ECON-C4200 - Econometrics II

Lecture 4: Difference in difference application: Bloom et al., 2015: Does working from home work? Evidence from a Chinese experiment

Otto Toivanen

What was the setting?

- Ctrip, NASDAQ-listed, China's largest travel agency. 16K empl. Worth 5B$ at the time of experiment.
- Increasing rental cost in Shanghai.
- High attrition among employees (commute).
- Increased shirking.
- Young employees.

What was the setting?

- Four types of jobs involved in the experiment:

1. Order takers
2. Order placers
3. Order correctors
4. Night shift placing and correcting orders.

What was the setting?

- 5 shifts a week.
- Team work.
- Earnings: flat wage + bonus linear in (volume, quality, shift type).
- Flat wage slightly >50% of total avg. earnings.
- The experiment changed only location of work.

WFH differences between treated and controls

- Treatment = 4 days home, 1 in office. 9 months.
- Commuting time.
- Supervisor support.
- Work environment.
- = the treatment.

WFH treatment

- It is crucial the researcher understands the treatment.
- Dichotomous versus multivalued treatment.
- Ask yourself: Is the treatment the same for everybody here? (e.g. what about differences in commuting time?).

WFH setup

- 996 employees in the Shanghai call center asked if they want to volunteer.
- 503 interested.
- 249 eligible ($=$tenure $>$ 6months, broadband at home, independent workspace at home).
- Treated and controls could not switch during experiment.

Who volunteers?

TABLE I
WFH VOLUNTEERS

Dependent variable: volunteer to work from home	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Sample mean
Children	0.123**		0.054	0.075	0.081		0.084	0.08
	(0.056)		(0.083)	(0.083)	(0.083)		(0.084)	
Married[a]		0.095**	0.012	0.054	0.052		0.057	0.15
		(0.044)	(0.065)	(0.066)	(0.066)		(0.068)	
Daily commute (minutes[a])			0.062**	0.062**	0.071**		0.072**	80.6
			(0.030)	(0.031)	(0.032)		(0.0032)	
Own bedroom			0.095***	0.088**	0.089**		0.089**	0.60
			(0.035)	(0.035)	(0.036)		(0.037)	
Tertiary education and above				−0.080**	−0.088***		−0.086**	0.42
				(0.033)	(0.033)		(0.034)	
Tenure (months[a])				−0.268***	−0.415***		−0.401***	25.0
				(0.080)	(0.110)		(0.117)	
Gross wage (¥1,000)					0.048**	−0.019	0.048**	2.86
					(0.024)	(0.017)	(0.024)	
Age							−0.002	23.2
							(0.007)	
Male							0.010	0.32
							(0.036)	
Number of employees	994	994	994	994	994	994	994	994

R-squared round 0.03, suggesting volunteering based mostly on “unobservables”

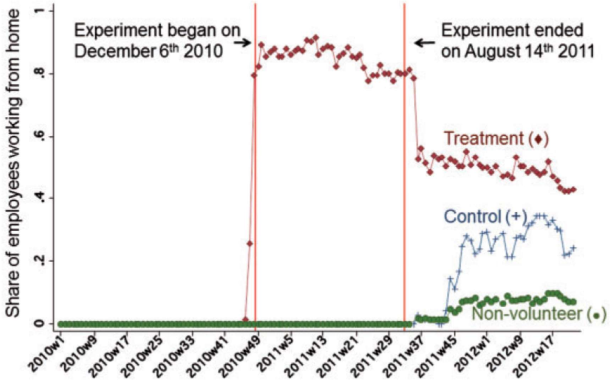
WFH randomization

- Rule: even birthdays $\rightarrow$ treatment, odd $\rightarrow$ control.
- Rule determined by lottery by the CEO.
- How to test success in randomization?
- Look at observable characteristics.
- T-tests between treatment and control groups show 1(#kids) the only statistically significant difference out of 18 characteristics (Appendix table A6).

WFH post experiment choices

- 50% of treated moved back to office.
- 35% of controls moved home.
- 10% of those who did not volunteer moved home.
- Notice how **endogenous sorting** yields a different distribution of individuals than randomization.

