Lecture 3: Differences-in-Differences

Fabian Waldinger

Topics Covered in Lecture

1. Review of fixed effects regression models.
2. Differences-in-Differences Basics: Card & Krueger (1994).
3. Regression Differences-in-Differences.
4. Synthetic Controls: Abadie & Gardeazabal (2003).
5. Combining Differences-in-Differences with IV: Waldinger (2010).

Very Brief Review of Fixed Effects Models - Introductory Example

- Suppose you are interested in the question whether union workers earn higher wages.
- Problem: unionized workers may be different (e.g. higher skilled, more experienced) from non-unionized workers.
- Many of these factors will not be observable to the econometrician (standard omitted variable bias problem).
- Therefore the error term and union status will be correlated and OLS will be biased.

Very Brief Review of Fixed Effects Models

- We are interested whether Y_{it} (earnings) is affected by D_{it} (union status) which we assume to be randomly assigned.
- We also have time varying covariates X_{ii} (such as experience) and unobserved but fixed confounders A_i (e.g. ability).

$$E[Y_{0it}|A_{i,}X_{it}, t] = \alpha + \lambda_t + A_i'\gamma + X_{it}'\beta$$

- Assuming that the causal effect of union membership is additive and constant we also have:

$$E[Y_{1it}|A_{i,}X_{it}, t] = E[Y_{0it}|A_{i,}X_{it}, t] + \rho$$

- Together with the previous equation this implies:

$$E[Y_{it}|A_{i,}X_{it}, t] = \alpha + \lambda_t + \rho D_{it} + A_i'\gamma + X_{it}'\beta$$

Estimation of Fixed Effects Models

- This equation implies the following regression equation:

$$Y_{it} = \alpha_i + \lambda_t + \rho D_{it} + X_{it}'\beta + \varepsilon_{it} \tag{1}$$

where $\varepsilon_{it} = Y_{0it} - E[Y_{0it}|A_{i,}X_{it}, t]$ and $\alpha_i = \alpha + A_i'\gamma$

- Suppose you simply estimate this model with OLS (without including individual fixed effects).
- You therefore estimate:

$$Y_{it} = constant + \lambda_t + \rho D_{it} + X_{it}'\beta + \underbrace{\alpha_i + \varepsilon_{it}}_{uit}$$

- As α_i is correlated with union status D_{it} there is a correlation of D_{it} with the error term. This will lead to biased OLS estimates.

2 Ways of Estimating Fixed Effects Models

- A fixed effect model would address this problem because α_i would be included in the regression.
- D_{it} and the error term would therefore be uncorrelated and you would obtain an unbiased estimate of ρ.
- In practice there are two ways of estimating this fixed effects model:
    1. Demeaning (sometimes called "within estimator").
    2. First differencing.

Within-Estimator

- With *demeaning* you (or the computer) first calculate individual averages of the dependent variable and all explanatory variables.
- You then substract these averages from regression equation (1):

$$Y_{it} - \overline{Y}_i = \lambda_t - \overline{\lambda} + \rho(D_{it} - \overline{D}_i) + (X_{it} - \overline{X}_i)'\beta + (\varepsilon_{it} - \overline{\varepsilon}_i)$$

- Thus α_i drops out and therefore the error and the regressor would no longer be correlated.

First-Differencing

- An alternative way of estimating the fixed effects model is *first differencing* which would also get rid of the α_i.

$$\Delta Y_{it} = \Delta\lambda_t + \rho\Delta D_{it} + \Delta X'_{it}\beta + \Delta\varepsilon_{it}$$

- With 2 periods the two methods are algebraically the same. Otherwise not.
- Both should work, but with first differencing you introduce serial correlation of the error terms.
- Therefore demeaning is usually the best option.

The Effect of Unionization on Wages - OLS vs. FE

- Freeman (1984) analyzed unionization comparing OLS and FE models for a number of datasets:

Survey	OLS	Fixed Effects
CPS 74-75	0.19	0.09
NLSY 70-78	0.28	0.19
PSID 70-79	0.23	0.14
QES 73-77	0.14	0.16

- These results suggest that union workers are positively selected.

Measurement Error and Fixed Effects Models

- OLS results were larger than FE $\rightarrow$ selection may be important.
- Another plausible explanation is measurement error.
- Measurement error introduces attenuation bias.
- As the signal to noise ratio is smaller with fixed effects (as we just use the deviations from the mean as signal) measurement error is typically a more important problem in fixed effect models.
- In this case union status may be misreported for some individuals in each year. Observed year to year changes in union status for one individual may thus be mostly noise.

