods estimate treatment effects with the correct (positive) sign. This sample is chosen to be closer to the experimental sample.

Including 1974 earnings data results in much better estimates with the observational control groups. OLS, PSM, G-computation (SuperLearner), AIPW (SuperLearner), TMLE (SuperLearner), and DML (OLS/logit) compute fairly stable estimates across the samples. On the other hand, the estimates produced by IPW (logit), IPW (GRF), IPW (SuperLearner), the Lin estimator, and DML (GRF) vary widely across samples.

These results highlight the importance of selection on observables holding. Without including 1974 earnings as a covariate, it appears that selection on observables does not hold, as most methods provide highly inaccurate estimates with the wrong sign. Once 1974 earnings are included, most of the methods provide estimates much closer to the experimental values.

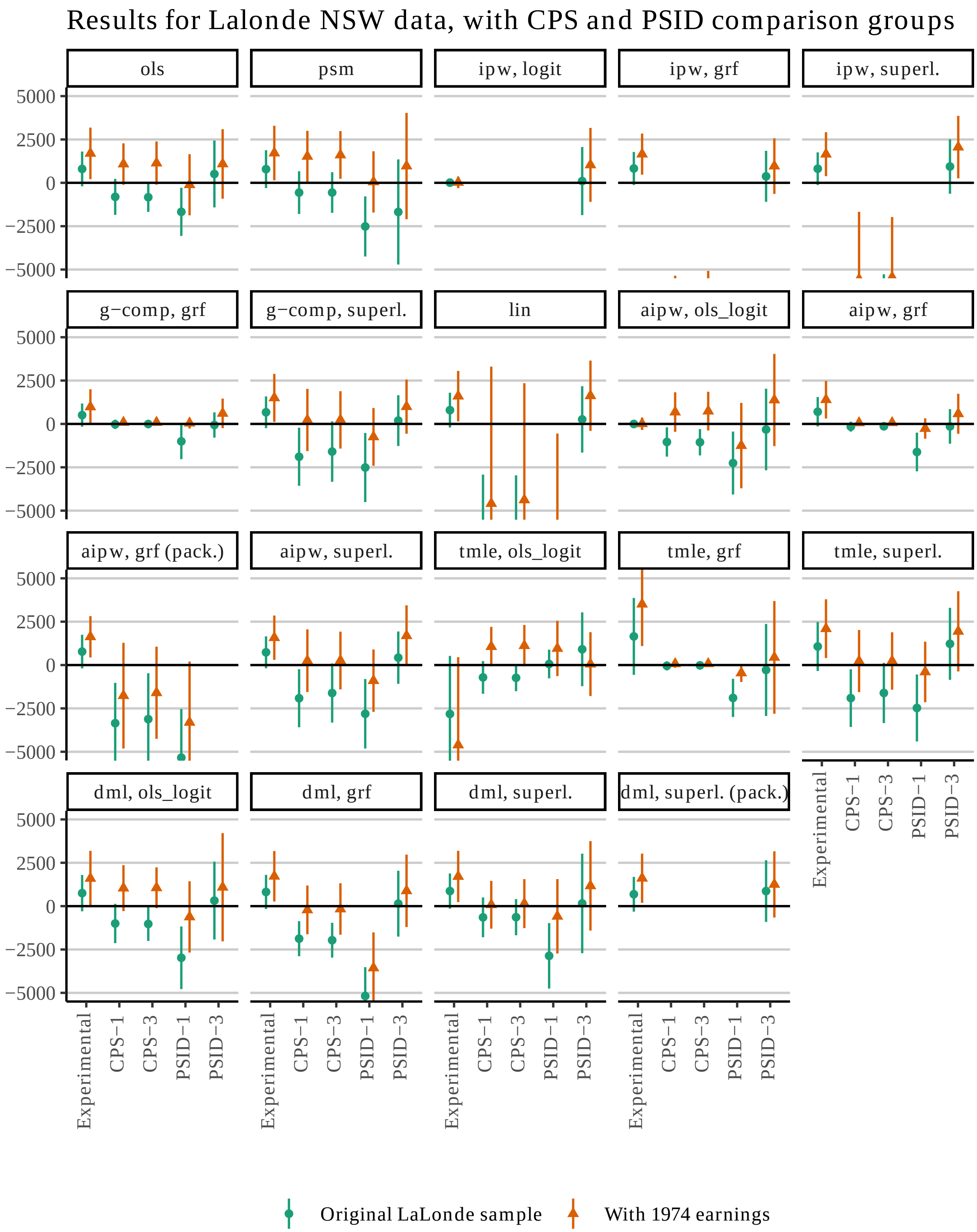

Figure A.1: ATE estimates and 95-percent bootstrap standard error confidence intervals for Lalonde NSW data as provided by Dehejia and Wahba (1999), with CPS and PSID comparison groups. Standard errors shown in parentheses. Covariates include age, education in years of schooling, earnings in 1975, and dichotomous variables for Black and Hispanic race, married, not having a high school degree, and having no earnings in 1975. The "With 1974 earnings" estimates additionally include earnings in 1974 as a covariate, along with an indicator for having no earnings in 1974.

