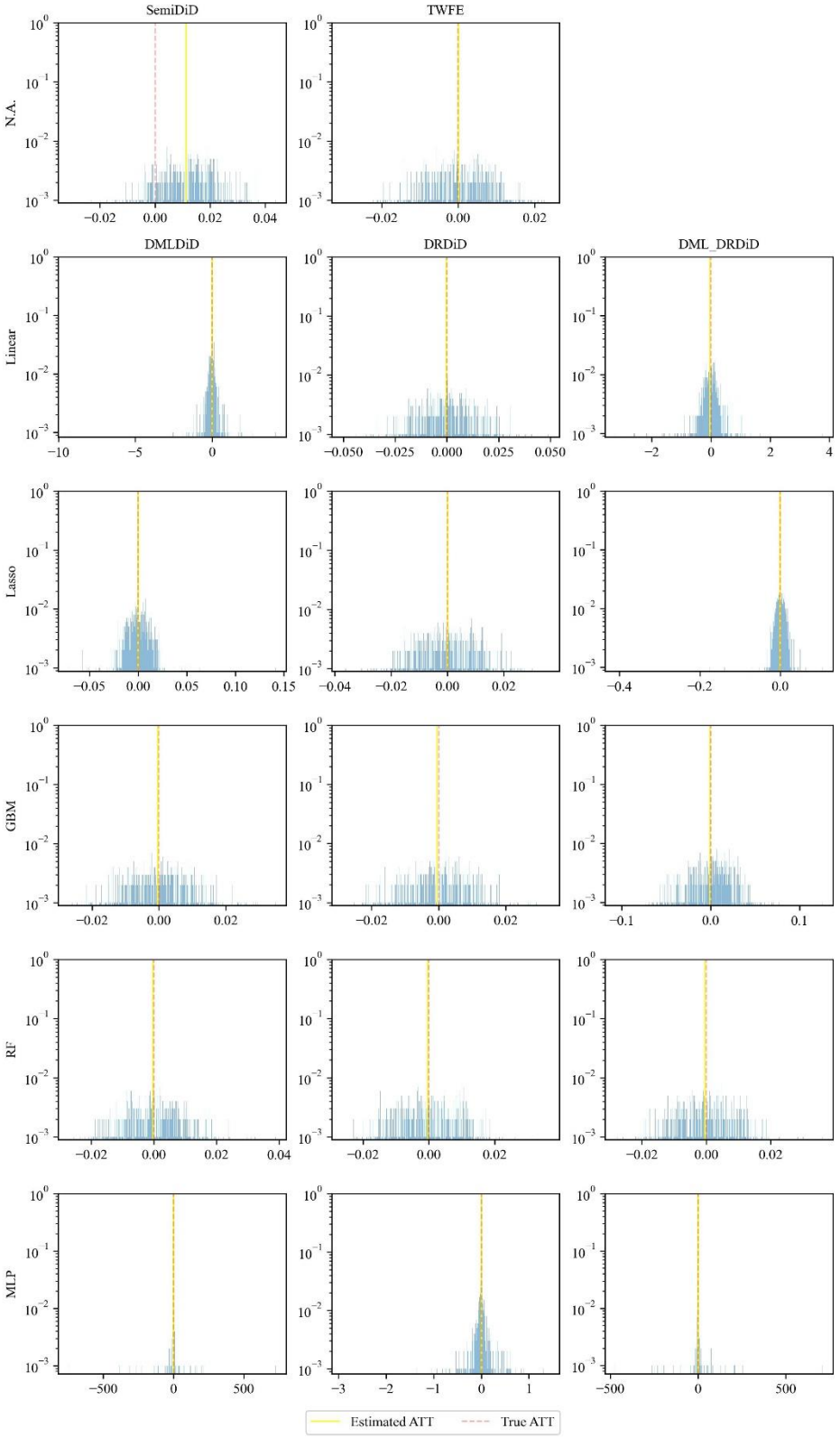
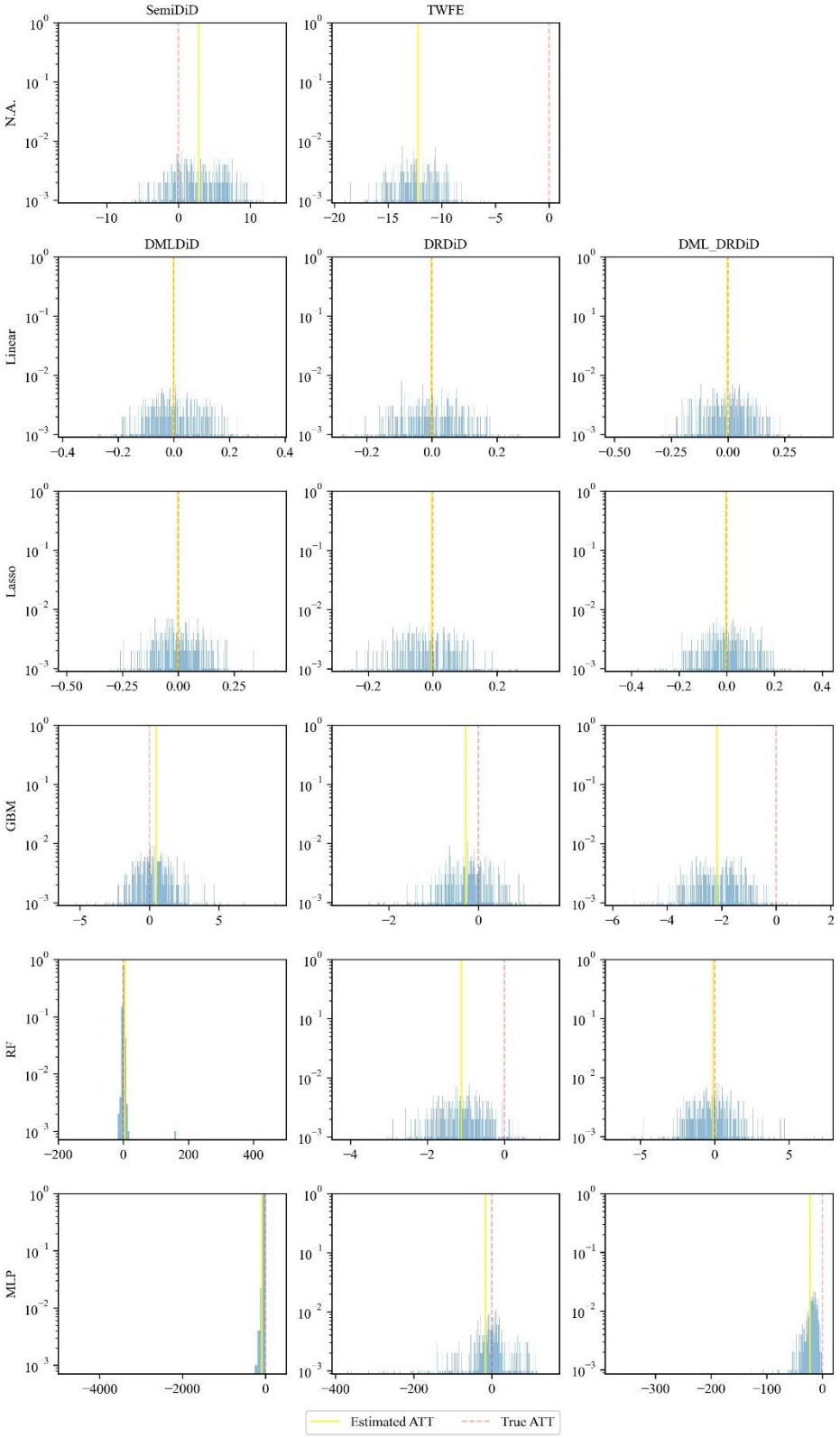
Appendix

