

Econometrics of Policy Evaluation: Difference-in-Differences

Cristian Huse

- We can't always randomize
 - e.g., estimating the impact of a “past” program
- We can try to find a “natural experiment” that allows us to identify the impact of a policy
 - e.g., an unexpected change in policy
 - e.g., a policy that only affects 16 year-olds but not 15 year-olds
- In general, can exploit variation of policies in time and space to identify...
 - Which is the group affected by the policy change (“treatment”); and
 - Which is the group that is not affected (“comparison” or “control”)
- As usual, the quality of the control group determines the quality of the evaluation

Motivation

- One of the most frequent causes of endogeneity in empirics is **omitted variables**
 - **Reminder:** Some variable that the econometrician does not observe but that is correlated with other explanatory variables, a variable that is jointly determined with the dependent variable...
- This happens mainly due to the considerable (unobserved) heterogeneity present in many settings
 - Individuals, firms can differ in so many ways! (think about it for one second)
- Difference-in-Differences (DD) can be a powerful tool to identify causal effects

Intuition

- DD is a **quasi-experimental** technique used to understand the effect of a sharp change in the economic environment or government policy
 - Example: Leary (2009) use changes in bank funding constraints to assess the effect of capital markets frictions on corporate capital structure decisions
 - Compares impact on firms with access (affected) pre-change to those without access (not affected) pre-change
- Used in conjunction with a **natural experiment** in which nature does the randomization for us
 - **Key:** transparent, exogenous, source of variation that determines treatment assignment (e.g. policy changes, government randomization etc.)
 - **Goal:** be as close as possible to a perfect random experiment

Real Example

- At the end of 1991, Delaware passes a law that significantly streamlines bankruptcy proceedings to make litigation less costly and time-consuming
 - Must assume that this is a random (or outcome unrelated) event, i.e. the event should not be a response to pre-existing differences between the treatment and control group
 - e.g. Delaware firms are much more likely to enter financial distress
 - Must understand what caused the event to occur – institutional knowledge
- This law change offers a potentially useful setting with which to test our hypothesized relation between bankruptcy costs and capital structure
 - How do we empirically test the relation? (three possibilities)
 - Cross-Sectional Difference After Treatment
 - Time-Series Difference Within the Treatment Group
 - DD estimator

Cross-Sectional Difference After Treatment

- **Idea:** compare the average leverage of firms registered in Delaware to that of firms registered elsewhere
- This can be accomplished via a cross-sectional regression

$$y_i = \beta_0 + \beta_1 1(treat_i) + \epsilon_i$$

where y_i = leverage for firm i
in 1992, and $1(treat_i) = 1$ if firm is registered in Delaware

- Assuming $E[\epsilon_i | 1(treat_i)] = 0$:

$$E[y_i | 1(treat_i) = 0] = \beta_0$$

$$E[y_i | 1(treat_i) = 1] = \beta_0 + \beta_1$$

thus $E[y_i | 1(treat_i) = 1] - E[y_i | 1(treat_i) = 0] = \beta_1$

- The estimate is just the difference in average leverage in 1992 for the treatment group (Delaware firms) and control group (non-Delaware firms)

Potential Concerns

- **Main concern:** The existence of other factors affecting leverage cross-sectionally
 - Econometrically, the concern lies with our assumption of $E[\epsilon_i | 1(treat_i)] = 0$ (exogeneity of the treatment)
- What could threaten this assumption and, consequentially, the internal validity of the estimate?
 - What if firms in Delaware are different (e.g. more capital intensive or more profitable) relative to firms elsewhere?
 - Problem is that firms with more physical capital tend to be more levered so our assumption is violated because capital intensity is part of ϵ and is correlated with treatment status
 - In other words, even if the law was never passed, we would expect firms in Delaware to have higher leverage than other firms because of genuine unobserved differences between firms in Delaware and elsewhere

What to Do?

- **Heterogeneity**
 - Observed (age, schooling, ...)
 - Unobserved (e.g., quality of education, motivation, temperament)
- **One Solution:** Insert control variables (e.g., net plant, property, and equipment divided by assets) in the regression
 - But, these are just proxies – account for observed heterogeneity
 - There is still some heterogeneity between treatment and control groups that is part of the error term and might be correlated with our treatment indicator
 - Unfortunately, unobserved differences can subsist (unless we can find an instrument...)

Time-Series Difference Within the Treatment Group

- **Idea:** compare the average leverage of firms registered in Delaware in 1991 to that in 1992 – avoids heterogeneous firm concern
- This can be accomplished via a two-period panel regression using only Delaware firms

$$y_{it} = \beta_0 + \beta_1 1(post_{it}) + \epsilon_{it}$$

where $1(post) = 1$ if $year = 1992$ and 0 if $year = 1991$

- Assuming $E[\epsilon_{it} | 1(post_{it})] = 0$:

$$E[y_{it} | 1(post_{it}) = 0] = \beta_0$$

$$E[y_{it} | 1(post_{it}) = 1] = \beta_0 + \beta_1$$

thus $E[y_{it} | 1(post_{it}) = 1] - E[y_{it} | 1(post_{it}) = 0] = \beta_1$

- Our estimate is just the difference in average leverage for Delaware firms in 1992 (the post-treatment era) and 1991 (the pre-treatment era)

Potential Concerns

- **Main concern:** The existence of other factors affecting leverage over time.
 - e.g. increase in the supply of credit due to financial innovation
 - Leverage would likely have increased for firms even without the passage of the law
 - Bias is positive in favor of supporting a treatment effect
- This is just another form of omitted variables bias, as before
 - We could insert control variables but difficult to measure all perfectly
- Also, bias can work in both directions
 - e.g., 1992 may be a period of tight credit, which leads to a decline in debt usage and offsetting effect

Difference-in Differences Estimator

- **Intuitive idea:** Combine the positive features of both estimators
 - Cross-sectional estimator avoids omitted unobserved common trends
 - Time-series estimator avoids omitted unobserved cross-sectional differences
- The DD estimator does precisely that!

$$y_{it} = \beta_0 + \beta_1 1(treat_{it}) + \beta_2 1(post_{it}) + \underbrace{\beta_3 1(treat_{it}) \times 1(post_{it})}_{DD\ estimator} + \epsilon_{it}$$

- We need a full panel of firms consisting of Delaware ($1(treat) = 1$) and Non- Delaware ($1(treat) = 0$) registered firms observed before ($1(post) = 0$) and after ($1(post) = 1$) the passage of the law

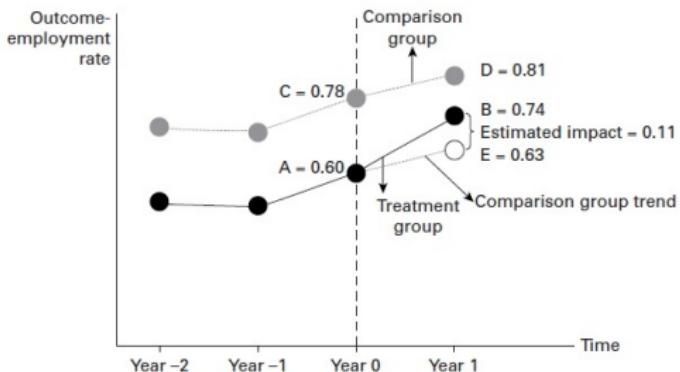
The Three Ways

- When pre-program and post-program data are available for a treatment group (TG) of program beneficiaries and a control (or comparison) group (CG) of non-beneficiaries, DD provides an alternative strategy to estimate program impacts
 - **Note:** Require a panel of units with measurements available both pre- and post-intervention
- There are three ways to look at DD
 - Graphically
 - In a box
 - Regression
- As before, the advantage of regression is that it provides standard errors, so one can readily perform inference, e.g. test hypotheses

Graphically

- Consider the following graph. What is the effect of the program?

Figure 7.1 The Difference-in-Differences Method



Note: All differences between points should be read as vertical differences in outcomes on the vertical axis.

Graphically

- Need to compute two differences:
 - $(\text{Treated in } t=1 - \text{Control in } t=1) - (\text{Treated in } t=0 - \text{Control in } t=0)$
Formally,
$$[(Y_{i,t=1}|P_i = 1) - (Y_{i,t=1}|P_i = 0)] - [(Y_{i,t=0}|P_i = 1) - (Y_{i,t=0}|P_i = 0)]$$
 or
$$[(Y_{i,t=1}|P_i = 1) - (Y_{i,t=0}|P_i = 1)] - [(Y_{i,t=1}|P_i = 0) - (Y_{i,t=0}|P_i = 0)]$$
 - $(\text{Treated in } t=1 - \text{Treated in } t=0) - (\text{Control in } t=1 - \text{Control in } t=0)$
Formally,
$$[(Y_{i,t=1}|P_i = 1) - (Y_{i,t=0}|P_i = 1)] - [(Y_{i,t=1}|P_i = 0) - (Y_{i,t=0}|P_i = 0)]$$
 - (Check for yourself that they are equivalent)
- Back to the graph, should obtain

$$\delta = (B - D) - (A - C)$$

In a Box

- Recall notation: $Y_{ti}, t = 0, 1; \bar{Y}_1 = \sum_{i=1}^{N_1} Y_{1i}$

		Groups		
		Treatment	Control	Treatment-Control
Pre	$E[Y_i P_i = 1, t = 0] = \alpha + \gamma$	$E[Y_i P_i = 0, t = 0] = \alpha$	$E[Y_i P_i = 1, t = 0] - E[Y_i P_i = 0, t = 0]$	$E[Y_i P_i = 1, t = 0] - E[Y_i P_i = 0, t = 0]$
Post	$E[Y_i P_i = 1, t = 1] = \alpha + \beta + \gamma + \delta$	$E[Y_i P_i = 0, t = 1] = \alpha + \beta$	$E[Y_i P_i = 1, t = 1] - E[Y_i P_i = 0, t = 1]$	$E[Y_i P_i = 1, t = 1] - E[Y_i P_i = 0, t = 1]$
Post	$E[Y_i P_i = 1, t = 1]$	$E[Y_i P_i = 0, t = 1]$	$E[Y_i P_i = 1, t = 1] - E[Y_i P_i = 0, t = 1]$	$E[Y_i P_i = 1, t = 1] - E[Y_i P_i = 0, t = 1]$
-	—	—	—	$E[Y_i P_i = 1, t = 0] - E[Y_i P_i = 0, t = 0]$
Pre	$E[Y_i P_i = 1, t = 0]$	$E[Y_i P_i = 0, t = 0]$		\uparrow Diff-in-diff \uparrow

$$\delta = (E[Y_i | P_i = 1, t = 1] - E[Y_i | P_i = 0, t = 1]) - (E[Y_i | P_i = 1, t = 0] - E[Y_i | P_i = 0, t = 0])$$

