A Life Cycle Analysis of Social Security

Carlos Lezama, Santiago Payró, Emiliano Ramírez
Dynamic Macroeconomics
ITAM

Spring 2022

1 Introduction

2 A Survey of DD Papers

3 Overrejection in DD Estimation

4 Solutions

5 Conclusion

Introduction

Introduction

Differences-in-Differences (DD) estimation consists of identifying a specific intervention or treatment. One then compares the difference in outcomes after and before the intervention for groups affected by the intervention to the same difference for unaffected groups.

The great appeal of DD estimation comes from its simplicity as well as its potential to circumvent many of the endogeneity problems that typically arise when making comparisons between heterogeneous individuals.

Obviously, DD estimation also has its limitations. It is appropriate when the interventions are as good as random, conditional on time and group fixed effects. Therefore, much of the debate around the validity of a DD estimate typically revolves around the possible endogeneity of the interventions themselves.

Introduction

Focus on issues relating to the *standard error* of the estimate.

DD estimates and their standard errors most often derive from using Ordinary Least Squares (OLS) in repeated cross sections (or a panel) of data on individuals in treatment and control groups for several years before and after a specific intervention.

One then typically estimates the following regression using OLS:

$$Y_{ist} = A_s + B_t + cX_{ist} + \beta I_{st} + \varepsilon_{ist}, \tag{1}$$

where the subscripts i, s, and t stand for the individual, group (such as a state), and time (such as a year), respectively. Furthermore, let Y be our outcome of interest, I be a dummy, A and B be fixed effects, X be relevant individual controls, and ε be an error term.

Introduction

The estimated impact of the intervention is then the OLS estimate $\hat{\beta}$. Standard errors used to form confidence interval for $\hat{\beta}$ are usually OLS standard errors.

Remark

Note that this is valid only under the very restrictive assumption that changes in the outcome variable over time would have been exactly the same in both treatment and control groups in the absence of the intervention.

Introduction

Three factors make serial correlation an especially important issue in the DD context:

1. DD estimation usually relies on fairly long time series.
2. The most commonly used dependent variables in DD estimation are typically highly positively serially correlated.
3. The treatment variable I_{st} changes itself very little within a state over time.

These three factors reinforce each other so that the standard error for $\hat{\beta}$ could severely understate the standard deviation of $\hat{\beta}$.

Introduction

To assess the extent of this problem, they examine how DD performs on placebo laws, where treated states and year of passage are chosen at random.

Placebo tests diagnose problems with research designs in observational studies. When a researcher estimates a treatment effect based on observational data, the estimator may be biased by confounders, model misspecification, differential measurement error, or other flaws; the researcher may also have constructed confidence intervals incorrectly, such that we would reject the null hypothesis too frequently (or infrequently) under the null. A placebo test checks for an association that should be absent if the research design is sound but not otherwise. Placebo tests can thus be seen as a strategy for checking the soundness of a research finding and, more broadly, improving causal inference.

A Survey of DD Papers

A Survey of DD Papers

Papers were classified as "DD" if they focus on specific interventions and use units unaffected by the law as a control group.

Table 1: Survey of DD papers

Number of DD papers	92
Number with more than 2 periods of data	69
Number which collapse data into before-after	4
Number with potential serial correlation problem	65
Number with some serial correlation correction	5
GLS	4
Arbitrary variance-covariance matrix	1
Number with potential clustering problem	80
Number which deal with it	36

A Survey of DD Papers

Table 2: Distribution of time span for papers with more than 2 periods

Percentile	Value
1%	3
5%	3
10%	4
25%	5.75
50%	11
75%	21.5
90%	36
95%	51
99%	83
Average	16.5

A Survey of DD Papers

Table 3: Most commonly used dependent variables

Employment	18
Wages	13
Health/medical expenditure	8
Unemployment	6
Fertility/teen motherhood	4
Insurance	4
Poverty	3
Consumption/savings	3

A Survey of DD Papers

Table 4: Informal techniques used to assess endogeneity

Graph dynamics of effect	15
See if effect is persistent	2
Attempt to do triple-differences (DDD)	11
Include time trend specific to treated states	7
Look for effect prior to intervention	3
Include lagged dependent variable	3

Overrejection in DD Estimation

Data

The survey above suggests that most DD papers may report standard errors that understate the standard deviation of the DD estimator. To illustrate the magnitude of the problem they turn to a sample of women's wages from the Current Population Survey (CPS).

