

Estimation of pre- and post-treatment Average Treatment Effects (ATEs) with binary time-varying treatment using Stata

Giovanni Cerulli
CNR-IRCrES

Marco Ventura
ISTAT

Methodological and Data Quality Division, Rome, Italy

Italian Stata Users Group meeting
Florence 16th Nov. 2017

- 1 Motivations
- 2 Our contribution
- 3 The econometric set up
- 4 Testing for the parallel trend assumption
- 5 the Stata syntax of `ddid`
- 6 An application on simulated data
- 7 Further developments

Motivations (1)

- Main question: are public policy programs effective?
- If yes how long and to what extent?
- Fundamental problem: treated individuals not randomly selected but rather self-selected
- (possible) solution: recovering the Average Treatment Effect (ATE) from panel data, Diff-in-Diff.

THE AIM OF THE WORK IS:

- to provide a Stata routine, `ddid`, which implements a generalization of the Difference-In-Differences (DID) estimator
- to provide a user friendly Stata routine to estimate the pre- and post-intervention effects
- to implement diagnostic tests for the parallel trend assumption
- to facilitate provide useful means for plotting the results in a easy-to-read graphical representation

The econometric set up (1)

Let us consider a binary treatment indicator

$$D_{it} = \begin{cases} 1 & \text{if unit } i \text{ is treated at time } t \\ 0 & \text{if unit } i \text{ is not treated at time } t \end{cases}$$

and an outcome equation with **contemporaneous** treatment plus **lags** and **leads**

$$Y_{it} = \mu_{it} + \beta_{-1}D_{it-1} + \beta_0D_{it} + \beta_{+1}D_{it+1} + \gamma\mathbf{x}_{it} + u_{it} \quad (1)$$

the β_{+1} coefficient measures the impact of the treatment one period before the treatment occurred and β_{-1} measures the impact of treatment one period after the treatment occurred.

The econometric set up (2)

let us assume that treatment can occur only once over the interval $[t - 1, t + 1]$ so that we can define the following sequences of possible treatments:

$$\{w^j\} = \{D_{it-1}, D_{it}, D_{it+1}\} = \begin{cases} w^1 = (0, 0, 0) \\ w^2 = (1, 0, 0) \\ w^3 = (0, 1, 0) \\ w^4 = (0, 0, 1) \end{cases}$$

The sequence w^1 is the usual benchmark of no-treatment.

The generic treatment sequence is indicated by w^j (with $j = 1, \dots, 4$) and the associated potential outcome as $Y(w^j)$.

The “Average Treatment Effect between two potential outcomes, w^j and w^k ” $Y(w^j)$ and $Y(w^k)$ ” is defined as:

$$ATE_{jk} = E[Y_{it}(w^j) - Y_{it}(w^k)] \quad \forall (i, t) \quad (2)$$

The econometric set up (3)

with treatment occurring only in one period out of three, and one lag and one lead we can define six possible ATEs:

$$\begin{bmatrix} & w_1 & w_2 & w_3 & w_4 \\ w_1 & - & & & \\ w_2 & ATE_{21} & - & & \\ w_3 & ATE_{31} & ATE_{32} & - & \\ w_4 & ATE_{41} & ATE_{42} & ATE_{43} & - \end{bmatrix}$$

The generic ATE_{ij} represents the ATE of the sequence i against the counterfactual sequence j . Obviously $ATE_{ij} = -ATE_{ji}$.

The econometric set up (4)

Using equation (1) and the definition of w^j , with $j = 1, \dots, 4$, it is possible to rewrite the ATEs

$$ATE_{21} = E(Y_{it}|w_2) - E(Y_{it}|w_1)] = (\bar{\mu} + \beta_{-1} + \gamma\bar{x}) - (\bar{\mu} + \gamma\bar{x}) = \beta_{-1}$$

$$ATE_{31} = E(Y_{it}|w_3) - E(Y_{it}|w_1)] = \beta_0$$

$$ATE_{41} = E(Y_{it}|w_4) - E(Y_{it}|w_1)] = \beta_{+1}$$

$$ATE_{32} = E(Y_{it}|w_3) - E(Y_{it}|w_2)] = \beta_0 - \beta_{-1}$$

$$ATE_{42} = E(Y_{it}|w_4) - E(Y_{it}|w_2)] = \beta_{+1} - \beta_{-1}$$

$$ATE_{43} = E(Y_{it}|w_4) - E(Y_{it}|w_3)] = \beta_{+1} - \beta_0$$

The ATEs have a straightforward interpretation:

The econometric set up (5)

- $\beta_{+1} \neq 0$. Treatment delivered at t affects the outcome at $t - 1$. Current treatment has an effect on past outcome (anticipatory effect). Therefore, the pre-treatment period is affected by the current treatment.
- $\beta_0 \neq 0$. Treatment delivered at t affects the outcome at t , simultaneous effect.
- $\beta_{-1} \neq 0$. Treatment delivered at t affects the outcome at $t + 1$. Current treatment has an effect on future outcomes (lagged effect). Therefore, the post-treatment period is affected by current treatment.