- (Check for yourself that they are equivalent)

Regression Framework

- Assume two time periods, $t = 0, 1$ which denote pre- and post-intervention periods, respectively
- The corresponding regression reads

$$Y_{it} = \alpha + \beta \cdot 1(t = 1) + \gamma \cdot 1(P_i = 1) + \delta \cdot 1(t = 1) \cdot 1(P_i = 1) + \epsilon_{it}$$

- Need to calculate the following

- $E(Y_{i1}|P_i = 1) =$
- $E(Y_{i0}|P_i = 1) =$
- $E(Y_{i1}|P_i = 0) =$
- $E(Y_{i0}|P_i = 0) =$
- $DD =$

$$[E(Y_{i1}|P_i = 1) - E(Y_{i0}|P_i = 1)] - [E(Y_{i1}|P_i = 0) - E(Y_{i0}|P_i = 0)] = ?$$

Regression Framework

- Recall

$$Y_{it} = \alpha + \beta \cdot 1(t=1) + \gamma \cdot 1(P_i=1) + \delta \cdot 1(t=1) \cdot 1(P_i=1) + \epsilon_{it}$$

- Need to calculate the following

- $E(Y_{i1}|P_i=1) = \alpha + \beta + \gamma + \delta$

- $E(Y_{i0}|P_i=1) = \alpha + \gamma$

- $E(Y_{i1}|P_i=0) = \alpha + \beta$

- $E(Y_{i0}|P_i=0) = \alpha$

- $DD =$

$$[E(Y_{i1}|P_i=1) - E(Y_{i0}|P_i=1)] - [E(Y_{i1}|P_i=0) - E(Y_{i0}|P_i=0)] = \delta$$

- (Again, check for yourself the calculations)

Connecting the Three Ways

- Recall notation: Y_{ti} , $t = 0, 1$; $\bar{Y}_1 = \sum_{i=1}^{N_1} Y_{1i}$

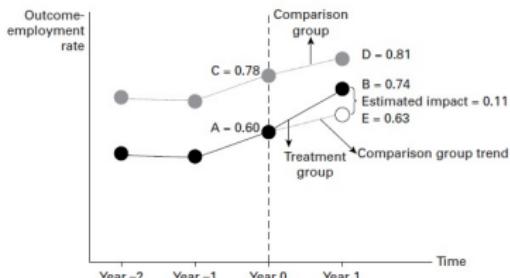
Groups			
	Treatment	Control	Treatment-Control
Pre	$E[Y_i P_i = 1, t = 0] = \alpha + \gamma$	$E[Y_i P_i = 0, t = 0] = \alpha$	γ
Post	$E[Y_i P_i = 1, t = 1] = \alpha + \beta + \gamma + \delta$	$E[Y_i P_i = 0, t = 1] = \alpha + \beta$	$\gamma + \delta$
Post	$\alpha + \beta + \gamma + \delta$	$\alpha + \beta$	$\gamma + \delta$
-	$-$	$-$	$-$
Pre	α	$\alpha + \gamma$	γ
	$\beta + \gamma + \delta$	$\beta - \gamma$	δ

$$\delta = (E[Y_i | P_i = 1, t = 1] - E[Y_i | P_i = 0, t = 1]) - (E[Y_i | P_i = 1, t = 0] - [Y_i | P_i = 0, t = 0])$$

The Core Identifying Assumption: Parallel Trends

- **Definition:** In the absence of the treatment, the average outcome for the Treatment Group would have followed the same trend as the average outcome for the Control Group (see “Comparison group trend” below).

Figure 7.1 The Difference-in-Differences Method



Note: All differences between points should be read as vertical differences in outcomes on the vertical axis.

- **Key Point:** This is an assumption about an **unobserved counterfactual**. We can never prove it, but we can test its plausibility.

A Modern Empirical Toolkit: Testing for Parallel Trends

- Plot example – raw data:



Notes: Data cover 2,673,771 unique patent classes filed by US inventors across 166 research fields defined at the level of USPTO classes. *Research fields of émigrés* cover 60 classes which include at least one patent between 1920 and 1970 by either a German or Austrian émigré to the United States. *Research fields of other German chemists* cover 106 USPTO classes which include at least one patent between 1920 and 1970 by another German chemist but include no patents by émigrés.

- **Test 1 (The Gold Standard):** Visual Event Study based on, say

$$Y_{it} = \alpha_i + \lambda_t + \sum \delta_k (1[t = k] \times 1[i \in TG]) + \theta x_{it} + \varepsilon_{it}$$

- With multiple pre-treatment periods, we can plot the “effect” in each period. **We should find that the pre-treatment coefficients are all flat and zero.**

A Modern Empirical Toolkit: Testing for Parallel Trends

- Plot example – event study:

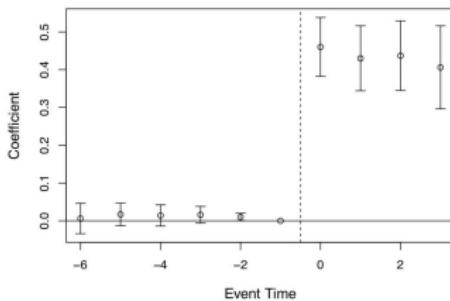


Figure 9.4: Estimates of Medicaid expansion's effects on eligibility using leads and lags in an event study model. Reprint from Miller et al. (2019).

- The bars show the effect of the interactions $1[t = k] \times 1[i \in TG]$ (interactions between time and treatment group dummies):
 - **Non-significant pre-policy** and **significant post-policy effects** support the parallel trends assumption
 - Groups differ **after** the policy (and only due to the policy!).
- This is the most important test of the parallel trends assumption.

A Modern Empirical Toolkit: Testing for Parallel Trends

- **The Problem:** OLS standard errors assume all observations are independent. In a DD, they are not:
 - **Serial Correlation:** Outcomes for the same unit over time are correlated (e.g., a firm's leverage this year is related to its leverage last year).
 - **Group Correlation:** Outcomes for units within the same group are correlated (e.g., all firms in Delaware are subject to the same state shocks).
- **Evidence:** Ignoring this leads to standard errors that are far too small, causing us to find “significant” effects that are just noise (Bertrand, Duflo, Mullainathan 2004).
- **The Solution:** Always cluster your standard errors at the level of the unit that the policy varies by (e.g., state, county, airport, school).
 - Rule of thumb: Number of cluster should be at least 40 for the asymptotics to work. Otherwise, robust standard errors.

A Modern Empirical Toolkit: Testing for Parallel Trends

- **Test 2: Placebo Tests.**

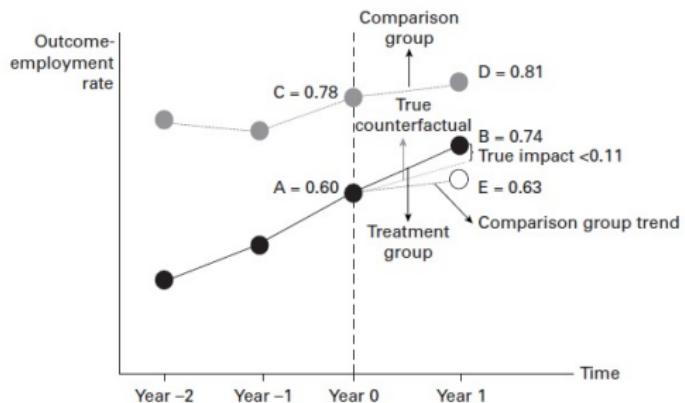
- **Placebo Outcome:** Re-run the DD on an outcome variable you know shouldn't be affected by the policy. The placebo-DD estimate should be zero.
- **Placebo Groups:** Re-run the DD using two different groups that were both untreated. The placebo-DD estimate should be zero.
- **Placebo Time:** If you have many pre-treatment periods, run a DD using only pre-treatment data (e.g., set a "fake" treatment year). The placebo-DD should be zero.

Caveats

- DD is valid only when the policy change has an **immediate** impact on the outcome variable
- If there is a delay in the impact of the policy change, we do need to use lagged treatment variables
- DD attributes any differences in trends between the treatment and control groups, that occur at the same time as the intervention, to that intervention
 - If there are other factors that affect the difference in trends between the two groups, then the estimation will be biased!
- Fixed effects control for
 - Fixed group effects
 - e.g. Farmers who own their land, farmers who don't own their land
 - Effects that are common to all groups at one particular point in time, or "common trends"
 - e.g. The 2006 drought affected all farmers, regardless of who owns the land

Violation of Common Trends Assumption

Figure 7.2 Difference-in-Differences When Outcome Trends Differ



DD in a Regression Framework

Stata Example 8. Difference-in-Differences in a Regression Framework

* In this method, you compare the change in health expenditures over time
 * between enrolled and nonenrolled households in the treatment localities.

