Statistics/Data Analysis

Description
help diff

Title

diff —— Difference in differences estimation

Syntax

diff *outcome_var* [*if*] [*in*] [*weight*] ,[*options*]

Description

diff runs several difference in differences (diff-in-diff) treatment effect estimations of a given outcome variable from a pooled baseline and follow up dataset: Single Diff-in-Diff, Diff-in-Diff accounting for covariates, Kernel Propensity Score Matching diff-in-diff, and the Quantile Diff-in-Diff. **diff** is also suitable for estimating repeated cross sections diff-in-diff (including the **kernel** option) and the triple difference-in-differences analysis.

Options

options	Description
Model - Required	
period(*varname***)**	Indicates the binary period variable (0: before; 1: after). Note: if your data contains a periodical frequency (monthly, quarterly, yearly, etc.), it is suggested to specify option **period(***varname***)** and include a binary variable for each frequency in option **cov(***varlist***)**.
treated(*varname***)**	Indicates the binary treatment variable (0: controls; 1:treated).
Optional	
cov(*varlist***)**	Specifies the pre-treatment covariates of the model. Also use this option to specify time fixed-effects in the case of multiple time-frequency data (e.g. monthly, yearly, quarterly, etc.). When option **kernel** is selected these variables are used to estimate the propensity score.
kernel	Performs the Kernel-based Propensity Score Matching diff-in-diff. This option generates the variable _weights containing the weights derived from the Kernel Propensity Score Matching, and _ps when the Propensity Score is not supplied in **pscore(***varname***)**, following Leuven and Sianesi (2014). This option requires the *id(varname)* of each unit or individual except under the repeated cross section **rcs**) setting.
id(*varname***)**	Option *kernel* requires the supply of the identification variable.
bw(#)	Supplied bandwidth of the Kernel function. The default bandwidth is 0.06.
ktype(*kernel***)**	Specifies the type of the Kernel function. The types are *epanechnikov* (the default), *gaussian*, *biweight*, *uniform* and *tricube*.
rcs	Indicates that the **kernel** is set for repeated cross section. This option does not require option **id(***varname***)**. Option **rcs** strongly assumes that covariates in **cov(***varlist***)** do not vary over time.
qdid(*quantile***)**	Performs the Quantile Difference in Differences estimation at the specified quantile from 0.1 to 0.9 (quantile 0.5 performs the QDID at the medeian). You may combine this option with **kernel** and **cov**. **qdid** does not support weights nor robust

	standard errors. This option uses **[R] qreg** and **[R] bsqreg** for bootstrapped standard errors
pscore(*varname***)**	Supplied Propensity Score.
logit	Specifies logit estimation of the Propensity Score. The default is Probit.
support	Performs **diff** on the common support of the propensity score given the option **kernel**.
addcov(*varlist***)**	Indicates additional covariates in addition to those specified in the estimation of the propensity score. Also use this option to specify time fixed-effects in the case of multiple time-frequency data (e.g. monthly, yearly, quarterly, etc.).
ddd(*varname***)**	Additional category for triple difference estimation. **treated(***varname***)** is deemed as the first category and **ddd(***varname***)** the second category. This option is not compatible with options **kernel**, **test** or **qdid(***quantile***)**.

SE/Robust

cluster(*varname***)**	Calculates clustered Std. Errors by *varname* .
robust	Calculates robust Std. Errors.
bs	performs a Bootstrap estimation of coefficients and standard errors.
reps(*int***)**	Specifies the number of repetitions when the **bs** is selected. The default are 50 repetitions.

Balancing test

test	Performs a balancing t-test of the difference in the means of the covariates between the control and treated groups in period == 0. The option *test* combined with *kernel* performs the balancing t-test with the weighted covariates. See **[R] ttest**

Reporting

report	Displays the inference of the included covariates or the estimation of the Propensity Score when option **kernel** is specified.
nostar	Removes the inference stars from the p-values.

