

# 14. Double-Difference Method

---

The matching methods discussed in previous exercises are meant to reduce bias by choosing the treatment and comparison groups on the basis of observable characteristics. They are usually implemented after the program has been operating for some time and survey data have been collected. Another powerful form of measuring the impact of a program is by using panel data, collected from a baseline survey before the program was implemented and after the program has been operating for some time. These two surveys should be comparable in the questions and survey methods used and must be administered to both participants and nonparticipants. Using the panel data allows elimination of unobserved variable bias, provided that it does not change over time.<sup>1</sup>

This approach, the double-difference (DD, also commonly known as difference-in-difference) method has been popular in nonexperimental evaluations. The DD method estimates the difference in the outcome during the postintervention period between a treatment group and comparison group relative to the outcomes observed during a preintervention baseline survey.

## Simplest Implementation: Simple Comparison Using “ttest”

The simplest way of calculating the DD estimator is to manually take the difference in outcomes between treatment and control between the surveys. The panel data `hh_9198.dta` are used for this purpose. The following commands open the data file and create a new 1991-level outcome variable (per capita expenditure) to make it available in observations of both years. Then, only 1998 observations are kept, and a log of per capita expenditure variable is created; the difference between 1998 and 1991 per capita expenditures (log form) is created.

```
use ..\data\hh_9198;
gen exptot0=exptot if year==0;
egen exptot91=max(exptot0), by(nh);
keep if year==1;
gen lexptot91=ln(1+exptot91);
gen lexptot98=ln(1+exptot);
gen lexptot9891=lexptot98-lexptot91;
```

The following command (“ttest”) takes the difference variable of outcomes created earlier (“lexptot9891”) and compares it for microcredit participants and nonparticipants. In essence, it creates a second difference of “lexptot9891” for those with `dfmfd=1`

and those with  $dfmfd==0$ . This second difference gives the estimate of the impact of females' microcredit program participation on per capita expenditure.

```
ttest lexptot9891, by(dfmfd);
```

The result shows that microcredit program participation by females increases per capita consumption by 11.1 percent and that this impact is significant at a less than 1 percent level.<sup>2</sup>

Two-sample t-test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	391	.1473188	.0269923	.5337372	.0942502	.2003873
1	435	.2586952	.024194	.5046057	.2111432	.3062472
combined	826	.2059734	.018137	.5212616	.1703733	.2415735
<b>diff</b>		<b>-.1113764</b>	<b>.03614</b>		<b>-.1823136</b>	<b>-.0404392</b>

Degrees of freedom: 824

Ho: mean(0) - mean(1) = diff = 0

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
t = -3.0818	t = -3.0818	t = -3.0818
P < t = 0.0011	P >  t  = 0.0021	P > t = 0.998

## Regression Implementation

Instead of manually taking the difference of the outcomes, DD can be implemented using a regression. On the basis of the discussion in Ravallion (2008), the DD estimate can be calculated from the regression

$$Y_{it} = a + DD.T_i t + \beta T_i + \delta t_i + \epsilon_{it},$$

where  $T$  is the treatment variable,  $t$  is the time dummy, and the coefficient of the interaction of  $T$  and  $t$  (DD) gives the estimate of the impact of treatment on outcome  $Y$ .

The following commands open the panel data file, create the log of outcome variable, and create a 1998-level participation variable available to both years—that is, those who participate in microcredit programs in 1998 are the assumed treatment group.

```
use hh_9198, clear;
gen lexptot=ln(1+exptot);
gen dfmfd1=dfmfd==1 & year==1;
egen dfmfd98=max(dfmfd1), by(nh);
```

The next command creates the interaction variable of treatment and time dummy (year in this case, which is 0 for 1991 and 1 for 1998).

```
gen dfmfdyr=dfmfd98*year;
```

The next command runs the actual regression that implements the DD method:

```
reg lexptot year dfmfd98 dfmfdyr;
```

The results show the same impact of female participation in microfinance programs on households' annual total per capita expenditures as obtained in the earlier exercise:

