The Stata Journal (2019)
19, Number 3, pp. 551–565 DOI: 10.1177/1536867X19874224

Estimation of pre- and posttreatment average treatment effects with binary time-varying treatment using Stata

Giovanni Cerulli
CNR-IRCrES
National Research Council of Italy
Research Institute on Sustainable Economic Growth
Rome, Italy
giovanni.cerulli@ircres.cnr.it

Marco Ventura
Department of Economics and Law
Sapienza University of Rome
Rome, Italy
marco.ventura@uniroma1.it

Abstract. In this article, we describe `tvdiff`, a community-contributed command that implements a generalization of the difference-in-differences estimator to the case of binary time-varying treatment with pre- and postintervention periods. `tvdiff` is flexible and can accommodate many actual situations, enabling the user to specify the number of pre- and postintervention periods and a graphical representation of the estimated coefficients. In addition, `tvdiff` provides two distinct tests for the necessary condition of the identification of causal effects, namely, two tests for the so-called parallel-trend assumption. `tvdiff` is intended to simplify applied works on program evaluation and causal inference when longitudinal data are available.

Keywords: st0566, tvdiff, treatment effects, difference-in-differences

1 Introduction

In this article, we present the community-contributed command `tvdiff`, which implements a generalization of the difference-in-differences (DID) estimator to the case of (many) pre- and postintervention periods. `tvdiff` estimates average treatment effects (ATEs) when the treatment variable is binary and varying over time, allowing the user to estimate pre- and postintervention effects by selecting the number of pre- and postintervention periods. The results are automatically plotted in an easy-to-read graph. Furthermore, to assess the reliability of the main identification hypothesis, `tvdiff` allows one to test the so-called parallel-trend or common-trend assumption implied by the underlying econometric model. To accomplish this, `tvdiff` performs two tests: one using time leads and one using an additional time-trend variable.

This article is organized as follows. Section 2 introduces the econometrics underlying the estimation model and shows a graph of the estimated coefficients. Section 3 explains the rationale underlying parallel-trend tests and shows how to carry them out. Section 4 presents the syntax of `tvdiff` and a detailed explanation of the command's options. Section 5 shows the use of `tvdiff` by comparing ordinary least squares (OLS) with fixed-effects estimation results on a simulated data-generating process (DGP), assuming hidden selection bias. Section 6 provides an application to real data, measuring the effect of public education expenditure on income equality at the country level. Section 7 concludes the article.

2 The model

We focus on the estimation of treatment effects in the presence of binary time-varying treatment. Such a setting characterizes several economic and social policies and medical trials delivered over time. For example, one could be interested in assessing whether a certain treatment had an impact on a given target variable with some delay and whether anticipatory effects took place. To formalize this setting, let us start by considering a binary treatment indicator for individual i at time t:

$$D_{it} = \begin{cases} 1 \text{ if unit } i \text{ is treated at time } t \\ 0 \text{ otherwise} \end{cases}$$

Let us also assume an outcome equation with contemporaneous treatment plus one lag and one lead:

$$Y_{it} = \mu_{it} + \beta_{-1} D_{it-1} + \beta_0 D_{it} + \beta_{+1} D_{it+1} + \boldsymbol{\gamma} \mathbf{x}_{it} + u_{it} \tag{1}$$

In (1), the β_{+1} coefficient measures the impact of the treatment one period before its occurrence, and β_{-1} measures the impact of treatment one period after it. $\mathbf{x}_{it}$ is a vector of covariates, $\boldsymbol{\gamma}$ is its conformable coefficient vector, and μ_{it} represents a fixed effect. Autor (2003) provided a first application of the treatment model implied by (1).[1]

For now, 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,

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

where the sequence w^1 is the usual benchmark of no treatment. The generic sequence is denoted as w^j (with $j = 1, \ldots, J$ and $J = 4$) and the associated potential outcome

1. Actually, the notation may seem slightly deceptive because the coefficient of the lead measures the impact of the treatment one period before its occurrence, while the coefficient of the lagged treatment variable measures the impact of treatment one period after treatment. These counter-intuitive effects stem from how the dataset is built when lags and leads are included. Indeed, when a lag is introduced, the $\mathbf{Y}$ vector is shifted one period ahead so that the Y_{t-1} figure refers to time t. Similarly for the lead, the Y_{t+1} figure is shifted backward in correspondence with time t.

as $Y(w^j)$. The "ATE between the two potential outcomes $Y(w^j)$ and $Y(w^k)$" can be easily defined as

$$\text{ATE}_{jk} = E\left\{Y_{it}\left(w^j\right) - Y_{it}\left(w^k\right)\right\}$$

for $j, k = 1, \ldots, 4$ and $j \neq k$.

