

# Package ‘wdsmatch’

April 21, 2026

**Type** Package

**Title** Weighted Double Score Matching for Survey-Weighted Causal Inference

**Version** 0.1.1

**Description** Implements weighted double score matching (WDSM) for estimating population-level causal effects from complex survey data. Combines propensity scores and prognostic scores with survey design weights for matching, survey-weighted imputation within match sets, and Hajek normalization to target the population average treatment effect (PATE) and the population average treatment effect on the treated (PATT). Supports both retrospective (treatment-dependent) and prospective (treatment-independent) sampling designs. Achieves double robustness: consistent estimation when either the propensity score or prognostic score model is correctly specified. Provides polynomial sieve bias correction and linearization-based multinomial bootstrap variance estimation that preserves the survey-weighted matching structure without re-matching. Methods are described in Zeng, Tong, Tong, Lu, Mukherjee, and Li (2026, under review) ``Where to weight? Estimating population causal effects with weighted double score matching in complex surveys".

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.3.2

**Depends** R (>= 3.5.0)

**Imports** stats

**Suggests** testthat (>= 3.0.0), knitr, rmarkdown

**Config/testthat/edition** 3

**URL** <https://github.com/ykzeng-yale/wdsmatch>

**BugReports** <https://github.com/ykzeng-yale/wdsmatch/issues>

**NeedsCompilation** no

**Author** Yukang Zeng [aut, cre],  
 Guangyu Tong [aut],  
 Jiaqi Tong [aut],  
 Haidong Lu [aut],  
 Bhramar Mukherjee [aut],  
 Fan Li [aut]

**Maintainer** Yukang Zeng <ykzeng2019@gmail.com>

**Repository** CRAN

**Date/Publication** 2026-04-21 19:00:02 UTC

## Contents

print.wdsmatch . . . . .	2
summary.wdsmatch . . . . .	3
survey_obs . . . . .	3
wdsmatchATE . . . . .	4
wdsmatchATT . . . . .	6

**Index** **9**

---

print.wdsmatch	<i>Print method for wdsmatch objects</i>
----------------	--

---

## Description

Print method for wdsmatch objects

## Usage

```
## S3 method for class 'wdsmatch'
print(x, digits = 4, ...)
```

## Arguments

x	A wdsmatch object returned by <code>wdsmatchATE</code> or <code>wdsmatchATT</code> .
digits	Number of significant digits.
...	Additional arguments (ignored).

## Value

Invisibly returns the input object `x`. Called for its side effect of printing a formatted summary to the console, including the point estimate, standard error, confidence interval, number of matches, and sample sizes.

---

summary.wdsmatch	<i>Summary method for wdsmatch objects</i>
------------------	--

---

**Description**

Summary method for wdsmatch objects

**Usage**

```
## S3 method for class 'wdsmatch'
summary(object, ...)
```

**Arguments**

object	A wdsmatch object returned by <a href="#">wdsmatchATE</a> or <a href="#">wdsmatchATT</a> .
...	Additional arguments (ignored).

**Value**

Invisibly returns the input object. Called for its side effect of printing a formatted summary to the console.

---

survey_obs	<i>Simulated Survey Observational Data</i>
------------	--

---

**Description**

A simulated dataset drawn from a survey-weighted observational study with treatment-dependent (retrospective) sampling. Contains 6 covariates, a binary treatment indicator, observed outcome, and survey design weights. The true PATE is approximately 0.8 and the true PATT approximately 1.0.

**Usage**

```
survey_obs
```

**Format**

A data frame with approximately 120 rows and 9 variables:

**Y** Observed outcome (continuous).

**Z** Binary treatment indicator (1 = treated, 0 = control).

**X1, X2, X3, X4, X5, X6** Pre-treatment covariates (continuous). The true propensity and prognostic models include an X1 : X2 interaction.

**survey\_weight** Survey design weight (inverse selection probability).

## Details

Generated by a simulation where:

- Treatment assignment:  $P(Z = 1|X) = \text{logit}^{-1}(0.3 + 0.6X_1 + 0.4X_2 - 0.3X_3 + 0.2X_1X_2)$ .
- Outcome model:  $Y(0) = 1 + X_1 + 0.5X_2 - 0.3X_3 + 0.2X_4 + 0.3X_1X_2 + \varepsilon$ , with treatment effect  $\tau(X) = 0.8 + 0.2X_1$ .
- Survey selection: treatment-dependent (retrospective) with  $P(S = 1|Z, X) = \text{logit}^{-1}(-2 + 0.3Z + 0.2X_1 + 0.15X_2)$ .

