The performance of estimators based on the propensity score

Martin Huber, Michael Lechner and Conny Wunsch

Swiss Institute for Empirical Economic Research

University of St.Gallen

Revised version: June 2012

Date this version has been printed: 19 November 2012

Supplemental material (paper published in the Journal of Econometrics, 2013)

Appendices B, C, D & E (internet appendices)

Address for correspondence: Martin Huber, Michael Lechner, Conny Wunsch, Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Varnbüelstrasse 14, CH-9000 St. Gallen, Switzerland, Michael.Lechner@unisg.ch, www.sew.unisg.ch/lechner.

Appendix B: Additional details on the Monte Carlo design

Appendix B.1Probit and Tobit results for the outcome equations

Table B.1: Probit and Tobit results for the outcome equations

Dependent variable:	Employment (Probit)				Earnings (OLS)			
	Participants		Nonparticipants		Participants		Nonparticipants	
Independent variables	ME	SE	ME	SE	Coeff.	SE	Coeff.	SE
Constant term	0.04	0.15	0.40	0.02	1451	576	-602	66
Age / 10	-0.17	0.19	-0.68	0.03	-154	305	1176	33
… squared / 1000	0.05	0.04	0.05	0.01	-3	60	-27	7
20 - 25 years old	-0.52	0.22	-0.24	0.04	-18	45	33	7
Woman	-0.03	0.03	-0.01	0.00	37	49	145	8
Not German	-0.02	0.02	0.01	0.00	-159	70	-203	9
Secondary degree	-0.09	0.03	-0.03	0.01	242	66	116	10
University entrance qualification	-0.05	0.04	-0.05	0.01	-157	70	-174	9
No vocational degree	0.07	0.03	0.06	0.01	-29	61	-32	9
At least one child in household	-0.03	0.04	-0.05	0.01	28	101	-72	16
Last occupation: Non-skilled worker	-0.01	0.03	-0.05	0.01	125	47	57	7
Last occupation: Salaried worker	0.00	0.05	-0.01	0.01	129	55	38	9
Last occupation: Part time	0.03	0.02	0.03	0.00	131	89	34	14
UI benefits: 0	0.04	0.03	0.01	0.01	-913	240	-408	32
> 650 EUR per month	0.05	0.04	0.02	0.01	656	600	426	101
Last 10 years before UE: share empl.	-0.54	0.12	-0.21	0.02	559	353	72	45
share unemployed	0.69	0.30	0.30	0.06	-144	112	-173	19
share in programme	0.35	0.21	0.27	0.03	-184	90	-402	12
Last y. before UE: share in minor empl	-0.07	0.06	-0.10	0.01	337	54	181	8
share part-time employed	-0.06	0.04	-0.18	0.01	168	50	138	7
share out-of-the labour force	0.10	0.03	0.07	0.01	-341	71	-180	9
Entering UE in 2000	0.06	0.02	0.05	0.00	37	50	-60	7
2001	-0.14	0.04	-0.07	0.01	-116	63	-353	8
2003	0.02	0.03	-0.04	0.00	12	65	-19	10
Share of population close to big city	-0.05	0.03	-0.20	0.00	-8	58	61	9
Health impairments	-0.03	0.03	-0.01	0.01	-273	98	-511	14
Never out of labour force	0.01	0.03	0.02	0.01	6	39	77	6
Part time in last 10 years	-0.33	0.05	-0.59	0.00	1342	620	-30	66
Never employed	0.01	0.02	0.03	0.00	705	98	872	15
Duration of last employment > 1 year	-0.03	0.05	-0.04	0.01	-100	383	-1849	39
Average earnings last 10 years when employed / 1000	0.41	0.18	0.14	0.02	66	89	130	13
Women x age / 10	-0.50	0.23	-0.15	0.03	729	353	-198	38
x squared / 1000	-0.04	0.05	0.00	0.01	-807	468	426	50
x no vocational degree	-0.06	0.04	-0.12	0.01	-34	94	110	12
x at least one child in household	-0.25	0.21	-0.16	0.03	-457	89	-365	14
x share minor employ. last year	-0.08	0.06	-0.06	0.01	-674	362	-258	49
x share OLF last year	-0.14	0.06	-0.09	0.01	-142	113	67	16
x av. earn in last 10 y. if employed	-0.09	0.05	-0.01	0.01	-198	123	-272	19
x entering UE in 2003	0.04	0.15	0.40	0.02	40	90	68	12
Number of observations	3266		114349		3266		114349	
(Pseudo) R^2	0.13		0.36		0.36		0.41	

Note: ME: Marginal effects estimated at the mean on the explanatory variables.

Appendix B.2: Wald tests

Table B.2: Wald tests

	Additional matching variables		Nonlinear and interaction terms	
	Test statistic (df)	p-value in %	Test statistic (df)	p-value in %
		Participation equation (Probit)		
Full sample	27 (2)	0	671 (10)	0
		Outcome equation employment (Probit)		
Participants	4 (2)	11	23 (10)	1
Nonparticipants	53 (2)	0	1944 (10)	0
		Outcome equation earnings (OLS)		
Participants	55 (2)	0	39 (10)	0
Nonparticipants	3441 (2)	0	5348 (10)	0

Note: Test statistic is distributed as $\chi^2(df)$ with df degrees of freedom.

Appendix C: Trimming, common support, and further descriptive statistics

Appendix C.1: Distribution of propensity scores for treated and control observations in the population

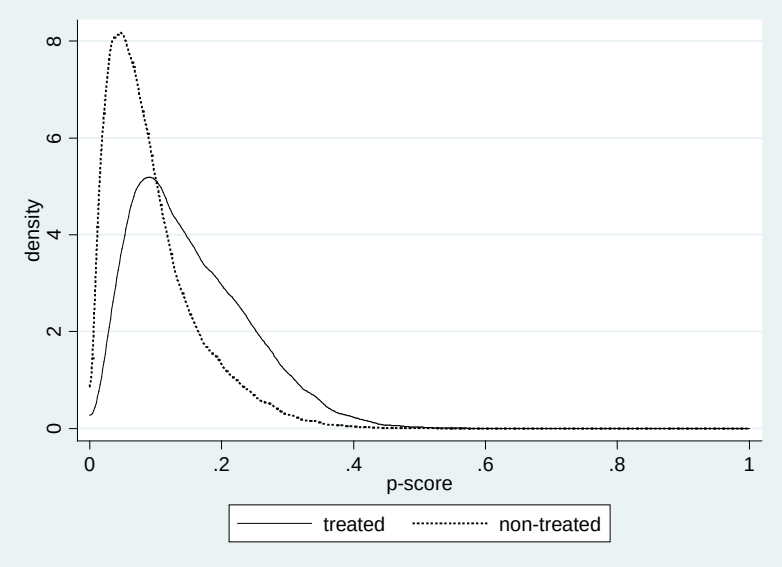
Table C.1: Number of control observations with propensity scores larger then p-score at particular quantile in treated population

DGP	Quantile of propensity score for treated				
	80 %	90 %	95 %	97.5 %	99 %
C10N	25366	10319	4404	1983	680
C10R	85903	42957	21455	10648	4077
C10S	4737	1516	507	178	41
C50N	16317	6485	2554	1018	348
C50R	49305	24626	12193	6165	2461
C50S	2092	497	116	33	6
C90N	2178	746	272	97	33
C90R	9704	4740	2377	1193	461
C90S	46	4	0	0	0
M10N	30457	12675	5567	2464	969
M10R	86767	43213	21755	10822	4239
M10S	8124	2817	995	432	141
M50N	19685	8136	3468	1523	546
M50R	49253	24469	12257	6184	2523
M50S	5214	1647	588	240	67
M90N	2980	1082	438	183	64
M90R	9631	4786	2348	1161	449
M90S	489	102	28	5	2

Note: Based on sample of size 500'000. DGP numbers are generated as follows: 1st digit: C=correctly specified p-score, M=misspecified p-score; 2nd and 3rd digit: % treated, 4th digit N=normal selection, R=random selection, S=strong selection. Information is independent of trimming level, because no trimming occurs for a sample size of 500'000.

Note that Table C.1 and the following figures are based on one draw of 500'000 in the control pool. The same simulation procedures that are applied to the sample sizes used in the Monte Carlos study are applied to this very large sample. The two selection processes with large shares of treated and strong selection lead to very thin tails in the control population. Nevertheless, the expected number of controls at each quantile of the treated is always positive (although very, very small for C22 leading to 0 observations for the 95% quantile of the treated in this particular draw; see also Figure C.9).

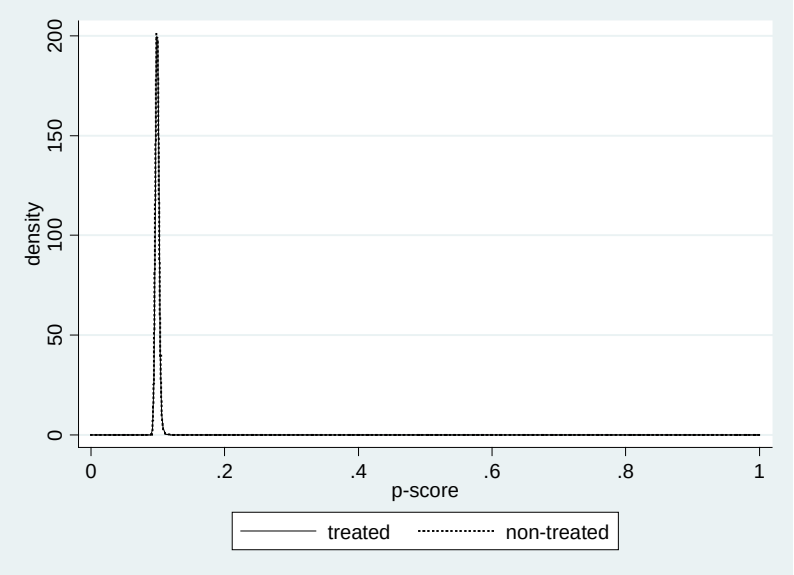
Figure C.1: Densities of propensity scores of treated and control observations:
4800 observations, correctly specified p-score, 10% treated, normal selection



Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.2: Densities of propensity scores of treated and control observations:

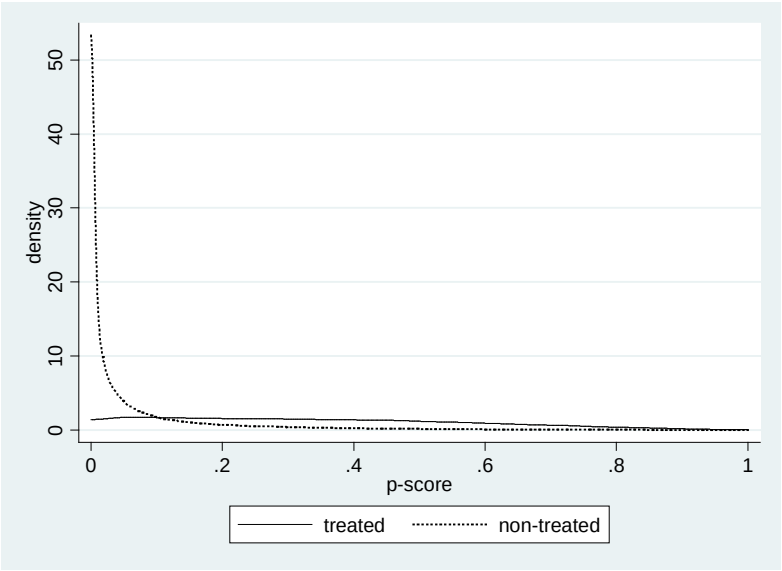
4800 observations, correctly specified p-score, 10% treated, random selection



Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

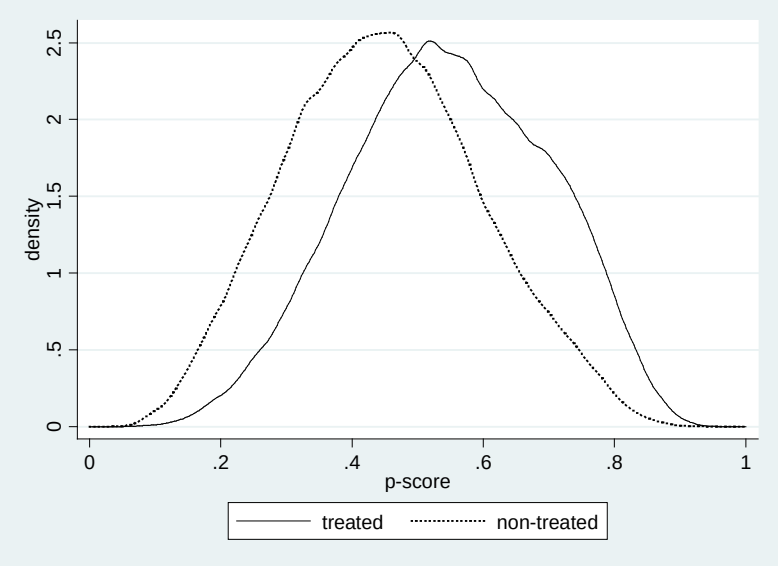
Figure C.3: Densities of propensity scores of treated and control observations:

4800 observations, correctly specified p-score, 10% treated, strong selection



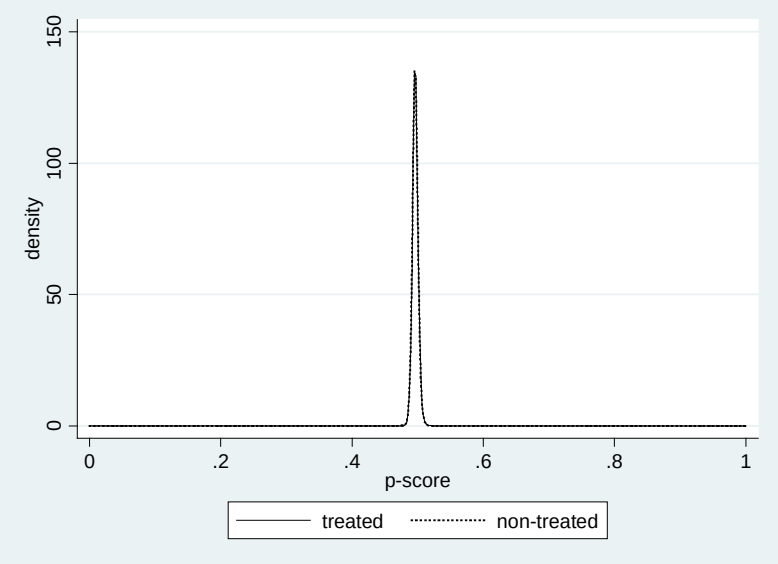
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.4: Densities of propensity scores of treated and control observations: 4800 observations, correctly specified p-score, 50% treated, normal selection



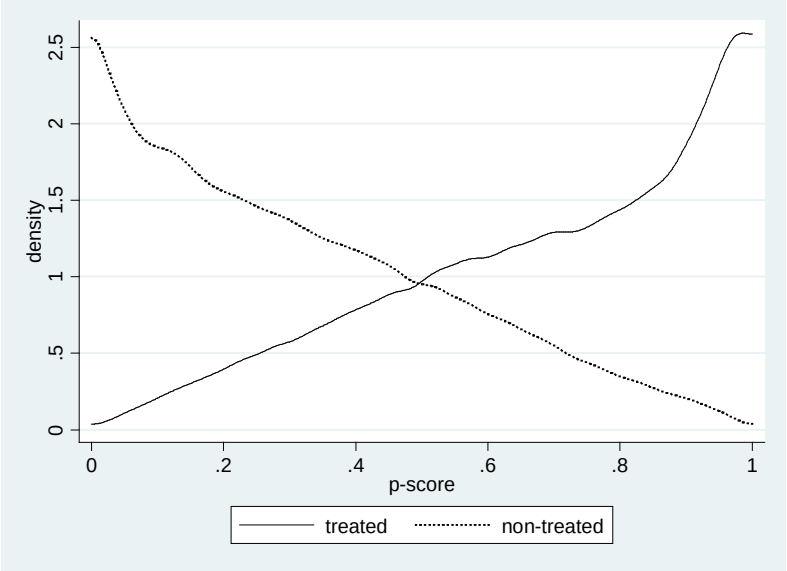
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.5: Densities of propensity scores of treated and control observations: 4800 observations, correctly specified p-score, 50% treated, random selection



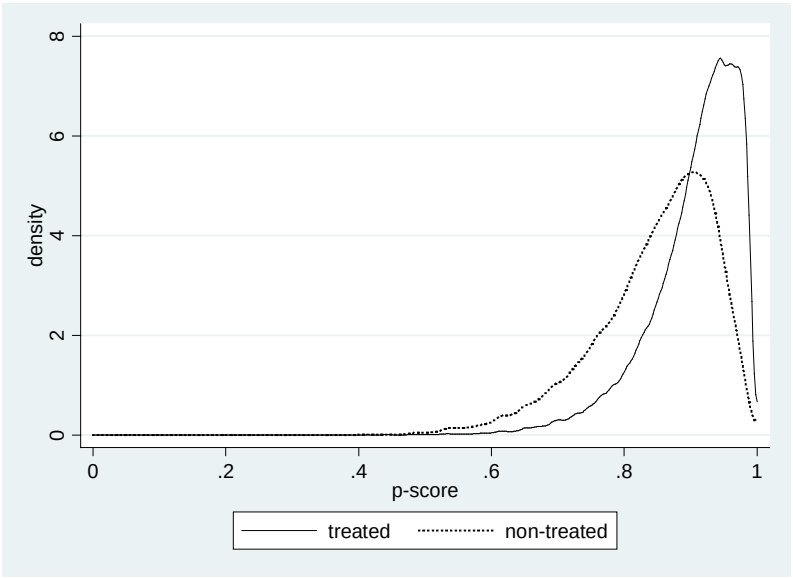
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.6: Densities of propensity scores of treated and control observations: 4800 observations, correctly specified p-score, 50% treated, strong selection



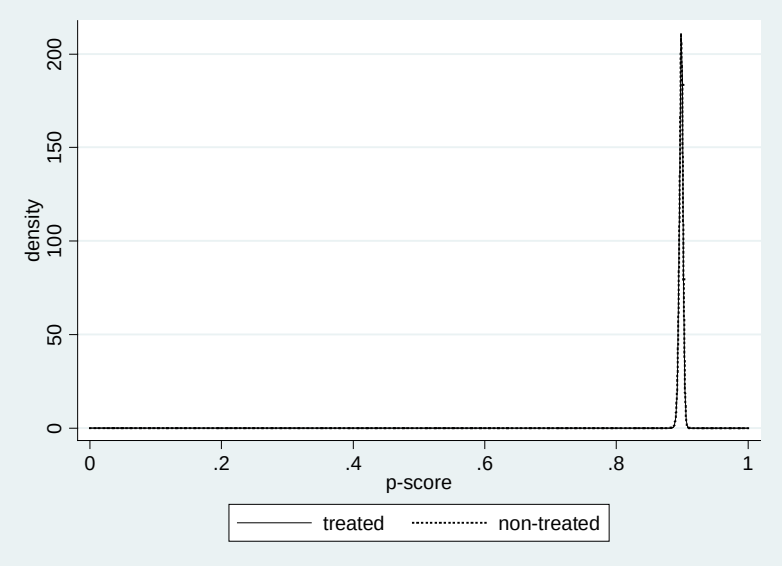
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.7: Densities of propensity scores of treated and control observations: 4800 observations, correctly specified p-score, 90% treated, normal selection



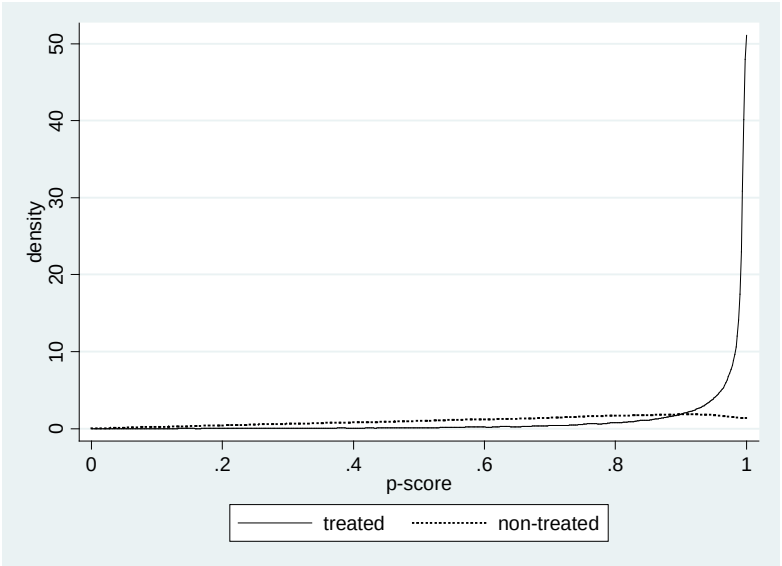
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.8: Densities of propensity scores of treated and control observations: 4800 observations, correctly specified p-score, 90% treated, random selection



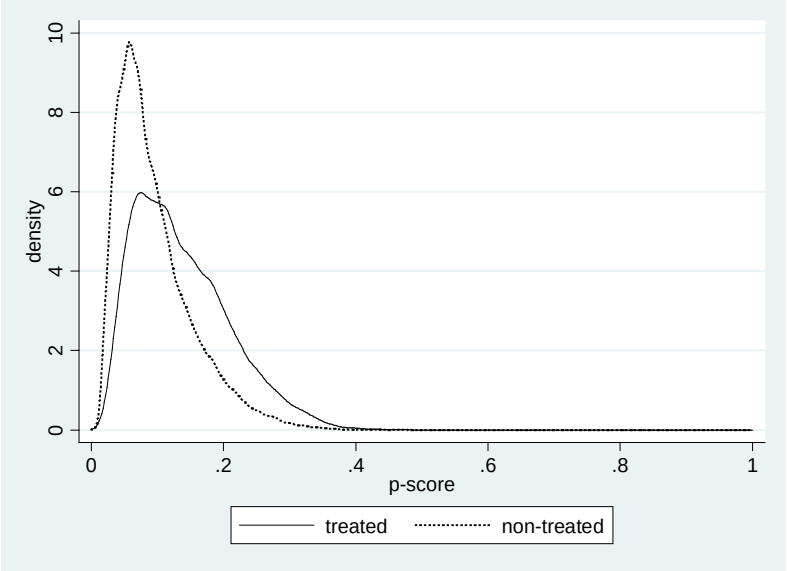
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.9: Densities of propensity scores of treated and control observations: 4800 observations, correctly specified p-score, 90% treated, strong selection



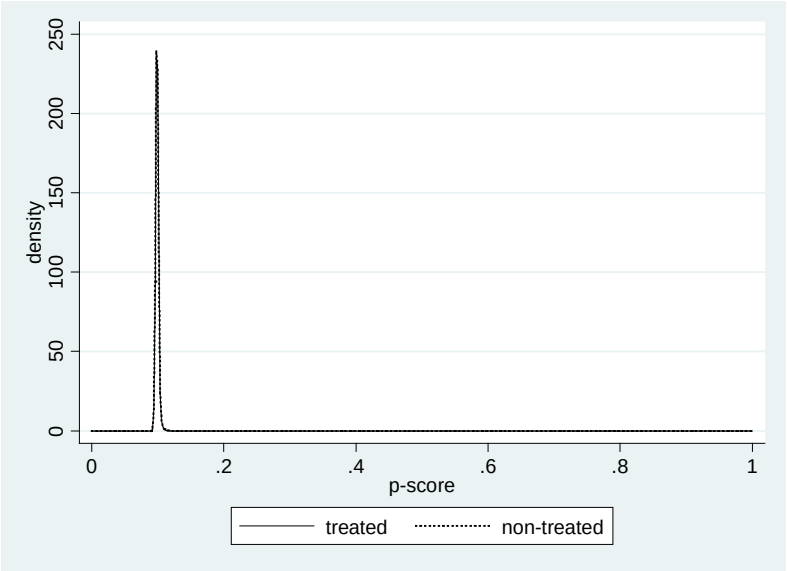
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.10: Densities of propensity scores of treated and control observations: 4800 observations, misspecified p-score, 10% treated, normal selection