Under conditional mean independence—that is, conditioning on both $\mathbf{x}_{it}$ and the fixed effect μ_{it}—we have

$$\begin{aligned}\text{ATE}_{jk} &= E_{\mathbf{x},\mu}\left\{\text{ATE}_{jk}\left(\mathbf{x}_{it}, \mu_{it}\right)\right\} = E_{\mathbf{x},\mu}\left[E\left\{Y_{it}(w^j, \mathbf{x}_{it}, \mu_{it}) - Y_{it}(w^k, \mathbf{x}_{it}, \mu_{it})\right\}\right] \\ &= E_{\mathbf{x},\mu}\left\{E\left(Y_{it}|w^j, \mathbf{x}_{it}, \mu_{it}\right) - E\left(Y_{it}|w^k, \mathbf{x}_{it}, \mu_{it}\right)\right\}\end{aligned}$$

In such a model—with treatment occurring only once out of three periods, plus one lag and one lead of the treatment variable—we can define six possible ATEs that, for ease of reference, we collect in a matrix,

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

where the generic ATE_{jk} represents the ATE of the sequence j against the counterfactual sequence k. Obviously, $\text{ATE}_{jk} = -\text{ATE}_{kj}$. Using (1) and the definition of w^j with $j = 1, \ldots, 4$, we can show that

$$\begin{aligned}\text{ATE}_{21} &= E(Y_{it}|w_2) - E(Y_{it}|w_1) = (\overline{\mu} + \beta_{-1} + \boldsymbol{\gamma}\overline{\mathbf{x}}) - (\overline{\mu} + \boldsymbol{\gamma}\overline{\mathbf{x}}) = \beta_{-1} \\ \text{ATE}_{31} &= E(Y_{it}|w_3) - E(Y_{it}|w_1) = \beta_0 \\ \text{ATE}_{41} &= E(Y_{it}|w_4) - E(Y_{it}|w_1) = \beta_{+1} \\ \text{ATE}_{32} &= E(Y_{it}|w_3) - E(Y_{it}|w_2) = \beta_0 - \beta_{-1} \\ \text{ATE}_{42} &= E(Y_{it}|w_4) - E(Y_{it}|w_2) = \beta_{+1} - \beta_{-1} \\ \text{ATE}_{43} &= E(Y_{it}|w_4) - E(Y_{it}|w_3) = \beta_{+1} - \beta_0\end{aligned}$$

In general, one obtains a number of ATEs equal to $(J^2 - J)/2$, where J is the number of treatment sequences; in our example, we have $(4^2 - 4)/2 = 6$ ATEs. An important advantage of a dynamic treatment model is the ability to graphically plot the evolution of the treatment effects over time. To this end, let us define the predictions of Y_{it} given the sequence of treatments as

$$E(Y_{it}|D_{it-1}, D_{it}, D_{it+1}, t) = \overline{\mu}_t + \beta_{-1}D_{it-1} + \beta_0 D_{it} + \beta_{+1}D_{it+1} + \boldsymbol{\gamma}\overline{\mathbf{x}}_t \qquad (3)$$

Consistently with the econometric practice, to make (3) computable, we assume additive separability; that is, $\mu_{it} = \theta_i + \delta_t$, where θ_i and δ_t represent individual and time-fixed effects, respectively. It follows that $\overline{\mu}_{it} \equiv E(\mu_{it}|D_{it-1}, D_{it}, D_{it+1}, t) = E(\mu_{it}|t) = E(\theta_i + \delta_t|t) = \overline{\theta} + \delta_t$.

