persuasio: Estimating the Effect of Persuasion in Stata

Sung Jae Jun
Department of Economics
The Pennsylvania State University
University Park, PA
suj14@psu.edu

Sokbae Lee
Department of Economics
Columbia University
New York, NY
sl3841@columbia.edu

Abstract. We present the `persuasio` command that implements some of estimation methods developed in Jun and Lee (2019).

Keywords: `persuasio`, persuasion, partial identification, treatment effects

1 Introduction

This document accompanies Jun and Lee (2019) and presents the `persuasio` command that implements some of estimation methods developed in Jun and Lee (2019). A Stata module is currently under development and available at `https://github.com/sokbae/persuasio`. It is also posted at the Statistical Software Components (SSC) archive and should be installed from within Stata by typing `ssc install persuasio`.[1]

2 The Commands

In this section, we present the syntax of each command, its available options, and its stored results.

2.1 The persuasio command

2.1.1 Syntax

`persuasio subcommand` *varlist* [*if*] [*in*] [`, level(`#`) model(`*string*`)`
`method(`*string*`) nboot(`#`) title(`*string*`)`]

where `persuasio` is a wrapper that calls a variety of subroutines via `subcommand` that has several options:
First of all, `apr` and `lpr` refer to subcommands for a data scenario where binary outcomes Y_i, binary treatments T_i, and binary instruments Z_i are observed (with covariates X_i if exist) for each observational unit i. Specifically, `apr` and `lpr` provide causal inference on the average persuasion rate (APR) and the local persuasion rate (LPR), respectively. Second, `yz` is concerned with a different data scenario where persuasive treatment T_i is unobserved. In this case, bounds on the APR are the same as those on the LPR. It provides causal inference for the APR and hence, for the LPR trivially. Finally, `calc` is designed for the case where summary statistics on $\Pr(Y_i = 1|Z_i = z)$

[1] The current version is `0.1.0` (30 January 2021).

Table 1: Overview of the `persuasio` command

Subcommand	Description
`apr`	inference on the average persuasion rate (APR) when $(Y_i, T_i, Z_i)_{i=1}^n$ are observed
`lpr`	inference on the local persuasion rate (LPR) when $(Y_i, T_i, Z_i)_{i=1}^n$ are observed
`yz`	inference on APR and LPR when $(Y_i, Z_i)_{i=1}^n$ are observed
`calc`	bound estimates on APR and LPR with summary statistics

and $\Pr(T_i = 1|Z_i = z)$ for each $z = 0, 1$ are available. It provides the lower and upper bounds on both the APR and the LPR.

2.1.2 Options

`model`(*string*) specifies a regression model of Y_i on Z_i and X_i. This option is only relevant when X_i is present. The default option is "no_interaction" between Z_i and X_i. When "interaction" is selected, full interactions between Z_i and X_i 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.

- When X_i is present, it needs to be set as `method`("bootstrap"); otherwise, the confidence interval will be missing.
- When X_i 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. 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. Generally, normality-based bootstrap confidence interval is not recommended because bootstrap standard errors can be unreasonably large in applications.

`title`(*string*) specifies a title.

All these options are irrelevant for subcommands `calc`.

2.1.3 Stored results

Subcommand `apr` calls the `persuasio4ytz` command, `lpr` command `persuasio4ytz2lpr`, `yz` command `persuasio4yz`, and `calc` command `calc4persuasio`, respectively. See below for these commands for details on stored results.

2.2 The persuasio4ytz command

2.2.1 Syntax

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

where *depvar*, *treatvar* and *instrvar* refer to binary outcomes Y_i, binary treatments T_i, and binary instruments Z_i, respectively. This command is for the case where persuasive treatment (T_i) is observed, using estimates of the lower and upper bounds on the average persuasion rate (APR) via the `aprlb` and `aprub` commands. The variables should be in the order of *depvar*, *treatvar*, *instrvar* and *covariates*, where *covariates* are optional.

See Sections 2.3 and 2.4 for details regarding estimation of the lower and upper bounds on the APR. The `persuasio4ytz` command provides two alternative methods for inference. First, using asymptotic normality, a confidence interval for the APR is set by

$$[\widehat{\theta}_L - \mathrm{cv}(1-\alpha)\widehat{s}_L, \widehat{\theta}_U + \mathrm{cv}(1-\alpha)\widehat{s}_U],$$

where $\widehat{\theta}_L$ and $\widehat{\theta}_U$ are the estimates of the lower and upper bounds, $\widehat{s}_L$ and $\widehat{s}_U$ are the corresponding standard errors, and $\mathrm{cv}(1-\alpha)$ is the critical value obtained via the method of Stoye (2009). This approach is currently available only for the case where there are no covariates in the model. With X_i, the standard error formula based on asymptotic normality is complex and furthermore, inference based on asymptotic normality may not provide an accurate approximation when the sample size is not large.

