Propensity Scores for Multiple Treatments

A Tutorial on the MNPS Command for Stata Users

Matthew Cefalu, Maya Buenaventura

For more information on this publication, visit www.rand.org/t/TL170z1

Published by the RAND Corporation, Santa Monica, Calif.

Propensity scores for Multiple Treatments: A Tutorial on the mnps Command for Stata Users

Matthew Cefalu and Maya Buenaventura[1]
RAND Corporation

November 2016

1 Introduction

1.1 Brief overview

The Toolkit for Weighting and Analysis of Nonequivalent Groups, **twang**, was designed to make causal estimates when comparing two treatment groups. The package was developed in the R statistical computing and graphics environment and ported to Stata through a family of commands available at http://www.rand.org/statistics/twang/downloads.html.

The **twang** package in R also handles more than two treatment conditions through the **mnps** function, which stands for multinomial propensity scores. McCaffrey et al. (2013) describe the methodology behind **mnps**. The purpose of this document is to describe the syntax and features related to the implementation of the **mnps** command in Stata.

At a high level, the **mnps** command decomposes the propensity score estimation into several applications of the **ps** command, which was designed for the standard dichotomous treatment setting. For this reason, users who are new to **twang** are encouraged to learn about the **ps** command before using the **mnps** command. A tutorial describing the use of **twang** commands for comparing two treatments is found at http://www.rand.org/statistics/twang/stata-tutorial.html. The information in that tutorial, including directions on installing R, will not be repeated here. This tutorial is draws heavily on tutorials written for R and SAS users, which can be found at http://www.rand.org/statistics/twang/tutorials.html.

1.2 Set-up

If you have not already done so, you will need to download **twang** ado files and supporting documents to a folder on your computer. The files are available at http://www.rand.org/content/dam/rand/www/external/statistics/twang/stata/twang_stata_tutorial.zip They include the following files (among others):

- twang Stata package – files containing Stata programs and help files with details on implementing the commands
- aod.dta – Example dataset from the McCaffrey et al. (2013) Study
- mnps_tutorial_code_using_macros.do – Stata code from examples presented in this tutorial using global macros to keep track of the location of adofiles, data, and output

The datasets and example code will be useful for trying the code presented in this tutorial. Those files are not necessary for you to run your own applications.

[1] The development of this software and tutorial was funded by National Institute of Drug Abuse grants number 1R01DA015697 (PI: McCaffrey) and 1R01DA034065 (PIs: Griffin/McCaffrey).

To use the ado files and help files, you can place the files in the PERSONAL ado-path directory, which can be identified using **adopath**. Alternately, you can place the files in any directory, and add the directory to ado-path using the command[2]

```
. adopath + "C:\Users\uname\adofile"
```

After adding the directory to ado-path, the help files can be opened using the help command together with the command that you need help with (*e.g.*, "**help mnps**"). Note that **adopath** temporarily adds the directory to the ado-path — you must rerun the command each time Stata is opened.

The ado files will run code in R and import the results into your Stata session. All files created by the macro are stored in the directory specified by the user in the **objpath** option as seen in subsequent sections. Any files in this directory created from previous calls of **mnps** will be overwritten.

The ado files also create a "Twang" folder in a standard location based on the operating system. In Windows, the folder is in the user's AppData\Local folder (C:\Users\username\AppData\Local\Twang would be the default for a user with the "username" as his or her username). In Mac OS-X, the folder is in the user's Library folder (/Users/username/Library/Twang). This folder will remain on the user's hard drive until it is removed. Users can remove the folder using any method they would normally use for removing a folder when they no longer plan to use **twang**.

2 An ATE example

To demonstrate how to implement an analysis using the **mnps** command in Stata, we use a random subset of the data described in McCaffrey et al. (2013). This truncated dataset is called aod.dta. As noted in Section 1, the macros and data can be downloaded from http://www.rand.org/statistics/twang/downloads.html. This example study includes three treatment groups, and the data include records for 200 youths in each treatment group of an alcohol and other drug treatment evaluation.

For the AOD dataset, the variable treat contains the treatment information, which take the values community, metcbt5, and scy. The other variables included in the dataset are:

- suf12: outcome variable, substance use frequency at 12-month follow-up
- illact: pretreatment covariate, illicit activities scale
- crimjust: pretreatment covariate, criminal justice involvement
- subprob: pretreatment covariate, substance use problem scale
- subdep: pretreatment covariate, substance use dependence scale
- white: pretreatment covariate, indicator for non-Hispanic white youth

In an observational study with multiple treatments, there are several quantities that one may be interested in estimating. The estimands that are most commonly of interest are the average treatment effect on the population (ATE) and the average treatment effect on the treated (ATT). The differences between these quantities are explained at length in McCaffrey et al. (2013), but in brief the ATE answers the question of how, on average, the outcome of interest would change if everyone in the population of interest had been assigned to a particular treatment relative to if they had all received another single treatment. The ATT answers the question of how the average outcome would change if everyone who received one particular treatment had instead received another treatment. We first demonstrate the use of **mnps** when ATE is the effect of interest and then turn to using the function to support estimation of ATT.

[2] Users might be tempted to copy and paste code from this PDF document into an editor to run this example code. We advise against this. Text from the PDF file may not appear the same in a text editor as it does in the PDF file; symbols or spaces may be added. The file "mnps_tutorial_code.do" file that is available with the twang ado files contains all the code from this tutorial in a text file. Analysts can use that file to copy the code and run the examples.

2.1 Estimating the weights

The following is an example of the command for running **mnps** to obtain propensity score weights for three or more treatment groups[3]:

```
. mnps treat illact crimjust subprob subdep white, ///
     ntrees(3000) intdepth(3) shrinkage(0.01) ///
     stopmethod(es.mean ks.mean) estimand(ATE) ///
     rcmd(C:\Program Files\R\R-3.3.1\bin\Rscript.exe) ///
     objpath(C:\Users\username\twang\output) ///
     plotname(C:\users\username\twang\output\mnps_example_plot.pdf)
. save C:\Users\username\twang\output\aod_ate_wgts, replace
```

The syntax of the **mnps** command requires some discussion. All of the variables that are to be used in the model are listed after the command name and before the comma. The first variable is the treatment variable, which as noted earlier is "treat". The variable specified by the treatment variable must take on at least three values indicating three or more treatment groups. If your study involves only two treatment conditions, then the **ps** command must be used instead of **mnps**. The treatment variable is followed by the names of the covariates.

There are a number of options related to fine-tuning the generalized boosted model GBM upon which the **twang** methods are built. The option **ntrees** is the maximum number of iterations that the GBM will run. There will be a warning if the estimated optimal number of iterations is too close to the bound selected in this option because it indicates that balance may improve if more complex models (i.e., those with more trees or a larger value for **ntrees**) are considered. The user should increase **ntrees** or decrease **shrinkage** and rerun it if this warning appears. The option **intdepth** controls the level of interactions allowed in the GBM, with larger values specifying more complex models. We specified a value of 2, indicating that the algorithm will consider all two-way interactions between covariates. The GBM estimation algorithm uses **shrinkage** to enhance the smoothness of resulting model. The **shrinkage** option controls the amount of **shrinkage**. Small values such as 0.005 or 0.001 yield smooth fits but require greater values of **ntrees** to achieve adequate fits. Computational time increases inversely with **shrinkage**. Additional details on **ntrees**, **intdepth**, and **shrinkage** can be found in McCaffrey, Ridgeway, and Morral (2004).

The balance criteria used to tune the propensity score model are specified in the **stopmethod** argument. As with the **ps** command, four stopping rule balance criteria are available for **mnps**. They are **es.mean**, **es.max**, **ks.mean**, and **ks.max**. The four stopping rules are defined by two components: a balance metric for each covariate and a rule for summarizing across covariates. A balance metric summarizes the difference between two univariate distributions of a single pretreatment variable (*e.g.*, illicit activities scale). The stopping rules in **twang** use two balance metrics: absolute standardized mean difference (ASMD; also referred to as the absolute standardized bias or the effect size (ES)) and the Kolmogorov-Smirnov (KS) statistic. The stopping rules use two different rules for summarizing across covariates: the mean of the covariate balance metrics ("mean") or the maximum of the balance metrics ("max"). The first piece of the stopping rule name identifies the balance metric (ES or KS) and the second piece specifies the rule for summarizing across balance metrics. For instance, **es.mean** uses the effect size or ASMD and summarizes across variables with the mean and the **ks.max** uses the KS statistics to assess balances and summarizes using the maximum across variables. The other two stopping rules use the remaining two combinations of balance metrics and summary statistics. In this example, we chose to examine both **es.mean** and **ks.mean**.

[3] This code will generate an error message from R that says "Optimal number of iterations is close to the specified n.trees. n.trees is likely set too small and better balance might be obtainable by setting n.trees to be larger." This error message can be ignored for now. As discussed later in this Section, after running the **mnps** command, a useful first step in the assessment of the weights is to make sure that the models run for a sufficiently large number of iterations in order to optimize the balance statistics of interest. In this example, 3,000 iterations is sufficient.

The **estimand** option is used to indicate whether the analyst is interested in estimating the average treatment effect (ATE) or the average treatment effect on the treated (ATT).

The options **rcmd** specifies the R program executable file for running R. The location of the file is determined through the installation of R. The default setup of R Version 3.3.1 on Windows 7 resulted in the executable being "C:\Program Files\R\R 3.3.1\bin\Rscript.exe". For other versions of R "3.3.1" needs to be replaced by the version number. If the analyst has added R to the path environmental variable then the path does not need to be included in the **rcmd** option. The specification of **rcmd** is not necessary for Mac OS-X, but the default location of the executable is "/Library/Frameworks/R.framework/Versions/3.3/Resources/bin/Rscript".