## Source

Simulated data; see `data-raw/make_survey_data.R`.

## Examples

```
data(survey_obs)
head(survey_obs)

# Estimate PATE
fit <- wdsmatchATE(Y = survey_obs$Y, X = survey_obs[, 3:8],
                  Z = survey_obs$Z, weights = survey_obs$survey_weight,
                  M = 3, varest = FALSE)

fit
```

---

wdsmatchATE	<i>Weighted Double Score Matching Estimator for Population Average Treatment Effect</i>
-------------	---

---

## Description

Estimates the population average treatment effect (PATE) using weighted double score matching (WDSM) with survey design weights. The method matches treated and control units on arm-specific double scores  $D_z(X) = (e(X), \text{Psi}_z(X))$  for  $z$  in  $\{0,1\}$ , imputes missing potential outcomes via survey-weighted averaging within match sets, and aggregates using Hajek normalization. Polynomial sieve bias correction removes the finite-sample matching discrepancy. Variance estimation uses a linearization-based multinomial bootstrap that re-estimates score parameters while preserving the original matching structure and survey-weighted reuse frequencies.

## Usage

```
wdsmatchATE(
  Y,
  X,
  Z,
  weights,
  M = 5,
```

```

ps = NULL,
pg = NULL,
model.ps = NULL,
model.pg = NULL,
sampling = c("retrospective", "prospective"),
use.bias.correction = TRUE,
varest = TRUE,
boots = 200,
alpha = 0.05
)

```

### Arguments

<code>Y</code>	Numeric vector of observed outcomes.
<code>X</code>	Numeric matrix or data frame of covariates.
<code>Z</code>	Binary treatment assignment indicator (1 = treated, 0 = control).
<code>weights</code>	Numeric vector of survey design weights. Required.
<code>M</code>	Number of nearest neighbors for matching (default 5).
<code>ps</code>	Numeric vector of pre-estimated propensity scores. If NULL (default), estimated internally using <code>model.ps</code> .
<code>pg</code>	Numeric matrix of pre-estimated prognostic scores with columns <code>psi0</code> (control) and <code>psi1</code> (treated). If NULL (default), estimated internally using <code>model.pg</code> .
<code>model.ps</code>	Formula for propensity score model (e.g., $Z \sim X1 + X2$ ). If NULL, uses all columns of <code>X</code> .
<code>model.pg</code>	Formula for prognostic score model (e.g., $Y \sim X1 + X2$ ). If NULL, uses all columns of <code>X</code> .
<code>sampling</code>	Character: "retrospective" (default) for treatment-dependent sampling (survey-weighted PS estimation), or "prospective" for treatment-independent sampling (unweighted PS estimation).
<code>use.bias.correction</code>	Logical: apply polynomial sieve bias correction (default TRUE).
<code>varest</code>	Logical: compute bootstrap variance estimate and confidence interval (default TRUE).
<code>boots</code>	Number of multinomial bootstrap replicates (default 200).
<code>alpha</code>	Significance level for confidence intervals (default 0.05).

### Details

The estimator achieves double robustness: it is consistent when either the propensity score model or the prognostic score model is correctly specified.

Under retrospective sampling (`sampling = "retrospective"`), the propensity score is estimated with survey weights to recover the population-level treatment assignment mechanism. Under prospective sampling (`sampling = "prospective"`), the propensity score is estimated without survey weights. Prognostic scores are always estimated without survey weights, as the conditional outcome mean is invariant to the sampling design.

The sieve basis uses log-odds of the propensity score to match the coordinate system used in matching distance computation.

### Value

A list with components:

estimate	Point estimate of PATE.
se	Bootstrap standard error (if varest = TRUE).
ci	Confidence interval as c(lower, upper) (if varest = TRUE).
boot.estimates	Vector of bootstrap replicate estimates (if varest = TRUE).
M	Number of matches used.
n	Sample size.
n.treated	Number of treated units.
n.control	Number of control units.
call	The matched call.