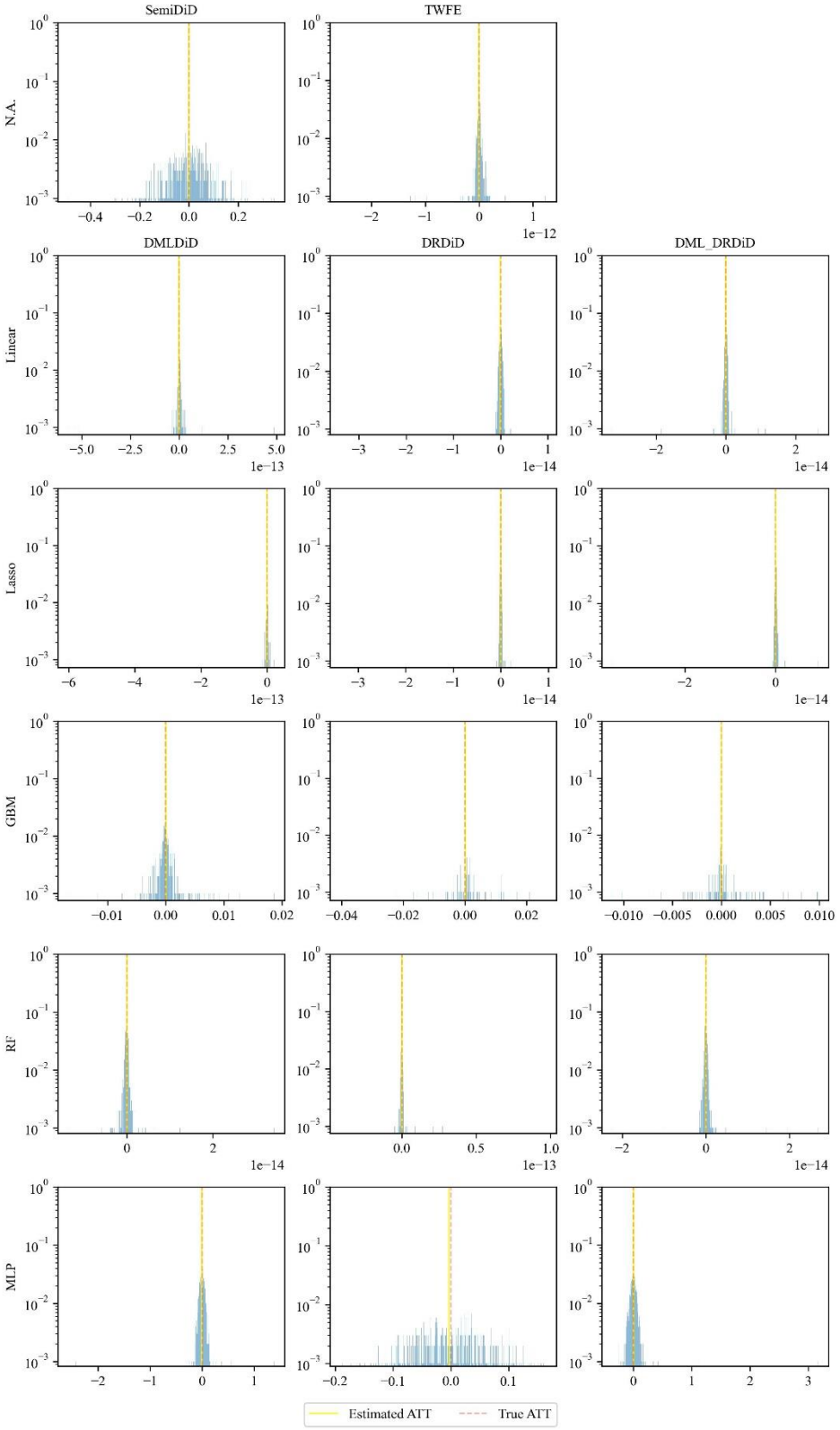
This is the appendix to the paper titled ‘Machine learning-based causal inference for travel behavior analysis: a difference-in-differences framework’. It contains two figures, Figure S 1 and Figure S 2, showing the detailed distributions of estimation biases using different estimators and machine learning algorithms and four tables (Table S 1-Table S 4) documenting the full lists of variables and results of the two empirical case studies.


SemiDiD
TWFE
DMLDiD
DRDiD
DML_DRDiD
N.A.
Linear
Lasso
GBM
RF
MLP
Estimated ATT
True ATT

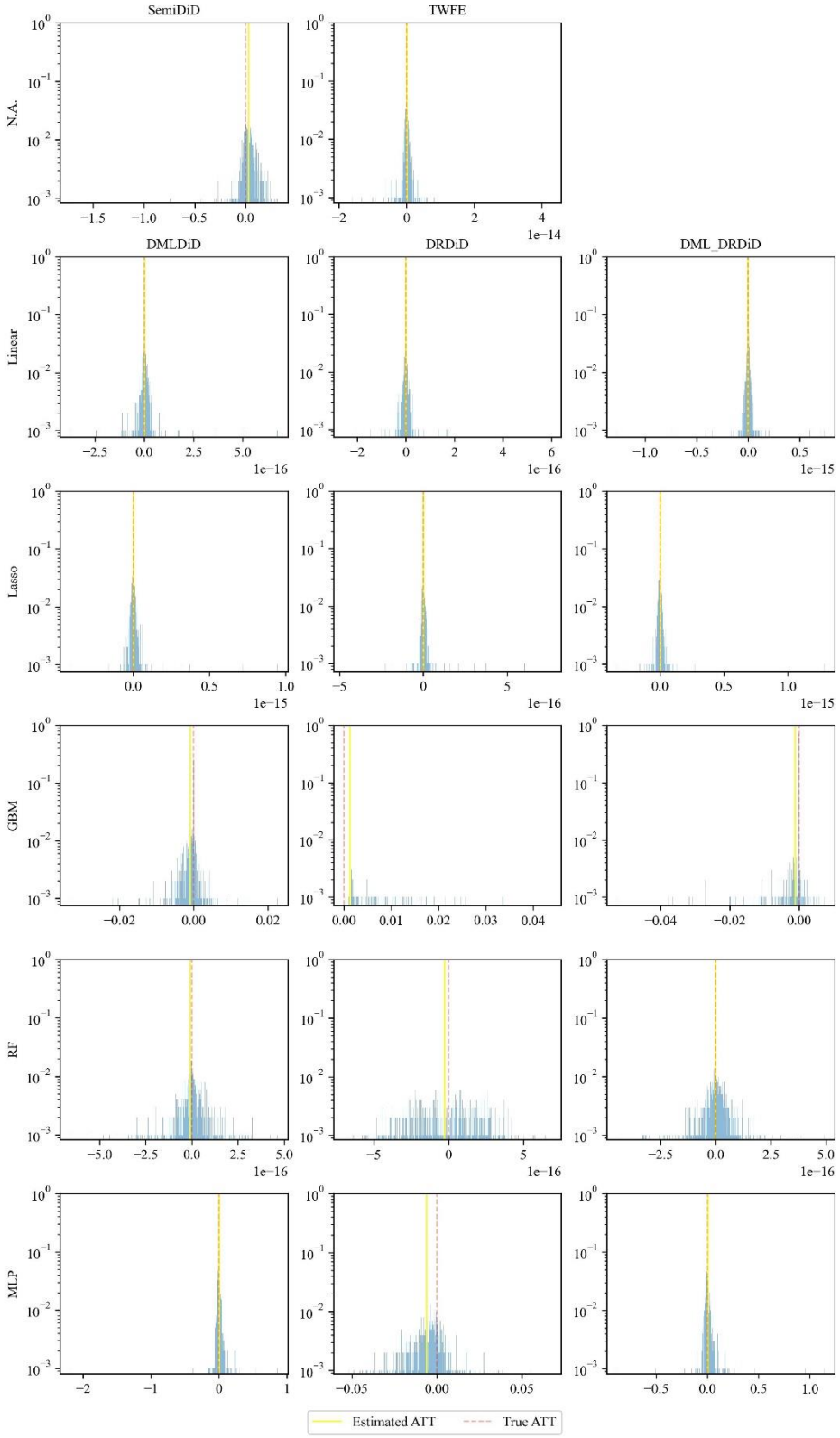
(a) Scenario A


SemiDiD
TWFE
DMLDiD
DRDiD
DML_DRDiD
N.A.
Linear
Lasso
GBM
RF
MLP
Estimated ATT
True ATT

