

Example 6.4.1 Matching with bias-corrected and Robust variance estimators

```
> setwd("~/propensity scores/book/Chapter6")
> library("foreign", lib.loc="C:/Program Files/R/R-2.13.2/library")
> data <- read.dta("cds_pc97.dta")
> rstudio::viewData(data)
> library("psych", lib.loc="C:/Program Files/R/R-2.13.2/library")
> describe(data)
```

	var	n	mean	sd	median	trimmed	mad	min
kid	1	606	3561777.09	2294886.33	2708535.00	3565607.31	3432960.30	4180
pcg_id	2	606	3561793.42	2294869.67	2708506.00	3565622.63	3433081.87	4179
age97	3	606	8.49	2.03	8.00	8.39	2.97	5
pcss97	4	606	102.45	16.10	101.00	102.81	16.31	38
k_adc	5	606	12.16	23.17	0.00	6.39	0.00	0
pcged97	6	606	12.55	1.83	12.00	12.51	1.48	7
pcg_adc	7	606	0.92	1.91	0.00	0.40	0.00	0
mratio96	8	606	2.45	2.43	2.06	2.11	1.58	0
black	9	606	0.50	0.50	0.50	0.50	0.74	0
male	10	606	0.53	0.50	1.00	0.54	0.00	0
kuse	11	606	0.31	0.46	0.00	0.26	0.00	0

```
> describe(data, se=TRUE)
```

	max	range	skew	kurtosis	se
kid	6845039.00	6840859.00	0.06	-1.64	93223.38
pcg_id	6845006.00	6840827.00	0.06	-1.64	93222.71
age97	13.00	8.00	0.25	-1.09	0.08
pcss97	154.00	116.00	-0.24	0.44	0.65
k_adc	100.00	100.00	1.94	2.76	0.94
pcged97	17.00	10.00	0.18	0.30	0.07
pcg_adc	7.00	7.00	2.10	3.21	0.08
mratio96	26.84	26.84	4.35	31.31	0.10
black	1.00	1.00	0.00	-2.01	0.02
male	1.00	1.00	-0.13	-1.99	0.02
kuse	1.00	1.00	0.82	-1.33	0.02

```
> library("Matching", lib.loc="C:/Program Files/R/R-2.13.2/library")
Loading required package: rgenoud
## rgenoud (Version 5.7-3, Build Date: 2011-05-04)
## See http://sekhon.berkeley.edu/rgenoud for additional documentation.
Loading required package: MASS
##
## Matching (Version 4.8-0, Build Date: 2012/01/19)
## See http://sekhon.berkeley.edu/matching for additional documentation.
## Please cite software as:
## Jasjeet S. Sekhon. 2011. "Multivariate and Propensity Score Matching
## Software with Automated Balance Optimization: The Matching package for R."
## Journal of Statistical Software, 42(7): 1-52.
##

> cov <- data.frame(data$male, data$black, data$age97, data$pcged97, data$mratio96,
data$pcg_adc)
> SATE_est <- Match(Y = data$pcss97, Tr= data$kuse, X=cov, Z = cov, Weight = 1, M=4,
BiasAdjust=TRUE, estimand = "ATE" )
> summary(SATE_est)
```

```
Estimate... -4.7038
AI SE..... 1.7475
T-stat..... -2.6917
p.val..... 0.0071083

Original number of observations..... 606
Original number of treated obs..... 188
Matched number of observations..... 606
Matched number of observations (unweighted). 2441
> SATT_est <- Match(Y = data$pcss97, Tr= data$kuse, X=cov, Z = cov, Weight = 1, M=4,
BiasAdjust=TRUE, estimand = "ATT" )
> summary(SATT_est)
```

```
Estimate... -5.2297
AI SE..... 1.7053
```

```
T-stat..... -3.0667
p.val..... 0.0021645
```

```
Original number of observations..... 606
Original number of treated obs..... 188
Matched number of observations..... 188
Matched number of observations (unweighted). 754
```

```
> SATC_est <- Match(Y = data$pcss97, Tr= data$kuse, X=cov, Z = cov, Weight = 1, M=4,
BiasAdjust=TRUE, estimand ="ATC")
> summary(SATC_est)
```

```
Estimate... -4.4673
AI SE..... 2.1267
T-stat..... -2.1005
p.val..... 0.035683
```

```
Original number of observations..... 606
Original number of control obs..... 418
Matched number of observations..... 418
Matched number of observations (unweighted). 1687
```

