Package 'CompCausal'

November 28, 2026

Type Package

Title Inferring Causal Effects in Comprehensive Cohort Studies

Version 0.1.0.5

Depends R (>= 4.1.0)

Description Implements semiparametric methods for estimating comprehensive cohort causal effects in comprehensive cohort studies with unmeasured confounding in the observational arm and outcomes that may be missing at random. The package provides tools for estimating comprehensive cohort causal effects (CCCE), randomized trial causal effects (RTCE), and patient preference causal effects (PPCE). For methodological details, see Xu S, Nabi R, Underwood M, Scharfstein D (2026). ``Inferring Comprehensive Cohort Causal Effects in the Presence of Unmeasured Confounding and Missing Outcomes." <doi:10.48550/arXiv.2606.28741>.

URL https://github.com/UofUEpiBio/CompCausal, https://uofuepibio.github.io/CompCausal/

BugReports https://github.com/UofUEpiBio/CompCausal/issues

License MIT + file LICENSE

Encoding UTF-8

Imports utils, splines, stats, betareg, methods, dplyr, ggplot2, MAVE, assertthat, dfoptim, ManifoldOptim, mgcv, purrr, Rcpp, rlang, matrixStats

Suggests tinytest, quarto, knitr

LazyData true

VignetteBuilder quarto

Config/roxygen2/version 8.0.0

RoxygenNote 8.0.0

NeedsCompilation no

Author Shiyao Xu [aut, cre] (ORCID: <https://orcid.org/0000-0002-1746-5977>), George Vega Yon [ctb] (ORCID: <https://orcid.org/0000-0002-3171-0844>), Daniel Scharfstein [pdr, aut] (ORCID: <https://orcid.org/0000-0001-7482-9653>)

Maintainer Shiyao Xu <jennifer.xu@hsc.utah.edu>

Repository CRAN

Date/Publication 2026-11-28 15:20:33 UTC

Contents

ccohort 2
est_psi_exchange 3
print.est_psi 5

Index **8**

`ccohort` *Simulated Data for the Comprehensive Cohort*

Description

Simulated Data for the Comprehensive Cohort

Usage

```
ccohort
```

Format

A data frame wtih 563 observations and 8 variables:

Y Continuous outcome ranging 0-100. Missing outcome as NA.

M Binary missingness indicator for Y: M=1 if Y=Y_obs, M=0 if Y=NA.

R Binary randomization consent indicator (1 for RCT, 0 for PPS).

t Binary treatment assignment.

age Continuous baseline covariate.

womac_bq Baseline WOMAC pain score, ranging 0-100.

expectationb Baseline covariate: Expected pain a year from now (much/a little worse, about the same, a little/much better/free of pain)

ChronicPainb Baseline covariate: Chronic pain grade (grade 1-2, grade 3-4)

Examples

```
data(ccohort)
```

`est_psi_exchange` *One-step, split sample estimator for E[Y(t)], E[Y(t)|R=0], under sensitivity analysis for exchangability assumption*

Description

One-step, split sample estimator for E[Y(t)], E[Y(t)|R=0], under sensitivity analysis for exchangability assumption

Usage

```
est_psi_exchange(
  Y,
  M,
  R,
  X,
  t,
  trt,
  gamma,
  fold,
  seed,
  IF_output,
  simple_trunc,
  quant,
  kernel,
  method = "optim",
  single_index_method,
  use_mave = TRUE,
  s_t_y = NULL
)
```

Arguments

`Y`	Numeric outcome vector. Missing values are internally replaced with `0` prior to model fitting.
`M`	Binary indicator for observed outcome (`1` = observed, `0` = missing).
`R`	Binary group indicator used to stratify nuisance and outcome models.
`X`	Data frame or matrix of baseline covariates.
`t`	Treatment assignment vector.
`trt`	Treatment level for which the target estimand is computed.
`gamma`	Numeric vector of sensitivity parameters.
`fold`	Number of cross-fitting folds.
`seed`	Optional integer random seed for fold assignment. Use `NULL` to leave RNG state unchanged.

IF_output	Logical; if TRUE, include influence-function vectors in the returned list.
simple_trunc	Logical; if TRUE, apply quantile truncation to inverse probability weights. If FALSE, apply IF truncation diagnostics.
quant	Numeric in (0, 1) used as the upper quantile for simple weight truncation when simple_trunc = TRUE.
kernel	Characters; Kernel used for SIMs. K2_Biweight for Epanechnikov kernel, dnorm for Gaussian kernel.
method	Characters; Optimization method used for SIMs. Choices are: optim, nlminb, nmk. Note that method is set to optim if single_index_method=norm1coef.
single_index_method	Characters; Three implementations for SIMs: fixed_bandwidth for setting bandwidth to 1, fixed_coef for setting the first coefficient to 1, and norm1coef for setting the norm of coefficients to 1.
use_mave	Logical; if TRUE, use Minimum Average Variance Estimation (MAVE) method for initial coefficients value for SIMs. If FALSE, use sliced inverse regression. Default is TRUE.
s_t_y	A function of Y in the exponential tilting model. If NULL, s_t_y is set to pnorm((y

Value

A named list of estimates and uncertainty summaries for each value in gamma. Core elements include point estimates (est, est_R0), variance estimates (var, var_R0), and confidence interval bounds (lowerCI*, upperCI*). Additional components depend on simple_trunc and IF_output:

- simple_trunc = TRUE: returns quantile-weight-truncated summaries only.
- simple_trunc = FALSE: additionally returns truncated summaries and truncated IF objects when requested.
- IF_output = TRUE: includes influence-function lists (IF*) and, when relevant, truncated IF lists (IF_trunc*).