Source	SS	df	MS	Number of obs =	1652
Model	20.2263902	3	6.74213005	F( 3, 1648) =	32.18
Residual	345.321048	1648	.209539471	Prob > F =	0.0000
				R-squared =	0.0553
				Adj R-squared =	0.0536
Total	365.547438	1651	.221409714	Root MSE =	.45775

lexptot	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
year	.1473188	.0327386	4.50	0.000	.0831052	.2115323
dfmfd98	-.1145671	.0318999	-3.59	0.000	-.1771358	-.0519984
<b>dfmfdyr</b>	<b>.1113764</b>	<b>.0451133</b>	<b>2.47</b>	<b>0.014</b>	<b>.0228909</b>	<b>.1998619</b>
_cons	8.310481	.0231497	358.99	0.000	8.265075	8.355887

A basic assumption behind the simple implementation of DD is that other covariates do not change across the years. But if those variables do vary, they should be controlled for in the regression to get the net effect of program participation on the outcome. So the regression model is extended by including other covariates that may affect the outcomes of interest:

```
reg lexptot year dfmfd98 dfmfdyr sexhead agehead educhead  
inland vaccaccess pcirr rice wheat milk oil egg [pw=weight];
```

By holding other factors constant, one sees that the impact of the microfinance programs has changed from significant to insignificant ( $t = 0.97$ ).

Regression with robust standard errors	Number of obs =	1652
	F( 14, 1637) =	24.90
	Prob > F =	0.0000
	R-squared =	0.2826
	Root MSE =	.42765

lexptot	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
year	.2768099	.0679939	4.07	0.000	.1434456	.4101741
dfmfd98	.0012122	.0326585	0.04	0.970	-.0628446	.0652691
<b>dfmfdyr</b>	<b>.0514655</b>	<b>.0530814</b>	<b>0.97</b>	<b>0.332</b>	<b>-.0526491</b>	<b>.1555802</b>

sexhead		-.0455035	.053903	-0.84	0.399	-.1512296	.0602227
agehead		.0017445	.0011041	1.58	0.114	-.0004212	.0039102
educhead		.0385333	.0049841	7.73	0.000	.0287575	.0483092
lnland		.226467	.0309236	7.32	0.000	.165813	.2871209
vaccess		-.011292	.0498495	-0.23	0.821	-.1090674	.0864835
pcirr		.0628715	.0453625	1.39	0.166	-.0261031	.1518461
rice		-.0023961	.0109958	-0.22	0.828	-.0239634	.0191712
wheat		.0071376	.0120905	0.59	0.555	-.0165769	.0308521
milk		.0158481	.005106	3.10	0.002	.0058332	.025863
oil		.0011434	.0031013	0.37	0.712	-.0049395	.0072263
egg		.1458875	.0475718	3.07	0.002	.0525794	.2391956
_cons		7.399387	.2715525	27.25	0.000	6.86676	7.932014

## Checking Robustness of DD with Fixed-Effects Regression

Another way to measure the DD estimate is to use a fixed-effects regression instead of ordinary least squares (OLS). Fixed-effects regression controls for household’s unobserved and time-invariant characteristics that may influence the outcome variable. The Stata “xtreg” command is used to run fixed-effects regression. In particular, with the “fe” option, it fits fixed-effect models.

Following is the demonstration of fixed-effects regression using the simple model:

```
xtreg lexptot year dfmfd98 dfmfdyr, fe i(nh)
```

The results showed again a significant positive impact of female participation:

```
Fixed-effects (within) regression      Number of obs      =   1652
Group variable (i): nh                Number of groups   =    826

R-sq:  within  =  0.1450              Obs per group:  min =     2
      between  =  0.0061                    avg  =    2.0
      overall   =  0.0415                    max  =     2

                                         F(2,824)          =   9.90
corr(u_i, Xb) = -0.0379                Prob > F           =  0.0000
```

lexptot		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
year		.1473188	.0262266	5.62	0.000	.0958399 .1987976
dfmfd98		(dropped)				
<b>dfmfdyr</b>		<b>.1113764</b>	<b>.03614</b>	<b>3.08</b>	<b>0.002</b>	<b>.0404392 .1823136</b>
_cons		8.250146	.0127593	646.60	0.000	8.225101 8.27519
sigma_u		.38132289				
sigma_e		.36670395				
rho		.51953588	(fraction of variance due to u_i)			

```
F test that all u_i=0:  F(825, 824) = 2.11  Prob > F = 0.0000
```

By including other covariates in the regression, the fixed-effects model can be extended in the following way:

```
xtreg lexptot year dfmfd98 dfmfdyr sexhead agehead educhead
lnland vaccess pcirr rice wheat milk oil egg, fe i(nh);
```

