

Lab Session: Matching

Cristian Huse

Introduction

This lab session is about **Matching**. The structure of the session is as follows:

- Preparation;
 - Load libraries, setwd, load data (.dta format)
- Example 13. Propensity Score Matching (PSM) Estimates
 - Here, we estimate a probit model and compare matched individuals
- PSM using the Nearest-Neighbour Method
 - 1-NN, 2-NN, and 1-NN with logit instead of probit model
- PSM Using Full Matching

Please note that there are different libraries implementing propensity score matching and we selectively cover only one of them.

Preparation

```
## Initialize ####
```

```
rm(list=ls())
```

```
#Load libraries
```

```
library(AER) #for ivreg
```

```
## Warning: package 'car' was built under R version 4.1.2
```

```
## Warning: package 'lmtest' was built under R version 4.1.2
```

```
library(fixest) #for feols etc
```

```
## Warning: package 'fixest' was built under R version 4.1.2
```

```
library(haven) #for read_dta
```

```
library(Matching)
```

```
library(MatchIt)
```

```
library(modelsummary) #for neat tables
```

```
## Warning: package 'modelsummary' was built under R version 4.1.2
```

```
library(optmatch)
library(panelr)
library(tidyverse)
```

```
## Warning: package 'readr' was built under R version 4.1.2
```

Propensity Score Matching (PSM)

Assume we have a group of treated and another of non-treated individuals for which we observe a number of characteristics X_i at baseline, i.e., before the programme is rolled out. One way to quantify treatment effects would be to estimate a model of the form

$$Y_i = \alpha + \delta P_i + \gamma X_i + \varepsilon_i$$

This strategy would provide a causal estimate for the effect of the program only in very particular cases and might suffer from sample selection, omitted variable bias, and other problems (recall the FTCI).

Matching parallels the standard regression approach by matching observations in treatment and control groups which are “similar” in terms of characteristics. To the extent that observed characteristics are the drivers of heterogeneity (as opposed to unobserved heterogeneity), matching is a good strategy, being often combined with other quasi-experimental methods such as DD. (Note: Matching per se doesn’t ensure causal estimates, but is a way to match to a, say, treated subject, its “closest” non-treated subjects.)

PSM produces a (propensity) score using a regression of programme participation on a set of observed (pre-policy) covariates. Then units in the treatment and comparison group with the closest propensity scores are matched, and differences in outcomes are calculated within each matched pair. The matching procedure is then repeated for all individuals in the treatment group, and averages in differences in outcomes within pairs are computed.

There exist different methods to calculate the “closest match” (see slides).

Example 13. Propensity Score Matching Estimates

This example illustrates how to implement PSM in the HISP data. First, estimate a probit (alternatively, logit) model where the dependent variable is an indicator of programme participation and the covariates are individual characteristics **pre-programme**. The output of the probit regression shows which variables are the strongest predictors of program participation.

Before performing any estimation, one needs to **reshape** the data. The intuition is as follows. In the **long format**, each household identifier (the id) appears in two rounds (the waves, 0 and 1), i.e., it is computed as if observed twice (pre- and post-treatment). What we want to match individuals on is on pre-programme characteristics, therefore the reshape operation. Please make sure to have a look at the data before and after reshaping it so you understand what you are doing.

```
#reshape the database
##panelr::panel_data
# Informs who are i (id) and t (wave)
##panelr::widen_panel
# Will make round "disappear"
```

```
# Note it may take some time for this reshape command to run

eval_wide.df <- panel_data(evaluation.df, id = household_identifider, wave = round) %>%
  widen_panel(separator = "_") %>%
  subset(select = c(-age_hh_1, -age_sp_1, -educ_hh_1, -educ_sp_1, -hospital_1)) %>%
  rename(age_hh = age_hh_0, age_sp = age_sp_0, educ_hh = educ_hh_0,
         educ_sp = educ_sp_0, hospital = hospital_0)
```

Recall that we are comparing health expenditures at follow-up (post-policy) between treated individuals (enrolled households) and a set of matched non-treated individuals (non-enrolled households) from both treatment and control villages.

The probit estimates highlight the significance of variables such as household size and dwellings with dirt floors in predicting participation in HISP. The propensity score is obtained by computing the predicted values from this first stage.

```
ex13_glm <- glm(enrolled ~ age_hh + age_sp + educ_hh + educ_sp + female_hh +
  indigenous + hhsize + dirtfloor + bathroom + land +
  hospital_distance,
  family = binomial(link = "probit"),
  data = eval_wide.df)

modelsummary(ex13_glm,
  vcov = ~ locality_identifider,
  stars = c("*" = .1, "**" = .05, "***" = .01),
  fmt = 3,
  gof_omit = "AIC|BIC|Log.Lik.|R2 Adj.|R2 Within|R2 Pseudo|F")
```

The following figure plots the propensity scores for the treatment and control groups. Once the propensity scores are calculated, the next step is to find for each individual in the treatment group, a comparable “match” with a similar propensity score in the comparison group, i.e., an individual with characteristics such that its likelihood of participating in the program are the same as the treatment individual for which a match is sought.

Therefore, matching generally occurs in the area of **common support** where the propensity scores for the treatment and comparison groups overlap. Therefore it is important to check to which extent the **common support assumption** holds empirically.

Once matching is performed, differences in outcomes within pairs are computed and the averages of these differences are obtained to provide estimates of treatment effects. A technical but important note is that matching will typically recover the ATT (Average Treatment Effect on the Treated) instead of the ATE. Intuitively, the ATE would be estimable if one were not to discard observations from the sample or if one were to target full sample balance (dropping observations would go against the idea of a random sample, on which the ATE is intuitively based). Therefore, one should not get surprised with the default option of the type **estimand** = “ATT” in Matching packages.

```
#drop observations omitted in probit estimation, then append fitted values
eval_wide.df <- subset(eval_wide.df,
  household_identifider !=
    subset(eval_wide.df,select=-hospital)
    [!complete.cases(subset(eval_wide.df,
      select = -hospital)),]$household_identifider) %>%
  mutate(pscore = ex13_glm$fitted.values)
```

	Model 1
(Intercept)	−0.497* (0.254)
age_hh	−0.013*** (0.002)
age_sp	−0.008*** (0.003)
educ_hh	−0.022** (0.010)
educ_sp	−0.016 (0.011)
female_hh	−0.020 (0.062)
indigenous	0.161 (0.121)
hhsiz	0.119*** (0.012)
dirtfloor	0.376*** (0.076)
bathroom	−0.125* (0.068)
land	−0.028*** (0.009)
hospital_distance	0.002 (0.002)
Num.Obs.	9913
Std.Errors	by: locality_identifier
* p < 0.1, ** p < 0.05, *** p < 0.01	