(b) Scenario B

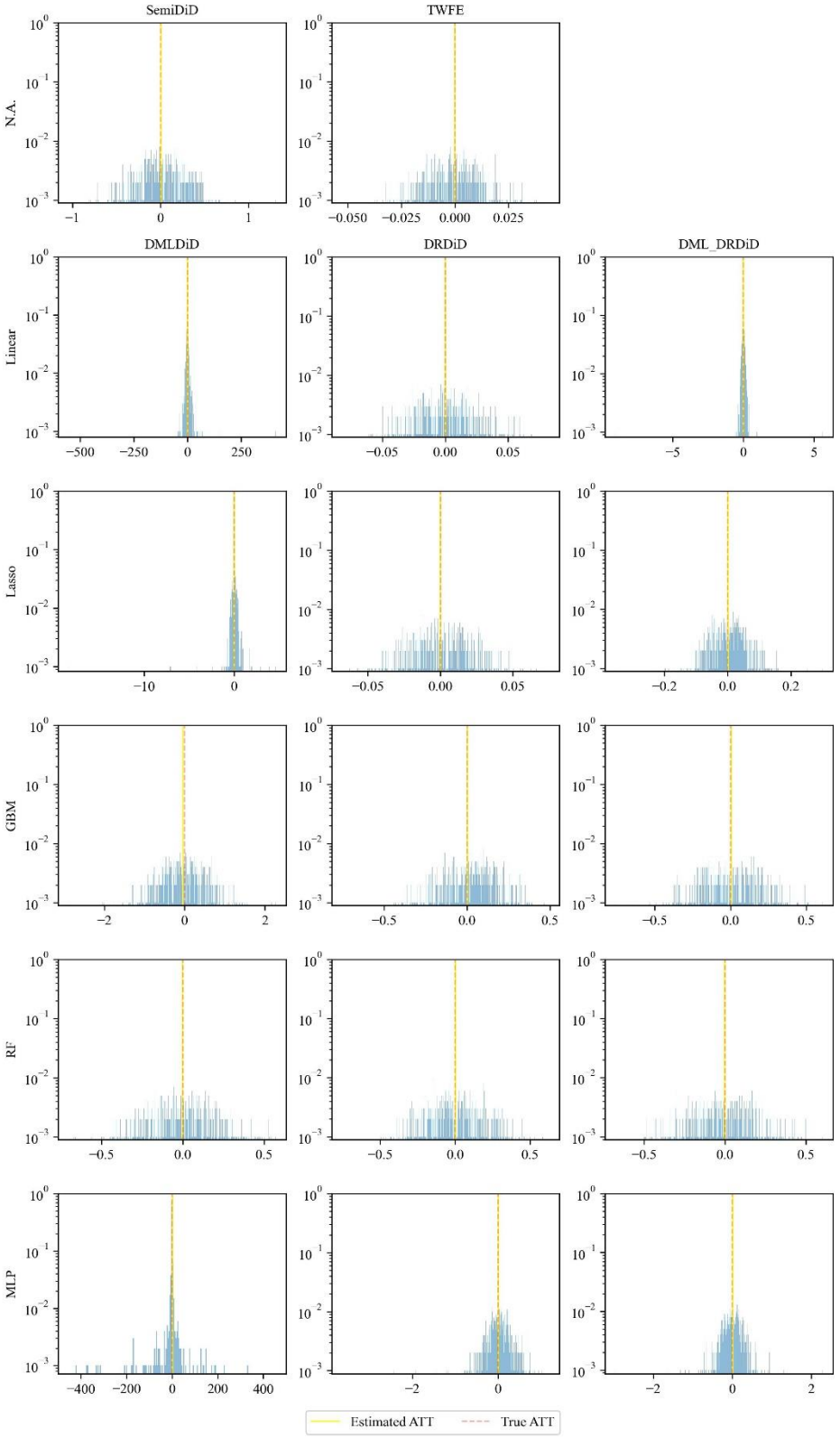

SemiDiD
TWFE
DMLDiD
DRDiD
DML_DRDiD
N.A.
Linear
Lasso
GBM
RF
MLP
1e−12
1e−13
1e−14
Estimated ATT
True ATT