Differences-in-Differences: Card & Krueger (1994)

- Suppose you are interested in the effect of minimum wages on employment (a classic and controversial question in labour economics).
- In a competitive labour market, increases in the minimum wage would move us up a downward-sloping labour demand curve.
  $\rightarrow$ employment would fall.

Differences-in-Differences: Card & Krueger (1994)

- Card & Krueger (1994) analyse the effect of a minimum wage increase in New Jersey using a differences-in-differences methodology.
- In February 1992 NJ increased the state minimum wage from \$4.25 to \$5.05. Pennsylvania's minimum wage stayed at \$4.25.

- They surveyed about 400 fast food stores both in NJ and in PA both before and after the minimum wage increase in NJ.

- DD is a version of fixed effects estimation. To see this more formally:
  Y_{1ist} : employment at restaurant i, state s, time t with a high w^{min}.
  Y_{0ist} : employment at restaurant i, state s, time t with a low w^{min}.
- In practice of course we only see one or the other.
- We then assume that:

$$E[Y_{0ist}|s,t] = \gamma_s + \lambda_t$$

- In the absence of a minimum wage change, employment is determined by the sum of a time-invariant state effect γ_s and a year effect λ_t that is common across states.
- Let D_{st} be a dummy for high-minimum wage states and periods.
- Assuming $E[Y_{1ist} - Y_{0ist}|s,t] = \delta$ is the treatment effect, observed employment can be written:

$$Y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \varepsilon_{ist}$$

Differences-in-Differences Strategy II

- In New Jersey:
  - Employment in February is:
    $E[Y_{ist}|s = NJ, t = Feb] = \gamma_{NJ} + \lambda_{Feb}$
  - Employment in November is:
    $E[Y_{ist}|s = NJ, t = Nov] = \gamma_{NJ} + \lambda_{Nov} + \delta$
  - the difference between February and November is:
    $E[Y_{ist}|s = NJ, t = N] - E[Y_{ist}|s = NJ, t = F] = \lambda_N - \lambda_F + \delta$
- In Pennsylvania:
  - Employment in February is:
    $E[Y_{ist}|s = PA, t = Feb] = \gamma_{PA} + \lambda_{Feb}$
  - Employment in November is:
    $E[Y_{ist}|s = PA, t = Nov] = \gamma_{PA} + \lambda_{Nov}$
  - the difference between February and November is:
    $E[Y_{ist}|s = PA, t = Nov] - E[Y_{ist}|s = PA, t = Feb] = \lambda_{Nov} - \lambda_{Feb}$

Differences-in-Differences Strategy

- The differences-in-differences strategy amounts to comparing the change in employment in NJ to the change in employment in PA.
- The population differences-in-differences are:

$$E[Y_{ist}|s = NJ, t = N] - E[Y_{ist}|s = NJ, t = F]$$
$$-E[Y_{ist}|s = PA, t = Nov] - E[Y_{ist}|s = PA, t = Feb] = \delta$$

- This is estimated using the sample analog of the population means.

Differences-in-Differences Table

Variable	PA (i)	NJ (ii)	Difference, NJ − PA (iii)
	Stores by state		
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	−2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	−0.14 (1.07)
3. Change in mean FTE employment	−2.16 (1.25)	0.59 (0.54)	$\boxed{2.76\ (1.36)}$

Surprisingly, employment rose in NJ relative to PA after the minimum wage change.

Regression DD

- We can estimate the differences-in-differences estimator in a regression framework.
- Advantages:
  - It is easy to calculate standard errors.
  - We can control for other variables which may reduce the residual variance (lead to smaller standard errors).
  - It is easy to include multiple periods.
  - We can study treatments with different treatment intensity. (e.g. varying increases in the minimum wage for different states).
- The typical regression model that we estimate is:

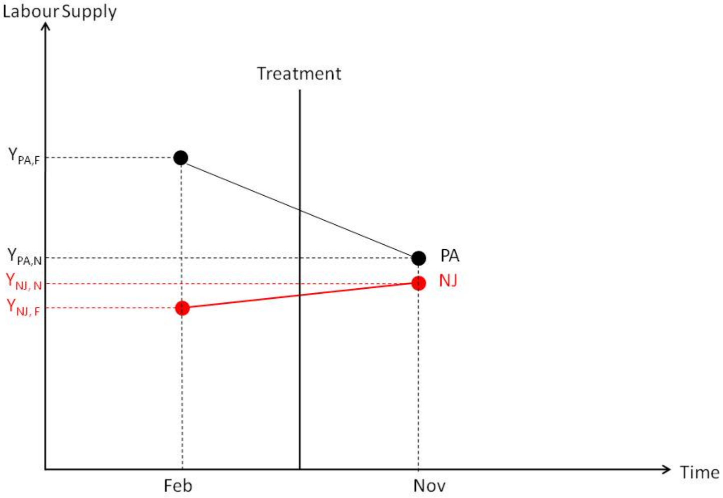
$$\text{Outcome}_{it} = \beta_1 + \beta_2 \ \text{Treat}_i + \beta_3 \ \text{Post}_t + \beta_4 \ (\text{Treat * Post})_{it} + \varepsilon$$