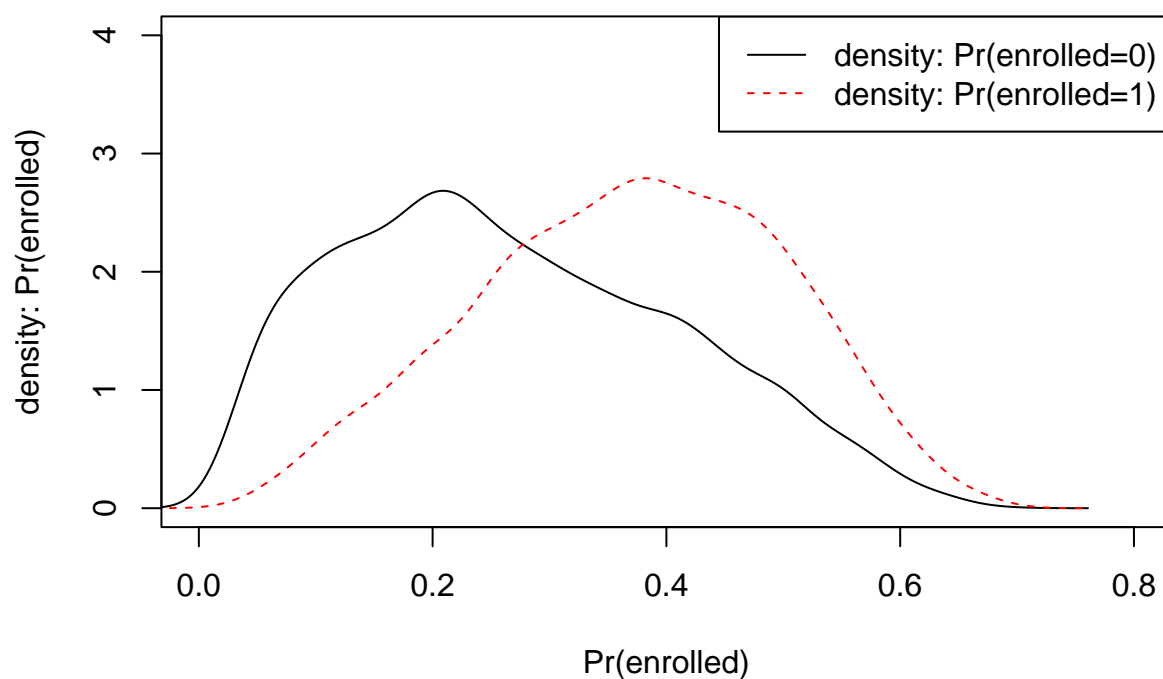
```

# Density estimates separately for subsamples of enrolled and non-enrolled
eval_wide_1.df <- subset(eval_wide.df, enrolled == 1)
eval_wide_0.df <- subset(eval_wide.df, enrolled == 0)
ex13_density1 <- density(eval_wide_1.df$pscore)
ex13_density0 <- density(eval_wide_0.df$pscore)

# Collect x (possibly a grid of values instead of the actual data) and y estimates
take1 <- ex13_density1$x
den1 <- ex13_density1$y
take0 <- ex13_density0$x
den0 <- ex13_density0$y

# Plot
plot(take0, den0, type = "l", xlim = c(0, 0.8), ylim = c(0, 4),
     xlab = "Pr(enrolled)", ylab = "density: Pr(enrolled)",
     lty = 1)
lines(take1, den1, col = "red", lty = 2)
legend("topright",
     legend = c("density: Pr(enrolled=0)", "density: Pr(enrolled=1)"),
     col = c("black", "red"), lty = c(1, 2), )

```



PSM Using the Package [MatchIt](#)

Checking for Balance

We start estimating a probit and checking balance. To do so, we specify the “first-stage” equation (`link = “probit”`) and call the library to check balance. The key statistics to focus on are **Std. Mean Diff.**, **Var. Ratio**, and **eCDF**. One should aim for values close to zero for the first and third, and values close to one for the middle one as measures of good balance.

```
# Notation
Y <- eval_wide.df$health_expenditures_1
Tr <- eval_wide.df$enrolled

# Note: method = NULL
m.out0 <- matchit(Tr ~ age_hh + age_sp + educ_hh + educ_sp + female_hh + indigenous +
                  hhsize + dirtfloor + bathroom + land + hospital_distance,
                  distance = "glm",
                  link = "probit",
                  method = NULL,
                  data = eval_wide.df)

summary(m.out0)
```

```
##
## Call:
## matchit(formula = Tr ~ age_hh + age_sp + educ_hh + educ_sp +
##         female_hh + indigenous + hhsize + dirtfloor + bathroom +
##         land + hospital_distance, data = eval_wide.df, method = NULL,
##         distance = "glm", link = "probit")
##
## Summary of Balance for All Data:
##               Means Treated Means Control Std. Mean Diff. Var. Ratio
## distance                0.3694         0.2684         0.7781    0.8190
## age_hh                   41.6566         48.1384        -0.4731    0.7798
## age_sp                   36.8364         41.6117        -0.4098    0.8042
## educ_hh                   2.9712          2.7745         0.0743    0.8960
## educ_sp                   2.7033          2.5810         0.0494    0.9257
## female_hh                 0.0732          0.1101        -0.1416         .
## indigenous                0.4291          0.3203         0.2199         .
## hhsize                    5.7699          4.9263         0.4226    0.8030
## dirtfloor                 0.7217          0.5533         0.3756         .
## bathroom                 0.5735          0.6340        -0.1223         .
## land                      1.6771          2.2511        -0.2172    0.6391
## hospital_distance        109.2043        103.6626         0.1324    0.9917
##               eCDF Mean eCDF Max
## distance                0.2003    0.3119
## age_hh                   0.0881    0.2074
## age_sp                   0.0648    0.2100
## educ_hh                   0.0218    0.0623
## educ_sp                   0.0141    0.0690
## female_hh                 0.0369    0.0369
## indigenous                0.1088    0.1088
## hhsize                    0.0652    0.1833
```

```
## dirtfloor          0.1683    0.1683
## bathroom          0.0605    0.0605
## land              0.0241    0.0882
## hospital_distance  0.0396    0.0830
##
##
## Sample Sizes:
##           Control Treated
## All           6949     2964
## Matched       6949     2964
## Unmatched      0         0
## Discarded      0         0
```

Matching Methods

There is a variety of matching methods, whose technical study is outside the scope of this session. Classical ones include **Nearest Neighbour Matching** (`method = "nearest"`) and **Caliper matching** (`caliper`), see link above for discussion and details. Intuitively, NN selects the K observations closest to a given element of the treatment group (K is chosen by the econometrician) whereas caliper matching selects all observations k such that they are within a radius r a given element of the treatment group (formally, k such that $\|x_k - x_0\| < r$ for each x_0 belonging to the treatment group, d chosen by the econometrician).

In what follows, we first perform the classical 1:1 NN matching method (one element of the control group is assigned to an element of the treatment group) and assess balance. We then specify a regression using the object generated by the **MatchIt** package as input in the next step – in practice, it carries with it the matched pairs. Note that there are several options available when it comes to matching and that both syntax and output make it conformable with **modelsummary**.

Looking at the diagnostics, extreme propensity scores are more scarce than intermediate ones, as expected. Comparing eQQ plots we get a sense that the matching procedure is doing a reasonably good job, at least for the three selected variables (`age_hh`, `hhsz`, `hospital_distance`) – observations of treated vs control units are mostly within the confidence bounds. You might want to examine QQ plots of other variables, especially those for which the diagnostics **Std. Mean Diff.**, **Var. Ratio** or **eCDF** raise concerns.

