Bootstrapping Propensity Score Analysis

As the popularity of propensity score methods for estimating causal effects in observational studies increase, the choices researchers have for which methods to use has also increased. Estimated treatment effects may be sensitive to choice of method. One approach to test the sensitivity of method choice is to test the null hypothesis more than once using more than one method (Rosenbaum, 2012). With the wide availability of high power computers resampling methods such as bootstrapping (Efron, 1979) have become popular for providing more estimates of the sampling distribution. This poster introduces the `PSAboot` R package that provides functions for bootstrapping propensity score methods. It deviates from traditional bootstrapping methods by allowing for different sampling specifications for treatment and control groups, mainly to ensure the ratio of treatment-to-control observations are consistent. This approach can also be used in situations where there is imbalance between the number of treatment and control observations by allowing for up and/or down sampling. Lastly, by estimating balance statistics and treatment effects for each bootstrap sample, we can compare the distributions across multiple propensity score methods to examine the relative performance of these methods.

Bootstrapping Process

The PSAboot function uses a stratified bootstrap approach that gets treatment and balance estimates across multiple methods for each bootstrap sample.

1. Estimate the effects using the full dataset (i.e. the non-bootstrapped analysis).
2. Draw *M* stratified bootstrap samples. Stratified on the treatment variable so that each bootstrap sample has the same number of treatment and control observations.
3. For each bootstrap sample, estimate the treatment effect for each method.
4. Evaluate balance for each method.
5. Provide an overall pooled estimate across all bootstrap samples.

Propensity Score Methods

The PSAboot package implements a number of common methods for estimating propensity scores and treatment effects, but can easily be extended. The default methods include:

1. Ctree - Stratification using classification trees.
2. Matching - Matching using the Matching package.
3. MatchIt - Matching using the MatchIt package.
4. Rpart - Stratification using the rpart recursive partitioning method.
5. Stratification - Stratification using quintiles on propensity scores estimated using logistic regression.
6. Weighting - Propensity score weights estimated using logistic regression applied to a linear regression.

PSAboot Function Parameters

- **`Tr`** - numeric (0 or 1) or logical vector of treatment indicators.
- **`Y`** - vector for the outcome variable.
- **`X`** - matrix or data frame of covariates used to estimate the propensity scores.
- **`M`** - number of bootstrap samples to generate (default is 100).
- **`formu`** - formula used for estimating propensity scores (optional). If omitted all variables in `X` will be used.
- **`control.ratio`** - the maximum ratio of control observations to sample relative to the treatment units. That is, assuming that the $n_{control} > n_{treatment}$, the number of control observations sampled will not exceed `control.ratio * ` $n_{treatment}$.
- **`control.sample.size`** - the size of each bootstrap sample of control units. The default is calculated based upon the `control.ratio` parameter. If specified, `control.ratio` is ignored.
- **`control.replace`** - whether to use replacement when sampling from control units (default is TRUE).
- **`treated.sample.size`** - the size of each bootstrap sample of treatment units. The default uses all treatment units for each bootstrap sample.
- **`treated.replace`** - whether to use replacement when sampling from treated units (default is TRUE).
- **`methods`** - a named vector of functions for each PSA method to use.
- **`seed`** - random seed. Each iteration, i, will use a seed of seed + i.
- **`parallel`** - whether to run the bootstrap samples in parallel.

R Code - Using the PSAboot Package

```
library(PSAboot)
data(tutoring, package="TriMatch")
tutoring.formu <- treat2 ~ Gender + Ethnicity + Military + ESL +
   EdMother + EdFather + Age + Employment + Income + Transfer + GPA
tutoring$treat2 <- tutoring$treat != "Control"
covariates <- tutoring[,all.vars(tutoring.formu)[-1]]

tutoring.boot <- PSAboot(Tr = tutoring$treat2,    # Vector of treatment indicators
                         Y = tutoring$Grade,      # Outcome variable
                         X = covariates,          # Data frame of covariates
                         seed = 2112)             # For reproducibility
summary(tutoring.boot)                            # Numeric results

tutoring.balance <- balance(tutoring.boot)        # Calculate balance statistics
plot(tutoring.balance)                            # Figure 1
boxplot(tutoring.balance)                         # Figure 2

boxplot(tutoring.boot)                            # Figure 3
plot(tutoring.boot)                               # Figure 4
matrixplot(tutoring.boot)                         # Figure 5
```

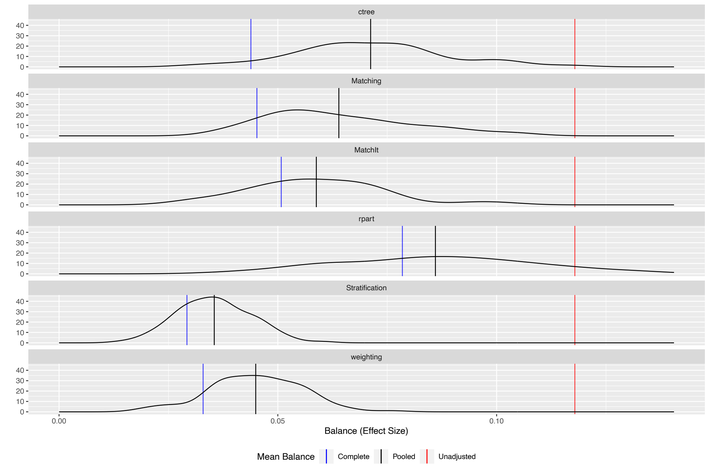

Figure 1. Distribution of overall balance by method.
This figure provides the distribution of the overall effect size (using the mean by default) across all bootstrap samples by method. The vertical red line is the average unadjusted effect size, the blue line is the average adjusted effect size using the the complete dataset (i.e. before bootstrapping), and the black line is the average adjusted effect size for all the bootstrap samples.