The option **plotname** gives the name for a file of default diagnostic plots that **mnps** creates. Creation of the plots is optional. If **plotname** is not given, then no plots are created. If the option contains a path, then the file with plots will be stored in the folder specified by it. Otherwise the file will be stored in the folder specified in the option **objpath.** In our example, we specify **plotname** as "C:\Users\username\twang\output\mnps_example_plot.pdf" so the plots will be stored in that folder.

The final option **objpath** specifies a folder where files created by the command to run the **twang** functions in R and return the results to Stata are stored. Namely, an "R object" ("mnps.RData") with the GBM fit information and a log of the R session ("mnps.Rout"). The **objpath** option is required for running the **mnps** command and must reference an existing directory.

The propensity score-based weights created by the **mnps** command can be found in new variables. In this example, we save this new dataset as "aod_ate_wgts" for future use. There is one weight variable for each stopping rule specified in **stopmethod**. The weight variables are named according to the stopping rule and estimand so that in this example there is a new variable in the dataset named 'esmeanate' with the weights from a GBM with the iterations chosen to minimize the mean standardized bias (effect size) and a second weight variable 'ksmeanate' with the weights from a GBM with the iterations chosen to minimize the mean KS statistic.

An essential component of propensity score weighting is the assessment of the quality of the weights. The **twang** commands provide both graphical and tabular displays to support that assessment. The default plots generated by specifying the **plotname** option in **mnps** are essential component of the initial exploration of the balance. However, the default is to create 5 different diagnostic plots, which may have multiple pages for some of the default graphics. The set of plots may become large when there are several treatment options and does not give users full control over the plots that **twang** provides. The **mnplot** command provides greater control over the plots to be created. It can be used to generate specific types of plots in separate files that can be helpful for identifying where balance problems occur, when they occur.

After running the **mnps** macro, a useful first step in the assessment of the weights before using them to estimate treatment effects is to make sure that we let the models run for a sufficiently large number of iterations in order to optimize the balance statistics of interest. We do this using the convergence plots created by **twang** to visually determine whether any of the balance measures of interest still appear to be decreasing at the maximum number of iterations specified by the **ntrees** argument, which we set to 3,000 for this example (10,000 iterations is the default). The convergence plots are one of the default plot types created by specifying the **plotname** option in **mnps**. The following code is an example of using **mnplot** to produce the convergence plot. This code produces just the convergence plots and saves them to a file:

```
. mnplot, plotname(mnps_example_plot_1-3.pdf) ///
      multipage plots(1)
```

The option **multipage** specifies that each panel created by the plot function be placed on a separate page. The comparison of multiple treatments often results in plot with multiple panels. By default the panels are placed on a single page and this can result in plots with problematic aspect ratios and which are difficult to read. Using the **multipage** option avoids this problem.

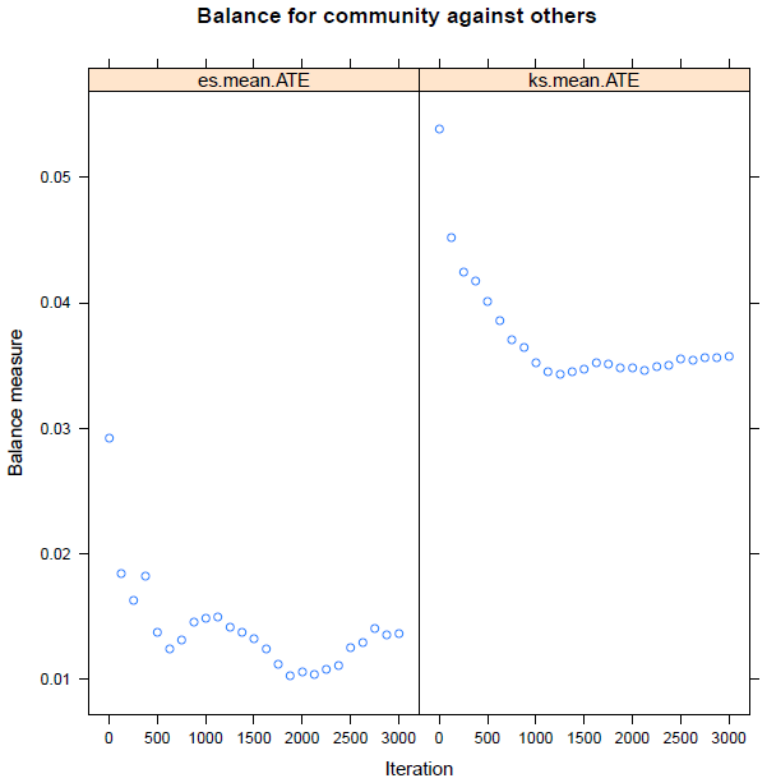

Figure 1: Example of an optimization plot for both stopping rules (**`es.mean`** and **`ks.mean`**) for estimating the propensity scores for comparing the Community treatment conditions to the combination of the other two to generate ATE weights for the AOD dataset.

As noted above, **`mnps`** estimates weights through repeated use of the **`ps`** function and comparison of each treatment to the pooled sample of other treatments. Thus, there is one convergence plot corresponding to each of those fits. Each plot is then further divided into one panel for each stopping rule used in the estimation. Since we used the "**`es.mean`**" and "**`ks.mean`**" stopping rules there are two panels in each plot. Figures 1 to 3 show the output from the **`mnplot`** code above. In these figures, it appears that each of the balance measures are optimized with substantially fewer than 3,000 iterations, so we do not have evidence that we should re-run the **`mnps`** command with a larger number of iterations.

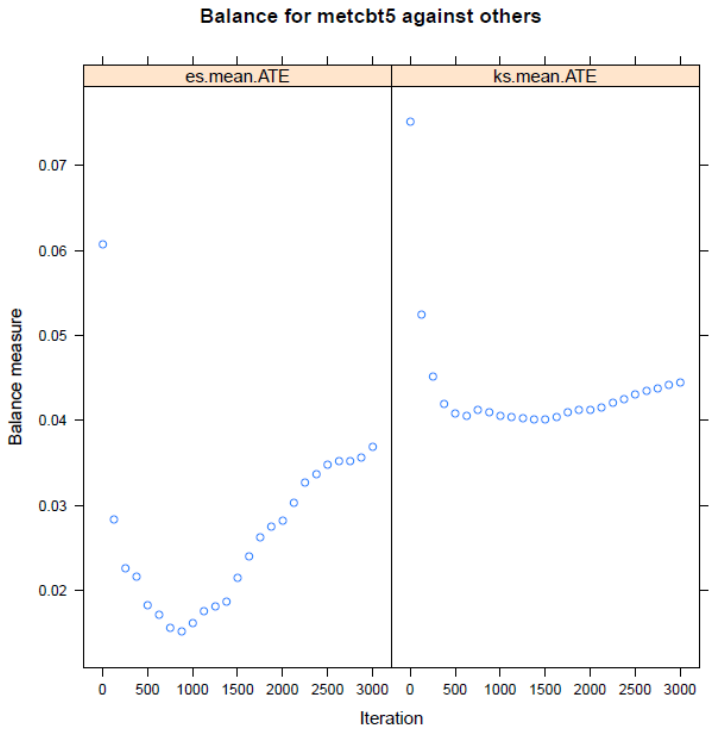

Figure 2: Example of an optimization plot for both stopping rules (**`es.mean`** and **`ks.mean`**) for estimating the propensity scores for comparing the MET/CBT-5 treatment conditions to the combination of the other two to generate ATE weights for the AOD dataset.

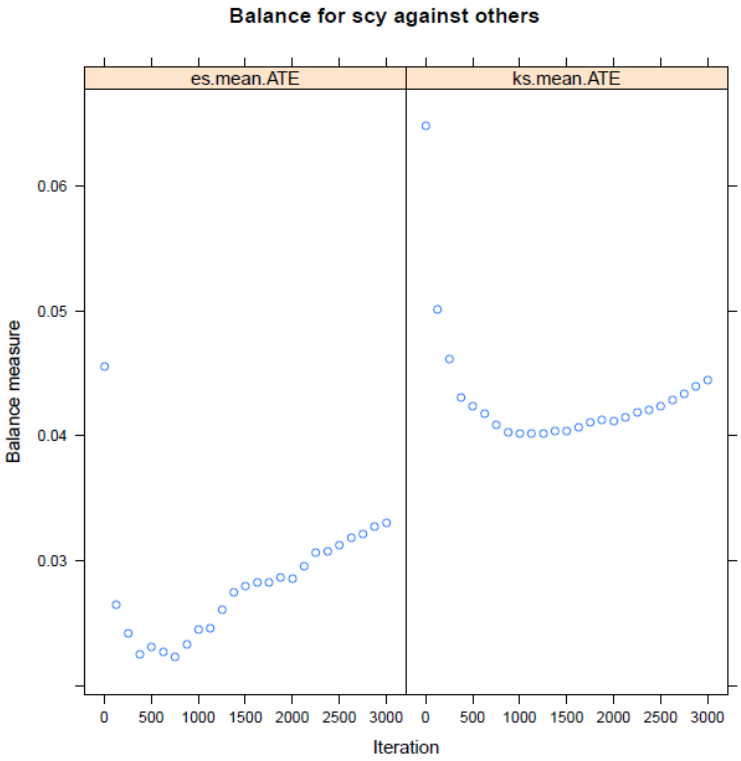

Figure 3: Example of an optimization plot for both stopping rules (**es.mean** and **ks.mean**) for estimating the propensity scores for comparing the SCY treatment conditions to the combination of the other two to generate ATE weights for the AOD dataset.

A useful second step is to check the key assumption in propensity score analyses that each experimental unit has a non-zero probability of receiving each treatment. The plausibility of this assumption may be assessed by examining the overlap of the empirical propensity score distributions. This diagnostic is available by setting the **plots** option to "2" when running the **mnplot** command. We use the **subset** option to specify which stopping rule we wish present in the plot.[4] The default panel layout for this plot results in readable figures so we do not use **multipage** option.

```
. mnplot, plotname(mnps_example_plot_4.pdf) ///
      plots(2) subset(1)
```

The code above produces Figure 4. The overlap plot uses data from only one stopping rule, by default the one that comes first alphabetically. Hence, without specifying **subset(1)**, the overlap for the weights from the model fit using that rule would be compared. To see the results for the "**ks.mean**" rule requires using **subset(2)**.

