The Stata Journal (2016)
16, Number 2, pp. 482–490

Abadie's semiparametric difference-in-differences estimator

Kenneth Houngbedji
Paris School of Economics
Paris, France
kenneth.houngbedji@psemail.eu

Abstract. The difference-in-differences estimator measures the effect of a treatment or policy intervention by comparing change over time of the outcome variable across treatment groups. To interpret the estimate as a causal effect, this strategy requires that, in the absence of the treatment, the outcome variable followed the same trend in treated and untreated groups. This assumption may be implausible if selection for treatment is correlated with characteristics that affect the dynamic of the outcome variable. In this article, I describe the command `asdid`, which implements the semiparametric difference-in-differences (SDID) estimator of Abadie (2005, *Review of Economic Studies* 72: 1–19). The SDID is a reweighing technique that addresses the imbalance of characteristics between treated and untreated groups. Hence, it makes the parallel trend assumption more credible. In addition, the SDID estimator allows the use of covariates to describe how the average effect of the treatment varies for different groups of the treated population.

Keywords: st0442, absdid, semiparametric estimations, difference-in-differences, propensity score

1 The semiparametric difference-in-differences estimator

Let's consider the general setting of studies of causal effects used by Rosenbaum and Rubin (1983). We want to estimate the causal effect of a treatment on a variable of interest $\mathbf{y}$ at some time t. Each participant has two potential outcomes: $\mathbf{y}_{1t}$ and $\mathbf{y}_{0t}$. $\mathbf{y}_{1t}$ is the value of $\mathbf{y}$ if the participant received the treatment by time t. $\mathbf{y}_{0t}$ is the value of $\mathbf{y}$ if the participant had not received the treatment by time t. $\mathbf{d}$ is an indicator of whether or not a participant was treated by time t. At time $t = 0$, which is the baseline b, no one is treated. At time $t \neq 0$, $\mathbf{d}$ is equal to 1 for a treated participant and is equal to 0 otherwise. We want to estimate the average treatment effect on the treated (ATT):

$$\text{ATT} \equiv \mathbb{E}\left(\mathbf{y}_{1t} - \mathbf{y}_{0t} \,\middle|\, \mathbf{d} = 1\right)$$

Because $\mathbf{y}_{0t}$ is never observed for a treated participant, the ATT cannot be directly estimated. Assume $\mathbf{y}_{0b}$ is the value of $\mathbf{y}$ at time $t = 0$—that is, the baseline. $\mathbf{x}_b$ is a set of pretreatment characteristics, $\Delta\mathbf{y}_t \equiv \mathbf{y}_t - \mathbf{y}_b$ is the change of $\mathbf{y}$ between time t and the baseline b, and $\pi(\mathbf{x}_b) \equiv \mathbb{P}\left(\mathbf{d} = 1 \,\middle|\, \mathbf{x}_b\right)$ is the conditional probability to be in the treatment group (also called the propensity score). Abadie (2005) shows that the sample analog of

$$\mathbb{E}\left\{\frac{\Delta \mathbf{y}_{t}}{\mathbb{P}(\mathbf{d}=1)} \times \frac{\mathbf{d}-\pi(\mathbf{x}_{b})}{1-\pi(\mathbf{x}_{b})}\right\} \tag{1}$$

gives an unbiased estimate of the ATT if (2) and (3) hold.

$$\mathbb{E}\left(\mathbf{y}_{0t}-\mathbf{y}_{0b}\,\middle|\,\mathbf{d}=1\,,\,\mathbf{x}_{b}\right)=\mathbb{E}\left(\mathbf{y}_{0t}-\mathbf{y}_{0b}\,\middle|\,\mathbf{d}=0\,,\,\mathbf{x}_{b}\right) \tag{2}$$

$$\mathbb{P}(\mathbf{d}=1)>0 \quad \text{and} \quad \pi(\mathbf{x}_{b})<1 \tag{3}$$

The estimator is a weighted average of the difference of trend—$\Delta \mathbf{y}_{t}$—across treatment groups. It proceeds by reweighing the trend for the untreated participants based on their propensity score $\pi(\mathbf{x}_{b})$. Because $\{\pi(\mathbf{x}_{b})\}/\{1-\pi(\mathbf{x}_{b})\}$ is an increasing function of $\pi(\mathbf{x}_{b})$, untreated participants with a higher propensity score are given a higher weight.