To keep things simple, let us restrict our attention only to the case of two specific treatment sequences,

$$\begin{aligned}w^T &= \{\ldots, D_{it-2} = 0, D_{it-1} = 0, D_{it} = 1, D_{it+1} = 0, D_{it+2} = 0, \ldots\} \\ w^C &= \{\ldots, D_{it-2} = 0, D_{it-1} = 0, D_{it} = 0, D_{it+1} = 0, D_{it+2} = 0, \ldots\}\end{aligned}$$

where w^T indicates the sequence in which treatment occurs only at time t and w^C indicates the no-treatment case. By setting $A_{it} = \{D_{it-1}, D_{it}, D_{it+1}, t\}$ and iterating (3) one period back and one period forward, one obtains the prediction of Y at $t-1$, t, and $t+1$,

$$\begin{aligned} E(Y_{it-1}|A_{it-1}) &= \overline{\mu}_{t-1} + \beta_{-1}D_{it-2} + \beta_0 D_{it-1} + \beta_{+1}D_{it} + \boldsymbol{\gamma}\overline{\mathbf{x}}_{t-1} \\ E(Y_{it}|A_{it}) &= \overline{\mu}_{t} + \beta_{-1}D_{it-1} + \beta_0 D_{it} + \beta_{+1}D_{it+1} + \boldsymbol{\gamma}\overline{\mathbf{x}}_{t} \\ E(Y_{it+1}|A_{it+1}) &= \overline{\mu}_{t+1} + \beta_{-1}D_{it} + \beta_0 D_{it+1} + \beta_{+1}D_{it+2} + \boldsymbol{\gamma}\overline{\mathbf{x}}_{t+1} \end{aligned}$$

which can be used to calculate the expected outcome over $t-1, t, t+1$ conditional on w^T and w^C. Thus,

- for w^T, we have

$$\begin{aligned} &E(Y_{it-1}|w^T = \ldots, 0, 0, 1, 0, 0, \ldots) = \overline{\mu}_{t-1} + \beta_{+1} + \boldsymbol{\gamma}\overline{\mathbf{x}}_{t-1} \\ &E(Y_{it}|w^T \quad = \ldots, 0, 0, 1, 0, 0, \ldots) = \overline{\mu}_{t} + \beta_{0} + \boldsymbol{\gamma}\overline{\mathbf{x}}_{t} \\ &E(Y_{it+1}|w^T = \ldots, 0, 0, 1, 0, 0, \ldots) = \overline{\mu}_{t+1} + \beta_{-1} + \boldsymbol{\gamma}\overline{\mathbf{x}}_{t+1} \end{aligned}$$

- for w^C, we have

$$\begin{aligned} &E(Y_{it-1}|w^C = \ldots, 0, 0, 0, 0, 0, \ldots) = \overline{\mu}_{t-1} + \boldsymbol{\gamma}\overline{\mathbf{x}}_{t-1} \\ &E(Y_{it}|w^C \quad = \ldots, 0, 0, 0, 0, 0, \ldots) = \overline{\mu}_{t} + \boldsymbol{\gamma}\overline{\mathbf{x}}_{t} \\ &E(Y_{it+1}|w^C = \ldots, 0, 0, 0, 0, 0, \ldots) = \overline{\mu}_{t+1} + \boldsymbol{\gamma}\overline{\mathbf{x}}_{t+1} \end{aligned}$$

We can now plot these predictions over time (figure 1) and depict these situations:

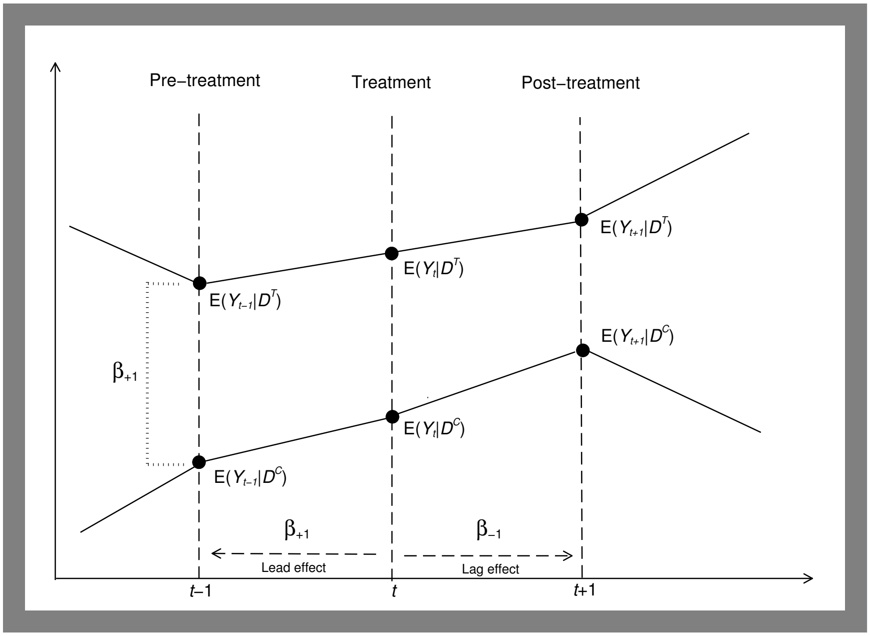

Figure 1. Pre- $(t-1)$ and post- $(t+1)$ treatment effect of a policy delivered at t. Source: Cerulli (2015, 202).