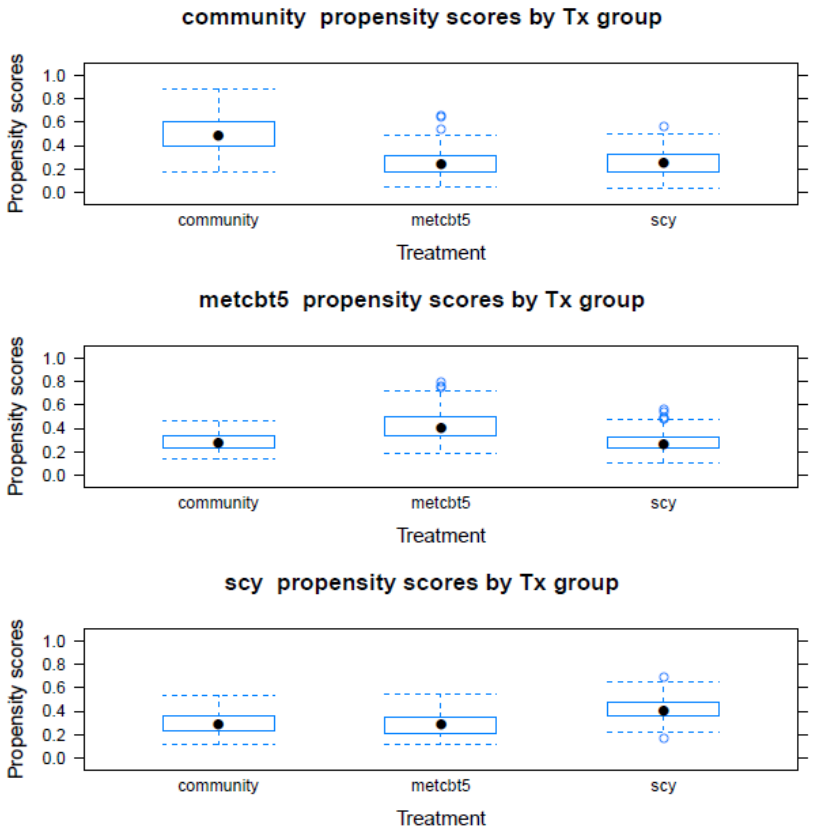

Figure 4: Example of an overlap boxplot for the **es.mean** stopping rule for estimating the propensity scores to generate ATE weights for the AOD dataset.

[4] The value for the **subset** argument can be a character variable with the name of the stopping rule, as was used in the example code, or a number corresponding to the stopping rule. Stopping rules are numbered by the alphabetical ordering among the rules specified in the **mnps** call.

As shown in Figure 4, the overlap assumption generally seems to be met in our example, although there should be some concern that adolescents in the metcbt5 and scy conditions do not overlap well with the community group given the top most graphic. See McCaffrey et al. (2013) for more details on this issue.

2.2 Graphical assessments of balance

As with the **ps** and **plot** commands for the binary treatment setting, **mnps** and **mnplot** also can generate plots to display information on commonly-used balance statistics. Checking balance is an essential part of any propensity score analysis and must be done thoroughly prior to moving into outcome analyses. The **twang** Stata commands provide the user with two ways to assess balance: graphical displays or tabular displays. Here we discuss how to create graphical displays of balance. Graphical displays can be produced by setting the **plots** argument for **mnplot** equal to "3", "4", or "5". In particular, when the **plots** argument is set equal to "3", **mnplot** provides comparisons of the absolute standardized mean differences (ASMD) between the treatment groups on the pretreatment covariates, before and after weighting. When the **plots** argument is set equal to "4", the display is of *t*-test and chi-squared statistic *p*-values from comparing the two groups before and after weighting. Setting the argument to "**5**" generates the corresponding *p*-value plots for tests of the KS statistics. However, whereas there is a single plot for these balance diagnostics in the binary treatment setting, in the multiple treatment case, one can either examine a plot for each of the pairwise comparisons (*e.g.*, Community versus the others, MET/CBT-5 versus the others, or SCY versus the others), or summarize the balance statistics, in some way, across the treatment conditions. As a default, the **mnplot** macro returns the maximum of the pairwise balance statistics across treatment groups for each of the covariates:

```
. mnplot, plotname(mnps_example_plot_5.pdf) plots(3)
```

The code above produces Figure 5. As shown in that figure, after propensity score weighting, the maximum ASMD decreases for all pretreatment covariates. The statistically significant difference (before taking the maximum across treatment groups) is indicated by the solid circle. One may see the balance plots for the individual fits by using the **nopairwisemax** option as in the following code.

```
. mnplot, plotname(mnps_example_plot_6.pdf) ///
      plots(3) nopairwisemax figurerows(3)
```

The additional **figurerows(3)** argument instructs the function to spread the plots over three rows, as shown in Figure 6. By default, the plots would be arranged in a single row rather than a single column. This produces an unreadable figure. We note here that red lines represent pretreatment covariates for which the pairwise ASMDs increase after weighting.

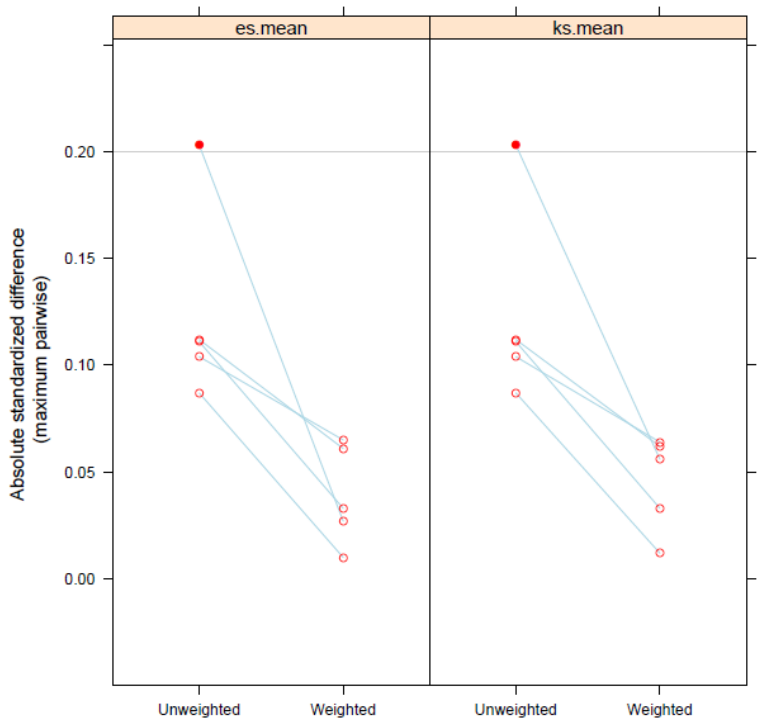

Figure 5: Example of a standardized effect size plot for estimating the propensity scores to generate ATE weights for the AOD dataset.

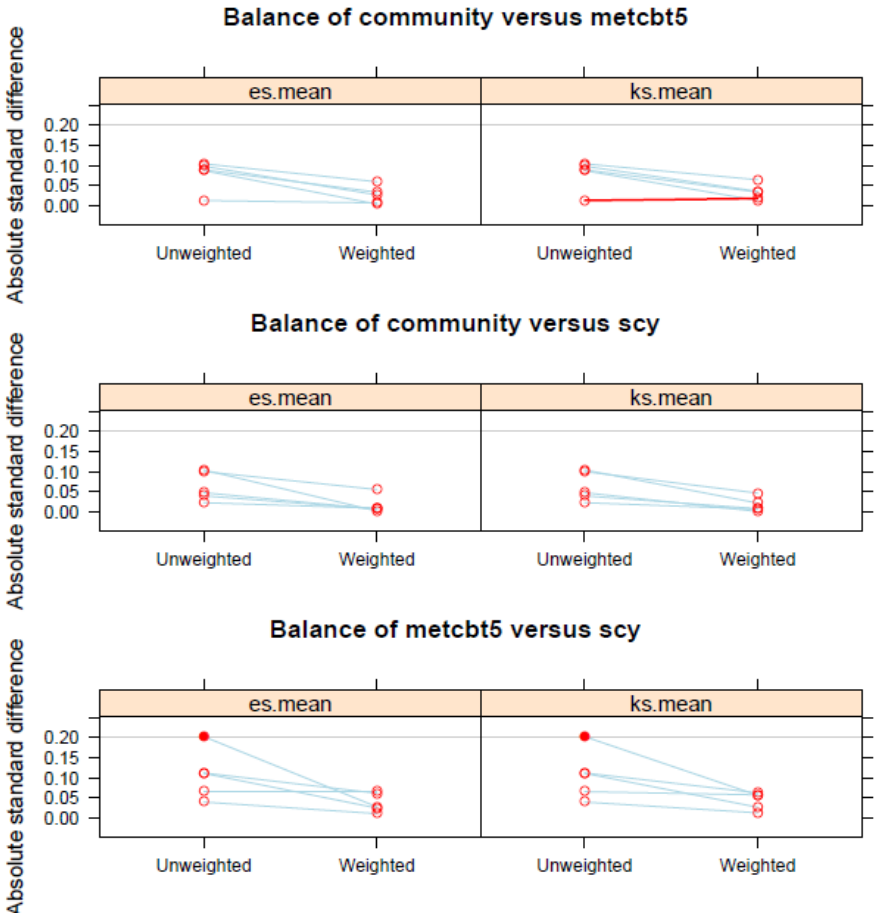

Figure 6: Example of a standardized effect size plot for each pairwise comparison of treatments for estimating the propensity scores to generate ATE weights for the AOD dataset.