Treatment = a dummy if the observation is in the treatment group
Post = post treatment dummy

Regression DD - Card & Krueger

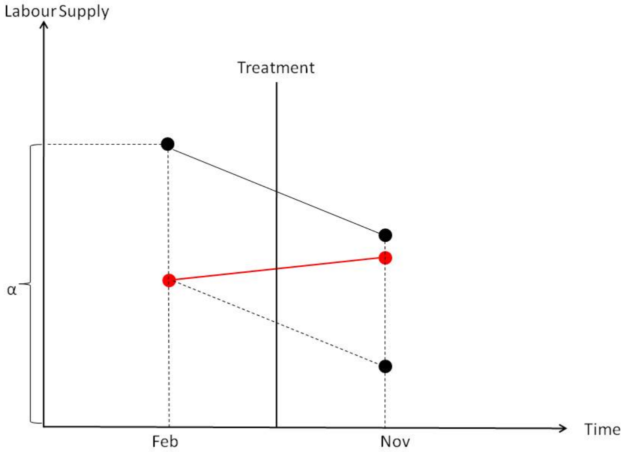
- In the Card & Krueger case the equivalent regression model would be: $Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ_s * d_t) + \varepsilon_{ist}$
  - *NJ* is a dummy which is equal to 1 if the observation is from NJ.
  - *d* is a dummy which is equal to 1 if the observation is from November (post).
- This equation takes the following values.
  - PA Pre: α
  - PA Post: $\alpha + \lambda$
  - NJ Pre: $\alpha + \gamma$
  - NJ Post: $\alpha + \gamma + \lambda + \delta$
- Differences-in-Differences estimate: (NJ Post - NJ Pre) - (PA Post - PA Pre) $= \delta$

Graph - Observed Data



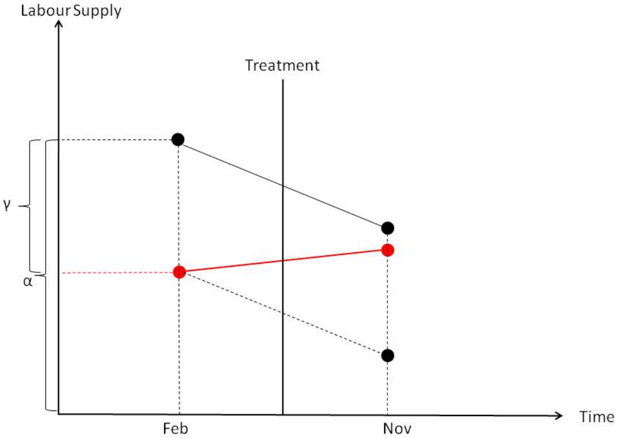
Graph - DD

$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ_s * d_t) + \varepsilon_{ist}$$



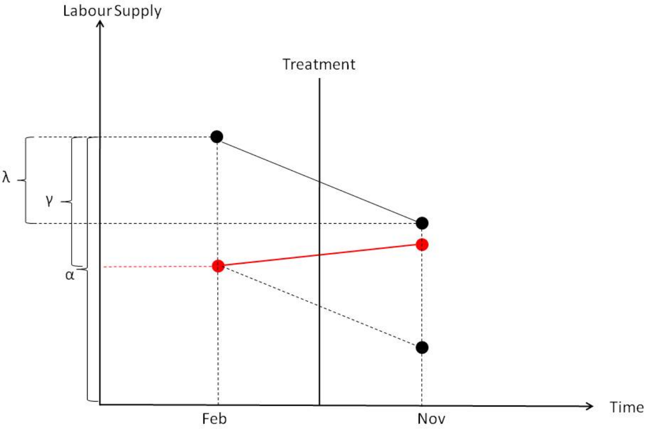
Graph - DD

$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ_s * d_t) + \varepsilon_{ist}$$