Table A.1: ATE estimates for Lalonde NSW data as provided by Dehejia and Wahba (1999), with CPS and PSID comparison groups. Bootstrap standard errors shown in parentheses. Covariates include age, education in years of schooling, earnings in 1975, and dichotomous variables for Black and Hispanic race, married, not having a high school degree, and having no earnings in 1975. The "With 1974 earnings" estimates additionally include earnings in 1974 as a covariate, along with an indicator for having no earnings in 1974.

’74 earnings?	Method	Experimental	CPS-1	CPS-3	PSID-1	PSID-3
No	ols	802 (511)	-809 (530)	-831 (430)	-1672 (709)	510 (983)
No	psm	788 (557)	-564 (628)	-560 (598)	-2515 (884)	-1680 (1546)
No	ipw, logit	11 (64)	-7299 (673)	-7275 (685)	-10357 (1508)	101 (1002)
No	ipw, grf	827 (485)	-7905 (679)	-7827 (606)	-13333 (1074)	375 (750)
No	ipw, superl.	814 (480)	-7607 (962)	-7338 (1054)	-13654 (1185)	939 (800)
No	g-comp, grf	509 (341)	-22 (138)	-6 (59)	-1004 (524)	-60 (372)
No	g-comp, superl.	673 (467)	-1891 (853)	-1594 (889)	-2514 (1017)	192 (747)
No	lin	795 (512)	-5801 (1469)	-5789 (1440)	-11100 (1537)	262 (976)
No	aipw, ols_logit	-2 (62)	-1041 (430)	-1057 (387)	-2256 (925)	-320 (1200)
No	aipw, grf	702 (432)	-156 (146)	-135 (89)	-1617 (569)	-144 (509)
No	aipw, grf (pack.)	776 (494)	-3354 (1186)	-3120 (1351)	-5340 (1432)	NA (NA)
No	aipw, superl.	734 (469)	-1915 (854)	-1614 (870)	-2811 (1022)	426 (769)
No	tmle, ols_logit	-2819 (1707)	-714 (481)	-737 (393)	59 (421)	909 (1086)
No	tmle, grf	1650 (1131)	-47 (135)	-20 (83)	-1896 (562)	-284 (1354)
No	tmle, superl.	1069 (720)	-1907 (849)	-1611 (886)	-2476 (985)	1223 (1059)
No	dml, ols_logit	750 (534)	-1001 (578)	-1026 (501)	-2977 (921)	320 (1145)
No	dml, grf	817 (502)	-1876 (515)	-1964 (512)	-5180 (847)	142 (968)
No	dml, superl.	869 (517)	-647 (585)	-636 (531)	-2866 (963)	157 (1464)
No	dml, superl. (pack.)	686 (512)	-14774 (433)	-14754 (399)	-9610 (725)	868 (907)
Yes	ols	1698 (758)	1083 (609)	1140 (635)	-111 (899)	1089 (1022)
Yes	psm	1717 (803)	1525 (751)	1606 (701)	54 (901)	968 (1563)
Yes	ipw, logit	23 (169)	-8115 (704)	-7998 (584)	-11317 (1910)	1034 (1088)
Yes	ipw, grf	1655 (605)	-7480 (1083)	-7277 (1121)	-12658 (2031)	965 (817)
Yes	ipw, superl.	1652 (646)	-5560 (1984)	-5444 (1772)	-11066 (2513)	2060 (919)
Yes	g-comp, grf	981 (520)	87 (73)	85 (65)	45 (157)	609 (434)
Yes	g-comp, superl.	1502 (706)	228 (916)	233 (844)	-746 (849)	996 (796)
Yes	lin	1603 (742)	-4590 (4028)	-4380 (3432)	-8717 (4165)	1625 (1035)
Yes	aipw, ols_logit	10 (182)	684 (581)	738 (570)	-1249 (1256)	1381 (1357)
Yes	aipw, grf	1398 (553)	60 (92)	69 (87)	-264 (299)	586 (588)
Yes	aipw, grf (pack.)	1629 (608)	-1766 (1555)	-1598 (1355)	-3306 (1789)	NA (NA)
Yes	aipw, superl.	1576 (652)	251 (921)	260 (850)	-901 (917)	1689 (894)
Yes	tmle, ols_logit	-4609 (2583)	1062 (582)	1115 (609)	956 (813)	54 (940)
Yes	tmle, grf	3509 (1231)	72 (117)	75 (97)	-455 (265)	441 (1658)
Yes	tmle, superl.	2096 (864)	229 (914)	235 (844)	-398 (890)	1943 (1180)
Yes	dml, ols_logit	1599 (810)	1038 (676)	1059 (601)	-621 (1048)	1090 (1591)
Yes	dml, grf	1723 (742)	-217 (717)	-164 (757)	-3565 (1047)	884 (1065)
Yes	dml, superl.	1711 (754)	80 (702)	147 (720)	-585 (1094)	1170 (1316)
Yes	dml, superl. (pack.)	1610 (724)	-17317 (460)	-17235 (609)	-11908 (831)	1252 (974)

B Replications with Fixed Effects

The tables below provide estimates from the ASR replications presented in the main text, but with the addition of two other models. These implement methods proposed by Clarke & Polselli (2024) that incorporate individual or group fixed effects into DML. The authors' package, `XTDML`, builds on the `DoubleML` package. I use two of their procedures. First, correlated random effects (CRE) models explicitly models the correlation between the group-level components in both the treatment and outcome models. Second, the authors' "hybrid" model combines CRE with a within-group estimator that partials out group means from all variables, similar to the function of fixed effects in OLS.

