

Syntax for **psasvy** in Stata

Selma Walther*

May 2016

psasvy – Calculate treatment effects or bounds under proportional selection of observables and unobservables (as in **psacalc**) for data that has been survey-set (**svyset**). It also allows the use of the **subpop** command.

Acknowledgment - This program is a modification of **psacalc** by Oster (2015). It is assumed that the user is familiar with the program **psacalc** and the associated help file.

1 Syntax

psasvy *indep_var type, outcome(varname) [options1]*
options1 Description

mcontrol(varlist): controls to be included in all regressions

rmax(#): maximum R-squared; default is *rmax=1*

delta(#): value of delta if requesting exact calculation of treatment effect or maximum R-squared; default is *delta=1*

beta(#): value of beta if requesting a value of delta or *rmax* to match exact beta; default is *beta=0*

controls (varlist): controls to be included in *controlled* regression i.e. the observables that inform on unobservables

outcome (varname): the name of the outcome variable

2 Description

psasvy evaluates the potential degree of omitted variable bias as in the Stata program **psacalc**, but for data that has been survey set using the Stata command **svyset**. Detailed information can be found in the help file for **psacalc**. Details of the calculation and theory are in Oster (2015).

*Department of Economics and Nuffield College, University of Oxford. Email address: selma.walther@economics.ox.ac.uk.

psasvy differs from psacalc in the following ways:

- 1) Data must be svyset by the user
- 2) The user must specify the outcome variable in "outcome(*varname*)" and the controls in "controls(*varlist*)". psasvy does not allow calculation after a controlled regression; instead, the controlled regression is estimated by psasvy.
- 3) There is no weights option as weights are assumed to have already been declared in svyset.
- 4) There is an optional subpopulation command "subpop(*varname*)", which allows the user to specify a variable which determines the subpopulation that is being used. If subpop(*varname*) is specified, then the program estimates "svy, subpop(*varname*): regress". See the Stata help file for **subpop** for more details.

3 Example

In this example, the data is survey set with the village primary sampling unit, household probability weights and rural and urban strata. The regression is the effect of the availability of a village market on household expenditure, with controls for education and age in all regressions, and controls for variables such as household size, land size, and the household distance to the nearest road, in the ‘controlled’ regression. The latter controls inform on the unobservables. The variable sample is a dummy variable that equals one if the household is in the using sample, and zero otherwise. An R^2 -max of 0.7 is set (see Oster 2014 for guidance on choosing an R^2 -max). The command estimates a bound on the treatment effect.

```
svyset villageID [pweight=hh_wgt], strata(urban) || _n
```

```
psasvy villagemarket set, mcontrol(education age) outcome(expenditure) controls(HHsize  
landsize maize tobacco distanceroad rainfall temperature) subpop(sample) rmax(0.7)
```

4 References

Oster, Emily. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Mimeo* (2015).