More specifically, data on:

- Women in their fourth interview month in the Merged Outgoing Rotation Group of the CPS.
- Years: 1979 – 1999.
- Age: 25 – 50 y/o.
- Information on weekly earnings, employment status, education, age, and state of residence.

Data

Summary

The sample contains:

- Nearly 900,000 observations.
- Approximately 540,000 women report strictly positive weekly earnings.
- wage $=$ log(weekly earnings).

This generates ($50 \times 21 = 1050$) state-year cells, with each cell containing on average a little more than 500 women with strictly positive earnings.

Methodology

The correlogram of the wage residuals was informative enough to estimate first, second, and third autocorrelation coefficients for the mean state-year residuals from a regression of wages on state and year dummies such that they equal 0.51, 0.44, and 0.31, respectively (obtained by a simple OLS regression of the residuals on the corresponding lagged residuals) — which are high and statistically significant.

Methodology

Subsequently, in the DD context:

1. Randomly, draw a year $\sim \mathcal{U}(1985, 1995)$.
2. Select exactly half states (25) at random and designate them as "affected" by the law such that

$$I_{st} = \begin{cases} 1 & \text{for all women that live in an affected state} \\ & \text{after the intervention date,} \\ 0 & \text{otherwise.} \end{cases}$$

3. Estimate equation (1) using OLS on these placebo laws.

Methodology

If OLS were to provide consistent standard errors, we would expect to reject the null hypothesis of no effect ($\beta = 0$) roughly 5 percent of the time when using a threshold of 1.96 for the absolute t-statistic.

Methodology

A. CPS DATA

Data	$\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$	Modifications	Rejection rate: No effect	Rejection rate: 2% effect
1) CPS micro, log wage			.675 (.027)	.855 (.020)
2) CPS micro, log wage		Cluster at state-year level	.44 (.029)	.74 (.025)
3) CPS agg, log wage	.509, .440, .332		.435 (.029)	.72 (.026)
4) CPS agg, log wage	.509, .440, .332	Sampling w/replacement	.49 (.025)	.663 (.024)
5) CPS agg, log wage	.509, .440, .332	Serially uncorrelated laws	.05 (.011)	.988 (.006)
6) CPS agg, employment	.470, .418, .367		.46 (.025)	.88 (.016)
7) CPS agg, hours worked	.151, .114, .063		.265 (.022)	.280 (.022)
8) CPS agg, changes in log wage	−.046, .032, .002		0	.978 (.007)

Methodology

B. MONTE CARLO SIMULATIONS WITH SAMPLING FROM AR(1) DISTRIBUTION

Data	ρ	Modifications	Rejection rate	
			No effect	2% effect
9) AR(1)	.8		.373	.725
			(.028)	(.026)
10) AR(1)	0		.053	.783
			(.013)	(.024)
11) AR(1)	.2		.123	.738
			(.019)	(.025)
12) AR(1)	.4		.19	.713
			(.023)	(.026)
13) AR(1)	.6		.333	.700
			(.027)	(.026)
14) AR(1)	−.4		.008	.7
			(.005)	(.026)

Methodology

These results demonstrate that, in the presence of positive serial correlation, conventional DD estimation leads to gross overestimation of t-statistics and significance levels.

Solutions

Solutions

VARYING N AND T

Data	N	T	Rejection rate: No effect	Rejection rate: 2% effect
		A. CPS DATA		
1) CPS aggregate	50	21	.49	.663
			(.025)	(.024)
2) CPS aggregate	20	21	.39	.54
			(.024)	(.025)
3) CPS aggregate	10	21	.443	.510
			(.025)	(.025)
4) CPS aggregate	6	21	.383	.433
			(.025)	(.025)
5) CPS aggregate	50	11	.20	.638
			(.020)	(.024)
6) CPS aggregate	50	7	.15	.635
			(.017)	(.024)
7) CPS aggregate	50	5	.078	.5
			(.013)	(.025)
8) CPS aggregate	50	3	.048	.363
			(.011)	(.024)
9) CPS aggregate	50	2	.055	.28
			(.011)	(.022)

Parametric Methods

A first possible solution to the serial correlation problem would be to specify an autocorrelation structure for the error term, estimate its parameters, and use these parameters to compute standard errors.