Finally, note that while the estimate doesn't exactly nail the original ones, at $-\$10.02$ their values are pretty close (regard the remark on ATE vs. ATT, which is the estimand used here).

```
m.out1 <- matchit(Tr ~ age_hh + age_sp + educ_hh + educ_sp +
                  female_hh + indigenous + hhsz + dirtfloor +
                  bathroom + land + hospital_distance,
                  data = eval_wide.df,
                  method = "nearest", distance = "glm",
                  link = "probit")

m.out1
```

```
## A matchit object
## - method: 1:1 nearest neighbor matching without replacement
## - distance: Propensity score
## - estimated with probit regression
## - number of obs.: 9913 (original), 5928 (matched)
## - target estimand: ATT
## - covariates: age_hh, age_sp, educ_hh, educ_sp, female_hh, indigenous, hhsz, dirtfloor, bathroom,
```

```
# Tools for assessing balance
summary(m.out1, un = FALSE)
```

```
##
## Call:
## matchit(formula = Tr ~ age_hh + age_sp + educ_hh + educ_sp +
##   female_hh + indigenous + hhsize + dirtfloor + bathroom +
##   land + hospital_distance, data = eval_wide.df, method = "nearest",
##   distance = "glm", link = "probit")
##
## Summary of Balance for Matched Data:
##
```

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio
distance	0.3694	0.3687	0.0054	1.0139
age_hh	41.6566	41.7868	-0.0095	1.0867
age_sp	36.8364	36.9916	-0.0133	1.1144
educ_hh	2.9712	2.9549	0.0061	0.9272
educ_sp	2.7033	2.7129	-0.0039	0.8918
female_hh	0.0732	0.0732	0.0000	.
indigenous	0.4291	0.4234	0.0116	.
hhsize	5.7699	5.7885	-0.0093	0.8692
dirtfloor	0.7217	0.7254	-0.0083	.
bathroom	0.5735	0.5732	0.0007	.
land	1.6771	1.7011	-0.0091	1.0159
hospital_distance	109.2043	108.5100	0.0166	0.9994

```
##
```

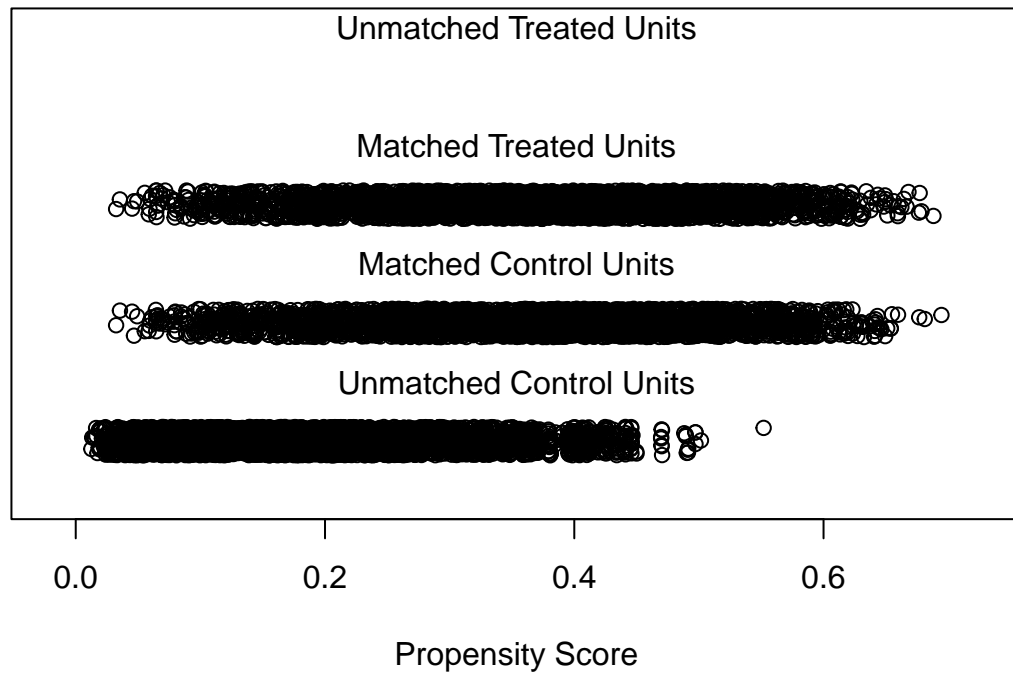
	eCDF Mean	eCDF Max	Std. Pair Dist.
distance	0.0007	0.0084	0.0056
age_hh	0.0115	0.0381	0.9087
age_sp	0.0104	0.0371	0.9332
educ_hh	0.0094	0.0240	1.0893
educ_sp	0.0073	0.0216	1.0664
female_hh	0.0000	0.0000	0.1289
indigenous	0.0057	0.0057	0.9032
hhsize	0.0106	0.0229	0.9946
dirtfloor	0.0037	0.0037	0.6858
bathroom	0.0003	0.0003	0.9817
land	0.0046	0.0277	0.8103
hospital_distance	0.0104	0.0486	1.0886

```
##
## Sample Sizes:
##
```

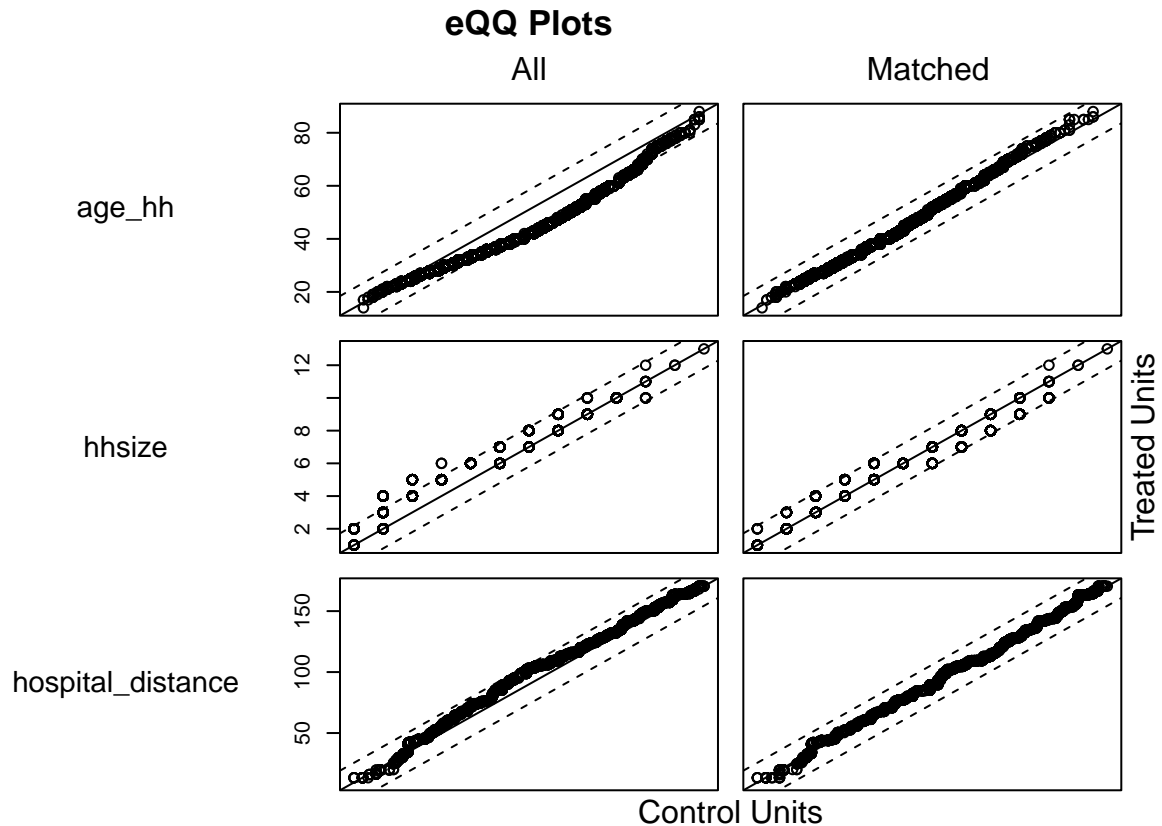
	Control	Treated
All	6949	2964
Matched	2964	2964
Unmatched	3985	0
Discarded	0	0

```
plot(m.out1, type = "jitter", interactive = FALSE)
```


