A Simple Regression Model for the Policy Effect Identification Using Alternative Diff-in-Diff Assumptions

Ricardo Mora and Iliana Reggio

Universidad Carlos III de Madrid

2012 Spanish Stata Users Group meeting

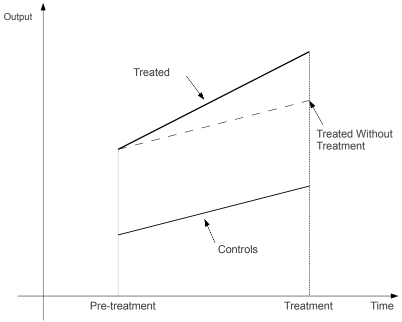
Outline

1. Introduction
2. The DID Model and the Parallel Paths Assumption
3. An alternative Assumption
4. A Look at Current Practice
5. Conclusions

The Parallel-Paths Assumption

- Difference-In-Differences (DID) estimators are widely used in economics to evaluate the impact of a policy
- The crucial assumption is referred to as the "Parallel-Paths assumption"

Without treatment, the average change for the treated would have been equal to the observed average change for the controls



Alternative Methods

- It is accepted that the Parallel-Path assumption is strong
- Several authors have analyzed the validity of DID assumptions and provided new methods and tests
  - Angrist and Krueger (1999) argue that it is essential to validate that trends did not differ before treatment
  - Athey and Imbens (2006) and Bonhomme and Sauder (2011) generalize the approach and identify the entire counter-factual distribution of potential outcomes
  - Donald and Lang (2007) and Bertrand et al. (2004) address problems with standard methods for computing standard errors
  - Abadie (2006) and Blundell et al. (2004) discuss adjusting for exogenous covariates using propensity score methods

Our Proposal

- For applications in which more than one pre-treatment periods are available, we propose a simple regression model in which
  - a set of estimators based on alternative DID trend assumptions can be easily computed
  - it is possible to test the validity of some assumptions and the equivalence of results
- We provide an evaluation of how relevant the alternative assumptions are by applying the method to data from several recent papers
  - results and their significance vary depending on the assumption actually used
  - sometimes, the identifying assumption is not clearly stated
  - even more, sometimes authors wrongly claim to be relying on one assumption but they actually assume a different one when they perform the estimation

The Basic Model with Only Three Periods

- Assume that we have information for two periods before treatment $(t = 1, 2)$ and one period after treatment $(t = 3)$

The Basic Model with Only Three Periods

- Assume that we have information for two periods before treatment $(t = 1, 2)$ and one period after treatment $(t = 3)$
- $Y_0(t)$: outcome in period t under no treatment
- $Y_1(t)$: outcome in period t under treatment
- $D = \begin{cases} 1 & \text{if individual is treated} \\ 0 & \text{if individual is a control} \end{cases}$

The Basic Model with Only Three Periods

- Assume that we have information for two periods before treatment $(t = 1, 2)$ and one period after treatment $(t = 3)$
- $Y_0(t)$: outcome in period t under no treatment
- $Y_1(t)$: outcome in period t under treatment
- $D = \begin{cases} 1 & \text{if individual is treated} \\ 0 & \text{if individual is a control} \end{cases}$
- Observed outcome: $Y(t) = Y_1(t)D + Y_0(t)(1 - D)$

The Basic Model with Only Three Periods

- Assume that we have information for two periods before treatment ($t = 1, 2$) and one period after treatment ($t = 3$)
- $Y_0(t)$: outcome in period t under no treatment
- $Y_1(t)$: outcome in period t under treatment
- $D = \begin{cases} 1 & \text{if individual is treated} \\ 0 & \text{if individual is a control} \end{cases}$
- Observed outcome: $Y(t) = Y_1(t)D + Y_0(t)(1 - D)$

$$\alpha_{ATT} = E[Y_1(3) - Y_0(3) \,|D = 1]$$

The Basic Model with Only Three Periods

- Assume that we have information for two periods before treatment $(t = 1, 2)$ and one period after treatment $(t = 3)$
- $Y_0(t)$: outcome in period t under no treatment
- $Y_1(t)$: outcome in period t under treatment
- $D = \begin{cases} 1 & \text{if individual is treated} \\ 0 & \text{if individual is a control} \end{cases}$
- Observed outcome: $Y(t) = Y_1(t)D + Y_0(t)(1 - D)$

$$\alpha_{ATT} = E[Y_1(3) - Y_0(3)\,|D = 1]$$

As $Y_0(3)$ is not observable for the treated, the identification strategy is to estimate $E[Y_0(3)\,|D = 1]$ using information from the sample of controls