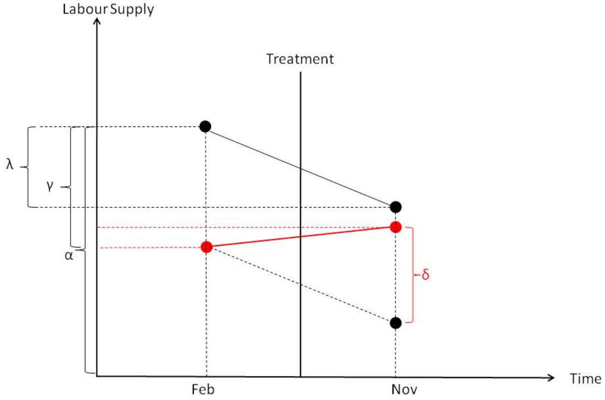
Graph - DD

$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ_s * d_t) + \varepsilon_{ist}$$



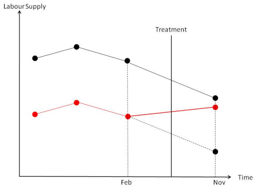
Graph - DD

$$Y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta(NJ_s * d_t) + \varepsilon_{ist}$$



Key Assumption of Any DD Strategy: Common Trends

- The key assumption for any DD strategy is that the outcome in treatment and control group would follow the same time trend in the absence of the treatment.
- This does not mean that they have to have the same mean of the outcome!
- Common trend assumption is difficult to verify but one often uses pre-treatment data to show that the trends are the same.
- Even if pre-trends are the same one still has to worry about other policies changing at the same time.



Regression DD Including Leads and Lags

- Including leads into the DD model is an easy way to analyze pre-trends.
- Lags can be included to analyze whether the treatment effect changes over time after treatment.
- The estimated regression would be:

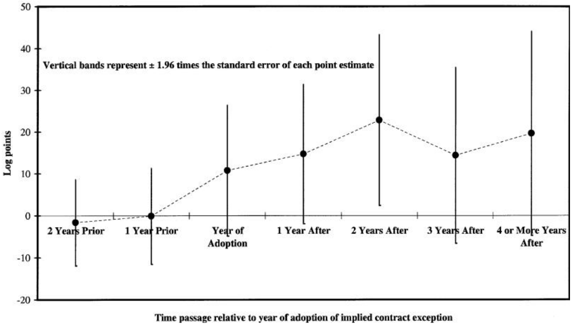
$$Y_{ist} = \gamma_s + \lambda_t + \sum_{\tau=-q}^{-1} \delta_\tau D_{s\tau} + \sum_{\tau=0}^{m} \delta_\tau D_{s\tau} + X_{ist} + \varepsilon_{ist}$$

  - treatment occurs in year 0.
  - includes q leads or anticipatory effects.
  - includes m leads or post treatment effects.

Study Including Leads and Lags - Author (2003)

- Autor (2003) includes both leads and lags in a DD model analyzing the effect of increased employment protection on the firm's use of temporary help workers.
- In the US employers can usually hire and fire workers at will.
- Some states courts have made some exceptions to this employment at will rule and have thus increased employment protection.
- Different states have passed these exeptions at different points in time.
- The standard thing to do is to normalize the adoption year to 0.
- Autor then analyzes the effect of these exeptions on the use of temporary help workers.

Results



- The leads are very close to 0. → no evidence for anticipatory effects (good news for the common trends assumption).
- The lags show that the effect increases during the first years of the treatment and then remains relatively constant.

Standard Errors in DD Strategies

- Many papers using a DD strategy use data from many years (not only 1 pre and 1 post period).
- The variables of interest in many of these setups only vary at a group level (say state) and outcome variables are often serially correlated.
- In the Card and Krueger study for example, it is very likely that employment in each state is not only correlated within the state but also serially correlated.
- As Bertrand, Duflo, and Mullainathan (2004) point out, conventional standard errors often severely understate the standard deviation of the estimators.

Standard Errors in DD Strategies - Practical Solutions

- Bertrand, Duflo, and Mullainathan propose the following solutions:
  1. Block bootstrapping standard errors (if you analyze states the block should be the states and you would sample whole states with replacing for the bootstrapping).
  2. Clustering standard errors at the group level. (in STATA one would simply add `cl(state)` to the regression equation if one analyzes state level variation).
  3. Aggregating the data into one pre and one post period.
     Literally works only if there is only one treatment date. With staggered treatment dates one should adopt the following procedure:
     - Regress Y_{st} on state FE, year FE, and relevant covariates.
     - Obtain residuals from the treatment states only and divide them into 2 groups: pre and post treatment.
     - Then regress the two groups of residuals on a post dummy.
- Correct treatment of standard errors sometimes makes the number of groups very small: in the Card and Krueger study the number of groups is only 2.