Alternatively, when the covariates are present, a bootstrap confidence interval for the APR is set by

$$[\widehat{\theta}^*_L(\alpha), \widehat{\theta}^*_U(1-\alpha)],$$

where $\widehat{\theta}^*_L(\alpha)$ is the α quantile of the bootstrap estimates of θ_L, $\widehat{\theta}^*_U(\alpha)$ is the $1-\alpha$ quantile of the bootstrap estimates of θ_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`.

2.2.2 Options

Options are identical to those in Section 2.1.2.

2.2.3 Stored results

Matrices

`e(apr_est)`	(1×2 matrix) lower and upper bounds on the average persuasion rate
`e(apr_ci)`	(1×2 matrix) confidence interval for the average persuasion rate

Macros

`e(cilevel)`	confidence level
`inference_method`	inference method: "normal" or "bootstrap"

2.3 The aprlb command

2.3.1 Syntax

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

where *depvar* and *instrvar* refer to binary outcomes Y_i and binary instruments Z_i, respectively. The variables should be in the order of *depvar*, *instrvar* and *covariates*, where *covariates* are optional.

There are two cases: (i) *covariates* (X_i) are absent and (ii) *covariates* are present. Without X_i, the lower bound (θ_L) on the APR is defined by

$$\theta_L = \frac{\Pr(Y_i = 1|Z_i = 1) - \Pr(Y_i = 1|Z_i = 0)}{1 - \Pr(Y_i = 1|Z_i = 1)}, \tag{1}$$

and the upper bound (θ_U) on the APR is defined by

$$\theta_U = \frac{\mathbb{E}[A_i|Z_i = 1] - \mathbb{E}[B_i|Z_i = 0]}{1 - \mathbb{E}[B_i|Z_i = 0]}, \tag{2}$$

where $A_i := 1(Y_i = 1, T_i = 1) + 1(T_i = 0)$ and $B_i := 1(Y_i = 1, T_i = 0)$. Here, $1(\cdot)$ is the usual indicator function. The lower bound is estimated by the following procedure:

1. $\Pr(Y_i = 1|Z_i = 1)$ and $\Pr(Y_i = 1|Z_i = 0)$ are estimated by regressing Y_i on Z_i;
2. θ_L in (1) is computed using the estimates obtained above;
3. the standard error is computed via STATA command `nlcom`.

With X_i, the lower bound on the APR is now defined by

$$\theta_L = \mathbb{E}[\theta_L(X_i)],$$

where

$$\theta_L(x) := \frac{\Pr(Y_i = 1|Z_i = 1, X_i = x) - \Pr(Y_i = 1|Z_i = 0, X_i = x)}{1 - \Pr(Y_i = 1|Z_i = 1, X_i = x)}.$$

The lower bound is estimated by the following modified procedure. If `model`("no_interaction") is selected (default choice),

1. $\Pr(Y_i = 1|Z_i, X_i)$ is estimated by regressing Y_i on Z_i and X_i.

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

1a. $\Pr(Y_i = 1|Z_i = 1, X_i)$ is estimated by regressing Y_i on X_i given $Z_i = 1$;

1b. $\Pr(Y_i = 1|Z_i = 0, X_i)$ is estimated by regressing Y_i on X_i given $Z_i = 0$.

After step 1, both options are followed by:

2. for each X_i in the estimation sample, $\theta_L(X_i)$ is evaluated;

3. the estimates of $\theta_L(X_i)$ are averaged to estimate θ_L.

When covariates (X_i) are present, the standard error is missing because an analytic formula for the standard error is complex. Bootstrap inference is implemented when this package's `persuasio` command is called to conduct inference. It is recommended to use the `persuasio` command instead of calling the `aprub` command directly.

2.3.2 Options

`model`(*string*) specifies a regression model of Y_i on Z_i and X_i. This option is only relevant when X_i is present. The default option is "no_interaction" between Z_i and X_i. When "interaction" is selected, full interactions between Z_i and X_i are allowed.

`title`(*string*) specifies a title.

2.3.3 Stored results

Scalars

`e(N)`	sample size
`e(lb_coef)`	estimate of the lower bound on the average persuasion rate
`e(lb_se)`	standard error of the lower bound on the average persuasion rate

Macros

`e(outcome)`	variable name of the binary outcome variable
`e(instrument)`	variable name of the binary instrumental variable
`e(covariates)`	variable name(s) of the covariates if they exist
`e(model)`	regression model specification ("no_interaction" or "interaction")

Functions

`e(sample)`	1 if the observations are used for estimation, and 0 otherwise.