The Parallel-Paths Assumption

Let Δ denote the first difference operator

The Parallel-Paths Assumption

Let Δ denote the first difference operator

(Conditional) Parallel-Paths

$$E[\Delta Y_0(1)\,|X, D = 1] = E[\Delta Y_0(1)\,|X, D = 0]$$

where X is a vector of covariates

The Parallel-Paths Assumption

Let Δ denote the first difference operator

(Conditional) Parallel-Paths

$$E[\Delta Y_0(1)\,|X, D=1] = E[\Delta Y_0(1)\,|X, D=0]$$

where X is a vector of covariates

- Essentially, PP states that the average change in output among the treated if untreated is equal to the observable average change among comparable controls

The Parallel-Paths Assumption

Let Δ denote the first difference operator

(Conditional) Parallel-Paths

$$E[\Delta Y_0(1)\,|X, D=1] = E[\Delta Y_0(1)\,|X, D=0]$$

where X is a vector of covariates

- Essentially, PP states that the average change in output among the treated if untreated is equal to the observable average change among comparable controls

Under Parallel-Paths

$$E[Y_0(3)\,|X, D=1] = E[Y(2)\,|X, D=1] + E[\Delta Y(3)\,|X, D=0]$$

The Parallel-Paths Assumption

Let Δ denote the first difference operator

(Conditional) Parallel-Paths

$$E[\Delta Y_0(1)\,|X, D=1] = E[\Delta Y_0(1)\,|X, D=0]$$

where X is a vector of covariates

- Essentially, PP states that the average change in output among the treated if untreated is equal to the observable average change among comparable controls

Under Parallel-Paths

$$E[Y_0(3)\,|X, D=1] = E[Y(2)\,|X, D=1] + E[\Delta Y(3)\,|X, D=0]$$

$$\alpha_{ATT}(X) = E[\Delta Y(3)\,|X, D=1] - E[\Delta Y(3)\,|X, D=0]$$

The DID Estimator Using Linear Regression

- Choosing $t = 1$ as the reference period, assuming linearity we have that

$$E[Y(t)\,|X, D] = \gamma + \gamma^D D + \gamma_2 I_2 + \gamma_3 I_3 + \gamma_2^D D I_2 + \gamma_3^D D I_3 + \beta X$$

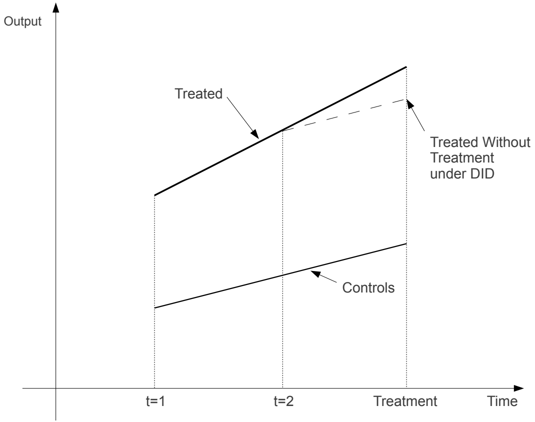
where I_t is period t dummy

- Although this approach is robust to different pre-treatment time trends, the policy effect is still identified by PP:

$$\alpha_{ATT} = \gamma_3^D - \gamma_2^D \equiv \Delta\gamma_3^D$$

Parallel Growths

Consider the situation whereby controls and treated have different but constant trends before and after treatment



Parallel Growth vs PP

- With no change in trends under no-treatment, a correct assumption is

Parallel Growths

$$E[\Delta^2 Y_0(3)\,|X, D=1] = E[\Delta^2 Y_0(3)\,|X, D=0]$$

where $\Delta^2 Y_0(t) = \Delta Y_0(t) - \Delta Y_0(t-1)$.

Parallel Growth vs PP

- With no change in trends under no-treatment, a correct assumption is

Parallel Growths

$$E[\Delta^2 Y_0(3)\,|X, D=1] = E[\Delta^2 Y_0(3)\,|X, D=0]$$

where $\Delta^2 Y_0(t) = \Delta Y_0(t) - \Delta Y_0(t-1)$.

- PG is just PP on output changes

Parallel Growth vs PP

- With no change in trends under no-treatment, a correct assumption is

Parallel Growths

$$E[\Delta^2 Y_0(3)\,|X, D = 1] = E[\Delta^2 Y_0(3)\,|X, D = 0]$$

where $\Delta^2 Y_0(t) = \Delta Y_0(t) - \Delta Y_0(t-1)$.