Parallel trend assumption: Test 1

In the spirit of Granger (1969) if D_{it} causes $Y_{it} \implies, \beta_{+j} = 0$ for $j = 1, \dots, J$ in an equation like (1). **NO anticipatory effects**

$$H_0 : \beta_{+1} = \beta_{+2} = \dots = \beta_{+J} = 0 \quad (3)$$

BEWARE: rejecting H_0 would invalidate the causal interpretation of the estimates, but ...

not rejecting H_0 implies only that a necessary condition for the parallel trend assumption holds.

The necessary and sufficient condition still remains untestable being formulated on counterfactual unobservable quantities.

parallel trend assumption: Test 2

Another way to test for the necessary condition of the parallel trend assumption

Drop lags and leads from equation (1) and augment it with the time trend variable t , and the interaction between D_{it} and t .

If the coefficient of the interaction term turns out to be statistically equal to zero, one can reasonably expect the parallel trend to hold.

See Angrist and Pischke (2009, pp. 238–239)

Proof: let us write down the following potential outcome model:

$$\begin{cases} Y_{0,it} = \mu_0 + \lambda_0 t + \gamma \mathbf{x}_{it} + \theta_i + u_{0,it} \\ Y_{1,it} = \mu_1 + \lambda_1 t + \gamma \mathbf{x}_{it} + \theta_i + u_{1,it} \\ Y_{it} = Y_{0,it} + D_{it} (Y_{1,it} - Y_{0,it}) \end{cases}$$

By substituting the first two equations into the third, we obtain:

$$Y_{it} = \mu_0 + \lambda_0 t + \gamma \mathbf{x}_{it} + D_{it}(\mu_1 - \mu_0) + D_{it}t(\lambda_1 - \lambda_0) + \theta_i + \eta_{it}$$

with $\eta_{it} = [u_{0,it} + D_{it} (u_{1,it} - u_{0,it})]$.

parallel trend assumption: Test 2

in a more compact form:

$$Y_{it} = \mu_0 + \lambda_0 t + \gamma \mathbf{x}_{it} + D_{it} \mu + D_{it} t \cdot \lambda + \theta_i + \eta_{it} \quad (4)$$

estimable by FE, and the following test can be performed:

$$H_0 : \lambda = 0$$

if H_0 is accepted, we can reasonably hold that the (necessary condition for the) parallel trend assumption is satisfied.

This test can be generalized assuming also quadratic or cubic time trend.

The Stata syntax of `ddid` (1)

```
ddid outcome treatment [varlist] [if] [in] [weight], model(modeltype)  
pre(#) post(#) [test_tt graph save_graph(graphname) vce(vcetype)]
```

fweights, *iwweights*, and *pweights* are allowed;
where:

- *outcome*: is the target variable over which measuring the impact of the treatment.
- *treatment*: is the binary treatment variable taking 1 for treated, and 0 for untreated units.
- *varlist*: is the set of pre-treatment (or observable confounding) variables.

The Stata syntax of `ddid` (2)

Options

- **model**(*modeltype*) specifies the estimation model, where *modeltype* must be one out of these two alternatives: “fe” (fixed effects), or “ols” (ordinary least squares). It is always required to specify one model.
- **pre**(#) allows to specify the number (#) of pre-treatment periods.
- **post**(#) allows to specify the number (#) of post-treatment periods.
- **test_tt** allows for performing the parallel-trend test using the time-trend approach. The default is to use the leads.
- **graph** allows for a graphical representation of results. It uses the `coefplot` command implemented by Jann (2014).
- **save_graph**(*graphname*) permits to save the graph as *graphname*.
- **vce**(*vcetype*) allows for robust and clustered regression standard errors in model’s estimates.

The Stata syntax of `ddid` (3)

ddid creates a number of variables:

_D_L1, ..., _D_Lm: are the lags of the treatment variable, with m equal to $\#$ in the **post**($\#$) option.

_D_F1, ..., _D_Fp: are the leads of the treatment variable, with p equal to $\#$ in the **pre**($\#$) option

and returns the following scalars:

`e(N)` is the total number of (used) observations.

`e(N1)` is the number of (used) treated units.

`e(N0)` is the number of (used) untreated units.

`e(ate)` is the value of the (contemporaneous) ATE.

REMEMBER: (i) the treatment has to be a 0/1 binary variable;
(ii) before running **ddid**, one has to install the **coefplot** user-written Stata command (Jann, 2014).