(c) Scenario C

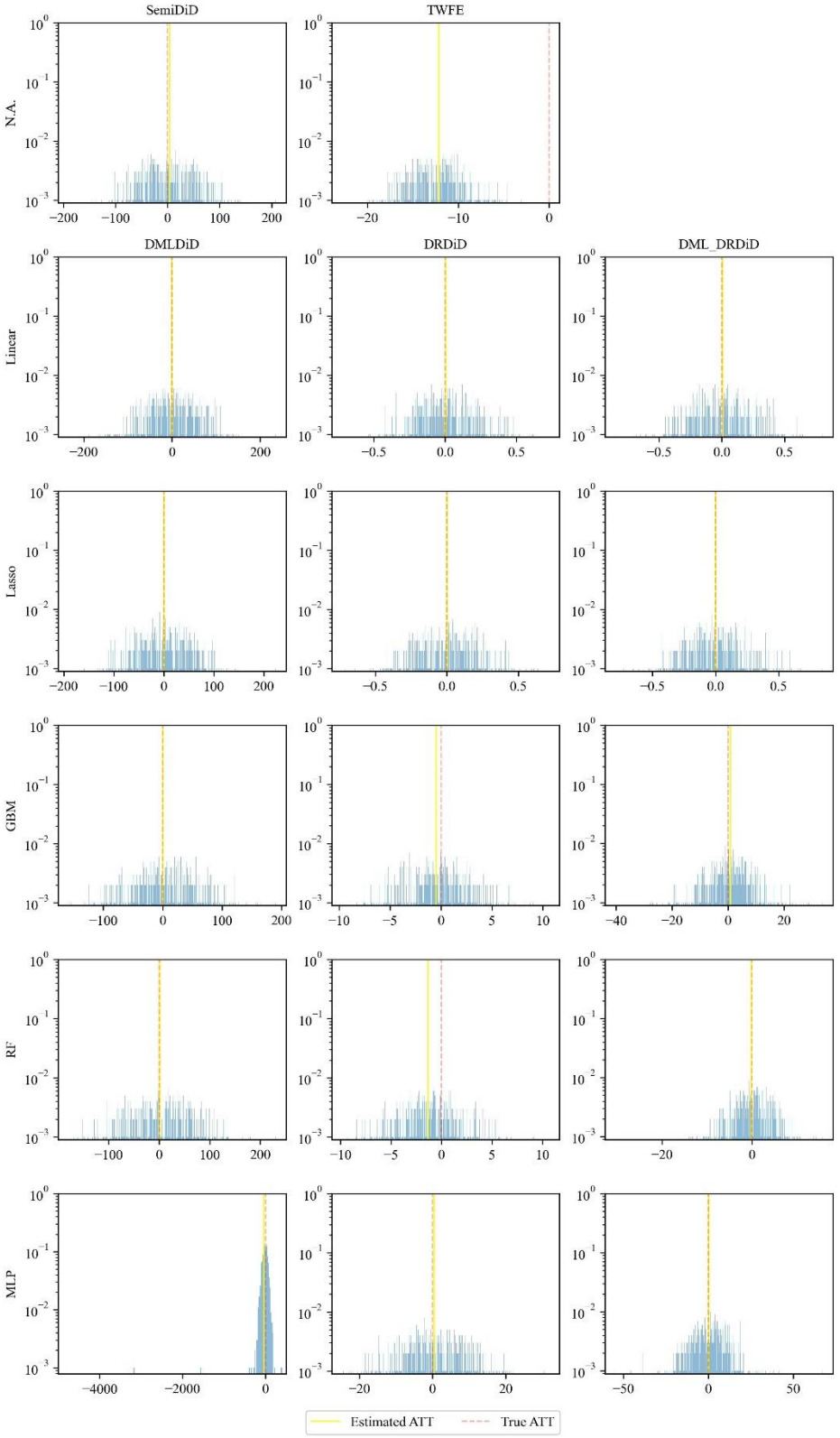


(d) Scenario D

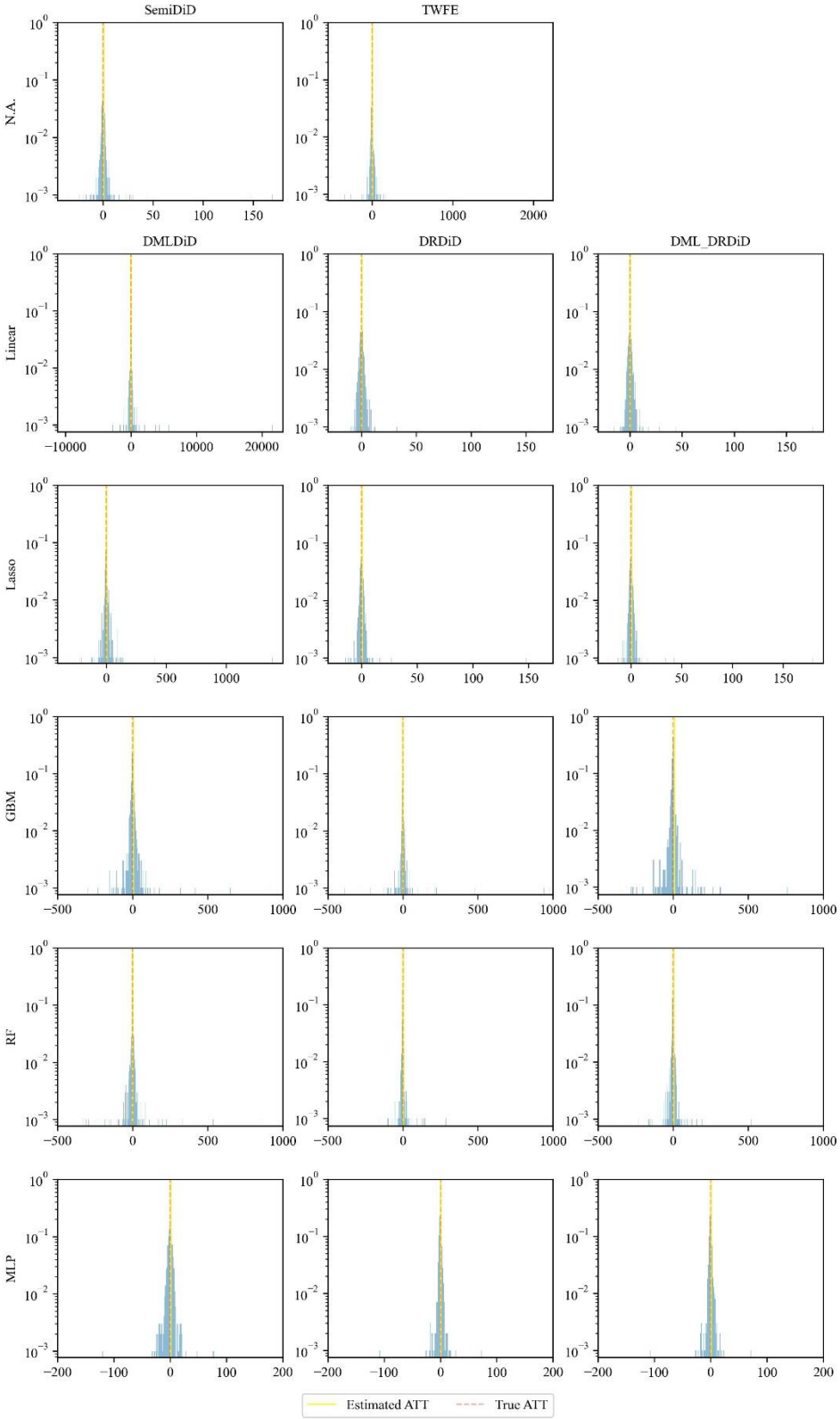
Figure S 1. Distributions of ATT estimates in the simulation study for panel data


SemiDiD
TWFE
DMLDiD
DRDiD
DML_DRDiD
N.A.
Linear
Lasso
GBM
RF
MLP
Estimated ATT
True ATT

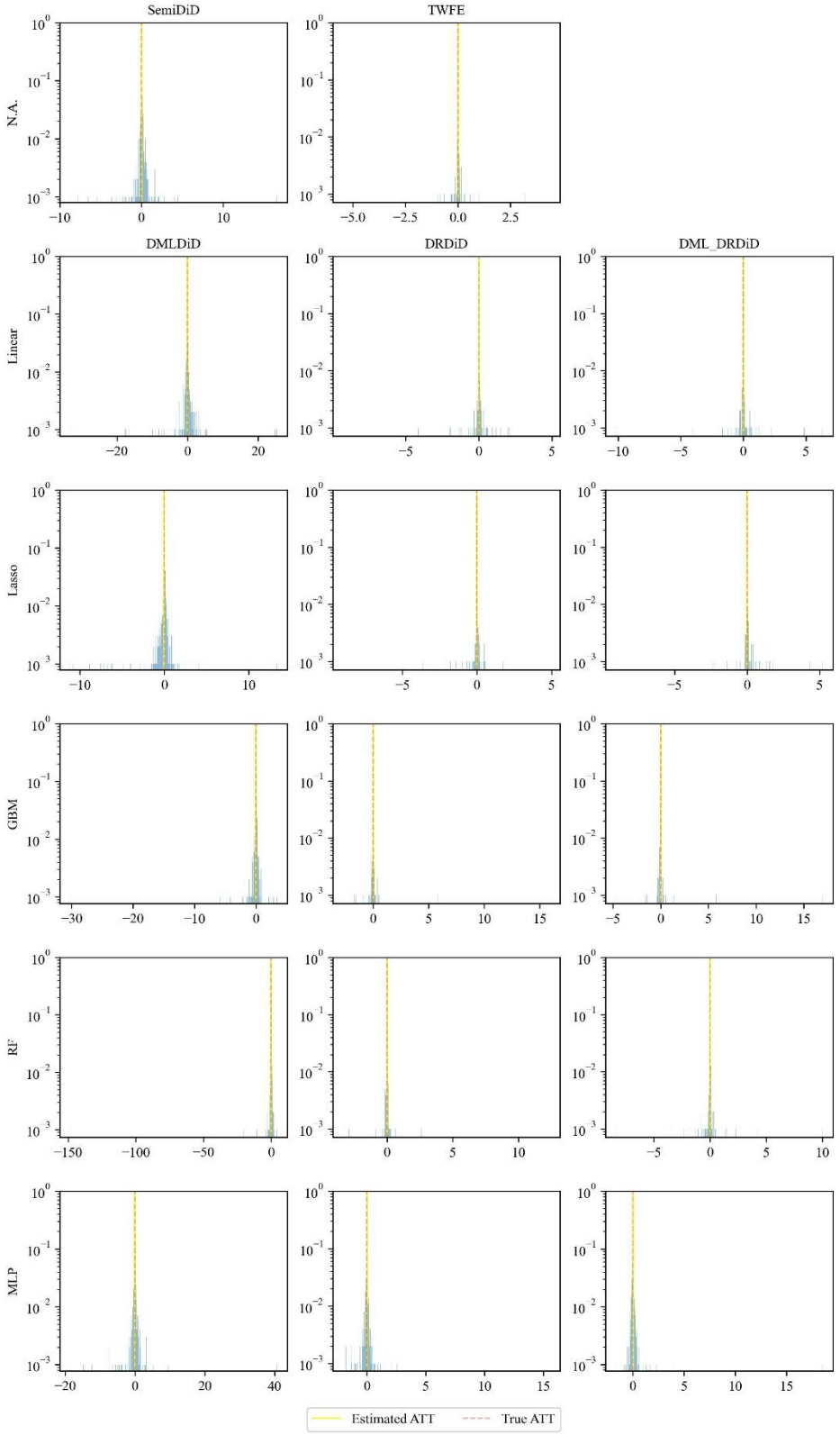
(a) Scenario A


SemiDiD
TWFE
DMLDiD
DRDiD
DML_DRDiD
N.A.
Linear
Lasso
GBM
RF
MLP
Estimated ATT
True ATT

(b) Scenario B


SemiDiD
TWFE
DMLDiD
DRDiD
DML_DRDiD
N.A.
Linear
Lasso
GBM
RF
MLP
Estimated ATT
True ATT

(c) Scenario C



(d) Scenario D

Figure S 2. Distributions of ATT estimates in the simulation study for repeated cross-section data

Table S 1. Full variables included in the panel empirical case study

Variable	Original variable	Code/range	Original code	Note
Treatment				
Newly provided work-from-home option since pandemic	wfh_pre wfh_now	0: No 1: Yes	No Question not displayed to respondent Yes	Computed using two original variables 1 if wfh_pre != 'Yes' & wfh_now = 'Yes' 0 otherwise
Covariates				
Household size	hhsize	1-11	1-11	
Household vehicle ownership	hhveh_harm	0: No 1: Yes	0-3 4 or more	Aggregated to binary
Number of children in household	nchildren		0-9	
Home ownership	tenure_harm		Other, please explain Own with a mortgage Own without a mortgage Rent	
Moving since pandemic	home_move		No Yes, to a new metropolitan area Yes, within the same metropolitan area	
Age	age	18-87	18-87	Has zero-frequency values
Gender	gender	0: Male 1: Female	Female Male	Recoded
Is student	studentjs	0: No	No	

Variable	Original variable	Code/range	Original code	Note
Driver's license	driver	1: Yes	Yes	
Bike ownership	bike			
Concerned for severe reaction to Covid-19	att_covid_selfsevere		Neutral Somewhat agree Somewhat disagree Strongly agree Strongly disagree	Aggregated to binary with 'Somewhat/ Strongly agree' as 1
Think everyone should stay home during pandemic	att_covid_stayhome			
Feel community is well-prepared for pandemic	att_covid_commdisasters			
Think society is overreacting to pandemic	att_covid_overreact			
Like working from home	att_wfh_likewfh			
Home-work distance	pre_work_com_dist	0-2153	0-2153	In miles
White/Caucasian race	race_1	0: No 1: Yes	Not selected White/Caucasian	Recoded
Black/African American race	race_2		Black/African American Not selected	
American Indian and Alaska Native race	race_3		American Indian and Alaska Native Not selected	
Asian race	race_4		Asian Not selected	
Native Hawaiian or Other Pacific Islander race	race_5		Native Hawaiian or Other Pacific Islander Not selected	
Household income categories	hhincome	0: Not in this	$10,000 to $14,999	One-hot encoded into ascending binaries