Exporting results

You can export your results with outreg2. Run the following command after **diff** with double difference:

```
outreg2 using table_diff, ctitle(`r(depvar)') addstat(Mean control t(0),
    r(mean_c0), Mean treated t(0), r(mean_t0), Diff t(0), r(diff0), Mean
    control t(1), r(mean_c1), Mean treated t(1), r(mean_t1), Diff t(1),
    r(diff1)) label excel keep(_diff) nocons
```

Run the following command after **diff** with triple difference:

```
outreg2 using output, ctitle(`r(depvar)') addstat(Mean control - A t(0),
    r(mean_c0a), Mean control - B t(0), r(mean_c0b), Mean treated A -
    t(0), r(mean_t0a), Mean treated B - t(0), r(mean_t0b), Diff t(0),
    r(diff0), Mean ), r(mean_c1a), Mean control - B t(1), r(mean_c1b),
    Mean treated - A t(1), r(mean_t1a), Mean treated - B t(1),
    r(mean_t1b), Diff t(1), r(diff1)) label excel keep(_diff) nocons
    dec(4)
```

Results will be stored in the working directory (also see **help outreg2** for further options).

Example

1. Diff-in-Diff with no covariates.
We use the dataset form Card & Krueger (1994)*.

```
use "http://fmwww.bc.edu/repec/bocode/c/CardKrueger1994.dta"

diff fte, t(treated) p(t)
```

For bootstrapped std. err.:

```
diff fte, t(treated) p(t) bs rep(50)
```

2. Diff-in-Diff with covariates.

```
diff fte, t(treated) p(t) cov(bk kfc roys)
diff fte, t(treated) p(t) cov(bk kfc roys) report
diff fte, t(treated) p(t) cov(bk kfc roys) report bs
```

3. Kernel Propensity Score Diff-in-Diff.

```
diff fte, t(treated) p(t) cov(bk kfc roys) kernel id(id)
diff fte, t(treated) p(t) cov(bk kfc roys) kernel id(id) support
diff fte, t(treated) p(t) cov(bk kfc roys) kernel id(id) support addco
> v(wendys)
diff fte, t(treated) p(t) kernel id(id) ktype(gaussian) pscore( ps)
diff fte, t(treated) p(t) kernel id(id) ktype(gaussian) pscore( ps) bs
>  reps(50)
```

3. Kernel Propensity Score Diff-in-Diff (Repeated Cross Section - rcs).

```
diff fte, t(treated) p(t) cov(bk kfc roys) kernel rcs
diff fte, t(treated) p(t) cov(bk kfc roys) kernel rcs support
diff fte, t(treated) p(t) cov(bk kfc roys) kernel rcs support addcov(w
> endys)
diff fte, t(treated) p(t) kernel rcs ktype(gaussian) pscore(_ps)
diff fte, t(treated) p(t) cov(bk kfc roys) kernel rcs support addcov(w
> endys) bs reps(50)
```

4. Quantile Diff-in-Diff.

```
diff fte, t(treated) p(t) qdid(0.25)
diff fte, t(treated) p(t) qdid(0.50)
diff fte, t(treated) p(t) qdid(0.75)
diff fte, t(treated) p(t) qdid(0.50) cov(bk kfc roys)
diff fte, t(treated) p(t) qdid(0.50) cov(bk kfc roys) kernel id(id)
diff fte, t(treated) p(t) qdid(0.50) cov(bk kfc roys) kernel rcs
```

5. Balancing test of covariates.

```
diff fte, t(treated) p(t) cov(bk kfc roys wendys) test
diff fte, t(treated) p(t) cov(bk kfc roys wendys) test id(id) kernel
diff fte, t(treated) p(t) cov(bk kfc roys wendys) test kernel rcs
```