Abadie (2005) suggests to approximate the propensity score $\pi(\mathbf{x}_{b})$ semiparametrically using a polynomial series of the predictors. Thereafter, the values predicted are plugged into the sample analogue of (1). Even though the approximation improves for higher polynomial order, the estimation becomes less precise. It is also possible to estimate $\pi(\mathbf{x}_{b})$ with the series logit estimator (SLE) (see Hirano, Imbens, and Ridder [2003]). This method uses a logit specification to constrain the estimated propensity score to vary between 0 and 1.

Consider, for instance, that $\widehat{\pi}(\mathbf{x}_{b})$ is the approximated propensity score and k is the order the polynomial function used to approximate $\pi(\mathbf{x}_{b})$. The approximation of $\pi(\mathbf{x}_{b})$ produced by the linear probability model (LPM) can be written as

$$\widehat{\pi}(\mathbf{x}_{b})=\widehat{\gamma}_{0}+\widehat{\gamma}_{1}\times\mathbf{x}_{1}+\sum_{i=1}^{\mathrm{k}}\widehat{\gamma}_{2i}\times\mathbf{x}_{2}^{i}$$

where $\mathbf{x}_{1}$ is a binary variable, $\mathbf{x}_{2}^{i}=\prod_{j=1}^{i}\mathbf{x}_{2}$, and $\mathbf{x}_{2}$ is a continuous variable. The coefficients $\widehat{\gamma}_{0},\widehat{\gamma}_{1},\widehat{\gamma}_{21},\ldots,\widehat{\gamma}_{2i},\ldots,\widehat{\gamma}_{2k}$ are estimated using an ordinary least-squares estimator.

With an SLE estimator approach, the propensity score $\pi(\mathbf{x}_{b})$ is estimated as follows:

$$\widehat{\pi}(\mathbf{x}_{b})=\Lambda\left(\widehat{\gamma}_{0}+\widehat{\gamma}_{1}\times\mathbf{x}_{1}+\sum_{k=1}^{\mathrm{K}}\widehat{\gamma}_{2k}\times\mathbf{x}_{2}^{k}\right)$$

where $\Lambda(x)=\exp(x)/\{1+\exp(x)\}$ is the logistic function. Higher order binary variables—like $\mathbf{x}_{1}$—are not considered because $\mathbf{x}_{1}^{k}=\mathbf{x}_{1}$ for any value $k>1$.

Independently of the approximation method used, the errors related to the estimation of the propensity scores are considered when estimating the standard error of the ATT as described in Abadie (2005). Other estimators use the propensity score to estimate the ATT. The kernel matching and nearest neighbor matching estimators are among the most widely used estimators for quasi experimental identification. However, both estimators assume that the propensity score is given, not estimated, and produce on average estimates with smaller standard errors than the estimator of Abadie.

2 The absdid command

The `absdid` command is the Stata equivalent of a MATLAB code written by Abadie (2005) in an empirical application of the semiparametric difference-in-differences (SDID) estimator.[1] `absdid` estimates the ATT by comparing change over time of the outcome of interest across treatment groups while adjusting for differences between treatment groups on the observable characteristics at baseline that are correlated to the propensity score.

2.1 Syntax

`absdid` *depvar* [*if*] [*in*], `tvar(`*varname*`)` `xvar(`*varlist*`)` [`yxvar(`*varlist*`)` `order(`#`)` `sle` `csinf(`#`)` `csup(`#`)`]

depvar is a variable that represents the change of the outcome of interest between baseline and post treatment for each observation.

2.2 Options

`tvar(`*varname*`)` is the binary treatment variable. It takes the value `1` when the observation is treated and takes the value `0` otherwise. `tvar()` is required.

`xvar(`*varlist*`)` are the control variables. They can be either continuous or binary and are used to estimate the propensity score. `xvar()` is required.

`yxvar(`*varlist*`)` is a list of variables that can modify the treatment effect. By default, the treatment effect is assumed to be constant.