2.4 The aprub command

2.4.1 Syntax

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

where *depvar*, *treatvar* and *instrvar* refer to binary outcomes Y_i and binary instruments Z_i, respectively. The variables should be in the order of *depvar*, *treatvar*, *instrvar* and *covariates*, where *covariates* are optional.

There are two cases: (i) *covariates* (X_i) are absent and (ii) *covariates* are present. Define $A_i := 1(Y_i = 1, T_i = 1) + 1(T_i = 0)$ and $B_i := 1(Y_i = 1, T_i = 0)$. Without X_i, the upper bound is estimated by the following procedure:

1. $\mathbb{E}[A_i|Z_i = 1]$ is estimated by regressing A_i on Z_i; $\mathbb{E}[B_i|Z_i = 0]$ is estimated by regressing B_i on Z_i;

2. θ_U in (2) is computed using the estimates obtained above;

3. the standard error is computed via STATA command `nlcom`.

With X_i, the upper bound on the APR is now defined by

$$\theta_U = \mathbb{E}[\theta_U(X_i)],$$

where

$$\theta_U(x) := \frac{\mathbb{E}[A_i|Z_i = 1, X_i = x] - \mathbb{E}[B_i|Z_i = 0, X_i = x]}{1 - \mathbb{E}[B_i|Z_i = 0, X_i = x]}.$$

Then, the upper bound is estimated by the following modified procedure. If `model`("no_interaction") is selected (default choice),

1. $\mathbb{E}[A_i|Z_i = 1, X_i]$ is estimated by regressing A_i on Z_i and X_i; $\mathbb{E}[B_i|Z_i = 0, X_i]$ is estimated by regressing B_i on Z_i and X_i.

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

1a. $\mathbb{E}[A_i|Z_i = 1, X_i]$ is estimated by regressing A_i on X_i given $Z_i = 1$;

1b. $\mathbb{E}[B_i|Z_i = 0, X_i]$ is estimated by regressing B_i on X_i given $Z_i = 0$.

After step 1, both options are followed by:

2. for each X_i in the estimation sample, $\theta_U(X_i)$ is evaluated;

3. the estimates of $\theta_U(X_i)$ are averaged to estimate θ_U.

As in the `aprlb` command, the standard error is missing when covariates are present, because an analytic formula for the standard error is complex. Bootstrap inference is implemented when this package's `persuasio` command is called to conduct inference. It is recommended to use the `persuasio` command instead of calling the `aprub` command directly.

2.4.2 Options

Options are identical to those in Section 2.3.2.

2.4.3 Stored results

Scalars

`e(N)`	sample size
`e(ub_coef)`	estimate of the upper bound on the average persuasion rate
`e(ub_se)`	standard error of the upper bound on the average persuasion rate

Macros

`e(outcome)`	variable name of the binary outcome variable
`e(instrument)`	variable name of the binary instrumental variable
`e(covariates)`	variable name(s) of the covariates if they exist
`e(model)`	regression model specification ("no_interaction" or "interaction")

Functions

`e(sample)`	1 if the observations are used for estimation, and 0 otherwise.

2.5 The persuasio4ytz2lpr command

2.5.1 Syntax

`persuasio4ytz2lpr` *depvar treatvar instrvar* [*covariates*] [*if*] [*in*] [, `level(`*#*`)`
`model(`*string*`)` `method(`*string*`)` `nboot(`*#*`)` `title(`*string*`)`]

where *depvar*, *treatvar* and *instrvar* refer to binary outcomes Y_i, binary treatments T_i, and binary instruments Z_i, respectively. This command is to carry out inference on the local persuasion rate (LPR) via this package's command `lpr4ytz` when persuasive treatment (T_i) is observed along with binary outcomes (Y_i) and instruments (Z_i). The variables should be in the order of *depvar*, *treatvar*, *instrvar* and *covariates*, where *covariates* are optional.

See Section 2.6 for details regarding estimation of the LPR. The `persuasio4ytz2lpr` command provides two alternative methods for inference. First, a confidence interval for the LPR is obtained via the usual normal approximation. This approach is currently unavailable for the most general case where there exist interactions between Z_i and X_i in the regression model. Second, the bootstrap procedure is implemented via STATA command `bootstrap`.

2.5.2 Options

Options are identical to those in Section 2.1.2.