Results show that, after controlling for the effects of time-invariant unobserved factors, female participation in microcredit has a 9.1 percent positive impact on household's per capita consumption, and the impact is very significant.

```
Fixed-effects (within) regression      Number of obs      =    1652
Group variable (i): nh                 Number of groups   =     826

R-sq: within= 0.1715                   Obs per group: min =      2
      between= 0.1914                   avg =                2.0
      overall  = 0.1737                 max =                2

                                      F(13,813)          =   12.95
corr(u_i, Xb) = 0.1222                 Prob > F           =   0.0000
```

lexptot	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
year	.2211178	.063087	3.50	0.000	.0972851	.3449504
dfmfd98	(dropped)					
<b>dfmfdyr</b>	<b>.0906308</b>	<b>.0367358</b>	<b>2.47</b>	<b>0.014</b>	<b>.0185226</b>	<b>.1627391</b>
sexhead	-.0577238	.0722968	-0.80	0.425	-.1996342	.0841866
agehead	-.0003766	.0016985	-0.22	0.825	-.0037106	.0029574
educhead	.0137419	.0082935	1.66	0.098	-.0025373	.030021
lnland	.1381659	.0619682	2.23	0.026	.0165293	.2598025
vaccess	-.0932955	.053396	-1.75	0.081	-.1981057	.0115147
pcirr	.0823594	.0642728	1.28	0.200	-.0438009	.2085196
rice	.0107911	.010209	1.06	0.291	-.0092481	.0308303
wheat	-.0227681	.0123379	-1.85	0.065	-.046986	.0014498
milk	-.0014743	.0064578	-0.23	0.819	-.0141503	.0112016
oil	.0038546	.0031366	1.23	0.219	-.0023022	.0100113
egg	.1439482	.047915	3.00	0.003	.0498965	.238
_cons	7.853111	.2482708	31.63	0.000	7.365784	8.340439
sigma_u	.34608097					
sigma_e	.3634207					
rho	.47557527 (fraction of variance due to u_i)					
-----						
F test that all u_i=0: F(825, 813) = 1.59 Prob > F = 0.0000						

### Applying the DD Method in Cross-Sectional Data

DD can be applied to cross-sectional data, too, not just panel data. The idea is very similar to the one used in panel data. Instead of a comparison between years, program and nonprogram villages are compared, and instead of a comparison between participants and nonparticipants, target and nontarget groups are compared.

Accordingly, the 1991 data hh\_91.dta are used. Create a dummy variable called “target” for those who are eligible to participate in microcredit programs (that is, those who have less than 50 decimals of land). Then, create a village program dummy (“progvill”) for those villages that are

```
use ..\data\hh_91,clear;
gen lexptot=ln(1+exptot);
gen lnland=ln(1+hhlanddb/100);
gen target=hhlanddb<50;
gen progvill=thanaid<25;
```

Then, generate a variable interacting the program village and target:

```
gen progtarget=progvill*target
```

Then, calculate the DD estimate by regressing log of total per capita expenditure against program village, target, and their interaction:

```
. reg lexptot progvill target progtarget
```

The results show that the impact of microcredit program placement on the target group is not significant ( $t = -0.61$ ).

Source	SS	df	MS	Number of obs = 826		
Model	10.9420259	3	3.64734195	F( 3, 822)	=	27.38
Residual	109.485295	822	.133193789	Prob > F	=	0.0000
				R-squared	=	0.0909
				Adj R-squared	=	0.0875
Total	120.427321	825	.14597251	Root MSE	=	.36496

  

lexptot	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
progvill	-.0646577	.0770632	-0.84	0.402	-.2159215	.086606
target	-.2996852	.0815261	-3.68	0.000	-.459709	-.1396614
<b>progtarget</b>	<b>.0529438</b>	<b>.0867976</b>	<b>0.61</b>	<b>0.542</b>	<b>-.1174272</b>	<b>.2233147</b>
_cons	8.485567	.0729914	116.25	0.000	8.342296	8.628839

The coefficient of the impact variable (“progtarget”), which is 0.053, does not give the actual impact of microcredit programs; it has to be adjusted by dividing by the proportion of target households in program villages. The following command can be used to find the proportion:

```
sum target if progvill==1;
```

Of the households in program villages, 68.9 percent belong to the target group. Therefore, the regression coefficient of “progtarget” is divided by this value, giving 0.077, which is the true impact of microcredit programs on the target population, even though it is not significant.