6. Triple differences (consider bk is a second treatment category).

```
diff fte, t(treated) p(t) ddd(bk)
```

Saved results

diff saves the following list of scalars in **r()**:

r(N)	total number of observations.
r(N_t0)	number of observations in period == 0.
r(N_t1)	number of observations in period == 1.
r(R2)	R-square
r(mean_c0)	mean of *output_var* of the control group in period == 0
r(mean_c0a)	mean of *output_var* of the control group A in period == 0
r(mean_c0b)	mean of *output_var* of the control group B in period == 0
r(mean_t0)	mean of *output_var* of the treated group in period == 0
r(mean_t0a)	mean of *output_var* of the treated group A in period == 0
r(mean_t0b)	mean of *output_var* of the treated group B in period == 0
r(diff0)	difference of the mean of *output_var* between treated and control groups in period == 0
r(mean_c1)	mean of *output_var* of the control group in period == 1
r(mean_c1a)	mean of *output_var* of the control group A in period == 1
r(mean_c1b)	mean of *output_var* of the control group B in period == 1
r(mean_t1)	mean of *output_var* of the treated group in period == 1
r(mean_t1a)	mean of *output_var* of the treated group A in period == 1
r(mean_t1b)	mean of *output_var* of the treated group B in period == 1
r(diff1)	difference of the mean of *output_var* between treated and control groups in period == 1
r(did)	differences in differences - Treatment Effect
r(se_c0)	Standard error of the mean of *output_var* of the control group in period == 0
r(se_c0a)	Standard error of the mean of *output_var* of the control group A in period == 0
r(se_c0b)	Standard error of the mean of *output_var* of the control group B in period == 0
r(se_t0)	standard errors of the mean of *output_var* of the treated group in period == 0
r(se_t0a)	standard errors of the mean of *output_var* of the treated group A in period == 0
r(se_t0b)	standard errors of the mean of *output_var* of the treated group B in period == 0
r(se_d0)	standard Errors of the difference of *output_var* between the treated and control groups in period == 0
r(se_c1)	standard errors of the mean of *output_var* of the control group in period == 1
r(se_c1a)	standard errors of the mean of *output_var* of the control group A in period == 1
r(se_c1b)	standard errors of the mean of *output_var* of the control group B in period == 1
r(se_t1)	standard errors of the mean of *output_var* of the treated group in period == 1
r(se_t1a)	standard errors of the mean of *output_var* of the treated group A in period == 1
r(se_t1b)	standard errors of the mean of *output_var* of the treated group B in period == 1
r(se_d1)	standard errors of the difference of *output_var* between the treated and control groups in == 0
r(se_dd)	standard errors of the difference in difference
r(se_dd)	standard errors of the difference in difference

Recommended references

Single diff-in-diff:

Card, D., Krueger, A. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania". The American Economic Review, Vol. 84, No. 4 (Sep., 1994), pp. 772-793.

Kernel diff-in-diff:

Heckman, J., Ichimura, H., Todd, P. "Matching As an Econometric Evaluation Estimator". The Review of Economic Studies, Vol. 65, No. 2 (Apr., 1998), pp. 261-294.

Leuven, E., Sianesi, B. 2014. "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing, Statistical Software Components". Boston College Department of Economics.

Kernel diff-in-diff (repeated cross section):

Blundell, R., Dias, M. "Alternative Approaches to Evaluation in Empirical Microeconomics". Journal of Human Resources, Vol. 44, No. 3 (Jun., 2009), pp. 565-640.

Quantile diff-in-diff:

Meyer, B., Viscusi, W. "Workers' Compensation and Injury Duration: Evidence from a Natural Experiment". The American Economic Review, Vol. 85, No.3 (Jun., 1995), pp. 322-340.

Triple difference in differences:

Imbens, G., Wooldridge, J. "Difference-in-Differences Estimation. Lecture Notes 10, Summer '07". NBER (Jul., 2007), pp. 322-340.

Author

Juan M. Villa
Global Development Institute
The University of Manchester
juan.villa@manchester.ac.uk

Colpensiones

Please cite as: Villa, J.M., 2016. diff: Simplifying the estimation of difference-in-differences treatment effects. Stata Journal 16, pp. 52-71.

This version: December - 2019. Click here periodically to get the lastest version.
*Acknowledgements to Kit Baum for valuable comments. The Kernel matching is based on the command psmatch2 developved by Edwin Leuven and Barbara Sianesi.