Distribution of Propensity Scores



```
plot(m.out1, type = "qq", interactive = FALSE,  
     which.xs = c("age_hh", "hhsz", "hospital_distance"))
```



```
m.data1 <- match.data(m.out1)

# Specify the regression
fit1 <- lm(health_expenditures_1 ~ enrolled +
          age_hh + age_sp + educ_hh + educ_sp +
          female_hh + indigenous + hhsize + dirtfloor +
          bathroom + land + hospital_distance,
          data = m.data1,
          weights = weights)

#coeftest(fit1, vcov. = vcovCL, cluster = ~locality_identifier)

model1 <- list("(1) 1-NN" = fit1)

modelsummary(model1,
             vcov = ~ locality_identifier,
             stars = c("*" = .1, "**" = .05, "***" = .01),
             fmt = 3,
             gof_omit = "AIC|BIC|Log.Lik.|R2 Adj.|R2 Within|R2 Pseudo|F")
```

As an alternative, we will adjust the NN-method setting so now two (**ratio = 2**) instead of only one neighbour are selected for each treated observation. (You can check yourself that if you select **ratio = 3**, not every observation will get three neighbours). Note that you can experiment with a number of additional settings which we won't be cover, e.g., **distance**.

When looking at the diagnostics, you will notice that the QQ plots don't necessarily look better than the

	(1) 1-NN
(Intercept)	28.093*** (0.912)
enrolled	-10.019*** (0.304)
age_hh	0.045*** (0.017)
age_sp	0.002 (0.019)
educ_hh	0.005 (0.047)
educ_sp	-0.088* (0.052)
female_hh	0.454 (0.463)
indigenous	-1.844*** (0.347)
hhsz	-1.663*** (0.063)
dirtfloor	-1.900*** (0.270)
bathroom	0.406 (0.256)
land	0.031 (0.042)
hospital_distance	-0.005 (0.004)
Num.Obs.	5928
R2	0.417
Std.Errors	by: locality_identifier
* p < 0.1, ** p < 0.05, *** p < 0.01	

previous ones for the variables selected. Moreover, at $-\$10.0$, the treatment effect hardly changes when compared to the previous case.

```
m.out2 <- matchit(Tr ~ age_hh + age_sp + educ_hh + educ_sp +
  female_hh + indigenous + hhsize + dirtfloor +
  bathroom + land + hospital_distance,
  data = eval_wide.df,
  method = "nearest", distance = "glm",
  link = "probit",
  ratio = 2)

m.out2
```

```
## A matchit object
## - method: 2:1 nearest neighbor matching without replacement
## - distance: Propensity score
## - estimated with probit regression
## - number of obs.: 9913 (original), 8892 (matched)
## - target estimand: ATT
## - covariates: age_hh, age_sp, educ_hh, educ_sp, female_hh, indigenous, hhsize, dirtfloor, bathroom,
```

```
# Tools for assessing balance
summary(m.out2, un = FALSE)
```

```
##
## Call:
## matchit(formula = Tr ~ age_hh + age_sp + educ_hh + educ_sp +
##   female_hh + indigenous + hhsize + dirtfloor + bathroom +
##   land + hospital_distance, data = eval_wide.df, method = "nearest",
##   distance = "glm", link = "probit", ratio = 2)
##
## Summary of Balance for Matched Data:
##
```

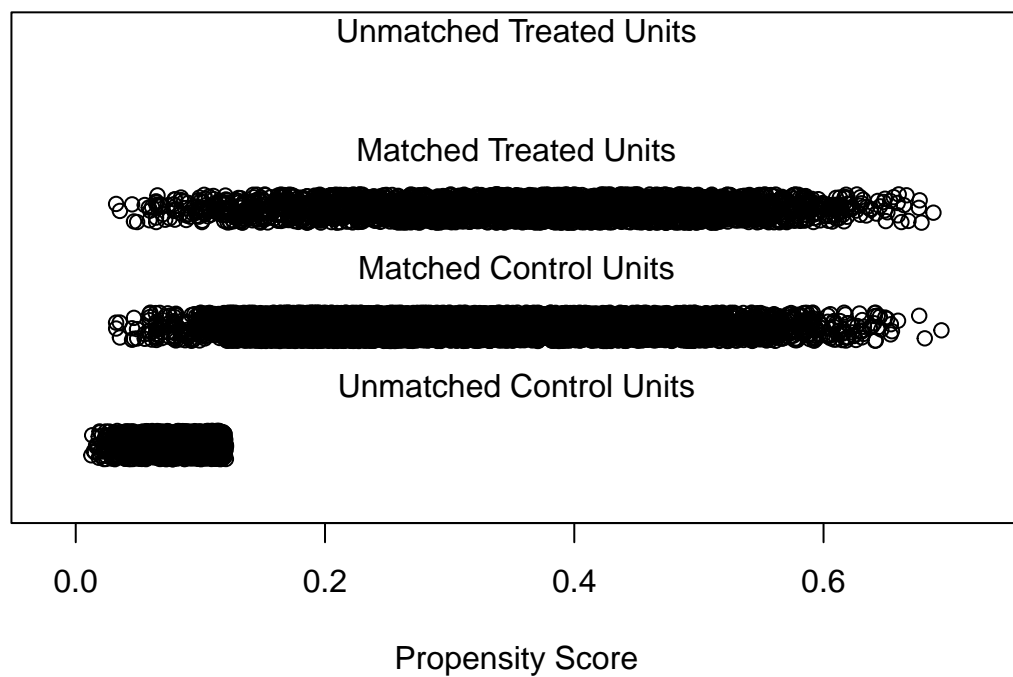
	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio
distance	0.3694	0.3017	0.5217	1.0243
age_hh	41.6566	45.1400	-0.2543	0.9297
age_sp	36.8364	39.3826	-0.2185	0.9887
educ_hh	2.9712	2.9301	0.0155	0.8724
educ_sp	2.7033	2.7248	-0.0087	0.8812
female_hh	0.0732	0.0970	-0.0913	.
indigenous	0.4291	0.3548	0.1503	.
hhsize	5.7699	5.2669	0.2520	0.8544
dirtfloor	0.7217	0.6127	0.2431	.
bathroom	0.5735	0.6115	-0.0767	.
land	1.6771	1.9389	-0.0991	0.8401
hospital_distance	109.2043	105.4741	0.0891	0.9912

```
##
## eCDF Mean eCDF Max Std. Pair Dist.
## distance      0.1350  0.2414      0.5218
## age_hh         0.0481  0.1412      1.0037
## age_sp         0.0364  0.1427      1.0122
## educ_hh        0.0131  0.0339      1.1124
## educ_sp        0.0107  0.0378      1.0822
## female_hh      0.0238  0.0238      0.5822
## indigenous     0.0744  0.0744      0.9015
```