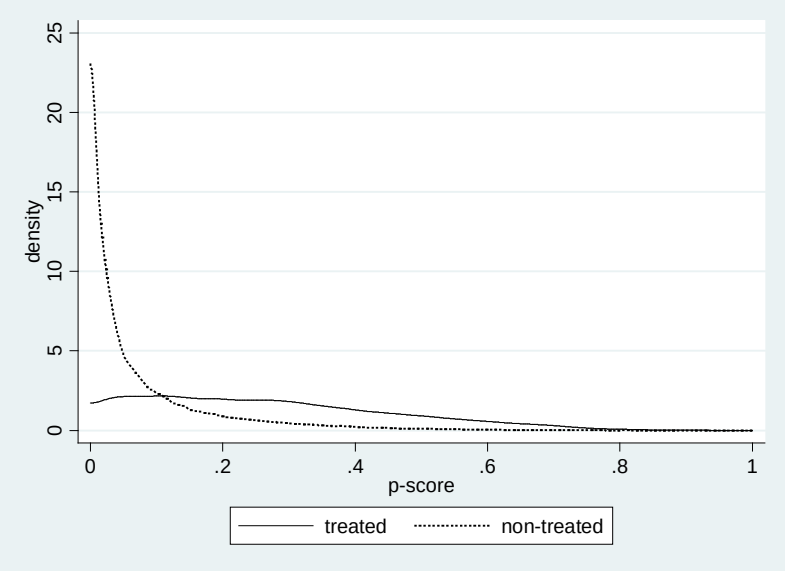
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.11: Densities of propensity scores of treated and control observations: 4800 observations, misspecified p-score, 10% treated, random selection



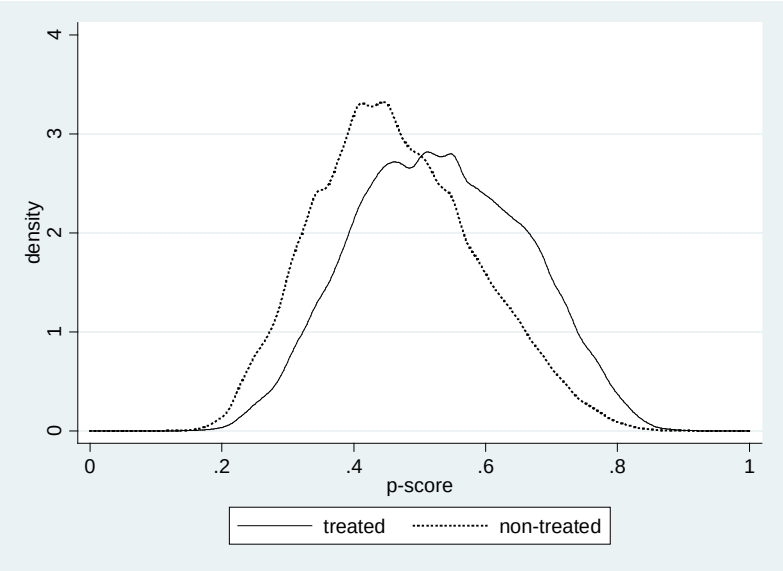
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.12: Densities of propensity scores of treated and control observations: 4800 observations, misspecified p-score, 10% treated, strong selection



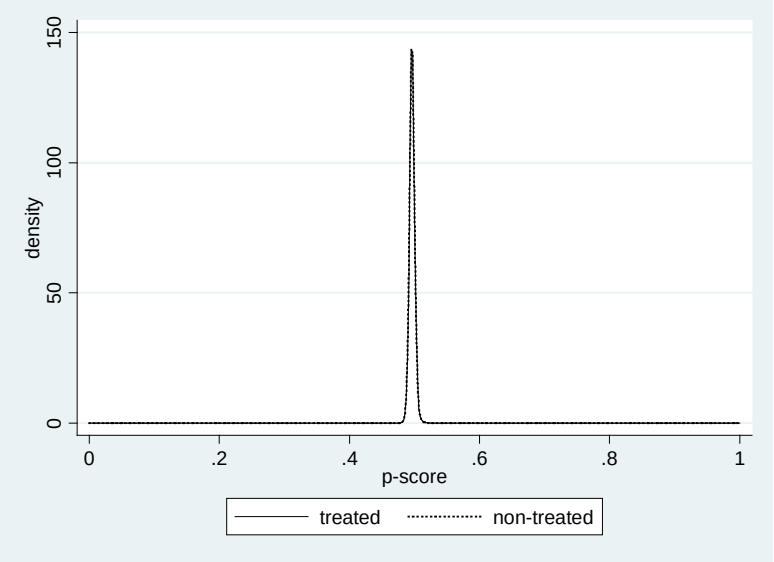
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.13: Densities of propensity scores of treated and control observations: 4800 observations, misspecified p-score, 50% treated, normal selection



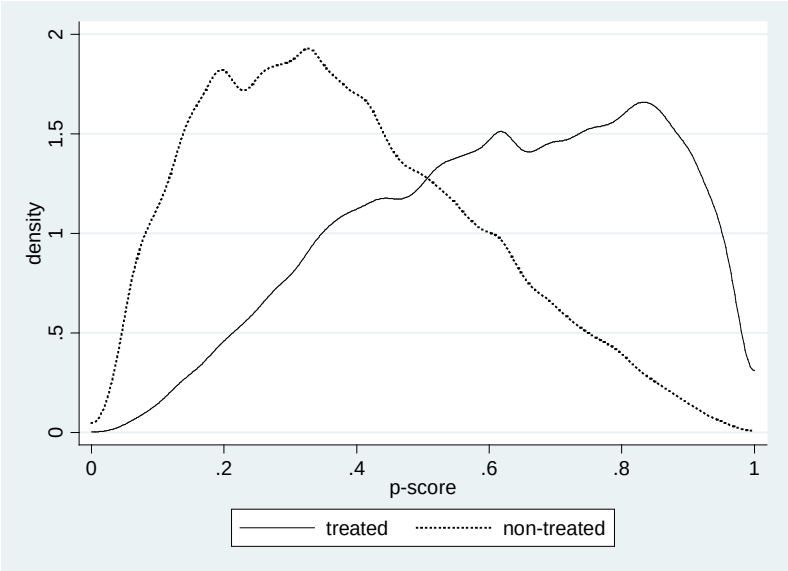
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.14: Densities of propensity scores of treated and control observations: 4800 observations, misspecified p-score, 50% treated, random selection



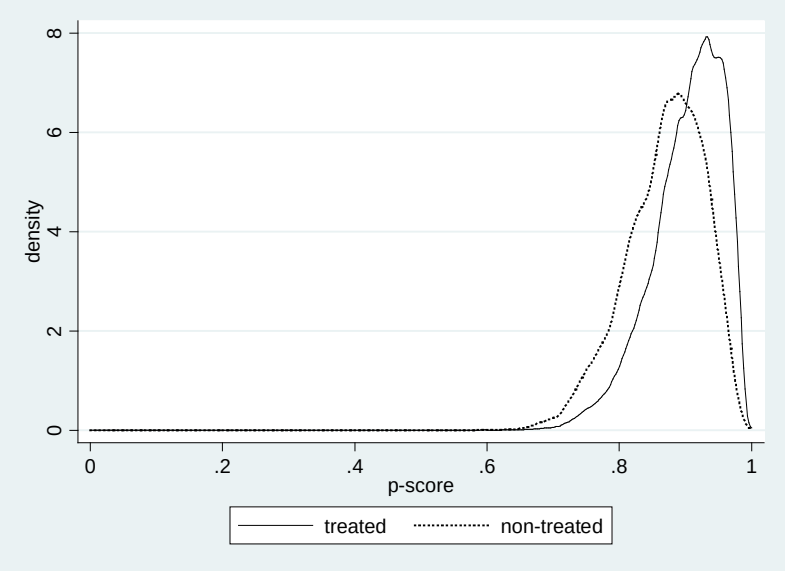
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.15: Densities of propensity scores of treated and control observations: 4800 observations, misspecified p-score, 50% treated, strong selection



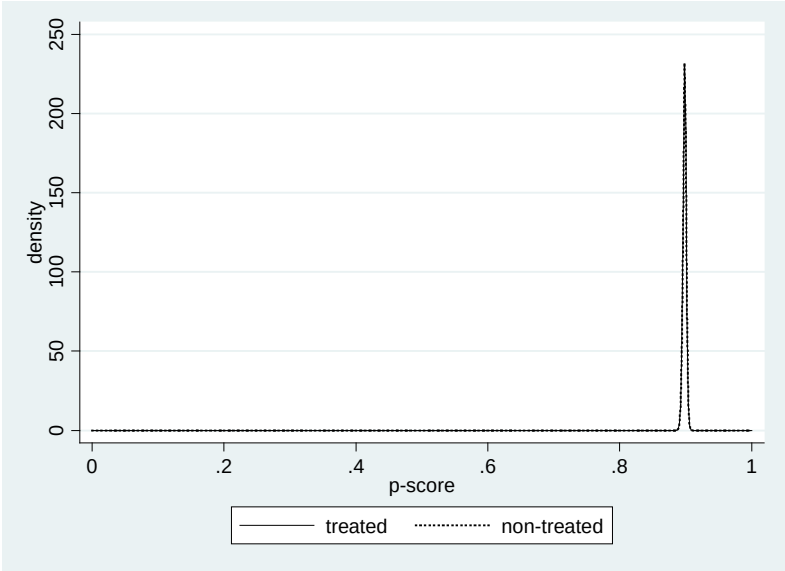
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.16: Densities of propensity scores of treated and control observations: 4800 observations, misspecified p-score, 90% treated, normal selection



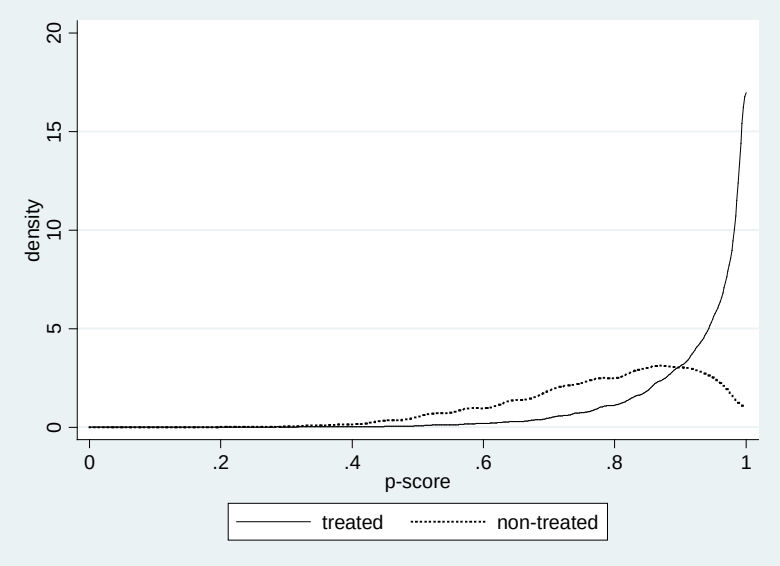
Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.17: Densities of propensity scores of treated and control observations: 4800 observations, misspecified p-score, 90% treated, random selection



Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Figure C.18: Densities of propensity scores of treated and control observations: 4800 observations, misspecified p-score, 90% treated, strong selection



Note: Drawn from a sample of size 500000. Densities are based on kernel density estimation with the Gaussian kernel (with the bandwidth chosen according to the rule of thumb) and boundary correction using the reflection method, see Silverman (1986).