Table B.1: Replication of Aksoy et al. (2022) Table 2, models for outcome of Islamic Votes: "Effect of Fasting Hours (Daylength) during Ramadan on Various Outcome Variables Based on Regression Models That Include Fixed Effects for Provinces and Election Years"

	(1)	(2)	(3)
Original	7.159 **	7.349 **	5.317 **
	(2.539)	(2.491)	(1.855)
AIPW (GRF)	2.058 ***	2.535	1.942 *
	(0.434)	(1.597)	(0.909)
DML (SuperLearner)	2.025 ***	2.931 ***	2.910 ***
	(0.286)	(0.397)	(0.200)
DML (SuperLearner), CRE FE	2.025 ***	2.969 ***	2.900 ***
	(0.290)	(0.399)	(0.142)
DML (SuperLearner), hybrid FE	2.027 ***	2.965 ***	2.900 ***
	(0.036)	(0.059)	(0.053)
Covariates	No	Yes	Yes
Lagged dependent variable	No	No	Yes
Election year fixed effects	Yes	Yes	Yes
Province fixed effects	Yes	Yes	Yes

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Covariates include GDP growth, population, turnout, number of parliamentary seats in province in Model 2 and 3; Models 3 also adjusts for the lagged value of the dependent variable (i.e., one election lagged Islamic votes); cluster robust standard errors are in parentheses. The AIPW (GRF) model does not drop incomplete observations.

Table B.2: Replication of Biegert et al. (2023) Table 3: "Average Marginal Effects from Logistic Regression Models of All-Star Nomination"

	(1)	(2)	(3)	(4)	(5)	(6)
Original	0.607 ***	0.445 ***	0.048 ***	0.035 ***	0.024 ***	0.020 ***
	(0.026)	(0.031)	(0.008)	(0.007)	(0.006)	(0.005)
AIPW (GRF)	0.610 ***	0.538 ***	0.298 ***	0.183 ***	0.168 ***	0.172 ***
	(0.027)	(0.028)	(0.023)	(0.020)	(0.011)	(0.016)
DML (SuperLearner)		0.659 ***	0.590 ***	0.572 ***	0.568 ***	0.566 ***
		(0.019)	(0.021)	(0.021)	(0.021)	(0.021)
DML (SuperLearner), CRE FE		0.548 ***	0.507 ***	0.498 ***	0.493 ***	0.489 ***
		(0.026)	(0.025)	(0.025)	(0.025)	(0.025)
DML (SuperLearner), hybrid FE		0.433 ***	0.422 ***	0.413 ***	0.408 ***	0.405 ***
		(0.022)	(0.023)	(0.023)	(0.023)	(0.023)
Baseline confounders	No	Yes	Yes	Yes	Yes	Yes
Prior situation + performance	No	No	Yes	Yes	Yes	Yes
Current performance	No	No	No	Yes	Yes	Yes
Current situation	No	No	No	No	Yes	Yes
Cumul. AS + cumul. mediators	No	No	No	No	No	Yes

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. With no control variables, Model 1 could not be estimated using DML with a SuperLearner. Model 2 adjusts for year (fixed effects in original model), height (cm), position, age at league entry, being Black, and NBA tenure. Model 3 adds controls for the previous year's average points per 36 minutes, average assists per 36 minutes, average rebounds per 36 minutes, minutes played, whether the team reached playoffs, the team's win percentage, and whether it was a big market team. Model 4 adds controls for current average points, average assists, and average rebounds per 36 minutes. Model 5 additionally controls for current minutes played, whether the team reaches the playoffs, the team win percentage, and whether it is a big market team. Model 6 adds controls for cumulative all-star nominations, average points per 36 minutes, average assists per 36 minutes, average rebounds per 36 minutes, minutes played, team reached playoffs rate, team win percentage, and big market team rate. Player clustered standard errors are in parentheses. None of the original models include player fixed effects. The AIPW (GRF) model does not drop incomplete observations.

