

Matching Estimators in R

Peter Sun

September 31, 2021

Contents

1	Load Packages	2
2	Description of Dataset	2
3	Load Data	2
4	Breusch-Pagan Test for Heteroskedasticity	3
5	Matching Estimators	4
5.1	Define Outcome (Y), Treatment Index (Tr), and Variables to Match On (X)	4
5.2	Get Estimators Individually	4
5.3	Get All Estimators	5
5.4	Specify Variables in the Bias-Corrected Matching	6
6	Bonus: Check Covariate Balance	7

1 Load Packages

The Breusch-Pagan test against heteroskedasticity requires the `lmtest` package, and a variety of matching estimators are implemented in the `Matching` package.

```
library(tidyverse)
library(haven)
library(lmtest)
library(Matching)
library(kableExtra)
library(cobalt)
select <- dplyr::select
```

2 Description of Dataset

This example uses the 1997 Child Development Supplement (CDS) to the Panel Study of Income Dynamics (PSID) and the core PSID annual data from 1968 to 1997.

The dependent variable in this dataset is `pcss97`, a passage comprehension score. Higher scores on this measure indicate higher academic achievement. The treatment variable is `kuse` or children who ever used Aid to Families With Dependent Children (AFDC). The covariates or matching variables are:

- `male`: Child's Gender: Male (Reference: Female)
- `black`: Child's Race: African American (Reference: Other)
- `age97`: Child's Age in 1997
- `pcged97`: Caregiver's Education in 1997 (Years of Schooling)
- `mratio96`: Ratio of Family Income to Poverty Line in 1996
- `pcg_adc`: Caregiver's History of Using Welfare (Number of Years)

3 Load Data

```
d <- read_dta("data/cds_pcss97.dta") %>%
  haven::zap_formats() %>%
  sjlabelled::remove_all_labels() %>%
  as_tibble()
head(d) %>%
  kbl(booktabs = T, linesep = "", digits = 2) %>%
  kable_styling(position = "center") %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

kid	pcg_id	age97	pcss97	k_adc	pcged97	pcg_adc	mratio96	black	male	kuse
4180	4179	12	93	7.69	9	0	0.41	0	0	1
5032	5170	12	122	0.00	12	0	5.00	0	0	0
7041	7030	6	93	0.00	11	2	1.07	0	1	0
10033	10177	7	101	8.33	12	0	2.40	1	0	1
10034	10006	6	127	0.00	14	0	1.00	1	0	0
14030	14172	11	103	0.00	12	0	3.77	0	1	0

4 Breusch-Pagan Test for Heteroskedasticity

The homoskedastic variance estimator assumes that the unit-level treatment effect is constant and that the conditional variance of $Y_i(w)$ given X_i does not vary with either covariates or the treatment. The Breusch-Pagan test is performed for each of the seven independent variables and showed that child's age was statistically significant, $p < .01$, indicating that the conditional variance of the dependent variable was not constant across levels of child's age. Therefore, the robust variance estimator should be used in the next stage to allow for heteroskedasticity.

```
# Regress outcome on treatment and matching variables using OLS
m0 <- lm(pc97 ~ kuse + male + black + age97 + pcged97 + mratio96 + pcg_adc, data = d)

# Function for Bresuch-Pagan test
get_bpctest <- function(data, lm.model, var) {
  b <- lmtest::bptest(
    lm.model,
    as.formula(paste0("~", var)),
    data = data,
    studentize = F
  )
  return(
    tibble(
      variable = var,
      statistic = b$statistic,
      df = b$parameter,
      p.value = b$p.value
    )
  )
}

# Get results from Bresuch-Pagan test
map_dfr(
  c("kuse", "male", "black", "age97", "pcged97", "mratio96", "pcg_adc"),
  get_bpctest,
  data = d,
  lm.model = m0
) %>%
  kbl(booktabs = T, linesep = "",
      caption = "Breusch-Pagan Test Results") %>%
  kable_styling(position = "center") %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 1: Breusch-Pagan Test Results

variable	statistic	df	p.value
kuse	0.0096209	1	0.9218637
male	3.0708544	1	0.0797076
black	1.4344983	1	0.2310317
age97	25.2698389	1	0.0000005
pcged97	0.1890527	1	0.6637062
mratio96	0.6182384	1	0.4317027
pcg_adc	0.2910502	1	0.5895483

5 Matching Estimators

Of the six matching variables, four are continuous and two are categorical, therefore bias-corrected matching estimator is necessary to correct for bias corresponding to the matching discrepancies between matched units and their matches on the four continuous covariates. Regression adjustment can be used with the `BiasAdjust = T` argument.

By default, the `Matching::Match()` function uses the matching variables to make bias adjustments. However, these covariates can be specified using the `Z` argument (example shown below).

If `Var.calc = 0`, then homoskedasticity is assumed. Use `Var.calc = 4` to request the robust variance estimator using four matches.

The `estimand` argument is by default “ATT,” but can be set to the “ATE” or “ATC”. The `sample` argument is a logical flag for whether the population or sample variance is returned.

Results from the `Matching::Match()` function are identical to Stata’s `nnmatch` program.