- If $\beta_{+1} \neq 0$, treatment delivered at time t affects the outcome at time $t-1$. Current treatment has an effect on past outcomes (anticipatory effect). Therefore, the pretreatment period is affected by current treatment.
- If $\beta_0 \neq 0$, treatment delivered at time t affects the outcome at time t, generating contemporaneous effects.
- If $\beta_{-1} \neq 0$, treatment delivered at time t affects the outcome at time $t+1$. Current treatment has an effect on future outcomes (lagged effect). Therefore, the posttreatment period is affected by current treatment.

3 Testing the parallel-trend assumption

The pattern of the leads is also important to check for causality in the spirit of Granger (1969). Indeed, conditional on $\mathbf{x}_{it}$ and the fixed effect, if D_{it} causes Y_{it}, the leads should not be jointly different from zero in an equation like (1). A test for checking whether all the β_{+s}'s are jointly equal to zero, $s = 1, \ldots, S$, indirectly tests whether the parallel-trend assumption holds. Formally, we can define such a test as

$$H_0\colon \beta_{+1} = \beta_{+2} = \cdots = \beta_{+S} = 0$$

Note that rejecting H_0 invalidates the causal interpretation of the estimates, while not rejecting H_0 implies only a necessary condition for the parallel-trend to hold because the necessary and sufficient condition still remains untestable, being formulated on counterfactual unobservable quantities.

Another approach to test the parallel-trend assumption (still a necessary condition) requires dropping lags and leads from (1) and augmenting it with the time-trend variable t and its interaction with D_{it}. If the coefficient of the interaction term is statistically not significant, one can reasonably expect the parallel-trend assumption to hold (see Angrist and Pischke [2009, 238–239]).

To provide ground for such a test, let us write down the following potential-outcome model:

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

We allow again for an individual fixed effect (θ_i) and a time effect (δ_t), with the parameters λ_1 and λ_0 being the treated and untreated time trends, respectively. This way, by plugging the first two equations into the third one, we obtain

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

with $\eta_{it} = [u_{0,it} + D_{it}\left(u_{1,it} - u_{0,it}\right)]$. Equivalently, we can write the previous equation as

$$Y_{it} = \mu_0 + \lambda_0 t + \gamma \mathbf{x}_{it} + D_{it}\mu + D_{it} \times t \times \lambda + \theta_i + \delta_t + \eta_{it}$$

which can be consistently estimated by a fixed-effects regression where the significance test for $\lambda = (\lambda_1 - \lambda_0)$ provides a test for the parallel-trend assumption. Accepting the null H_0: $\lambda = 0$ implies accepting that the parallel-trend assumption is not violated whenever one assumes no "anticipation effects").

Finally, note that we can extend the previous test by also considering quadratic or even cubic time trends.

4 The tvdiff command

4.1 Description

`tvdiff` estimates ATEs in setups like the one just discussed, namely, when treatment is binary and varying over time. Using `tvdiff`, the user can estimate the pre- and postintervention effects by selecting the pre- and postintervention periods and plot the results in an easy-to-read graph. To assess the reliability of the causal interpretation of the results achieved by the user's specified model, `tvdiff` allows one to test for the parallel-trend assumption using two tests, that is, the joint test on the leads and the time-trend test. `tvdiff` is a generalization of the DID approach to the case of many post- and preintervention times.

4.2 Syntax

The syntax of the command is as follows:

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

outcome is the target variable measuring the impact of treatment.

treatment is the binary treatment variable taking a value of 1 for treated units and a value of 0 for untreated units.

varlist is the set of pretreatment (or observable confounding) variables.

`aweight`s, `fweight`s, and `pweight`s are allowed; see [U] **11.1.6 weight**.

