

Package: PND.heter.cluster (via r-universe)

June 1, 2026

Type Package

Title Estimating the Cluster Specific Treatment Effects in Partially Nested Designs

Version 0.1.0

Maintainer Xiao Liu <xiao.liu@austin.utexas.edu>

Description Implements the methods for assessing heterogeneous cluster-specific treatment effects in partially nested designs as described in Liu (2024) <doi:10.1037/met0000723>. The estimation uses the multiply robust method, allowing for the use of machine learning methods in model estimation (e.g., random forest, neural network, and the super learner ensemble). Partially nested designs (also known as partially clustered designs) are designs where individuals in the treatment arm are assigned to clusters (e.g., teachers, tutoring groups, therapists), whereas individuals in the control arm have no such clustering.

Depends R (>= 4.0.0)

Imports stats, mvtnorm, SuperLearner, ranger, xgboost, nnet, origami, boot, tidyverse, dplyr, purrr, magrittr, glue

Suggests testthat, knitr, rmarkdown

URL <https://github.com/xliu12/PND.heter>

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.3.1

Config/pak/sysreqs

cmake libfontconfig1-dev libfreetype6-dev libfribidi-dev make libharfbuzz-dev libicu-dev libjpeg-dev libpng-dev libtiff-dev libuv1-dev libwebp-dev libxml2-dev libssl-dev libx11-dev zlib1g-dev

Repository <https://xliu12.r-universe.dev>

Date/Publication 2025-06-03 13:38:14 UTC

RemoteUrl <https://github.com/xliu12/pnd.heter>

RemoteRef HEAD

RemoteSha 7099a2096dbf6f24a133af82ced0793a1bc981d6

RemoteSubdir PND.heter.cluster

Contents

atekCl	2
balance	4
cluster.specific.ate	5
data_in	7
partially_nested_data_example	8

Index	10
--------------	-----------

atekCl	<i>Estimation of the cluster-specific treatment effects in the partially nested design.</i>
--------	---

Description

Estimation of the cluster-specific treatment effects in the partially nested design.

Usage

```
atekCl(
  data_in,
  tname,
  Kname,
  Yname,
  Xnames,
  Yfamily = "gaussian",
  learners_tt = c("SL.glm"),
  learners_k = c("SL.multinom"),
  learners_y = c("SL.glm"),
  sensitivity = NULL,
  cv_folds = 4L,
  seed = NULL
)
```

Arguments

data_in	A data.frame containing all necessary variables.
tname	[character] A character string of the column name of the treatment variable. The treatment variable should be dummy-coded, with 1 for the (clustered) treatment arm and 0 for the (non-clustered) control arm.

Kname	[character] A character string of the column name of the cluster assignment variable. This variable should be coded as 0 for individuals in the control arm, the arm without the cluster assignment.
Yname	[character] A character string of the column name of the outcome variable
Xnames	[character] A character vector of the column names of the baseline covariates.
Yfamily	[numeric(1)] Variable type of the outcome, with Yfamily = "gaussian" for continuous outcome, and Yfamily = "binomial" for binary outcome.
learners_tt	[character] A character vector of methods for estimating the treatment model, chosen from the SuperLearner R package. Default is "SL.glm", a generalized linear model for the binary treatment variable. Other available methods can be found using the R function SuperLearner::listWrappers().
learners_k	[character] A character string of a method for estimating the cluster assignment model, which can be one of "SL.multinom" (default), "SL.xgboost.modified", "SL.ranger.modified", and "SL.nnet.modified". Default is "SL.multinom", the multinomial regression (nnet::multinom) for the categorical cluster assignment using the treatment arm data. The other options are "SL.xgboost.modified" (gradient boosted model, xgboost::xgboost), "SL.ranger.modified" (random forest model, ranger::ranger), and "SL.nnet.modified" (neural network model, "SL.nnet.modified") modified for fitting categorical response variable of type multinomial.
learners_y	[character] A character vector of methods for estimating the outcome model, chosen from the SuperLearner R package. Default is "SL.glm", a generalized linear model for the outcome variable, with family specified by Yfamily. Other available methods can be found using the R function SuperLearner::listWrappers().
sensitivity	Specification for sensitivity parameter values on the standardized mean difference scale, which can be NULL (default) or "small_to_medium". If NULL, no sensitivity analysis will be run. If "small_to_medium", the function will run a sensitivity analysis for the cluster assignment ignorability assumption, and the sensitivity parameter values indicate a deviation from this assumption of magnitude 0.1 and 0.3 standardized mean difference.
cv_folds	[numeric(1)] The number of cross-fitting folds. Default is 4.
seed	An integer that is used as argument by the set.seed() for offsetting the random number generator. Default is to leave the random number generator alone.