2.5.3 Stored results

Macros

`e(lpr_est)`	(1×1 matrix) estimate of the local persuasion rate
`e(lpr_ci)`	(1×2 matrix) confidence interval for the local persuasion rate

Matrices

`e(cilevel)`	confidence level
`inference_method`	inference method: "normal" or "bootstrap"

2.6 The lpr4ytz command

2.6.1 Syntax

`lpr4ytz` *depvar treatvar instrvar* [*covariates*] [*if*] [*in*] [, `level(`*#*`)`
`model(`*string*`)` `method(`*string*`)` `nboot(`*#*`)` `title(`*string*`)`]

where *depvar*, *treatvar* and *instrvar* refer to binary outcomes Y_i, binary treatments T_i, and binary instruments Z_i, respectively. This command is to estimate the local persuasion rate (LPR) when persuasive treatment (T_i) is observed along with binary outcomes (Y_i) and instruments (Z_i). The variables should be in the order of *depvar*, *treatvar*, *instrvar* and *covariates*, where *covariates* are optional.

There are two cases: (i) *covariates* (X_i) are absent and (ii) *covariates* are present. Without X_i, the LPR is defined by

$$\theta^* = \frac{\Pr(Y_i = 1|Z_i = 1) - \Pr(Y_i = 1|Z_i = 0)}{\Pr(Y_i = 0, T_i = 0|Z_i = 0) - \Pr(Y_i = 0, T_i = 0|Z_i = 1)}.$$

The estimate and its standard error are obtained by the following procedure:

1. the numerator of the LPR is estimated by regressing Y_i on Z_i;
2. the denominator is estimated by regressing $(1 - Y_i)(1 - T_i)$ on Z_i;
3. the estimate of the LPR is obtained as the ratio;
4. the standard error is computed via STATA command `nlcom`.

With X_i, the LPR is defined by

$$\theta^* = \frac{\mathbb{E}\left[\theta^*(X_i)\{e(1|X_i) - e(0|X_i)\}\right]}{\mathbb{E}[e(1|X_i) - e(0|X_i)]}, \tag{3}$$

where

$$\theta^*(x) := \frac{\Pr(Y_i = 1|Z_i = 1, X_i = x) - \Pr(Y_i = 1|Z_i = 0, X_i = x)}{\Pr(Y_i = 0, T_i = 0|Z_i = 0, X_i = x) - \Pr(Y_i = 0, T_i = 0|Z_i = 1, X_i = x)}, \tag{4}$$

$$e(t|x) := \Pr(T_i = t|Z_i = 1, X_i = x) - \Pr(T_i = t|Z_i = 0, X_i = x) \quad \text{for } t = 0, 1. \tag{5}$$

The estimate is obtained by the following procedure. If `model`("no_interaction") is selected (default choice),

1. the numerator of the LPR is estimated by regressing Y_i on Z_i and X_i;
2. the denominator is estimated by regressing $(1 - Y_i)(1 - T_i)$ on Z_i and X_i;
3. the estimate of the LPR is obtained as the ratio;
4. the standard error is computed via STATA command `nlcom`.

Note that in this case, $\theta^*(x)$ does not depend on x because of the linear regression model specification.

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

1. $\Pr(Y_i = 1|Z_i = z, X_i)$ is estimated by regressing Y_i on X_i given $z = 0, 1$;
2. $\Pr(Y_i = 0, T_i = 0|Z_i = z, X_i)$ is estimated by regressing $(1 - Y_i)(1 - T_i)$ on X_i given $z = 0, 1$;
3. $\Pr(T_i = 1|Z_i = z, X_i)$ is estimated by regressing T_i on X_i given $z = 0, 1$;
4. for each X_i in the estimation sample, estimates of $\theta^*(X_i)$ in (4) and $e(1|X_i) - e(0|X_i)$ in (5) are obtained;
5. then, the sample analog of θ^* in (3) is constructed.

2.6.2 Options

Options are identical to those in Section 2.3.2.

2.6.3 Stored results

Scalars	
`e(N)`	sample size
`e(lpr_coef)`	estimate of the local persuasion rate
`e(lpr_se)`	standard error of the estimate of the local persuasion rate
Macros	
`e(outcome)`	variable name of the binary outcome variable
`e(instrument)`	variable name of the binary instrumental variable
`e(covariates)`	variable name(s) of the covariates if they exist
`e(model)`	regression model specification ("no_interaction" or "interaction")
Functions	
`e(sample)`	1 if the observations are used for estimation, and 0 otherwise.

2.7 The persuasio4yz command

2.7.1 Syntax

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

where *depvar* and *instrvar* refer to binary outcomes Y_i and binary instruments Z_i, respectively. This command is for the case where persuasive treatment (T_i) is unobserved, using an estimate of the lower bound on the average persuasion rate (APR) via the `aprlb` command. The variables should be in the order of *depvar*, *instrvar* and *covariates*, where *covariates* are optional.