Examples

```
library(CompCausal)
data(ccohort)
out <- with(ccohort, {
        est_psi_exchange(Y, M, R, X = data.frame(age, womac_bq, expectationb, ChronicPainb),
            t, trt = 1, gamma = c(0, 0.5), fold = 5, seed = 1, IF_output = FALSE,
            simple_trunc = TRUE, quant = 0.99, kernel="dnorm",
            single_index_method="norm1coef", method="optim")})
```

print.est_psi *One-step, split sample estimator for E[Y(t)], E[Y(t)|R=1], E[Y(t)|R=0],*

Description

Estimates study-specific and overall outcome means (and difference from baseline) using cross-fitting with single index models (SIMs) and nuisance models for treatment and outcome missingness (`mgcv::gam`). The function also computes influence-function-based variances, confidence intervals, and optional truncated influence-function diagnostics.

Usage

```
## S3 method for class 'est_psi'
print(x, rounding = 3, ...)

print_effects(object_t1, object_t0, rounding = 3, ...)

est_psi(
  Y,
  M,
  R,
  X,
  t,
  trt,
  gamma,
  fold,
  seed,
  IF_output,
  simple_trunc,
  quant,
  kernel,
  single_index_method,
  method = "optim",
  use_mave = TRUE,
  s_t_y = NULL,
  coef_g.fit = NULL,
  coef_t_R0.fit = NULL,
  coef_t_R1.fit = NULL,
  coef_M_R0.fit = NULL,
  coef_M_R1.fit = NULL
)
```

Arguments

x	An object of class `est_psi`, as returned by `est_psi`.

rounding	Integer; number of decimal places to round estimates and variances in the printed output. Default is 3.
...	Additional arguments (not currently used).
object_t1	An object of class est_psi for the treatment group (e.g., trt = 1), as returned by est_psi.
object_t0	An object of class est_psi for the control group (e.g., trt = 0), as returned by est_psi.
Y	Numeric outcome vector. Missing values are internally replaced with 0 prior to model fitting.
M	Binary indicator for observed outcome (1 = observed, 0 = missing).
R	Binary randomization consent indicator (1 for RCT, 0 for PPS).
X	Data frame or matrix of baseline covariates.
t	Treatment assignment vector.
trt	Treatment level for which the target estimand is computed.
gamma	Numeric vector of sensitivity parameters.
fold	Number of cross-fitting folds.
seed	Optional integer random seed for fold assignment. Use NULL to leave RNG state unchanged.
IF_output	Logical; if TRUE, include influence-function vectors in the returned list.
simple_trunc	Logical; if TRUE, apply quantile truncation to inverse probability weights. If FALSE, apply IF truncation diagnostics.
quant	Numeric in (0, 1) used as the upper quantile for simple weight truncation when simple_trunc = TRUE.
kernel	Characters; Kernel used for SIMs. K2_Biweight for Epanechnikov kernel, dnorm for Gaussian kernel.
single_index_method	Characters; Three implementations for SIMs: fixed_bandwidth for setting bandwidth to 1, fixed_coef for setting the first coefficient to 1, and norm1coef for setting the norm of coefficients to 1.
method	Characters; Optimization method used for SIMs. Choices are: optim, nlminb, nmk. Note that method is set to optim if single_index_method=norm1coef.
use_mave	Logical; if TRUE, use Minimum Average Variance Estimation (MAVE) method for initial coefficients value for SIMs. If FALSE, use sliced inverse regression. Default is TRUE.
s_t_y	A function of Y in the exponential tilting model. If NULL, s_t_y is set to pnorm((y-60)/25).
coef_g.fit	Optional starting values for a treatment model; currently retained for interface compatibility.
coef_t_R0.fit	Optional starting coefficients for treatment model fit in t=trt and R = 0stratum.
coef_t_R1.fit	Optional starting coefficients for treatment model fit in t=trt and R = 1 stratum.
coef_M_R0.fit	Optional starting coefficients for missingness model fit in R = 0 stratum.
coef_M_R1.fit	Optional starting coefficients for missingness model fit in R = 1 stratum.

Details

The procedure uses sample-splitting/cross-fitting to reduce overfitting bias in nuisance estimation. Outcome regressions are fit with single index models and then integrated over estimated conditional distributions to obtain conditional means and sensitivity-adjusted moments.

Value

A named list of class `est_psi` that contains estimates and uncertainty summaries for each value in `gamma`. Core elements include point estimates (`est`, `est_R1`, `est_R0`), variance estimates (`var`, `var_R1`, `var_R0`), and confidence interval bounds (`lowerCI*`, `upperCI*`). Additional components depend on `simple_trunc` and `IF_output`:

- `simple_trunc = TRUE`: returns quantile-weight-truncated summaries only.
- `simple_trunc = FALSE`: additionally returns truncated summaries and truncated IF objects when requested.
- `IF_output = TRUE`: includes influence-function lists (`IF*`) and, when relevant, truncated IF lists (`IF_trunc*`).

Examples

```
library(CompCausal)
data(ccohort)
out <- with(ccohort, {
            est_psi(Y, M, R, X = data.frame(age, womac_bq, expectationb, ChronicPainb),
            t, trt = 1, gamma = c(0, 0.5), fold = 5, seed = 1, IF_output = FALSE,
            simple_trunc = TRUE, quant = 0.99, kernel="dnorm",
            single_index_method="norm1coef", method="optim")})
```

Index

∗ **datasets**
ccohort, 2

ccohort, 2

est_psi, *5*, *6*
est_psi *(*print.est_psi*)*, 5
est_psi_exchange, 3

print.est_psi, 5
print_effects *(*print.est_psi*)*, 5