WFH worker allocation



Why not just rely on randomization?

fn#17:*Because we have a randomized intervention we can examine either the* ***difference between treatment and control*** *(evaluated over the experimental period), or* ***the difference of differences*** *(evaluated as the change in performance between treatment and control over the experimental period versus the pre-exper- imental period). Since employees have* ***large preexisting cross-sectional variations in performance****, we appear to obtain* ***more accurate (lower mean-squared error) estimations*** *from using the difference in differences specification, estimated using* ***the panel with employee fixed effects****. However, comparing columns (1) and (2) we see the* ***estimators are quantitatively similar*** *and within 1 standard deviation of each other.*

Estimation equation

$$Employee\ Performance_{i,t} = \alpha\, Treat_i \times Experiment_t + \beta_t + \gamma_i + \epsilon_{it} \quad (1)$$

- $Employee\ Performance_{i,t}$ = performance of employee i during week t.
- $Treat_i$ = dummy taking value 1 if individual i in the treatment group (even-numbered birthday).
- $Experiment_t$ = dummy that equals 1 for the experimental period Dec. 6 - Aug. 14.
- $Employee\ Performance_{i,t}$ = work performance. Alternative measures:
  1. log of weekly phone calls answered;
  2. log of phone calls answered / minute on the phone;
  3. log of weekly sum of minutes on the phone;
  4. overall performance z-score. (mean zero, std. one, based on pre-experiment performance).
- β_t = full set of week dummies; γ_i = full set of individual FE.

Estimation results

TABLE II
THE PERFORMANCE IMPACT OF WFH

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable	Overall performance	Overall performance	Phone calls	Phone calls	Phone calls per minute	Minutes on the phone	Gross wage
Period	Pre and during experiment	During experiment	Pre and during experiment	Pre and during experiment	Pre and during experiment	Pre and during experiment	Pre and during experiment
Dependent normalization	*z*-score	*z*-score	*z*-score	log	log	log	log
$Experiment_t$*$Treatment_i$	0.232***		0.248***	0.120***	0.032**	0.088***	0.094***
	(0.063)		(0.058)	(0.025)	(0.001)	(0.027)	(0.032)
$Treatment_i$		0.184**					
		(0.086)					
Number of employees	249	249	134	134	134	134	249
Number of time periods	85	37	85	85	85	85	20
Individual fixed effects	Yes	No	Yes	Yes	Yes	Yes	Yes
Observations	17,806	7,476	9,426	9,426	9,426	9,426	4,648

Performance differences over time

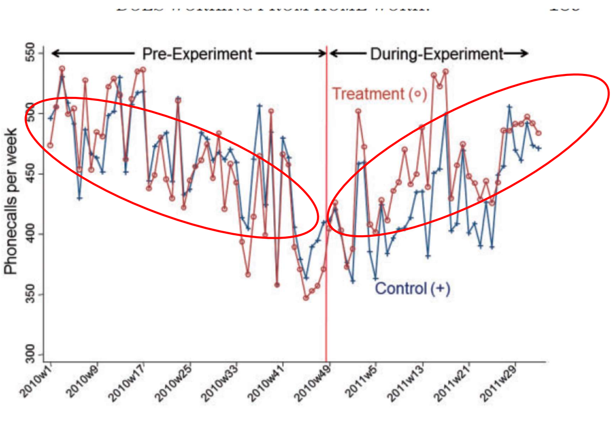


FIGURE VI

Why does the treatment effect not start immediately?

pp. 188: *Interestingly, the difference in performance was greatest during the middle of the experiment, from about two to six months. It seems* ***the smaller rise*** *in performance during the* ***first two months*** *was due to* ***installation and learning effects****. It took several weeks for all the IT and logistical bugs to be addressed.*

Why does the treatment effect melt away?

pp. 189: *The gradual decline in the performance gap from six months onward reflects two trends. First,* ***poorly performing*** *employees in the* ***control group*** *were* ***more likely to quit*** *than those in the treatment group (see Section IV.B and Table VIII), boosting the control group's performance absolutely and relative to the treatment group. Second, from surveys and interviews we learned that some* ***employees in the treatment group felt lonely*** *working at home after a few months and wanted to return to the office but could not because of the experimental design. This potentially affected their motivation.*