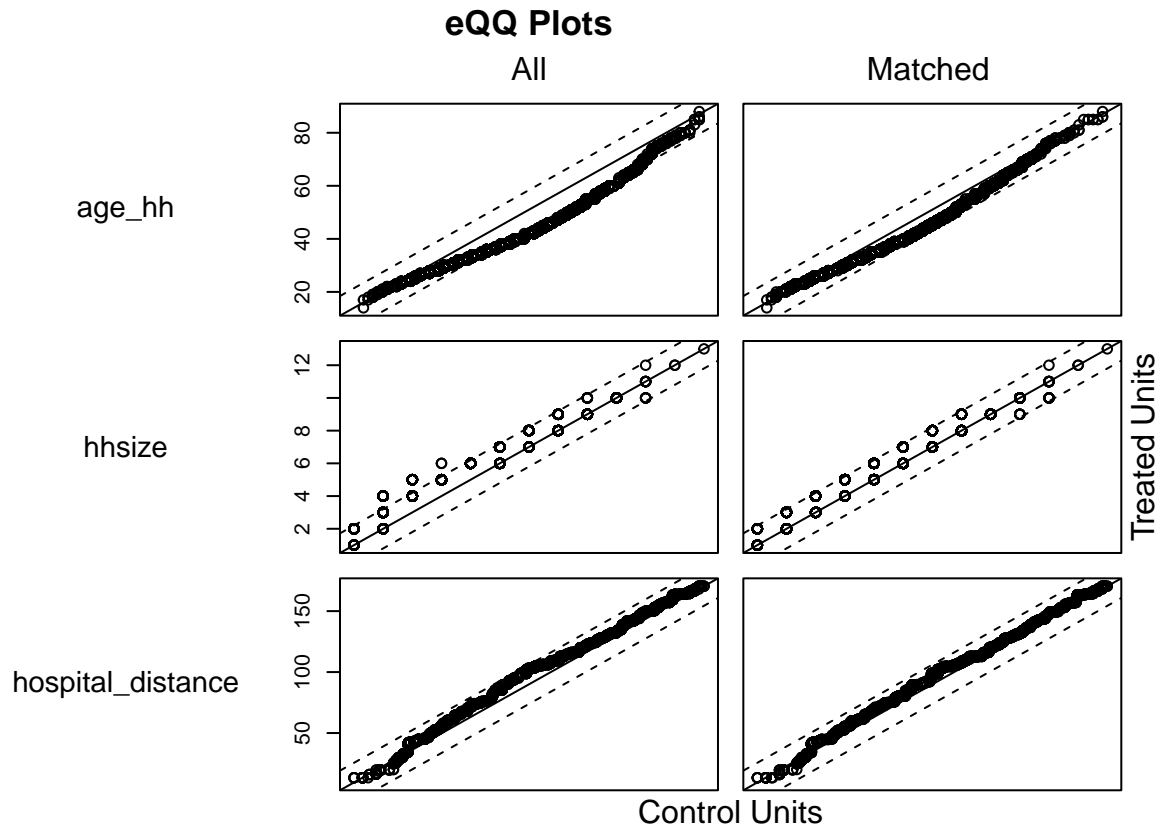
```
## hhsize          0.0394  0.1145      1.0484
## dirtfloor       0.1090  0.1090      0.8484
## bathroom       0.0380  0.0380      0.9568
## land           0.0114  0.0558      0.8729
## hospital_distance 0.0269  0.0609      1.1168
##
## Sample Sizes:
##           Control Treated
## All         6949    2964
## Matched     5928    2964
## Unmatched   1021      0
## Discarded      0      0
```

```
plot(m.out2, type = "jitter", interactive = FALSE)
```

Distribution of Propensity Scores



```
plot(m.out2, type = "qq", interactive = FALSE,
      which.xs = c("age_hh", "hhsize", "hospital_distance"))
```



```
m.data2 <- match.data(m.out2)

# Specify the regression
fit2 <- lm(health_expenditures_1 ~ enrolled +
          age_hh + age_sp + educ_hh + educ_sp +
          female_hh + indigenous + hhsize + dirtfloor +
          bathroom + land + hospital_distance,
          data = m.data2,
          weights = weights)

#coeftest(fit2, vcov. = vcovCL, cluster = ~locality_identifier)

model2 <- list("(2) 2-NN" = fit2)

modelsummary(model2,
             vcov = ~ locality_identifier,
             stars = c("*" = .1, "**" = .05, "***" = .01),
             fmt = 3,
             gof_omit = "AIC|BIC|Log.Lik.|R2 Adj.|R2 Within|R2 Pseudo|F")
```

Another moving part when performing matching is the regression model used in the first-stage regression. Therefore, we now investigate what happens when one replaces a probit with a logit model in Specification (1). These models differ when it comes to the underlying distributional assumption, with the probit model assuming a Normal distribution and the logit model assuming a Type-1 extreme value distribution. (We will discuss those models in detail in the future.)

	(2) 2-NN
(Intercept)	28.170*** (0.791)
enrolled	−9.997*** (0.285)
age_hh	0.045*** (0.014)
age_sp	0.010 (0.017)
educ_hh	−0.005 (0.042)
educ_sp	−0.023 (0.045)
female_hh	0.828** (0.395)
indigenous	−2.079*** (0.356)
hhsz	−1.799*** (0.054)
dirtfloor	−1.613*** (0.246)
bathroom	0.304 (0.244)
land	0.061 (0.038)
hospital_distance	−0.004 (0.004)
Num.Obs.	8892
R2	0.396
Std.Errors	by: locality__identifier
* p < 0.1, ** p < 0.05, *** p < 0.01	

Looking at the results, the estimates are very close when comparing those models (-10.02 vs. -10.18), which is not uncommon, at least in this specific time of model (binary choice).

```
m.out3 <- matchit(Tr ~ age_hh + age_sp + educ_hh + educ_sp +
  female_hh + indigenous + hhsiz + dirtfloor +
  bathroom + land + hospital_distance,
  data = eval_wide.df,
  method = "nearest", distance = "glm",
  link = "logit")
```

```
m.out3
```

```
## A matchit object
## - method: 1:1 nearest neighbor matching without replacement
## - distance: Propensity score
## - estimated with logistic regression
## - number of obs.: 9913 (original), 5928 (matched)
## - target estimand: ATT
## - covariates: age_hh, age_sp, educ_hh, educ_sp, female_hh, indigenous, hhsiz, dirtfloor, bathroom,
```

```
# Tools for assessing balance
summary(m.out3, un = FALSE)
```

```
##
## Call:
## matchit(formula = Tr ~ age_hh + age_sp + educ_hh + educ_sp +
##   female_hh + indigenous + hhsiz + dirtfloor + bathroom +
##   land + hospital_distance, data = eval_wide.df, method = "nearest",
##   distance = "glm", link = "logit")
##
## Summary of Balance for Matched Data:
##
```