Variable	Original variable	Code/range	Original code	Note
		category 1: In this category	$100,000 to $124,999 $125,000 to $149,999 $15,000 to $24,999 $150,000 to $199,999 $200,000 or more $25,000 to $34,999 $35,000 to $49,999 $50,000 to $74,999 $75,000 to $99,999 Less than $10,000	with 'Less than $10,000' as reference
Job industry categories	jobcat_pre_harm		Clerical or administrative support I prefer not to answer Manufacturing, construction, maintenance, or farming Professional, managerial, or technical Question not displayed to respondent Sales or service Something else	One-hot encoded into unordered binaries with 'I prefer not to answer' as reference
Education attainment categories	educ		Bachelor's degree(s) or some graduate school Completed graduate degree(s) Completed high school or GED	One-hot encoded into binaries with 'Some grade/high school' as reference

Variable	Original variable	Code/range	Original code	Note
			Some college or technical school Some grade/high school	
Outcomes				
No. of days a week commuting to work before pandemic	pre_work_com_days	0-7	0-7	
No. of days commuting to week the past week	now_work_com_days			
Time commuting to work before pandemic	pre_work_pri_time	0-480	0-480	In minutes
Time commuting to work now	now_work_pri_time	0-320	0-320	
Commuting mode before pandemic categories	pre_work_pri_mode_harm	0: Not in this category 1: In this category	Bicycle or scooter Other mode Private vehicle Question not displayed to respondent Transit Walks	One-hot encoded into unordered binaries with 'Question not displayed to respondent' as reference
Commuting mode now categories	now_work_pri_mode_harm			

Table S 2. Full results of the panel empirical case study

Outcome	Estimator	Model	All workers (N=4,733)		Only commuters (N=2,374)	
			ATT (S.E.)	95% confidence interval	ATT (S.E.)	95% confidence interval
No. of days commuting to work in past week	TWFE	N.A.	-1.942*** (0.093)	[-2.124, -1.761]	-1.366*** (0.106)	[-1.573, -1.16]
	SemiDiD	N.A.	-1.25*** (0.08)	[-1.407, -1.094]	-0.834*** (0.112)	[-1.053, -0.615]
	DMLDiD	Linear	-1.371*** (0.078)	[-1.525, -1.218]	-1.076*** (0.106)	[-1.283, -0.869]
		Lasso	-1.454*** (0.088)	[-1.626, -1.282]	-1.174*** (0.13)	[-1.43, -0.919]
		GBM	-1.198*** (0.094)	[-1.382, -1.013]	-1.0*** (0.165)	[-1.324, -0.676]
		RF	-1.31*** (0.077)	[-1.46, -1.159]	-1.135*** (0.101)	[-1.333, -0.938]
		MLP	-1.547*** (0.079)	[-1.701, -1.393]	-0.941*** (0.099)	[-1.135, -0.747]
	DRDID	Linear	-1.408*** (0.076)	[-1.556, -1.26]	-1.095*** (0.099)	[-1.288, -0.902]
		Lasso	-1.434*** (0.08)	[-1.59, -1.277]	-1.195*** (0.11)	[-1.41, -0.979]
		GBM	-1.241*** (0.068)	[-1.373, -1.108]	-1.035*** (0.087)	[-1.205, -0.865]
		RF	-1.319*** (0.067)	[-1.449, -1.188]	-1.112*** (0.088)	[-1.284, -0.94]
		MLP	-1.476*** (0.07)	[-1.614, -1.339]	-1.366*** (0.094)	[-1.55, -1.182]
	DML_DRDID	Linear	-1.421*** (0.078)	[-1.575, -1.268]	-1.044*** (0.109)	[-1.258, -0.83]
		Lasso	-1.422*** (0.087)	[-1.593, -1.251]	-1.223*** (0.133)	[-1.483, -0.963]
		GBM	-1.256*** (0.102)	[-1.457, -1.056]	-1.001*** (0.214)	[-1.421, -0.581]
		RF	-1.33*** (0.077)	[-1.481, -1.18]	-1.172*** (0.1)	[-1.367, -0.976]
		MLP	-1.535*** (0.106)	[-1.743, -1.327]	-1.548*** (0.126)	[-1.796, -1.301]

Table S 2. (continue)

Outcome	Estimator	Model	All workers (N=4,733)		Only commuters (N=2,374)	
			ATT (S.E.)	95% confidence interval	ATT (S.E.)	95% confidence interval
Commuting time to work (minutes)	TWFE	N.A.	-7.99*** (1.003)	[-9.956, -6.024]	-2.513 (1.551)	[-5.553, 0.527]
	SemiDiD	N.A.	-3.663*** (0.87)	[-5.368, -1.958]	-0.1 (1.277)	[-2.604, 2.405]
	DMLDiD	Linear	-5.062*** (0.854)	[-6.737, -3.387]	-1.535 (1.195)	[-3.878, 0.807]
		Lasso	-5.838*** (0.961)	[-7.722, -3.954]	-2.508* (1.444)	[-5.34, 0.324]
		GBM	-4.732*** (1.029)	[-6.749, -2.715]	3.377 (4.025)	[-4.516, 11.27]
		RF	-4.123*** (0.86)	[-5.809, -2.438]	-2.11* (1.232)	[-4.527, 0.306]
		MLP	-6.995*** (0.908)	[-8.775, -5.215]	4.72*** (1.222)	[2.324, 7.116]
	DRDID	Linear	-4.893*** (0.832)	[-6.525, -3.262]	-1.827 (1.165)	[-4.113, 0.458]
		Lasso	-5.548*** (0.866)	[-7.246, -3.851]	-2.539** (1.272)	[-5.032, -0.046]
		GBM	-4.082*** (0.717)	[-5.488, -2.676]	-1.698 (1.078)	[-3.813, 0.417]
		RF	-3.691*** (0.731)	[-5.124, -2.259]	-1.534 (1.081)	[-3.653, 0.585]
		MLP	-7.915*** (0.86)	[-9.6, -6.23]	-2.128* (1.133)	[-4.35, 0.094]
	DML_DRDID	Linear	-5.235*** (0.846)	[-6.893, -3.577]	-1.082 (1.234)	[-3.501, 1.337]
		Lasso	-6.248*** (1.026)	[-8.259, -4.236]	-3.063** (1.47)	[-5.946, -0.18]
		GBM	-3.966*** (1.066)	[-6.055, -1.877]	-0.231 (1.819)	[-3.798, 3.336]
		RF	-4.0*** (0.856)	[-5.679, -2.32]	-1.538 (1.208)	[-3.906, 0.83]
		MLP	-5.944*** (0.912)	[-7.732, -4.156]	2.662 (1.815)	[-0.896, 6.221]

Table S 2. (continue)

Outcome	Estimator	Model	All workers (N=4,733)		Only commuters (N=2,374)	
			ATT (S.E.)	95% confidence interval	ATT (S.E.)	95% confidence interval
Private car as commuting mode	TWFE	N.A.	-0.24*** (0.021)	[-0.281, -0.198]	-0.002 (0.022)	[-0.045, 0.041]
	SemiDiD	N.A.	-0.175*** (0.02)	[-0.214, -0.137]	-0.029 (0.019)	[-0.067, 0.01]
	DMLDiD	Linear	-0.166*** (0.02)	[-0.204, -0.127]	0.011 (0.019)	[-0.027, 0.049]
		Lasso	-0.172*** (0.021)	[-0.214, -0.13]	-0.003 (0.022)	[-0.047, 0.041]
		GBM	-0.161*** (0.025)	[-0.209, -0.112]	-0.037 (0.031)	[-0.097, 0.024]
		RF	-0.157*** (0.019)	[-0.194, -0.119]	0.012 (0.018)	[-0.024, 0.048]
		MLP	-0.205*** (0.022)	[-0.248, -0.162]	-0.016 (0.052)	[-0.118, 0.087]
	DRDID	Linear	-0.168*** (0.019)	[-0.205, -0.13]	0.004 (0.017)	[-0.03, 0.039]
		Lasso	-0.171*** (0.02)	[-0.21, -0.133]	-0.001 (0.019)	[-0.04, 0.037]
		GBM	-0.137*** (0.018)	[-0.171, -0.102]	0.004 (0.017)	[-0.028, 0.037]
		RF	-0.147*** (0.017)	[-0.181, -0.113]	0.005 (0.016)	[-0.026, 0.037]
		MLP	-0.24*** (0.019)	[-0.276, -0.203]	-0.005 (0.017)	[-0.038, 0.028]
	DML_DRDID	Linear	-0.167*** (0.019)	[-0.205, -0.129]	0.005 (0.019)	[-0.032, 0.042]
		Lasso	-0.172*** (0.021)	[-0.213, -0.131]	-0.003 (0.022)	[-0.046, 0.041]
		GBM	-0.126*** (0.024)	[-0.173, -0.078]	0.038 (0.025)	[-0.011, 0.087]
		RF	-0.151*** (0.019)	[-0.189, -0.113]	0.012 (0.017)	[-0.022, 0.046]
		MLP	0.268*** (0.058)	[0.155, 0.382]	-0.068** (0.033)	[-0.133, -0.004]

