Clustered Standard Errors

1. The Attraction of "Differences in Differences"
2. Grouped Errors Across Individuals
3. Serially Correlated Errors

1. The Attraction of Differences in Differences Estimates

- Typically evaluate programs which differ across groups, such as U.S. States e.g., effect of changes in state minimum wage laws or state welfare programs on earnings or unemployment
- Treat selection (heterogeneity) bias by removing state effects (one “diff”)
- Treat common economic fluctuations by removing year effects (the other “diff”)
- Hence the appealing nickname “diffs in diffs”

2. The Grouped Error Problem:

- Binary covariates define groups within which errors are potentially correlated (e.g., cities, states, years, states after treatment, self-employed, etc..)
  - remember that errors contain unobserved variables
- $Y_{ist} = A_{s\cancel{t}} + B_t + cX_{ist} + \beta I_{st} + \varepsilon_{ist}$,
- s are groups (perhaps states)
- t is time
- I is an indicator for treatment, which occurs as the group x time level
- ε is an error term, which is not necessary iid.

2. Grouped Errors Across Individuals
E.g., Minimum wages on NJ/Penn border

- Card and Krueger (1994) looked at the effects of minimum wages on employment in fast-food restaurants near the NJ – Penn border.
- Data collected before and after NJ raised its' minimum wage by 80 cents (in 1992).
- i - restaurant, s – state, t – time
- S=2, T=2, N is large.
- They found small positive effects within a small confidence interval of zero.

2. Grouped Errors Across Individuals E.g., Mariel Boatlift

- Card (1990) looked at the effects of a surprise supply shock of immigrants to Miami due to a temporary lifting of emigration restrictions by Cuba in 1980.
- He estimates the effect of the boatlift on unemployment and wages of low skill workers in Miami using four other cities as comparisons (Atlanta, Houston, LA and Tampa-St. Petersburg) with CPS data.
- i - individual, s – city, t – time
- S=5, T~=2, N is large.
- He finds no statistically significant effect on employment or wages of the labor supply shock.

2. Grouped Errors Across Individuals

- How big does the number of groups (S, or S*T) have to be?
- $Y_{ist} = a_{st} + d_t + cZ_{ist} + \beta I_{st} + \varepsilon_{ist}$,
- Donald and Lang (2004): In the (plausible) case where we have some within-group correlation, and under generous assumptions the t-statistics converge to a normal distribution at rate S*T no matter what N is.
- Intuition: Imagine that within s,t groups the errors are perfectly correlated. Then you might as well aggregate and run the regression with S*T observations.
- Intuition: 2 step estimator
- If group and time effects are included, with normally distributed group-time specific errors under generous assumptions, the t-statistics have a t distribution with S*T-S-T degrees of freedom, no matter what N is. (Table 3)
- Donald-Lang suggested estimator has this flavor. (Table 3)
- Alternative: collapse into s,t groups
- 3 issues: consistent s.e., efficient s.e. and distribution of t-stat in small samples

Distribution of t-ratio, 4 d.o.f, $\beta = 0$

TABLE 3 MONTE CARLO ESTIMATION Distribution of t-statistics (4 groups, 2500 observations per group)					
	99^{th} percentile	95^{th} percentile	90^{th} percentile	% > 1.645	% > 1.96
OLS (conventional standard errors)					
No Z	13.01	9.93	8.40	74.5	69.8
Z	13.01	9.93	8.40	74.5	69.9
OLS (Eicker-White standard errors)					
No Z	15.56	6.74	4.58	39.6	33.0
Z	14.58	6.74	4.58	39.6	33.0
Feasible GLS (random effects)					
No Z	7.34	4.00	2.82	23.8	18.6
Z	9.74	6.07	4.30	32.7	27.4
Two-Step					
No Z	9.72	4.28	2.92	24.1	18.9
Z	9.75	4.28	2.92	24.1	18.9

When N=250 the simulated distribution is almost identical

3. Correlations over time in panels

- $Y_{ist} = A_{s\cancel{t}} + B_t + cX_{ist} + \beta I_{st} + \varepsilon_{ist}$,
- S are groups (perhaps states)
- t is time
- I is an indicator for treatment, which occurs as the group x time level
- Correlations within group, period (i.e., s,t) cells only is very restrictive.
- In general we want to allow correlations over time as well (within s but not within t)

Lots of DD papers

T is large

The variables tend to be serially corr.