Below, we change the weights from the inverse of the variance matrix to the inverse of the variance-covariance matrix (i.e., Mahalanobis metric distance)

```
> SATE_M <- Match(Y = data$pcss97, Tr= data$kuse, X=cov, Z = cov, Weight = 2, M=4,
BiasAdjust=TRUE, estimand = "ATE" )
> summary(SATE_M)
```

```
Estimate... -4.7911
AI SE..... 1.6881
T-stat..... -2.8382
p.val..... 0.0045365
```

```
Original number of observations..... 606
Original number of treated obs..... 188
Matched number of observations..... 606
Matched number of observations (unweighted). 2427
```

```
> SATT_M <- Match(Y = data$pcss97, Tr= data$kuse, X=cov, Z = cov, Weight = 2, M=4,
BiasAdjust=TRUE, estimand ="ATT")
> summary(SATT_M)
```

```
Estimate... -5.812
AI SE..... 1.6553
T-stat..... -3.511
p.val..... 0.00044639
```

```
Original number of observations..... 606
Original number of treated obs..... 188
Matched number of observations..... 188
Matched number of observations (unweighted). 753
```

```
> SATC_M <- Match(Y = data$pcss97, Tr= data$kuse, X=cov, Z = cov, Weight = 2, M=4,
BiasAdjust=TRUE, estimand ="ATC")
> summary(SATC_M)
```

```
Estimate... -4.332
AI SE..... 2.0263
T-stat..... -2.1379
p.val..... 0.032523
```

```
Original number of observations..... 606
Original number of control obs..... 418
Matched number of observations..... 418
Matched number of observations (unweighted). 1674
```

Below, we change the number of matches from 4 to 1 and go back to use the inverse of the variance matrix

```
> SATE_Match1 <- Match(Y = data$pcss97, Tr= data$kuse, X=cov, Z = cov, Weight = 1,
M=1, BiasAdjust=TRUE, estimand = "ATE" )
> summary(SATE_Match1)
```

```
Estimate... -4.9616
AI SE..... 2.1467
T-stat..... -2.3112
p.val..... 0.020821
```

```
Original number of observations..... 606
Original number of treated obs..... 188
Matched number of observations..... 606
Matched number of observations (unweighted). 609
```

```
> SATT_Match1 <- Match(Y = data$pcss97, Tr= data$kuse, X=cov, Z = cov, Weight = 1,
M=1, BiasAdjust=TRUE, estimand = "ATT")
> summary(SATT_Match1)
```

```
Estimate... -6.1842
AI SE..... 2.0756
T-stat..... -2.9795
p.val..... 0.0028873
```

```
Original number of observations..... 606
Original number of treated obs..... 188
Matched number of observations..... 188
Matched number of observations (unweighted). 189
```

```
> SATC_Match1 <- Match(Y = data$pcss97, Tr= data$kuse, X=cov, Z = cov, Weight = 1,
M=1, BiasAdjust=TRUE, estimand = "ATC")
> summary(SATC_Match1)
```

```
Estimate... -4.4117
AI SE..... 2.7737
T-stat..... -1.5905
p.val..... 0.11172
```

```
Original number of observations..... 606
Original number of control obs..... 418
Matched number of observations..... 418
Matched number of observations (unweighted). 420
```

Below, we use a function included in this program named MatchBalance. This function, according to the manual, provides “Tests for Univariate and Multivariate Balance”

```
> mb_SATE_est <- MatchBalance(kuse ~ male + black + age97 + pcged97 + mratio96 +
pcg_adc, data = data, match.out=SATE_est, ks = TRUE, nboots=50)
```

```
***** (V1) male *****
```

	Before Matching	After Matching
mean treatment.....	0.54787	0.52888
mean control.....	0.52392	0.53383
std mean diff.....	4.7991	-0.99094
mean raw eQQ diff.....	0.026596	0.004916
med raw eQQ diff.....	0	0

max	raw eQQ diff.....	1	1
mean	eCDF diff.....	0.011974	0.002458
med	eCDF diff.....	0.011974	0.002458
max	eCDF diff.....	0.023949	0.004916

var ratio (Tr/Co).....	0.99603	1.0012
T-test p-value.....	0.58529	0.22058

***** (V2) black *****

	Before Matching	After Matching
mean treatment.....	0.79787	0.52244
mean control.....	0.36603	0.49794
std mean diff.....	107.25	4.9019