	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio
distance	0.3696	0.3688	0.0057	1.0151
age_hh	41.6566	41.8343	-0.0130	1.1041
age_sp	36.8364	36.9831	-0.0126	1.1502
educ_hh	2.9712	2.9490	0.0084	0.9680
educ_sp	2.7033	2.7468	-0.0176	0.9118
female_hh	0.0732	0.0749	-0.0065	.
indigenous	0.4291	0.4231	0.0123	.
hhsiz	5.7699	5.7557	0.0071	0.8712
dirtfloor	0.7217	0.7338	-0.0271	.
bathroom	0.5735	0.5735	0.0000	.
land	1.6771	1.6660	0.0042	1.0923
hospital_distance	109.2043	109.0837	0.0029	1.0222

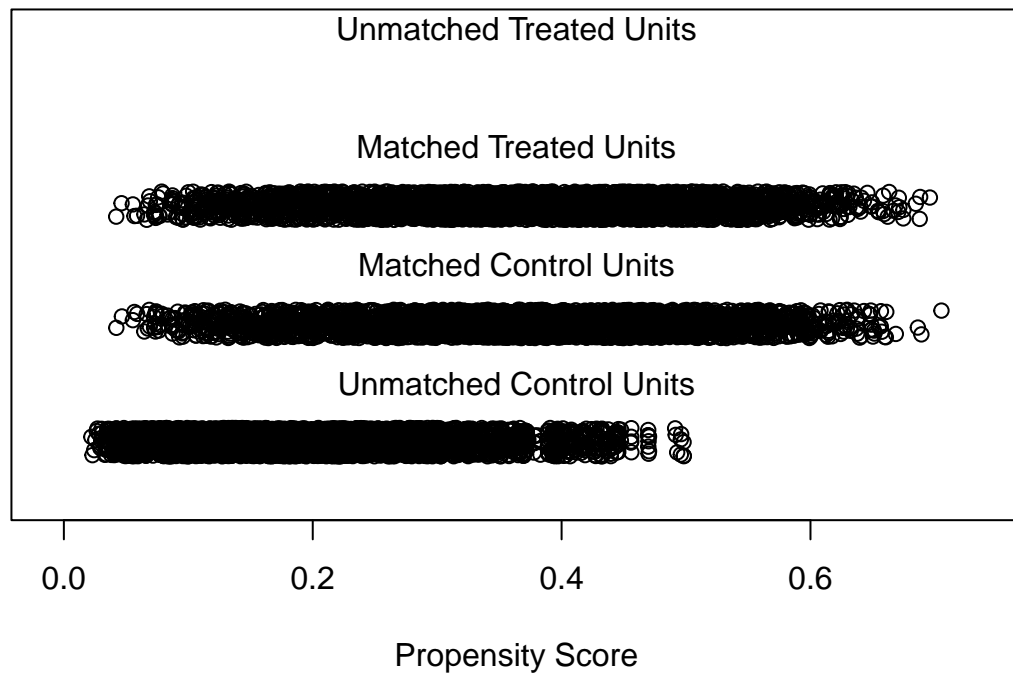
```
##
## eCDF Mean eCDF Max Std. Pair Dist.
## distance 0.0008 0.0098 0.0060
## age_hh 0.0129 0.0395 0.9051
## age_sp 0.0122 0.0398 0.9425
## educ_hh 0.0073 0.0179 1.0749
## educ_sp 0.0070 0.0192 1.0816
## female_hh 0.0017 0.0017 0.4961
## indigenous 0.0061 0.0061 0.8984
## hhsiz 0.0104 0.0273 1.0205
```



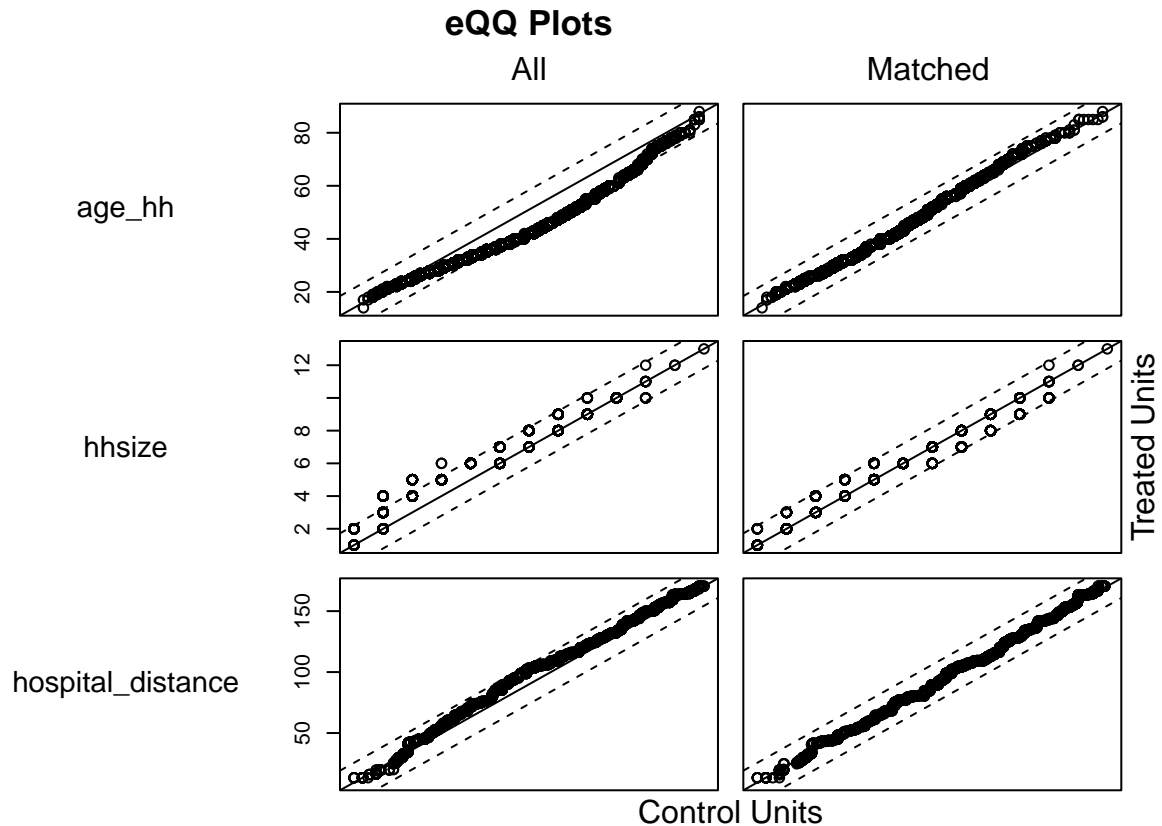
```
## dirtfloor          0.0121  0.0121      0.6865
## bathroom          0.0000  0.0000      0.4777
## land              0.0037  0.0199      0.8050
## hospital_distance  0.0101  0.0489      1.0946
##
## Sample Sizes:
##           Control Treated
## All           6949    2964
## Matched       2964    2964
## Unmatched     3985      0
## Discarded      0      0
```

```
plot(m.out3, type = "jitter", interactive = FALSE)
```