Variable	Obs	Mean	Std. Dev.	Min	Max
target	700	.6885714	.4634087	0	1

As before, the regression model can be specified adjusting for covariates that affect the outcomes of interest:

```
reg lexptot progwill target progtarget sexhead agehead educhead lnland
vaccess pcirr rice wheat milk oil egg [pw=weight];
```

Holding other factors constant, one finds no change in the significance level of microcredit impacts on households' annual total per capita expenditures:

```
Regression with robust standard errors          Number of obs =      826
                                                F( 14,   811) =    11.03
                                                Prob > F      =    0.0000
                                                R-squared     =    0.3236
                                                Root MSE     =    .35757
```

lexptot	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
progwill	-.001756	.0793878	-0.02	0.982	-.1575857 .1540738
target	.0214491	.0911074	0.24	0.814	-.1573849 .2002832
<b>progtarget</b>	<b>-.0102772</b>	<b>.0895501</b>	<b>-0.11</b>	<b>0.909</b>	<b>-.1860545 .1655</b>
sexhead	-.019398	.0743026	-0.26	0.794	-.1652462 .1264502
agehead	-.0001666	.0014126	-0.12	0.906	-.0029394 .0026062
educhead	.0263119	.0060213	4.37	0.000	.0144927 .0381311
lnland	.268622	.0513087	5.24	0.000	.1679084 .3693356
vaccess	-.0098224	.0695396	-0.14	0.888	-.1463211 .1266764
pcirr	.0007576	.0571461	0.01	0.989	-.1114141 .1129294
rice	-.0082217	.0160899	-0.51	0.610	-.0398044 .023361
wheat	.0206119	.0146325	1.41	0.159	-.0081101 .049334
milk	.0227563	.0059707	3.81	0.000	.0110365 .0344761
oil	-.0067235	.0039718	-1.69	0.091	-.0145196 .0010727
egg	.1182376	.0569364	2.08	0.038	.0064775 .2299978
_cons	7.827818	.3696557	21.18	0.000	7.102223 8.553413

Again, fixed-effects regression can be used instead of OLS to check the robustness of the results. However, with cross-sectional data, household-level fixed effects cannot be run, because each household appears only once in the data. Therefore, a village-level fixed-effects regression is run:

```
xtreg lexptot progwill target progtarget, fe i(vill)
```

This time there is a negative (insignificant) impact of microcredit programs on household per capita expenditure:

```
Fixed-effects (within) regression          Number of obs      =    826
Group variable (i): vill                  Number of groups   =     87

R-sq: within  = 0.1088                    Obs per group: min =     4
      between  = 0.0240                    avg               =    9.5
      overall  = 0.0901                    max               =    15

                                                F(2,737)          =    44.98
corr(u_i, Xb) = -0.0350                    Prob > F          =    0.0000
```

```

-----
      lexptot |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
      progvill | (dropped)
      target  |  -.2531591   .0801025    -3.16  0.002   - .4104155   -.0959028
progtarget | -.0134339 .0854701 -0.16 0.875 -.1812278 .15436
      _cons   |   8.436668   .0232409   363.01  0.000    8.391041    8.482294
-----+-----
      sigma_u | .16994272
      sigma_e | .3419746
      rho     | .1980463     (fraction of variance due to u_i)
-----
F test that all u_i=0:      F(86, 737) =      2.32          Prob > F = 0.0000

```

The same fixed-effects regression is run after including other covariates:

```

xtreg lexptot progvill target progtarget sexhead agehead educhead
      ead lnland, fe i(vill)

```

Again, no change is seen in the significance level:

```

Fixed-effects (within) regression           Number of obs   =      826
Group variable (i): vill                   Number of groups =      87

R-sq:  within = 0.2258                      Obs per group:  min =      4
      between = 0.0643                          avg   =      9.5
      overall  = 0.1887                          max   =     15

                                           F(6,733)       =     35.62
corr(u_i, Xb) = -0.0497                    Prob > F       =     0.0000