mean	raw eQQ diff.....	0.43085	0.02458
med	raw eQQ diff.....	0	0
max	raw eQQ diff.....	1	1

mean	eCDF diff.....	0.21592	0.01229
med	eCDF diff.....	0.21592	0.01229
max	eCDF diff.....	0.43184	0.02458

var ratio (Tr/Co).....	0.69703	0.998
T-test p-value.....	< 2.22e-16	0.00010625

***** (V3) age97 *****

	Before Matching	After Matching
mean treatment.....	8.5904	8.4524
mean control.....	8.445	8.4503
std mean diff.....	7.0257	0.10781

mean	raw eQQ diff.....	0.16489	0.070053
med	raw eQQ diff.....	0	0
max	raw eQQ diff.....	1	1

mean	eCDF diff.....	0.017841	0.0077837
med	eCDF diff.....	0.010969	0.008603
max	eCDF diff.....	0.052097	0.016387

var ratio (Tr/Co).....	1.063	0.93848
T-test p-value.....	0.41991	0.96248
KS Bootstrap p-value..	0.6	0.48
KS Naive p-value.....	0.87308	0.89849
KS Statistic.....	0.052097	0.016387

***** (V4) pcged97 *****

	Before Matching	After Matching
mean treatment.....	11.617	12.083
mean control.....	12.974	12.667
std mean diff.....	-87.609	-44.278

mean	raw eQQ diff.....	1.3457	0.58746
med	raw eQQ diff.....	1	0
max	raw eQQ diff.....	3	2

mean	eCDF diff.....	0.12333	0.053406
med	eCDF diff.....	0.11269	0.016387
max	eCDF diff.....	0.31436	0.16837

var ratio (Tr/Co).....	0.74142	0.56975
T-test p-value.....	< 2.22e-16	< 2.22e-16
KS Bootstrap p-value..	< 2.22e-16	< 2.22e-16
KS Naive p-value.....	1.4785e-11	< 2.22e-16
KS Statistic.....	0.31436	0.16837

***** (V5) mratio96 *****

	Before Matching	After Matching
--	-----------------	----------------

mean treatment.....	1.0688	1.4899
mean control.....	3.0708	2.6497
std mean diff.....	-199.25	-105.88
mean raw eQQ diff.....	2.0076	1.1581
med raw eQQ diff.....	1.8335	0.83069
max raw eQQ diff.....	18.554	18.554
mean eCDF diff.....	0.36023	0.19717
med eCDF diff.....	0.41917	0.2331
max eCDF diff.....	0.5961	0.356
var ratio (Tr/Co).....	0.14627	0.22291
T-test p-value.....	< 2.22e-16	< 2.22e-16
KS Bootstrap p-value..	< 2.22e-16	< 2.22e-16
KS Naive p-value.....	< 2.22e-16	< 2.22e-16
KS Statistic.....	0.5961	0.356

***** (V6) pcg_adc *****

	Before Matching	After Matching
mean treatment.....	1.8298	0.88952
mean control.....	0.50478	0.79785
std mean diff.....	52.163	4.8167
mean raw eQQ diff.....	1.3298	0.09914
med raw eQQ diff.....	0	0
max raw eQQ diff.....	6	2
mean eCDF diff.....	0.16563	0.012392
med eCDF diff.....	0.17968	0.0096272
max eCDF diff.....	0.26631	0.038509
var ratio (Tr/Co).....	3.5102	1.2166
T-test p-value.....	1.2444e-10	0.007065
KS Bootstrap p-value..	< 2.22e-16	< 2.22e-16
KS Naive p-value.....	2.0539e-08	0.053574
KS Statistic.....	0.26631	0.038509

Before Matching Minimum p.value: < 2.22e-16
 Variable Name(s): black pcged97 mratio96 pcg_adc Number(s): 2 4 5 6

After Matching Minimum p.value: < 2.22e-16
 Variable Name(s): pcged97 mratio96 pcg_adc Number(s): 4 5 6

Below, for the same estimate (SATE with inverse variance), but M=1

```
> mb_SATE_M1 <- MatchBalance(kuse ~ male + black + age97 + pcged97 + mratio96 +
pcg_adc, data = data, match.out=SATE_Match1, ks = TRUE, nboots=50)
```