Parametric Methods

PARAMETRIC SOLUTIONS

Data	Technique	Estimated $\hat{\rho}_1$	Rejection rate: No effect	Rejection rate: 2% Effect
		A. CPS DATA		
1) CPS aggregate	OLS		.49	.663
			(.025)	(.024)
2) CPS aggregate	Standard AR(1) correction	.381	.24	.66
			(.021)	(.024)
3) CPS aggregate	AR(1) correction imposing $\rho = .8$		.18	.363
			(.019)	(.024)
		B. OTHER DATA GENERATING PROCESSES		
4) AR(1), $\rho = .8$	OLS		.373	.765
			(.028)	(.024)
5) AR(1), $\rho = .8$	Standard AR(1) correction	.622	.205	.715
			(.023)	(.026)
6) AR(1), $\rho = .8$	AR(1) correction imposing $\rho = .8$		.06	.323
			(.023)	(.027)
7) AR(2), $\rho_1 = .55$ $\rho_2 = .35$	Standard AR(1) correction	.444	.305	.625
			(.027)	(.028)
8) AR(1) + white noise, $\rho = .95$, noise/signal = .13	Standard AR(1) correction	.301	.385	.4
			(.028)	(.028)

Block Bootstrap

This variant of bootstrap maintains the autocorrelation structure by keeping all the observations that belong to the same group (e.g., state) together. In practice, we bootstrap the t-statistic as follows. For each placebo intervention we compute the absolute t-statistic $t = |\hat{\beta}/\text{SE}(\hat{\beta})|$, using the OLS estimate of β and its standard error. We then construct a bootstrap sample by drawing with replacement 50 matrices $(\bar{Y}_s, V_s)$, where $\bar{Y}_s$ is the entire time series of observations for state s, and V_s is the matrix of state dummies, time dummies, and treatment dummy for state s.

Block Bootstrap

We then run OLS on this sample. obtain an estimate $\hat{\beta}_r$ and construct the absolute t-statistic $t_r = |(\hat{\beta}_r - \hat{\beta})/\text{SE}(\hat{\beta}_r)|$.

Remark

The difference between the distribution of t_r and the sampling distribution of t becomes small as N goes to infinity, even in presence of arbitrary autocorrelation within states and heteroskedasticity.

Block Bootstrap

BLOCK BOOTSTRAP

Data	Technique	N	Rejection rate: No effect	Rejection rate: 2% effect
	A. CPS DATA			
1) CPS aggregate	OLS	50	.43	.735
			(.025)	(.022)
2) CPS aggregate	Block bootstrap	50	.065	.26
			(.013)	(.022)
3) CPS aggregate	OLS	20	.385	.595
			(.022)	(.025)
4) CPS aggregate	Block bootstrap	20	.13	.19
			(.017)	(.020)
5) CPS aggregate	OLS	10	.385	.48
			(.024)	(.024)
6) CPS aggregate	Block bootstrap	10	.225	.25
			(.021)	(.022)
7) CPS aggregate	OLS	6	.48	.435
			(.025)	(.025)
8) CPS aggregate	Block bootstrap	6	.435	.375
			(.022)	(.025)

Ignoring Time Series Information

Simple Aggregation

One could simply average the data before and after the law and run equation (1) on this averaged outcome variable in a panel of length 2 — this solution will work only for laws that are passed at the same time for all the treated states.

Residual Aggregation

First, one can regress Y_{st} on state fixed effects, year dummies, and any relevant covariates. One can then divide the residuals of the treatment states only into two groups: residuals from years before the laws, and residuals from years after the laws. The estimate of the laws' effect and its standard error can then be obtained from an OLS regression in this two-period panel. It also does well when the laws are staggered over time.