4.3 Options

`model(`*modeltype*`)` specifies the estimation model, where *modeltype* must be `fe` (fixed effects) or `ols` (OLS). `model()` is required.

`pre(`#`)` specifies the number (#) of pretreatment periods. `pre()` is required.

`post(`#`)` specifies the number (#) of posttreatment periods. `post()` is required.

`test_tt` performs the parallel-trend test using the time-trend approach. The default is to use the leads.

`graph` allows for a graph of the results. It uses the `coefplot` command implemented by Jann (2014).

`save_graph(`*graphname*`)` allows one to save the graph as *graphname*.

`vce(`*vcetype*`)` allows for robust and clustered regression standard errors in the model's estimates.

4.4 Remarks

`tvdiff` creates the following variables:

- `_D_L1`, ..., `_D_L`*m* are the lags of the treatment variable, with m equal to # in the `post(`#`)` option.
- `_D_F1`, ..., `_D_F`*p* are the leads of the treatment variable, with p equal to # in the `pre(`#`)` option.

Finally, note that i) the treatment has to be a 0/1 binary variable (`1` = treated, `0` = untreated) and that ii) before running `tvdiff`, one has to install the community-contributed command `coefplot` (Jann 2014).

4.5 Stored results

`tvdiff` stores the following in `e()`:

Scalars

`e(N)`	total number of (used) observations
`e(N1)`	number of (used) treated units
`e(N0)`	number of (used) untreated units
`e(ate)`	value of the (contemporaneous) ATE

5 An application using simulated data in the presence of selection bias

This example shows how to correctly run `tvdiff` and shows how it solves the selection bias that often arises in real causal inference applications.

For this purpose, we design a simulated DGP allowing for a nonzero correlation between the "selection equation" or "treatment equation" (the D-equation) and the "outcome equation" (the y-equation), due to the presence of unobservable selection as captured by an individual specific effect acting as confounder.

Consider the same treatment setting of (1); that is, only one lead and one lag are included. Exclude, without loss of generality, observable confounders $\mathbf{x}$. Below, we show that (1) can be derived from a generalized potential-outcome model made of three treatments, that is, the treatment sequences set out in (2):

$$\begin{aligned} y_{it} = y_{it}(w^1) + D_{it-1}\left\{y_{it}\left(w^2\right) - y_{it}\left(w^1\right)\right\} + D_{it}\left\{y_{it}\left(w^3\right) - y_{it}\left(w^1\right)\right\} \\ + D_{it+1}\left\{y_{it}\left(w^4\right) - y_{it}\left(w^1\right)\right\} \end{aligned} \tag{4}$$

Assume that the potential outcome takes on this form

$$y_{it}(w^j) = \beta_j + u_{it}^j = \beta_j + \varepsilon_{it}^j + c_i \tag{5}$$

where $u_{it}^j = c_i + \varepsilon_{it}^j$. c_i represents the individual fixed effect, ε_{it}^j represents a pure random shock, and β_j is a parameter to be estimated. By substituting (5) into (4), we can see that

$$y_{it} = (\beta_1 + c_i) + D_{it-1}(\beta_2 - \beta_1) + D_{it}(\beta_3 - \beta_1) + D_{it+1}(\beta_4 - \beta_1) + \eta_{it} \tag{6}$$

where $\eta_{it} = \varepsilon_{it}^1 + D_{it-1}(\varepsilon_{it}^2 - \varepsilon_{it}^1) + D_{it}(\varepsilon_{it}^3 - \varepsilon_{it}^1) + D_{it+1}(\varepsilon_{it}^4 - \varepsilon_{it}^1)$.

Equation (6) is equivalent to (1), apart from $\mathbf{x}_{it}$, under the specification $\beta_{-1} = \beta_2 - \beta_1$, $\beta_0 = \beta_3 - \beta_1$, $\beta_{+1} = \beta_4 - \beta_1$, $\mu_{it} = \beta_1 + c_i$, and finally, $u_{it} = \eta_{it}$.