*Select the relevant data

```
use "evaluation.dta", clear
keep if treatment_locality==1
```

```
gen eligible_round=eligible*round
```

```
reg health_expenditures eligible_round round eligible, cl(locality_identifier)
```

Linear regression

Number of obs =	9919
F(3, 99) =	813.98
Prob > F =	0.0000
R-squared =	0.3436
Root MSE =	7.9128

(Std. Err. adjusted for 100 clusters in locality_identifier)

	Robust					
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
health_expen~s						
eligible_round	-8.162931	.3191368	-25.58	0.000	-8.796168	-7.529695
round	1.513416	.3564533	4.25	0.000	.8061355	2.220697
eligible	-6.3018	.193082	-32.64	0.000	-6.684917	-5.918684
_cons	20.79149	.1722887	120.68	0.000	20.44964	21.13335

- Note: round = treatment period, eligible = treatment group

DD in a Multivariate Regression Framework

Stata Example 9. Difference-in-Differences in a Multivariate Regression Framework

```
reg health_expenditures eligible_round round eligible age_hh age_sp educ_hh educ_sp
female_hh indigenous hhszie dirlfloor bathroom land hospital_distance,
cl(locality_identifier)
```

Linear regression

Number of obs = 9919
F(14, 99) = 2410.28
Prob > F = 0.0000
R-squared = 0.5516
Root MSE = 6.5437

(Std. Err. adjusted for 100 clusters in locality_identifier)

		Robust				
		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
health_expendits						
eligible_round		-8.161499	.3197482	-25.52	0.000	-8.795949 -7.527049
round		1.450526	.3558662	4.08	0.000	.74441 2.156641
eligible		-1.51276	.129937	-11.64	0.000	-1.770583 -1.254937
age_hh		.0804852	.0113711	7.08	0.000	.0579224 .103048
age_sp		-.0197229	.0129787	-1.52	0.132	-.0454754 .0060297
educ_hh		.0599944	.0290694	2.06	0.042	.0023144 .1176743
educ_sp		-.0765127	.0339694	-2.25	0.027	-.1439153 -.0091101
female_hh		1.103935	.3157136	3.50	0.001	.4774905 1.730379
indigenous		-2.311985	.2361846	-9.79	0.000	-2.780627 -1.843344
hhszie		-1.994729	.0391445	-50.96	0.000	-2.0724 -1.917058
dirlfloor		-2.299839	.1632436	-14.09	0.000	-2.62375 -1.975929
bathroom		.5000436	.157629	3.17	0.002	.1872735 .8128137
land		.0909001	.028528	3.19	0.002	.0342943 .1475058
hospital_distance		-.0031917	.0030591	-1.04	0.299	-.0092617 .0028783
_cons		27.39458	.5526554	49.57	0.000	26.29799 28.49117

DD Alternatives

Stata Example 10. Calculating Difference-in-Difference Estimates by Taking the Difference between Before-After Differences in the Treatment and Comparison Groups

```

xtset household_identifier round
    panel variable: household_identifier (unbalanced)
    time variable: round, 0 to 1
    delta: 1 unit

gen xtenrolled=0

replace xtenrolled=1 if enrolled==1 & round==1

xtreg health_expenditures xtenrolled round if treatment_locality==1, fe vce(cluster
locality_identifier)

Fixed-effects (within) regression                               Number of obs      =      9919
Group variable: household_~r                                Number of groups   =      4960

R-sq:  within = 0.2437                                         Obs per group: min =         1
          between = 0.3852                                         avg =        2.0
          overall = 0.2805                                         max =         2

                                                F(2,99)           =     633.70
corr(u_i, Xb)  = 0.2632                                         Prob > F        = 0.0000

                                                (Std. Err. adjusted for 100 clusters in locality_identifier)
-----
|           Robust
health_exp~s |   Coef.  Std. Err.      t    P>|t|   [95% Conf. Interval]
-----+
xtenrolled | -8.163337  .3190874   -25.58  0.000   -8.796476  -7.530198
round |  1.513416  .3564353    4.25  0.000   .8061711   2.220661
_cons | 17.02477  .1203165   141.50  0.000   16.78603   17.2635
-----+
sigma_u | 7.1354916
sigma_e | 6.5169842
rho | .54521089  (fraction of variance due to u_i)
-----+

```

DD Alternatives II

```

Stata Example 11. Single-Differences Estimates for Difference-in-Differences

keep health_expenditures treatment_locality locality_identifier enrolled
household_identifier round

reshape wide health_expenditures enrolled, i(household_identifier) j(round)
(note: j = 0 1)

Data                                long    ->    wide
-----
Number of obs.                      9919   ->    4960
Number of variables                  6       ->    7
j variable (2 values)                round   ->    (dropped)
xij variables:
          health_expenditures      ->    health_expenditures0
health_expenditures1                 enrolled  ->    enrolled0 enrolled1
-----
gen dy = health_expenditures1 - health_expenditures0

replace enrolled0=0

gen dp = enrolled1-enrolled0

reg dy dp if treatment_locality==1, cl(locality_identifier)

Linear regression
                                         Number of obs =     4959
                                         F( 1,    99) =  654.51
                                         Prob > F = 0.0000
                                         R-squared = 0.1588
                                         Root MSE = 9.2164

                                         (Std. Err. adjusted for 100 clusters in locality_identifier)
-----
          |      Robust
          |      Coef.  Std. Err.      t  P>|t|  [95% Conf. Interval]
-----+
dp |  -8.163337  .3190874  -25.58  0.000  -8.796476  -7.530198
_cons |  1.513416  .3564353  4.25  0.000   .8061711  2.220661
-----+

```

- Note:

- There are several implementations in R and Stata

Overview

- Paper: Cawley, Willage, and Frisvold (2018). Pass-Through of a Tax on Sugar-Sweetened Beverages at the Philadelphia International Airport. JAMA
 - **Research question:** Effect of tax increase on sugar-sweetened beverages in Philadelphia raised retail prices of soft drinks at the Philadelphia International Airport 1 month after implementation
 - **Motivation:** Health effects of sugar ingestion, effectiveness of tax
 - **Empirical strategy:** The Philadelphia International Airport straddles the city border, with some terminals in Philadelphia that are subject to the beverage tax, and other terminals in Tinicum that are not. Retailers in the airport were visited once before the tax (December 21, 2016) and twice after the tax took effect (January 14, 2017, and February 5, 2017).

TG	CG
----	----
- **Results:**
 - Stores on both sides raised prices, but pass-through was incomplete
 - Stores where the tax actually increased responded more strongly, but the “competing stores” also responded

Results

- Motivation here is to show how this can be just a simple exercise
 - Find research question
 - Collect data
 - Estimate (“box framework” + standard errors)

Table 1.

Mean Price and Mean Change in Price of Sugar-Sweetened Beverages (SSBs) at the Tinicum Side (Untaxed) vs Philadelphia Side (Taxed) of the Philadelphia International Airport^a

Time Point	Tinicum Side Price		Philadelphia Side Price		Difference in Mean Change (95% CI), ¢/oz
	Mean (95% CI), ¢/oz	Mean Change vs December 2016 (n = 10 Stores)	Mean (95% CI), ¢/oz	Mean Change vs December 2016 (n = 21 Stores)	
Before new tax on SSBs					
December 2016	12.37 (10.83-13.91)		12.53 (11.80-13.25)		
After new tax on SSBs					
January 2017	12.78 (11.07-14.48)	0.41 (-0.08 to 0.89)	13.44 (12.59-14.29)	0.91 (0.60 to 1.23)	0.51 (-0.01 to 1.03)
February 2017	12.93 (11.21-14.64)	0.56 (0.03 to 1.09)	13.92 (13.18-14.66)	1.39 (1.20 to 1.58)	0.83 (0.33 to 1.33)

^aStandard errors are clustered at store location level.

Overview

- Paper: Bollinger, Leslie, and Sorensen (2011). Calorie Posting in Chain Restaurants. American Economic Journal: Economic Policy.
- Research question: Effect of mandatory calorie posting on consumers' purchase decisions and company profits (Starbucks)
- Motivation: Obesity is a major health issue, regulators passed a law requiring nutrition labeling
- Empirical challenge: Getting the data, identifying control group
- Empirical strategy: Use passing of the law and a DD methodology to make a causal statement on the effects of requiring nutrition labeling
- Results:
 - Average calories per transaction fall by 6 percent, mostly due to changes in consumers' food (rather than beverages) choices
 - On average, no impact on Starbucks profit
 - For the subset of stores located close to their competitor Dunkin Donuts, the effect of calorie posting is actually to increase Starbucks revenue

Institutional Background

- Between 1995 and 2008, the fraction of Americans who were obese rose from 15.9 percent to 26.6 percent and, according to the OECD, the United States is the most obese nation in the world
- Policy: Mandatory posting of calories on menus in chain restaurants – implemented in New York City (NYC) from 1 April, 2008
- Aim: measure the effect of the NYC law on consumers' caloric purchases, and analyze the mechanism underlying the effect
 - On the one hand it may seem obvious that increasing the provision of nutrition information to consumers would help them to purchase healthier food
 - On the other hand, consumers at chain restaurants (especially fast food chains) may care mostly about convenience, price, and taste, with calories being relatively unimportant. Consumers who do care about calories may already be well-informed, since calorie information is already widely available

Data

- Two datasets
 - Observe every transaction at Starbucks company stores in NYC from January 1, 2008 to February 28, 2009. Plus every transaction at Starbucks company stores in Boston and Philadelphia, where there was no calorie posting
 - A large sample Starbucks cardholders (inside and outside of NYC) tracked over the same period of time
 - Allowing to examine the impact of calorie posting at the individual level
 - *What is the difference between these datasets? What about selection?*
- Plus in-store customer surveys performed before and after the introduction of a calorie posting law in Seattle on January 1, 2009
 - Provide evidence about how knowledgeable people were about calories at Starbucks before and after the law change
 - Survey of consumers at the same points in time in control locations where there was no calorie posting

Empirical Specification

- DD specification reads

$$y_{sct} = x_{sct}\beta + \gamma POST_{ct} + \varepsilon_{sct}$$

where y_{sct} is a measure of calories per transaction at store s in city c on day t , $POST_{ct}$ is a dummy equal to one if calories were posted (i.e., NYC stores after April 1, 2008), and x_{sct} includes week FEs (to control for seasonality), day-of-week FEs, holiday FEs, temperature (also squared), and precipitation (also squared)

- Why weather controls?