- PG is just PP on output changes

$$E[Y_0(3)\,|X, D = 1] =$$
$$E[Y(2)\,|X, D = 1] + E[\Delta Y(2)\,|X, D = 1] + E[\Delta^2 Y(3)\,|X, D = 0]$$

- Under PG the counter-factual trend is the previous period growth plus the acceleration experienced by the controls

Parallel Growth vs PP

- With no change in trends under no-treatment, a correct assumption is

Parallel Growths

$$E[\Delta^2 Y_0(3)\,|X, D = 1] = E[\Delta^2 Y_0(3)\,|X, D = 0]$$

where $\Delta^2 Y_0(t) = \Delta Y_0(t) - \Delta Y_0(t-1)$.

- PG is just PP on output changes

$$E[Y_0(3)\,|X, D = 1] =$$
$$E[Y(2)\,|X, D = 1] + E[\Delta Y(2)\,|X, D = 1] + E[\Delta^2 Y(3)\,|X, D = 0]$$

- Under PG the counter-factual trend is the previous period growth plus the acceleration experienced by the controls
- Hence, PG allows for differing trends before and also after treatment while PP only allows for different trends before treatment

Estimation Under PG

The natural estimator for α_{ATT} under PG is not the DID estimator

Estimation Under PG

The natural estimator for α_{ATT} under PG is not the DID estimator

Under PG, α_{ATT} equals the "diff-in-double-diff operator", d2d,

$$\alpha_{ATT}(X) = E[\Delta^2 Y(3)\,|X, D=1] - E[\Delta^2 Y(3)\,|X, D=0] \equiv \alpha^{d2d}_{ATT}(X)$$

Estimation Under PG

The natural estimator for α_{ATT} under PG is not the DID estimator

Under PG, α_{ATT} equals the "diff-in-double-diff operator", d2d,

$$\alpha_{ATT}(X) = E[\Delta^2 Y(3)\,|X, D=1] - E[\Delta^2 Y(3)\,|X, D=0] \equiv \alpha_{ATT}^{d2d}(X)$$

Moreover,

$$\alpha_{ATT}^{d2d}(X) = \alpha_{ATT}^{DID}(X) \iff E[\Delta Y(2)\,|X, D=1] = E[\Delta Y(2)\,|X, D=0]$$

Estimation Under PG

The natural estimator for α_{ATT} under PG is not the DID estimator

Under PG, α_{ATT} equals the "diff-in-double-diff operator", d2d,

$$\alpha_{ATT}(X) = E[\Delta^2 Y(3)\,|X, D=1] - E[\Delta^2 Y(3)\,|X, D=0] \equiv \alpha_{ATT}^{d2d}(X)$$

Moreover,

$$\alpha_{ATT}^{d2d}(X) = \alpha_{ATT}^{DID}(X) \iff E[\Delta Y(2)\,|X, D=1] = E[\Delta Y(2)\,|X, D=0]$$

- Thus, in the presence of pre-treatment differing trends, one of the two estimators must be inconsistent

Regression Techniques

Under linearity, regression techniques can also be used to directly obtain $\hat{\alpha}^{d2d}_{ATT}$ and its standard error

Regression Techniques

Under linearity, regression techniques can also be used to directly obtain $\hat{\alpha}_{ATT}^{d2d}$ and its standard error

- Consider the simple model

$$Y(t) = \gamma + \gamma^D D + \gamma_2 I_2 + \gamma_3 I_3 + \gamma_2^D D I_2 + \gamma_3^D D I_3 + \beta X + u(t)$$

where $E[u(t)\,|X, D] = 0$

Regression Techniques

Under linearity, regression techniques can also be used to directly obtain $\hat{\alpha}_{ATT}^{d2d}$ and its standard error

- Consider the simple model

$$Y(t) = \gamma + \gamma^D D + \gamma_2 I_2 + \gamma_3 I_3 + \gamma_2^D DI_2 + \gamma_3^D DI_3 + \beta X + u(t)$$

where $E[u(t)\,|X, D] = 0$

- Under PP, $\alpha_{ATT} = \gamma_3^D - \gamma_2^D$
- Under PG, $\alpha_{ATT} = \gamma_3^D - 2\gamma_2^D$

Regression Techniques

Under linearity, regression techniques can also be used to directly obtain $\hat{\alpha}_{ATT}^{d2d}$ and its standard error

- Consider the simple model

$$Y(t) = \gamma + \gamma^D D + \gamma_2 I_2 + \gamma_3 I_3 + \gamma_2^D D I_2 + \gamma_3^D D I_3 + \beta X + u(t)$$

where $E[u(t)\,|X, D] = 0$

- Under PP, $\alpha_{ATT} = \gamma_3^D - \gamma_2^D$
- Under PG, $\alpha_{ATT} = \gamma_3^D - 2\gamma_2^D$