Setting the **`plots`** argument equal to "4" displays *t*-test or chi-squared statistic pairwise minimum *p*-values for differences between each of the individual treatment groups and observations in all other treatment groups. The following command produces the results shown in Figure 7. As seen in that figure, the pairwise minimum *p*-values all increase after propensity score weighting.

```
. mnplot, plotname(mnps_example_plot_7.pdf) plots(4)
```

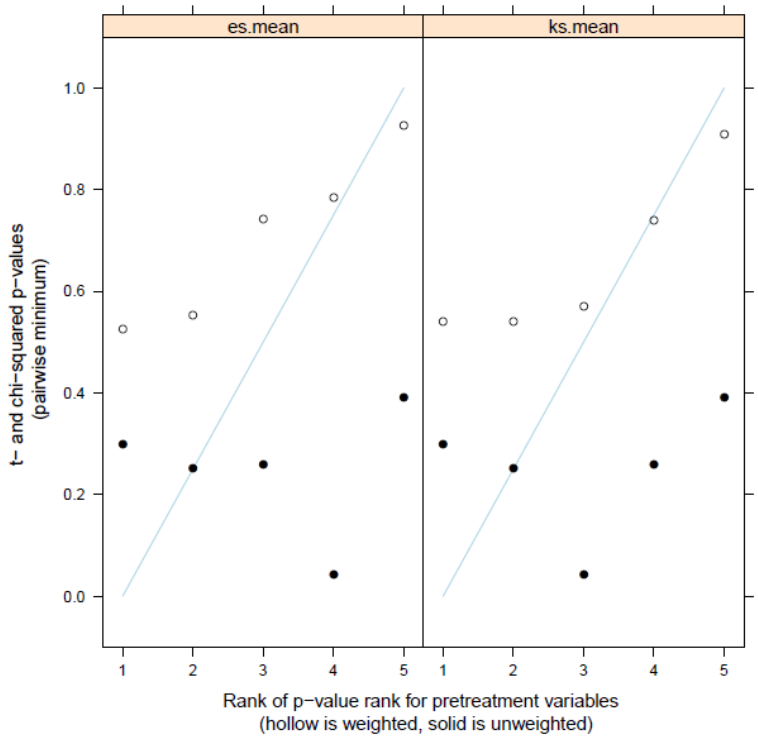

Figure 7: Example of a p-value plot for t-tests and Chi-square tests for estimating the propensity scores to generate ATE weights for the AOD dataset.

Some of the figures include many frames, which can result in figures that are difficult to read. There are three arguments to the **`mnplot`** to control the placement of multiple panels across pages in the graphics file. First, the **`treatments`** argument can be used to specify only comparisons that involve a specific treatment level or, in the ATE case, only comparisons between two specified treatment levels. Similarly, the **`singleplot`** argument can be used to plot only a single page from a call that will generate multiple pages of plots or multiple frames on a single page. For example, **`singleplot(2)`** would display only the second page of those produced by the plot command (see figure below). Finally, as described previously, specifying the **`multipage`** option prints in succession the frames generated by the plot function.

The following code and corresponding figure (Figure 8) demonstrate using these arguments to plot the p-values for the KS tests when comparing the treatment levels of community and scy. By specifying the **`nopairwisemax`** option, each pairwise comparison of treatment will be plotted, and by specifying the **`multipage`** option, each comparison will be on a separate page. The comparisons are ordered alphabetically so the first page is community versus metcbt5, the second is community versus scy, and the third is metcbt5 versus scy. By setting **`singleplot(2)`**, we select the p-values for the tests from the comparison of community with scy.

```
. mnplot, plotname(mnps_example_plot_8.pdf) ///
      plots(5) nopairwisemax ///
      multipage singleplot(2)
```

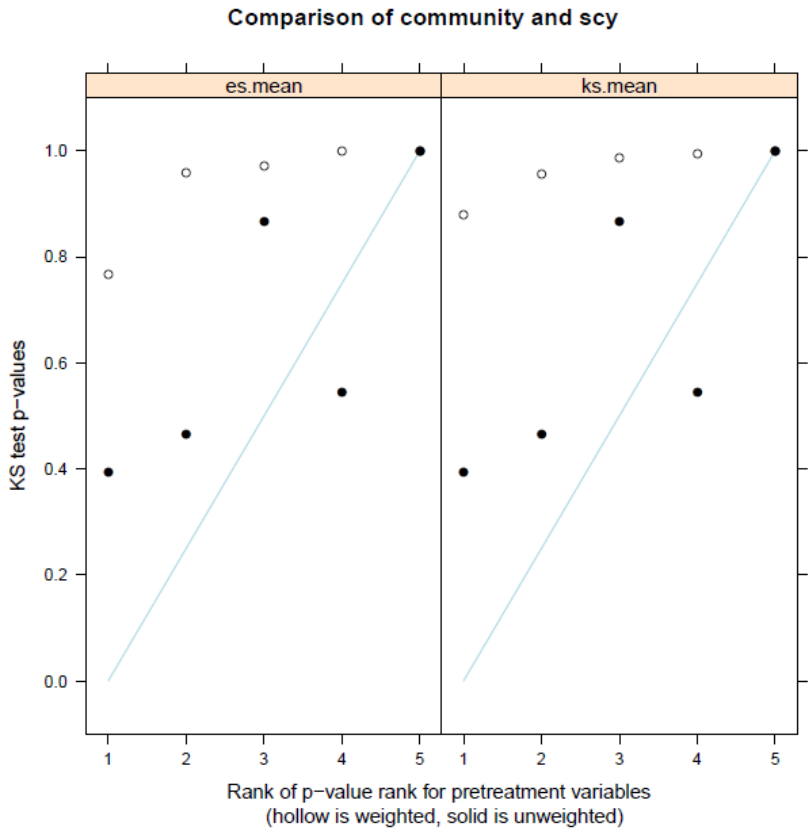

Figure 8: Example of a p-value plot KS tests for estimating the propensity scores to generate ATE weights for the AOD dataset, using control arguments to plot only the tests for the comparison of community to scy.

2.3 Tabular assessments of balance

There are two primary ways to obtain tabular assessments of balance via the **twang** macros in Stata. First, the **mnps** command returns a number of useful tabular summaries of the balance as part of its default output when running the command. Additionally, the user can make use of the **balance** command to obtain more fine-tuned and simplistic summaries of balance.

By default, the **mnps** command produces a Summary of Pairwise Comparisons table and a Sample Sizes and Effective Sample Sizes table. The Summary of Pairwise Comparisons table presents the maximum values of the balance statistics and minimum values of p-values across all covariates and all pairwise comparisons of the treatments (*e.g.*, in our case study Community versus MET/CBT-5; Community versus SCY; MET/CBT-5 versus SCY). There is one line in the table for the comparisons prior to weighting and one line for weighting with the weights generated by the model selected by each stopping rule. The table generated by our example **mnps** code follows. In our example, we use the "**es.mean**" and "**ks.mean**" stopping rules so the summary table has three rows. The columns correspond to the maximum ASMD (maxstd~z), the minimum p-value from pairwise tests (minp), the maximum KS statistic (maxks), and the minimum p-value from the KS test (minksp~l). The summary allow us to quickly see how the maximum ASMDs have gotten smaller and minimum *p*-values have gotten larger, as desired, after propensity score weighting.

```
Summary of pairwise comparisons
   ------------------------------------------------
              maxstd~z     minp    maxks minksp~l
   ------------------------------------------------
   unw           .2027    .0416      .13     .068
```

```
es_mean        .0653     .5253     .0666     .7664
ks_mean        .0645     .5395     .0646     .7985
-------------------------------------------------
```

The Sample Sizes and Effective Sample Sizes table has one row per treatment group and one column for the original sample size ("n") and then one column for the effective sample size for the weights generated using each stopping rule. As shown in the output from our **mnps** call, this summary table allows for a quick check of whether the weights are highly variable and could potentially yield a very imprecise treatment effect estimate. In this case the effective sample sizes are all close to the actual sample size because the groups are fairly well balanced even before weighting.

```
Sample sizes and effective sample sizes
-------------------------------------
                     n essesm~n essksm~n
-------------------------------------
community          200    184.5    187.5
metcbt5            200    186.2    183.4
scy                200    189.5    185.7
-------------------------------------
```

The **balance** command produces a table of balance statistics for each covariate for each pairwise comparison of treatments. A table generated using the **balance** command for our **mnps** example follows. This table contains balance information for unweighted comparisons and weighted comparisons using the weights generated under each specified stopping rule. Each record contains:

- tmt1 – the name of the first treatment group in the pairwise comparison; names are sorted alphabetically
- tmt2 – the name of the second treatment group in the pairwise comparison
- var – the name of the covariate being assessed
- mean1 – the covariate mean for the first treatment group
- mean2 – the covariate mean for the second treatment group
- popsd – the pooled within sample standard deviation from all treatment groups
- stdeffsz – the ASMD or absolute effect size equal to the absolute value of the difference in the group means divided by the pop sd
- p – the p-value of the t-test (continuous variables) or the Chi-squared test (categorical variables)
- ks – the KS statistic for comparing the covariate distribution for the two groups
- kspval – the approximate p-value for testing the KS statistic
- stopmethod – the stop method used for generating the weights or "unw" for the unweighted comparison.