An application on simulated data (1)

```
. clear
. set obs 5
. set seed 10101
. gen id=_n
. expand 50
. drop in 1/5
. bys id: gen time=_n+1999
. gen D=rbinomial(1,0.4)
. gen x1=rnormal(1,7)
. tsset id time
  forvalues i=1/6{
    gen L'i'_x=L'i'.x1
  }
```

An application on simulated data (2)

```
bys id: gen y0=5+1*x+ rnormal()  
bys id: gen y1=100+5*x+90*L1_x+90*L2_x+120*L3_x+100*L4_x+ ///  
90*L5_x+90*L6_x+rnormal()  
  
gen A=6*x+rnormal()  
replace D=1 if A>=15  
replace D=0 if A<15  
gen y=y0+D*(y1-y0)  
  
tsset id time  
xi: ddid y D x, model(fe) pre(6) post(6) vce(robust) graph test_tt
```

An application on simulated data (3)

```
between = 0.3825          avg =      37.0
overall = 0.3865          max =      37

corr(u_1, Xb) = -0.1395      F(4,4)      =      .
                               Prob > F       =      .

                               (Std. Err. adjusted for 5 clusters in id)
```

y	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
_D_F6	-72.46167	180.2637	-0.40	0.708	-572.9538	428.0305
_D_F5	60.09743	63.70172	0.94	0.399	-116.7669	236.9618
_D_F4	22.75842	278.3109	0.08	0.939	-749.9566	795.4735
_D_F3	-143.4906	158.6749	-0.90	0.417	-584.0427	297.0614
_D_F2	395.8175	257.9225	1.53	0.200	-320.2905	1111.926
_D_F1	-87.25493	186.4611	-0.47	0.664	-604.9539	430.444
D	805.2605	215.0375	3.74	0.020	208.2206	1402.3
_D_L1	323.9564	123.6558	2.62	0.059	-19.36713	667.2799
_D_L2	595.6533	183.0705	3.25	0.031	87.36797	1103.939
_D_L3	494.4453	123.5111	4.00	0.016	151.5236	837.367
_D_L4	446.2026	156.2224	2.86	0.046	12.45967	879.9456
_D_L5	499.1779	211.6438	2.36	0.078	-88.43944	1086.795
_D_L6	301.861	73.76806	4.09	0.015	97.04799	506.6739
x1	-9.918519	18.52512	-0.54	0.621	-61.35249	41.51545
_cons	-1107.346	307.6728	-3.60	0.023	-1961.582	-253.1091
sigma_u	161.27232					
sigma_e	935.12866					
rho	.0288834	(fraction of variance due to u_1)				

An application on simulated data (4)

```
*****  
***** Test for 'parallel trend' using the 'leads' *****  
*****
```

```
( 1)  _D_F6 = 0
```

```
( 2)  _D_F5 = 0
```

```
( 3)  _D_F4 = 0
```

```
( 4)  _D_F3 = 0
```

```
( 5)  _D_F2 = 0
```

```
( 6)  _D_F1 = 0
```

```
Constraint 2 dropped
```

```
Constraint 6 dropped
```

```
F( 4, 4) = 0.42
```

```
Prob > F = 0.7875
```

```
RESULT: 'Parallel-trend' passed
```

```
*****
```

An application on simulated data (5)

```
*****  
***** Test for 'parallel trend' using the 'time-trend' *****  
*****
```

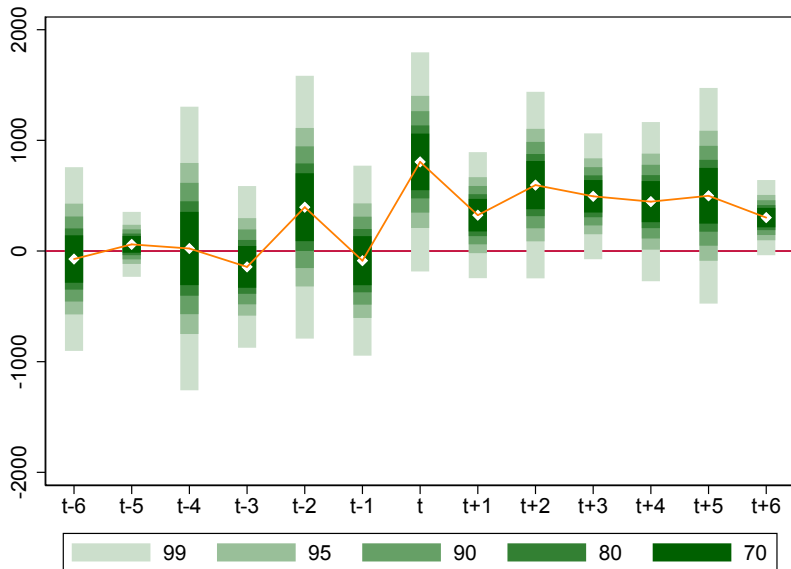
```
( 1)  _DT = 0
```

```
      F( 1,      4) =      1.44  
      Prob > F =      0.2961
```

```
RESULT: 'Parallel-trend' passed
```

```
*****
```

An application on simulated data (6)



An application on simulated data (7)

The option graph provides a graphical representation of the results plotting the lags and leads coefficients with 99, 95, 90, 80, and 70 confidence intervals.

The pre-treatment pattern lays around zero

The post-treatment pattern shows the positive effect of the (simulated) policy with a value laying around 500.

Assuming the sufficient condition of parallel trend to hold, one can conclude that this policy has generated positive effects.

Further developments

- ① non binary treatment;
- ② more than one treatment over the sequence w^j , with $j = 1, \dots, 4$;