Appendix C.2: Trimming and common support in the samples

Table C.2: Number of deleted non-treated observations for different levels of trimming and different DGP's

Data generating processes			Trimming levels						
Magnitude of selection	Share of treated in %	Sample size	4%	5%	6%	7%	8%	9%	10%
Correct specification of the propensity score									
Random	10	1200	-	-	-	-	-	-	-
		4800	-	-	-	-	-	-	-
	50	300	0.48	0.21	0.10	0.06	0.03	0.02	0.01
		1200	-	-	-	-	-	-	-
		4800	-	-	-	-	-	-	-
	90	1200	0.40	0.17	0.09	0.05	0.03	0.02	0.01
		4800	-	-	-	-	-	-	-
Observed	10	1200	0.00	-	-	-	-	-	-
		4800	-	-	-	-	-	-	-
	50	300	1.66	0.94	0.58	0.37	0.26	0.19	0.13
		1200	0.01	0.00	0.00	0.00	0.00	-	-
		4800	-	-	-	-	-	-	-
	90	1200	2.39	1.49	1.00	0.72	0.53	0.40	0.30
		4800	0.14	0.06	0.03	0.02	0.01	0.01	0.00
Strong	10	1200	0.73	0.41	0.25	0.17	0.11	0.08	0.06
		4800	0.02	0.01	0.01	0.00	0.00	0.00	-
	50	300	3.99	2.86	2.17	1.71	1.40	1.16	0.98
		1200	1.45	0.97	0.68	0.51	0.39	0.30	0.24
		4800	0.34	0.21	0.14	0.10	0.08	0.06	0.05
	90	1200	4.30	3.26	2.57	2.08	1.74	1.48	1.27
		4800	2.34	1.69	1.31	1.04	0.86	0.74	0.62
Functional misspecification of the propensity score									
Random	10	1200	-	-	-	-	-	-	-
		4800	-	-	-	-	-	-	-
	50	300	0.21	0.09	0.04	0.02	0.01	0.01	0.01
		1200	-	-	-	-	-	-	-
		4800	-	-	-	-	-	-	-
	90	1200	0.20	0.08	0.04	0.02	0.01	0.01	0.00
		4800	-	-	-	-	-	-	-
Observed	10	1200	-	-	-	-	-	-	-
		4800	-	-	-	-	-	-	-
	50	300	0.92	0.47	0.26	0.16	0.10	0.06	0.04
		1200	0.00	-	-	-	-	-	-
		4800	-	-	-	-	-	-	-
	90	1200	1.57	0.90	0.57	0.38	0.26	1.57	0.90
		4800	0.03	0.01	0.01	0.00	0.00	0.00	0.00
Strong	10	1200	0.15	0.06	0.03	0.02	0.01	0.15	0.06
		4800	-	-	-	-	-	-	-
	50	300	2.93	1.96	1.38	1.03	0.78	0.61	0.48
		1200	0.49	0.29	0.19	0.13	0.09	0.07	0.05
		4800	0.06	0.03	0.02	0.01	0.00	0.00	0.00
	90	1200	3.44	2.44	1.82	1.42	1.15	0.95	0.79
		4800	1.26	0.85	0.62	0.48	0.39	0.31	0.27

Note: See also note on Table 3.1. '-': no observations are removed. '0.00': average number of observations removed < 0.005.

Table C.3: Support features of the DGPs with and without trimming - mean statistics

DGP	trimming	% treated remaining after trimming	% controls remaining after trimming	# of treated out of support	share (%) of treated in support	largest weight of control obs.	2nd largest weight of control obs.
300C50N	none	100.00%	100.00%	12.61	91.57%	6.84%	5.58%
300C50N	6%	96.91%	99.61%	13.29	90.88%	4.69%	4.25%
300C50N	4%	91.15%	98.89%	10.74	92.22%	3.81%	3.57%
300C50R	none	100.00%	100.00%	7.35	95.08%	4.03%	3.46%
300C50R	6%	99.79%	99.93%	7.86	94.73%	3.60%	3.21%
300C50R	4%	98.66%	99.68%	8.23	94.43%	3.14%	2.90%
300C50S	none	100.00%	100.00%	39.03	73.96%	17.32%	12.72%
300C50S	6%	69.02%	98.54%	15.34	85.90%	6.87%	6.28%
300C50S	4%	59.59%	97.32%	10.41	88.85%	5.60%	5.25%
300M50N	none	100.00%	100.00%	8.67	94.20%	5.15%	4.32%
300M50N	6%	99.18%	99.82%	9.74	93.44%	4.12%	3.70%
300M50N	4%	96.52%	99.37%	9.51	93.43%	3.44%	3.20%
300M50R	none	100.00%	100.00%	5.17	96.54%	3.23%	2.83%
300M50R	6%	99.96%	99.97%	5.39	96.39%	3.06%	2.74%
300M50R	4%	99.65%	99.86%	5.85	96.07%	2.80%	2.57%
300M50S	none	100.00%	100.00%	21.60	85.60%	11.36%	8.71%
300M50S	6%	87.21%	99.08%	16.76	87.53%	5.65%	5.16%
300M50S	4%	77.59%	98.04%	11.34	90.53%	4.53%	4.25%
1200C10N	none	100.00%	100.00%	1.26	98.94%	1.11%	1.00%
1200C10N	6%	100.00%	100.00%	1.26	98.94%	1.11%	1.00%
1200C10N	4%	100.00%	100.00%	1.27	98.94%	1.11%	1.00%
1200C10R	none	100.00%	100.00%	0.54	99.55%	0.52%	0.48%
1200C10R	6%	100.00%	100.00%	0.54	99.55%	0.52%	0.48%
1200C10R	4%	100.00%	100.00%	0.54	99.55%	0.52%	0.48%
1200C10S	none	100.00%	100.00%	6.06	94.97%	5.09%	4.11%
1200C10S	6%	99.45%	99.98%	6.84	94.30%	3.85%	3.38%
1200C10S	4%	97.83%	99.93%	6.87	94.19%	3.18%	2.90%
1200C50N	none	100.00%	100.00%	8.68	98.54%	1.46%	1.28%
1200C50N	6%	100.00%	100.00%	8.71	98.54%	1.45%	1.27%
1200C50N	4%	100.00%	100.00%	8.81	98.52%	1.43%	1.26%
1200C50R	none	100.00%	100.00%	3.22	99.46%	0.53%	0.49%
1200C50R	6%	100.00%	100.00%	3.22	99.46%	0.53%	0.49%
1200C50R	4%	100.00%	100.00%	3.22	99.46%	0.53%	0.49%
1200C50S	none	100.00%	100.00%	58.40	90.24%	8.29%	6.26%
1200C50S	6%	96.09%	99.89%	58.20	90.00%	4.53%	3.99%
1200C50S	4%	91.26%	99.76%	45.72	91.82%	3.58%	3.26%
1200C90N	none	100.00%	100.00%	126.51	88.29%	9.17%	7.23%
1200C90N	6%	92.32%	99.15%	114.93	88.69%	5.24%	4.77%
1200C90N	4%	83.74%	97.98%	80.70	91.30%	4.20%	3.95%
1200C90R	none	100.00%	100.00%	53.33	95.06%	3.85%	3.32%
1200C90R	6%	99.84%	99.93%	56.86	94.72%	3.48%	3.11%
1200C90R	4%	98.88%	99.66%	60.05	94.37%	3.08%	2.84%
1200C90S	none	100.00%	100.00%	436.74	59.53%	21.92%	15.65%
1200C90S	6%	51.91%	97.84%	98.55	83.77%	7.89%	7.16%

Table C.3 to be continued.

Table C.3: Support features of the DGPs with and without trimming - mean stats (cont.)

DGP	trimming	% treated remaining after trimming	% controls remaining after trimming	# of treated out of support	share (%) of treated in support	largest weight of control obs.	2nd largest weight of control obs.
1200C90S	4%	44.19%	96.37%	68.41	86.80%	6.49%	6.02%
1200M10N	none	100.00%	100.00%	0.95	99.20%	0.84%	0.76%
1200M10N	6%	100.00%	100.00%	0.95	99.20%	0.84%	0.76%
1200M10N	4%	100.00%	100.00%	0.95	99.20%	0.84%	0.76%
1200M10R	none	100.00%	100.00%	0.45	99.62%	0.41%	0.38%
1200M10R	6%	100.00%	100.00%	0.45	99.62%	0.41%	0.38%
1200M10R	4%	100.00%	100.00%	0.45	99.62%	0.41%	0.38%
1200M10S	none	100.00%	100.00%	3.59	97.03%	2.85%	2.42%
1200M10S	6%	99.98%	100.00%	3.73	96.91%	2.67%	2.32%
1200M10S	4%	99.87%	99.99%	4.01	96.67%	2.44%	2.17%
1200M50N	none	100.00%	100.00%	6.26	98.95%	1.06%	0.95%
1200M50N	6%	100.00%	100.00%	6.26	98.95%	1.06%	0.95%
1200M50N	4%	100.00%	100.00%	6.27	98.95%	1.06%	0.95%
1200M50R	none	100.00%	100.00%	2.57	99.57%	0.44%	0.41%
1200M50R	6%	100.00%	100.00%	2.57	99.57%	0.44%	0.41%
1200M50R	4%	100.00%	100.00%	2.57	99.57%	0.44%	0.41%
1200M50S	none	100.00%	100.00%	25.34	95.77%	4.52%	3.53%
1200M50S	6%	99.78%	99.97%	29.06	95.13%	3.34%	2.86%
1200M50S	4%	99.12%	99.92%	31.09	94.77%	2.81%	2.50%
1200M90N	none	100.00%	100.00%	84.53	92.17%	6.80%	5.53%
1200M90N	6%	97.38%	99.53%	93.12	91.18%	4.66%	4.21%
1200M90N	4%	91.98%	98.66%	79.08	92.14%	3.76%	3.52%
1200M90R	none	100.00%	100.00%	38.62	96.42%	3.17%	2.80%
1200M90R	6%	99.95%	99.97%	40.09	96.28%	3.01%	2.71%
1200M90R	4%	99.66%	99.83%	43.31	95.97%	2.78%	2.56%
1200M90S	none	100.00%	100.00%	231.56	78.55%	15.64%	11.45%
1200M90S	6%	76.42%	98.46%	119.77	86.18%	6.30%	5.72%
1200M90S	4%	67.08%	97.12%	81.13	89.29%	5.07%	4.74%
4800C10N	none	100.00%	100.00%	1.09	99.77%	0.23%	0.21%
4800C10N	6%	100.00%	100.00%	1.09	99.77%	0.23%	0.21%
4800C10N	4%	100.00%	100.00%	1.09	99.77%	0.23%	0.21%
4800C10R	none	100.00%	100.00%	0.31	99.94%	0.07%	0.06%
4800C10R	6%	100.00%	100.00%	0.31	99.94%	0.07%	0.06%
4800C10R	4%	100.00%	100.00%	0.31	99.94%	0.07%	0.06%
4800C10S	none	100.00%	100.00%	7.91	98.35%	1.59%	1.33%
4800C10S	6%	100.00%	100.00%	7.95	98.35%	1.56%	1.31%
4800C10S	4%	100.00%	100.00%	8.10	98.32%	1.50%	1.28%
4800C50N	none	100.00%	100.00%	8.67	99.64%	0.37%	0.33%
4800C50N	6%	100.00%	100.00%	8.67	99.64%	0.37%	0.33%
4800C50N	4%	100.00%	100.00%	8.67	99.64%	0.37%	0.33%
4800C50R	none	100.00%	100.00%	1.83	99.92%	0.08%	0.08%
4800C50R	6%	100.00%	100.00%	1.83	99.92%	0.08%	0.08%
4800C50R	4%	100.00%	100.00%	1.83	99.92%	0.08%	0.08%
4800C50S	none	100.00%	100.00%	101.55	95.76%	4.15%	3.11%
4800C50S	6%	99.82%	99.99%	110.75	95.36%	2.90%	2.42%

Table C.3 to be continued.

Table C.3: Support features of the DGPs with and without trimming - mean stats (cont.)