- Note that they are equal if and only if $\gamma_2^D = 0$

A General Framework with Many Periods

Consider the case whereby we have information on the outcome in $T-1$ periods before treatment and one period under treatment

A General Framework with Many Periods

Consider the case whereby we have information on the outcome in $T-1$ periods before treatment and one period under treatment

- Using linearity, and choosing $t=1$ as the reference period, we have that

$$E[Y(t)\,|X,D] = \gamma + \gamma^D D + \sum_{\tau=2}^{T}\left[\gamma_\tau + \gamma_\tau^D D\right] I_\tau + \sum_{\tau=1}^{T}\beta_{x(\tau)}X(\tau)I_\tau$$

A General Framework with Many Periods

Consider the case whereby we have information on the outcome in $T-1$ periods before treatment and one period under treatment

- Using linearity, and choosing $t=1$ as the reference period, we have that

$$E[Y(t)\,|X,D] = \gamma + \gamma^{D}D + \sum_{\tau=2}^{T}\left[\gamma_{\tau} + \gamma_{\tau}^{D}D\right] I_{\tau} + \sum_{\tau=1}^{T}\beta_{x(\tau)}X(\tau)I_{\tau}$$

$$\alpha_{ATT}^{d1d} = \gamma_{T}^{D} - \gamma_{T-1}^{D} \equiv \Delta\gamma_{T}^{D}$$

$$\alpha_{ATT}^{d2d} = \gamma_{T}^{D} - 2\gamma_{T-1}^{D} + \gamma_{T-2}^{D} \equiv \Delta^{2}\gamma_{T}^{D}$$

A General DqD Assumption

Parallel-q

$$E[\Delta^q Y_0(T)\,|D=1] = E[\Delta^q Y_0(T)\,|D=0], \qquad q<T$$

- The dqd operator is defined as

$$\alpha_{ATT}^{dqd} \equiv E[\Delta^q Y(T)\,|D=1] - E[\Delta^q Y(T)\,|D=0]$$

Under Parallel-q

$$\alpha_{ATT} = \alpha_{ATT}^{dqd} = \Delta^q \gamma_T^D$$

$$\alpha_{ATT}^{d(q-1)d} = \alpha_{ATT}^{dqd} \iff$$
$$E[\Delta^{q-1} Y(T-1)\,|D=1] - E[\Delta^{q-1} Y(T-1)\,|D=0]$$

DqD vs. D(q-1)D Estimators

$$\hat{\alpha}^{dqd}_{ATT} = \Delta^q \hat{\gamma}^D_T$$

- Under P(q) and general conditions, $\hat{\alpha}^{dqd}_{ATT}$ will be consistent and asymptotically normal.
- With our simple specification, we can easily:
  - obtain $\hat{\alpha}^{dqd}_{ATT}$ and its standard errors for every possible q
  - test how different they are
  - test for pre-treatment trends
  - the approach is generalized to the situation whereby there are many periods before treatment, many periods with effects similar as at treatment, and many periods after treatment

A Look at Current Practice

- In this Section, we provide an evaluation of how relevant the alternative Parallel-q assumptions are by applying the methods to data from several recent papers
- We look for papers which satisfy the following conditions:
  - There is an application of DID
  - The sample includes more than one period before treatment
  - Data is made available
  - Paper is published in the period $2009:2012$ in one of the following 10 Economics journals: AEJ:AE, AER, JAppEcon, JEcon, JEEA, JLabEc, JPE, QJE, REcoStat, REconStud
- We program the estimation of the model and the specification tests using Stata

A Typical Stata Output Using dqd

```
. dqd insch `lista' if sample_A & blackrural, treated(D_A) time(census) begin(2) end(3) cluster(cntyid)

DqD Policy Evaluation

    Output: insch
    Sample Period:    0:3
    First Period of Treatment:    2
    Last Period of Only-Treatment:    3

    Panel A:     Common Trends          Estimated Policy Effects with Time Dummies
                 H0: DqD(t0)=0                t0+1                               t0+2
----------------------------------------------------------------------------------------------------
    D1D          -.0171305                .0681937                           .0654742
                  (0.0245)               (0.0293)                           (0.0215)
    D2D              0                    .0853242                           .0144109
                  (0.0000)               (0.0444)                           (0.0363)
----------------------------------------------------------------------------------------------------

    Panel B: Equivalence Tests
                                              t0+1                               t0+2
----------------------------------------------------------------------------------------------------
 H0: D1D = D2D                           -.0171305                           .0510632
                                         (0.0245)                           (0.0307)
----------------------------------------------------------------------------------------------------
 Clustered Standard Errors in parenthesis
```