Ignoring Time Series Information

IGNORING TIME SERIES DATA

Data	Technique	N	Rejection rate: No effect	Rejection rate: 2% effect
	A. CPS DATA			
1) CPS agg	OLS	50	.49	.663
			(.025)	(.024)
2) CPS agg	Simple aggregation	50	.053	.163
			(.011)	(.018)
3) CPS agg	Residual aggregation	50	.058	.173
			(.011)	(.019)
4) CPS agg, staggered laws	Residual aggregation	50	.048	.363
			(.011)	(.024)
5) CPS agg	OLS	20	.39	.54
			(.025)	(.025)
6) CPS agg	Simple aggregation	20	.050	.088
			(.011)	(.014)
7) CPS agg	Residual aggregation	20	.06	.183
			(.011)	(.019)
8) CPS agg, staggered laws	Residual aggregation	20	.048	.130
			(.011)	(.017)

Empirical Variance-Covariance Matrix

Suppose that the autocorrelation process is the same across all states and that there is **no** cross-section **heteroskedasticity**. In this case, if the data are sorted by states, and years, the variance-covariance matrix of the error term is block diagonal. Each of these blocks is symmetric, and the element $(i, i+j)$ is the correlation between ε_i and ε_{i-j}. We can therefore use the variation across the 50 states to estimate each element of this matrix, and use this estimated matrix to compute standard errors.

Remark

Under our initial assumption, this method will produce consistent estimates of the standard error as $N \longrightarrow \infty$.

Empirical Variance-Covariance Matrix

EMPIRICAL VARIANCE-COVARIANCE MATRIX

Data	Technique	N	Rejection rate: No effect	Rejection rate: 2% effect
	A. CPS DATA			
1) CPS aggregate	OLS	50	.49	.663
			(.025)	(.024)
2) CPS aggregate	Empirical variance	50	.055	.243
			(.011)	(.021)
3) CPS aggregate	OLS	20	.39	.54
			(.024)	(.025)
4) CPS aggregate	Empirical variance	20	.08	.138
			(.013)	(.017)
5) CPS aggregate	OLS	10	.443	.510
			(.025)	(.025)
6) CPS aggregate	Empirical variance	10	.105	.145
			(.015)	(.018)
7) CPS aggregate	OLS	6	.383	.433
			(.025)	(.025)
8) CPS aggregate	Empirical variance	6	.153	.185
			(.018)	(.019)

Arbitrary Variance-Covariance Matrix

Since the assumption in the previous method is likely to be violated in practice, this method can be generalized to an estimator of the variance-covariance matrix which is consistent in the presence of any correlation pattern within states over time.

Arbitrary Variance-Covariance Matrix

This estimator for the variance-covariance matrix is given by

$$W = (V'V)^{-1} \left(\sum_{j=1}^{N} u_j' u_j \right) (V'V)^{-1},$$

where V is matrix of independent variables (year dummies, state dummies and treatment dummy). Furthermore, u_j is defined as follows

$$u_j = \sum_{t=1}^{T} e_{jt} v_{jt},$$

where e_{jt} is the estimated residual for state i at time t, and v_{jt} is the row vector of dependent variables.

Arbitrary Variance-Covariance Matrix

ARBITRARY VARIANCE-COVARIANCE MATRIX

Data	Technique	N	Rejection rate	
			No effect	2% effect
A. CPS DATA				
1) CPS aggregate	OLS	50	.49	.663
			(.025)	(.024)
2) CPS aggregate	Cluster	50	.063	.268
			(.012)	(.022)
3) CPS aggregate	OLS	20	.385	.535
			(.024)	(.025)
4) CPS aggregate	Cluster	20	.058	.13
			(.011)	(.017)
5) CPS aggregate	OLS	10	.443	.51
			(.025)	(.025)
6) CPS aggregate	Cluster	10	.08	.12
			(.014)	(.016)
7) CPS aggregate	OLS	6	.383	.433
			(.024)	(.025)
8) CPS aggregate	Cluster	6	.115	.118
			(.016)	(.016)

Conclusion

Conclusion

This study suggests that, because of serial correlation, conventional DD standard errors may grossly understate the standard deviation of the estimated treatment effects, leading to serious overestimation of t-statistics and significance levels. In other words, it is possible that too many false rejections of the null hypothesis of no effect have taken place.