The following code demonstrates the use of the **balance** command to create a balance table summarizing balance of covariates before and after weighting:

```
. balance, unweighted weighted
```

Balance Table Summarizing Balance of Covariates before and after Weighting for the ATE Example, generated using the **balance** command

Unweighted

tmt1	tmt2	var	mean1	mean2	popsd	stdeffsz	p	ks	kspval	stopme~d
community	metcbt5	illact	.097	.007	1.014	.089	.385	.1	.27	unw
community	metcbt5	crimjust	-.065	.037	1.041	.098	.328	.105	.221	unw
community	metcbt5	subprob	-.06	.026	.985	.087	.39	.09	.394	unw
community	metcbt5	subdep	.046	.058	1.031	.012	.91	.055	.924	unw
community	metcbt5	white	.16	.2	.383	.104	.298	.04	.997	unw
community	scy	illact	.097	.12	1.014	.022	.823	.06	.866	unw
community	scy	crimjust	-.065	-.174	1.041	.104	.295	.08	.545	unw
community	scy	subprob	-.06	-.013	.985	.047	.631	.09	.394	unw
community	scy	subdep	.046	-.058	1.031	.1	.312	.085	.466	unw
community	scy	white	.16	.175	.383	.039	.688	.015	1	unw
metcbt5	scy	illact	.007	.12	1.014	.111	.259	.11	.178	unw

	metcbt5	scy	crimjust	.037	-.174	1.041	.203	.042	.13	.068	unw
	metcbt5	scy	subprob	.026	-.013	.985	.039	.696	.065	.793	unw
	metcbt5	scy	subdep	.058	-.058	1.031	.112	.251	.09	.394	unw
	metcbt5	scy	white	.2	.175	.383	.065	.523	.025	1	unw

Weighted: esmean

	tmt1	tmt2	var	mean1	mean2	popsd	stdeffsz	p	ks	kspval	stopme~d
	community	metcbt5	illact	.085	.052	1.014	.033	.742	.057	.896	esmean
	community	metcbt5	crimjust	-.092	-.065	1.041	.026	.793	.054	.931	esmean
	community	metcbt5	subprob	-.013	-.016	.985	.003	.974	.062	.831	esmean
	community	metcbt5	subdep	.015	.021	1.031	.006	.958	.05	.965	esmean
	community	metcbt5	white	.173	.195	.383	.059	.582	.023	1	esmean
	community	scy	illact	.085	.077	1.014	.008	.937	.048	.97	esmean
	community	scy	crimjust	-.092	-.093	1.041	.001	.989	.037	.998	esmean
	community	scy	subprob	-.013	-.007	.985	.006	.949	.067	.766	esmean
	community	scy	subdep	.015	-.042	1.031	.055	.582	.051	.957	esmean
	community	scy	white	.173	.17	.383	.006	.95	.002	1	esmean
	metcbt5	scy	illact	.052	.077	1.014	.024	.812	.065	.793	esmean
	metcbt5	scy	crimjust	-.065	-.093	1.041	.027	.783	.057	.896	esmean
	metcbt5	scy	subprob	-.016	-.007	.985	.01	.925	.036	.999	esmean
	metcbt5	scy	subdep	.021	-.042	1.031	.061	.553	.065	.794	esmean
	metcbt5	scy	white	.195	.17	.383	.065	.525	.025	1	esmean

Weighted: ksmean

	tmt1	tmt2	var	mean1	mean2	popsd	stdeffsz	p	ks	kspval	stopme~d
	community	metcbt5	illact	.083	.05	1.014	.033	.74	.062	.839	ksmean
	community	metcbt5	crimjust	-.084	-.048	1.041	.035	.723	.052	.95	ksmean
	community	metcbt5	subprob	-.001	-.012	.985	.012	.908	.053	.934	ksmean
	community	metcbt5	subdep	.007	.024	1.031	.017	.873	.049	.966	ksmean
	community	metcbt5	white	.169	.194	.383	.064	.539	.025	1	ksmean
	community	scy	illact	.083	.077	1.014	.006	.95	.045	.985	ksmean
	community	scy	crimjust	-.084	-.106	1.041	.021	.83	.041	.995	ksmean
	community	scy	subprob	-.001	-.001	.985	0	1	.058	.879	ksmean
	community	scy	subdep	.007	-.04	1.031	.046	.65	.051	.955	ksmean
	community	scy	white	.169	.172	.383	.007	.946	.003	1	ksmean
	metcbt5	scy	illact	.05	.077	1.014	.026	.797	.062	.832	ksmean
	metcbt5	scy	crimjust	-.048	-.106	1.041	.056	.57	.064	.809	ksmean
	metcbt5	scy	subprob	-.012	-.001	.985	.012	.911	.035	.999	ksmean
	metcbt5	scy	subdep	.024	-.04	1.031	.062	.541	.065	.799	ksmean
	metcbt5	scy	white	.194	.172	.383	.057	.58	.022	1	ksmean

For propensity score analyses with multiple treatments, the balance table information returned can be quite overwhelming and, with many covariates, sorting through that information can be challenging. More parsimonious versions of the summaries are available using the **`collapseto`** option with the **`balance`** command.

The following code produces a table of balance diagnostics, by pretreatment covariate and stop method:

```
. balance, unweighted weighted collapseto(covariate)
```

Unweighted

	var	maxstd~z	minp	maxks	minksp~l	stopme~d
	crimjust	.203	.042	.13	.068	unw
	illact	.111	.259	.11	.178	unw
	subdep	.112	.251	.09	.394	unw
	subprob	.087	.39	.09	.394	unw
	white	.104	.298	.04	.997	unw

Weighted: esmean

	var	maxstd~z	minp	maxks	minksp~l	stopme~d
	crimjust	.027	.783	.057	.896	esmean
	illact	.033	.742	.065	.793	esmean
	subdep	.061	.553	.065	.794	esmean
	subprob	.01	.925	.067	.766	esmean

```
|    white       .065   .525    .025          1     esmean |
+-------------------------------------------------------------+
```

```
Weighted: ksmean
```

```
+-------------------------------------------------------------+
|      var   maxstd~z   minp   maxks   minksp~l   stopme~d |
|-------------------------------------------------------------|
| crimjust       .056    .57    .064       .809     ksmean |
|   illact       .033    .74    .062       .832     ksmean |
|   subdep       .062   .541    .065       .799     ksmean |
|  subprob       .012   .908    .058       .879     ksmean |
|    white       .064   .539    .025          1     ksmean |
+-------------------------------------------------------------+
```

The following code produces a table of balance diagnostics, by stop method.

```
.balance, unweighted weighted collapseto(stop.method)
```

```
Unweighted
```

```
+-----------------------------------------------+
| maxstd~z   minp   maxks   minksp~l   stopme~d |
|-----------------------------------------------|
|     .203   .042     .13       .068        unw |
+-----------------------------------------------+
```

```
Weighted: esmean
```

```
+-----------------------------------------------+
| maxstd~z   minp   maxks   minksp~l   stopme~d |
|-----------------------------------------------|
|     .065   .525    .067       .766     esmean |
+-----------------------------------------------+
```

```
Weighted: ksmean
```

```
+-----------------------------------------------+
| maxstd~z   minp   maxks   minksp~l   stopme~d |
|-----------------------------------------------|
|     .064   .539    .065       .799     ksmean |
+-----------------------------------------------+
```

After examining the graphical and tabular diagnostics provided by **twang**, we can analyze the outcome variable using the propensity score weights generated by the **mnps** function. Although two stop methods were specified initially ("**es.mean**" and "**ks.mean**"), at this point we have to commit to a single set of weights. From the balance table above, we see that the balance properties are very similar for the two stopping rules, and from the Sample Sizes and Effective Sample Sizes Table, we see that the effective sample sizes (ESS) are similar as well. Hence, we expect the two stop methods to give similar results; we choose to analyze the data with the **es.mean** weights.

2.4 Estimating treatment effects

As in McCaffrey et al. (2013) we consider estimating treatment effects on suf12, the substance frequency scale which measures frequency of substance use during the past 90 days prior to the 12-month follow-up visits for individuals in the study. The propensity score adjusted test can be computed using **regress** along with Stata's built-in weighting features (**svyset** followed by the **svy:** prefix). We start with an analysis using the weights derived from the GBM selected to minimize the mean standardized bias ("**es.mean**" stopping rule). We use

svyset to declare that our weighting variable is **esmeanate**. The propensity score adjusted results are then estimated using the **svy** prefix with **regress.**

Note that before proceeding, we need to first generate a variable "treat2" that is a labeled numeric variable instead of a string. Also note that we are changing the base level of treat2 to the third treatment group ("scy") using the **fvset** command.

```
. use "C:\Users\username\twang\output\aod_ate_wgts", clear
. encode treat, generate(treat2)
. svyset [pweight=esmeanate]
. fvset base 3 treat2
. svy: regress suf12 i.treat2
```

```
Survey: Linear regression

        Number of strata   =         1                  Number of obs      =       600
        Number of PSUs     =       600                  Population size    = 1427.9672
                                                        Design df          =       599
                                                        F(   2,    598)     =      1.00
                                                        Prob > F           =    0.3685
                                                        R-squared          =    0.0036

        ------------------------------------------------------------------------------
                             Linearized
               suf12 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
        -------------+----------------------------------------------------------------
              treat2 |
          community  |  -.0646436   .0999812    -0.65   0.518    -.2609999    .1317127
            metcbt5  |   .0839397   .1093243     0.77   0.443    -.1307658    .2986453
               _cons |  -.0344831    .073887    -0.47   0.641    -.1795921    .1106259
        ------------------------------------------------------------------------------
```

As noted above, we set the base level of treat2 to the third treatment group ("scy"). Consequently, the estimated effect for metcbt5 equals the weighted mean for the metcbt5 sample less the weighted mean for the scy sample, where both means are weighted to match the overall sample. Similarly, the effect for community equals the difference in the weighted means for the community and scy samples. The coefficients estimate the causal effects of Community vs. SCY and MET/CBT-5 vs. SCY, respectively, assuming there are no unobserved confounders.

Using this small subset of the data, we are unable to detect differences in the treatment group means. In the context of this application, the signs of the estimates correspond to higher substance use frequency for youths exposed to MET/CBT-5 relative to SCY and lower use for youth exposed to Community relative to SCY. The estimate statement is estimating the average treatment effect of Community relative to MET/CBT-5 for all the youths in the population. Youth exposed to Community have lower use but the estimate is not statistically significant.

