Package 'longterm'

November 28, 2021

Title Semiparametric Estimation of Long-Term Treatment Effects

Version 0.0.0.9000

Description This package provides code for semiparametric estimation of long-term treatment effects though the combination of short-term experimental and long-term observational datasets. In particular, this package is appropriate for settings in which only short-term outcomes are observed in an experimental data set with exogenously assigned treatment, both short-term and long-term outcomes are observed in an observational data set, where treatment assignment may be confounded, and the researcher is willing to assume that the causal relationships between treatment assignment and the short-term and long-term outcomes share the same unobserved confounding variables in the observational sample. The implementation is based on Chen and Ritzwoller (2021) <arXiv:2107.14405>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

URL https://https://github.com/DavidRitzwoller/longterm

BugReports https://https://github.com/DavidRitzwoller/longterm/issues

Depends R (>= 2.10),
rlang,
caret,
glmnet,
grf,
xgboost,
dplyr,
tidyselect,
haven

Suggests knitr,
rmarkdown

VignetteBuilder knitr

R topics documented:

graduation *Outcomes from Graduation Experiment in Banerejee et. al (2015)*

Description

This dataset contains a subset of the publicly available data from Banerjee et al (2015). It contains the outcomes from a randomized evaluation of the long-term effects of a poverty alleviation program implemented in the Singh region of Pakistan.

Usage

```
graduation
```

Format

A data frame with 854 rows corresponding to households 67 columns corresponding to the variables:

id_hh Household ID

treatment Assignment to Treatment

***_bsl** Pre-treatment covariateds

***_end** Two-year post-treatment outcomes

***_fup** Three-year post-treatment outcomes

Source

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/NHIXNT

References

A. Banerjee et al., Science 348, 1260799 (2015). doi: 10.1126/science.1260799

longterm *Semiparametric Estimation of Long-Term Treatment Effects*

Description

`longterm` estimates the long-term average treatment effect of a binary treatment on a scalar long-term outcome using the semiparametric estimator developed in Chen and Ritzwoller (2021).

Usage

```
longterm(
  data,
  S_vars,
  X_vars,
  Y_var,
  obs,
  estimand,
  type,
  prop_lb = 0.01,
  prop_ub = 0.99,
  alpha = 0.05,
  te_lb = -500,
  te_ub = 500,
  cross_fit_fold = 5,
  nuisance_cv_fold = 5,
  grf_honesty = FALSE,
  grf_tune_parameters = "all",
  grf_num_threads = 1,
  xgb_cv_rounds = 100,
  xgb_eta = 0.1,
  xgb_max_depth = 2,
  xgb_threads = 1
)
```

Arguments

data	A data frame containing the indicators "treatment," denoting whether treatment was assigned, and "observe," denoting whether the observation was from the observational or experimental sample, in addition to pretreatment covariates, short-term outcomes , and long-term outcomes. If an observation is missing values of a particular variable by construction, e.g., long-term outcomes are not observed in the experimental sample, then these values should be coded with any nonmissing value and will not contribute to estimation.
S_vars	A list containing strings of the names of the short-term outcome variables.
X_vars	A list containing strings of the names of the pre-treatment covariates.
Y_var	A string giving the name of the long-term outcome of interest.
obs	A boolean variable specifying whether treatment is observed in the long-term sample.
estimand	A boolean variable specifying whether the estimand of interest is the long-term average treatment effect in the observational population, as opposed to the experimental population.
type	A string specifying how nuisance functions should be estimated. "glmnet" specifies cross-validated lasso. "grf" specifies generalized random forests. "xgboost" specifies XGBoost.
prop_lb	A float specifying the lower threshold for propensity score estimates.
prop_ub	A float specifying the upper threshold for propensity score estimates.
alpha	One minus the nominal coverage probability of the confidence intervals.
te_lb	A float giving a lower bound for the long-term average treatment effect.

`te_ub`	A float giving an upper bound for the long-term average treatment effect.
`cross_fit_fold`	An integer giving the number of folds for the cross-fit estimation of nuisance parameters.
`nuisance_cv_fold`	An interger giving the number of folds for cross-validating nuisance parameter estimates.
`grf_honesty`	A boolean variable setting the `honesty` parameter for `grf` estimation.
`grf_tune_parameters`	A string variable setting the `tune.parameters` parameter for `grf` estimation.
`grf_num_threads`	An integer variable setting the `num.threads` parameter for `grf` estimation.
`xgb_cv_rounds`	An integer variable setting maximum number of rounds for `xgboost` estimation.
`xgb_eta`	A float variable setting `eta` parameter for `xgboost` estimation.
`xgb_max_depth`	A float variable setting `max_depth` parameter for `xgboost` estimation.
`xgb_threads`	An integer variable setting the `nthread` parameter for `xgboost` estimation.

Value

Returns a list with three components:

`hat_tau` Estimate of the long-term average treatment effect.

`se` Estimate of the standard error of the estimator.

`ci` A vector giving the lower and upper points of the confidence interval.

References

Chen, J., & Ritzwoller, D. M. (2021). Semiparametric Estimation of Long-Term Treatment Effects. arXiv preprint arXiv:2107.14405.

Index