There are two cases: (i) *covariates* (X_i) are absent and (ii) *covariates* are present. Without X_i, the lower bound (θ_L) on the APR is defined in (1). When persuasive treatment (T_i) is unobserved, the upper bound on the APR is simply 1. The lower bound is estimated by the same procedure as described in Section 2.2. Then, a confidence interval for the APR is set by

$$[\widehat{\theta}_L - \mathrm{cv}_{\text{one-sided}}(1-\alpha)\widehat{s}_L, 1],$$

where $\widehat{\theta}_L$ is the estimate of the lower bound, $\widehat{s}_L$ is the standard error, and $\mathrm{cv}_{\text{one-sided}}(1-\alpha)$ is the one-sided standard normal critical value (e.g., $\mathrm{cv}_{\text{one-sided}}(1-\alpha) = 1.645$ for $\alpha = 0.05$).

With X_i, the lower bound on the APR is estimated as as described in Section 2.2. Then, a bootstrap confidence interval for the APR is set by

$$[\widehat{\theta}^*_L(\alpha), 1],$$

where $\widehat{\theta}^*_L(\alpha)$ is the α quantile of the bootstrap estimates of θ_L, and $1-\alpha$ is the confidence level. The bootstrap procedure is implemented via STATA command `bootstrap`.

2.7.2 Options

Options are identical to those in Section 2.1.2.

2.7.3 Stored results

Stored results are identical to those in Section 2.2.3.

2.8 The calc4persuasio command

2.8.1 Syntax

`calc4persuasio y1 y0` $\left[\texttt{e1 e0}\right]$

where `calc4persuasio` calculates the effect of persuasion when information on $\Pr(Y_i = 1|Z_i = z)$ and optionally $\Pr(T_i = 1|Z_i = z)$ for each $z = 0, 1$ is available. The inputs to this command are $\texttt{y1}, \texttt{y0}, \texttt{e1}$ and `e0`. They are all scalars and refer to the estimates of $\Pr(Y_i = 1|Z_i = 1)$, $\Pr(Y_i = 1|Z_i = 0)$, $\Pr(T_i = 1|Z_i = 1)$, and $\Pr(T_i = 1|Z_i = 0)$. The outputs of this command are the lower and upper bounds on the average persuasion rate (APR) as well as the lower and upper bounds on the local persuasion rate (LPR).

There are two cases: (i) all four inputs are given and (ii) only `y1` and `y0` are given. In case (i), `calc4persuasio` provides the following bounds.

1. The lower bound on the APR is defined by
$$\frac{\texttt{y1} - \texttt{y0}}{1 - \texttt{y0}}.$$
2. The upper bound on the APR is given by
$$\frac{\min(1, \texttt{y1} + 1 - \texttt{e1}) - \max(0, \texttt{y0} - \texttt{e0})}{1 - \max(0, \texttt{y0} - \texttt{e0})}.$$
3. The lower bound on the LPR is given by
$$\max\left\{\frac{\texttt{y1} - \texttt{y0}}{1 - \texttt{y0}}, \frac{\texttt{y1} - \texttt{y0}}{\texttt{e1} - \texttt{e0}}\right\}.$$
4. The upper bound on the LPR is simply 1.

In case (ii), `calc4persuasio` provides the following bounds.

1. The lower bound on the APR equals that on LPR and is given by
$$\frac{\texttt{y1} - \texttt{y0}}{1 - \texttt{y0}}.$$
2. The upper bound on both the APR and LPR is simply 1.

2.8.2 Options

There are no options available for this command.

2.8.3 Stored results

Scalars

`r(apr_lb)`	estimate of the lower bound on the average persuasion rate
`r(apr_ub)`	estimate of the upper bound on the average persuasion rate
`r(lpr_lb)`	estimate of the lower bound on the local persuasion rate
`r(lpr_ub)`	estimate of the upper bound on the local persuasion rate

3 An Example

In this section, we exemplify the `persuasio` command with the data from Gerber et al. (2009, GKB hereafter), who report findings from a field experiment to measure the effect of political news. This dataset is used in Jun and Lee (2019) to illustrate the identification results on the effect of persuasion.