Empirical Specification

- DD specification reads

$$y_{sct} = x_{sct}\beta + \gamma POST_{ct} + \varepsilon_{sct}$$

where y_{sct} is a measure of calories per transaction at store s in city c on day t , $POST_{ct}$ is a dummy equal to one if calories were posted (i.e., NYC stores after April 1, 2008), and x_{sct} includes week FEs (to control for seasonality), day-of-week FEs, holiday FEs, temperature (also squared), and precipitation (also squared)

- Why weather controls?
Important determinants of beverage demand, weather may vary between the three cities in the analysis on any given day
- Estimate separate versions using transaction and cardholder data
 - Regressions using the transaction data include store FEs, and regressions using the cardholder data include individual FEs
 - In both cases, identification of the effect of calorie posting stems from within-city variation over time

Summary Statistics

- Comparison of transaction vs. cardholder data, and treatment vs. control groups

TABLE 1—SUMMARY STATISTICS FOR TRANSACTION DATA AND CARDHOLDER DATA
(*Prior to policy change*)

	Transaction data		Cardholder data	
	New York City	Boston & Philadelphia	New York City	Boston & Philadelphia
Avg. weekly transactions per store	1.00	0.77	1.00	1.90
Avg. weekly revenue per store	1.00	0.74	1.00	1.87
Percent transactions with brewed coffee	1.00	1.00	1.00	0.80
Percent transactions with beverage	1.00	1.01	1.00	0.98
Percent transactions with food	1.00	0.96	1.00	1.06
Avg. num. items per transaction	1.00	0.99	1.00	1.01
Avg. num. drink items per transaction	1.00	1.00	1.00	1.01
Avg. num. food items per transaction	1.00	0.93	1.00	1.05
Food attach rate	1.00	1.00	1.00	1.00
Avg. dollars per transaction	1.00	0.94	1.00	0.97
Avg. calories per transaction	1.00	1.03	1.00	1.14
Avg. drink calories per transaction	1.00	1.09	1.00	1.23
Avg. food calories per transaction	1.00	0.94	1.00	0.99

Notes: Variables have been normalized (first and third columns equal 1.00) to preserve confidentiality of the data. All statistics are based on data prior to calorie posting in NYC (April 1, 2008). “Brewed coffee” does not include barista-made beverages (such as a cafe latte). “Food attach rate” is defined as the probability of purchasing a food item conditional on purchasing a beverage. The statistics related to calories (the bottom three rows) are based only on transactions with at least one beverage or food item.

TABLE 2—CHANGES IN CARDHOLDERS’ BEVERAGE CHOICES FOLLOWING MANDATORY CALORIE POSTING (*Treatment and control results shown separately*)

	Smaller size	Same size	Larger size	Total
Lower calories per ounce	2.26	7.08	2.27	11.61
	1.87	9.44	1.76	13.06
Same calories per ounce	4.51	67.57	4.12	76.20
	4.39	66.74	3.95	75.08
Higher calories per ounce	1.59	6.54	4.05	12.19
	1.54	6.92	3.40	11.86
Total	8.36	81.20	10.45	100.00
	7.79	83.20	9.11	100.00

Notes: Based on cardholder dataset. Entries are based on each individuals most common beverage choice before versus after calorie posting in NYC. The top entry in each cell relates to individuals in NYC, and the bottom entry in each cell relates to individuals in Boston and Philadelphia. For example, 2.26 percent of individuals in NYC changed their beverage choice (following calorie posting) to a beverage that has lower calories per ounce and also a smaller size. The number for Boston and Philadelphia is 1.87 percent. Pearson’s chi-square test fails to reject that the cell proportions for NYC are equal to those for Boston and Philadelphia (p -value of 0.11).

Common Trends?

- No apparent difference pre-policy

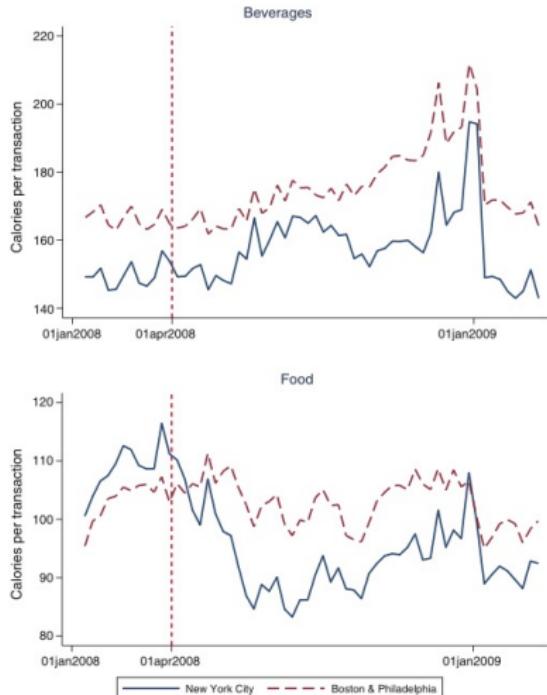


FIGURE 1. CALORIES PER TRANSACTION

Main Results

- Reduction in calories occurs mostly via food instead of beverages

TABLE 3—ESTIMATES OF THE EFFECT OF MANDATORY CALORIE POSTING ON LOG
(Calories per transaction)

	Transaction data	Cardholder data
log (beverage calories)	-0.003*** (0.001)	0.008 (0.005)
log (food calories)	-0.147*** (0.002)	-0.119*** (0.008)
log (beverages + food)	-0.060*** (0.001)	-0.051*** (0.005)
Observations	118,480	1,511,516

Notes: Each reported coefficient estimate is obtained from a separate regression. The rows represent different dependent variables and the columns correspond to the transaction data and the cardholder data, respectively. An observation in the transaction data regressions is a store-day combination. An observation in the cardholder data regressions is a cardholder transaction. We exclude transactions that do not include at least one beverage or food item. All regressions include week fixed effects, day-of-week fixed effects, weather controls (temperature, temperature-squared, precipitation, and precipitation-squared). Additionally, regressions using the transaction data include store fixed effects, and regressions using the cardholder data include individual fixed effects. In the first column, the R^2 ranges from 0.73 to 0.85, and in the second column the R^2 ranges from 0.27 to 0.64.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Substitution Effects

- Figure suggests substitution from high to low calorie items

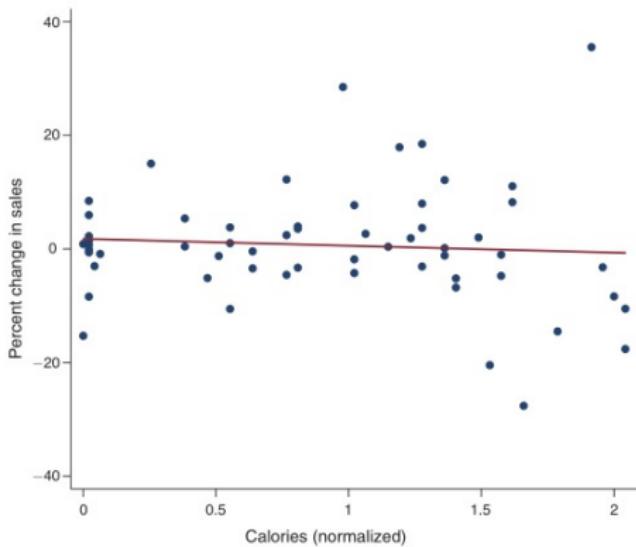


FIGURE 3. ESTIMATED SALES CHANGES FOR 60 MOST POPULAR MENU ITEMS

Notes: For each of the 60 most popular menu items we separately ran regressions of log daily sales on an indicator for calorie posting, plus store, week, and day-of-week fixed effects, holiday dummies, and weather controls. Transactions for the control cities were included to control for seasonal variation in product demand. We plot each coefficient on the calorie posting variable, against the normalized calories of the menu item. Calories are normalized in the figure to preserve product-level confidentiality of the data.

Differential Responses: Food vs. Beverages

- Reduction in items purchased and food items vs. increase in beverage items
- Reduction in calories from food items

TABLE 4—ESTIMATES OF THE EFFECT OF CALORIE POSTING ON ITEMS PER TRANSACTION AND CALORIES PER SINGLE BEVERAGE OR SINGLE FOOD ITEM TRANSACTION

	Items per transaction	
	Transaction data	Cardholder data
Number of beverages	0.005*** (0.001)	-0.002 (0.002)
Number of food items	-0.029*** (0.001)	-0.021*** (0.002)
Beverages + food items	-0.027*** (0.003)	-0.017*** (0.003)
Calories per item purchased		
	Calories per item purchased	
	Transaction data	Cardholder data
log (beverage calories per beverage)	-0.008*** (0.001)	0.004 (0.004)
log (food calories per food item)	-0.039*** (0.001)	-0.165*** (0.014)

Notes: Each reported coefficient estimate in this table is obtained from a separate regression. All specifications include the same controls as in Table 3. In the top panel (items per transaction), we utilize 118,480 store-day combinations for the regressions in the transaction data column, and we obtain R^2 's ranging from 0.27 to 0.82. The regressions using the cardholder data in the top panel are based on 1,511,516 observations, and the R^2 's vary between 0.26 and 0.37. In the bottom panel, examining log(calories per item purchased), we condition the sample on transactions with at least one beverage (second to bottom row) or at least one food item (bottom row). In the transaction data column an observation is a store-day combination, and the number of observations is 118,480 in both cases (R^2 's are 0.83 and 0.64, respectively). In the cardholder data column in the bottom panel there are 1,486,839 observations of transactions with at least one beverage and 233,575 observations of transactions with at least one food item. The R^2 's in these regressions are 0.70 and 0.33, respectively.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Heterogeneity

- Stronger effects for wealthy, better-educated, females, medium/high calorie consumers

TABLE 5—HETEROGENEITY IN THE IMPACT OF CALORIE POSTING ON LOG
(*Calories per transaction*)

	(1)	(2)	(3)	(4)
Posting	-0.102*** (0.011)	-0.032*** (0.006)	-0.058*** (0.007)	0.147*** (0.006)
Posting × median income (in \$100,000)	-0.012** (0.006)			
Posting × percent with college degree	-0.020** (0.010)			
Posting × percent aged 20–45	0.001 (0.001)			
Posting × percent female	-0.001 (0.001)	-0.049*** (0.006)		
Posting × high frequency customer			0.011 (0.007)	
Posting × medium caloric customer				-0.298*** (0.008)
Posting × high caloric customer				-0.444*** (0.007)
Observations	94,997	1,511,516	1,511,516	1,511,516
R ²	0.81	0.56	0.56	0.56
Transaction data	Yes	No	No	No
Cardholder data	No	Yes	Yes	Yes