So are std. errors consistent?

TABLE I
SURVEY OF DD PAPERS[a]

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	
Distribution of time span for papers with more than 2 periods	Average	16.5
	Percentile	Value
	1%	3
	5%	3
	10%	4
	25%	5.75
	50%	11
	75%	21.5
	90%	36
	95%	51
	99%	83
Most commonly used dependent variables	Number	
Employment	18	
Wages	13	
Health/medical expenditure	8	
Unemployment	6	
Fertility/teen motherhood	4	
Insurance	4	
Poverty	3	
Consumption/savings	3	
Informal techniques used to assess endogeneity	Number	
Graph dynamics of effect	15	
See if effect is persistent	2	
DDD	11	
Include time trend specific to treated states	7	
Look for effect prior to intervention	3	
Include lagged dependent variable	3	
Number with potential clustering problem	80	
Number which deal with it	36	

Data come from a survey of all articles in six journals between 1990 and 2000: the *American Economic Review*, the *Industrial Labor Relations Review*, the *Journal of Labor Economics*, the *Journal of Political Economy*, the *Journal of Public Economics*, and the *Quarterly Journal of Economics*. We define an article as "Difference-in-Difference" if it (1) examines the effect of a specific intervention and (2) uses units unaffected by the intervention as a control group.

Placebo Binary "Laws"

- Randomly choose a year between 79-99 & randomly assign a law to 25 states til end of 99
- Rej. rate is % for which t>1.96

TABLE II
DD REJECTION RATES FOR PLACEBO LAWS

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)

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

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

a. Unless mentioned otherwise under "Modifications," reported in the last two columns are the OLS rejection rates of the null hypothesis of no effect (at the 5 percent significance level) on the intervention variable for randomly generated placebo interventions as described in text. The data used in the last column were altered to simulate a true 2 percent effect of the intervention. The number of simulations for each cell is at least 200 and typically 400.

b. CPS data are data for women between 25 and 50 in the fourth interview month of the Merged Outgoing Rotation Group for the years 1979 to 1999. In rows 3 to 8 of Panel A, data are aggregated to state-year level cells after controlling for demographic variables (four education dummies and a quartic in age). For each simulation in rows 1 through 3, we use the observed CPS data. For each simulation in rows 4 through 8, the data generating process is the state-level empirical distribution of the CPS data that puts a probability of 1/50 on the different states' outcomes (see text for details). For each simulation in Panel B, the data generating process is an AR(1) model with normal disturbances chosen to match the CPS state female wage variances (see text for details). $\hat{\rho}_i$ refer to the estimated autocorrelation parameter of lag i. ρ refers to the autocorrelation parameter in the AR(1) model.

c. All regressions include, in addition to the intervention variable, state and year fixed effects. The individual level regressions also include demographic controls.

Placebo Binary "Laws"

- Type I error is worst when T is large

TABLE III
VARYING N AND T

Data	N	T	Rejection rate: No effect	Rejection rate: 2% effect
A. CPS DATA				
1) CPS aggregate	50	21	.49 (.025)	.663 (.024)
2) CPS aggregate	20	21	.39 (.024)	.54 (.025)
3) CPS aggregate	10	21	.443 (.025)	.510 (.025)
4) CPS aggregate	6	21	.383 (.025)	.433 (.025)
5) CPS aggregate	50	11	.20 (.020)	.638 (.024)
6) CPS aggregate	50	7	.15 (.017)	.635 (.024)
7) CPS aggregate	50	5	.078 (.013)	.5 (.025)
8) CPS aggregate	50	3	.048 (.011)	.363 (.024)
9) CPS aggregate	50	2	.055 (.011)	.28 (.022)
B. MONTE CARLO SIMULATIONS WITH SAMPLING FROM AR(1) DISTRIBUTION				
10) AR(1), $\rho = .8$	50	21	.35 (.028)	.638 (.028)
11) AR(1), $\rho = .8$	20	21	.35 (.028)	.538 (.029)
12) AR(1), $\rho = .8$	10	21	.3975 (.028)	.505 (.029)
13) AR(1), $\rho = .8$	6	21	.393 (.028)	.5 (.029)
14) AR(1), $\rho = .8$	50	11	.335 (.027)	.588 (.028)
15) AR(1), $\rho = .8$	50	5	.175 (.022)	.5525 (.029)
16) AR(1), $\rho = .8$	50	3	.09 (.017)	.435 (.029)
17) AR(1), $\rho = .8$	50	50	.4975 (.029)	.855 (.020)