DGP	trimming	% treated remaining after trimming	% controls remaining after trimming	# of treated out of support	share (%) of treated in support	largest weight of control obs.	2nd largest weight of control obs.
4800C50S	4%	99.38%	99.99%	113.42	95.24%	2.47%	2.14%
4800C90N	none	100.00%	100.00%	127.30	97.05%	2.79%	2.34%
4800C90N	6%	99.98%	99.99%	131.86	96.95%	2.58%	2.22%
4800C90N	4%	99.87%	99.97%	140.69	96.74%	2.35%	2.07%
4800C90R	none	100.00%	100.00%	24.30	99.44%	0.56%	0.52%
4800C90R	6%	100.00%	100.00%	24.30	99.44%	0.56%	0.52%
4800C90R	4%	100.00%	100.00%	24.30	99.44%	0.56%	0.52%
4800C90S	none	100.00%	100.00%	996.11	76.92%	14.88%	10.55%
4800C90S	6%	79.92%	99.72%	511.59	86.19%	5.52%	4.90%
4800C90S	4%	72.31%	99.50%	326.57	90.13%	4.39%	3.99%
4800M10N	none	100.00%	100.00%	0.79	99.83%	0.17%	0.16%
4800M10N	6%	100.00%	100.00%	0.79	99.83%	0.17%	0.16%
4800M10N	4%	100.00%	100.00%	0.79	99.83%	0.17%	0.16%
4800M10R	none	100.00%	100.00%	0.25	99.95%	0.06%	0.05%
4800M10R	6%	100.00%	100.00%	0.25	99.95%	0.06%	0.05%
4800M10R	4%	100.00%	100.00%	0.25	99.95%	0.06%	0.05%
4800M10S	none	100.00%	100.00%	3.88	99.19%	0.83%	0.73%
4800M10S	6%	100.00%	100.00%	3.88	99.19%	0.83%	0.73%
4800M10S	4%	100.00%	100.00%	3.90	99.19%	0.82%	0.73%
4800M50N	none	100.00%	100.00%	5.73	99.76%	0.27%	0.25%
4800M50N	6%	100.00%	100.00%	5.73	99.76%	0.27%	0.25%
4800M50N	4%	100.00%	100.00%	5.73	99.76%	0.27%	0.25%
4800M50R	none	100.00%	100.00%	1.64	99.93%	0.07%	0.07%
4800M50R	6%	100.00%	100.00%	1.64	99.93%	0.07%	0.07%
4800M50R	4%	100.00%	100.00%	1.64	99.93%	0.07%	0.07%
4800M50S	none	100.00%	100.00%	30.49	98.73%	1.96%	1.55%
4800M50S	6%	100.00%	100.00%	31.08	98.70%	1.84%	1.48%
4800M50S	4%	99.99%	100.00%	32.43	98.65%	1.70%	1.40%
4800M90N	none	100.00%	100.00%	75.12	98.26%	1.91%	1.64%
4800M90N	6%	100.00%	100.00%	75.81	98.24%	1.88%	1.62%
4800M90N	4%	99.99%	99.99%	78.25	98.19%	1.80%	1.58%
4800M90R	none	100.00%	100.00%	20.32	99.53%	0.47%	0.44%
4800M90R	6%	100.00%	100.00%	20.32	99.53%	0.47%	0.44%
4800M90R	4%	100.00%	100.00%	20.32	99.53%	0.47%	0.44%
4800M90S	none	100.00%	100.00%	397.52	90.79%	9.08%	6.52%
4800M90S	6%	97.07%	99.87%	434.08	89.73%	4.32%	3.75%
4800M90S	4%	92.99%	99.74%	364.19	91.12%	3.43%	3.10%

Note: Means over all simulations. DGP numbers are generated as follows: 1st four digits: sample size, C=correctly specified p-score, M=misspecified p-score; 6-7th digit % treated, N=normal selection, R=random selection, S=strong selection. 'in support' / 'out of support' denote treated that have smaller / larger values than the largest p-score of any non-treated observation remaining after trimming.

Table C.4: Support features of the DGPs with and without trimming - median statistics

DGP	trimming	% treated remaining after trimming	% controls remaining after trimming	# of treated out of support	share (%) of treated in support	largest weight of control obs.	2nd largest weight of control obs.
300C50N	none	100.00%	100.00%	11	92.36%	5.78%	5.04%
300C50N	6%	100.00%	100.00%	12	91.50%	4.70%	4.24%
300C50N	4%	91.49%	98.77%	9	93.13%	3.83%	3.58%
300C50R	none	100.00%	100.00%	6	95.74%	3.59%	3.22%
300C50R	6%	100.00%	100.00%	7	95.36%	3.47%	3.14%
300C50R	4%	100.00%	100.00%	7	95.03%	3.17%	2.91%
300C50S	none	100.00%	100.00%	38	74.31%	13.73%	11.13%
300C50S	6%	68.03%	98.64%	12	87.62%	6.66%	6.14%
300C50S	4%	58.72%	97.33%	8	90.54%	5.42%	5.12%
300M50N	none	100.00%	100.00%	7	94.96%	4.44%	3.93%
300M50N	6%	100.00%	100.00%	9	94.16%	4.05%	3.64%
300M50N	4%	98.04%	99.35%	8	94.12%	3.48%	3.21%
300M50R	none	100.00%	100.00%	4	97.22%	2.90%	2.64%
300M50R	6%	100.00%	100.00%	4	97.04%	2.87%	2.62%
300M50R	4%	100.00%	100.00%	5	96.64%	2.76%	2.54%
300M50S	none	100.00%	100.00%	20	86.31%	9.03%	7.59%
300M50S	6%	86.93%	99.31%	14	88.73%	5.59%	5.11%
300M50S	4%	77.40%	98.04%	9	91.84%	4.45%	4.19%
1200C10N	none	100.00%	100.00%	1	99.18%	1.02%	0.94%
1200C10N	6%	100.00%	100.00%	1	99.18%	1.02%	0.94%
1200C10N	4%	100.00%	100.00%	1	99.18%	1.02%	0.94%
1200C10R	none	100.00%	100.00%	0	100.00%	0.48%	0.45%
1200C10R	6%	100.00%	100.00%	0	100.00%	0.48%	0.45%
1200C10R	4%	100.00%	100.00%	0	100.00%	0.48%	0.45%
1200C10S	none	100.00%	100.00%	5	95.61%	4.24%	3.65%
1200C10S	6%	100.00%	100.00%	6	94.92%	3.74%	3.32%
1200C10S	4%	99.22%	99.91%	6	94.74%	3.21%	2.90%
1200C50N	none	100.00%	100.00%	7	98.82%	1.31%	1.19%
1200C50N	6%	100.00%	100.00%	7	98.82%	1.31%	1.19%
1200C50N	4%	100.00%	100.00%	7	98.81%	1.30%	1.19%
1200C50R	none	100.00%	100.00%	2	99.65%	0.49%	0.47%
1200C50R	6%	100.00%	100.00%	2	99.65%	0.49%	0.47%
1200C50R	4%	100.00%	100.00%	2	99.65%	0.49%	0.47%
1200C50S	none	100.00%	100.00%	56	90.65%	6.32%	5.26%
1200C50S	6%	97.35%	99.84%	58	90.07%	4.48%	3.95%
1200C50S	4%	90.44%	99.83%	38	92.87%	3.57%	3.25%
1200C90N	none	100.00%	100.00%	119	89.01%	7.47%	6.38%
1200C90N	6%	92.53%	99.18%	104	89.40%	5.25%	4.73%
1200C90N	4%	83.46%	98.20%	67	92.46%	4.16%	3.91%
1200C90R	none	100.00%	100.00%	46	95.71%	3.39%	3.07%
1200C90R	6%	100.00%	100.00%	50	95.37%	3.30%	3.01%
1200C90R	4%	100.00%	100.00%	53	95.00%	3.08%	2.83%
1200C90S	none	100.00%	100.00%	442	58.97%	17.22%	13.60%
1200C90S	6%	50.09%	97.83%	75	85.67%	7.42%	6.85%
1200C90S	4%	42.75%	96.43%	50	88.86%	6.11%	5.75%
1200M10N	none	100.00%	100.00%	1	99.26%	0.77%	0.71%

Table C.4 to be continued.

Table C.4: Support features of the DGPs with and without trimming - median stats (cont.)

DGP	trimming	% treated remaining after trimming	% controls remaining after trimming	# of treated out of support	share (%) of treated in support	largest weight of control obs.	2nd largest weight of control obs.
1200M10N	6%	100.00%	100.00%	1	99.26%	0.77%	0.71%
1200M10N	4%	100.00%	100.00%	1	99.26%	0.77%	0.71%
1200M10R	none	100.00%	100.00%	0	100.00%	0.38%	0.36%
1200M10R	6%	100.00%	100.00%	0	100.00%	0.38%	0.36%
1200M10R	4%	100.00%	100.00%	0	100.00%	0.38%	0.36%
1200M10S	none	100.00%	100.00%	3	97.56%	2.47%	2.20%
1200M10S	6%	100.00%	100.00%	3	97.46%	2.44%	2.18%
1200M10S	4%	100.00%	100.00%	3	97.22%	2.36%	2.12%
1200M50N	none	100.00%	100.00%	5	99.17%	0.97%	0.90%
1200M50N	6%	100.00%	100.00%	5	99.17%	0.97%	0.90%
1200M50N	4%	100.00%	100.00%	5	99.17%	0.97%	0.90%
1200M50R	none	100.00%	100.00%	2	99.67%	0.41%	0.40%
1200M50R	6%	100.00%	100.00%	2	99.67%	0.41%	0.40%
1200M50R	4%	100.00%	100.00%	2	99.67%	0.41%	0.40%
1200M50S	none	100.00%	100.00%	22	96.31%	3.53%	3.01%
1200M50S	6%	100.00%	100.00%	26	95.62%	3.15%	2.73%
1200M50S	4%	100.00%	100.00%	29	95.14%	2.78%	2.47%
1200M90N	none	100.00%	100.00%	76	92.91%	5.65%	4.94%
1200M90N	6%	100.00%	100.00%	86	91.77%	4.67%	4.19%
1200M90N	4%	92.20%	99.03%	70	92.84%	3.78%	3.52%
1200M90R	none	100.00%	100.00%	32	96.98%	2.84%	2.61%
1200M90R	6%	100.00%	100.00%	34	96.85%	2.82%	2.59%
1200M90R	4%	100.00%	100.00%	38	96.49%	2.74%	2.53%
1200M90S	none	100.00%	100.00%	231	78.60%	11.86%	9.68%
1200M90S	6%	75.00%	98.41%	96	87.84%	6.12%	5.58%
1200M90S	4%	66.16%	97.27%	65	90.78%	4.89%	4.61%
4800C10N	none	100.00%	100.00%	1	99.80%	0.22%	0.20%
4800C10N	6%	100.00%	100.00%	1	99.80%	0.22%	0.20%
4800C10N	4%	100.00%	100.00%	1	99.80%	0.22%	0.20%
4800C10R	none	100.00%	100.00%	0	100.00%	0.06%	0.06%
4800C10R	6%	100.00%	100.00%	0	100.00%	0.06%	0.06%
4800C10R	4%	100.00%	100.00%	0	100.00%	0.06%	0.06%
4800C10S	none	100.00%	100.00%	7	98.61%	1.34%	1.18%
4800C10S	6%	100.00%	100.00%	7	98.60%	1.34%	1.18%
4800C10S	4%	100.00%	100.00%	7	98.58%	1.33%	1.17%
4800C50N	none	100.00%	100.00%	6	99.75%	0.34%	0.32%
4800C50N	6%	100.00%	100.00%	6	99.75%	0.34%	0.32%
4800C50N	4%	100.00%	100.00%	6	99.75%	0.34%	0.32%
4800C50R	none	100.00%	100.00%	1	99.96%	0.08%	0.08%
4800C50R	6%	100.00%	100.00%	1	99.96%	0.08%	0.08%
4800C50R	4%	100.00%	100.00%	1	99.96%	0.08%	0.08%
4800C50S	none	100.00%	100.00%	97	95.95%	2.89%	2.46%
4800C50S	6%	100.00%	100.00%	103	95.64%	2.63%	2.26%
4800C50S	4%	100.00%	100.00%	109	95.43%	2.41%	2.09%
4800C90N	none	100.00%	100.00%	111	97.41%	2.39%	2.12%
4800C90N	6%	100.00%	100.00%	117	97.29%	2.35%	2.08%

Table C.4 to be continued.

