

cace_serv_bias.R

Documentation for Running the R Program

August 2022

Peter Z. Schochet (pschochet@mathematica-mpr.com)

cace_serv_bias.R is an R function that estimates treatment effects and standard errors for complier average causal effect (CACE) estimands using inverse probability weighting (IPW) estimators developed in Schochet (Journal of Causal Inference, 2022). The methods pertain to randomized controlled trials (RCTs) with *service receipt bias*, where (i) controls can access services related to the treatment; (ii) the treatment affects the offered services; and (iii) some in the intention-to-treat sample do not receive services (that is, noncompliance exists). The IPW estimators pertain to various complier populations defined by their potential service receipt decisions in the treatment and control conditions. The program allows for both clustered and non-clustered RCTs, and adjusts standard errors for estimation error in the IPW weights from the propensity score logit models predicting service receipt.

INPUTS

Enter inputs using the `c()` function

Variable	Example input	Description
<code>data_csv</code>	<code><- c('serv_rec.csv')</code>	Input csv data file. A complete case analysis is conducted where missing data are excluded for all input variables provided below.
<code>serv_var</code>	<code><- c('r')</code>	Service receipt indicator (1 = received, 0 = not received)
<code>y_var</code>	<code><- c('y','y1')</code>	Outcome variables
<code>trt_var</code>	<code><- c('t')</code>	Treatment indicator (1 = treatment group, 0 = control group)
<code>xt_logit</code>	<code><- c('x1','x2')</code>	Covariates for the service receipt propensity score logit model for the <i>treatment</i> group (covariates are required)
<code>xc_logit</code>	<code><- c('x1','x3')</code>	Covariates for the service receipt propensity score logit model for the <i>control</i> group (covariates are required)
<code>x_wls</code>	<code><- c('x1','x2','x3')</code> or <code>c(0)</code>	Covariates for the weighted least squares models, <code>c(0)</code> = no covariates
<code>clus_var</code>	<code><- c('clus_id')</code> or <code>c(0)</code>	Name of clustering variable (cluster IDs) for clustered RCTs <code>c(0)</code> = non-clustered RCTs
<code>wgt</code>	<code><- c('wgt1')</code> or <code>c(0)</code>	Name of weight to adjust for the sample design or missing data <code>c(0)</code> = no weights
<code>marg</code>	<code><- c(1)</code> or <code>c(0)</code>	<code>c(1)</code> = marginal logit model that conditions on the covariates only <code>c(0)</code> = random effects logit model that also conditions on random intercepts (for clustered RCTs only)
<code>se_ipw</code>	<code><- c(1)</code> or <code>c(0)</code>	<code>c(1)</code> = standard errors adjust for error in the IPW weights <code>c(0)</code> = no standard error adjustments (could run faster)
<code>out_regr</code>	<code><- c('cace regr log.txt')</code>	Name of output txt file summarizing data and regression results
<code>out_est</code>	<code><- c('cace est.txt')</code>	Name of output txt file containing CACE estimation results

RUNNING THE PROGRAM

The program was developed using Version R 4.1.3 and tested using R Studio 2022.07.1.

Install the required libraries

Before running the program, you will need to install two R packages from the official R repository ([CRAN](#)): [lme4](#) and [estimatr](#). These packages can be installed, for example, using the `install.packages("lme4")` command and similarly for [estimatr](#). If not installed, you may be asked if you want them installed the first time you run the program.

Steps for running the program

Set the working directory

```
setwd("C:/MyDirectory")
```

Enter the inputs from above

Call the cace_serv_bias.R script from the working directory

```
source("cace_serv_bias.R")
```

Call the function to conduct the analysis

```
cace_serv_bias(data_csv,serv_var,y_var,trt_var,xt_logit,xc_logit,x_wls,  
               clus_var,wgt,marg,se_ipw,out_regr,out_est)
```

You can view the output text files using Notepad

You can call the function again with different inputs, where you only need to re-specify the inputs that you want to change. However, make sure to provide new inputs for the output txt files (out_regr and out_est) or the old output files will be overwritten.

TECHINICAL NOTES

1. The program assumes the data have been cleaned. The program does not check for invalid data values. It may not run properly with a csv data file with invalid data elements or missing data codes. However, the program removes observations with (i) missing values for any of the input variables, (ii) treatment and service receipt indicator values not equal to 0 or 1, and (iii) negative weights.
2. If input weights are provided, the program normalizes them to sum to the treatment and control group sample sizes. They are then multiplied by the IPW weights to estimate treatment effects.
3. In calculating standard errors using the generalized estimating equation approach (which generates robust estimators), the program does not apply a small sample adjustment that is sometimes applied in R and Stata packages. This is because more research is needed to develop appropriate adjustment factors in the causal mediation RCT context that is the focus of the program.
4. Adjusting standard errors for the estimation error in the IPW weights might increase program running time. This will especially be the case for non-clustered designs with large samples.
5. In calculating the CACE_TC1 treatment effect estimator, the program uses the treatment group estimate for the CACE-TC population share.