Performance differences in a cross-section (3 months into treatment)

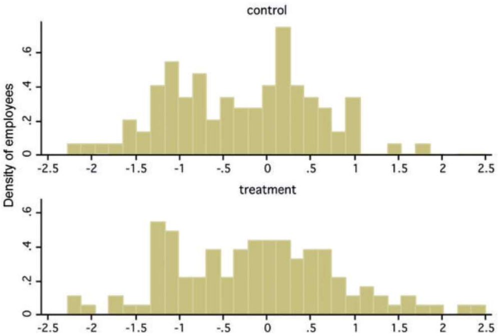


FIGURE VII

Cross-Sectional Performance Spread During the Experiment

Histograms of the performance z-score for the treatment and control groups after 3 months into experiment (SD = 1 across individuals in the pre-experimental data).

Individual labor supply

TABLE III

WFH PRODUCTIVITY

Variables	(1) Minutes on the phone	(2) Minutes on the phone/days worked	(3) Days worked	(4) Minutes on the phone	(5) Minutes on the phone/days worked	(6) Days worked
*Experiment$_t$*Treatment$_i$*	0.088***	0.063***	0.025**	0.069**	0.049*	0.021
	(0.027)	(0.024)	(0.012)	(0.030)	(0.027)	(0.013)
*Experiment$_t$*Treatment$_i$**				0.069*	0.055*	0.014
[total commute > 120 min]$_i$				(0.036)	(0.031)	(0.017)
Number of employees	134	134	134	134	134	134
Number of weeks	85	85	85	85	85	85
Observations	9,426	9,426	9,426	9,426	9,426	9,426

Notes. The regressions are run at the individual by week level, with a full set of individual and week fixed effects. *Experiment*treatment* is the interaction of the period of the experimentation (December 6, 2010, until August 14, 2011) by an individual having an even birthdate (2nd, 4th, 6th, etc. day of the month). The pre-experiment period refers to January 1, 2010, until November 28, 2010. During the experiment period refers to December 6, 2010, to August 14, 2011. In columns (4)–(6), *Experiment* × *Treatment* is further interacted with a dummy variable indicating whether an employee's total daily commute (to and from work) is longer than 120 minutes (21.3% of employees have a commute longer than 120 minutes). Standard errors are clustered at the individual level. Once employees quit they are dropped from the data. *** denotes 1% significance, ** 5% significance, and * 10% significance. Minutes on the phone are recorded from the call logs.

Why does the treatment effect melt away?

pp. 192: *Column (2) shows that about three quarters of the difference in the time on the phone was accounted for by the* ***treatment group's spending more time on the phone*** *per day worked. This is because: (i)* ***they started work more punctually****, a phenomenon they attributed to avoiding the impact of events like bad traffic or the heavy snow in Shanghai in February 2011;18 (ii)* ***they could schedule personal matters****, like doctor's appointments, in the time they saved by not commuting (rather than having to leave early); and (iii)* ***they took shorter breaks*** *during the day because breaks (for lunch or toilet) were less time-consuming at home*

Was the effect due to improvement in the treatment group of decline of performance in the control group?

TABLE IV

THE IMPACT OF WFH AGAINST NAN TONG AND NONEXPERIMENTAL EMPLOYEES

Variables	(1) Overall performance (*z*-score)	(2) Phone calls (*z*-score)	(3) Overall performance (*z*-score)	(4) Phone calls (*z*-score)
Comparison group	Nan Tong	Nan Tong	Nonexperiment	Nonexperiment
$Experiment_t$*$treatment_i$	0.194***	0.281***	0.302***	0.312***
	(0.047)	(0.048)	(0.060)	(0.064)
$Experiment_t$*$control_i$	−0.035	−0.011	0.066	0.019
	(0.048)	(0.043)	(0.061)	(0.061)
Observations	99,753	86,589	27,823	15,261

Internal validity

- Failure to randomize: No evidence of this.
- Non-compliance by the subjects: Compliance in the treatment group 80-90%.
- Attrition: May have affected the control group (17% in the treatment, 35% in the control group).