Table B.3: Replication of Nussio (2024) Table 2: "Individual-Level Analysis: Community Ties and Lynching Participation"

	(1)	(2)	(3)	(4)
Original	0.036 ***	0.033 ***	0.027 ***	0.032 ***
	(0.006)	(0.006)	(0.007)	(0.008)
AIPW (GRF)	0.035 ***	0.033 ***	0.025 ***	0.025 ***
	(0.006)	(0.006)	(0.006)	(0.006)
DML (SuperLearner)	-0.005	-0.008	-0.015 *	-0.015 *
	(0.006)	(0.006)	(0.007)	(0.007)
DML (SuperLearner), CRE FE	-0.009	-0.012	-0.019 **	-0.019 **
	(0.007)	(0.007)	(0.007)	(0.007)
DML (SuperLearner), hybrid FE	-0.010	-0.013 *	-0.020 ***	-0.020 ***
	(0.005)	(0.005)	(0.005)	(0.005)
Colonia FE	No	No	No	Yes
Control variables	No	Some	All	All

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Models estimate the effect of the log number of names that respondents know in their colonia (neighborhood) on whether they had ever participate in a lynching. Model 1 controls for the extent to which people trust others living in their neighborhood. Model 2 adds controls for education, age, female, number of light bulbs, and unemployment. Model 3 additionally controls for employment, being Catholic or non-religious, whether someone has participated in a fight, whether parents live in the colonia, trust in government, garbage on the street, and residential street block. Colonia clustered standard errors are in parentheses. Model 4 adds colonia fixed effects. The AIPW (GRF) model does not drop incomplete observations.

Table B.4: Replication of Nussio (2024) Table 4: "Aggregate-Level Analysis: Community Ties and Lynching Rate"

	(1)	(2)	(3)	(4)
Original	2.126 ***	2.538 ***	2.178 ***	1.310 ***
	(0.601)	(0.572)	(0.580)	(0.382)
AIPW (GRF)	1.716 **	2.365 ***	2.449 ***	2.449 ***
	(0.639)	(0.645)	(0.670)	(0.670)
DML (SuperLearner)	2.284 ***	2.673 ***	2.374 ***	2.374 ***
	(0.598)	(0.573)	(0.627)	(0.627)
DML (SuperLearner), CRE FE	1.452 **	1.773 ***	1.537 **	1.537 **
	(0.539)	(0.508)	(0.517)	(0.517)
DML (SuperLearner), hybrid FE	1.128	1.247 *	1.176 *	1.176 *
	(0.608)	(0.562)	(0.555)	(0.555)
Control variables	No	Some	All	All
Estado FE	No	No	No	Yes
Estado clustered SE	Yes	Yes	Yes	No

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Models estimate the effect of neighborly cooperation – operationalized by the proportion of respondents who think that most neighbors help each other in problems related to electric lighting – on the log municipal lynching rate. Model 1 adds controls for the proportion of respondents in a municipality that trust their neighbors. Model controls for population, area in square kilometers, poverty rate, Gini coefficient, share indigenous people, non-religious population, homicide rate, and robbery rate. Model 3 adds controls for household victimization and trust in army. Model 4 adds estado (state) fixed effects. Municipality clustered standard errors are in paratheses. The AIPW (GRF) model does not drop incomplete observations.

Table B.5: Replication of Nussio (2024) Table 5: "Natural Experiment: Earthquake Exposure and Lynching"

	(1)	(2)	(3)	(4)
Original	0.108 ***	0.113 ***	0.087 ***	0.126 ***
	(0.019)	(0.021)	(0.019)	(0.026)
AIPW (GRF)	0.122 ***	0.153 ***	0.087 ***	0.135 ***
	(0.019)	(0.022)	(0.019)	(0.021)
DML (SuperLearner)	0.257 ***	0.210 ***	0.231 ***	0.204 ***
	(0.013)	(0.016)	(0.013)	(0.017)
DML (SuperLearner), CRE FE	0.292 ***	0.256 ***	0.269 ***	0.271 ***
	(0.012)	(0.018)	(0.012)	(0.020)
DML (SuperLearner), hybrid FE	0.285 ***	0.248 ***	0.264 ***	0.266 ***
	(0.015)	(0.021)	(0.014)	(0.024)
Control variables	No	Yes	No	Yes

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Models estimate the effect of the Puebla earthquake on September 19, 2017, on the number of lynching events. Data is a panel of Mexican municipalities from 2000 to 2020. Specific independent variables include having an earthquake within 250 km (Models 1 and 2) or having earthquake damage (Models 3 and 4), each interacted with a post-2017 indicator variable. Municipality clustered standard errors are in parentheses. The original models include two-way fixed effects for year and municipality. Model 2 and 4 additionally control for homicides, robberies, kidnappings, and infant mortality. Machine learning models control for year. All models drop incomplete observations; Models 1, and 3 have 51,597 observations, while Models 2, and 4 have 17,225 observations. Due to complications with continuous treatment effects, I do not present the original Models 5 and 6, where the independent variable was distance from the earthquake.