Selected Papers

Selected Papers

Paper	Year	Journal	Issue
Aaronson & Mazumber	2011	JPE	Did Rosenwald rural schools improve educational gains of rural southern blacks
Abramitzky et al.	2011	AEJ:AE	Does local male abundance lead to men marrying women of lower social classes?
Currie & Walker	2011	AEJ:AE	Does E-ZPass affect pollution and infant death?
De Jong et al.	2011	JEEA	Does screening of disability insurance applications reduce sickness absenteeism and DI applications?
Jayachandran et al.	2010	AEJ:AE	Did the introduction of sulfa drugs in the 1930s decreased US mortality?
Furman & Stern	2011	AER	Is an article accessible through a Biological Research Center more likely to be cited?
Moser & Voena	2012	AER	Did US compulsory licensing from the 1917 TWEA affect the number of patents by US inventors?
Redding et al.	2011	Rev Econ Stat	Did Berlin and Frankfurt Airport air passenger shares switch roles after WWII?

Summary of Results.

	Method	Outcome	Result	d1d	Common	d2d	Equivalence
Aaronson & Mazumber	DID	School attendance	+ (***)	+	0.008 (0.031)	+	0.003 (0.029)
Abramitzky et al.	DID	Social gap	- (**)	+			
Currie & Walker	DID	Car Pollution (NO_2)	-(***)	- (***)	0.128 (0.163)	- (***)	0.193 (0.037)
Currie & Walker	DID	No-Car Pollution (NS_2)	+	+ (*)	0.699 (0.299)	+ (***)	$-.126$ (0.069)
De Jong et al.	DID	Sickness absenteeism	-(**)	-(**)	0.0007 (0.001)	-(***)	0.0007 (0.001)
De Jong et al.	DID	DI Applications	-	-	0.0011 (0.0005)	-(***)	0.0011 (0.0005)
Furman & Stern	DID	Forward Citations	+(***)	+	1.077 (0.248)	-	0.306 (0.125)
Jayachandran et al.	d2d	Maternal Mortality	-(**)	- (***)	0.271 (0.000)	- (***)	$-.045$ (0.000)
Jayachandran et al.	d2d	Pneumonia/influenza	-	- (***)	0.421 (0.000)	- (***)	0.123 (0.000)
Jayachandran et al.	d2d	Scarlet Fever	-(**)	- (***)	0.351 (0.000)	- (***)	$-.085$ (0.000)
Moser & Voena	DID	Patents by US inventors	+ (***)	+(*)	$-.222$ (0.063)	-	0.109 (0.050)
Moser & Voena	DID, no controls	Patents by US inventors	+ (***)	+	$-.126$ (0.061)	-	0.027 (0.045)
Redding et al.	DID in trends	Passenger shares	-(***)	- (***)	15.478 (0.000)	- (***)	5.827 (0.000)

Conclusions

- How trends are modeled matters: in 6 out of 13 cases, the significance of the results are affected by the trend specification
  - In five cases, significance is lost with a more flexible trend specification
- Which dqd assumption is used matters even more: in 10 out of the 13 cases, the estimated effect is significantly different
- DID is not particularly well supported with a flexible test for a common trend before treatment: only in 3 out of the 9 relevant cases, we could not reject a common trend before treatment

Thank you

Abadie, A., 2006. Semiparametric Difference-in-Differences Estimators. Review of Economic Studies 72, 1–19.

Angrist, J. D., Krueger, A. B., 1999. Empirical strategies in labor economics. In: Ashenfelter, O. C., Card, D. (Eds.), Handbook of Labor Economics. Vol. 3A. Elsevier, pp. 1277–1366.
URL http://www.sciencedirect.com/science/article/pii/S1573446399030047

Athey, S., Imbens, G. W., 2006. Identification and Inference in Nonlinear Difference-in-Differences Models. Econometrica 74 (2), 431–497.
URL http://dx.doi.org/10.1111/j.1468-0262.2006.00668.x

Bertrand, M., Duflo, E., Mullainathan, S., 2004. How Much Should We Trust Differences-in-Differences Estimates? The Quarterly Journal of Economics 119 (1), 249–275.
URL http://econpapers.repec.org/RePEc:tpr:qjecon:v:119:y:2004:i:1:p:249-275

Blundell, R., Dias, M., Meghir, C., Reenen, J., 2004. Evaluating the employment impact of a mandatory job search program. Journal of the European Economic Association 2 (4), 569–606.

Bonhomme, S., Sauder, U., 2011. Recovering distributions in