

MATCHING ESTIMATORS IN STATA

nnmatch

- The original version was called "match" instead of "nnmatch" (for "nearest neighbor"). This is a covariate-metric matching routine, with standardized Euclidean distance as the default metric. Mahalanobis distance can be invoked by adding the switch "metric(maha)". User-defined distance metrics can also be employed. The *nnmatch* routine does not handle propensity score matching.
- You can upload the *nnmatch* routine to your own computer in three simple steps:
 - Make sure your computer is connected to the Internet.
 - Start up Stata
 - Issue the following command from the Stata command window: *ssc*
install nnmatch
- Solely designed for nearest-neighbor matching (hence the "nn" prefix in its name). The default is one-to-one matching; one-to-many matching is accomplished by adding the "m(#)" switch
- The "robust(#v)" switch provides a nonparametric heteroskedasticity-consistent standard error for the estimate of treatment effect. The input "#v" represents the number of matches over which the robust estimate is based (usually, but not necessarily, the same number as the number of controls matched to each treatment observation).
- Offers the "exact" switch to perform exact matches as in perfect algorithmic matching.
- Other switches for bias adjustment and matching with replacement. See help file and the reference for details.
- Reference: A. Abadie, D. Drukker, J. Herr, & G. Imbens (2001), "Implementing Matching Estimators for Average Treatment Effects", *The Stata Journal*, 1(1):1-18

psmatch2

- A propensity score matching routine that also offers Mahalanobis matching as an option. Remarkably general and flexible.
- You can upload *psmatch2* to your own computer in three simple steps:
 - Make sure your computer is connected to the Internet.
 - Start up Stata
 - Issue the following command from the Stata command window:
ssc install psmatch2
- *psmatch2* supports a dialog box that makes using this routine easy for anyone familiar with the matching literature. To execute the dialog box, enter "*db psmatch2*". A schematic map of the *psmatch2* dialog box can be found on p. 4, below.

- Once familiar with *psmatch2*, you may find it easier to issue commands directly by modifying earlier commands rather than working through the dialog box.
- By default, the propensity score is estimated via a probit model. To use logit instead, add the "logit" switch
- Offers k-nearest-neighbors, radius/caliper, as well as two sophisticated nonparametric matching options (kernel matching and local linear regression matching).
- Generates several useful output variables that will appear in your Stata variable list after executing the command:
 - *_pscore*: The propensity score for each observation
 - *_support*: A dummy variable indicating whether or not an observation lies on the common support of the covariates for treatment and control.
 - *_outcome_variable*: For every treatment observation, this is the value of the matched outcome
- The *psmatch2* routine does calculate the standard error for the overall treatment effect, but there is little or no theory to justify textbook formulas, especially when conducting one-to-many matching. Users are expected to bootstrap the standard error. In Stata this can be done by embedding the *psmatch2* command within Stata's bootstrap command ("bs"). The bootstrap literature recommends specifying between 100 to 500 bootstrap replications, something that you accomplished by adding the switch "reps(n)" to the *bs* command line – without quotation marks – where 'n' indicates the number of replications desired (50 is the rather low default).
 - Notice that if you choose to use the dialog box to execute *psmatch2*, Stata generates the Stata line command for you, storing it in your "Review" window in case you want to use it as a shortcut in the future. To embed your *psmatch2* command within the *bs* command, first retrieve the *psmatch2* line command from the "Review" window and prefix it with the appropriate *bs* commands.
 - Example for Stata 9:
 - *bs r(att), reps(500)*: [your *psmatch2* command]
 - The "r(att)" argument in the command prefix tells bootstrap to focus on bootstrapping the value of the variable for the "average treatment effect on the treated"
 - The *bs* routine provides:
 - An estimate of the overall treatment effect. This is the mean of the estimated treatment effect drawn from the bootstrap replications, a figure that does not necessarily match the original estimated treatment effect.
 - An estimate of the bias associated with bootstrapping this mean effect.
 - A bootstrap estimate of the standard error of the treatment effect.

- A conventional confidence interval for the treatment effect, but one based on the bootstrap estimates of mean and standard error (labeled the "Normal" confidence interval).
 - To carry out a bootstrapping exercise to estimate the standard error of the ATT or ATE, prefix the *psmatch2* command that the dialog box created for you with "*bs att, reps(500):*"
- Using the *bs estat* post-estimation command after running the bootstrap provides you with additional information:
 - A nonparametric (robust) confidence interval based on percentiles of the bootstrap distribution.
 - A bias-corrected estimate of the confidence interval of the treatment effect.
- The *psmatch2* routine comes packaged with a useful companion routine, *psgraph*, that provides an elegant comparison of the distribution of the propensity scores for the treatments and controls. With it you can visualize the degree of common support shared by the treatment and control groups, respectively. All you need to do is issue the *psgraph* command immediately after the results of the *psmatch2* command are displayed.
- Also included with *psmatch2* is a separate routine, *pstest*, which compares the means of the covariates that you specify before and after matching. See its help file for details
- Reference: B. Sianesi (2001), "Implementing Propensity Score Matching Estimators with STATA", University College London
 - Available on the course Blackboard.
 - In outline form and poorly written, but this reference still provides some insights.

**THE PSMATCH2 DIALOG BOX:
THE STRUCTURE OF THE "db ps2match" FEATURE IN STATA 8**
(options for a basic, simple run are italicized)

MATCHING METRIC

- *Propensity Score*
 - *Estimate score*
 - *Provide score*
- *Mahalanobis*

PROPENSITY SCORE

ESTIMATION SETTINGS

- *Variable List* (covariates for the logit model)
- *Logit estimation*

OUTCOME

- *Treatment variable*
- *Outcome variable*
- *Also estimate ATE*

MATCHING METHOD

- *Nearest neighbor*
 - *# neighbors*
- *Kernel*
 - *Kernel type*
 - *Bandwidth*
- *Local linear regression*
- *Match without replacement*
(default = with replacement)

SUPPORT/TRIM

- *Common support*
- *Trim tails*
- *Caliper*