C Tables of Results

Table C.1: Main datasets: Results of Monte Carlo simulations using the first 20 datasets from Dorie et al. (2019), 10 replications each. Percent bias is calculated as the estimator's bias as a percentage of its standard error, rmse is root mean squared error, mae is median absolute error, and comp_time is median computation time measured in seconds for each dataset.

method	estimator	bias	percent_bias	rmse	mae	comp_time	fail_count
ols	NA	0.250	0.157	0.74	0.41	0.061	0
psm	NA	0.203	0.131	0.86	0.53	0.668	6
ipw	logit	-6.690	-1.590	8.14	6.37	0.560	0
ipw	grf	0.433	0.266	0.81	0.46	32.227	0
ipw	superlearner	0.389	0.234	0.73	0.43	130.013	0
g-comp	grf	-0.115	-0.073	0.50	0.26	32.222	0
g-comp	superlearner	0.074	0.048	0.35	0.10	130.006	0
lin	NA	0.209	0.117	0.59	0.28	0.152	0
aipw	ols_logit	-8.240	-1.547	10.06	7.63	0.555	0
aipw	grf	0.060	0.039	0.45	0.22	32.223	0
aipw	grf (pack.)	0.252	0.166	0.51	0.25	8.560	2
aipw	superlearner	0.072	0.047	0.34	0.11	130.007	0
tmle	ols_logit	-1.583	-0.676	2.27	1.47	0.575	0
tmle	grf	0.349	0.230	0.58	0.33	32.241	0
tmle	superlearner	0.073	0.047	0.34	0.10	130.027	0
dml	ols_logit	0.311	0.192	0.79	0.42	0.665	0
dml	grf	0.380	0.247	0.86	0.51	31.574	0
dml	superlearner	0.152	0.068	1.64	0.46	129.160	0
dml	superlearner (pack.)	0.284	0.167	1.12	0.52	7.910	0

Table C.2: Linear datasets: Results of Monte Carlo simulations using the two datasets from Dorie et al. (2019), with linear data generating processes, 100 replications each ("linear"). Percent bias is calculated as the estimator's bias as a percentage of its standard error, rmse is root mean squared error, mae is median absolute error, and comp_time is median computation time measured in seconds for each dataset.

method	estimator	bias	percent_bias	rmse	mae	comp_time	fail_count
ols	NA	-0.024	-0.017	0.39	0.123	0.058	0
psm	NA	-0.053	-0.039	0.52	0.125	0.570	12
ipw	logit	-2.383	-0.897	3.42	2.256	0.596	0
ipw	grf	0.680	0.428	0.98	0.763	30.264	0
ipw	superlearner	0.480	0.309	0.72	0.517	126.091	0
g-comp	grf	-0.343	-0.219	0.61	0.390	30.259	0
g-comp	superlearner	-0.042	-0.029	0.25	0.077	126.085	0
lin	NA	0.027	0.017	0.28	0.072	0.155	0
aipw	ols_logit	-3.435	-1.012	4.75	2.062	0.592	0
aipw	grf	-0.035	-0.024	0.40	0.228	30.261	0
aipw	grf (pack.)	0.184	0.126	0.36	0.208	9.315	0
aipw	superlearner	-0.034	-0.023	0.25	0.076	126.086	0
tmle	ols_logit	-0.536	-0.339	0.91	0.329	0.609	0
tmle	grf	0.305	0.220	0.48	0.331	30.276	0
tmle	superlearner	-0.023	-0.016	0.25	0.077	126.102	0
dml	ols_logit	0.029	0.020	0.39	0.140	0.679	0
dml	grf	0.496	0.336	0.76	0.576	28.689	0
dml	superlearner	0.098	0.069	0.44	0.228	129.669	0
dml	superlearner (pack.)	-0.008	-0.006	0.41	0.155	7.632	0

Table C.3: Sample size: Results of Monte Carlo simulations using dataset 7 from Dorie et al. (2019) with varying sample sizes, 20 replications each. Percent bias is calculated as the estimator's bias as a percentage of its standard error, rmse is root mean squared error, mae is median absolute error, and comp_time is median computation time measured in seconds for each dataset.