Table S 2. (continue)

Outcome	Estimator	Model	All workers (N=4,733)		Only commuters (N=2,374)	
			ATT (S.E.)	95% confidence interval	ATT (S.E.)	95% confidence interval
Transit as commuting mode	TWFE	N.A.	-0.045*** (0.011)	[-0.067, -0.022]	-0.034** (0.016)	[-0.065, -0.003]
	SemiDiD	N.A.	-0.004 (0.012)	[-0.028, 0.02]	-0.013 (0.016)	[-0.044, 0.018]
	DMLDiD	Linear	-0.019 (0.012)	[-0.043, 0.005]	-0.032** (0.014)	[-0.06, -0.004]
		Lasso	-0.024* (0.013)	[-0.049, 0.002]	-0.032* (0.018)	[-0.067, 0.002]
		GBM	-0.016 (0.014)	[-0.044, 0.012]	-0.02 (0.019)	[-0.058, 0.018]
		RF	-0.017 (0.012)	[-0.04, 0.006]	-0.039*** (0.014)	[-0.067, -0.011]
		MLP	-0.028** (0.012)	[-0.052, -0.005]	-0.035 (0.027)	[-0.087, 0.018]
	DRDID	Linear	-0.017 (0.012)	[-0.04, 0.007]	-0.028** (0.014)	[-0.056, -0.0]
		Lasso	-0.019 (0.012)	[-0.043, 0.005]	-0.032** (0.016)	[-0.063, -0.001]
		GBM	-0.019* (0.011)	[-0.04, 0.001]	-0.026* (0.014)	[-0.053, 0.0]
		RF	-0.013 (0.01)	[-0.033, 0.008]	-0.029** (0.013)	[-0.055, -0.003]
		MLP	-0.044*** (0.011)	[-0.067, -0.022]	-0.03** (0.013)	[-0.057, -0.004]
	DML_DRDID	Linear	-0.018 (0.012)	[-0.042, 0.006]	40.155 (38.917)	[-36.159, 116.469]
		Lasso	-0.018 (0.013)	[-0.044, 0.007]	-0.034* (0.018)	[-0.069, 0.001]
		GBM	-0.016 (0.019)	[-0.054, 0.021]	-0.035* (0.018)	[-0.071, 0.001]
		RF	-0.019 (0.012)	[-0.043, 0.004]	-0.033** (0.014)	[-0.062, -0.005]
		MLP	-0.029** (0.014)	[-0.057, -0.002]	-0.022 (0.015)	[-0.051, 0.007]

Table S 2. (continue)

Outcome	Estimator	Model	All workers (N=4,733)		Only commuters (N=2,374)	
			ATT (S.E.)	95% confidence interval	ATT (S.E.)	95% confidence interval
Bicycle as commuting mode	TWFE	N.A.	0.007 (0.007)	[-0.007, 0.022]	0.003 (0.012)	[-0.021, 0.027]
	SemiDiD	N.A.	0.019*** (0.007)	[0.005, 0.032]	0.001 (0.011)	[-0.02, 0.023]
	DMLDiD	Linear	0.019*** (0.007)	[0.006, 0.033]	-0.014 (0.011)	[-0.035, 0.007]
		Lasso	0.014* (0.007)	[-0.0, 0.029]	0.001 (0.013)	[-0.023, 0.026]
		GBM	0.025*** (0.009)	[0.006, 0.043]	-0.005 (0.017)	[-0.038, 0.027]
		RF	0.012** (0.006)	[0.0, 0.025]	0.0 (0.01)	[-0.019, 0.019]
		MLP	-0.005 (0.009)	[-0.024, 0.013]	2.11*** (0.13)	[1.855, 2.364]
	DRDID	Linear	0.019*** (0.007)	[0.006, 0.032]	-0.004 (0.009)	[-0.022, 0.014]
		Lasso	0.011 (0.007)	[-0.003, 0.024]	0.003 (0.011)	[-0.019, 0.024]
		GBM	0.015*** (0.005)	[0.005, 0.026]	-0.002 (0.008)	[-0.019, 0.014]
		RF	0.014*** (0.005)	[0.005, 0.024]	-0.005 (0.008)	[-0.02, 0.01]
		MLP	0.005 (0.006)	[-0.006, 0.017]	0.002 (0.008)	[-0.014, 0.018]
	DML_DRDID	Linear	0.012* (0.007)	[-0.001, 0.026]	-0.0 (0.011)	[-0.021, 0.021]
		Lasso	0.014* (0.007)	[-0.0, 0.029]	0.007 (0.014)	[-0.02, 0.033]
		GBM	0.011 (0.008)	[-0.005, 0.027]	0.022 (0.013)	[-0.004, 0.048]
		RF	0.013** (0.006)	[0.001, 0.026]	0.002 (0.011)	[-0.018, 0.023]
		MLP	-0.026** (0.012)	[-0.049, -0.003]	-0.009 (0.015)	[-0.038, 0.019]

Table S 2. (continue)

Outcome	Estimator	Model	All workers (N=4,733)		Only commuters (N=2,374)	
			ATT (S.E.)	95% confidence interval	ATT (S.E.)	95% confidence interval
Walk as commuting mode	TWFE	N.A.	-0.0 (0.007)	[-0.014, 0.014]	0.013 (0.013)	[-0.012, 0.037]
	SemiDiD	N.A.	0.011 (0.007)	[-0.002, 0.024]	0.019* (0.011)	[-0.003, 0.041]
	DMLDiD	Linear	0.001 (0.006)	[-0.011, 0.013]	-0.001 (0.018)	[-0.037, 0.035]
		Lasso	0.004 (0.007)	[-0.01, 0.019]	0.015 (0.013)	[-0.011, 0.041]
		GBM	0.004 (0.008)	[-0.012, 0.021]	0.015 (0.019)	[-0.022, 0.053]
		RF	0.003 (0.006)	[-0.008, 0.014]	0.011 (0.01)	[-0.009, 0.032]
		MLP	0.001 (0.006)	[-0.011, 0.013]	0.01 (0.023)	[-0.036, 0.057]
	DRDID	Linear	0.004 (0.006)	[-0.009, 0.016]	0.013 (0.009)	[-0.005, 0.031]
		Lasso	0.004 (0.007)	[-0.009, 0.018]	0.015 (0.011)	[-0.008, 0.037]
		GBM	0.008 (0.005)	[-0.002, 0.017]	0.015* (0.009)	[-0.002, 0.033]
		RF	0.005 (0.005)	[-0.005, 0.014]	0.014 (0.008)	[-0.003, 0.03]
		MLP	0.0 (0.006)	[-0.011, 0.012]	1.011*** (0.087)	[0.841, 1.182]
	DML_DRDID	Linear	0.005 (0.006)	[-0.008, 0.017]	0.024** (0.01)	[0.004, 0.043]
		Lasso	0.004 (0.007)	[-0.01, 0.019]	0.015 (0.014)	[-0.012, 0.042]
		GBM	0.011* (0.006)	[-0.001, 0.024]	0.025** (0.012)	[0.003, 0.048]
		RF	0.002 (0.006)	[-0.009, 0.014]	0.018* (0.01)	[-0.002, 0.037]
		MLP	0.002 (0.007)	[-0.011, 0.015]	0.196*** (0.024)	[0.149, 0.242]

S.E.: standard error

* significant at 0.1 level

** significant at 0.05 level

*** significant at 0.01 level

Table S 3. Full variables included in the repeated cross-section empirical case study