`order(`#`)` represents the order of the polynomial function used to estimate the propensity score. It takes integer values and the default is `order(1)`.

`sle` forces the use of a logistic specification to estimate the propensity score (see Hirano, Imbens, and Ridder [2003]). This ensures, for instance, that the estimated propensity score is always greater than 0 and less than 1. By default, the propensity score is estimated with a linear regression.

`csinf(`#`)` drops the observations of which the propensity score is less than #. The default is `csinf(0)`.

`csup(`#`)` drops the observations of which the propensity score is greater than #. The default is `csup(1)`.

1. The original code is tailored to measure the effect of union membership on wages for workers. It is available at http://www.hks.harvard.edu/fs/aabadie/cdid_union.m.

3 Example

To illustrate how `absdid` works, let's reproduce the application exercise available on Abadie's website. We estimate the effect of participation in a worker union on wages of unionized female workers. The data used are an excerpt of the current population survey—a U.S. government monthly survey of unemployment and labor force participation. The data consist of female workers observed in 1996 and resurveyed in 1997 (see table 1). The workers were not unionized in 1996, so we can identify the union–wage effect on the workers who joined a worker union between 1996 and 1997.

Let $w_{1,97}$ be the wage of a worker in 1997 if she joined a worker union, and let $w_{0,97}$ be the wage if she had not joined a union. Because wage variations are traditionally modeled through a lognormal distribution, the parameter of interest is as follows:

$$\text{ATT}\left\{\log(w)\right\} \equiv \mathbb{E}\left\{\log(w_{1,97}) - \log(w_{0,97}) \,\middle|\, \mathbf{union}_{97} = 1\right\}$$

For simplicity, we report estimates of $\text{ATT}\{\log(w)\}$ and interpret the results as the percentage effect of worker union on wage.[2]

If female workers were randomly selected to join a union in 1997, one could estimate $\text{ATT}\{\log(w)\}$ by comparing the log of wages of unionized and nonunionized workers in 1997. To account for the female workers who joined a union in 1997 differing from those who remained nonunionized with respect to age, education level, and race—see table 1—we use an SDID approach.

Assume that, in the absence of worker unions, the wage dynamics of unionized workers would have been similar to that of nonunionized workers with the same age, education level, race, state of residence, and sector of activity. If that assumption holds, we can use the `absdid` command to compute the SDID estimator of the union–wage effect for female workers.

2. Actually, a more accurate estimate of the percentage effect of worker union on wage can be obtained using the transformation suggested by Kennedy (1981).

Table 1. Characteristics of female workers across treatment groups

Variables	Entire sample	Unionized	Non-unionized	Diff.
Union coverage in 1997	0.05			
	[0.22]			
Wage variables:				
Log wage in 1997	2.36	2.43	2.36	0.07 ***
	[0.52]	[0.49]	[0.53]	(0.02)
Log wage in 1996	2.30	2.34	2.30	0.04 **
	[0.54]	[0.52]	[0.54]	(0.02)
Covariates in 1996:				
Age (years)	39.33	40.37	39.27	1.09 ***
	[11.01]	[10.55]	[11.03]	(0.37)
High school	0.93	0.92	0.93	−0.01
	[0.26]	[0.27]	[0.26]	(0.01)
College	0.25	0.35	0.24	0.10 ***
	[0.43]	[0.48]	[0.43]	(0.01)
African American	0.10	0.19	0.09	0.10 ***
	[0.29]	[0.39]	[0.29]	(0.01)
Hispanic	0.06	0.07	0.06	0.01
	[0.24]	[0.26]	[0.24]	(0.01)
Married	0.63	0.63	0.63	−0.00
	[0.48]	[0.48]	[0.48]	(0.02)
Number of workers	18,470	958	17,512	18,470

Notes: Standard deviations are in brackets. Standard errors are in parentheses. Significance levels are denoted as follows: * p<0.10, ** p<0.05, and *** p<0.01.