method	estimator	size	bias	percent_bias	rmse	mae	comp_time	fail_count
ols	NA	150	NaN	NaN	NaN	NA	0.010	20
ols	NA	300	-0.125	-0.148	1.63	1.769	0.012	17
ols	NA	600	-0.042	-0.034	1.03	0.471	0.018	7
ols	NA	1200	0.215	0.171	0.95	0.590	0.025	0
ols	NA	2400	0.186	0.149	0.83	0.680	0.035	0
ols	NA	4802	0.186	0.150	0.79	0.692	0.061	0
ols	NA	9604	0.275	0.185	0.65	0.477	0.115	0
ols	NA	24010	0.253	0.136	0.62	0.440	0.273	0
ols	NA	48020	0.228	0.242	0.57	0.357	0.597	0
ols	NA	96040	0.183	0.121	0.52	0.314	1.090	0
psm	NA	150	NaN	NaN	NaN	NA	0.013	20
psm	NA	300	NaN	NaN	NaN	NA	0.015	20
psm	NA	600	NaN	NaN	NaN	NA	0.085	20
psm	NA	1200	0.441	0.230	0.55	0.514	0.151	17
psm	NA	2400	0.201	0.150	1.01	0.626	0.274	6
psm	NA	4802	0.174	0.137	0.93	0.761	0.579	1
psm	NA	9604	0.203	0.140	0.66	0.467	1.528	1
psm	NA	24010	0.190	0.105	0.55	0.318	7.664	1
psm	NA	48020	0.198	0.223	0.60	0.334	27.511	1
psm	NA	96040	0.128	0.093	0.37	0.165	113.044	1
ipw	logit	150	0.331	0.185	1.46	1.008	0.077	0
ipw	logit	300	-4.856	-1.395	5.99	3.923	0.098	0
ipw	logit	600	-5.981	-1.964	6.87	5.217	0.140	0
ipw	logit	1200	-7.081	-2.511	7.73	7.662	0.206	0
ipw	logit	2400	-7.201	-2.436	7.88	7.618	0.319	0
ipw	logit	4802	-7.888	-2.757	8.52	8.004	0.529	0
ipw	logit	9604	-9.370	-2.434	10.30	9.599	1.135	0
ipw	logit	24010	-6.516	-2.179	7.49	6.492	1.783	0
ipw	logit	48020	-8.914	-2.652	9.66	8.666	3.513	0
ipw	logit	96040	-8.635	-1.779	9.88	8.161	7.372	0
ipw	grf	150	0.351	0.206	1.24	0.787	1.613	0
ipw	grf	300	0.377	0.248	1.17	0.588	2.533	0
ipw	grf	600	0.351	0.248	0.99	0.421	4.428	0
ipw	grf	1200	0.425	0.303	0.99	0.552	7.752	0
ipw	grf	2400	0.334	0.241	0.79	0.576	15.160	0
ipw	grf	4802	0.320	0.240	0.69	0.492	36.832	0
ipw	grf	9604	0.244	0.148	0.60	0.268	77.877	0
ipw	grf	24010	0.299	0.162	0.59	0.294	217.183	0
ipw	grf	48020	0.172	0.143	0.56	0.390	589.802	0
ipw	grf	96040	0.128	0.077	0.34	0.182	1872.404	0
ipw	superlearner	150	0.294	0.180	1.18	0.869	10.070	0
ipw	superlearner	300	0.334	0.238	1.02	0.661	12.480	0
ipw	superlearner	600	0.367	0.264	0.98	0.451	21.676	0
ipw	superlearner	1200	0.392	0.285	0.94	0.583	37.717	0
ipw	superlearner	2400	0.342	0.255	0.76	0.519	67.716	0
ipw	superlearner	4802	0.319	0.248	0.65	0.449	125.311	0
ipw	superlearner	9604	0.249	0.152	0.56	0.157	251.578	0

ipw	superlearner	24010	0.273	0.148	0.51	0.288	623.468	0
ipw	superlearner	48020	0.103	0.090	0.46	0.322	1175.452	0
ipw	superlearner	96040	0.021	0.013	0.31	0.131	2240.567	0
g-comp	grf	150	-2.439	-1.828	2.77	2.528	1.611	0
g-comp	grf	300	-1.758	-1.177	2.08	1.951	2.531	0
g-comp	grf	600	-1.089	-0.769	1.42	1.187	4.426	0
g-comp	grf	1200	-0.634	-0.498	0.99	0.655	7.751	0
g-comp	grf	2400	-0.386	-0.324	0.71	0.378	15.157	0
g-comp	grf	4802	-0.250	-0.219	0.56	0.192	36.828	0
g-comp	grf	9604	-0.127	-0.073	0.37	0.239	77.868	0
g-comp	grf	24010	-0.079	-0.045	0.27	0.156	217.163	0
g-comp	grf	48020	-0.117	-0.115	0.24	0.103	589.773	0
g-comp	grf	96040	-0.154	-0.098	0.26	0.123	1872.343	0
g-comp	superlearner	150	-0.567	-0.385	1.33	0.607	10.035	0
g-comp	superlearner	300	-0.359	-0.283	0.89	0.528	12.479	0
g-comp	superlearner	600	-0.197	-0.157	0.75	0.293	21.673	0
g-comp	superlearner	1200	0.048	0.041	0.60	0.301	37.716	0
g-comp	superlearner	2400	0.006	0.005	0.38	0.150	67.708	0
g-comp	superlearner	4802	-0.004	-0.003	0.28	0.089	125.308	0
g-comp	superlearner	9604	0.122	0.078	0.33	0.107	251.571	0
g-comp	superlearner	24010	0.019	0.011	0.14	0.050	623.449	0
g-comp	superlearner	48020	-0.010	-0.010	0.25	0.053	1175.419	0
g-comp	superlearner	96040	0.041	0.025	0.21	0.151	2240.497	0
lin	NA	150	NaN	NaN	NaN	NA	0.011	20
lin	NA	300	-1.027	-0.557	2.93	2.322	0.011	17
lin	NA	600	0.175	0.138	1.16	0.550	0.030	7
lin	NA	1200	0.310	0.225	0.94	0.493	0.048	0
lin	NA	2400	0.255	0.184	0.73	0.540	0.078	0
lin	NA	4802	0.230	0.170	0.72	0.517	0.157	0
lin	NA	9604	0.258	0.158	0.58	0.440	0.325	0
lin	NA	24010	0.402	0.206	0.69	0.489	0.824	0
lin	NA	48020	0.111	0.112	0.51	0.354	1.646	0
lin	NA	96040	0.007	0.004	0.89	0.464	3.359	0
aipw	ols_logit	150	-0.368	-0.140	1.86	0.681	0.075	0
aipw	ols_logit	300	-6.828	-0.814	10.31	4.078	0.097	0
aipw	ols_logit	600	-6.623	-1.890	7.53	5.425	0.139	0
aipw	ols_logit	1200	-8.480	-2.252	9.23	8.810	0.204	0
aipw	ols_logit	2400	-8.606	-2.188	9.48	8.856	0.318	0
aipw	ols_logit	4802	-9.584	-2.455	10.41	10.369	0.527	0
aipw	ols_logit	9604	-11.622	-2.343	12.81	11.258	1.127	0
aipw	ols_logit	24010	-7.479	-2.309	8.59	6.902	1.772	0
aipw	ols_logit	48020	-11.523	-2.400	12.66	11.548	3.481	0
aipw	ols_logit	96040	-10.762	-1.780	12.29	9.426	7.334	0
aipw	grf	150	-0.517	-0.349	1.32	0.927	1.612	0
aipw	grf	300	-0.383	-0.285	0.98	0.581	2.532	0
aipw	grf	600	-0.290	-0.222	0.81	0.330	4.426	0
aipw	grf	1200	-0.116	-0.092	0.66	0.183	7.751	0
aipw	grf	2400	-0.086	-0.071	0.50	0.213	15.158	0
aipw	grf	4802	-0.073	-0.063	0.42	0.189	36.829	0
aipw	grf	9604	-0.035	-0.021	0.29	0.227	77.869	0
aipw	grf	24010	-0.001	-0.001	0.24	0.087	217.166	0
aipw	grf	48020	-0.075	-0.074	0.20	0.055	589.777	0
aipw	grf	96040	-0.132	-0.084	0.24	0.107	1872.350	0
aipw	grf (pack.)	150	0.157	0.101	1.23	0.976	0.162	0