Notes: Each column is a separate regression. In all cases the dependent variable is log (calories per transaction). Regressions based on the transaction data also include store fixed effects, week and day of week dummies, and weather controls. Regressions based on the cardholder data also include individual fixed effects, week dummies and weather controls. In the last column, “medium caloric” customers are defined as customers for whom average calories per transaction in the pre-calorie-posting period was between 125–250. “High caloric” customers had average calories per transaction above 250.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Effect on Revenues

- No significant effect on revenue – but this is a net effect of several components!
- Regulation results in Starbucks poaching consumers from DunDon

TABLE 7—EFFECT OF MANDATORY CALORIE POSTING ON REVENUES

	log (daily store revenue)		log (daily store transactions)		log (drink revenue)		log (food revenue)	
	(1)	(2)	(3)	(4)	(5)	(6)		
Calorie posting	0.005 (0.004)	-0.000 (0.004)	0.014*** (0.004)	0.009** (0.004)	0.011*** (0.004)		-0.074*** (0.004)	
Posting × Dunkin Donuts nearby		0.033*** (0.006)		0.032*** (0.006)	0.038*** (0.006)		0.020*** (0.005)	
R ²	0.71	0.71	0.72	0.72	0.70		0.76	

Notes: Each column is a separate regression with dependent variables as specified at the top of each pair of columns. All regressions are based on the transaction data and include store, week and day of week fixed effects and weather controls. “Dunkin Donuts nearby” is a dummy equaling one if there is a Dunkin Donuts located within 100 meters of each Starbucks. There are 118,480 observations in each regression.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Commuters vs. Non-commuters

- Robust policy effects, even outside of NYC

TABLE 9—EFFECTS OF CALORIE POSTING ON COMMUTERS' LOG
(*Calories per transaction*)

	(1)	(2)
Non-commuters:		
NYC store after 01April08	-0.060*** (0.006)	-0.060*** (0.006)
Commuters:		
NYC store after 01April08	-0.077*** (0.011)	-0.077*** (0.011)
Non-NYC store after 01April08	-0.120* (0.067)	
No prior visits to posting stores		-0.015 (0.116)
One or more visits to posting stores		-0.124* (0.068)
Non-NYC store before 01April08	0.238*** (0.061)	0.238*** (0.061)
Observations	1,470,095	1,470,095
R ²	0.56	0.56

Notes: The regressions are based on the cardholder data, and include individual, week, and day-of-week fixed effects, and weather controls. An observation is a transaction, and the dependent variable is log(calories + 1). Robust standard errors in parentheses. Commuters are defined as cardholders whose visits to NYC stores comprised between 20 percent and 80 percent of their total visits.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Overview

- **Paper:** Auffhammer and Kellogg (2011). Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality. *American Economic Review*
 - Research question: Effect of US gasoline content regulations on ozone pollution
 - Motivation: Ozone has been linked to asthma, increased susceptibility to pneumonia and bronchitis, and damage to crops and natural vegetation; even marginal short-term changes in ozone concentrations can have substantial human mortality impacts
 - Empirical strategy: Use introduction of regulations and a DD methodology
 - Results:
 - Flexible federal gasoline standards, which allow refiners in choosing a compliance mechanism, did not improve air quality (Intuition: refiners minimize costs, no impact on air quality)
 - Inflexible Californian regulations, which require the removal of particularly harmful compounds significantly improved air quality – reduction in ground-level ozone concentrations by 16 percent in the severely polluted Los Angeles-San Diego area

Institutional Background

- **Policies:** Restrictions on the chemical composition of gasoline primarily intended to reduce VOC emissions from mobile sources
 - Substantial geographic variation in regulation: EPA tightness varies across states/counties, some local governments have implemented their own more stringent regulations
- Key policies of interest
 - RVP: Reid vapor pressure regulation (# refers to the VP limit)
 - RFG: Federal reformulated gasoline
 - CARB RFG: California Air Resources Board reformulated gasoline

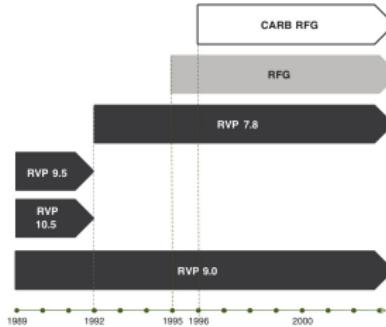


FIGURE 1. REGULATORY TIMELINE

Notes: RVP: Reid vapor pressure regulation (the number refers to the vapor pressure limit). RFG: Federal reformulated gasoline. CARB RFG: California Air Resources Board reformulated gasoline.

Institutional Background

- Cross-sectional variation in regulations (2006)

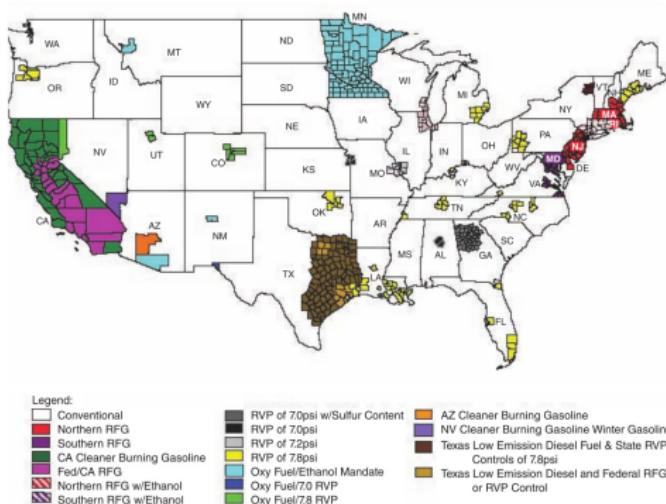


FIGURE 2. MAP OF RVP PHASE II AND RFG REGULATIONS AS OF 2006

Notes: Unshaded "conventional" gasoline areas are subject to the summertime RVP Phase II standard of 9.0 psi. Shaded areas in Minnesota, Colorado, Utah, and Montana have oxygenated gasoline for control of carbon monoxide pollution but do not have RVP or RFG. This study does not evaluate the effect of oxygenates on carbon monoxide pollution.

Source: EPA (Dec. 2006)

Data

- Data on ambient air concentrations of ozone from the EPA's Air Quality Standards database for 1989–2003
 - Hourly readings from the EPA's network of air quality monitors
 - Construct the daily maximum concentration and the daily eight-hour maximum
- Weather data measurements from the National Climatic Data Center's Cooperative Station Data (NOAA 2008), which provide daily minimum and maximum temperatures, rain, and snowfall. To assign to each air monitoring station the corresponding weather data...
 - ① Identify the ten closest weather stations to each pollution monitor, provided that each is less than 50 miles from the monitor and the elevation difference between the monitor and the station is less than 500 vertical feet
 - ② Among these, identify the closest station for which 50 percent of the pollution monitor's daily readings can be matched to the station's weather data. We then match the four climate variables for this station to the time series of ozone measurements

Data

- “Raw” data – Common trends?

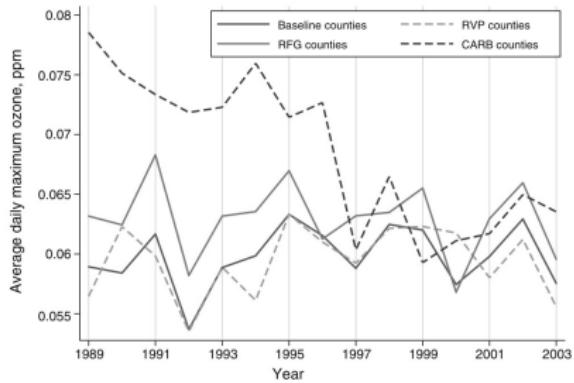


FIGURE 3. MEAN SUMMER OZONE CONCENTRATIONS
(Broken out by regulation type)

Note: For both Figures 3 and 4 data include only those monitors recording data in every summer.

Empirical Specification

- Basic DD specification takes the form

$$\ln(y_{it}) = \alpha \times \mathbf{Treat}_{ct} + \mu_i + \eta_{ry} + \varepsilon_{it}$$

where y_{it} is a measure of air pollution at monitor i on date t ,
 \mathbf{Treat}_{ct} is a vector of four variables indicating whether the county c in which monitor i is located is subject to one of four possible regulatory treatments at time t (excluding the baseline RVP standard), α is a four-element vector of parameters whose estimation is of primary interest, plus monitor FEs (μ_i), interaction of the four US census regions r with each year y (η_{ry})

- Identifying assumption: $E[\mathbf{Treat}_{ct}\varepsilon_{it} | \mu_i, \eta_{ry}] = 0$

Empirical Specification

- Generalized version given by

$$\ln(y_{it}) = \alpha \times \mathbf{Treat}_{ct} + \beta \mathbf{W}_{it} + \gamma_r \mathbf{D}_t + \delta I_{ct} + \theta \cdot \mathbf{Trend}_{rct} + \mu_i + \eta_{ry} + \varepsilon_{it}$$

where \mathbf{W}_{it} control for monitor-specific weather shocks and include a flexible polynomial in temperature and precipitation, as well as interactions of these variables with day-of-year and day-of-week; \mathbf{D}_t denotes a vector consisting of six dummy variables for day-of-week and a day-of-year variable; I_{ct} denotes county-level total annual personal income; \mathbf{Trend}_{rct} are linear time trends that are specific to treated and control counties within each census region

- Standard errors clustered on each state-year combination

Descriptives

- **Note:** Plots below are residuals obtained after controlling for weather, adding fixed-effects etc (see regressions above)
 - Authors aim to get common trends – successful?