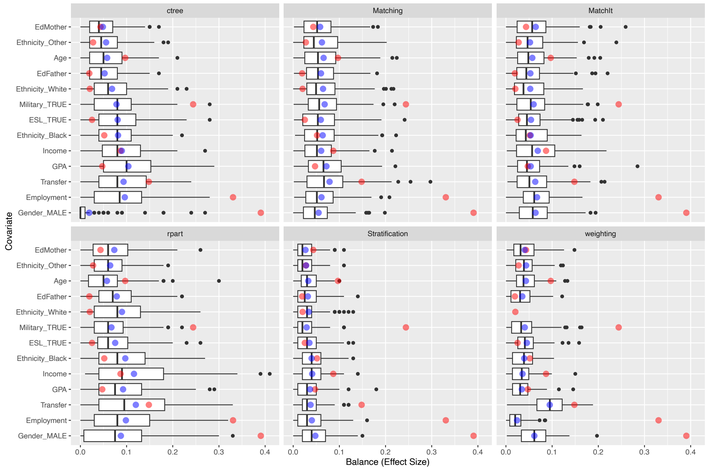

Figure 2. Box plot of covariate balance by covariate and method.
The individual boxes represent the distribution of adjusted effect size for covariates from the bootstrap samples with the mean adjusted effect size in blue. The unadjusted effect size is the red point.

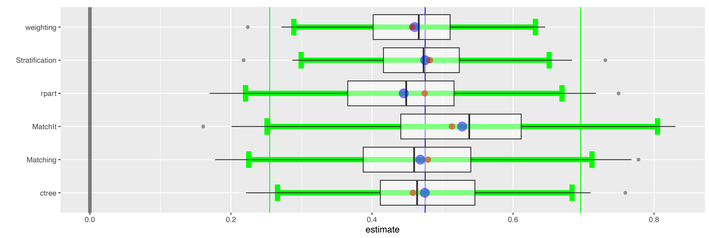

Figure 3. Box plot of treatment effects.
Box plot of the estimated treatment effects across all bootstrap samples. The green bars represent a 95% confidence interval from the bootstrap distributions for each method. Blue points are the estimated treatment effect using the non-bootstrap sample.

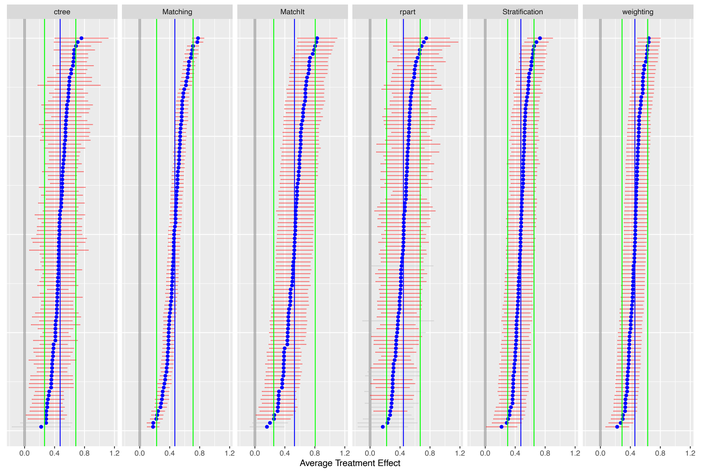

Figure 4. Bootstrap Treatment Estimates.
Each line represents the 95% confidence interval for the estimated treatment effect for each bootstrap sample (100 bootstrap samples in this example). Lines that are red are confidence intervals that do not span zero. The overall treatment effect is represented by the vertical black line with 90% confidence intervals in green.

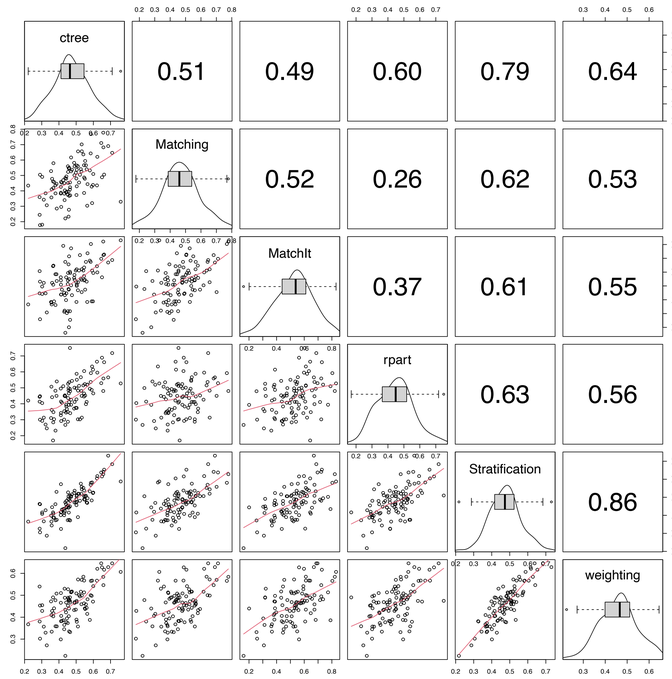

Figure 5. Matrix Plot.
This figure plots the distribution of treatment effects for each method in the diagonal, a scatter plot of treatment effects estimated with the same bootstrap sample across each pair of methods in the lower half, and the correlation between those treatment estimates by bootstrap samples in the upper half.

CUNY School of Professional Studies

Jason M. Bryer, Ph.D.
jason.bryer@cuny.edu
https://jbryer.github.io/PSAboot/

SCAN ME