Solutions: AR(1) correction

- N=50, T=21
- AR(1) biased for small T
- Process looks more like AR(2)

TABLE IV
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)

Solutions: Ignore TS Information

- correct size but loss of power

- Residual aggregation is a Frisch-Waugh exercise: first - regress on other variables, then - aggregate residuals before and after treatment

TABLE VI
IGNORING TIME SERIES DATA

Data	Technique	N	Rejection rate: No effect	Rejection rate: 2% effect
A. CPS DATA				
1) CPS agg	OLS	50	.49 (.025)	.663 (.024)
2) CPS agg	Simple aggregation	50	.053 (.011)	.163 (.018)
3) CPS agg	Residual aggregation	50	.058 (.011)	.173 (.019)
4) CPS agg, staggered laws	Residual aggregation	50	.048 (.011)	.363 (.024)
5) CPS agg	OLS	20	.39 (.025)	.54 (.025)
6) CPS agg	Simple aggregation	20	.050 (.011)	.088 (.014)
7) CPS agg	Residual aggregation	20	.06 (.011)	.183 (.019)
8) CPS agg, staggered laws	Residual aggregation	20	.048 (.011)	.130 (.017)
9) CPS agg	OLS	10	.443 (.025)	.51 (.025)
10) CPS agg	Simple aggregation	10	.053 (.011)	.065 (.012)
11) CPS agg	Residual aggregation	10	.093 (.014)	.178 (.019)
12) CPS agg, staggered laws	Residual aggregation	10	.088 (.014)	.128 (.017)
13) CPS agg	OLS	6	.383 (.024)	.433 (.024)
14) CPS agg	Simple aggregation	6	.068 (.013)	.07 (.013)
15) CPS agg	Residual aggregation	6	.11 (.016)	.123 (.016)
16) CPS agg, staggered laws	Residual aggregation	6	.09 (.014)	.138 (.017)
B. AR(1) DISTRIBUTION				
17) AR(1), $\rho = .8$	Simple aggregation	50	.050 (.013)	.243 (.025)
18) AR(1), $\rho = .8$	Residual aggregation	50	.045 (.012)	.235 (.024)
19) AR(1), $\rho = .8$, staggered laws	Residual aggregation	50	.075 (.015)	.355 (.028)

Solutions: "Cluster" within states (over time)

- simple, easy to implement
- Works well for N=10
- But this is only one data set and one variable (CPS, log weekly earnings)

-

TABLE VIII
ARBITRARY VARIANCE-COVARIANCE MATRIX

Data	Technique	N	Rejection rate: No effect	Rejection rate: 2% effect
A. CPS DATA				
1) CPS aggregate	OLS	50	.49 (.025)	.663 (.024)
2) CPS aggregate	Cluster	50	.063 (.012)	.268 (.022)
3) CPS aggregate	OLS	20	.385 (.024)	.535 (.025)
4) CPS aggregate	Cluster	20	.058 (.011)	.13 (.017)
5) CPS aggregate	OLS	10	.443 (.025)	.51 (.025)
6) CPS aggregate	Cluster	10	.08 (.014)	.12 (.016)
7) CPS aggregate	OLS	6	.383 (.024)	.433 (.025)
8) CPS aggregate	Cluster	6	.115 (.016)	.118 (.016)
B. AR(1) DISTRIBUTION				
9) AR(1), $\rho = .8$	Cluster	50	.045 (.012)	.275 (.026)
10) AR(1), $\rho = 0$	Cluster	50	.035 (.011)	.74 (.025)

Current Standard Practice

- Be conservative: cluster by group or time (not the interaction) and report the larger std. error
  - note: this may get size and power wrong
- Better.. you can cluster on both!
  Cameron, Gelbach, and Miller (2006, NBER Technical WP)
  method not coded in Stata yet, but you can get an .ado from Doug Miller's Stata page
  http://www.econ.ucdavis.edu/faculty/dlmiller/statafiles/
- Do you have enough groups for a normal approximation?
  .. Check with a "Wild Bootstrap" Cameron, Gelbach, Miller (*ReStat* 2008);
  .do file on Miller's page.
- May be argument for using Newey-West std. errors.

- Ask Gordon Dahl, who is working on a better method

Exam ?

- Wed Dec 7 in Granger room, 3PM