Semiparametric Estimation of Long-Term Treatment Effects

Jiafeng Chen and David Ritzwoller

January 16, 2022

The package `longterm` implements the semiparametric estimators of long-term average treatment effects developed in Chen and Ritzwoller [2021].

Setup

We observe two data sets. One data set contains measurements of the short-term outcomes of a randomized evaluation of a binary treatment of interest. The other data set contains observational measurements of the joint distribution of short-term and long-term outcomes. We are interested in estimating the average treatment effect of the treatment on a long-term outcome of interest, and are willing to impose the identifying assumptions specified in Chen and Ritzwoller [2021]. Chen and Ritzwoller [2021] develop a semiparametric estimator of this parameter in this context with the "Double/Debiased Machine Learning" construction of Chernozhukov et al. [2018].

Example

We illustrate the use of the package using a data set from Banerjee et al. [2015]. The dataset is included in the package as the list `graduation`. Run `?graduation` for a full description of the dataset. Further details on the contents and construction of this data set is available in the Supplemental Appendix to Chen and Ritzwoller [2021]. See `?longterm` for full descriptions of the arguments and returned values of `longterm`.

```
library(longterm)
#> Loading required package: rlang
#> Loading required package: caret
#> Loading required package: ggplot2
#> Loading required package: lattice
#> Loading required package: glmnet
#> Loading required package: Matrix
#> Loaded glmnet 4.0-2
#> Loading required package: grf
#> Loading required package: xgboost
#> Loading required package: dplyr
#>
#> Attaching package: 'dplyr'
#> The following object is masked from 'package:xgboost':
#>
```

```
#>     slice
#> The following objects are masked from 'package:stats':
#>
#>     filter, lag
#> The following objects are masked from 'package:base':
#>
#>     intersect, setdiff, setequal, union
#> Loading required package: tidyselect
#> Loading required package: haven

## Assign each observation to `experimental` and `observational` groups randomly
graduation$observe <- as.numeric(runif(nrow(graduation)) <= 0.5)

## Choose Pre-treatment covariates, Short-term Outcomes, and Long-term Outcome
X_vars <- c("ctotal_pcmonth_bsl", "cnonfood_pcmonth_bsl",
            "cfood_pcmonth_bsl",  "cdurable_pcmonth_bsl",
            "asset_index_bsl",    "asset_prod_index_bsl")
S_vars <- c("ctotal_pcmonth_end", "cnonfood_pcmonth_end",
            "cfood_pcmonth_end",  "cdurable_pcmonth_end",
            "asset_index_end",   "asset_prod_index_end")
Y_var  <- "ctotal_pcmonth_fup"

## Assume treatment is observed in the observational sample
# Average effect for the experimental population
est_1_1 <- longterm(graduation, S_vars = S_vars, X_vars = X_vars, Y_var = Y_var,
                    obs = TRUE, estimand = TRUE, type = "glmnet")
# Average effect for the observational population
est_1_0 <- longterm(graduation, S_vars = S_vars, X_vars = X_vars, Y_var = Y_var,
                    obs = TRUE, estimand = FALSE, type = "glmnet")

## Assume treatment is not observed in the observational sample
# Average effect for the experimental population
est_0_1 <- longterm(graduation, S_vars = S_vars, X_vars = X_vars, Y_var = Y_var,
                    obs = FALSE, estimand = TRUE, type = "glmnet")
# Average effect for the observational population
est_0_0 <- longterm(graduation, S_vars = S_vars, X_vars = X_vars, Y_var = Y_var,
                    obs = FALSE, estimand = FALSE, type = "glmnet")

## Display results
kable(data.frame(treatment_observed = c("Yes", "Yes", "No", "No"),
                 subpopulation = c("Observational", "Experimental",
                                   "Observational", "Experimental"),
                 estimate = c(est_1_1$hat_tau, est_1_0$hat_tau,
                              est_0_1$hat_tau, est_0_0$hat_tau),
                 standard_error = c(est_1_1$se, est_1_0$se,
                                    est_0_1$se, est_0_0$se),
                 lower_ci = c(est_1_1$ci[1], est_1_0$ci[1],
```

```
                              est_0_1$ci[1], est_0_0$ci[1]),
                  upper_ci = c(est_1_1$ci[2], est_1_0$ci[2],
                              est_0_1$ci[2], est_0_0$ci[2])))
```

treatment_observed	subpopulation	estimate	standard_error	lower_ci	upper_ci
Yes	Observational	2.610574	5.973235	-9.096752	14.317901
Yes	Experimental	3.516739	5.802154	-7.855274	14.888752
No	Observational	1.348587	2.194615	-2.952781	5.649954
No	Experimental	2.002766	1.958200	-1.835235	5.840767

References

Abhijit Banerjee, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236), 2015.

Jiafeng Chen and David M. Ritzwoller. Semiparametric estimation of long-term treatment effects, 2021.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters, 2018.