By using the definitions of ATE_{jk}, we can finally rewrite (6) as

$$y_{it} = \beta_1 + c_i + D_{it-1}\text{ATE}_{21} + D_{it}\text{ATE}_{31} + D_{it+1}\text{ATE}_{41} + \eta_{it}$$

We can now perform our simulation experiment showing that `tvdiff` with the option `model(fe)`, unlike the option `model(ols)`, can solve a selection-on-unobservables problem. The DGP is such that the individual effect c_i feeds into both the potential outcomes, through the error term, and the selection equation for D_{it}:

$$\begin{aligned} c_i &\sim \text{normal}(3,1) \\ u_{it}^j &= c_i^3 + \varepsilon_{it}^j \\ j &= 1,2,3,4 \\ \varepsilon_{it} &\sim \text{normal}(0,1) \\ D_{it} &= \mathbf{1}\left(10 \times c_i + \varepsilon_{it} > 0\right) \\ \beta_1 &= 10, \quad \beta_2 = 20, \quad \beta_3 = 30, \quad \beta_4 = 40 \end{aligned}$$

Observe that c_i enters the potential outcomes' random shocks, u_{it}^j, nonlinearly. This choice depends on the fact that D_{it} is modeled as binary and nonlinear within this DGP, thus requiring a nonlinear form of the fixed effect in the potential-outcome equations to produce substantial OLS bias. Below is the code for this DGP:

```
*** Stata code for the DGP ****************************************************
set obs 2000                  // set the number of individuals
set seed 101                  // set seed to obtain the same results
generate c=rnormal(3,1)       // time-invariant individual specific heterogeneity
generate id=_n                // generate individual specific ID
expand 100                    // create 100 observations for each initial
                              // observation
by id, sort: generate time=_n // year of observation
generate D=(rnormal()+10*c>0) // selection equation
generate u1= rnormal()+1*c^3  // correlated error of the potential outcome 1
generate u2= rnormal()+1*c^3  // correlated error of the potential outcome 2
generate u3= rnormal()+1*c^3  // correlated error of the potential outcome 3
generate u4= rnormal()+1*c^3  // correlated error of the potential outcome 4
generate y1=10+u1             // potential outcome 1
generate y2=20+u2             // potential outcome 2
generate y3=30+u3             // potential outcome 3
generate y4=40+u4             // potential outcome 4
tsset id time                 // tsset the data
generate D_L1 = L1.D          // generate one lag of D
generate D_F1=F1.D            // generate one lead of D
// Generate the observable y using the POM
generate y=y1+D_L1*(y2-y1)+D*(y3-y1)+D_F1*(y4-y1)
*******************************************************************************
```

We run `tvdiff` twice, once with the option `model(ols)` and once with `model(fe)`:

```
. ***************************************************************************
. * tvdiff --> OLS estimates
. ***************************************************************************
. tvdiff y D, model(ols) pre(1) post(1) vce(robust)
  (output omitted)
```

y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
_D_F1	64.83851	.5517372	117.52	0.000	63.75712	65.9199
D	56.39676	.6552312	86.07	0.000	55.11252	57.681
_D_L1	44.52236	.9123655	48.80	0.000	42.73415	46.31058
_cons	-60.43861	1.253538	-48.21	0.000	-62.89551	-57.98171

```
  (output omitted)
. ***************************************************************************
. * tvdiff --> fixed-effects estimates
. ***************************************************************************
. tvdiff y D, model(fe) pre(1) post(1) vce(robust)
  (output omitted)
                                  (Std. Err. adjusted for 2,000 clusters in id)
```

y	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
_D_F1	29.5395	.4757364	62.09	0.000	28.60651	30.4725
D	21.09776	.5077086	41.55	0.000	20.10206	22.09345
_D_L1	9.222757	.1645905	56.03	0.000	8.89997	9.545544
_cons	45.45666	.1965579	231.26	0.000	45.07118	45.84214

(output omitted)

As expected, because of the correlation between the outcome equation and the selection equation entailed by this DGP, OLS estimates are severely biased. The OLS coefficient of the lead—expected to be equal to 30—is in fact equal to about 65, and large biases also arise for the contemporaneous and the lagged coefficients (respectively, about 56 instead of 20 and about 44 instead of 10). On the contrary, the fixed-effects estimator performs well, with all the coefficients close to the true coefficients, thus showing that it effectively solves the selection bias underlying this DGP. Introducing exogenous variables within the previous DGP does not change these results.

6 An application to the effect of public education expenditure on income equality

In this section, we provide an application of `tvdiff` to real data. We apply `tvdiff` to measure the effect of public education effort on income equality at the country level. To this end, we use the longitudinal data of Castellacci and Natera (2011) to build `cana.dta`. This dataset is a rich and complete set of 41 indicators for 134 countries observed over the 1980–2008 period for a total of 3,886 country-year observations. Castellacci and Natera's (2011) data are publicly available and allow for detailed cross-country analyses of national systems, growth, and development.