```

```

-----
      lexptot |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
      progvill | (dropped)
      target  |   .0326157   .0818661     0.40  0.690   - .1281043   .1933357
progtarget | -.0081697 .07999 -0.10 0.919 -.1652066 .1488671
      sexhead |  -.0051257   .0568657    -0.09  0.928   - .1167648   .1065134
      agehead |   .0001635   .0010231     0.16  0.873   - .0018451   .0021721
      educhead | .0229979    .0039722     5.79  0.000    .0151997   .0307962
      lnland  | .2732536    .0385588     7.09  0.000    .1975548   .3489523
      _cons   |   8.072129   .0806635   100.07  0.000    7.91377    8.230488
-----+-----
      sigma_u | .16666988
      sigma_e | .3196088
      rho     | .21380081     (fraction of variance due to u_i)
-----
F test that all u_i=0:      F(86, 733) =      2.55          Prob > F = 0.0000

```

## Taking into Account Initial Conditions

Even though DD implementation through regression (OLS or fixed effects) controls for household- and community-level covariates, the initial conditions during the baseline survey may have a separate influence on the subsequent changes in outcome or assignment to the treatment. Ignoring the separate effect of initial conditions therefore may bias the DD estimates.

Including the initial conditions in the regression is tricky, however. Because the baseline observations in the panel sample already contain initial characteristics, extra variables for initial conditions cannot be added directly. One way to add initial conditions is to take into account an alternate implementation of the fixed-effects regression. In this implementation, difference variables are created for all variables (outcome and covariates) between the years, and then these difference variables are used in regression instead of the original variables. In this modified data set, initial condition variables can be added as extra regressors without a colinearity problem.

The following commands create the difference variables from the panel data `hh_9198`:

```
sort nh year;
by nh: gen dlexptot=lexptot[2]-lexptot[1];
by nh: gen ddmfd98= dmfd98[2]- dmfd98[1];
by nh: gen dmmfd98= dmmfd98[2]- dmmfd98[1];
by nh: gen ddfmfd98= dfmfd98[2]- dfmfd98[1];
by nh: gen ddmfdyr= dmfdyr[2]- dmfdyr[1];
by nh: gen dmmfdyr= dmmfdyr[2]- dmmfdyr[1];
by nh: gen ddfmfdyr= dfmfdyr[2]- dfmfdyr[1];
by nh: gen dsexhead= sexhead[2]- sexhead[1];
by nh: gen dagehead= agehead[2]- agehead[1];
by nh: gen deduchehead= educhead[2]- educhead[1];
by nh: gen dlndland= lnland[2]- lnland[1];
by nh: gen dvaccess= vaccess[2]- vaccess[1];
by nh: gen dpcirr= pcirr[2]- pcirr[1];
by nh: gen drice= rice[2]- rice[1];
by nh: gen dwhtflr= whtflr[2]- whtflr[1];
by nh: gen dmilk= milk[2]- milk[1];
by nh: gen dmustoil= mustoil[2]- mustoil[1];
by nh: gen dhenegg= henegg[2]- henegg[1];
```

Stata creates these difference variables for both years. Then an OLS regression is run with the difference variables plus the original covariates as additional regressors, restricting the sample to the baseline year (`year = 0`). This is done because the baseline year contains both the difference variables and the initial condition variables.

```
reg dlexptot ddfmfd98 ddfmfdyr dsexhead dagehead deduchehead
dlndland dvaccess dpcirr drice dwhtflr dmilk dmustoil dhenegg
sexhead agehead educhead lnland vaccess pcirr rice whtflr milk
mustoil henegg if year==0 [pw=weight];
```

The results show that, after controlling for the initial conditions, the impact of microcredit participation disappears ( $t = 1.42$ ):

Regression with robust standard errors	Number of obs =	826
	F( 23, 802) =	2.93
	Prob > F =	0.0000
	R-squared =	0.0917
	Root MSE =	.51074