Value

A list containing the following components:

`ate_K` A `data.frame` of the estimation results.
The columns "ate_k", "std_error", "CI_lower", and "CI_upper" contain the estimate, standard error estimate, and lower and upper bounds of the 0.95 confidence interval of the cluster-specific treatment effect for the cluster (indicated by column "cluster") in the same row.

`cv_components` A `data.frame` of nuisance model estimates.

`sens_results` NULL if the argument `sensitivity = NULL`.
If the argument `sensitivity = "small_to_medium"` is specified, `sens_results` is a list of four data frames, containing the estimation results with the sensitivity parameter value (standardized mean difference) being 0.1, 0.3, -0.1, -0.3.

Examples

```
library(tidyverse)
library(SuperLearner)
library(glue)
library(nnet)

# data
data(data_in)
data_in <- data_in

# baseline covariates
Xnames <- c(grep("X_dat", colnames(data_in), value = TRUE))

estimates_ate_K <- PND.heter.cluster::atekCl(
  data_in = data_in,
  tname = "tt", # treatment variable
  Kname = "K", # cluster assignment variable, coded as 0 for
               # individuals in the (non-clustered) control arm
  Yname = "Y", # outcome variable
  Xnames = Xnames,
  seed = 12345
)
estimates_ate_K$ate_K
```

balance	<i>Checking covariate balance based on estimated cluster assignment probabilities (principal score) and treatment assignment probabilities (propensity score).</i>
---------	--

Description

Checking covariate balance based on estimated cluster assignment probabilities (principal score) and treatment assignment probabilities (propensity score).

Usage

```
balance(data_in, atekCl_results, covariate_names = "X_dat.1", tname, Kname)
```

Arguments

`data_in` A data.frame containing all necessary variables.

`atekCl_results` [list]
A list returned from the R function `atekCl()`.

`covariate_names` [character]
A character vector of the column names of the baseline covariates for checking balance.

`tname` [character]
A character string of the column name of the treatment variable. The treatment variable should be dummy-coded, with 1 for the (clustered) treatment arm and 0 for the (non-clustered) control arm.

`Kname` [character]
A character string of the column name of the cluster assignment variable. This variable should be coded as 0 for individuals in the control arm, the arm without the cluster assignment.

Value

A data.frame containing the covariate balance measures (`smd`, standardized mean difference) between each cluster in the treatment arm and the control arm, both before and after the weighting adjustment.

`cluster.specific.ate` *Estimation of the cluster-specific treatment effects in the partially nested design*

Description

Estimation of the cluster-specific treatment effects in the partially nested design

Usage

```
cluster.specific.ate(
  data_in,
  Xnames,
  estimator = c("trt-cluster", "trt-y", "cluster-y", "triple-robust (linear)",
    "triple-robust (dml)"),
  y1model_lme = FALSE,
  randomized.tt = FALSE,
  randomized.tprop = NULL
)
```

Arguments

data_in	A data.frame containing the observed data. In "data_in", column "tt" is the treatment assignment ("tt" is coded as 0 for individuals in the control arm and as 1 for individuals in the treatment arm); column "K" is the cluster assignment in the treatment arm ("K" is coded as 1, 2, ..., J for each individual in the treatment arm with J being the number of clusters, and "K" is coded as 0 for individuals in the control arm); column "Y" is the outcome. The other columns are baseline covariates (X).
Xnames	A character vector of the names of the columns in "data_in" that correspond to baseline covariates (X)
estimator	A character vector of the names of the estimators to use for estimating the cluster-specific treatment effects (ATE _k , k = 1,...,J). The estimators currently supported include those described in Liu (2023), including (i) trt-cluster, (ii) trt-y, (iii) cluster-y, (iv) triply-robust (linear), and (v) triply-robust (dml). The estimators (i)-(iv) are implemented with the parametric models where linear terms of the baseline covariates are included. the estimator (v) triply-robust (dml) is implemented with the double machine learning procedure (Chernozhukov et al., 2018 < https://doi.org/10.1111/ectj.12097 >) with two-fold cross-fitting and data-adaptive packages; specifically, for estimating the cluster assignment probability, the "xgboost" R package is used < https://cran.r-project.org/package=xgboost >; for estimating the treatment probability and the outcome mean, an ensemble of algorithms is used, including boosted trees (via the "xgboost" R package), random forest (via the "ranger" R package < https://cran.r-project.org/package=ranger >), and generalized additive model (via the "gam" R package < https://cran.r-project.org/package=xgboost >.), implemented with the super learner ensembling procedure (via the "SuperLearner" R package < https://cran.r-project.org/package=SuperLearner >).
y1model_lme	Whether to use the random-effects outcome regression for the estimators (ii) trt-y, (iii) cluster-y, and (iv) triply-robust (linear), where the outcome mean is involved. If "y1model_lme = TRUE", the random-intercept outcome regression (with the cluster-mean covariates and cluster-mean centered covariates) will be used; if "y1model_lme = FALSE", the fixed-effects outcome regression will be used.
randomized.tt	Whether the treatment assignment is randomized. If "randomized.tt = TRUE", the treatment probability will be a constant specified by the argument "randomized.ttprop", which is the proportion of individuals randomized to the treatment arm.
randomized.ttprop	The proportion of individuals randomized to the treatment arm.

