

version 0.1.0

Title

persuasio4yz — Conducts causal inference on persuasive effects for binary outcome y and binary instrument z

Syntax

```
persuasio4yz depvar instrvar [covariates] [if] [in] [, level(#)  
model(string) method(string) nboot(#) title(string)]
```

Options

<i>option</i>	<i>Description</i>
level (#)	Set confidence level; default is level (95)
model (string)	Regression model when <i>covariates</i> are present
method (string)	Inference method; default is method ("normal")
nboot (#)	Perform # bootstrap replications
title (string)	Title of estimation

Description

persuasio4yz conducts causal inference on persuasive effects.

It is assumed that binary outcome y and binary instrument z are observed. This command is for the case when persuasive treatment (t) is unobserved, using estimates of the lower bound on the average persuasion rate (APR) via this package's command **aprlb**.

varlist should include *depvar* *instrvar* *covariates* in order. Here, *depvar* is binary outcomes (y), *instrvar* is binary instruments (z), and *covariates* (x) are optional.

When treatment t is unobserved, the upper bound on the APR is simply 1.

There are two cases: (i) *covariates* are absent and (ii) *covariates* are present.

- If x are absent, the lower bound (**theta_L**) on the APR is defined by

$$\mathbf{theta_L} = \{\Pr(y=1|z=1) - \Pr(y=1|z=0)\} / \{1 - \Pr(y=1|z=0)\}.$$

The estimate and confidence interval are obtained by the following procedure:

1. $\Pr(y=1|z=1)$ and $\Pr(y=1|z=0)$ are estimated by regressing y on z .
2. **theta_L** is computed using the estimates obtained above.
3. The standard error is computed via STATA command **nlcom**.
4. Then, a confidence interval for the APR is set by

$$[\text{est} - \text{cv} * \text{se} , 1],$$

where *est* is the estimate, *se* is the standard error, and *cv* is the one-sided standard normal critical value (e.g., *cv* = 1.645 for **level**(95)).

- If x are present, the lower bound (**theta_L**) on the APR is defined by

$$\mathbf{theta_L} = E[\mathbf{theta_L}(x)],$$

where

$$\mathbf{theta_L}(x) = \{\Pr(y=1|z=1,x) - \Pr(y=1|z=0,x)\} / \{1 - \Pr(y=1|z=0,x)\}.$$

The estimate is obtained by the following procedure.

If **model**("no_interaction") is selected (default choice),

1. $\Pr(y=1|z,x)$ is estimated by regressing y on z and x .

Alternatively, if **model**("interaction") is selected,

- 1a. $\Pr(y=1|z=1,x)$ is estimated by regressing y on x given $z = 1$.
- 1b. $\Pr(y=1|z=0,x)$ is estimated by regressing y on x given $z = 0$.

After step 1, both options are followed by:

2. For each x in the estimation sample, **theta_L**(x) is evaluated.
3. The estimates of **theta_L**(x) are averaged to estimate **theta_L**.
4. A bootstrap confidence interval for the APR is set by

$$[\text{bs_est}(\alpha) , 1],$$

where **bs_est**(*alpha*) is the *alpha* quantile of the bootstrap estimates of **theta_L** and $1 - \alpha$ is the confidence level.

The bootstrap procedure is implemented via STATA command **bootstrap**.

Options

model(*string*) specifies a regression model of y on z and x .

This option is only relevant when x is present. The default option is "no_interaction" between z and x . When "interaction" is selected, full interactions between z and x are allowed; this is accomplished by estimating $\Pr(y=1|z=1,x)$ and $\Pr(y=1|z=0,x)$, separately.

level(*#*) sets confidence level; default is **level**(95).

method(*string*) refers the method for inference.

The default option is **method**("normal"). By the nature of identification, one-sided confidence intervals are produced.

1. When x are present, it needs to be set as **method**("bootstrap"); otherwise, the confidence interval will be missing.
2. When x are absent, both options yield non-missing confidence intervals.

nboot(*#*) chooses the number of bootstrap replications.

The default option is **nboot**(50). It is only relevant when **method**("bootstrap") is selected.

title(*string*) specifies the title of estimation.

Remarks

It is recommended to use **nboot**(*#*) with *#* at least 1000. A default choice of 50 is meant to check the code initially because it may take a long time to run the bootstrap part when there are a large number of covariates. The bootstrap confidence interval is based on percentile bootstrap. A use of normality-based bootstrap confidence interval is not recommended because bootstrap standard errors can be unreasonably large in applications.

Examples

We first call the dataset included in the package.

```
. use GKB, clear
```

The first example conducts inference on the APR without covariates, using normal approximation.

```
. persuasio4yz voteddem_all post, level(80) method("normal")
```

The second example conducts bootstrap inference on the APR.

```
. persuasio4yz voteddem_all post, level(80) method("bootstrap")  
nboot(1000)
```

The third example conducts bootstrap inference on the APR with a covariate, MZwave2, interacting with the instrument, post.

```
. persuasio4yz voteddem_all post MZwave2, level(80)  
model("interaction") method("bootstrap") nboot(1000)
```

Stored results

Matrices

e(lb_est): (1*2 matrix) bounds on the average persuasion rate in the form of [lb, 1]

e(lb_ci): (1*2 matrix) confidence interval for the average persuasion rate in the form of [lb_ci, 1]

Macros

e(cilevel): confidence level

e(inference_method): inference method: "normal" or "bootstrap"

Authors

Sung Jae Jun, Penn State University, <sjun@psu.edu>

Sokbae Lee, Columbia University, <sl3841@columbia.edu>

License

References

Sung Jae Jun and Sokbae Lee (2019), Identifying the Effect of Persuasion, [arXiv:1812.02276](https://arxiv.org/abs/1812.02276) [[econ.EM](#)]