Table C.4: Support features of the DGPs with and without trimming - median stats (cont.)

DGP	trimming	% treated remaining after trimming	% controls remaining after trimming	# of treated out of support	share (%) of treated in support	largest weight of control obs.	2nd largest weight of control obs.
4800C90N	4%	100.00%	100.00%	126	97.05%	2.25%	2.02%
4800C90R	none	100.00%	100.00%	19	99.56%	0.52%	0.49%
4800C90R	6%	100.00%	100.00%	19	99.56%	0.52%	0.49%
4800C90R	4%	100.00%	100.00%	19	99.56%	0.52%	0.49%
4800C90S	none	100.00%	100.00%	1017	76.42%	10.40%	8.43%
4800C90S	6%	76.50%	99.79%	368	88.80%	5.31%	4.77%
4800C90S	4%	70.61%	99.57%	249	91.81%	4.20%	3.88%
4800M10N	none	100.00%	100.00%	0	100.00%	0.16%	0.16%
4800M10N	6%	100.00%	100.00%	0	100.00%	0.16%	0.16%
4800M10N	4%	100.00%	100.00%	0	100.00%	0.16%	0.16%
4800M10R	none	100.00%	100.00%	0	100.00%	0.05%	0.05%
4800M10R	6%	100.00%	100.00%	0	100.00%	0.05%	0.05%
4800M10R	4%	100.00%	100.00%	0	100.00%	0.05%	0.05%
4800M10S	none	100.00%	100.00%	3	99.38%	0.75%	0.69%
4800M10S	6%	100.00%	100.00%	3	99.38%	0.75%	0.69%
4800M10S	4%	100.00%	100.00%	3	99.38%	0.75%	0.68%
4800M50N	none	100.00%	100.00%	4	99.83%	0.25%	0.24%
4800M50N	6%	100.00%	100.00%	4	99.83%	0.25%	0.24%
4800M50N	4%	100.00%	100.00%	4	99.83%	0.25%	0.24%
4800M50R	none	100.00%	100.00%	1	99.96%	0.07%	0.07%
4800M50R	6%	100.00%	100.00%	1	99.96%	0.07%	0.07%
4800M50R	4%	100.00%	100.00%	1	99.96%	0.07%	0.07%
4800M50S	none	100.00%	100.00%	24	98.99%	1.55%	1.32%
4800M50S	6%	100.00%	100.00%	25	98.96%	1.53%	1.31%
4800M50S	4%	100.00%	100.00%	26	98.89%	1.48%	1.27%
4800M90N	none	100.00%	100.00%	62	98.55%	1.70%	1.52%
4800M90N	6%	100.00%	100.00%	64	98.52%	1.70%	1.51%
4800M90N	4%	100.00%	100.00%	66	98.47%	1.67%	1.50%
4800M90R	none	100.00%	100.00%	16	99.63%	0.44%	0.43%
4800M90R	6%	100.00%	100.00%	16	99.63%	0.44%	0.43%
4800M90R	4%	100.00%	100.00%	16	99.63%	0.44%	0.43%
4800M90S	none	100.00%	100.00%	383	91.11%	5.91%	4.81%
4800M90S	6%	100.00%	100.00%	431	89.77%	4.28%	3.74%
4800M90S	4%	92.03%	99.79%	320	91.94%	3.40%	3.07%

Note: Means over all simulations. DGP numbers are generated as follows: 1st four digits: sample size, C=correctly specified p-score, M=misspecified p-score; 6-7th digit % treated, N=normal selection, R=random selection, S=strong selection. 'in support' / 'out of support' denote treated that have smaller / larger values than the largest p-score of any non-treated observation remaining after trimming.

Appendix D: Further results from the simulations

Appendix D.1: Effect heterogeneity

Table D.1: Winners with/without heterogeneity (normal selection, 50% treated, no trimming)

		binary outcome				continuous outcome			
		no heterogeneity		heterogeneity		no heterogeneity		heterogeneity	
specification	N	Estimator	RMSE	Estimator	RMSE	Estimator	RMSE	Estimator	RMSE
correct	300	kernel (high)	6.03	kernel (high)	5.88	kernel (high)	148.30	kernel (high)	149.90
correct	1200	probit	2.72	kernel (high)	2.72	OLS	65.70	OLS	66.40
correct	4800	probit	1.33	probit	1.29	matching OLS	32.60	matching OLS	32.20
misspecified	300	matching logit	6.05	matching logit	5.99	kernel (high)	149.80	matching OLS	150.70
misspecified	1200	matching logit	2.92	matching logit	2.97	matching OLS	76.00	matching OLS	78.40
misspecified	4800	matching logit	1.80	matching logit	1.78	matching OLS	53.40	matching OLS	54.20

Note: Best estimator among those considered in Table 5.3. Inverse probability tilting (IPT) and OLS DR are not considered for this table.

Table D.2: Winners with/without heterogeneity (no selection, 50% treated, no trimming)

		binary outcome				continuous outcome			
		no heterogeneity		heterogeneity		no heterogeneity		heterogeneity	
specification	N	Estimator	RMSE	Estimator	RMSE	Estimator	RMSE	Estimator	RMSE
correct	300	matching logit	5.34	matching logit	5.32	matching OLS	121.20	matching OLS	122.50
correct	1200	probit	2.41	probit	2.45	OLS	54.50	OLS	55.30
correct	4800	probit	1.14	probit	1.17	IPW	26.60	IPW	26.80
misspecified	300	probit	5.24	probit	5.22	kernel (low)	120.40	kernel (low)	121.20
misspecified	1200	probit	2.39	probit	2.45	OLS	55.10	OLS	56.30
misspecified	4800	probit	1.14	probit	1.18	OLS	26.80	OLS	27.40

Note: Best estimator among those considered in Table 5.3. Inverse probability tilting (IPT) and OLS DR are not considered for this table.

Table D.3:Winners with/without heterogeneity (strong selection, 50% treated, no trimming)

		binary outcome				continuous outcome			
		no heterogeneity		heterogeneity		no heterogeneity		heterogeneity	
specification	N	Estimator	RMSE	Estimator	RMSE	Estimator	RMSE	Estim	RMSE
correct	300	matching logit	7.92	matching logit	7.90	IPW	199.80	IPW	203.60
correct	1200	probit	3.76	OLS	3.67	IPW	94.30	IPW	96.50
correct	4800	probit	1.79	probit	1.81	IPW	46.70	matching OLS	49.70
misspecified	300	matching logit	7.79	matching logit	7.77	matching OLS	206.20	IPW	209.20
misspecified	1200	matching logit	4.21	matching logit	4.13	matching OLS	129.80	matching OLS	131.40
misspecified	4800	matching logit	2.96	matching logit	2.98	kernel (high)	111.60	matching OLS	111.20

Note: Best estimator among those considered in Table 5.3. Inverse probability tilting (IPT) and OLS DR are not considered for this table.

Effect heterogeneity is modelled is modelled stochastically. For employment, the effect is such that if an individual is employed without the programme, there is a probability that she might not be employed with the programme. This probability is declining in the propensity score. If an individual is not employed without the programme there is probability that she will be employed with the programme. This probability is increasing in the propensity score. Therefore, the programme effect, which is around 10%-points on average, is increasing in the propensity score. For earnings, the first part of the effect comes directly from modelling employment. For those who turn out to be employed with the programme, a random share of mean earnings of the observed earnings distribution is added.

Note that the differences that appear in Tables D.1 to D.3 for the cases of effect heterogeneity and homogeneity are due to the impact of the trimming rule (which uses a cut-off value of 4% in this case). The latter changes the bias term for all estimators in the same direction by the same amount, because it affects the treated always in the same way. However, the impact of this shift on the squared biases coming from treated and controls is ambiguous.

Appendix D.2: Absolute bias and standard deviation of estimators

Table D.4: OLS regression analysis of the simulation results: determinants of absolute bias by class of estimator and sample size: employment outcome

Variables (all indicators)		IPW			Kernel			Matching			Parametric		
Sample Size		300	1200	4800	300	1200	4800	300	1200	4800	300	1200	4800
Constant		5.8	2.0	1.1	5.6	2.5	1.0	5.6	3.0	1.8	5.5	2.5	1.1
							Features of the data generating process						
Selection:	Random	-1.1	-1.3	-0.7	-0.6	-0.6	-0.8	-0.7	-0.7	-0.7	-0.8	-0.7	-0.6
	Observed	0	0	0	0	0	0	0	0	0	0	0	0
	Strong	2.8	3.1	2.8	2.2	2.3	2.4	2.1	2.2	2.1	1.8	1.8	1.7
Share treated:	10%	-	(1.0)	(0.2)	-	1.0	0.4	-	1.3	0.4	-	1.4	0.3
	50%	0	0	0	0	0	0	0	0	0	0	0	0
	90%	-	4.9	1.8	-	2.7	1.8	-	3.2	1.6	-	2.7	1.3
							Features of the estimators						
Misspecified p-score		(0.4)	(-0.8)	1.3	-0.6	0.3	1.1	-0.5	(-0.0)	0.8	(-0.2)	(0.2)	1.0
No trimming		0	0	0	0	0	0	0	0	0	0	0	0
Trimming max 6%		-1.3	(0.2)	(-0.3)	-0.4	-0.3	(-0.0)	-0.6	-0.7	-0.1	-0.5	-0.5	(-0.1)
Trimming max 4%		-1.6	(0.2)	(-0.5)	-0.6	-0.4	(-0.1)	-0.7	-0.8	-0.2	-0.6	-0.6	(-0.1)
Inverse probability tilting		(0.6)	1.4	(0.0)									
Bandwidth:	Low				(0.2)	(0.1)	(-0.0)						
	Cross validation				0	0	0						
	High				-0.5	(-0.1)	(0.0)						
	Rule of thumb				(0.2)	(0.1)	(0.0)						
	Local logit				0.7	0.2	(-0.1)						
Nearest neighbour								2.0	1.1	(-0.0)			
Radius matching:	Radius low							0.6	(0.2)	(-0.1)			
	medium							0	0	0			
	large							(0.0)	(0.1)	(0.1)			
No adjustment								0	0	0			
Regression adjustment								0.4	(-0.1)	-0.9			
Logit adjustment								-0.4	-0.4	(0.1)			
PScore instead of linear index								(0.1)	(0.1)	(0.1)			
Regression for treated											(-0.1)	0.3	(0.0)
Robust											0.4	(0.2)	(-0.0)
Probit											(0.1)	(-0.0)	-0.2
							Statistics						
R^2 (in %)		78	54	72	90	78	70	75	60	66	77	70	75
Number of observations		36	108	108	144	432	432	540	1620	1620	108	324	324

Note: Dependent variable: Bias. The estimation sample consists of the results of all simulations (all DGPs and all estimators) within the specified class of estimators and sample size. For the smallest sample size only simulations with 50% treated have been run. All coefficients are in %. Coefficients that are not significant at the 5% level (conventional OLS standard errors), appear in brackets.