Within these data, public education effort is measured as the (current and capital) total public expenditure on education as a percentage of gross domestic product (GDP) (variable `es12educe`), while income equality is measured as the complement of the Gini index (variable `sc8ginii`). In this application, we consider as controls the following covariates as found in Castellacci and Natera (2011):

- `i3teler`: **Telecommunication Revenue**. Revenue from the provision of telecommunications services such as fixed-line, mobile, and data, % of GDP.
- `i4elecc`: **Electric power consumption**. Production of power plants and combined heat and power plants less transmission, distribution, and transformation losses and own use by heat and power plants.
- `i6telecap`: **Mobile and fixed-line subscribers**. Total telephone subscribers (fixed line plus mobile) per 1000 inhabitants.
- `ec16openi`: **Openness Indicator**. (Import + Export)/GDP. PPP, 2000 USD.
- `sc20trust`: **Most people can be trusted**. Percentage of respondents who "agree" with this statement.
- `ec14credg`: **Domestic Credit by Banking Sector**. Includes all credit to various sectors on a gross basis, with the exception of credit to the central government, which is net, as a share of GDP.
- `pf20demoa`: **Index Democracy and Autocracy**. *Democracy*: political participation is full and competitive, executive recruitment is elective, constraints on the chief executive are substantial. *Autocracy*: it restricts or suppresses political participation. The index ranges from +10 (democratic) to −10 (autocratic).

The binary treatment D_{it} is defined as follows: consider the "within" median of public expenditure in education over GDP, namely, the median by country of the share of public expenditure in education over GDP for the 1980–2008 period. If in year t, country i performs a public expenditure in education larger than its "within" median, then $D_{it} = 1$ (the pair country-year is thus "treated"); otherwise, $D_{it} = 0$. In other words, the treatment is defined as the tendency of a country to boost its expenditure in education in a specific year compared with a baseline reference, measured as its median performance over the overall time span. The outcome y is measured as the "total public expenditure in education as a percentage of GDP".

We do this exercise by running the following code, where `tvdiff` is used with five pretreatment periods and nine posttreatment periods:

```
********************************************************************************
use cana.dta, clear
********************************************************************************
destring _all, replace
********************************************************************************
* TREATMENT
********************************************************************************
global S "es12educe" // public expenditure in education
by Country, sort: egen med_$S=median($S)
capture drop demed$S
generate demed_$S=$S-med_$S
capture drop d$S
generate d$S=.
replace d$S=1 if demed_$S>0 & demed_$S!=.
replace d$S=0 if demed_$S<=0
summarize d$S        // treatment dummy 0,1
global D d$S         // democracy (treatment)
********************************************************************************
* OUTCOME
********************************************************************************
generate equality=100-sc8ginii
global y "equality"  // equality
********************************************************************************
* COVARIATES
********************************************************************************
global x "i3teler i4elecc i6telecap ec16openi sc20trust ec14credg pf20demoa"
********************************************************************************
set scheme s1mono
encode Country, gen(Country_n)
tsset Country_n Year
tvdiff $y $D $x, model(fe) pre(5) post(9) vce(robust) ///
     graph save_graph(mygraph)
********************************************************************************
```

For brevity's sake, the regression outputs are omitted, ensuring that both parallel-trend tests are passed, and we focus on the graph in figure 2. This figure shows that from the time of treatment (that is, higher than the median education expenditure) onward, the ATE, given by the level of equality in income distribution, increases steeply and remains positive until the seventh year after treatment.