Distribution of Propensity Scores



```
plot(m.out3, type = "qq", interactive = FALSE,
     which.xs = c("age_hh", "hhsz", "hospital_distance"))
```



```
m.data3 <- match.data(m.out3)

# Specify the regression
fit3 <- lm(health_expenditures_1 ~ enrolled +
          age_hh + age_sp + educ_hh + educ_sp +
          female_hh + indigenous + hhsize + dirtfloor +
          bathroom + land + hospital_distance,
          data = m.data3,
          weights = weights)

#coeftest(fit3, vcov. = vcovCL, cluster = ~locality_identifier)

model3 <- list("(3) 1-NN w/ Logit" = fit3)

modelsummary(model3,
             vcov = ~ locality_identifier,
             stars = c("*" = .1, "**" = .05, "***" = .01),
             fmt = 3,
             gof_omit = "AIC|BIC|Log.Lik.|R2 Adj.|R2 Within|R2 Pseudo|F")
```

Finally, we will try a different matching method such as **full matching**, which matches every treated unit to at least one control and every control to at least one treated unit. The treatment effect in this case is lower yet close to previous estimates, at $-\$9.83$. (Please note that this computation can take longer than the previous one.)

All in all, we can say that the programme effects are largely robust to the methods used.

	(3) 1-NN w/ Logit
(Intercept)	28.235*** (0.927)
enrolled	-10.184*** (0.305)
age_hh	0.069*** (0.020)
age_sp	-0.014 (0.021)
educ_hh	0.034 (0.051)
educ_sp	-0.058 (0.047)
female_hh	0.803 (0.517)
indigenous	-1.785*** (0.352)
hhsz	-1.713*** (0.069)
dirtfloor	-1.932*** (0.289)
bathroom	0.327 (0.251)
land	0.070* (0.041)
hospital_distance	-0.008* (0.004)
Num.Obs.	5928
R2	0.416
Std.Errors	by: locality_identifier
* p < 0.1, ** p < 0.05, *** p < 0.01	

```
m.out4 <- matchit(Tr ~ age_hh + age_sp + educ_hh + educ_sp +
  female_hh + indigenous + hhsize + dirtfloor +
  bathroom + land + hospital_distance,
  data = eval_wide.df,
  method = "full", distance = "glm",
  link = "probit")
```

```
m.out4
```

```
## A matchit object
## - method: Optimal full matching
## - distance: Propensity score
## - estimated with probit regression
## - number of obs.: 9913 (original), 9913 (matched)
## - target estimand: ATT
## - covariates: age_hh, age_sp, educ_hh, educ_sp, female_hh, indigenous, hhsize, dirtfloor, bathroom,
```

Tools for assessing balance

```
summary(m.out4, un = FALSE)
```

```
##
## Call:
## matchit(formula = Tr ~ age_hh + age_sp + educ_hh + educ_sp +
##   female_hh + indigenous + hhsize + dirtfloor + bathroom +
##   land + hospital_distance, data = eval_wide.df, method = "full",
##   distance = "glm", link = "probit")
##
## Summary of Balance for Matched Data:
##
```

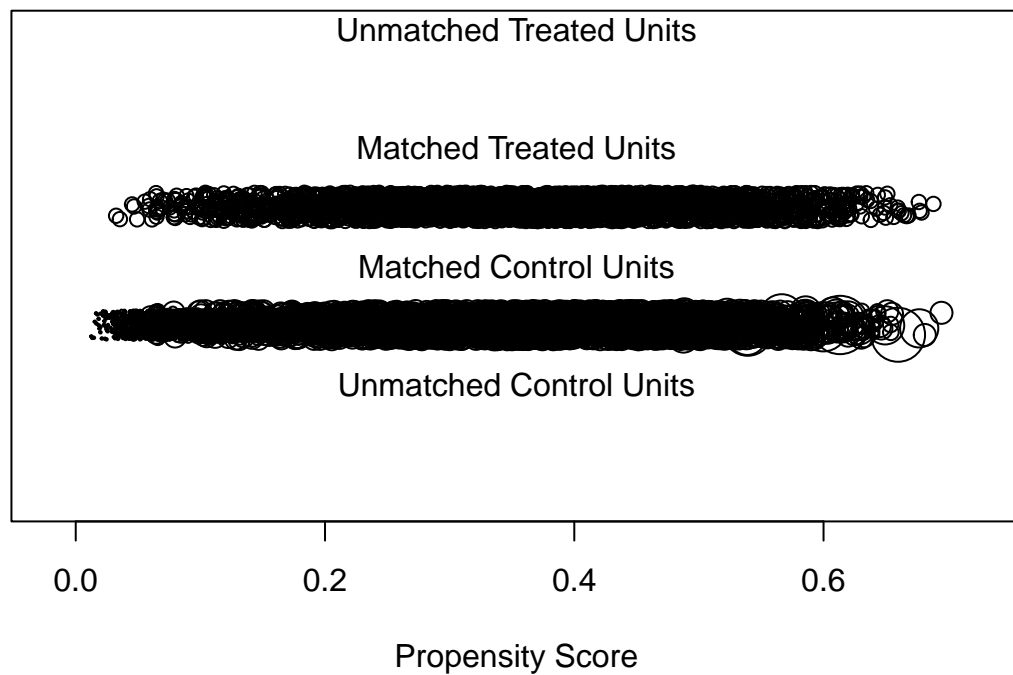
	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio
distance	0.3694	0.3694	0.0000	0.9998
age_hh	41.6566	42.4680	-0.0592	1.0960
age_sp	36.8364	37.5785	-0.0637	1.1043
educ_hh	2.9712	2.8861	0.0321	0.9502
educ_sp	2.7033	2.6201	0.0336	0.9623
female_hh	0.0732	0.0723	0.0033	.
indigenous	0.4291	0.4810	-0.1047	.
hhsize	5.7699	5.7217	0.0241	0.8945
dirtfloor	0.7217	0.7769	-0.1232	.
bathroom	0.5735	0.6133	-0.0805	.
land	1.6771	1.7863	-0.0413	0.9616
hospital_distance	109.2043	107.9099	0.0309	0.8823

```
##
## eCDF Mean eCDF Max Std. Pair Dist.
## distance 0.0005 0.0037 0.0048
## age_hh 0.0190 0.0689 0.9751
## age_sp 0.0171 0.0619 1.0519
## educ_hh 0.0109 0.0369 1.0375
## educ_sp 0.0083 0.0290 0.9947
## female_hh 0.0009 0.0009 0.7882
## indigenous 0.0518 0.0518 0.9179
## hhsize 0.0099 0.0261 0.9905
## dirtfloor 0.0552 0.0552 0.8987
## bathroom 0.0398 0.0398 0.9152
## land 0.0058 0.0405 1.1365
```