dlexptot	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
ddfmd98	(dropped)					
<b>ddfmdyr</b>	<b>.0619405</b>	<b>.0435103</b>	<b>1.42</b>	<b>0.155</b>	<b>-.0234671</b>	<b>.1473481</b>
dsexhead	-.0615416	.0871488	-0.71	0.480	-.2326083	.1095251
dagehead	.0013583	.0023165	0.59	0.558	-.0031889	.0059055
deduchead	.0153497	.0117889	1.30	0.193	-.0077909	.0384904
dlnland	.1260302	.0701158	1.80	0.073	-.011602	.2636624
dvaccess	-.1365889	.0702504	-1.94	0.052	-.2744853	.0013075
dpcirr	.1042085	.1124156	0.93	0.354	-.1164551	.3248721
dtrice	.0065267	.0147616	0.44	0.659	-.0224493	.0355027
dwheat	-.04828	.0261598	-1.85	0.065	-.0996297	.0030697
dmilk	-.0071707	.0143637	-0.50	0.618	-.0353656	.0210241
doil	.0137635	.0062199	2.21	0.027	.0015542	.0259727
degg	.1991899	.101613	1.96	0.050	-.0002689	.3986486
dsexhead	-.1157563	.0844686	-1.37	0.171	-.281562	.0500494
dagehead	.0054212	.002046	2.65	0.008	.001405	.0094375
educhead	.0230352	.008891	2.59	0.010	.0055828	.0404876
dlnland	-.0690961	.0545822	-1.27	0.206	-.1762369	.0380448
vaccess	-.1142214	.1065896	-1.07	0.284	-.323449	.0950062
dpcirr	.1471455	.109057	1.35	0.178	-.0669254	.3612164
dtrice	-.0047485	.0317983	-0.15	0.881	-.0671661	.0576691
dwheat	-.0337045	.0306002	-1.10	0.271	-.0937705	.0263614
dmilk	-.0047502	.0129723	-0.37	0.714	-.0302138	.0207134
doil	.0205757	.0083353	2.47	0.014	.0042142	.0369373
degg	.1015795	.1273284	0.80	0.425	-.1483568	.3515158
_cons	-.704969	.5861648	-1.20	0.229	-1.855567	.4456292

## The DD Method Combined with Propensity Score Matching

The DD method can be refined in a number of ways. One is by using propensity score matching (PSM) with the baseline data to make certain the comparison group is similar to the treatment group and then applying double differences to the matched sample. This way, the observable heterogeneity in the initial conditions can be dealt with.

Using the “pscore” command, the participation variable in 1998/99 (which is created here as “dfmfd98” for both years) is regressed with 1991/92 exogenous variables to obtain propensity scores from the baseline data. These commands are as follows:

```
use ..\data\hh_9198,clear;
gen lnland=ln(1+hhland/100);
gen dfmfd1=dfmfd==1 & year==1;
egen dfmfd98=max(dfmfd1), by(nh);
keep if year==0;
pscore dfmfd98 sexhead agehead educhead lnland vaccess
pcirr rice wheat milk oil egg [pw=weight], pscore(ps98)
blockid(blockf1) comsup level(0.001);
```

The balancing property of the PSM has been satisfied, which means that households with the same propensity scores have the same distributions of all covariates for all five

blocks. The region of common support is [.06030439, .78893426], and 26 observations have been dropped:

\*\*\*\*\*  
 Algorithm to estimate the propensity score  
 \*\*\*\*\*

The treatment is dfmfd98

dfmfd98	Freq.	Percent	Cum.
0	391	47.34	47.34
1	435	52.66	100.00
Total	826	100.00	

Estimation of the propensity score

(sum of wgt is 8.2233e+02)  
 Iteration 0: log pseudolikelihood = -554.25786  
 Iteration 1: log pseudolikelihood = -480.05123  
 Iteration 2: log pseudolikelihood = -475.25432  
 Iteration 3: log pseudolikelihood = -475.17443  
 Iteration 4: log pseudolikelihood = -475.1744

Probit estimates	Number of obs =	826
	Wald chi2(11) =	78.73
	Prob > chi2 =	0.0000
Log pseudolikelihood = -475.1744	Pseudo R2 =	0.1427

dfmfd98	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
sexhead	-.1512794	.2698723	-0.56	0.575	-.6802194	.3776605
agehead	-.0073102	.0046942	-1.56	0.119	-.0165106	.0018903
educhead	-.0261142	.018235	-1.43	0.152	-.0618542	.0096257
lnland	-.9010234	.137662	-6.55	0.000	-1.170836	-.6312109
vaccess	.2894359	.2626682	1.10	0.271	-.2253843	.804256
pcirr	.0367083	.1999013	0.18	0.854	-.3550911	.4285077
rice	.1682276	.0606261	2.77	0.006	.0494028	.2870525
wheat	.0603593	.0500646	1.21	0.228	-.0377655	.1584841
milk	-.0472819	.0205877	-2.30	0.022	-.087633	-.0069309
oil	.009133	.0141985	0.64	0.520	-.0186954	.0369615
egg	-.2991866	.184372	-1.62	0.105	-.660549	.0621759
_cons	-1.002465	1.241022	-0.81	0.419	-3.434823	1.429894