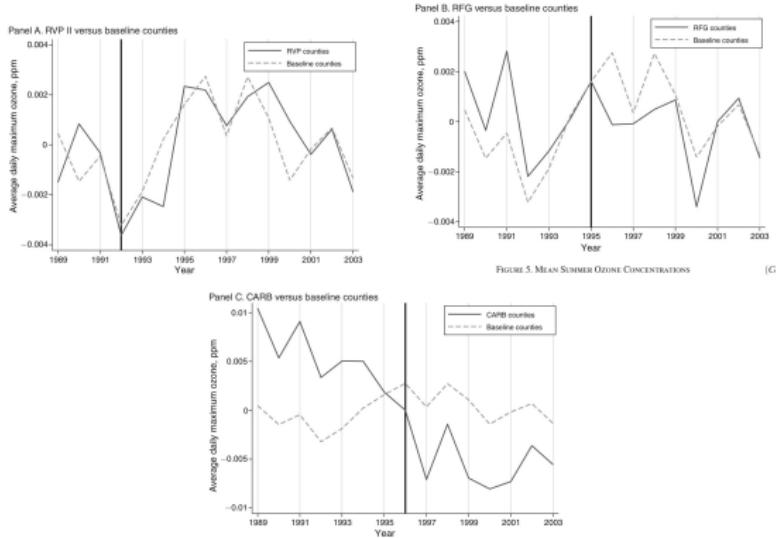


FIGURE 5. MEAN SUMMER OZONE CONCENTRATIONS
(Concluded)

Notes: Values plotted are averaged residuals of a regression of daily maximum ozone on weather variables W_t and D_t described in Section IIIA and monthly fixed effects. Solid lines are TREATED counties; dashed lines are BASELINE counties. Vertical bars indicate the first implementation of the indicated regulation. Data include only those months recording data in every summer.

Main Results: Baseline

TABLE 2—DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS

Regressand	Dependent var: ln(daily maximum ozone concentration)					ln(daily max 8 hour concentration)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RVP Phase I: 9.5 or 10.5 psi	0.016 (0.016)	0.013 (0.015)	0.015 (0.016)	0.001 (0.016)	0.004 (0.018)	0.018 (0.017)	0.015 (0.017)	0.004 (0.020)
RVP Phase II: 7.8 psi or lower	-0.007 (0.008)	-0.011 (0.007)	-0.007 (0.007)	-0.012 (0.009)	-0.012 (0.011)	-0.005 (0.008)	-0.009 (0.007)	-0.011 (0.012)
Federal RFG	-0.029 (0.009)***	-0.030 (0.007)***	-0.016 (0.008)**	-0.036 (0.011)***	-0.019 (0.012)	-0.028 (0.009)***	-0.028 (0.008)***	-0.022 (0.013)*
CARB gasoline	-0.095 (0.013)***	-0.090 (0.011)***	-0.077 (0.011)***	-0.065 (0.019)***	-0.064 (0.020)***	-0.090 (0.013)***	-0.086 (0.012)***	-0.063 (0.021)***
County income (\$ billion)	— —	— (0.285)***	-1.379 (0.245)	-0.225 (0.236)	-0.302 —	— —	— —	-0.102 (0.246)
Monitor FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region—year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region—DOW FEs	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Region FE— DOY interaction	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Weather controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Income	No	No	Yes	Yes	Yes	No	No	Yes
Regulation— region trends	No	No	No	Yes	Yes	No	No	Yes
Regulation— region quad trends	No	No	No	No	Yes	No	No	Yes
Observations	1,144,025	1,144,025	1,144,025	1,144,025	1,144,025	1,144,025	1,144,025	1,144,025
R ² (within-monitor)	0.024	0.261	0.261	0.264	0.264	0.026	0.255	0.259

Notes: Values shown are the coefficients of OLS regressions of the indicated dependent variable on the regressands. Standard errors clustered by state-year are in parentheses. All regulatory effects are relative to the omitted baseline of a 9.0 psi RVP standard. Sample uses data from June–August each year.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

- Insignificant effects for RVPs vs. significant for both Federal RFG and CARB, which happens to be the strongest

Main Results: Robustness

TABLE 3—DIFFERENCE-IN-DIFFERENCE ESTIMATION RESULTS: URBAN VERSUS SUBURBAN VERSUS RURAL

Regressand	Dependent var: ln(daily maximum ozone concentration)					
	Urban		Suburban		Rural	
	(1)	(2)	(3)	(4)	(5)	(6)
RVP Phase I: 9.5 or 10.5 psi	0.020 (0.021)	0.007 (0.021)	0.024 (0.018)	-0.001 (0.016)	0.011 (0.018)	5.0E-04 (0.022)
RVP Phase II: 7.8 psi or lower	0.002 (0.014)	0.002 (0.014)	-0.011 (0.009)	-0.020 (0.011)*	-0.006 (0.009)	-0.011 (0.011)
Federal RFG	-0.004 (0.013)	-0.032 (0.015)**	-0.025 (0.009)***	-0.049 (0.014)***	-0.018 (0.011)	-0.027 (0.013)**
CARB gasoline	-0.071 (0.016)***	-0.072 (0.025)***	-0.113 (0.014)***	-0.102 (0.022)***	-0.040 (0.013)***	-0.017 (0.024)
County income (\$ billion)	-1.236 (0.397)***	0.470 (0.428)	-1.614 (0.268)***	-0.727 (0.219)***	-1.145 (0.629)*	0.051 (0.841)
Monitor FEes	Yes	Yes	Yes	Yes	Yes	Yes
Region—year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Region—DOW FEs	Yes	Yes	Yes	Yes	Yes	Yes
Region FE—DOY interaction	Yes	Yes	Yes	Yes	Yes	Yes
Weather controls	Yes	Yes	Yes	Yes	Yes	Yes
Income	Yes	Yes	Yes	Yes	Yes	Yes
Regulation–region trends	No	Yes	No	Yes	No	Yes
Observations	222,982	222,982	490,539	490,539	430,504	430,504
R ² (within-monitor)	0.281	0.285	0.275	0.278	0.243	0.244

Notes: Values shown are the coefficients of OLS regressions of the indicated dependent variable on the regressands. Standard errors clustered by state-year are in parentheses. All regulatory effects are relative to the omitted baseline of a 9.0 psi RVP standard. Sample uses data from June through August each year.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

Main Results: Robustness

TABLE 4—DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS: MONITORS RECORDING DATA IN EVERY YEAR

Regressand	Dependent var: ln (daily maximum ozone concentration)					ln (daily max 8 hour concentration)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RVP Phase I: 9.5 or 10.5 psi	-0.009 (0.015)	-0.005 (0.015)	-0.005 (0.016)	-0.019 (0.016)	-0.014 (0.018)	-0.007 (0.016)	-0.003 (0.017)	-0.015 (0.020)
RVP Phase II: 7.8 psi or lower	-0.009 (0.009)	-0.014 (0.008)*	-0.009 (0.008)	-0.025 (0.010)**	-0.022 (0.012)*	-0.009 (0.009)	-0.013 (0.009)	-0.023 (0.013)*
Federal RFG	-0.031 (0.010)***	-0.034 (0.008)***	-0.018 (0.009)**	-0.055 (0.013)***	-0.031 (0.015)**	-0.031 (0.010)***	-0.034 (0.009)***	-0.036 (0.016)**
CARB gasoline	-0.148 (0.014)***	-0.144 (0.013)***	-0.118 (0.013)***	-0.134 (0.023)***	-0.123 (0.024)***	-0.139 (0.014)***	-0.137 (0.014)***	-0.121 (0.025)***
County income (\$ billion)	— —	— —	-1.851 (0.326)***	-0.289 (0.265)	-0.414 (0.247)*	— —	— —	-0.201 (0.270)
Monitor FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region—year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region—DOW FEs	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Region FE—DOY interaction	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Weather controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Income	No	No	Yes	Yes	Yes	No	No	Yes
Regulation— region trends	No	No	No	Yes	Yes	No	No	Yes
Regulation— region quad trends	No	No	No	No	Yes	No	No	Yes
Observations	455,084	455,084	455,084	455,084	455,084	455,084	455,084	455,084
R^2 (within-monitor)	0.031	0.278	0.280	0.285	0.285	0.030	0.270	0.280

Notes: Values shown are the coefficients of OLS regressions of the indicated dependent variable on the regressands. Standard errors clustered by state-year are in parentheses. All regulatory effects are relative to the omitted baseline of a 9.0 psi RVP standard. Sample uses data from June–August each year.

***Significant at the 1 percent level.

*Significant at the 5 percent level.

*Significant at the 10 percent level.

Wrapping-up

- The bar in terms of robustness is much higher nowadays than when the paper was published
 - Testing for common trends
 - Placebo tests
- Nevertheless, an interesting and important paper!

Advanced Topic: Staggered Treatment Timing

- In many cases, units are not treated at the same time. The Giroud (2013) paper below is a perfect example: new airline routes (the “treatment”) were introduced in different years.
- The “old” way to estimate this was the standard DD, also known as the Two-Way Fixed Effects (TWFE) estimator:

$$Y_{it} = \alpha_i + \gamma_t + \delta \times Treat_{it} + \varepsilon_{it}$$

- **The Problem:** Recent econometrics research (e.g., Goodman-Bacon, 2021; Callaway & Sant’Anna, 2021; Sun & Abraham, 2021) shows that this δ can be severely biased when treatment effects are heterogeneous. It gets “contaminated” by comparing early-treated units to late-treated units (who are not a valid control group).
- **The Solution:** Do not run a single TWFE regression. The modern, robust approach for **staggered interventions** is the **dynamic event study**, as shown in Giroud’s paper.