Table D.5: OLS regression analysis of the simulation results: determinants of the standard deviation by class of estimator and sample size: employment outcome

Variables (all indicators)		IPW			Kernel			Matching			Parametric		
Sample Size		300	1200	4800	300	1200	4800	300	1200	4800	300	1200	4800
Constant		7.3	(2.4)	(2.7)	7.1	3.2	1.6	6.6	2.9	1.2	7.1	4.2	1.3
Features of the data generating process													
Selection:	Random	-0.8	(-1.5)	(-0.3)	-0.6	-0.5	-0.3	-0.8	-0.6	-0.3	-0.8	(-0.7)	-0.3
	Observed	0	0	0	0	0	0	0	0	0	0	0	0
	Strong	3.1	2.3	3.2	2.4	1.7	1.0	2.2	1.8	1.2	1.7	2.0	0.9
Share treated:	10%	-	(1.4)	(-0.9)	-	1.4	0.6	-	1.9	0.9	-	1.9	0.8
	50%	0	0	0	0	0	0	0	0	0	0	0	0
	90%	-	7.2	(0.6)	-	2.7	1.3	-	4.3	1.8	-	3.9	1.5
Features of the estimators													
Misspecified p-score		(-0.5)	-2.6	(0.6)	-1.2	-0.7	-0.4	-1.0	-0.7	-0.3	-0.8	-1.0	-0.2
No trimming		0	0	0	0	0	0	0	0	0	0	0	0
Trimming max 6%		-1.4	(0.9)	(-2.0)	-0.5	(-0.2)	(-0.0)	-0.7	-0.8	-0.1	-0.7	-1.2	(-0.1)
Trimming max 4%		-1.6	(1.0)	(-2.1)	-0.7	-0.3	(-0.1)	-0.9	-1.0	-0.2	-0.8	-1.3	-0.2
Inverse probability tilting		(0.5)	2.6	(1.1)									
Bandwidth: Low					(0.2)	(0.2)	(0.1)						
Cross validation					0	0	0						
High					-0.6	-0.3	-0.1						
Rule of thumb					(0.3)	(0.1)	(0.1)						
Local logit					0.9	0.4	(0.1)						
Nearest neighbour								3.1	2.3	1.1			
Radius matching:	Radius low							0.9	0.5	0.2			
	medium							0	0	0			
	large							(-0.2)	(-0.1)	(-0.0)			
No adjustment								0	0	0			
Regression adjustment								1.2	1.3	0.2			
Logit adjustment								-0.9	-1.1	-0.1			
PScore instead of linear index								(0.1)	(0.1)	(-0.0)			
Regression for treated											(0.1)	(0.3)	(0.1)
Robust											0.3	(-0.6)	0.4
Probit											(-0.1)	(-0.8)	(0.1)
Statistics													
R^2 (in %)		79	47	14	68	76	72	71	58	62	82	27	70
Number of observations		36	108	108	144	432	432	540	1620	1620	108	324	324

Note: Dependent variable: Standard deviation. The estimation sample consists of the results of all simulations (all DGPs and all estimators) within the specified class of estimators and sample size. For the smallest sample size only simulations with 50% treated have been run. All coefficients are in %. Coefficients that are not significant at the 5% level (conventional OLS standard errors), appear in brackets.

Table D.6: OLS regression analysis of the simulation results: determinants of absolute bias by class of estimator and sample size: earnings outcome

Variables (all indicators)		IPW			Kernel			Matching			Parametric*)		
Sample Size		300	1200	4800	300	1200	4800	300	1200	4800	300	1200	4800
Constant		143	60	30	134	64	24	141	87	57	129	64	27
							Features of the data generating process						
Selection:	Random	-28	-38	-23	-25	-24	-24	-28	-30	-27	-25	-25	-22
	Observed	0	0	0	0	0	0	0	0	0	0	0	0
	Strong	68	74	71	51	61	68	54	60	64	63	58	57
Share treated:	10%	-	(22)	(-1)	-	24	(10)	-	30	7	-	21	(4)
	50%	0	0	0	0	0	0	0	0	0	0	0	0
	90%	-	102	45	-	60	50	-	67	42	-	72	35
							Features of the estimators						
Misspecified p-score		(9)	(-10)	32	(-1)	13	25	(-1)	11	33	(-3)	13	33
No trimming		0	0	0	0	0	0	0	0	0	0	0	0
Trimming max 6%		-29	(-2)	(-10)	(-6)	-9	(-2)	-17	-18	(-4)	-19	-18	(-3)
Trimming max 4%		-37	(-4)	(-12)	-11	-12	(-3)	-23	-23	-6	-24	-21	(-4)
Inverse probability tilting		(4)	28	(4)									
Bandwidth: Low					(-6)	(-1)	(4)						
Cross validation					0	0	0						
High					-9	(-3)	(2)						
Rule of thumb					(0)	(0)	(1)						
Nearest neighbour								36	12	-18			
Radius matching:	Radius low							11	(2)	(-4)			
	medium							0	0	0			
	large							(1)	(4)	8			
Regression adjustment								-7	-18	-37			
PScore instead of linear index								(2)	(0)	(-1)			
Regression for treated											(-0)	(0)	(0)
Robust											23	10	(-1)
							Statistics						
R^2 (in %)		81	65	71	90	78	70	73	63	70	77	70	75
Number of observations		36	108	108	72	216	216	360	1080	1080	72	216	216

Note: Dependent variable: Bias. The estimation sample consists of the results of all simulations (all DGPs and all estimators) within the specified class of estimators and sample size. For the smallest sample size only simulations with 50% treated have been run. Coefficients that are not significant at the 5% level (conventional OLS standard errors), appear in brackets. *): Heckit and Tobit estimates are very unstable and therefore excluded from the regressions presented in this table.

Table D.7: OLS regression analysis of the simulation results: determinants of the standard deviation by class of estimator and sample size: earnings outcome

Variables (all indicators)		IPW			Kernel			Matching			Parametric*)		
Sample Size		300	1200	4800	300	1200	4800	300	1200	4800	300	1200	4800
Constant		188	79	(93)	174	87	44	154	68	29	172	116	30
Features of the data generating process													
Selection:	Random	-25	-42	(-14)	-28	-19	-11	-26	-25	-10	-29	(-22)	-9
	Observed	0	0	0	0	0	0	0	0	0	0	0	0
	Strong	68	55	115	51	36	26	48	35	28	63	107	28
Share treated:	10%	-	(33)	(-56)	-	34	14	-	49	22	-	(37)	15
	50%	0	0	0	0	0	0	0	0	0	0	0	0
	90%	-	141	(-17)	-	49	29	-	96	38	-	162	32
Features of the estimators													
Misspecified p-score		-21	-50	(35)	-17	-11	-9	-23	-16	-6	-29	-58	(-3)
No trimming		0	0	0	0	0	0	0	0	0	0	0	0
Trimming max 6%		-35	(3)	(-84)	(-9)	(-4)	(-0)	-19	-21	(-2)	-29	-89	(-3)
Trimming max 4%		-42	(4)	(-86)	-15	-6	(-1)	-26	-28	-4	-36	-94	(-4)
Inverse probability tilting		(-5)	48	(49)									
Bandwidth: Low					(-7)	(-2)	(1)						
Cross validation					0	0	0						
High					-13	-8	-8						
Rule of thumb					(2)	(-0)	(-0)						
Nearest neighbour								81	63	28			
Radius matching:	Radius low							21	12	5			
	medium							0	0	0			
	large							(-7)	(-4)	(-2)			
Regression adjustment								27	29	5			
PScore instead of linear index								(5)	(3)	(0)			
Regression for treated											(-0)	(1)	(0)
Robust											42	(35)	14
Statistics													
R^2 (in %)		73	54	11	82	84	75	66	56	61	65	23	69
Number of observations		36	108	108	72	216	216	360	1080	1080	72	216	216

Note: Dependent variable: Standard deviation. The estimation sample consists of the results of all simulations (all DGPs and all estimators) within the specified class of estimators and sample size. For the smallest sample size only simulations with 50% treated have been run. Coefficients that are not significant at the 5% level (conventional OLS standard errors), appear in brackets. *): Heckit and Tobit estimates are very unstable and therefore excluded from the regressions presented in this table.

Appendix D.3: Comparison of Tobit, Heckit, and OLS for earnings

Table D.8: Comparison of OLS, Heckit and Tobit for earnings outcome

	No trimming			Trimming 4%		
	OLS	Heckit	Tobit	OLS	Heckit	Tobit
			Total sample			
RelRMSE	34	29391	3996	2	26397	815
Bias	52	624	117	46	658	71
Std. dev.	113	29671	4104	82	26655	901
			N = 300			
RelRMSE	12	145	86	0	111	82
Bias	48	46	75	42	39	83
Std. dev.	157	362	268	140	311	260
			N = 1200			
RelRMSE	56	59536	8055	0	53481	1609
Bias	54	1398	196	46	1482	87
Std. dev.	162	69060	9434	98	62042	1966
			N = 4800			
RelRMSE	39	31	43	32	25	37
Bias	51	43	52	47	41	50
Std. dev.	49	51	53	47	49	50
			10% treated			
RelRMSE	14	10	13	14	10	13
Bias	38	31	37	38	31	38
Std. dev.	77	77	77	77	77	77
			50% treated			
RelRMSE	19	92	64	7	44	41
Bias	45	40	58	43	38	56
Std. dev.	86	155	124	75	111	103
			90% treated			
RelRMSE	94	90500	12254	3	87216	631
Bias	75	2094	285	59	2377	75
Std. dev.	189	103539	14101	87	99775	812
			Correctly specified p-score			
RelRMSE	62	56670	8688	0	59856	1474
Bias	29	1048	166	22	1254	75
Std. dev.	137	50375	7793	84	53199	1390
			Misspecified p-score			
RelRMSE	16	8156	310	8	28	308
Bias	75	200	68	70	63	66
Std. dev.	88	8967	415	81	111	413
			Normal selection			
RelRMSE	7	17740	2204	1	1253	83
Bias	39	305	64	39	71	60
Std. dev.	89	16719	2147	82	1260	156
			No selection			
RelRMSE	3	100297	58	3	110159	59
Bias	1	1448	22	1	1791	23
Std. dev.	72	70354	108	72	77260	109
			Strong selection			
RelRMSE	78	1420	7660	10	1040	1806
Bias	115	119	265	98	113	129
Std. dev.	177	1940	10057	93	1445	2440

Note: Relative root mean square error is relative to best overall estimator (as in Table 5.4).

Appendix E: Further sensitivity checks

E.1 More flexible specification of propensity score

We also consider flexible data-driven parametric estimation, which may be regarded as a basic implementation of sieve regression, see for instance Chen (2007). We adapt the probit and OLS models to the sample size by selectively adding higher order and interaction terms in order to minimize the corrected Akaike information criterion (AIC).[1] The latter is defined as

$$2k - 2\ln(L) + \frac{2k(k+1)}{N_0 - k - 1},$$

where $\ln(L)$ denotes the log-likelihood function of the model (for OLS, standard normally distributed errors are assumed), k is the dimension of the regressors, interactions, and higher order terms, and N_0 is the number of nontreated observations used for the estimation. The leading terms $2k - 2\ln(L)$, which constitute the uncorrected AIC, imply that the criterion both increases and declines in the dimension due to the penalty $2k$ and the improvement of the likelihood function, respectively. Adding the correction term $\frac{2k(k+1)}{N_0 - k - 1}$ ameliorates the finite sample properties (see Hurvich and Tsai, 1989), as this additional penalty reduces the probability of undersmoothing in small data sets.

Note that when considering scenarios with misspecified propensity scores, we here induce potential misspecification by excluding any interaction terms in the models assessed by the corrected AIC, while higher order terms are allowed. In contrast, in scenarios with correct propensity score specification, both interaction and higher order terms may be flexibly combined. All in all, we consider more than 180 different model specifications. In our

[1] We thank an anonymous referee for this suggestion.

simulations however, standard probit and OLS dominate flexible estimation based on the corrected AIC in terms of RMSE. Therefore, the results for the latter are not reported in the main text, but are provided in Tables E.1 to E.3 for selected DGPs under 4% trimming.

