RobustiPy: An easy-to-use multiversal Python library

Daniel Valdenegro, Charles Rahal.

Demographic Science Unit, Nuffield Department of Population Health, University of Oxford.

Introduction

There has recently been an increasing amount of attention paid to the levels of uncertainty around estimates produced in the academic social sciences, with the issue of researcher-induced uncertainty addressed in several high-profile papers [1].

A key issue in researcher-induced uncertainty is the sensitivity of the estimates to different model specifications.

These model specifications usually include all possible combinations of 'control' variables, along with other analytical choices.

In order to address the sensitivity of these estimates, all possible specifications are typically computed and presented in a curve, showcasing the variability of the output space.

This has come to be known as 'multiverse analysis' [2] when examining multiple types of 'researcher degrees of freedom' or 'specification-curve analysis' [3] when exclusively considering specification choices.

Formal problem

Consider the association between variables Y and X, where a set of covariates Z can influence the relationship between the former as follows:

$$Y = F(X, \mathbf{Z}) + \epsilon$$

Let's define the set of reasonable operationalisations of Y, X, Z and F() as:

$$\overleftrightarrow{Y}, \overleftrightarrow{X}, \overleftrightarrow{\mathbf{Z}}, \overleftrightarrow{F()},$$

Then, we have:

$$\overleftrightarrow{Y}_{k_Y} = \overleftrightarrow{F}_{k_f}(\overleftrightarrow{X}_{k_X}, \overleftrightarrow{\mathbf{Z}}_{k_\mathbf{Z}}) + \epsilon.$$

Which corresponds to a single possible specification of the π set of all possible specifications. The total number of specifications can then be calculated as 2^N where N is:

$$N = n_{\overleftrightarrow{Y}} + n_{\overleftrightarrow{X}} + n_{\overleftrightarrow{\mathbf{Z}}} + n_{\overleftrightarrow{F()}}$$

The goal of a multiverse type analysis should be to have a large specification space π, since:

$$\lim_{n \to \infty} \Pr\left(E[\overleftrightarrow{y}^{n}_{\Pi}] = y\right) = 1$$

Usage

```
1 import os
2 from robustify.utils import prepare_union, prepare_asc
```

```
1 import matplotlib.pyplot as plt
2 from robustify.models import OLSRobust
```

```
1 y, c, x, data = prepare_union(os.path.join('data',
2                                            'input',
3                                            'nlsw88.dta'))
```

```
1 union_robust = OLSRobust(y=[y], x=[x], data=data)
```

```
1 union_robust.fit(controls=c,
2                  draws=100,
3                  replace=True)
```

```
Working... ──────────────────────────────────── 100% 0:06:34
```

```
1 union_results = union_robust.get_results()
```

```
1 union_results.plot(specs=[['hours', 'collgrad'],
2                           ['collgrad'],
3                           ['hours', 'age']],
4                    ic='hqic',
5                    figsize=(16, 8))
6 plt.show()
```

Download here!

Future direction

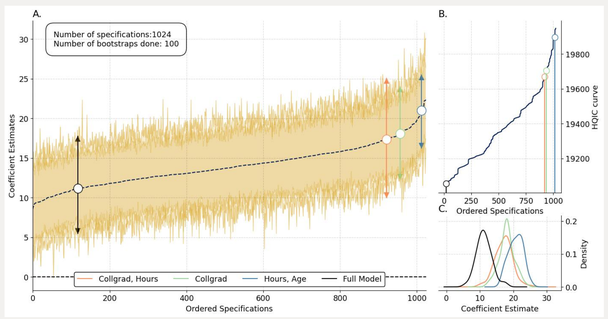
Hackathon to be held June 2024, more info in the link above!

Expanding the features based on user suggestion.

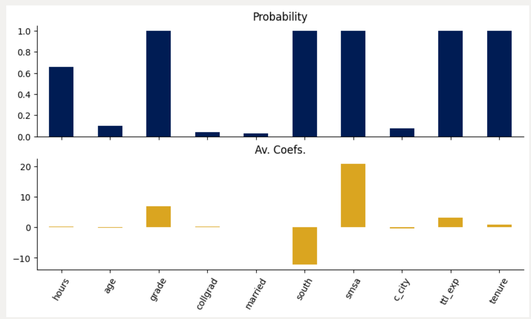
Port to R programing language coming soon!

Outputs

Customizable Specification Curve output with **many extra indicators**:



Bayesian Model Averaging for all covariates **by default**:



And much more!!!

References

[1] Breznau, Nate, Eike Mark Rinke, Alexander Wuttke, Hung HV Nguyen, Muna Adem, Jule Adriaans, Amalia Alvarez-Benjumea et al. "Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty." Proceedings of the National Academy of Sciences 119, no. 44 (2022): e2203150119.

[2] Steegen, Sara, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. "Increasing transparency through a multiverse analysis." Perspectives on Psychological Science 11, no. 5 (2016): 702-712.

[3] Simonsohn, Uri, Joseph P. Simmons, and Leif D. Nelson. "Specification curve analysis." Nature Human Behaviour 4, no. 11 (2020): 1208-1214.

Visit booth 312 and scan the code to find out more