Synthetic Control Methods

- In some cases, treatment and potential control groups do not follow parallel trends.
  $\rightarrow$ Standard DD method would lead to biased estimates.
- Abadie & Gardeazabal (2003) pioneered a synthetic control method when estimating the effects of the terrorist conflict in the Basque Country using other Spanish regions as a comparison group. (Card (1990) implicitly used a very similar approach in his Mariel boatlift paper investigating the effect of immigration on employment of natives).
- The basic idea behind synthetic controls is that a combination of units often provides a better comparison for the unit exposed to the intervention than any single unit alone.

Abadie & Gardeazabal (2003) - The Effect of Terrorism on Growth

- They want to evaluate whether Terrorism in the Basque Country had a negative effect on growth.
- They cannot use a standard DD method because none of the other Spanish regions followed the same time trend as the Basque Country.
- They therefore take a weighted average of other Spanish regions as a synthetic control group.

The Basque Country is Different from the Rest of Spain

	Basque Country (1)	Spain (2)
Real per capita GDP[a]	5,285.46	3,633.25
Investment ratio (percentage)[b]	24.65	21.79
Population density[c]	246.89	66.34
Sectoral shares (percentage)[d]		
Agriculture, forestry, and fishing	6.84	16.34
Energy and water	4.11	4.32
Industry	45.08	26.60
Construction and engineering	6.15	7.25
Marketable services	33.75	38.53
Nonmarketable services	4.07	6.97
Human capital (percentage)[e]		
Illiterates	3.32	11.66
Primary or without studies	85.97	80.15
High school	7.46	5.49
More than high school	3.26	2.70

The Synthetic Control Method

- They have J available control regions (the 16 Spanish regions other than the Basque Country).
- They want to assign weights $\mathrm{W} = (w_1, ..., w_J)'$ a (J x 1) to each region. ($wj \geq 0$ & $\sum wj = 1$; this ensures that there is no extrapolation outside the support of the growth predictors for the control regions).
- The weights are chosen so that the synthetic Basque country most closely resembles the actual one before terrorism.

- Let X_1 be a $(K \times 1)$ vector of pre-terrorism of K economic growth predictors (i.e. the values in the previous table: investment ratio, population density, ...) in the Basque Country.
- Let X_0 be a $(K \times J)$ matrix which contains the values of the same variables for the J possible control regions.
- Let V be a diagonal matrix with nonnegative components reflecting the relative importance of the different growth predictors.
- The vector of weights W* is then chosen to minimize:

$$(X_1 - X_0 W)' V (X_1 - X_0 W)$$

- They choose the matrix V such that the real per capita GDP path for the Basque Country during the 1960s (pre terrorism) is best reproduced by the resulting synthetic Basque Country.

The Synthetic Control Method - Details

- The optimal weights they get are: Catalonia: 0.8508, Madrid: 0.1492, and all other regions: 0.
- Alternatively they could have just chosen the weights to reproduce only the pre-terrorism growth path for the Basque country (and not the growth predictors as well. In that case they would have minimized:

$$(Z_1 - Z_0W)'(Z_1 - Z_0W)$$

- Z_1 is the (10 x 1) vector of pre-terrorism (1960-1969) GDP values for the Basque Country.
- Z_0 is the (10 x J) vector of pre-terrorism (1960-1969) GDP values for the J potential control regions.

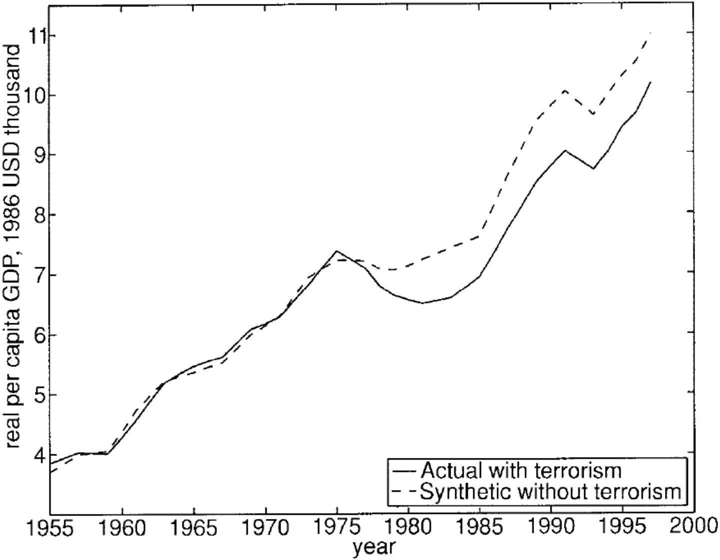
The Synthetic Basque Country Looks Similar