First, we need a variable (`dlwage`) that, as suggested in (1), measures the change of log wage between baseline and follow-up. Second, we need a binary variable (`union97`) that indicates treated and untreated observations. Third, we need a list of control variables among which unionized and nonunionized workers differ from one another; let's consider the variables `age`, `black`, `hispanic`, and `grade`, which report the age, ethnic background, and education level of the workers in 1996. With these inputs, we can estimate the SDID estimator of the union–wage effect for female workers:

```
. absdid dlwage, tvar(union97) xvar(age black hispanic married grade)
Abadie´s semi-parametric diff-in-diff           Number of obs     =      18469
```

dlwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATT						
_cons	.0361469	.0163367	2.21	0.027	.0041276	.0681663

`Number of obs` shows the number of observations used for the estimation that satisfy (3), that is, the observations for which the estimated propensity score is bigger than 0 and smaller than 1. Though the sample has 18,470 observations, only 18,469 are used to estimate the ATT. This suggests that 1 observation has an estimated propensity score that either is smaller than or equal to 0 or is bigger than or equal to 1. This is not surprising because, by default, `absdid` uses a linear regression to estimate the propensity score; hence, the predicted values often can be either negative or bigger than 1. To avoid any loss of information, we can add the `sle` option.[3]

```
. absdid dlwage, tvar(union97) xvar(age black hispanic married grade) sle
Abadie´s semi-parametric diff-in-diff           Number of obs     =      18470
```

dlwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATT						
_cons	.0364533	.0163435	2.23	0.026	.0044207	.0684859

To discard observations with very small or very large propensity scores, we can use the `csinf` and `csup` options to indicate the lowest and highest acceptable values of the propensity score. In the example below, we restrict the estimation of the ATT to female workers whose propensity score is between 0.01 and 0.99.

```
. absdid dlwage, tvar(union97) xvar(age black hispanic married grade)
> csinf(0.01) csup(0.99)
Abadie´s semi-parametric diff-in-diff           Number of obs     =      18447
```

dlwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATT						
_cons	.0362135	.0163374	2.22	0.027	.0041928	.0682343

Independently of the method used to estimate the propensity score, the outputs of `absdid` show a point estimate of the ATT when the union–wage premium is constant and does not vary with worker characteristics. Overall, the results suggest that joining a worker union increased the wage of female workers by 3.6% in 1997. The effect does not vary with the option `sle`.

3. When `sle` is used, some observations can still be left out of the propensity score estimation when there is perfect prediction. This is the case, for instance, when all the workers in a given industry are either unionized or nonunionized. In those cases, the ATT is estimated only for the observations for which the treatment status is not perfectly predicted by observed characteristics.

Similarly, we can also consider that the effect of being in a union on wage varies with worker characteristics. For instance, the union–wage premium may vary with the age of the worker. Experienced workers—based on age—are often scarce in the economy. As such, they have more bargaining power and may not need to join a worker union to negotiate their wage. Hence, we may expect the union–wage premium to decrease with the age of the worker. Likewise, the union–wage premium may also vary with education level. Workers who have not completed high school should expect a higher premium compared with similar workers who have completed either high school or college. We see below the command for estimating how the union premium for female workers varies with age and education level.

```
. absdid dlwage, tvar(union97) xvar(age black hispanic married grade)
> yxvar(age hschool college) sle
Abadie´s semi-parametric diff-in-diff           Number of obs     =      18470
```

dlwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATT						
age	-.0036144	.0016146	-2.24	0.025	-.0067789	-.0004499
hschool	-.3099432	.1043214	-2.97	0.003	-.5144095	-.105477
college	.0562573	.0374896	1.50	0.133	-.0172211	.1297356
_cons	.4582553	.1356808	3.38	0.001	.1923259	.7241847

As expected, the results indicate that union premium decreases with age and education level. Considering that the average female worker of the sample was 39 years old in 1996, joining a worker union should increase the wage of the average female worker by 31.8% (that is, $0.458 - 39 \times 0.0036 = 0.3176$). In contrast, the premium is estimated at 16.1% for a worker who was 50 years old in 1996. Likewise, compared with workers who have no diploma in 1996, the union premium decreases by 31% for workers whose highest diploma is high school. Surprisingly, there is no statistically significant difference between the union premium of workers with a college diploma and those with no diploma. This is likely because of the small sample size: only 7.3% of female workers with a college diploma joined a union between 1996 and 1997.