Internal validity

- Experimental effects (Hawthorne): employees motivated by the experiment (e.g. to make sure WFH is rolled out).
  - 131 individuals in the treatment group make this unlikely (each had small effect).
  - Returners (to office) and stayers of WFH had similar performance in the treatment group.
  - The performance gap grew after the experiment.
  - The firm rolled out WFH.

What happened after the experiment? Roll-out

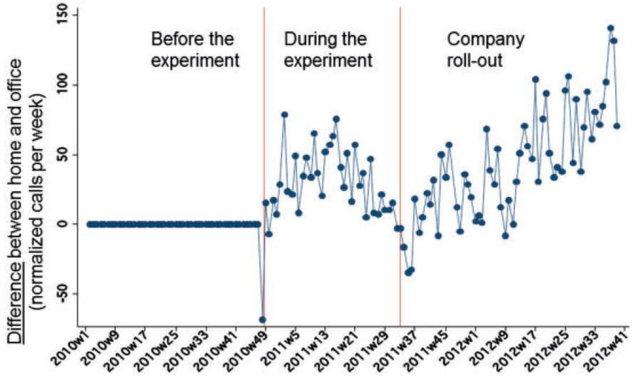


FIGURE VIII

Selection Further Increased the Performance Impact of Home Working During the Company Roll-Out

What happened after the experiment? Roll-out

TABLE V

SELECTION EFFECTS

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Performance	Performance	Performance	Log(phone calls)	Log(phone calls)	Log(phone calls)
Dependent normalization	*z*-score	*z*-score	*z*-score	Log	log	log
Sample	All	All	Balanced	All	All	Balanced
$Experiment_t$*$WFH_{i,t}$	0.244***	0.221***	0.174***	0.134***	0.125***	0.104***
	(0.059)	(0.049)	(0.057)	(0.029)	(0.035)	(0.041)
$Post\text{-}Experiment_t$*$WFH_{i,t}$		0.284***	0.245***		0.220***	0.203***
		(0.082)	(0.089)		(0.059)	(0.066)
Number of employees	249	249	150	134	134	73
Number of weeks	85	144	144	85	144	143
Observations	17,614	25,449	18,214	9,440	13,278	8,866

Notes. $WFH_{i,t}$ here is defined as working from home at least one day that week. The regressions are run at the individual by week level, with a full set of individual and week fixed effects. The pre-experiment period is January 1, 2010–November 28, 2010. *Experiment*WFH* is the interaction of the period of the experimentation (December 6, 2010–August 14, 2011) with an individual having worked from home at least one day a week by week. *Post-experiment*WFH* is the interaction of the period after the experimentation from August 14, 2011, until end of September 2012 with an individual having worked from home at least one day a week by week. Balanced panel drops anybody that quits before the end of March 2012. Once employees quit they are dropped from the data. Individually clustered standard errors *** denotes 1% significance, ** 5% significance, and * 10% significance.

WFH, profits and productivity

- Transforming percentages to profits (13% of performance, 9.2% of wages: $\rightarrow$ 230\$ / employee / year.
- Firm estimates 1 400$ capital cost savings.
- Also, firm estimates 260$ savings / employee / year in reduced turnover costs.

WFH, profits and productivity

- What is **T**otal **F**actor **P**roductivity (TFP)?
- Let's assume a Cobb-Douglas production function:

$$Y_i = e^{\omega_i} K_i^{\alpha} L_i^{\beta}$$

- Let's take logs:

$$\ln Y_i = \omega_i + \alpha \ln K_i + \beta \ln L_i$$

- ω_i is TFP.

WFH, profits and productivity

- Two TFP measures:
  1. Using the detailed data.
  2. "commonly used" TFP using more aggregate data.
- TFP change with detailed data: 21%.
- TFP change with agg. data: 28%.

WFH summary

- 13% improvement in performance.
- 9ppc came from working more minutes / shift.
- 4ppc from taking more calls / minute.
- Increased job satisfaction, attrition halved.
- Promotion rate — performance decreased.
- Post-experiment (endogenous) sorting almost doubled the gains to 22%.