	Basque Country (1)	Spain (2)	"Synthetic" Basque Country (3)
Real per capita GDP[a]	5,285.46	3,633.25	5,270.80
Investment ratio (percentage)[b]	24.65	21.79	21.58
Population density[c]	246.89	66.34	196.28
Sectoral shares (percentage)[d]			
Agriculture, forestry, and fishing	6.84	16.34	6.18
Energy and water	4.11	4.32	2.76
Industry	45.08	26.60	37.64
Construction and engineering	6.15	7.25	6.96
Marketable services	33.75	38.53	41.10
Nonmarketable services	4.07	6.97	5.37
Human capital (percentage)[e]			
Illiterates	3.32	11.66	7.65
Primary or without studies	85.97	80.15	82.33
High school	7.46	5.49	6.92
More than high school	3.26	2.70	3.10

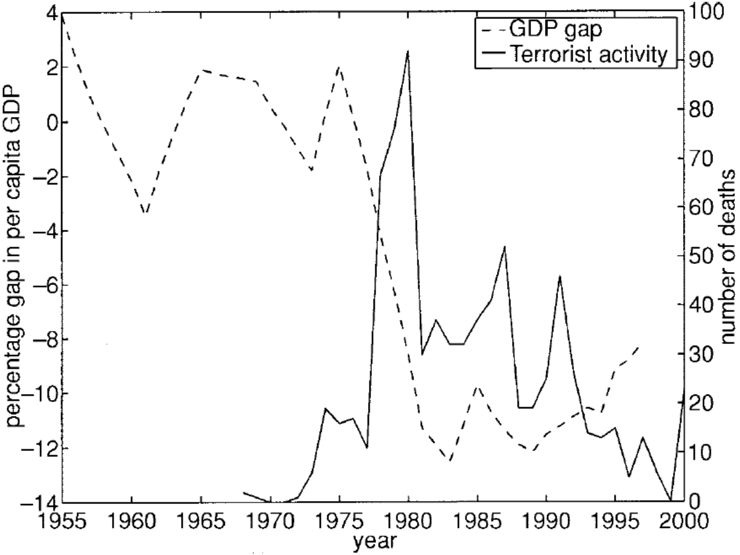
Constructing the Counterfactual Using the Weights

- Y_1 is a (T x 1) vector whose elements are the values of real per capital GDP values for T years in the Basque country.
- Y_0 is a (T x J) matrix whose elements are the values of real per capital GDP values for T years in the control regions.
- They then constructed the counterfactual GDP (in the absence of terrorism) as: $Y*_1 = Y_0 W*$

Growth in the Basque Country with and without Terrorism



Terrorist Activity and Estimated GDP Gap



Combining DD and IV

- Sometimes combining DD and IV methods can be quite useful.
- In a recent paper (Waldinger, 2010), I have done that to estimate the effect of faculty quality on the outcomes of PhD students.
- Estimating the effect of faculty quality on PhD student outcomes is challenging because of:
  1. Selection of good students into good universities.
  2. Omitted variables affecting both faculty quality and student outcomes.
  3. Measurement error in faculty quality.
- I address these issues by using the dismissal of scientists in Nazi Germany as an exogenous shock to faculty quality.
- The dismissal affected some departments very strongly, while other departments were not affected.

Historical Background

- Germany was the leading country for scientific research at the beginning of the 20th century.
- Immediately after gaining power in 1933 the new Nazi government dismissed all Jewish and ‘politically unreliable’ scholars from the German universities.

Reichsgesetzblatt

Teil I

1933 | Ausgegeben zu Berlin, den 7. April 1933 | Nr. 34

Inhalt: Gesetz zur Wiederherstellung des Berufsbeamtentums. Vom 7. April 1933.................. S. 175

Gesetz zur Wiederherstellung des Berufsbeamtentums. Vom 7. April 1933.

Die Reichsregierung hat das folgende Gesetz beschlossen, das hiermit verkündet wird:

§ 1

(1) Zur Wiederherstellung eines nationalen Berufsbeamtentums und zur Vereinfachung der Verwaltung können Beamte nach Maßgabe der folgenden Bestimmungen aus dem Amt entlassen werden, auch wenn die nach dem geltenden Recht hierfür erforderlichen Voraussetzungen nicht vorliegen.

(2) Als Beamte im Sinne dieses Gesetzes gelten unmittelbare und mittelbare Beamte des Reichs,

des jeweiligen Grundgehalts der von ihnen zuletzt bekleideten Stelle bewilligt werden; eine Nachversicherung nach Maßgabe der reichsgesetzlichen Sozialversicherung findet nicht statt.