To reproduce the results from table II of the empirical illustration available from Abadie's website, we need to consider other control variables that may affect the propensity score. We also need to increase the order of the polynomial function used to estimate the propensity score.

First, Abadie considers a larger list of control variables, including age, ethnic group, and fixed effects for education level, state of residence, sector of activity, and date of interview. Let's call this list `cvars` and save it in a macro:

```
. local cvars age black hispanic married i.grade i.state i.dind i.month
```

Second, Abadie uses a polynomial function of order 4 to estimate the propensity score. Using the control variables listed above and using 4 as the order of the polynomial function, we reproduce the results shown on Abadie's website for female workers:

```
. absdid dlwage, tvar(union97) xvar(`cvars´) order(4)
Abadie´s semi-parametric diff-in-diff           Number of obs     =      16374
```

dlwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATT						
_cons	.0327631	.0159989	2.05	0.041	.0014058	.0641203

```
. absdid dlwage, tvar(union97) xvar(`cvars´) yxvar(age hschool college) order(4)
Abadie´s semi-parametric diff-in-diff           Number of obs     =      16374
```

dlwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATT						
age	-.0031764	.001577	-2.01	0.044	-.0062673	-.0000856
hschool	-.1505565	.0648411	-2.32	0.020	-.2776427	-.0234703
college	.0388147	.0349236	1.11	0.266	-.0296343	.1072637
_cons	.2865646	.0955502	3.00	0.003	.0992897	.4738394

Those results are presented in columns (1) and (2) of table 2. They are similar to the union–wage premium for female workers found by Abadie in his empirical exercise.

Table 2. Effects of worker union on log of wage of female workers

Union premium (ATT)	LPM		SLE	
	(1)	(2)	(3)	(4)
Constant	0.0328 **	0.2866 ***	0.0399 **	0.3426 ***
	(0.0160)	(0.0956)	(0.0168)	(0.1082)
Age (years)		−0.0032 **		−0.0036 **
		(0.0016)		(0.0017)
High school		−0.1506 **		−0.1869 ***
		(0.0648)		(0.0724)
College		0.0388		0.0422
		(0.0349)		(0.0361)
Number of workers	16,374	16,374	18,273	18,273

Notes: Models (1) and (3) report estimates of the average union premium for unionized workers. Models (2) and (4) show how the union premium varies with worker age and education level. The average union premiums reported in (1) and (2) are estimated using a linear polynomial function of degree 4 to approximate the propensity score. The premiums reported in (3) and (4) are estimated using a logit specification of degree 4 to estimate the propensity score. Standard errors are in parentheses. Significance levels are denoted as follows: * p<0.10, ** p<0.05, and *** p<0.01.

4 Discussion

For a given set of control variables and predictors, the SDID estimates vary with the type of approximation used—`sle` or simple LPM (the default)—and the order of the polynomial approximation used—`order(#)`. To reduce the margin for arbitrage, one could use a cross validation technique to decide the combination of methods that best suits the semiparametric approximation of the propensity score. It can also help to consider that the LPM is likely to produce estimates of the propensity score that are either negative or greater than 1. When the SLE approximation is used, the observations for which the treatment status is perfectly predicted by a control variable are discarded from the estimation. In most cases, however, the sample size used to estimate the ATT is larger when the propensity score is approximated with the `sle` option.

Using our latest example as benchmark, table 2 shows how our estimates of the union premium for unionized workers vary depending on the type of approximation used.

To conclude, the SDID approach is mostly suited for longitudinal surveys with a baseline and follow-up rounds. To use `absdid`, the user needs to have a measure of the change of the main outcome variable over time for each observation along with treatment status and baseline characteristics.

5 References

Abadie, A. 2005. Semiparametric difference-in-differences estimators. *Review of Economic Studies* 72: 1–19.

Hirano, K., G. W. Imbens, and G. Ridder. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71: 1161–1189.

Kennedy, P. 1981. Estimation with correctly interpreted dummy variables in semilogarithmic equations. *American Economic Review* 71: 801.

Rosenbaum, P. R., and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41–55.

About the author

Kenneth Houngbedji is a researcher at the Paris School of Economics. His main research interests are studies of economic behavior and decision-making processes of households in developing countries to help design better public policies.