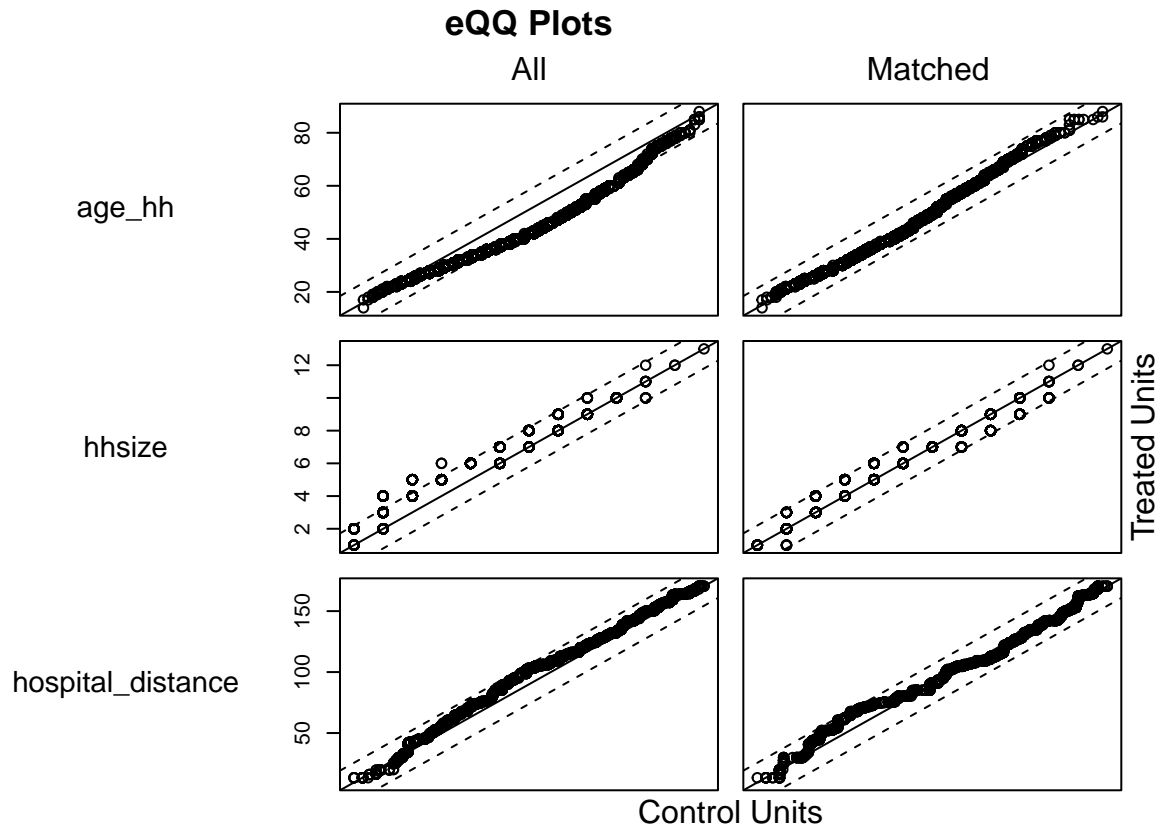
```
## hospital_distance    0.0233    0.0719          1.1691
##
## Sample Sizes:
##           Control Treated
## All           6949.     2964
## Matched (ESS) 3080.82    2964
## Matched       6949.     2964
## Unmatched           0.         0
## Discarded           0.         0
```

```
plot(m.out4, type = "jitter", interactive = FALSE)
```

Distribution of Propensity Scores



```
plot(m.out4, type = "qq", interactive = FALSE,
      which.xs = c("age_hh", "hhsizes", "hospital_distance"))
```



```
m.data4 <- match.data(m.out4)

# Specify the regression
fit4 <- lm(health_expenditures_1 ~ enrolled +
          age_hh + age_sp + educ_hh + educ_sp +
          female_hh + indigenous + hhsize + dirtfloor +
          bathroom + land + hospital_distance,
          data = m.data4,
          weights = weights)

#coeftest(fit3, vcov. = vcovCL, cluster = ~locality_identifier)

model4 <- list("(4) Full Matching" = fit4)

modelsummary(model4,
             vcov = ~ locality_identifier,
             stars = c("*" = .1, "**" = .05, "***" = .01),
             fmt = 3,
             gof_omit = "AIC|BIC|Log.Lik.|R2 Adj.|R2 Within|R2 Pseudo|F")
```

We end by comparing all specifications we estimated in this session. The main result is that the estimates are largely robust across models.

```
modelsummary(c(model1, model2, model3, model4),
             vcov = ~ locality_identifier,
```

(4) Full Matching	
(Intercept)	27.789*** (0.909)
enrolled	-9.827*** (0.298)
age_hh	0.060*** (0.016)
age_sp	-0.012 (0.018)
educ_hh	0.024 (0.049)
educ_sp	-0.012 (0.055)
female_hh	0.848 (0.602)
indigenous	-2.033*** (0.324)
hhsz	-1.625*** (0.058)
dirtfloor	-1.641*** (0.279)
bathroom	0.399* (0.238)
land	0.093** (0.043)
hospital_distance	-0.011*** (0.004)
Num.Obs.	9913
R2	0.381
Std.Errors	by: locality__identifier
* p < 0.1, ** p < 0.05, *** p < 0.01	

	(1) 1-NN	(2) 2-NN	(3) 1-NN w/ Logit	(4) Full Matching
(Intercept)	28.093*** (0.912)	28.170*** (0.791)	28.235*** (0.927)	27.789*** (0.909)
enrolled	-10.019*** (0.304)	-9.997*** (0.285)	-10.184*** (0.305)	-9.827*** (0.298)
age_hh	0.045*** (0.017)	0.045*** (0.014)	0.069*** (0.020)	0.060*** (0.016)
age_sp	0.002 (0.019)	0.010 (0.017)	-0.014 (0.021)	-0.012 (0.018)
educ_hh	0.005 (0.047)	-0.005 (0.042)	0.034 (0.051)	0.024 (0.049)
educ_sp	-0.088* (0.052)	-0.023 (0.045)	-0.058 (0.047)	-0.012 (0.055)
female_hh	0.454 (0.463)	0.828** (0.395)	0.803 (0.517)	0.848 (0.602)
indigenous	-1.844*** (0.347)	-2.079*** (0.356)	-1.785*** (0.352)	-2.033*** (0.324)
hhsizes	-1.663*** (0.063)	-1.799*** (0.054)	-1.713*** (0.069)	-1.625*** (0.058)
dirtfloor	-1.900*** (0.270)	-1.613*** (0.246)	-1.932*** (0.289)	-1.641*** (0.279)
bathroom	0.406 (0.256)	0.304 (0.244)	0.327 (0.251)	0.399* (0.238)
land	0.031 (0.042)	0.061 (0.038)	0.070* (0.041)	0.093** (0.043)
hospital_distance	-0.005 (0.004)	-0.004 (0.004)	-0.008* (0.004)	-0.011*** (0.004)
Num.Obs.	5928	8892	5928	9913
R2	0.417	0.396	0.416	0.381
Std.Errors	by: locality_identifier	by: locality_identifier	by: locality_identifier	by: locality_identifier

* p < 0.1, ** p < 0.05, *** p < 0.01

```
stars = c("*" = .1, "**" = .05, "***" = .01),
fmt = 3,
gof_omit = "AIC|BIC|Log.Lik.|R2 Adj.|R2 Within|R2 Pseudo|F")
```

References

Gertler, Paul J.; Martinez, Sebastian; Premand, Patrick; Rawlings, Laura B.; Vermeersch, Christel M. J. (2016). Impact Evaluation in Practice, Second Edition, Technical Companion (Version 1.0). Washington, DC: Inter-American Development Bank and World Bank.