After the user fits the outcome model in STATA, the user can get all the comparisons of interest using the **margins** command. In this way, it does not matter which base level we use in the model.

```
. margins treat2, pwcompare
```

```
Pairwise comparisons of adjusted predictions

        Number of strata   =         0                  Design df          =       599
                                                        Number of PSUs     =         0
                                                             Model VCE     : Linearized
                                              Expression   : Linear prediction, predict()

        ------------------------------------------------------------------------------
                                           Delta-method         Unadjusted
                                |   Contrast   Std. Err.     [95% Conf. Interval]
        ------------------------+-----------------------------------------------------
        treat2
          metcbt5 vs community  |   .1485833   .1050213     -.0576714     .354838
```

```
      scy vs community|   .0646436   .0999812     -.1317127    .2609999
      scy vs metcbt5  |  -.0839397   .1093243     -.2986453    .1307658
-------------------------------------------------------------------------
```

3 An ATT example

3.1 Estimating the weights

It is also possible to explore treatment effects on the treated (ATTs) using the **mnps** command. A key difference in the multiple treatment setting is that we must be clear as to which treatment condition "the treated" refers to. This is done through the **treatatt** argument. Here, we define the treatment group of interest to be the community group; thus, we are trying to draw inferences about the relative effectiveness of the three treatment groups for individuals like those who were enrolled in the community program.

```
. mnps treat illact crimjust subprob subdep white, ///
      ntrees(3000) intdepth(3) shrinkage(0.01) ///
      permtestiters(0) stopmethod(es.mean ks.mean) ///
      estimand(ATT) treatatt(community) ///
      rcmd(C:\Program Files\R\R-3.3.1\bin\Rscript.exe) ///
      objpath(C:\Users\username\twang\output)

. balance, summary unweighted weighted
. save C:\Users\username\twang\output\aod_att_wgts, replace
```

3.2 Graphical assessments of balance

The same basic graphical descriptions are available as in the ATE case, though it is important to note that these comparisons all assess balance relative to the "treatment" group rather than by comparing balance for all possible pairwise treatment group comparisons as is done with ATE. Specifying the **plotname** argument will generate the full set of default plots. Alternatively, the **mnplot** can be used to create specific plots, as it was for ATE case. The following code produces the graphics shown in Figures 9 and 10.

```
. mnplot, plotname(mnps_example_plot_9-10.pdf) ///
      multipage plots(1)
```

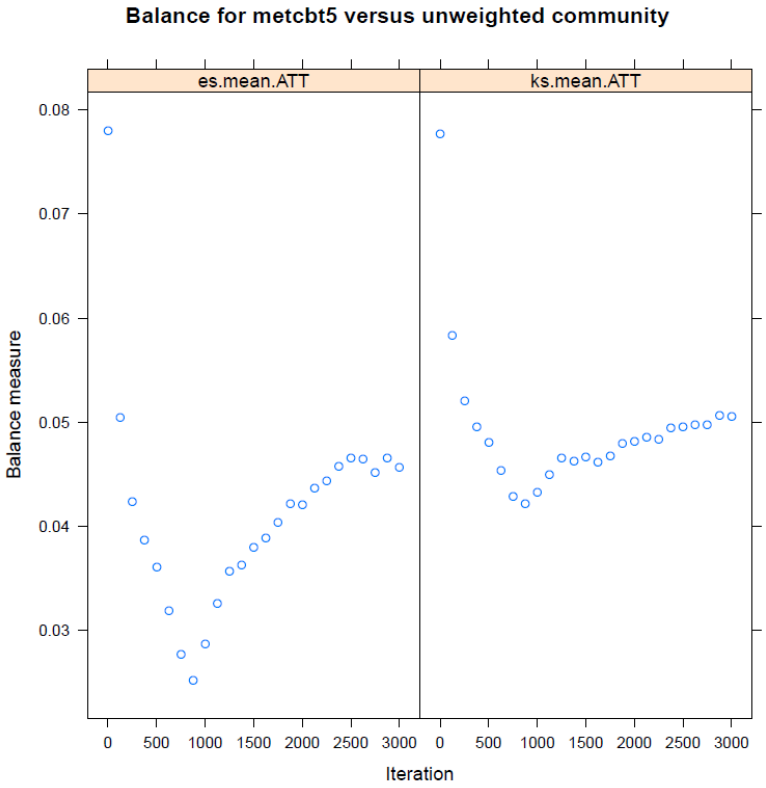

Figure 9: Example of an optimization plot for both stopping rules (**`es.mean`** and **`ks.mean`**) for estimating the propensity scores for comparing the MET/CBT-5 condition to the Community condition to generate ATT weights for the AOD dataset for a target population of those who received community treatment.

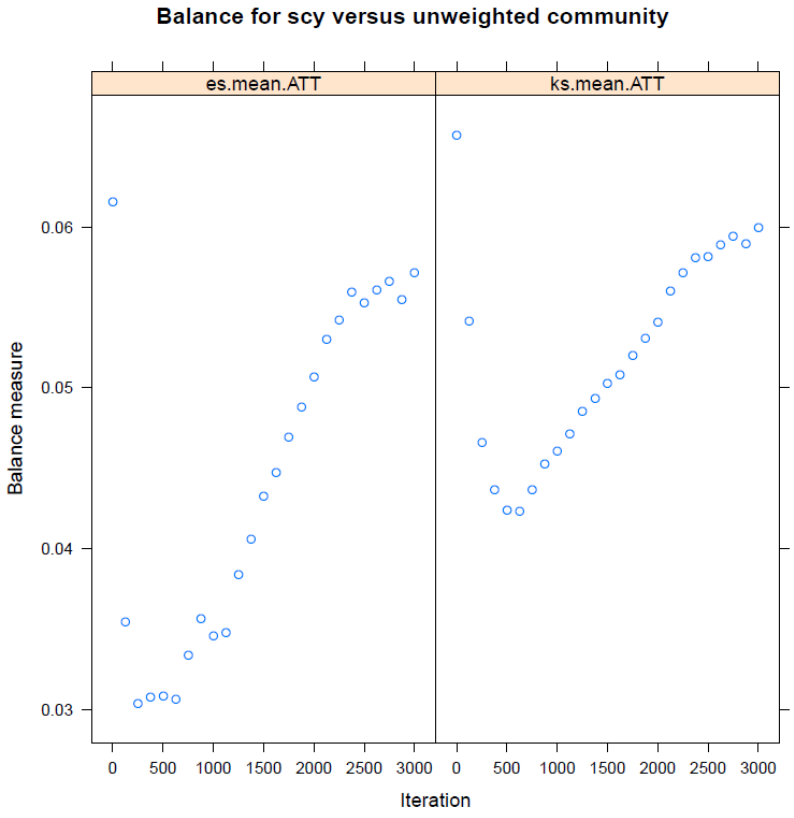

Figure 10: Example of an optimization plot for both stopping rules (**es.mean** and **ks.mean**) for estimating the propensity scores for comparing the SCY condition to the Community condition to generate ATT weights for the AOD dataset for a target population of those who received community treatment.

When the estimand is "ATT," there is one propensity score model fit for comparing each of the other treatments to the treatment specified by the **treatatt** argument. In this case, the target treatment is "community" so there is one model for comparing "metcbt5" to "community" and another for comparing "scy5" to "community". Consequently there is one optimization plot for the GBM model to compare "metcbt5" to "community" and another for comparing "scy" to "community". Similarly, we can look at the balance for each of the pairwise comparisons (here, SCY versus Community and MET/CBT5 versus Community) using the effect size plots (setting the **plots** argument to "3" or "es"). The following code produces Figure 11.

```
. mnplot, plotname(mnps_example_plot_11.pdf) ///
      multipage plots(3)
```

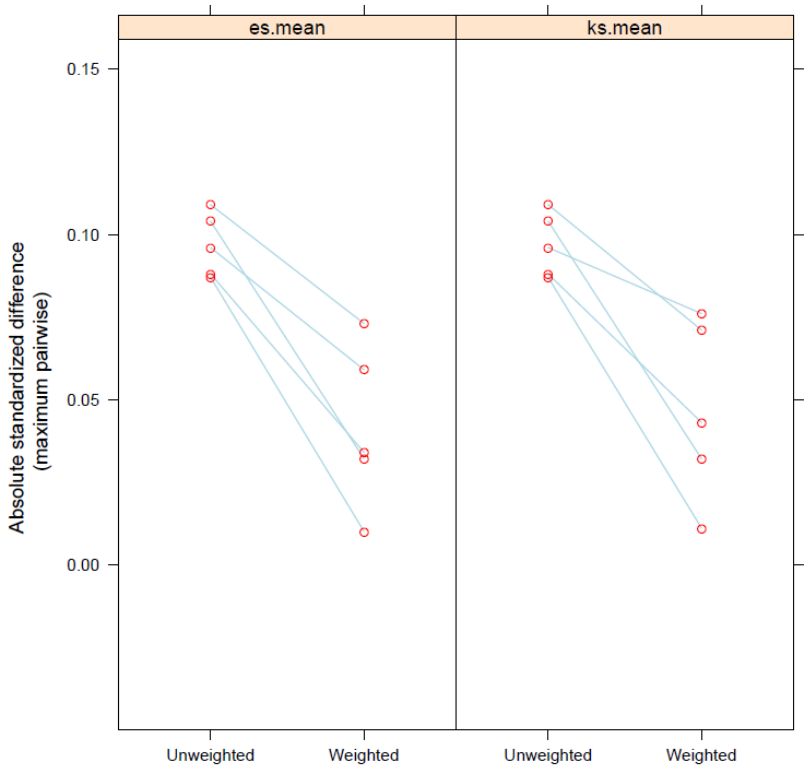

Figure 11: Example of an effect size plot for both stopping rules (**es.mean** and **ks.mean**) for comparing the MET/CBT-5 or the SCY condition to the Community condition to generate ATT weights for the AOD dataset for a target population of those who received community treatment. Plot of the maximum effect size across both comparisons for each covariate.