### Examples

```
data(survey_obs)
fit <- wdsmatchATE(
  Y = survey_obs$Y,
  X = survey_obs[, c("X1", "X2", "X3", "X4", "X5", "X6")],
  Z = survey_obs$Z,
  weights = survey_obs$survey_weight,
  M = 3,
  model.ps = Z ~ X1 + X2 + X3 + X4 + X5 + X6 + X1:X2,
  model.pg = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X1:X2,
  sampling = "retrospective",
  varest = FALSE
)
fit
```

---

wdsmatchATT

*Weighted Double Score Matching Estimator for Population Average Treatment Effect on the Treated*

---

### Description

Estimates the population average treatment effect on the treated (PATT) using weighted double score matching (WDSM) with survey design weights. Performs one-sided matching from treated to control units on the control-side double score  $D_0(X) = (e(X), \Psi_0(X))$ . Only the counterfactual control outcome  $Y(0)$  needs imputation; treated outcomes are directly observed. Aggregation uses Hajek normalization over treated-side survey weights. Polynomial sieve bias correction and linearization-based multinomial bootstrap with survey-weighted reuse frequencies are applied. PATT requires only one-sided unconfoundedness:  $Y(0)$  independent of  $Z$  given  $X$ .

**Usage**

```
wdsmatchATT(
  Y,
  X,
  Z,
  weights,
  M = 5,
  ps = NULL,
  pg = NULL,
  model.ps = NULL,
  model.pg = NULL,
  sampling = c("retrospective", "prospective"),
  use.bias.correction = TRUE,
  varest = TRUE,
  boots = 200,
  alpha = 0.05
)
```

**Arguments**

Y	Numeric vector of observed outcomes.
X	Numeric matrix or data frame of covariates.
Z	Binary treatment assignment indicator (1 = treated, 0 = control).
weights	Numeric vector of survey design weights. Required.
M	Number of nearest neighbors for matching (default 5).
ps	Numeric vector of pre-estimated propensity scores. If NULL (default), estimated internally using <code>model.ps</code> .
pg	Numeric matrix of pre-estimated prognostic scores with columns <code>psi0</code> (control) and <code>psi1</code> (treated). If NULL (default), estimated internally using <code>model.pg</code> .
<code>model.ps</code>	Formula for propensity score model (e.g., $Z \sim X1 + X2$ ). If NULL, uses all columns of X.
<code>model.pg</code>	Formula for prognostic score model (e.g., $Y \sim X1 + X2$ ). If NULL, uses all columns of X.
sampling	Character: "retrospective" (default) for treatment-dependent sampling (survey-weighted PS estimation), or "prospective" for treatment-independent sampling (unweighted PS estimation).
<code>use.bias.correction</code>	Logical: apply polynomial sieve bias correction (default TRUE).
varest	Logical: compute bootstrap variance estimate and confidence interval (default TRUE).
boots	Number of multinomial bootstrap replicates (default 200).
alpha	Significance level for confidence intervals (default 0.05).

**Value**

A list with components:

estimate	Point estimate of PATT.
se	Bootstrap standard error (if varest = TRUE).
ci	Confidence interval as c(lower, upper) (if varest = TRUE).
boot.estimates	Vector of bootstrap replicate estimates (if varest = TRUE).
M	Number of matches used.
n	Sample size.
n.treated	Number of treated units.
n.control	Number of control units.
call	The matched call.

**Examples**

```
data(survey_obs)
fit <- wdsmatchATT(
  Y = survey_obs$Y,
  X = survey_obs[, c("X1", "X2", "X3", "X4", "X5", "X6")],
  Z = survey_obs$Z,
  weights = survey_obs$survey_weight,
  M = 3,
  model.ps = Z ~ X1 + X2 + X3 + X4 + X5 + X6 + X1:X2,
  model.pg = Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X1:X2,
  sampling = "retrospective",
  varest = FALSE
)
fit
```

# Index

## \* datasets

survey\_obs, 3

print.wdsmatch, 2

summary.wdsmatch, 3

survey\_obs, 3

wdsmatchATE, 2, 3, 4

wdsmatchATT, 2, 3, 6