In the dataset, an offer of free subscription to *The Washington Post* is randomly offered in the field experiment. This provides a credible instrument: that is, $Z_i = 1$ if the ith individual received free subscription to *The Washington Post*, and $Z_i = 0$ if not (the variable name is `post`). As persuasive treatment, $T_i = 1$ if the ith individual reads a newspaper at least several times per week and $T_i = 0$ otherwise, which is a variable that GKB kept track of in a follow-up survey (the variable name is `readsome`). For the outcome variable, $Y_i = 1$ if the ith individual reported voting for the Democratic candidate in the 2005 gubernatorial election, and $Y_i = 0$ if the subject did not vote for the Democratic candidate or did not vote at all (the variable name is `voteddem_all`). The GKB data is summarized as follows:

```
. use GKB_stj, clear
. by post, sort: tab voteddem_all readsome
---------------------------------------------------------------------------

-> post = 0
 Voted for |
  Democrat |
   in 2005 |
       Gub |
 election, |   Reads paper at least
  set to 0 |     several times per
if did not |           week
      vote |         0          1 |     Total
-----------+----------------------+----------
         0 |       162        130 |       292
         1 |        46         77 |       123
-----------+----------------------+----------
     Total |       208        207 |       415

---------------------------------------------------------------------------

-> post = 1
 Voted for |
  Democrat |
   in 2005 |
       Gub |
 election, |   Reads paper at least
  set to 0 |     several times per
if did not |           week
      vote |         0          1 |     Total
-----------+----------------------+----------
         0 |        94         93 |       187
         1 |        31         68 |        99
-----------+----------------------+----------
     Total |       125        161 |       286
```

3.1 Without Covariates

We first consider the case where there are no covariates. The following Stata output shows estimation results for the average persuasion rate. Because the size of the sample extract we use is relatively modest ($n = 701$) for an interval-identified object, we report the 80% confidence intervals throughout the example.

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

------------------------------------------------------------------
persuasio4ytz: Causal inference on the Average Persuasion Rate
 when outcome, instrument and instrument are observed
------------------------------------------------------------------

 - Binary outcome: voteddem_all
 - Binary treatment: readsome
 - Binary instrument: post

-------------------------------------------------------------------
               Parameter | Bound Estimate      80% Conf. Interval
-------------------------------------------------------------------
 Average Persuasion Rate | .070732  .634288     .02881  .661059
-------------------------------------------------------------------

Note: 80% conf. interval is based on normal approximation
       using the method of Stoye (2009).
       Conf. interval is missing if covariates are present.
       Use option bootstrap for that case.

Reference: Jun and Lee (2019), arXiv:1812.02276 [econ.EM]
```

The average effect of persuasion by reading the newspaper is bounded between 7% and 63% and the 80% confidence interval is $[3\%, 66\%]$. Thus, uncertainty about the effect of persuasion mainly comes from the identification concerns at the population level due to partial identification rather than from sampling uncertainty. We now change the option for inference to bootstrap.

```
. set seed 339487731

. persuasio apr voteddem_all readsome post, ///
>         level(80) method("bootstrap") nboot(1000)

------------------------------------------------------------------
persuasio4ytz: Causal inference on the Average Persuasion Rate
 when outcome, instrument and instrument are observed
 along with covariates
------------------------------------------------------------------

 - Binary outcome: voteddem_all
 - Binary treatment: readsome
 - Binary instrument: post
 - Covariates (if exist):
 - Regression model (if specified):

(running aprlb on estimation sample)

Bootstrap replications (1000)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
```

```
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
..................................................   400
..................................................   450
..................................................   500
..................................................   550
..................................................   600
..................................................   650
..................................................   700
..................................................   750
..................................................   800
..................................................   850
..................................................   900
..................................................   950
..................................................  1000

Bootstrap results                               Number of obs     =        701
                                                Replications      =      1,000

      command:  aprlb voteddem_all post, model("")
         coef:  e(lb_coef)
(running aprub on estimation sample)

Bootstrap replications (1000)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
..................................................   400
..................................................   450
..................................................   500
..................................................   550
..................................................   600
..................................................   650
..................................................   700
..................................................   750
..................................................   800
..................................................   850
..................................................   900
..................................................   950
..................................................  1000

Bootstrap results                               Number of obs     =        701
                                                Replications      =      1,000

      command:  aprub voteddem_all readsome post, model("")
         coef:  e(ub_coef)
```

Parameter	Bound Estimate		80% Conf. Interval	
Average Persuasion Rate	.070732	.634288	.02957	.661439

Note: 80% conf. interval is based on percentile bootstrap.

```
        The conf. level is one-sided for the lower and upper bounds separately.