By default a call to **mnplot** with the **plots** argument equal to "3", as with the previous code, generates a plot of the maximum standardized effect across both comparisons (SCY versus Community and MET/CBT5 versus Community) for each covariate. This is useful for determining if balance is satisfactory or if there are problems, but it is not as useful for assessing the implications of balance problems if any exist. To probe the balance in more detail, as with ATE, separate plots for each pairwise comparison can be created by specifying the **pairwisemax** option for the **plotname** command. We also use the **multipage** option so that each plot will be on a separate page. The following code produces Figures 12 and 13.

```
. mnplot, plotname(mnps_example_plot_12-13.pdf) ///
      multipage nopairwisemax plots(3)
```

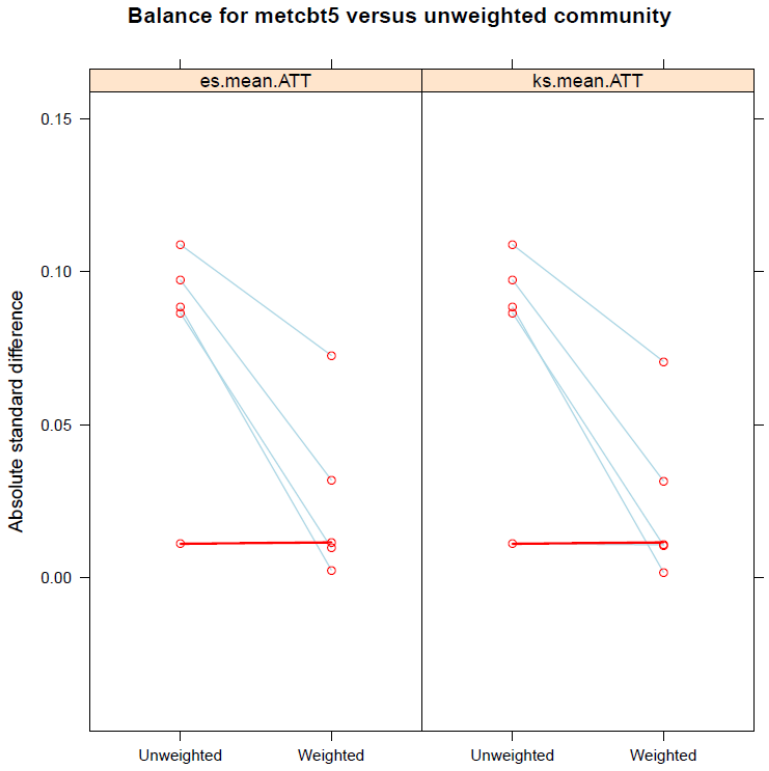

Figure 12: Example of an effect size plot for both stopping rules (**`es.mean`** and **`ks.mean`**) for comparing the MET/CBT-5 condition to the Community condition to generate ATT weights for the AOD dataset for a target population of those who received community treatment.

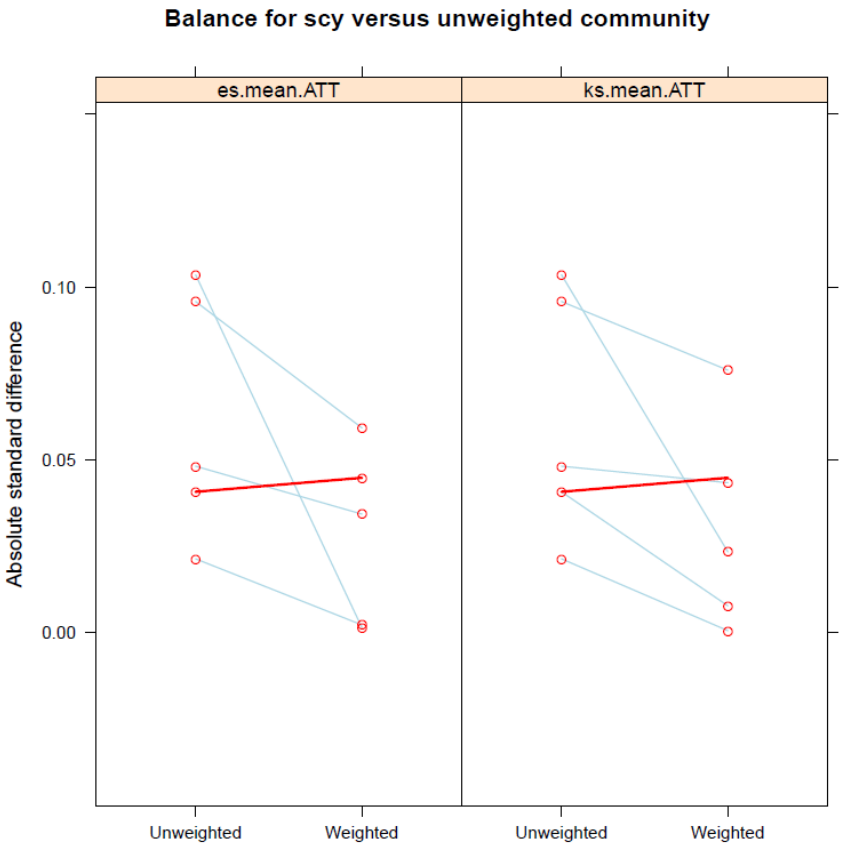

Figure 13: Example of an effect size plot for both stopping rules (**es.mean** and **ks.mean**) for comparing the SCY condition to the Community condition to generate ATT weights for the AOD dataset for a target population of those who received community treatment.

The p-value plots can also be useful as part of the assessment of balance. The minimum p-value across comparisons is the default and, like the effect size plot, the plots for each separate pairwise comparison can be created using the **nopairwisemax** option. The code below produces Figure 14. We include only the summary plot with minimum p-value for each covariate. In this example, with very small samples and well balanced groups, there are no statistically significant differences between the "metcbt5" or the "scy" samples and the "community" sample.

```
. mnplot, plotname(mnps_example_plot_14.pdf) ///
      plotformat(pdf) multipage plots(t)
```

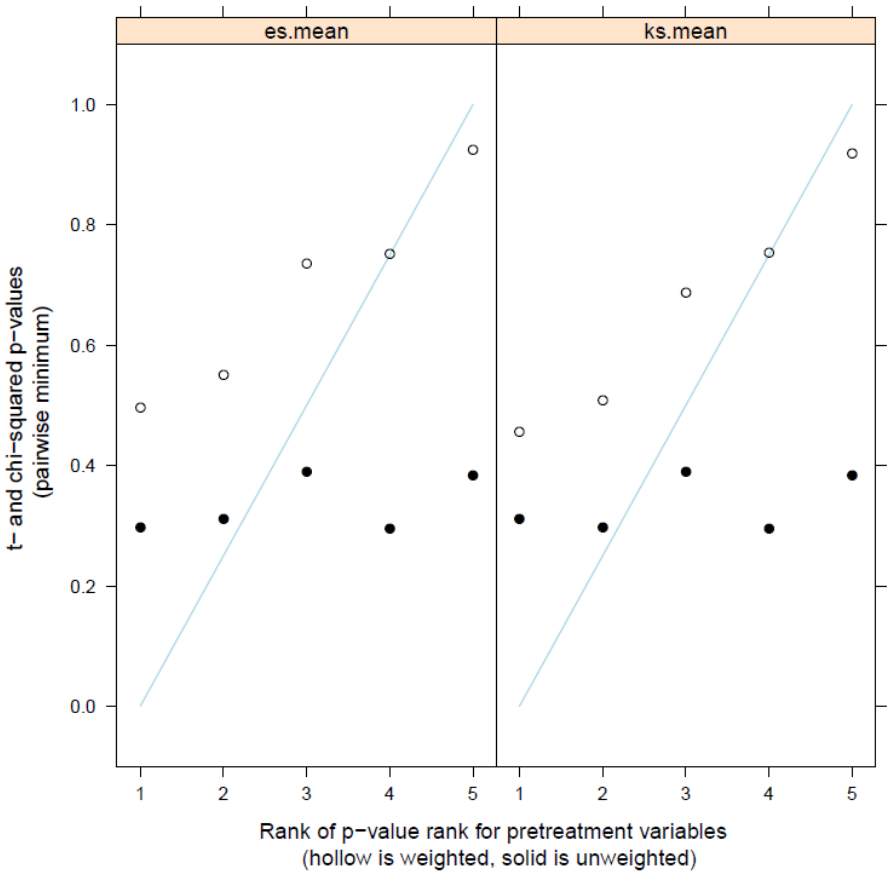

Figure 14: Example of a p-value plot for t-tests for both stopping rules (**`es.mean`** and **`ks.mean`**) for comparing the MET/CBT-5 and the SCY condition to the Community condition to generate ATT weights for the AOD dataset for a target population of those who received community treatment. Plot of the minimum p-value across both comparisons for each covariate.

3.3 Tabular assessments of balance

The **`mnps`** command prints default balance tables with the estimand equal to "ATT" like it did for "ATE". However, for ATT, it only reports pairwise comparisons that include the **`treatatt`** category. Only one summary table is created. It contains summary information for each comparison of a treatment to the target condition and for each stopping rule or unweighted. There is one line in the table for each combination of the alternative treatment. Each record includes:

- the stopping rule used in calculating the weights or "unw" for unweighted,
- ntreat and nctrl for the sample sizes of the target population and the comparison group, respectively,
- esstreat and essctrl, which equal the effective sample sizes for each group,
- maxes and meanes, the maximum and average absolute standardized effect sizes across the covariates,
- maxks, maxksp, and meanks, the maximum KS statistic across covariates, the p-value testing this (if requested), and the average KS statistics across covariates, and
- iter, which equals the number of iterations used in the GBM model chosen by the stopping rule of the record.