Variable	Original variable	Code/range	Original code in NHTS 2009	Original code in NHTS 2017	Note
Treatment					
Living in Dallas-Fort Worth-Arlington	HH_CBSA	0: No 1: Yes	12420 = Austin-Round Rock, TX 19100 = Dallas-Fort Worth-Arlington, TX 26420 = Houston-Sugar Land-Baytown, TX 41700 = San Antonio, TX	12420 = Austin-Round Rock, TX 19100 = Dallas-Fort Worth-Arlington, TX 26420 = Houston-The Woodlands-Sugar Land, TX 41700 = San Antonio-New Braunfels, TX	Aggregated to binary Non-Texan area code not listed
Covariates					
Household size	HHSIZE	1-12	1-14	1-13	
Household vehicle ownership	HHVEHCNT	0-10	0-15	0-12	
Home ownership	HOMEOWN	0: No 1: Yes	01 = Own 02 = Rent	01 = Own 02 = Rent	Recoded
No. of workers	WRKCOUNT	0-6	0-6	0-7	
Census tract-level population density	HTPPOPDN	50 300 750 1500 3000 7000	50 = 0-99 300 = 100-499 750 = 500-999 1500 = 1,000-1,999 3000 = 2,000-3,999 7000 = 4,000-9,999	50 = 0-99 300 = 100-499 750 = 500-999 1500 = 1,000-1,999 3000 = 2,000-3,999 7000 = 4,000-9,999	Per square mile
Census tract-level housing unit	HTRESDN				

Variable	Original variable	Code/range	Original code in NHTS 2009	Original code in NHTS 2017	Note
density		17000 30000	17000 = 10,000-24,999 30000 = 25,000-999,999	17000 = 10,000-24,999 30000 = 25,000-999,999	
Home-work distance	GCDWORK	0-9011.39	0-3899	0-9744.49	In miles
Household living in urban area	URBRUR	0: No 1: Yes	01 = Urban 02 = Rural	01 = Urban 02 = Rural	Recoded
Born in the U.S.	BORNINUS	0: No 1: Yes	01 = Yes 02 = No	01 = Yes 02 = No	Recoded
Driver status	DRIVER	0: No 1: Yes	01 = Yes, a driver 02 = No, not a driver	01 = Yes 02 = No	Recoded
Age	R_AGE	5-92	5-92	5-92	
Gender	R_SEX	0: Female 1: Male	01 = Male 02 = Female	01 = Male 02 = Female	Recoded
Worker status	WORKER	0: No 1: Yes	01 = Yes 02 = No	01 = Yes 02 = No	Recoded
Student status	SCHTYP	0: No 1: Yes	01 = Public 02 = Private 03 = Home schooled 04 = Not in school	01 = Public or private school 02 = Home schooled 03 = Not in school	Aggregated into binary
Household income categories	HHFAMINC	0: Not in this category 1: In this category	01 = < $5,000 02 = $5,000 - $9,999 03 = $10,000 - $14,999 04 = $15,000 - $19,999 05 = $20,000 - $24,999	01 = Less than $10,000 02 = $10,000 to $14,999 03 = $15,000 to $24,999 04 = $25,000 to $34,999 05 = $35,000 to $49,999	Adjusted to 2017 values & one-hot encoded into ascending binaries with 'Less than $10,000' as reference

Variable	Original variable	Code/range	Original code in NHTS 2009	Original code in NHTS 2017	Note
			06 = $25,000 - $29,999 07 = $30,000 - $34,999 08 = $35,000 - $39,999 09 = $40,000 - $44,999 10 = $45,000 - $49,999 11 = $50,000 - $54,999 12 = $55,000 - $59,999 13 = $60,000 - $64,999 14 = $65,000 - $69,999 15 = $70,000 - $74,999 16 = $75,000 - $79,999 17 = $80,000 - $99,999 18 = > = $100,000	06 = $50,000 to $74,999 07 = $75,000 to $99,999 08 = $100,000 to $124,999 09 = $125,000 to $149,999 10 = $150,000 to $199,999 11 = $200,000 or more	
Education attainment categories	EDUC	0: Not in this category 1: In this category	01 = Less than high school graduate 02 = High school graduate, include GED 03 = Some college or Associate's degree (Vocational) 04 = Bachelor's degree (BA, AB, BS) 05 = Graduate or Professional Degree	01 = Less than a high school graduate 02 = High school graduate or GED 03 = Some college or associates degree 04 = Bachelor's degree 05 = Graduate degree or professional degree	One-hot encoded into ascending binaries with ‘Less than a high school graduate’ as reference

Variable	Original variable	Code/range	Original code in NHTS 2009	Original code in NHTS 2017	Note
Job categories	OCCAT	0: Not in this category 1: In this category	01 = Sales / service 02 = Clerical / admin support 03 = Manuf, construct, maintenance, or farming 04 = Professional, managerial, or technical 97 = Other	01 = Sales or service 02 = Clerical or administrative support 03 = Manufacturing, construction, maintenance, or farming 04 = Professional, managerial, or technical 97 = Something else	One-hot encoded into ascending binaries with 'Sales or service' as reference
Life cycle categories	LIF_CYC	0: Not in this category 1: In this category	01 = one adult, no children 02 = 2+ adults, no children 03 = one adult, youngest child 0-5 04 = 2+ adults, youngest child 0-5 05 = one adult, youngest child 6-15 06 = 2+ adults, youngest child 6-15 07 = one adult, youngest child 16-21 08 = 2+ adults, youngest child 16-21 09 = one adult, retired, no	01 = one adult, no children 02 = 2+ adults, no children 03 = one adult, youngest child 0-5 04 = 2+ adults, youngest child 0-5 05 = one adult, youngest child 6-15 06 = 2+ adults, youngest child 6-15 07 = one adult, youngest child 16-21 08 = 2+ adults, youngest child 16-21 09 = one adult, retired, no	One-hot encoded into ascending binaries with 'one adult, no children' as reference

Variable	Original variable	Code/range	Original code in NHTS 2009	Original code in NHTS 2017	Note
			children 10 = 2+ adults, retired, no children	children 10 = 2+ adults, retired, no children	
Vehicle fuel type categories	FUELTYPE	0: Not in this category 1: In this category	1 = Diesel 2 = Natural Gas 3 = Electricity 4 = Motor Gasoline	01 = Gas 02 = Diesel 03 = Hybrid, electric or alternative fuel 97 = Some other fuel	Harmonized & one-hot encoded into ascending binaries with 'Motor Gasoline' as reference
Vehicle type categories	VEHTYPE	0: Not in this category 1: In this category	01 = Automobile/car/station wagon 02 = Van (mini, cargo, passenger) 03 = Sports utility vehicle 04 = Pickup truck 05 = Other truck 06 = RV (recreational vehicle) 07 = Motorcycle 08 = Golf cart 97 = Other	01 = Automobile/Car/Station Wagon 02 = Van (Mini/Cargo/Passenger) 03 = SUV (Santa Fe, Tahoe, Jeep, etc.) 04 = Pickup Truck 05 = Other Truck 06 = RV (Recreational Vehicle) 07 = Motorcycle/Motorbike 97 = Something Else	Harmonized & one-hot encoded into ascending binaries with 'Automobile/Car/Station Wagon' as reference
Outcomes					
No. of trips	CNTTDTR	0-26	0-27	0-50	
No. of motorcycle trips	MCUSED	0-99	0-99	0-99	Outlier detected
No. of bike trips	NBIKETRP	0-99	0-99	0-99	

Variable	Original variable	Code/range	Original code in NHTS 2009	Original code in NHTS 2017	Note
No. of public transit trips	PTUSED	0-99	0-180	0-30	
Commuting time	TIMETOWK	0-600	0-660	0-600	In minutes
Mileage last 12 months	YEARMILE	0-200000	0-200000	0-200000	In miles

Note: original codes include non-Texan samples

Table S 4. Full results of the repeated cross-section empirical case study

Outcome	Estimator	Model	ATT (S.E.)	95% confidence interval
No. of trips made	TWFE	N.A.	-0.045 (0.041)	[-0.125, 0.035]
	SemiDiD	N.A.	0.979*** (0.073)	[0.837, 1.121]
	DMLDiD	Linear	0.707*** (0.069)	[0.572, 0.842]
		Lasso	0.918*** (0.073)	[0.775, 1.061]
		GBM	0.747*** (0.068)	[0.613, 0.881]
		RF	0.734*** (0.067)	[0.602, 0.867]
		MLP	0.959*** (0.072)	[0.819, 1.1]
	DRDID	Linear	-0.058 (0.041)	[-0.139, 0.022]
		Lasso	-0.051 (0.043)	[-0.135, 0.033]
		GBM	-0.015 (0.039)	[-0.093, 0.062]
		RF	-0.033 (0.039)	[-0.109, 0.042]
		MLP	-0.06 (0.043)	[-0.145, 0.024]
	DML_DRDID	Linear	-0.05 (0.041)	[-0.131, 0.031]
		Lasso	-0.069 (0.043)	[-0.152, 0.015]
		GBM	-0.042 (0.043)	[-0.126, 0.042]
		RF	-0.018 (0.042)	[-0.1, 0.064]
		MLP	-0.06 (0.043)	[-0.144, 0.025]