aipw	grf (pack.)	300	0.126	0.091	0.95	0.611	0.326	0
aipw	grf (pack.)	600	0.091	0.072	0.79	0.346	0.675	0
aipw	grf (pack.)	1200	0.189	0.149	0.66	0.295	1.550	0
aipw	grf (pack.)	2400	0.137	0.110	0.44	0.256	3.633	0
aipw	grf (pack.)	4802	0.144	0.119	0.35	0.235	9.748	0
aipw	grf (pack.)	9604	0.164	0.098	0.25	0.226	24.451	1
aipw	grf (pack.)	24010	0.134	0.070	0.21	0.119	79.418	4
aipw	grf (pack.)	48020	0.047	0.058	0.15	0.140	208.004	5
aipw	grf (pack.)	96040	0.105	0.060	0.18	0.031	325.812	15
aipw	superlearner	150	-0.347	-0.238	1.24	0.590	10.035	0
aipw	superlearner	300	-0.283	-0.225	0.85	0.391	12.479	0
aipw	superlearner	600	-0.169	-0.133	0.77	0.268	21.674	0
aipw	superlearner	1200	0.057	0.048	0.62	0.300	37.717	0
aipw	superlearner	2400	0.011	0.009	0.38	0.149	67.708	0
aipw	superlearner	4802	-0.003	-0.002	0.27	0.081	125.308	0
aipw	superlearner	9604	0.102	0.064	0.28	0.095	251.572	0
aipw	superlearner	24010	0.023	0.013	0.14	0.050	623.451	0
aipw	superlearner	48020	-0.016	-0.016	0.20	0.050	1175.424	0
aipw	superlearner	96040	-0.020	-0.012	0.16	0.100	2240.504	0
tmle	ols_logit	150	-0.138	-0.077	1.33	0.587	0.080	0
tmle	ols_logit	300	-1.275	-0.650	2.12	0.954	0.101	0
tmle	ols_logit	600	-1.648	-0.850	2.24	1.452	0.147	0
tmle	ols_logit	1200	-1.350	-0.732	2.07	1.317	0.212	0
tmle	ols_logit	2400	-1.359	-0.653	2.04	1.065	0.327	0
tmle	ols_logit	4802	-1.283	-0.625	2.06	0.948	0.544	0
tmle	ols_logit	9604	-0.680	-0.311	1.37	1.063	1.185	0
tmle	ols_logit	24010	-1.124	-0.498	1.64	0.797	1.875	0
tmle	ols_logit	48020	-1.248	-0.622	1.94	0.923	3.693	0
tmle	ols_logit	96040	-0.864	-0.543	1.46	0.643	7.596	0
tmle	grf	150	1.186	0.693	1.81	1.649	1.615	0
tmle	grf	300	1.413	0.638	2.43	1.178	2.537	0
tmle	grf	600	0.795	0.537	1.37	0.871	4.430	0
tmle	grf	1200	0.573	0.422	0.94	0.704	7.756	0
tmle	grf	2400	0.341	0.265	0.64	0.541	15.167	0
tmle	grf	4802	0.225	0.183	0.40	0.416	36.842	0
tmle	grf	9604	0.137	0.085	0.34	0.255	77.898	0
tmle	grf	24010	0.113	0.064	0.28	0.143	217.227	0
tmle	grf	48020	-0.003	-0.003	0.16	0.071	589.889	0
tmle	grf	96040	-0.106	-0.067	0.23	0.113	1872.553	0
tmle	superlearner	150	-0.031	-0.022	1.23	0.757	10.039	0
tmle	superlearner	300	-0.110	-0.088	0.90	0.557	12.484	0
tmle	superlearner	600	-0.088	-0.070	0.81	0.303	21.679	0
tmle	superlearner	1200	0.089	0.075	0.66	0.298	37.722	0
tmle	superlearner	2400	0.030	0.023	0.40	0.150	67.717	0
tmle	superlearner	4802	0.005	0.004	0.29	0.098	125.321	0
tmle	superlearner	9604	0.086	0.054	0.26	0.091	251.605	0
tmle	superlearner	24010	0.014	0.008	0.14	0.048	623.518	0
tmle	superlearner	48020	0.002	0.002	0.14	0.058	1175.546	0
tmle	superlearner	96040	-0.140	-0.092	0.31	0.105	2240.722	0
dml	ols_logit	150	-56.640	-0.281	204.19	3.448	0.093	0
dml	ols_logit	300	-2.596	-0.162	15.70	2.425	0.140	0
dml	ols_logit	600	-0.263	-0.050	5.01	0.902	0.196	0
dml	ols_logit	1200	0.654	0.362	1.90	0.770	0.265	0
dml	ols_logit	2400	-0.291	-0.105	2.47	0.656	0.450	0

