

## Title

---

**persuasio4ytz** — Conduct causal inference on persuasive effects for binary outcomes  $y$ , binary treatments  $t$  and binary instruments  $z$

## Syntax

---

```
persuasio4ytz depvar treatvar instrvar [covariates] [if] [in] [,  
level(#) model(string) method(string) nboot(#) title(string)]
```

## Options

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

## Description

---

**persuasio4ytz** conducts causal inference on persuasive effects.

It is assumed that binary outcomes  $y$ , binary treatments  $t$ , and binary instruments  $z$  are observed. This command is for the case when persuasive treatment ( $t$ ) is observed, using estimates of the lower and upper bounds on the average persuasion rate (APR) via this package's commands **aprlb** and **aprub**.

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

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

- Without  $x$ , the lower bound (**theta\_L**) on the APR is defined by

$$\mathbf{theta\_L} = \{\Pr(y=1|z=1) - \Pr(y=1|z=0)\} / \{1 - \Pr(y=1|z=0)\},$$

and the upper bound (**theta\_U**) on the APR is defined by

$$\mathbf{theta\_U} = \{E[A|z=1] - E[B|z=0]\} / \{1 - E[B|z=0]\},$$

where  $A = 1(y=1, t=1) + 1 - 1(t=1)$  and  $B = 1(y=1, t=0)$ .

The lower bound is estimated 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**.

The upper bound is estimated by the following procedure:

1.  $E[A|z=1]$  is estimated by regressing  $A$  on  $z$ .
2.  $E[B|z=0]$  is estimated by regressing  $B$  on  $z$ .
3. **theta\_U** is computed using the estimates obtained above.
4. The standard error is computed via STATA command **nlcom**.

Then, a confidence interval for the APR is set by

$$[ \mathit{est\_lb} - \mathit{cv} * \mathit{se\_lb} , \mathit{est\_ub} + \mathit{cv} * \mathit{se\_ub} ],$$

where  $\mathit{est\_lb}$  and  $\mathit{est\_ub}$  are the estimates of the lower and upper bounds,  $\mathit{se\_lb}$  and  $\mathit{se\_ub}$  are the corresponding standard errors, and  $\mathit{cv}$  is the critical value obtained via the method of Stoye (2009).

- With  $x$ , the lower bound (**theta\_L**) on the APR is defined by

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

where

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

and the upper bound (**theta\_U**) on the APR is defined by

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

where

$$\mathbf{theta\_U}(x) = \{E[A|z=1, x] - E[B|z=0, x]\} / \{1 - E[B|z=0, x]\}.$$

The lower bound is estimated 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`.

The upper bound is estimated by the following procedure:

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

1.  $E[A|z=1,x]$  is estimated by regressing  $A$  on  $z$  and  $x$ .

2.  $E[B|z=0,x]$  is estimated by regressing  $B$  on  $z$  and  $x$ .

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

1.  $E[A|z=1,x]$  is estimated by regressing  $A$  on  $x$  given  $z = 1$ .

2.  $E[B|z=0,x]$  is estimated by regressing  $B$  on  $x$  given  $z = 0$ .

After step 1, both options are followed by:

3. For each  $x$  in the estimation sample, `theta_U(x)` is evaluated.

4. The estimates of `theta_U(x)` are averaged to estimate `theta_U`.

Then, a bootstrap confidence interval for the APR is set by

`[ bs_est_lb(alpha) , bs_est_ub(1 - alpha) ],`

where `bs_est_lb(alpha)` is the  $\alpha$  quantile of the bootstrap estimates of `theta_L`, `bs_est_ub(alpha)` is the  $1 - \alpha$  quantile of the bootstrap estimates of `theta_U`, and  $1 - \alpha$  is the confidence level.

The resulting coverage probability is  $1 - \alpha$  if the identified interval never reduces to a singleton set. More generally, it will be  $1 - 2*\alpha$  by Bonferroni correction.

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.

**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* is present, it needs to be set as **method**("bootstrap"); otherwise, the confidence interval will be missing.
2. When *x* is 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. 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.

```
. persuasio4ytz voteddem_all readsome post, level(80) method("normal")
```

The second example conducts bootstrap inference on the APR.

```
. persuasio4ytz voteddem_all readsome 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.

```
. persuasio4ytz voteddem_all readsome post MZwave2, level(80)
model("interaction") method("bootstrap") nboot(1000)
```

## **Stored results**

---

### **Matrices**

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

**e(apr\_ci)**: (1\*2 matrix) confidence interval for the average persuasion rate in the form of [lb\_ci, ub\_ci]

### **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**

---

GPL-3

## **References**

---

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