Examples

```

data(data_in)
data_in <- data_in
Xnames <- c(grep("X_dat", colnames(data_in), value = TRUE))

# estimates_ate_K <- cluster.specific.ate(
# data_in = data_in,
# Xnames = Xnames,

```

```
# estimator = c("trt-cluster",  
# "trt-y",  
# "cluster-y",  
# "triple-robust (linear)",  
# "triple-robust (dml)",  
# y1model_lme = FALSE,  
# randomized.tt = FALSE, randomized.ttprop = NULL  
# )
```

data_in

data_in

Description

A simulated dataset from the 2/1 partially nested design with treatment-included clustering

Usage

data_in

Format

A data frame with 400 rows and 8 variables:

Y Outcome.

K Cluster assignment in the treatment arm.

tt Treatment assignment. 1 for individuals assigned to the treatment arm. 0 for individuals assigned to the control arm. The control arm is unclustered.

X_dat.1 Baseline covariates.

X_dat.2 Baseline covariates.

X_dat.3 Baseline covariates.

X_dat.4 Baseline covariates.

id Individual id.

 partially_nested_data_example

partially_nested_data_example

Description

An example dataset with the 2/1 partially nested design where the clustering is induced by treatment delivery. The example was based on the public-use data of the National Center for Research on Early Childhood Education Teacher Professional Development Study (2007-2011; for details about the study, see this [website](https://www.childandfamilydataarchive.org/cfda/archives/cfda/studies/34848/versions/V2)). The participants were assigned to either the treatment or control arms. The treatment arm was a one-on-one, web-mediated consultancy intervention in which the participants received online coaching from one of $J = 12$ coaches; that is, each coach represents a cluster in this example. The control arm participants had no such clustering.

Usage

partially_nested_data_example

Format

A data frame with 308 rows and 8 variables:

Posttest_Instructional_Support The outcome variable, measuring the instructional support quality after the intervention program.

Coach_ID Coach (i.e., cluster) assignment for participants in the treatment arm.

Intervention_Assignment Treatment assignment. 1 for participants assigned to the treatment arm to receive the intervention program. 0 for participants assigned to the control arm. The control arm is unclustered.

X_gender Baseline covariates.

X_age Baseline covariates.

X_TRace_Black Baseline covariates.

X_TRace_Hispanic Baseline covariates.

X_TRace_White Baseline covariates.

X_Tses_aboveMiddle Baseline covariates.

X_TINTNEED Baseline covariates.

X_Tparedu_aboveHS Baseline covariates.

X_yrs_education Baseline covariates.

X_yrs_teaching_experience Baseline covariates.

X_CLASSPOV Baseline covariates.

X_Cheadstart Baseline covariates.

X_CpublicSCH Baseline covariates.

X_self_efficacy Baseline covariates.
X_pretest_emotional_support Baseline covariates.
X_pretest_organizational_support Baseline covariates.
X_pretest_instructional_support Baseline covariates.
X_extraversion Baseline covariates.
X_agreeableness Baseline covariates.
X_conscientiousness Baseline covariates.
id Participant id.

Index

* datasets

data_in, 7

partially_nested_data_example, 8

atekCl, 2

balance, 4

cluster_specific.ate, 5

data_in, 7

partially_nested_data_example, 8