Note: the common support option has been selected  
 The region of common support is [.06030439, .78893426]

Description of the estimated propensity score  
 in region of common support

Estimated propensity score			
Percentiles	Smallest		
1%	.0800224	.0603044	
5%	.1415098	.061277	
10%	.2124288	.0622054	Obs 800
25%	.3583033	.0647113	Sum of Wgt. 800

```

50%      .481352
75%      .570064      Largest      .7616697
90%      .6600336     .7650957
95%      .688278      .7716357
99%      .7515092     .7889343
Mean     .4579494
Std. Dev. .1612539
Variance .0260028
Skewness -.4881678
Kurtosis 2.637857

```

```

*****
Step 1: Identification of the optimal number of blocks
Use option detail if you want more detailed output
*****

```

The final number of blocks is 4

This number of blocks ensures that the mean propensity score is not different for treated and controls in each blocks

```

*****
Step 2: Test of balancing property of the propensity score
Use option detail if you want more detailed output
*****

```

The balancing property is satisfied

This table shows the inferior bound, the number of treated, and the number of controls for each block

Inferior of block of pscore	dfmfd98		Total
	0	1	
.0603044	53	16	69
.2	110	70	180
.4	151	250	401
.6	51	99	150
Total	365	435	800

Note: the common support option has been selected

```

*****
End of the algorithm to estimate the pscore
*****

```

The following commands keep the matched households in the baseline year and merge them with panel data to keep only the matched households in the panel sample:

```

keep if blockf1!=.;
keep nh;
sort nh;
merge nh using ..\data\hh_9198;
keep if _merge==3;

```

The next step is to implement the DD method as before. For this exercise, only the fixed-effects implementation is shown:

```

xtreg lexptot year dfmfd98 dfmfdyr sexhead agehead educhead
lnland vaccess pcirr rice wheat milk oil egg, fe i(nh);

```

The results show that applying PSM to DD retains the original positive impact of female participation in microcredit programs on household expenditure:

```

Fixed-effects (within) regression      Number of obs   =   1600
Group variable (i): nh                Number of groups =    800

R-sq: within   =   0.1791              Obs per group: min =     2
      between  =   0.1237                  avg   =    2.0
      overall   =   0.1434                  max   =     2

                                          F(13,787)      =   13.21
corr(u_i, Xb) = 0.0414                  Prob > F       =   0.0000
    
```

lexptot	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
year	.222509	.0639108	3.48	0.001	.0970532 .3479647
dfmfd98	(dropped)				
<b>dfmfdyr</b>	<b>.0925741</b>	<b>.0371517</b>	<b>2.49</b>	<b>0.013</b>	<b>.019646 .1655023</b>
sexhead	-.084584	.0739679	-1.14	0.253	-.2297818 .0606138
agehead	-.0003225	.001732	-0.19	0.852	-.0037223 .0030773
educhead	.0132322	.0084471	1.57	0.118	-.0033494 .0298138
lnland	.2003341	.0778701	2.57	0.010	.0474766 .3531917
vaccess	-.0857169	.0542065	-1.58	0.114	-.1921234 .0206896
pcirr	.083983	.0644159	1.30	0.193	-.0424644 .2104303
rice	.0131877	.0102657	1.28	0.199	-.0069638 .0333392
wheat	-.0272757	.0123259	-2.21	0.027	-.0514712 -.0030802
milk	-.0015386	.0064937	-0.24	0.813	-.0142857 .0112084
oil	.0047885	.0031592	1.52	0.130	-.001413 .0109899
egg	.1400882	.0485296	2.89	0.004	.0448254 .2353509
_cons	7.815588	.2504303	31.21	0.000	7.323998 8.307179
sigma_u	.33642591				
sigma_e	.36009944				
rho	.46605118	(fraction of variance due to u_i)			
F test that all u_i=0:		F(799, 787) =	1.58	Prob > F =	0.0000

## Notes

1. Panel data are not strictly needed for double-difference estimation. How this technique can be applied to cross-sectional data is shown later.
2. The negative sign in output means that outcome of participants (dfmfd = 1) is greater than that of nonparticipants (dfmfd = 0), thus implying that the participation impact is in fact positive.

## Reference

Ravallion, Martin. 2008. "Evaluating Anti-poverty Programs." In *Handbook of Development Economics*, vol. 4, ed. T. Paul Schultz and John Strauss, 3787–846. Amsterdam: North-Holland.