Reference: Jun and Lee (2019), arXiv:1812.02276 [econ.EM]
```

The 80% bootstrap confidence interval is almost the same as the previous confidence interval based on Stoye (2009), although the former is simply based on Bonferroni correction. This is because the width of the identified interval is so large that the method of Stoye (2009) is virtually equivalent to Bonferroni correction.

We now move to estimation of the local persuasion rate.

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

-----------------------------------------------------------------
persuasio4ytz2lpr: Causal inference on the local Persuasion rate
 when outcome, instrument and instrument are observed
-----------------------------------------------------------------

 - Binary outcome: voteddem_all
 - Binary treatment: readsome
 - Binary instrument: post
 - Covariates (if exist):
 - Regression model (if specified):

------------------------------------------------------------------
              Parameter | Estimate            80% Conf. Interval
------------------------------------------------------------------
   Local Persuasion Rate | .806747             .124333        1
------------------------------------------------------------------

Note: 80% conf. interval is based on normal approximation.
   Conf. interval is missing (given as [0,1])  if interactions are
   allowed between x and z. Use option bootstrap for that case.

Reference: Jun and Lee (2019), arXiv:1812.02276 [econ.EM]
```

The local persuasion rate for the group of compliers is point-estimated by 81%. It is interesting to note that the estimate of the LPR is so large that it is greater than the upper bound on the APR. This suggests that individuals are highly heterogeneous in this example and it is potentially misleading to draw a conclusion on the general population from the estimate on the compliers only. We also note that the LPR is point-identified but its confidence interval is quite large: $[12\%, 100\%]$. Thus, sampling uncertainty diminishes substantially the value of point identification for the LPR.

We now pretend the treatment variable is unobserved.

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

-----------------------------------------------------------------
persuasio4yz: Causal inference on the average persuasion rate
 when binary outcomes and binary instruments are observed
-----------------------------------------------------------------

 - Binary outcome: voteddem_all
 - Binary instrument: post
```

```
------------------------------------------------------------
   Parameter |  Estimate     [80% Conf. Interval]
-------------+----------------------------------------------
 Lower Bound |   .070732        .028802         1
------------------------------------------------------------

Note: 80% one-sided conf. interval is based on normal approximation.

Reference: Jun and Lee (2019), arXiv:1812.02276 [econ.EM]
```

The lower bound on the APR is the same as before and the left-end point of the confidence interval changes slightly compared to the case where subcommand `ytz` is used. However, the upper bound is trivially 1 because there is no information on the upper bound without observing the treatment variable `readsome`.

Lastly, we illustrate the `calc` subcommand.

```
. foreach var in voteddem_all readsome {
  2.         foreach treat in 0 1 {
  3.                 qui sum `var´ if post == `treat´
  4.                 scalar `var´_`treat´ = r(mean)
  5.         }
  6. }
. persuasio calc voteddem_all_1 voteddem_all_0 readsome_1 readsome_0
--------------------------------------------------------------
calc4persuasio: Calculate the effect of persuasion when info.
   on Pr(y=1|z) and/or Pr(t=1|z) for each z=0,1 is available

---------------------------------------------------------------
               Parameter |  Lower Bound       Upper Bound
-------------------------+-------------------------------------
 Average Persuasion Rate |   .070732           .783217
   Local Persuasion Rate |   .77591                  1
---------------------------------------------------------------
```

The scalars `voteddem_all_1` and `voteddem_all_0` are sample proportions corresponding to $\Pr(Y_i = 1|Z_i = z)$ for $z = 1, 0$. Similarly, the scalars `readsome_1` and `readsome_0` are sample proportions corresponding to $\Pr(T_i = 1|Z_i = z)$ for $z = 1, 0$. When the marginals of (Y_i, Z_i) and (T_i, Z_i) are used separately, the upper bound of θ increases from 63% to 78%. Further, θ_{local} is no longer point estimated but we only know that it is bounded between 78% and 100%. This difference illustrates the loss of identification power if we do not observe the joint distribution of (Y_i, T_i, Z_i). The subcommand `calc` is meant to be a quick calculator of the persuasion rates when the summary statistics are available. Hence, it does not provide confidence intervals. The following Stata output shows the result when only `voteddem_all_1` and `voteddem_all_0` are given.

```
. persuasio calc voteddem_all_1 voteddem_all_0
--------------------------------------------------------------
calc4persuasio: Calculate the effect of persuasion when info.
   on Pr(y=1|z) and/or Pr(t=1|z) for each z=0,1 is available
```

```
Parameter               | Lower Bound       Upper Bound
------------------------+-----------------------------------
Average Persuasion Rate |  .070732              1
  Local Persuasion Rate |  .070732              1
------------------------------------------------------------

Note: Exposure rates, Pr(t=1|z) for z=0,1, are missing.
```

If there is no information on $\Pr(T_i = 1|Z_i = z)$, the lower bounds on the APR and LPR are the same and their upper bounds are simply 1.

3.2 With Covariates

In this section, we illustrate how to use the `persuasio` command with covariates. As a simple example, we consider a scalar covariate: an indicator for wave of the study in the GKB dataset (the variable name is `MZwave2`). GKB launched their study in two waves each a week apart. It is not expected that this covariate would affect any of the estimation results reported in the previous subsection. The main purpose of this subsection is to explain the changes in the `persuasio` command when there exist covariates.

We start with the subcommand `apr`.