dml	ols_logit	4802	-0.098	-0.054	1.42	0.695	0.660	0
dml	ols_logit	9604	0.263	0.174	0.64	0.432	1.173	0
dml	ols_logit	24010	0.294	0.156	0.66	0.475	2.197	0
dml	ols_logit	48020	0.218	0.227	0.58	0.327	3.121	0
dml	ols_logit	96040	0.122	0.079	0.57	0.335	6.145	0
dml	grf	150	-0.444	-0.260	1.09	0.996	1.756	14
dml	grf	300	-0.270	-0.100	2.50	1.062	3.010	0
dml	grf	600	0.379	0.238	1.38	1.014	4.722	0
dml	grf	1200	0.593	0.356	1.48	0.619	8.475	0
dml	grf	2400	0.409	0.301	0.98	0.744	16.105	0
dml	grf	4802	0.423	0.306	0.96	0.626	31.840	0
dml	grf	9604	0.283	0.186	0.61	0.399	65.090	0
dml	grf	24010	0.179	0.097	0.49	0.377	174.612	0
dml	grf	48020	0.227	0.223	0.54	0.368	391.386	0
dml	grf	96040	0.188	0.119	0.32	0.210	855.111	0
dml	superlearner	150	-1.382	-0.229	5.75	1.051	9.933	0
dml	superlearner	300	0.366	0.206	1.76	1.072	18.926	0
dml	superlearner	600	1.394	0.356	3.98	0.898	23.790	0
dml	superlearner	1200	1.884	0.427	4.73	0.884	40.678	0
dml	superlearner	2400	1.729	0.423	4.38	1.166	75.058	0
dml	superlearner	4802	1.672	0.456	3.96	1.027	129.318	0
dml	superlearner	9604	0.110	0.062	1.15	0.586	242.361	0
dml	superlearner	24010	-0.081	-0.037	1.31	0.630	574.176	0
dml	superlearner	48020	-0.107	-0.028	3.49	0.573	1105.293	0
dml	superlearner	96040	-0.092	-0.041	1.33	0.340	2218.412	0
dml	superlearner (pack.)	150	-1.112	-0.270	3.91	1.089	5.118	0
dml	superlearner (pack.)	300	-0.896	-0.300	2.93	1.103	4.047	0
dml	superlearner (pack.)	600	-0.941	-0.270	3.41	0.921	4.497	0
dml	superlearner (pack.)	1200	-0.864	-0.262	3.15	0.812	4.392	0
dml	superlearner (pack.)	2400	-0.932	-0.281	3.17	0.848	4.934	0
dml	superlearner (pack.)	4802	-0.997	-0.288	3.33	0.875	6.043	0
dml	superlearner (pack.)	9604	0.240	0.147	0.87	0.505	9.210	0
dml	superlearner (pack.)	24010	0.276	0.137	1.01	0.522	18.139	0
dml	superlearner (pack.)	48020	0.365	0.146	2.59	0.489	32.800	0
dml	superlearner (pack.)	96040	0.344	0.197	1.65	0.476	64.226	0

D References

Clarke, P., & Polselli, A. (2024). *Double Machine Learning for Static Panel Models with Fixed Effects* (No. arXiv:2312.08174). arXiv. https://doi.org/10.48550/arXiv.2312.08174

Dehejia, R. H., & Wahba, S. (1999). Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, *94*(448), 1053–1062. https://doi.org/10.2307/2669919

LaLonde, R. J. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *The American Economic Review*, *76*(4), 604–620. https://www.jstor.org/stable/1806062