5.1 Define Outcome (Y), Treatment Index (Tr), and Variables to Match On (X)

```
Y <- d$pcss97
Tr <- d$kuse
X <- select(d, male, black, age97, pcged97, mratio96, pcg_adc)
```

5.2 Get Estimators Individually

Note that by default matching is done with replacement. However, this can be changed with the `replace` logical flag.

```
# Sample Average Treatment Effect (SATE)
summary(Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,
  estimand = "ATE", sample = T, replace = T)) # replace = T is default

# Population Average Treatment Effect (PATE)
summary(Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,
  estimand = "ATE", sample = F))

# Sample average treatment effect for the treated (SATT)
summary(Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,
  estimand = "ATT", sample = T))

# Population average treatment effect for the treated (PATT)
summary(Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,
  estimand = "ATT", sample = F))

# Sample average treatment effect for the controls (SATC)
summary(Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,
  estimand = "ATC", sample = T))

# Population average treatment effect for the controls (PATC)
summary(Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,
  estimand = "ATC", sample = F))
```

5.3 Get All Estimators

```
# Function for extracting estimate, SE, t-stat, and p-value from Match()
get_match <- function(estimand, sample, Z = X) {
  m <- Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = TRUE, Var.calc = 4,
    estimand = estimand, sample = sample)
  return(list(
    est = m$est[,1],
    se = m$se,
    t.stat = m$est[,1]/m$se,
    p = (1 - pnorm(abs(m$est[,1]/m$se))) * 2
  ))
}

# Estimate different matching estimators
tribble(
  ~estimator, ~estimand, ~sample,
  "SATE", "ATE", T,
  "PATE", "ATE", F,
  "SATT", "ATT", T,
  "PATT", "ATT", F,
  "SATC", "ATC", T,
  "PATC", "ATC", F
) %>%
  rowwise() %>%
  mutate(match = list(get_match(estimand, sample))) %>%
  unnest_wider(match) %>%
  select(-estimand, -sample) %>%
  kbl(booktabs = T, linesep = "",
    caption = "Bias-Corrected Matching Estimators with Robust Standard Errors") %>%
  kable_styling(position = "center") %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 2: Bias-Corrected Matching Estimators with Robust Standard Errors

estimator	est	se	t.stat	p
SATE	-4.703773	1.769696	-2.657956	0.0078616
PATE	-4.703773	1.765187	-2.664746	0.0077047
SATT	-5.229651	1.781217	-2.935999	0.0033248
PATT	-5.229651	1.720590	-3.039451	0.0023701
SATC	-4.467254	2.133536	-2.093827	0.0362754
PATC	-4.467254	2.135647	-2.091757	0.0364602

When looking at the ATE, the results suggest that childhood poverty strongly affected children's academic achievement. Children who used AFDC had a passage comprehension score 4.7 units lower than that of children who never used AFDC. When considering only the treated population, this treatment effect was much larger.

5.4 Specify Variables in the Bias-Corrected Matching

```
# Sample Average Treatment Effect (SATE)
Z <- select(d, age97, pcged97, mratio96, pcg_adc) # Covariates for bias adj.
summary(Match(Y = Y, Tr = Tr, X = X, Z = Z, M = 4,
             BiasAdjust = T, Var.calc = 4, estimand = "ATE", sample = T))

##
## Estimate... -4.4867
## AI SE..... 1.7697
## T-stat..... -2.5353
## p.val..... 0.011235
##
## Original number of observations..... 606
## Original number of treated obs..... 188
## Matched number of observations..... 606
## Matched number of observations (unweighted). 2441
```

6 Bonus: Check Covariate Balance

The `Matching::Match()` function works well in conjunction with the `cobalt::bal.tab()` function for checking covariate balance.

By default, the denominator for standardized mean differences uses a pooled estimate (square root of the average of the group variances) for ATE and the standard deviation of the treated group for ATT—both standard deviations are computed using the sample before matching. This option can also be manually set with the `s.d.denom` option.

```
# Covariate Balance for SATE
SATE <- Match(Y = Y, Tr = Tr, X = X, M = 4, BiasAdjust = T, Var.calc = 4,
             estimand = "ATE", sample = T)
bal.tab(SATE, kuse ~ male + black + age97 + pcged97 + mratio96 + pcg_adc, data = d,
        abs = T,
        un = T,
        thresholds = c(m = .1),
        s.d.denom = "pooled")
```

```
## Balance Measures
##           Type Diff.Un Diff.Adj      M.Threshold
## male      Binary 0.0239  0.0140    Balanced, <0.1
## black     Binary 0.4318  0.2999 Not Balanced, >0.1
## age97     Contin. 0.0713  0.0687    Balanced, <0.1
## pcged97   Contin. 0.8084  0.6257 Not Balanced, >0.1
## mratio96  Contin. 1.0066  0.7949 Not Balanced, >0.1
## pcg_adc   Contin. 0.6508  0.5069 Not Balanced, >0.1
##
## Balance tally for mean differences
##           count
## Balanced, <0.1      2
## Not Balanced, >0.1  4
##
## Variable with the greatest mean difference
## Variable Diff.Adj      M.Threshold
## mratio96  0.7949 Not Balanced, >0.1
##
## Sample sizes
##           Control Treated
## All           418.      188
## Matched (ESS)  318.46    188
## Matched (Unweighted) 418.      188
```