```
. persuasio apr voteddem_all readsome post MZwave2

-----------------------------------------------------------------
persuasio4ytz: Causal inference on the Average Persuasion Rate
 when outcome, instrument and instrument are observed
-----------------------------------------------------------------

 - Binary outcome: voteddem_all
 - Binary treatment: readsome
 - Binary instrument: post

------------------------------------------------------------------
              Parameter | Bound Estimate      95% Conf. Interval
------------------------+-----------------------------------------
 Average Persuasion Rate |  .071943  .635921          .         .
------------------------------------------------------------------

Note: 95% conf. interval is based on normal approximation
       using the method of Stoye (2009).
       Conf. interval is missing if covariates are present.
       Use option bootstrap for that case.

Reference: Jun and Lee (2019), arXiv:1812.02276 [econ.EM]
```

It can be seen that the confidence interval is missing in the Stata output. To obtain the confidence interval, we need to rely on bootstrap inference when there exist covariates in the regression model.

```
. set seed 339487731
. qui persuasio apr voteddem_all readsome post MZwave2, ///
```

```
>         level(80) method("bootstrap") nboot(1000)
. * display estimation results
. mat list e(apr_est)
e(apr_est)[1,2]
           c1         c2
r1  .07194256  .63592148
. mat list e(apr_ci)
e(apr_ci)[1,2]
           c1         c2
r1  .03040678  .66262102
. qui persuasio apr voteddem_all readsome post MZwave2, ///
>         level(80) model("interaction") method("bootstrap") nboot(1000)
. * display estimation results
. mat list e(apr_est)
e(apr_est)[1,2]
           c1         c2
r1  .07247338  .63866045
. mat list e(apr_ci)
e(apr_ci)[1,2]
           c1         c2
r1   .0279203  .66419843
```

The first specification adds the covariate `Mzwave2` separately, which is a default option; the second specification allows for interactions between the instrument `post` and the covariate `Mzwave2` by estimating the regression model of `voteddem_all` on `readsome` and `Mzwave2` separately for each value of `post`. The output of the `persuasio` command is suppressed, but its stored results are called via `e(apr_est)` and `e(apr_ci)`, which are bound estimates and the confidence interval for the average persuasion rate, respectively. The estimation results are almost identical between the two specifications and are also very similar to those without the covariate `Mzwave2`.

We now move to the local persuasion rate.

```
. persuasio lpr voteddem_all readsome post MZwave2, level(80)

-----------------------------------------------------------------
persuasio4ytz2lpr: Causal inference on the local Persuasion rate
 when outcome, instrument and instrument are observed
-----------------------------------------------------------------

 - Binary outcome: voteddem_all
 - Binary treatment: readsome
 - Binary instrument: post
 - Covariates (if exist): MZwave2
 - Regression model (if specified):

------------------------------------------------------------------
              Parameter | Estimate            80% Conf. Interval
------------------------------------------------------------------
   Local Persuasion Rate | .831206             .135306        1
------------------------------------------------------------------

Note: 80% conf. interval is based on normal approximation.
   Conf. interval is missing (given as [0,1])  if interactions are
```

```
   allowed between x and z. Use option bootstrap for that case.

Reference: Jun and Lee (2019), arXiv:1812.02276 [econ.EM]
. set seed 339487731
. qui persuasio lpr voteddem_all readsome post MZwave2, ///
>         level(80) model("interaction") method("bootstrap") nboot(1000)
. * display estimation results
. mat list e(lpr_est)
symmetric e(lpr_est)[1,1]
           c1
r1  .74301564
. mat list e(lpr_ci)
e(lpr_ci)[1,2]
    c1  c2
r1   0   1
```

The default option is to add covariates separately. In this case, inference based on normal approximation is used. The resulting point estimate is 83% with the 80% confidence interval $[14\%, 100\%]$. When interactions are allowed for, the point estimate is 74% and the confidence interval is the entire $[0\%, 100\%]$. As in the case without the covariate `Mzwave2`, sampling uncertainty associated with the LPR seems quite large.

4 References

Gerber, A. S., D. Karlan, and D. Bergan. 2009. Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics* 1(2): 35–52.

Jun, S. J., and S. Lee. 2019. Identifying the Effect of Persuasion. ArXiv:1812.02276. URL `https://arxiv.org/abs/1812.02276`.

Stoye, J. 2009. More on Confidence Regions for Partially Identified Parameters. *Econometrica* 77(4): 1299–1315.