In this example, both the MET/CBT-5 and SCY condition samples are very similar to the Community condition sample prior to any weighting and consequently the balance is excellent after weighting.

Summary

	ntreat	nctrl	esstreat	essctrl	maxes	meanes	maxks	maxksp	meanks	iter
metcbt5										
_unw	200	200	200	200	.1088	.0785	.105	.	.078	.
_es_mean~T	200	200	200	166.1	.0726	.0257	.0513	.	.0418	831
_ks_mean~T	200	200	200	165.5	.0707	.0251	.051	.	.0417	854
scy										
_unw	200	200	200	200	.1036	.062	.09	.	.066	.
_es_mean~T	200	200	200	187.5	.0593	.0285	.0692	.	.0497	185
_ks_mean~T	200	200	200	170.9	.076	.0303	.0736	.	.0421	576

In Section 3.1 above, to estimate ATT weights, the following command was included in the code:

```
. balance, summary unweighted weighted
```

This command prints the balance table for the individual covariates. The table includes the same statistics as the balance table for the **ps** command (txmn, txsd, ctmn, ctsd, stdeffsz, stat, p, ks, kspval equal to the target treatment group mean and standard deviation, the comparison condition mean and standard deviation, the standardized effect size, the t-statistic testing the mean differences between groups and its associated p-value, the KS statistic and its p-value). It also includes the name of the covariate (var), the treatment group variable value for the comparison group (control) and the stop method (stop method).

Stata Output: Balance table for individual covariates for the ATT example

Unweighted

var	txmn	txsd	ctmn	ctsd	stdeffsz	stat	p	ks	kspval	control	stopme~d
illact	.097	1.045	.007	1.035	.087	.87	.385	.1	.27	metcbt5	unw
crimjust	-.065	1.05	.037	1.038	-.097	-.98	.328	.105	.221	metcbt5	unw
subprob	-.06	.965	.026	1.019	-.088	-.861	.39	.09	.394	metcbt5	unw
subdep	.046	1.079	.058	1.047	-.011	-.113	.91	.055	.924	metcbt5	unw
white	.16	.368	.2	.401	-.109	-1.041	.298	.04	.997	metcbt5	unw
illact	.097	1.045	.12	.963	-.021	-.223	.823	.06	.866	scy	unw
crimjust	-.065	1.05	-.174	1.028	.104	1.048	.295	.08	.545	scy	unw
subprob	-.06	.965	-.013	.972	-.048	-.481	.631	.09	.394	scy	unw
subdep	.046	1.079	-.058	.964	.096	1.012	.312	.085	.466	scy	unw
white	.16	.368	.175	.381	-.041	-.401	.688	.015	1	scy	unw

Weighted: esmean

var	txmn	txsd	ctmn	ctsd	stdeffsz	stat	p	ks	kspval	control	stopme~d
illact	.097	1.045	.087	1.024	.01	.094	.925	.041	.995	metcbt5	esmean
crimjust	-.065	1.05	-.032	.998	-.032	-.317	.752	.051	.957	metcbt5	esmean
subprob	-.06	.965	-.062	.989	.003	.025	.98	.039	.998	metcbt5	esmean
subdep	.046	1.079	.058	1.049	-.012	-.112	.911	.051	.959	metcbt5	esmean
white	.16	.368	.187	.391	-.073	-.68	.497	.027	1	metcbt5	esmean
illact	.097	1.045	.1	1.005	-.002	-.023	.982	.056	.902	scy	esmean
crimjust	-.065	1.05	-.064	.995	-.002	-.016	.988	.052	.941	scy	esmean
subprob	-.06	.965	-.027	.967	-.034	-.336	.737	.055	.904	scy	esmean
subdep	.046	1.079	-.018	.993	.059	.596	.551	.069	.707	scy	esmean
white	.16	.368	.176	.382	-.045	-.433	.665	.016	1	scy	esmean

Weighted: ksmean

var	txmn	txsd	ctmn	ctsd	stdeffsz	stat	p	ks	kspval	control	stopme~d
illact	.097	1.045	.086	1.023	.011	.102	.919	.042	.995	metcbt5	ksmean
crimjust	-.065	1.05	-.032	.997	-.032	-.313	.754	.051	.959	metcbt5	ksmean
subprob	-.06	.965	-.062	.988	.002	.018	.986	.039	.997	metcbt5	ksmean
subdep	.046	1.079	.057	1.048	-.011	-.104	.917	.05	.963	metcbt5	ksmean
white	.16	.368	.186	.39	-.071	-.662	.509	.026	1	metcbt5	ksmean
illact	.097	1.045	.098	1.036	-.001	-.006	.995	.05	.96	scy	ksmean
crimjust	-.065	1.05	-.041	.973	-.023	-.235	.814	.039	.998	scy	ksmean
subprob	-.06	.965	-.018	.979	-.043	-.402	.688	.045	.987	scy	ksmean
subdep	.046	1.079	-.036	.994	.076	.744	.457	.074	.664	scy	ksmean
white	.16	.368	.163	.37	-.008	-.077	.939	.003	1	scy	ksmean

As with the ATE condition, the **balance** command allows for printing balance tables that summarize across groups using the **collapseto** option to collapse by covariate or stopping method.

3.4 Estimating treatment effects

The effects of interest are comparisons of each of the treatments to Community care. As with ATE, the propensity score adjusted test can be computed using **regress** along with Stata's built-in weighting features (**svyset** followed by the **svy:** prefix). We start with an analysis using the ATT weights derived from the GBM selected to minimize the mean standardized bias ("**es.mean**" stopping rule). We use **svyset** to declare that our weighting variable is **esmeanatt**. The propensity score adjusted results are then estimated using the **svy** prefix with **regress.**

As before, we need to first generate a variable "treat3" that is a labeled numeric variable instead of a string. Also, note that we are changing the base level of treat3 to the first treatment group (community) using the **fvset** command.

```
. use "$objpath/aod_att_wgts", clear
. encode treat, generate(treat3)
. svyset [pweight=esmeanatt]
. fvset base 1 treat3
. svy: regress suf12 i.treat3
```

```
Survey: Linear regression

Number of strata   =         1                  Number of obs      =       600
Number of PSUs     =       600                  Population size    = 534.83326
                                                Design df          =       599
                                                F(   2,    598)    =      1.86
                                                Prob > F           =    0.1569
                                                R-squared          =    0.0066

------------------------------------------------------------------------------
                             Linearized
       suf12 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treat3  |
      metcbt5|   .2007108   .1040874     1.93   0.054    -.0037098    .4051315
      scy    |   .0807567    .099009     0.82   0.415    -.1136904    .2752038
     _cons   |  -.1050526   .0638328    -1.65   0.100    -.2304159    .0203107
------------------------------------------------------------------------------
```

Note in this case that the estimated treatment effect of community on those exposed to the community treatment is slightly stronger than in the ATE case (high numbers are bad for the outcome variable). Although not statistically significant, such differences are compatible with the notion that the youths who actually received the community treatment responded more favorably to it than the "average" youth would have (where the average is taken across the whole collection of youths enrolled in the study). The discussion in McCaffrey et al. (2013) may be useful for determining whether the ATE or ATT is of greater interest in a particular application.

4 Conclusion

Often, more than two treatments are available to study participants. If the study is not randomized, analysts may be interested in using a propensity score approach. Previously, few tools existed to aide the analysis of such data, perhaps tempting analysts to ignore all but two of the treatment conditions. We hope that this extension to the

twang package will encourage more appropriate analyses of observational data with more than two treatment conditions.

5 Acknowledgements

The random subset of data was supported by the Center for Substance Abuse Treatment (CSAT), Substance Abuse and Mental Health Services Administration (SAMHSA) contract # 270-07-0191 using data provided by the following grantees: Adolescent Treatment Model (Study: ATM: CSAT/SAMHSA contracts # 270-98-7047, # 270-97-7011, #2 77-00-6500, # 270-2003- 00006 and grantees: TI-11894, TI-11874, TI-11892), the Effective Adolescent Treatment (Study: EAT; CSAT/SAMHSA contract # 270-2003-00006 and grantees: TI-15413, TI-15433, TI-15447, TI-15461, TI-15467, TI-15475, TI-15478, TI-15479, TI-15481, TI-15483, TI-15486, TI-15511, TI15514, TI-15545, TI-15562, TI-15670, TI-15671, TI-15672, TI-15674, TI-15678, TI-15682, TI-15686, TI-15415, TI-15421, TI-15438, TI-15446, TI-15458, TI-15466, TI-15469, TI-15485, TI-15489, TI-15524, TI-15527, TI-15577, TI-15584, TI-15586, TI-15677), and the Strengthening Communities-Youth (Study: SCY; CSAT/SAMHSA contracts # 277-00-6500, # 270-2003- 00006 and grantees: TI-13305, TI-13308, TI-13313, TI-13322, TI-13323, TI-13344, TI-13345, TI-13354). The authors thank these grantees and their participants for agreeing to share their data to support the development of the **mnps** functionality.

6 References

[1] Burgette, L, McCaffrey, DF and Griffin, BA. Propensity score estimation with boosted regression. In W. Pan and H. Bai (Eds.) Propensity score analysis: Fundamentals and developments. New York: Guildford Publications, Inc. 2015:49-73.

[2] McCaffrey, D.F., B.A. Griffin, D. Almirall, M.E. Slaughter, R. Ramchand, and L.F. Burgette (2013). "A tutorial on propensity score estimation for multiple treatments using generalized boosted models." *Statistics in Medicine*, *32* (19), 3388–3414.

[3] Ridgeway, G., D. McCaffrey, B.A. Griffin, and L. Burgette (2014). "twang: Toolkit for weighting and analysis of non-equivalent groups." Available at http://cran.r-project.org/web/packages/twang/vignettes/twang.pdf.