(4) Die Vorschriften der Abs. 2 und 3 finden auf Personen der im Abs. 1 bezeichneten Art, die bereits vor dem Inkrafttreten dieses Gesetzes in den Ruhestand getreten sind, entsprechende Anwendung.

§ 3

(1) Beamte, die nicht arischer Abstammung sind, sind in den Ruhestand (§§ 8 ff.) zu versetzen; soweit es sich um Ehrenbeamte handelt, sind sie aus dem Amtsverhältnis zu entlassen.

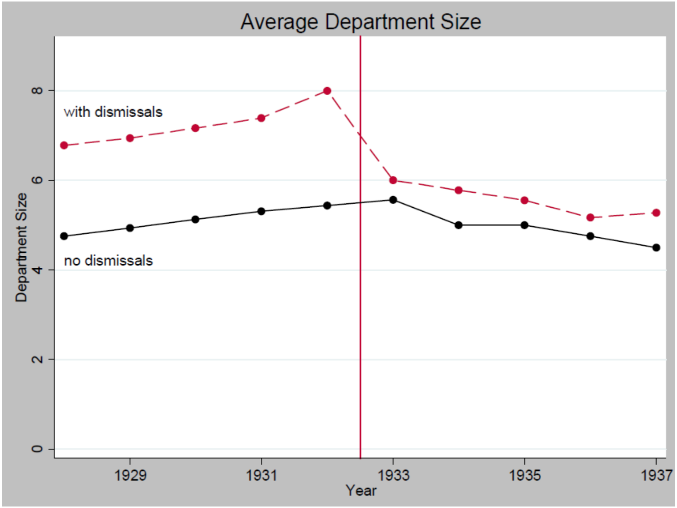
Dismissed Professors Across German Universities

University	Number of Professors Beginning of 1933	Dismissed 1933–34		Dismissal-Induced Change to Department Quality
		Number	Percentage	
Aachen TU	7	3	42.9	+
Berlin	13	5	38.5	−−
Berlin TU	14	2	14.3	+
Bonn	7	1	14.3	+
Braunschweig TU	3	0	0	0
Breslau	6	3	50.0	−−
Breslau TU	5	2	40.0	−−
Darmstadt TU	9	1	11.1	+
Dresden TU	10	0	0	0
Erlangen	3	0	0	0
Frankfurt	8	1	12.5	+
Freiburg	9	1	11.1	−
Giessen	7	1	14.3	+
Göttingen	17	10	58.8	−−
Greifswald	3	0	0	0

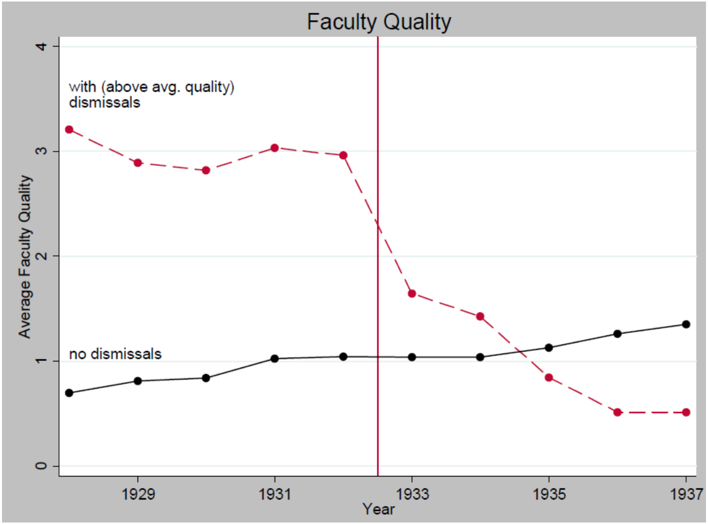
Dismissed Professors Across German Universities II

Halle	7	1	14.3	+
Hamburg	8	0	0	0
Hannover TU	6	0	0	0
Heidelberg	5	1	20.0	+
Jena	5	0	0	0
Karlsruhe TU	6	1	16.7	0
Kiel	5	2	40.0	+
Köln	6	2	33.3	+
Königsberg	5	2	40.0	–
Leipzig	8	2	25.0	–
Marburg	8	0	0	0
München	9	0	0	0
München TU	5	0	0	0
Münster	5	0	0	0
Rostock	2	0	0	0
Stuttgart TU	6	0	0	0
Tübingen	6	0	0	0
Würzburg	4	0	0	0