Overview

- **Paper:** Giroud (2013). Proximity and Investment: Evidence from Plant-Level Data. *Quarterly Journal of Economics*
 - Research question: How does proximity to headquarter affect plants' investment and productivity?
 - Motivation: No existing evidence on the role of proximity for firms internal decisions (evidence for arms-length relation such as banking or mutual fund investment)
 - Empirical Strategy: Use introduction of new airline routes as exogenous reduction in travel time, and thus increase in proximity
 - Results: Proximity matters! Higher investment and higher TFP!
- **Note:** This paper has lots of details and requires very careful reading; it is included here as inspiration due to the combination of research question, data, methodology
- **Note:** TFP (total factor productivity)
 - Portion of output not explained by traditionally measured inputs of labor and capital used in production
 - Calculated by dividing output by the weighted average of labor and capital input (weights 0.3 and 0.7, respectively)
 - If all inputs are accounted for, then TFP can be taken as a measure of a long-term technological change or technological dynamism

Data

- Plant-level data
 - Manufacturing plant data from US Census (1977-2005)
 - Identification of HQ and all the plants of a given firm
 - Single- and multi-units firms
 - Addresses of all plants (establishments)
 - 1.3 million plant-year observations
- Air travel data
 - Include all flights that have taken place between two airports in the US
 - Monthly data for each airline and route
 - e.g. United-Boston-Memphis
 - Origin, destination airports, flight duration, and aircraft type

Travel Time

- Assumption: Managers use the shortest travel time (pretty reasonable)
- Travel time between two ZIP codes (HQ and plant)
 - Compute travel time by car
 - Find the fastest airline route between two ZIP codes
 - ① Time by car between HQ and origin airport
 - ② Duration of the flight (including layover) (average across all flights)
 - ③ Time by car between destination airport and plant
 - Assumption: 1 hour at airports (including layover)

Identification Strategy

- “Consider a company headquartered in Boston with a plant in Memphis. In 1985, the fastest way to travel from Boston to Memphis was an indirect flight with one stopover in Atlanta. In 1986, Northwest Airlines opened a new hub in Memphis and started operating direct flights between Boston and Memphis. The introduction of this new airline route substantially reduced the travel time between the Boston headquarters and the Memphis plant and is therefore coded as a “treatment” of the Memphis plant in 1986.” (p. 863)
- “To measure the effect of this treatment (...), one could simply compare investment at the Memphis plant before and after 1986. However, other events in 1986 might have also affected investment at the Memphis plant. For instance, (...) a nationwide surge in investment due to favorable economic conditions or low interest rates. To account for this (...), I include a control group that consists of all plants that have not (yet) been treated. Due to the staggered nature of the introduction of new airline routes, this implies a plant remains in the control group until it is treated (which, for some plants, may be never). I then compare the difference in investment at the Memphis plant before and after 1986 with the difference in investment at the control plants before and after 1986. The difference between the two differences is the estimated effect of the introduction of the new airline route between BOS and MEM on investment at the Memphis plant.” (p. 870-1)

Specification

- DD specification

$$y_{ijlt} = \alpha_i + \alpha_t + \beta \times \text{treatment}_{it} + \gamma' X_{ijlt} + \varepsilon_{ijlt}$$

where i is plant, j is firm, l is location and t is time

- $\text{treatment} = 1$ if new route reducing travel time between plant and HQ has been introduced by time t
- H_0 : New route introduction does not matter ($\beta = 0$)
- Treated group: Plant experiencing a decrease in travel time
 - (5 years before and 5 years after, see paper for discussion/details)
- Control group: All the other plants (i.e., business as usual)

Identification Challenges

- New routes are endogenous
 - Could depend on local conditions (booming economy)
 - This could also be related with plant investment
 - Omitted variable problem
- BUT treatment is uniquely defined by two airport locations
 - Compare Boston (HQ)-Memphis (Treated) with Chicago (HQ)-Memphis (Control)
 - MSA-Year controls (mean of independent variable)
- Include firm-year controls (whole firm) to capture firm-level shocks
 - e.g. the whole firm invests more in a given year
- Plant-level shocks (new route because the plant is booming)
 - Look at investment before the event and find nothing significant

Summary Statistics

TABLE I
SUMMARY STATISTICS: PLANTS

	(1) All plants	(2) Eventually new airline route	(3) No new airline route
Total value of shipments	50,196 (360,930)	75,752 (222,685)	48,721 (367,270)
Capital stock	20,710 (106,473)	33,615 (118,024)	19,965 (105,719)
Employees	213 (568)	300 (638)	208 (564)
Distance to headquarters (miles)	312 (563)	854 (616)	281 (544)
Travel time (minutes)	126 (170)	362 (135)	113 (162)
Number of observations	1,291,280	70,467	1,220,813

TABLE II
SUMMARY STATISTICS: TRAVEL TIME REDUCTIONS

	(1) All	(2) Indirect to indirect	(3) Indirect to direct	(4) Direct to direct	(5) Road to flight
Distance to headquarters (miles)	854	1,211	942	726	191
Travel time before (minutes)	417	566	466	338	253
Travel time after (minutes)	314	420	339	266	206
Δ travel time (minutes)	-103	-146	-127	-72	-47
Δ travel time (%)	-25	-26	-27	-21	-19
Number of observations	10,533	1,911	3,469	4,544	609

Main Results

- Robust effect of treatment on both investment and productivity across specifications

TABLE III
THE EFFECT OF NEW AIRLINE ROUTES ON PLANT INVESTMENT AND PRODUCTIVITY

Dependent variable	Investment			TFP		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.008*** (0.001)	0.009*** (0.001)	0.010*** (0.001)	0.014*** (0.003)	0.013*** (0.003)	0.013*** (0.003)
MSA-year		0.153*** (0.022)	0.148*** (0.022)		0.080*** (0.012)	0.080*** (0.012)
Firm-year		0.205*** (0.006)	0.205*** (0.006)		0.186*** (0.005)	0.186*** (0.005)
Age		-0.060*** (0.002)	-0.061*** (0.002)		0.015*** (0.002)	0.018*** (0.003)
Size		0.029*** (0.001)			0.012*** (0.002)	
Plant fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Size \times year fixed effects	No	No	Yes	No	No	Yes
R-squared	0.39	0.41	0.41	0.60	0.61	0.61
Number of observations	1,291,280	1,291,280	1,291,280	1,291,280	1,291,280	1,291,280

Dynamic Effects

- Solution to the staggered DD problem (bias of TWFE):
 - **Point 1 (Pre-Trends):** The pre-treatment coefficients ($t < 0$) serve as our parallel trends test. Here, they are all close to zero, which is great.
 - **Point 2 (Dynamic Effects):** The coefficients for $t \geq 0$ show the dynamic effect of the treatment. The effect grows over time (a “Time-to-build” effect).

TABLE IV
DYNAMIC EFFECTS OF NEW AIRLINE ROUTES

Dependent variable	(1) Investment	(2) TFP
Treatment (-12 m, -6 m)	-0.000 (0.003)	-0.001 (0.005)
Treatment (-6 m, 0 m)	-0.001 (0.002)	-0.001 (0.004)
Treatment (0 m, 6 m)	0.003 (0.003)	0.001 (0.005)
Treatment (6 m, 12 m)	0.005** (0.002)	0.006 (0.005)
Treatment (12 m, 18 m)	0.013*** (0.003)	0.012** (0.005)
Treatment (18 m, 24 m)	0.014*** (0.002)	0.020*** (0.004)
Treatment (24 m, 30 m)	0.014*** (0.003)	0.020*** (0.005)
Treatment (30 m +)	0.009*** (0.002)	0.013*** (0.004)
MSA-year	0.153*** (0.022)	0.080*** (0.012)
Firm-year	0.205*** (0.006)	0.186*** (0.005)
Age	-0.060*** (0.002)	0.015*** (0.002)
Size	0.029*** (0.001)	0.012*** (0.002)
Plant fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
R-squared	0.41	0.61
Number of observations	1,291,280	1,291,280

Heterogeneity in Treatment Effect

- Larger effect if proximity combined with time-constrained managers
- Larger effect before IT innovations (think SAP, Skype etc) but still...

TABLE VIII
HEADQUARTERS' TIME CONSTRAINTS

Dependent variable	Managers/plants		Managers/total distance	
	(1) Investment	(2) TFP	(3) Investment	(4) TFP
Treatment \times high time constraints	0.012*** (0.002)	0.015*** (0.004)	0.013*** (0.002)	0.015*** (0.003)
Treatment \times low time constraints	0.006*** (0.002)	0.010** (0.004)	0.005** (0.002)	0.009* (0.005)
MSA-year	0.153*** (0.022)	0.080*** (0.012)	0.153*** (0.022)	0.080*** (0.012)
Firm-year	0.205*** (0.006)	0.186*** (0.005)	0.205*** (0.006)	0.186*** (0.005)
Age	-0.060*** (0.002)	0.015*** (0.002)	-0.060*** (0.002)	0.015*** (0.002)
Size	0.029*** (0.001)	0.012*** (0.002)	0.029*** (0.001)	0.012*** (0.002)
Plant fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
R-squared	0.41	0.61	0.41	0.61
Number of observations	1,291,280	1,291,280	1,291,280	1,291,280

TABLE IX
INNOVATIONS IN INFORMATION TECHNOLOGY

Dependent variable	(1) Investment	(2) TFP
Treatment \times pre 1986	0.013*** (0.002)	0.019*** (0.004)
Treatment \times between 1986 and 1995	0.010*** (0.002)	0.012*** (0.004)
Treatment \times post 1995	0.005** (0.002)	0.009* (0.005)
MSA-year	0.153*** (0.022)	0.080*** (0.012)
Firm-year	0.205*** (0.006)	0.186*** (0.005)
Age	-0.060*** (0.002)	0.015*** (0.002)
Size	0.029*** (0.001)	0.012*** (0.002)
Plant fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
R-squared	0.41	0.61
Number of observations	1,291,280	1,291,280

Comment

- This is the kind of paper which – if well-executed – is worth a replication for another country or the EU
 - Additional twists could be, e.g., whether languages, national borders matter? Differential impact of M&A?
- In this case, would result in a fantastic thesis

- Difference-in-differences compares the changes in outcomes over time between units that are enrolled in a program (the treatment group) and units that are not (the comparison/control group).
 - This allows us to correct for any differences between the treatment and control groups that are constant over time
- Instead of comparing outcomes between the treatment and control groups after the intervention, the DD method compares trends between the treatment and control groups

References

- Gertler et al (2016). Impact Evaluation in Practice, 2nd. Edition. Washington, DC: Inter-American Development Bank and World Bank
 - Chapter 7
- Gertler et al (2016). Impact Evaluation in Practice, 2nd. Edition, Technical Companion (Version 1.0). Washington, DC: Inter-American Development Bank and World Bank.
 - p. 20-25
- Auffhammer and Kellogg (2011). Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality. *American Economic Review* 101 (10): 2687-2722.
- Bollinger, Leslie, and Sorensen (2011). Calorie Posting in Chain Restaurants. *American Economic Journal: Economic Policy* 3, 91-128.
- Giroud (2013). Proximity and Investment: Evidence from Plant-Level Data. *The Quarterly Journal of Economics* 128, 861–915.
- Moser, Voena, and Waldinger (2014). German Jewish Émigrés and US Invention. *American Economic Review* 104(10): 3222–3255.

