Rough Draft of the `benchtestr` vignette for 200C - Spring 2021

Igor Geyn

5/26/2021

Introduction

This document serves to introduce the R package `benchtestr`. It is co-authored and co-mainted by Igor Geyn (PhD Student, Department of Political Science UCLA) and Shing Hon Lam (PhD Student, Department of Political Science UCLA). Original inspiration for the project comes from Professor Chad Hazlett (UCLA Political Science and Statistics) who taught 200C with Ciara Sterbenz in Spring 2021.

Purpose

As has been made famous by Dehejia and Wahba (1999), [citation], and [citation], benchmarking experimental findings with observational data is an important and informative exercise. When executed properly, an observational benchmark of an experimental study can reveal key information about the generalizability of the experiment. Unfortunately, while high-quality benchmarking data exists and has been leveraged in a number of past benchmarking efforts as well as meta analyses of benchmarking, this data is not readily available to the researcher.

This package compiles data across substantive fields including political science, medicine, and education to allow for robust benchmarking of causal estimates. Specifically, `benchtestr` enables the user to leverage common estimation techniques – including a range of matching approaches – on the above-mentioned datasets. `benchtestr` presents a number of intuitive functions, helpful defaults, and diverse datasets that can be tested using existing classical estimators (e.g., naive difference-in-means) as well as those defined by the user.

Vignette / Examples

The best way to learn `benchtestr` is by doing. Let's take a look at a couple of scenarios.

Example 1: The Kitchen Sink

Let's say you are working on an estimator that you think is effective for estimating the average treatment effect on the treated (i.e. ATT, ATET, etc.) You would like to see how this estimator performs in the experimental setting vs. in the observational setting. What would you do?

Follow a few easy steps:

1. **Examine the documentation on datasets.** Included with `benchtestr`, and made available on the package's Github repo, is documentation describing the data that's shipped with the package. This includes a summary of the accompanying paper's findings (where appropriate), a synopsis of the estimators and estimates in the original anlaysis, a brief description of the variables, and links to past benchmarking efforts (focused mostly on published academic work, for the time being).

2. **Make dataset selections.** Choose which datasets are most relevant and appropriate, or simply use the default option (all datasets) and examine the output.

3. **Make decision about matching.** Do you want to look at your balance on observed covariates across the selected datasets? Do you want to try different matching approaches? The default is to apply three

classical matching techniques and output two summaries of balance – a balance table and a Love plot – for each of the three technques.

4. **Run the corresponding `benchtestr` functions to get your results.** With your decisions in-hand, all you have to do is pass parameters to a series of `benchtestr` functions.

For example:

Example 2: The Balance Check

Example 3:

Example 4:

References

Publications

Brian J. Gaines and James H. Kuklinski, (2011), "Experimental Estimation of Heterogeneous Treatment Effects Related to Self-Selection," American Journal of Political Science 55(3): 724-736, doi:10.1111/j.1540-5907.2011.00518.x.

software

GK2011, Gaines and Kuklinski (2011) Estimators for Hybrid Experiments, *Github user*: `leeper`, *Name on Github*: Thomas J. Leeper. https://github.com/leeper/GK2011.