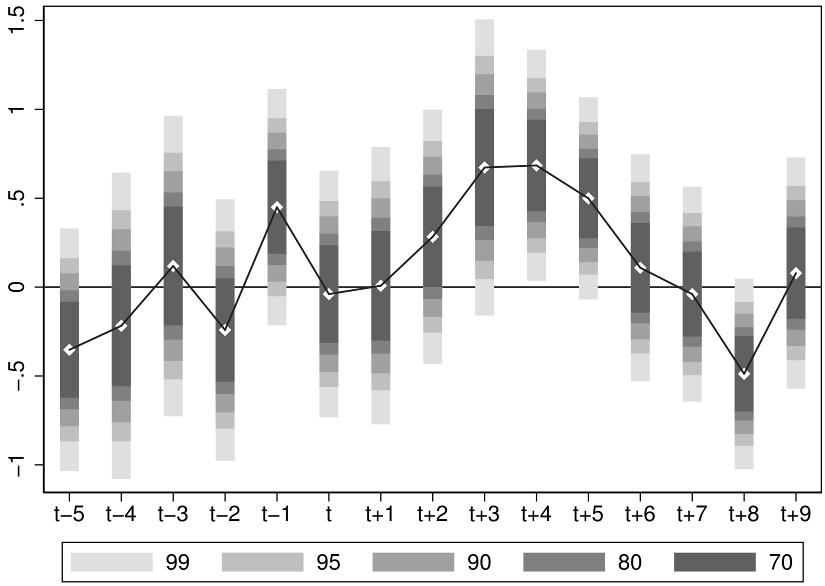

Figure 2. Graph of the pre- and posttreatment pattern for the relation between country investment in public education and income equality

The pattern is a sort of parabola, showing that the effect of one short increase in education expenditure above the median has a transitory effect tending to fade away around seven years after treatment. Considering that significance is relatively high after 3, 4, and 5 years from treatment time (t), this finding shows a quite sensible effect of public investment in education on income equality. More specifically, we see that the (average) equality index difference between treated and untreated reaches a value of around 0.5% three and four years after treatment and then decreases in the subsequent years.

Of course, other possible confounders may be present. However, the use of fixed-effects estimation should mitigate unobservable selection, thus making these results also sufficiently robust to selection on unobservables. This is one of the main strengths of DID that `tvdiff` helps to apply.

7 Conclusion

In this article, we presented `tvdiff`, a community-contributed command for estimating ATEs when the treatment is binary and varying over time. We introduced the econometrics underlying the model fit by `tvdiff` and showed the possibility to graph the estimated effects. Subsequently, we showed the logic of the two parallel trends tests and how to carry them out with `tvdiff`. We finally presented the command's syntax and provided two applications: one on simulated data and one on real data evaluating the effect of public education expenditure on income equality.

Note that one must be cautious when using this command for causal inference because both tests allow for testing only the necessary condition for identification to hold. Hence, if the parallel trend is supported by the tests, the user should validly motivate why the sufficient condition is expected to hold under the specific context of analysis.

We hope readers will find `tvdiff` useful for practical program evaluation in appropriate contexts. We envision further developments to extend this command to multivalued and continuous treatment settings.

8 Acknowledgments

We presented a first draft of this article at the 2017 Italian Stata Users Group meeting held in Florence on November 16–17, 2017. We thank the organizers and all the participants of the meeting for their useful comments.

9 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 19-3
. net install st0566      (to install program files, if available)
. net get st0566          (to install ancillary files, if available)
```

10 References

Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Autor, D. H. 2003. Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics* 21: 1–42.

Castellacci, F., and J. M. Natera. 2011. A new panel dataset for cross-country analyses of national systems, growth and development (CANA). MPRA Paper 28376, University Library of Munich. https://mpra.ub.uni-muenchen.de/28376/1/MPRA_paper_28376.pdf.

Cerulli, G. 2015. *Econometric Evaluation of Socio-Economic Programs: Theory and Applications*. Berlin: Springer.

Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37: 424–438.

Jann, B. 2014. Plotting regression coefficients and other estimates. *Stata Journal* 14: 708–737.

About the authors

Giovanni Cerulli is a researcher at the CNR-IRCrES, Research Institute on Sustainable Economic Growth, National Research Council of Italy, Unit of Rome. His research interest is in applied economics and econometrics, with a special focus on causal inference. His main field of application focuses on measuring the effects of technological policies on firms' performance. Cerulli has developed some original causal inference models, such as dose–response and treatment models with social interaction, and has also provided Stata implementations. He has published his articles in several high quality scientific journals and is currently editor-in-chief of the *International Journal of Computational Economics and Econometrics*.

Marco Ventura is a professor of econometrics at Sapienza, University of Rome. He holds a PhD in economics from the Faculty of Statistics, La Sapienza University of Rome, an MSc in finance from Birkbeck College, University of London, and a degree in economics from the Faculty of Economics, University of Rome Tor Vergata. He has obtained the Italian national scientific qualification in econometrics, applied economics, political economy, and economic policy. His main research interests are evaluation of public programs, R&D investment, innovation, and patent valuation.