Table S 4. (continue)

Outcome	Estimator	Model	ATT (S.E.)	95% confidence interval
No. of motorcycle trips past month	TWFE	N.A.	-0.018 (0.09)	[-0.195, 0.159]
	SemiDiD	N.A.	0.016 (0.101)	[-0.182, 0.214]
	DMLDiD	Linear	0.028 (0.107)	[-0.181, 0.237]
		Lasso	0.009 (0.103)	[-0.193, 0.212]
		GBM	0.05 (0.1)	[-0.147, 0.247]
		RF	0.02 (0.102)	[-0.179, 0.22]
		MLP	0.014 (0.104)	[-0.189, 0.217]
	DRDID	Linear	0.127 (0.101)	[-0.07, 0.324]
		Lasso	-0.036 (0.102)	[-0.237, 0.165]
		GBM	0.006 (0.083)	[-0.157, 0.169]
		RF	0.21** (0.085)	[0.044, 0.377]
		MLP	-0.033 (0.106)	[-0.24, 0.174]
	DML_DRDID	Linear	0.095 (0.101)	[-0.102, 0.293]
		Lasso	-0.025 (0.102)	[-0.225, 0.176]
		GBM	-0.006 (0.105)	[-0.212, 0.2]
		RF	0.099 (0.119)	[-0.133, 0.331]
		MLP	-0.025 (0.096)	[-0.212, 0.163]

Table S 4. (continue)

Outcome	Estimator	Model	ATT (S.E.)	95% confidence interval
No. of bicycle trips past week	TWFE	N.A.	0.041 (0.027)	[-0.012, 0.093]
	SemiDiD	N.A.	0.141*** (0.03)	[0.083, 0.199]
	DMLDiD	Linear	0.048* (0.027)	[-0.005, 0.101]
		Lasso	0.118*** (0.03)	[0.061, 0.176]
		GBM	0.055** (0.027)	[0.001, 0.108]
		RF	0.055** (0.027)	[0.002, 0.108]
		MLP	0.128*** (0.028)	[0.073, 0.183]
	DRDID	Linear	0.022 (0.027)	[-0.031, 0.075]
		Lasso	0.051* (0.027)	[-0.002, 0.105]
		GBM	0.008 (0.023)	[-0.037, 0.053]
		RF	-0.051** (0.021)	[-0.091, -0.011]
		MLP	0.026 (0.027)	[-0.028, 0.079]
	DML_DRDID	Linear	0.03 (0.026)	[-0.021, 0.081]
		Lasso	0.054** (0.027)	[0.001, 0.107]
		GBM	-0.007 (0.026)	[-0.059, 0.044]
		RF	-0.011 (0.027)	[-0.064, 0.042]
		MLP	0.029 (0.027)	[-0.024, 0.083]

Table S 4. (continue)

Outcome	Estimator	Model	ATT (S.E.)	95% confidence interval
No. of walk trips past week	TWFE	N.A.	0.076 (0.105)	[-0.131, 0.283]
	SemiDiD	N.A.	1.186*** (0.124)	[0.943, 1.429]
	DMLDiD	Linear	0.831*** (0.12)	[0.595, 1.066]
		Lasso	1.129*** (0.125)	[0.884, 1.374]
		GBM	0.842*** (0.12)	[0.607, 1.078]
		RF	0.876*** (0.117)	[0.647, 1.105]
		MLP	1.145*** (0.121)	[0.907, 1.383]
	DRDID	Linear	0.128 (0.103)	[-0.074, 0.33]
		Lasso	0.042 (0.104)	[-0.161, 0.246]
		GBM	0.095 (0.096)	[-0.094, 0.284]
		RF	0.015 (0.09)	[-0.162, 0.191]
		MLP	0.056 (0.104)	[-0.148, 0.259]
	DML_DRDID	Linear	0.139 (0.104)	[-0.065, 0.342]
		Lasso	0.039 (0.104)	[-0.164, 0.243]
		GBM	0.168 (0.109)	[-0.045, 0.381]
		RF	0.0 (0.103)	[-0.202, 0.202]
		MLP	0.049 (0.104)	[-0.154, 0.253]

Table S 4. (continue)

Outcome	Estimator	Model	ATT (S.E.)	95% confidence interval
No. of public transit trips past month	TWFE	N.A.	-0.17** (0.071)	[-0.309, -0.03]
	SemiDiD	N.A.	0.03 (0.079)	[-0.125, 0.184]
	DMLDiD	Linear	-0.018 (0.077)	[-0.17, 0.134]
		Lasso	0.046 (0.079)	[-0.109, 0.202]
		GBM	-0.004 (0.077)	[-0.155, 0.146]
		RF	-0.035 (0.076)	[-0.185, 0.114]
		MLP	0.053 (0.077)	[-0.098, 0.203]
	DRDID	Linear	-0.054 (0.076)	[-0.203, 0.094]
		Lasso	-0.117 (0.077)	[-0.268, 0.033]
		GBM	-0.161** (0.067)	[-0.293, -0.029]
		RF	-0.173*** (0.06)	[-0.291, -0.054]
		MLP	-0.163** (0.076)	[-0.312, -0.013]
	DML_DRDID	Linear	-0.059 (0.076)	[-0.209, 0.09]
		Lasso	-0.111 (0.076)	[-0.261, 0.039]
		GBM	-0.11 (0.078)	[-0.263, 0.042]
		RF	-0.174** (0.077)	[-0.326, -0.022]
		MLP	-0.133* (0.076)	[-0.283, 0.017]

Table S 4. (continue)

Outcome	Estimator	Model	ATT (S.E.)	95% confidence interval
Commuting time to work last week (minutes)	TWFE	N.A.	-0.122 (0.263)	[-0.638, 0.395]
	SemiDiD	N.A.	2.879*** (0.365)	[2.163, 3.594]
	DMLDiD	Linear	2.519*** (0.365)	[1.804, 3.235]
		Lasso	2.834*** (0.367)	[2.115, 3.553]
		GBM	2.656*** (0.371)	[1.929, 3.383]
		RF	2.414*** (0.359)	[1.711, 3.118]
		MLP	3.058*** (0.361)	[2.351, 3.765]
	DRDID	Linear	-0.392 (0.252)	[-0.886, 0.102]
		Lasso	-0.545** (0.258)	[-1.051, -0.039]
		GBM	-0.366** (0.173)	[-0.706, -0.026]
		RF	-0.444*** (0.152)	[-0.742, -0.146]
		MLP	-0.046 (0.316)	[-0.666, 0.574]
	DML_DRDID	Linear	-0.36 (0.257)	[-0.864, 0.143]
		Lasso	-0.508* (0.26)	[-1.017, 0.001]
		GBM	-0.431** (0.218)	[-0.858, -0.003]
		RF	-0.411* (0.21)	[-0.823, 0.0]
		MLP	-0.271 (0.311)	[-0.88, 0.338]

Table S 4. (continue)

Outcome	Estimator	Model	ATT (S.E.)	95% confidence interval
Driving mileage last 12 months (miles)	TWFE	N.A.	-348.608** (158.517)	[-659.301, -37.915]
	SemiDiD	N.A.	1817.404*** (219.054)	[1388.059, 2246.75]
	DMLDiD	Linear	1509.54*** (215.489)	[1087.181, 1931.899]
		Lasso	1668.14*** (220.769)	[1235.434, 2100.847]
		GBM	1518.888*** (219.532)	[1088.604, 1949.171]
		RF	1417.214*** (214.292)	[997.201, 1837.227]
		MLP	1801.592*** (215.953)	[1378.325, 2224.86]
	DRDID	Linear	-422.474*** (157.791)	[-731.743, -113.204]
		Lasso	-479.044*** (177.762)	[-827.458, -130.63]
		GBM	-406.874*** (142.664)	[-686.494, -127.253]
		RF	-421.922*** (136.618)	[-689.694, -154.15]
		MLP	-336.549* (180.949)	[-691.21, 18.112]
	DML_DRDID	Linear	-427.922*** (158.334)	[-738.257, -117.586]
		Lasso	-479.926*** (178.011)	[-828.827, -131.024]
		GBM	-443.734*** (165.401)	[-767.919, -119.548]
		RF	-457.496*** (162.552)	[-776.097, -138.894]
		MLP	-321.654* (179.382)	[-673.243, 29.935]

S.E.: standard error

* significant at 0.1 level

** significant at 0.05 level

*** significant at 0.01 level