***** (V1) male *****

	Before Matching	After Matching
mean treatment.....	0.54787	0.51815
mean control.....	0.52392	0.53135
std mean diff.....	4.7991	-2.6398
mean raw eQQ diff.....	0.026596	0.013136
med raw eQQ diff.....	0	0
max raw eQQ diff.....	1	1
mean eCDF diff.....	0.011974	0.0065681
med eCDF diff.....	0.011974	0.0065681
max eCDF diff.....	0.023949	0.013136
var ratio (Tr/Co).....	0.99603	1.0026
T-test p-value.....	0.58529	0.0045591

***** (V2) black *****

	Before Matching	After Matching
mean treatment.....	0.79787	0.52145
mean control.....	0.36603	0.5
std mean diff.....	107.25	4.2908
mean raw eQQ diff.....	0.43085	0.021346
med raw eQQ diff.....	0	0
max raw eQQ diff.....	1	1
mean eCDF diff.....	0.21592	0.010673
med eCDF diff.....	0.21592	0.010673
max eCDF diff.....	0.43184	0.021346
var ratio (Tr/Co).....	0.69703	0.99816
T-test p-value.....	< 2.22e-16	0.00029054

***** (V3) age97 *****

	Before Matching	After Matching
mean treatment.....	8.5904	8.4076
mean control.....	8.445	8.4472
std mean diff.....	7.0257	-2.0089
mean raw eQQ diff.....	0.16489	0.11494
med raw eQQ diff.....	0	0
max raw eQQ diff.....	1	1
mean eCDF diff.....	0.017841	0.012771
med eCDF diff.....	0.010969	0.013136
max eCDF diff.....	0.052097	0.027915
var ratio (Tr/Co).....	1.063	0.96571
T-test p-value.....	0.41991	0.2525
KS Bootstrap p-value..	0.38	0.7
KS Naive p-value.....	0.87308	0.9716
KS Statistic.....	0.052097	0.027915

***** (V4) pcged97 *****

	Before Matching	After Matching
mean treatment.....	11.617	12.252
mean control.....	12.974	12.64
std mean diff.....	-87.609	-26.761
mean raw eQQ diff.....	1.3457	0.39409
med raw eQQ diff.....	1	0
max raw eQQ diff.....	3	2
mean eCDF diff.....	0.12333	0.035826
med eCDF diff.....	0.11269	0.01642
max eCDF diff.....	0.31436	0.10837
var ratio (Tr/Co).....	0.74142	0.67088
T-test p-value.....	< 2.22e-16	< 2.22e-16
KS Bootstrap p-value..	< 2.22e-16	< 2.22e-16
KS Naive p-value.....	1.4785e-11	0.0015655
KS Statistic.....	0.31436	0.10837

***** (V5) mratio96 *****

	Before Matching	After Matching
mean treatment.....	1.0688	1.7233
mean control.....	3.0708	2.5984
std mean diff.....	-199.25	-64.343
mean raw eQQ diff.....	2.0076	0.87804
med raw eQQ diff.....	1.8335	0.59461
max raw eQQ diff.....	18.554	18.554
mean eCDF diff.....	0.36023	0.14385
med eCDF diff.....	0.41917	0.15435
max eCDF diff.....	0.5961	0.26273

var ratio (Tr/Co).....	0.14627	0.33929
T-test p-value.....	< 2.22e-16	< 2.22e-16
KS Bootstrap p-value..	< 2.22e-16	< 2.22e-16
KS Naive p-value.....	< 2.22e-16	< 2.22e-16
KS Statistic.....	0.5961	0.26273

***** (V6) pcg_adc *****

	Before Matching	After Matching
mean treatment.....	1.8298	0.90759
mean control.....	0.50478	0.86304
std mean diff.....	52.163	2.3089

mean raw eQQ diff.....	1.3298	0.073892
med raw eQQ diff.....	0	0
max raw eQQ diff.....	6	1

mean eCDF diff.....	0.16563	0.0092365
med eCDF diff.....	0.17968	0.0082102
max eCDF diff.....	0.26631	0.026273

var ratio (Tr/Co).....	3.5102	1.0709
T-test p-value.....	1.2444e-10	0.079201
KS Bootstrap p-value..	< 2.22e-16	0.38
KS Naive p-value.....	2.0539e-08	0.98456
KS Statistic.....	0.26631	0.026273

Before Matching Minimum p.value: < 2.22e-16

Variable Name(s): black pcged97 mratio96 pcg_adc Number(s): 2 4 5 6

After Matching Minimum p.value: < 2.22e-16

Variable Name(s): pcged97 mratio96 Number(s): 4 5