Table E.1: Est. based on corr. AIC (no effect het., normal selection, 50% treated, 4% trimming)

		binary outcome			continuous outcome		
specification	N	bias	std. dev.	RMSE	bias	std. dev.	RMSE
correct	300	2.02	6.58	6.88	15.2	316.6	317.0
correct	1200	0.03	2.74	2.74	9.6	65.4	66.1
correct	4800	0.01	1.33	1.33	10.7	32.1	33.9
misspecified	300	1.46	6.06	6.23	16.7	532.6	532.8
misspecified	1200	0.01	2.73	2.73	8.0	66.0	66.5
misspecified	4800	-0.12	1.30	1.31	6.8	32.2	32.9

Table E.2: Est. based on corr. AIC (no effect het., no selection, 50% treated, 4% trimming)

		binary outcome			continuous outcome		
specification	N	bias	std. dev.	RMSE	bias	std. dev.	RMSE
correct	300	1.25	5.82	5.95	-0.6	133.2	133.2
correct	1200	-0.09	2.41	2.41	0.3	54.6	54.6
correct	4800	-0.09	1.13	1.14	1.1	26.6	26.6
misspecified	300	0.43	5.38	5.40	-4.0	284.2	284.2
misspecified	1200	-0.14	2.38	2.38	0.3	54.6	54.6
misspecified	4800	-0.12	1.13	1.14	1.1	26.7	26.7

Table E.3: Est. based on corr. AIC (no effect het., strong selection, 50% treated, 4% trimming)

		binary outcome			continuous outcome		
specification	N	bias	std. dev.	RMSE	bias	std. dev.	RMSE
correct	300	2.62	7.88	8.30	47.3	172.7	179.0
correct	1200	0.39	3.55	3.57	43.6	81.6	92.5
correct	4800	0.10	1.79	1.79	48.4	41.2	63.5
misspecified	300	2.73	7.41	7.90	77.9	351.9	360.5
misspecified	1200	0.46	3.79	3.82	45.3	87.9	98.9
misspecified	4800	-0.20	1.67	1.68	38.2	40.9	55.9

E.2 Different ways to deal with the common support

Table E.4 shows variations of the trimming rules used in this paper. These variations concern the way how the control observations are handled. As the impact of trimming and common support routines are expected to be largest for the case with an extreme treatment control ratio and strong selection, it shows the case of 90% treated and strong selection (not available for N = 300).

The middle panel present the results for the baseline procedure used in this paper, i.e. controls are removed according to their weights. Next, if any control observation has been removed, then the value of the p-score that corresponds to the smallest p-score of the removed observations is computed and all treated with a p-score larger than this value are removed as well.

The upper panels show the case without removing any treated, while the lower panels show the case in which treated with a p-score larger than the largest remaining control observation are removed as well.

The results clearly show that in this rather extreme DGP not enforcing the support condition among the treated leads to considerable additional bias, while using an even stronger criterion seems to lead to some further improvements for all estimators but the tilting estimator.

Table E.4: Comparison of the properties of the selected estimators for different ways of dealing with treated observations: Correct p-score, 90% treated, strong selection

	Employment								Earnings							
			Kernel		Matching		Probit				Kernel		Matching		OLS	
	IPW	IPT	high	low	logit	pair		DR	IPW	IPT	high	low	OLS	pair		DR
Only deletion of controls in trimming procedure																
N = 1200																
RelRMSE	17	18	7	13	13	199	3*	217	21	21	4*	30	63	179	45	170
Bias	9.7	9.1	6.4	7.9	5.5	3.5	6.0	3.4	258	248	211	200	187	94	160	135
Std. dev.	7.7	8.6	9.4	9.0	10.7	31.4	9.2	33.3	168	185	161	264	372	706	333	676
Skew.	0.0	-0.1	-0.3	-0.4	0.7	0.4	-0.3	2.4	-0.4	-0.6	-0.1	-1.7	-1.1	-0.7	2.4	1.3
Kurtosis	3.0	3.1	3.3	5.9	3.4	1.2	3.7	60.1	4.2	4.7	3.2	9.4	6.0	3.0	69.8	65.4
N = 4800																
RelRMSE	35	42	26	66	25	261	0.2*	113	35	41	33	78	2*	242	8	80
Bias	4.7	4.7	3.9	6.6	4.0	0.7	1.3	0.7	132	133	149	133	58	38	103	45
Std. dev.	4.7	5.1	4.8	4.8	4.6	17.7	4.7	10.4	108	119	79	182	115	430	89	223
Skew.	0.2	0.0	-0.1	-1.9	0.3	0.6	0.2	0.0	-0.2	-0.2	-0.1	-1.4	-0.5	-0.7	0.0	-0.2
Kurtosis	3.3	3.1	2.7	12.8	3.2	1.6	3.1	3.4	3.4	3.3	3.2	4.7	5.0	3.1	3.1	3.5
Deletion of treated and controls in trimming procedure (used throughout this paper)																
N = 1200																
RelRMSE	29	511	20	17	26	98	B	2	18	372	B	16	8	84	8	46
Bias	0.8	20.0	1.4	2.2	1.6	0.0	0.2	0.5	19	361	52	49	25	3	45	9
Std. dev.	7.5	29.8	6.9	6.5	7.2	11.6	5.9	6.0	161	537	127	151	147	252	142	200
Skew.	0.0	0.8	-0.1	-0.2	0.0	0.3	0.1	0.2	-0.3	0.8	-0.1	-0.8	-0.2	-0.5	0.8	-0.8
Kurtosis	2.9	2.0	3.1	3.7	2.9	2.8	3.6	3.7	3.8	2.1	3.1	5.7	3.8	4.3	22.9	14.8
N = 4800																
RelRMSE	38	47	31	63	21	129	B	34	33	42	26	76	3	128	B	33
Bias	0.9	1.0	2.0	4.2	1.5	0.2	0.6	0.5	22	25	78	73	19	8	48	16
Std. dev.	4.7	5.0	4.1	3.8	3.9	7.9	3.4	4.6	109	117	71	127	84	190	68	110
Skew.	0.3	0.0	-0.1	-1.1	0.0	0.5	0.1	0.3	-0.2	-0.1	-0.1	-1.0	-0.2	-0.7	0.1	-0.2
Kurtosis	3.4	3.0	2.9	7.2	3.2	3.0	3.2	4.3	3.5	3.3	3.2	3.8	3.2	6.2	3.2	5.4
Deletion of treated and controls in trimming procedure, additionally common support enforced after trimming																
N = 1200																
RelRMSE	35	475	16	11	19	55	2*	65	29	358	B	8	6	42	9	41
Bias	0.7	17.1	0.8	1.5	1.2	0.0	1.0	0.1	20	300	31	31	19	2	33	1
Std. dev.	7.5	27.0	6.4	6.0	6.5	8.6	5.6	9.2	159	483	120	131	130	177	131	175
Skew.	0.0	0.8	-0.1	-0.1	0.0	0.0	0.2	1.1	-0.3	0.7	-0.1	-0.4	-0.1	-0.2	0.7	-0.7
Kurtosis	2.9	1.9	3.0	3.2	3.0	2.9	3.3	59.9	3.9	2.0	3.0	3.9	3.2	3.0	20.4	16.7
N = 4800																
RelRMSE	47	53	30	54	14	65	B	31	48	55	24	71	5	64	B	25
Bias	0.3	0.1	1.5	3.6	1.0	0.3	0.4	0.3	12	5	59	60	14	5	35	10
Std. dev.	4.6	4.8	3.8	3.3	3.5	5.2	3.1	4.1	106	111	67	107	74	118	63	89
Skew.	0.2	0.0	0.0	-0.5	0.0	0.1	0.0	0.1	-0.2	-0.1	-0.1	-0.9	-0.1	-0.2	-0.1	-0.1
Kurtosis	3.4	3.1	2.9	4.4	3.0	3.1	3.0	3.4	3.5	3.6	3.2	3.5	3.0	3.4	3.1	3.4

Note: RelRMSE: Difference in relative root mean squared error in % compared to best estimator. Bias and standard deviation for employment is given in %. All results based on relative trimming level of 4%. *The best estimator is this estimator with 6% trimming.

E.3 Including 'too many' covariates in the propensity score

In this section, we investigate the behaviour of the estimators when the specification of the propensity score is more flexibly specified than required by the parametric model used to

simulate the treated in the Monte Carlo study. This has been implemented by including additionally squares of the seven continuous variables of the true model.

Table E.5: Comparison of the properties of the selected estimators for model with overspecified propensity score: Correct p-score, 50% treated, normal selection

	Employment								Earnings							
			Kernel		Matching		Probit				Kernel		Matching		OLS	
	IPW	IPT	high	low	logit	pair		DR	IPW	IPT	high	low	OLS	pair		DR
Correctly specified p-score																
N = 300																
RelRMSE	2	13	B	5	5	62	10	14	5	15	B	8	5	64	0.1	15
Bias	0.5	0.5	0.2	0.2	1.0	0.1	1.7	1.8	11	17	7	4	10	4	17	1
Std. dev.	6.1	6.7	6.0	6.3	6.2	9.7	6.3	6.6	148	162	142	153	148	233	141	164
Skew.	0.1	0.2	0.0	0.1	0.1	0.2	0.0	0.0	-0.1	-0.1	0.0	-0.1	-0.9	-0.3	0.0	0.3
Kurtosis	3.0	3.6	3.0	3.1	3.1	5.9	2.9	2.8	3.0	3.1	3.0	3.4	17.5	4.9	4.4	9.8
N = 1200																
RelRMSE	6	4	5	8	6	42	B	5	10	11	7	12	1	51	0.001*	3
Bias	0.0	0.0	0.5	0.5	0.6	0.0	0.0	0.1	0.0	3	16	10	5	0.0	10	5
Std. dev.	2.9	2.8	2.8	2.9	2.8	3.9	2.7	2.9	72	73	68	73	66	99	65	67
Skew.	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0
Kurtosis	3.0	3.0	3.0	3.0	3.0	3.0	2.9	3.0	3.0	3.0	2.9	3.0	2.8	3.3	3.0	3.0
N = 4800																
RelRMSE	5	3	5	9	5	45	B	3	8	3	10	13	B	44	4	6
Bias	0.0	0.0	0.1	0.2	0.1	0.1	0.0	0.1	1	1	11	2	1	0.0	11	7
Std. dev.	1.4	1.4	1.4	1.4	1.4	1.9	1.3	1.4	35	34	34	37	33	47	32	34
Skew.	-0.1	0.0	-0.1	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0
Kurtosis	3.1	2.7	3.1	2.8	3.0	3.0	3.0	3.0	3.0	3.0	3.2	3.0	3.3	3.0	3.1	3.2
Too many explanatory variables in p-score																
N = 300																
RelRMSE	2	5	5**	6	10	106	22	257	4	7	5**	7	33	117	75	240
Bias	1.8	1.6	0.0	0.2	0.9	0.5	3.1	0.2	42	39	8	12	33	8	21	7
Std. dev.	6.2	6.5	6.7	6.8	6.9	13.1	7.1	22.7	153	159	161	162	201	331	266	520
Skew.	0.1	0.1	0.1	0.1	0.2	0.4	0.0	-1.8	0.0	-0.1	-0.1	-0.1	-2.7	-0.6	-1.1	-3.6
Kurtosis	3.1	3.2	3.1	3.1	3.0	3.4	2.8	255	3.1	3.2	3.1	3.2	45.3	3.9	80.2	278
N = 1200																
RelRMSE	6	4	4	8	7	42	0.01*	7	9	10	5	10	1	50	0.01*	4
Bias	0.0	0.0	0.5	0.6	0.6	0.0	0.1	0.1	1	1	16	11	6	1	11	5
Std. dev.	2.9	2.8	2.8	2.9	2.9	3.9	2.7	2.9	73	74	69	73	67	101	66	69
Skew.	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0
Kurtosis	3.0	3.0	2.9	3.0	2.9	3.0	2.9	3.1	2.9	2.9	3.0	3.1	2.9	3.1	3.0	3.1
N = 4800																
RelRMSE	6	3	5	10	6	43	B	7	8	2	10	14	B	46	5	6
Bias	0.0	0.0	0.1	0.2	0.2	0.1	0.1	0.2	1	1	11	4	0	2	13	8
Std. dev.	1.4	1.4	1.4	1.4	1.4	1.9	1.3	1.4	35	33	34	37	33	48	32	34
Skew.	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.1	-0.1	0.0	0.0	0.0
Kurtosis	3.1	2.7	3.1	3.0	3.2	3.1	3.0	3.0	3.0	3.0	3.1	2.9	3.3	2.9	3.1	3.1

Note: RelRMSE: Difference in relative root mean squared error in % compared to best estimator. Bias and standard deviation for employment is given in %. All results based on relative trimming level of 4%. *The best estimator is this estimator with 6% trimming. **The best estimator is this estimator without trimming.

Additional reference

Hurvich, C. M. and C. Tsai (1989): “Regression and Time Series Model Selection in Small Samples,” *Biometrika*, 76, 297-307.