Overview

- **Paper:** Moser et al (2014). German Jewish Émigrés and US Invention. American Economic Review
 - Research question: Effect of migration (Jewish Émigrés) on (US) science?
 - Motivation: Jewish (and other) scientists fled Nazi Germany and oftentimes emigrated to the US – what was their contribution?
 - Empirical challenge: choice of field (within chemistry) might be endogenous, e.g. war-related research
 - (Paper explains why they don't focus on Physics– in short, data!)
 - Empirical strategy: Use a natural experiment and a difference-in-differences methodology to make a causal statement on the effect of migration on science, as measured by patent filings
 - Results: Patenting by US inventors increased by 31 percent in émigré fields. Inventor-level data indicate that émigrés encouraged innovation by attracting new researchers to their fields, rather than by increasing the productivity of incumbent inventors.

Natural Experiment

- On April 7, 1933 – only 67 days after the Nazis assumed power in Germany – the Law for the Restoration of the Professional Civil Service required that “*Civil servants who are not of Aryan descent are to be placed in retirement*”
- After the annexation of Austria in 1938, dismissals were extended to Austrian universities, so that the term “German scientists” in this paper includes chemists from both countries

Empirical Strategy

- DD regressions compare changes in US patenting by US inventors in research fields of German Jewish émigrés with changes in US patenting by US inventors in fields of other German chemists
 - Allows to control for a potential increase in US invention in fields where German chemists – who had dominated chemical research in the early twentieth century – were active inventors
- **Concern:** Baseline estimates biased if the US attracted more productive scientists or if the émigrés were more likely to work in fields in which US inventors would become more productive
 - Historical evidence: émigrés to the US may have been negatively selected, because Britain, was the first refuge for many émigrés, and established universities offered employment opportunities to the most prominent dismissed German scientists
 - Historical accounts also suggest that selection into research fields may have been negative because anti-Semitism in the United States restricted access to the most promising fields.
 - e.g. US firm Du Pont rejected the “father” of modern biochemistry Carl Neuberg, because he “looked” too Jewish

Empirical Strategy cont'd

- IV regressions use the pre-1933 fields of dismissed chemists as an instrument for the fields of émigrés to the US
 - Pre-1933 research fields were determined before the Nazis' rise to power and did not depend on expectations about the types of research which would become productive in the US after 1933
 - Consistent with historical accounts of negative selection, IV estimates imply a 71 percent increase in patenting, which implies that OLS (ordinary least squares) estimates underestimate, rather than overestimate, the true effects of the émigrés on US invention

Data

- A lot of data work, combining different sources, some of which historical
 - Émigré and Other Chemistry Professors at German and Austrian Universities
 - US Patents of Émigré and Other German Professors (1920–1970)
 - Matching Patents with USPTO Classes
 - US Patents by US Inventors per Class and Year
 - Individual-Level Patent Histories for US Patentees

Effects of Émigrés on Domestic Invention in the US

- Baseline OLS regression

$$\text{Patents by US Inventor}_{c,t} = \alpha_0 + \beta \text{émigré class}_c \cdot \text{post}_t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$$

where the dep. var. counts US patents by domestic inventors in technology class c and year t between 1920-1970. The indicator variable émigré class_c equals 1 if technology class c includes at least one patent between 1920-1970 by a German Jewish émigré to the US; post_t equals 1 starting with the year when dismissals first occurred in Germany (1933) and in Austria (1938)

- Control group: USPTO technology classes which include patents by other Germany chemists but not the émigrés
- $X_{c,t}$ includes controls for variation in patenting at the level of research fields and years (see paper for details)
- Year FEs δ control for unobs. variation in patenting over time which is common across technologies, and class FEs f control for unobs. variation in patenting across technologies which is constant over time

Common Trends?

- No difference in leverage trends between treatment and control before 1933 – Good news!

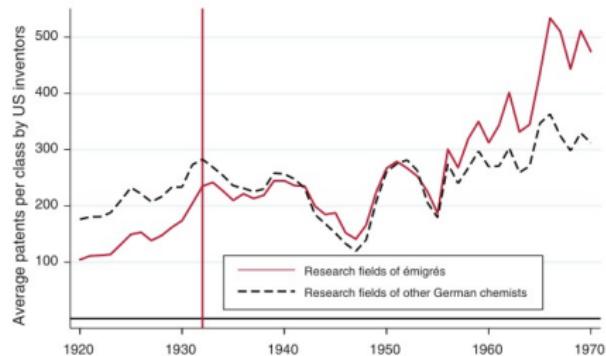


FIGURE 2A. US PATENTS PER CLASS AND YEAR BY DOMESTIC US INVENTORS
IN RESEARCH FIELDS OF ÉMIGRÉS AND OTHER GERMAN CHEMISTS

Notes: Data cover 2,073,771 patent-main class combinations by US inventors across 166 research fields defined at the level of USPTO classes. *Research fields of émigrés* cover 60 classes which include at least one patent between 1920 and 1970 by a German or Austrian émigré to the United States. *Research fields of other German chemists* cover 106 USPTO classes which include at least one patent between 1920 and 1970 by another German chemist but include no patents by émigrés.

OLS Results

- OLS estimates imply that the arrival of émigré chemists increased US patenting by a minimum of 31 percent [Col. (4): 75.4/240.9]
 - Col. (1): In classes that include at least 1 émigré patent, domestic inventors produced 105.2 additional patents per year after 1933, compared with classes that include at least 1 patent by another German chemist

TABLE 2—ORDINARY LEAST SQUARES REGRESSIONS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Émigré class × post</i>	105.222*** (22.203)	91.712*** (19.212)	84.803*** (18.950)	75.439*** (19.326)				
<i>Number émigré patents × post</i>				5.848* (3.058)	4.992* (2.561)	4.527** (2.182)	3.991** (1.956)	
Number foreign patents	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Quadratic class age	No	No	Yes	Yes	No	No	Yes	Yes
Patent pools	No	No	No	Yes	No	No	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,466	8,466	8,466	8,466	8,466	8,466	8,466	8,466
R ²	0.783	0.845	0.849	0.851	0.779	0.842	0.846	0.848

Notes: The dependent variable is patents by US inventors per USPTO class and year, excluding patents by émigrés. *Émigré class* equals 1 for classes which include at least one US patent by an émigré. *Number émigré patents* measures the number of US patents by émigrés in class *c*. Classes without émigré patents form the control group. The dummy variable *Post* equals 1 for years after the dismissals. *Number of foreign patents* counts US patents by foreign nationals in class *c* and year *t*. *Quadratic class age* is a second-degree polynomial for years since the first patent in class *c*. The indicator variable *patent pools* equals 1 for classes which were affected by a patent pool.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

IV Regressions

- Estimate

$$\text{Patents by US Inventor}_{c,t} = \alpha_0 + \beta \text{pre} - 1933 \text{ dismissed class}_c + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$$

where the indicator variable $\text{pre} - 1933 \text{ dismissed class}_c$ equals 1 for technology classes c which include at least one pre-1933 patent by a dismissed German chemist

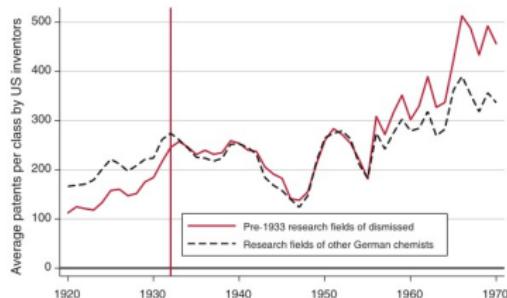


FIGURE 4. PATENTS BY DOMESTIC INVENTORS IN RESEARCH FIELDS IN WHICH DISMISSED CHEMISTS WERE ACTIVE BEFORE 1933

Notes: Data cover 2,073,771 patent-main class combinations by US inventors across 166 research fields defined at the level of USPTO classes. *Pre-1933 research fields of dismissed chemists* cover 48 classes which include at least one patent between 1920 and 1932 by a dismissed chemist. *Research fields of other German chemists* cover 118 USPTO technology classes which include at least one patent by another German chemist, but include no pre-1933 patents by dismissed chemists.

TABLE 4—INSTRUMENTAL VARIABLE REGRESSIONS

	(1)	(2)	(3)	(4)
Émigré class \times post	218.707*** (60.614)	170.136*** (57.992)		
Number émigré patents \times post			25.717*** (8.750)	17.137** (6.909)
Number foreign patents	No	Yes	No	Yes
Quadratic class age	No	Yes	No	Yes
Patent pools	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Class fixed effects	Yes	Yes	Yes	Yes
Observations	8,466	8,466	8,466	8,466

Notes: The dependent variable is patents by US inventors per USPTO class and year, excluding US patents by émigrés. *Émigré class* equals 1 for classes that include at least one US patent by an émigré. *Number émigré patents* measures the number of US patents by émigrés in class c . Classes without émigré patents form the control. The dummy variable *Post* equals 1 for years after the dismissals. Instruments are *Dismissed class \times post* (columns 1 and 2) and *number dismissed patents \times post* (columns 3 and 4). *Dismissed class* equals 1 for classes that include at least one pre-1933 US patent by a dismissed chemist. *Number dismissed patents* indicates the number of pre-1933 US patents by dismissed chemists in each class. *Number of foreign patents* counts US patents by foreign nationals in class c and year t . *Quadratic class age* is a second-degree polynomial for years since the first patent in class c . The indicator variable *patent pools* equals 1 for classes that were affected by a patent pool.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Note

- See paper for details, robustness, discussion of mechanism, individual-level results etc