Effect of Dismissals on Department Size



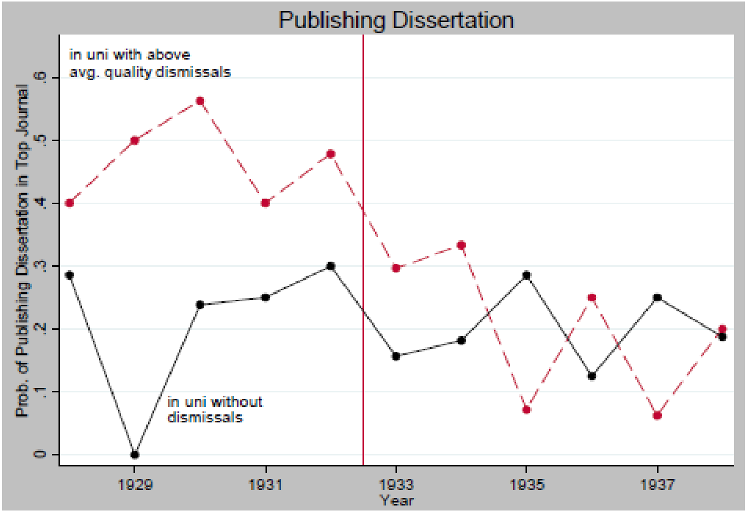
Effect of Dismissals on Faculty Quality



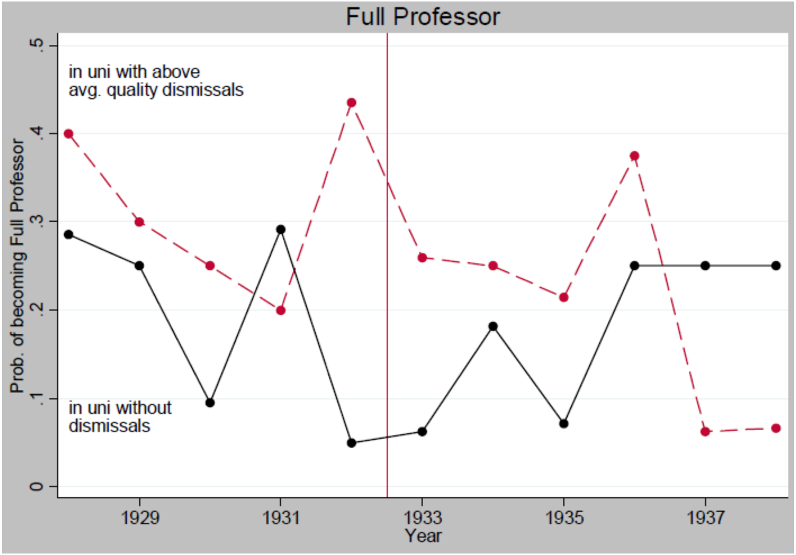
Panel Data on PhD graduates from German Universities

- I obtained a panel dataset of all mathematics PhD students graduating from all German universities between 1923 and 1938 and use the dismissal as exogenous variation in faculty quality.
- The empirical strategy essentially compares changes in outcomes of PhD students in affected department before and after 1933 to changes in outcomes in unaffected departments.
- I investigate the following outcomes:
  1. Whether former PhD student publishes dissertation in a top journal.
  2. Whether former PhD student ever becomes full professor.
  3. # of lifetime citations.
  4. Positive lifetime citations.

Reduced Form Graphical Analysis - Publishing Dissertation



Reduced Form Graphical Analysis - Full Professor



Reduced Form Graphical Analysis - Lifetime Citations

Reduced Form Estimates

- The reduced form of the dismissal effect is essentially a DD estimator.

$$
\begin{aligned}
\text{Outcome}_{idt} = \beta_1 &+ \beta_2(\text{Dismissal induced Reduction in Faculty Quality})_{dt} \\
&+ \beta_3(\text{Dismissal induced increase in Student/Faculty Ratio})_{dt} \\
&+ \beta_4 \text{ Female}_{idt} + \beta_5\text{Foreign}_{idt} + \beta_6 \text{ CohortFE}_t + \beta_5 \text{ DepFE}_d + \varepsilon_{idt}
\end{aligned}
$$

- Dismissal induced Reduction in Faculty Quality is 0 until 1933 and equal to the dismissal induced fall in faculty quality after 1933 (and remains 0 in departments without dismissals).
- Dismissal induced increase in Student/Faculty Ratio is also 0 until 1933 but equal to the dismissal induced increase in student/faculty ratio after 1933
  $\rightarrow$ essentially a differences-in-differences estimator but with different treatment intensities.

Reduced Form Estimates

	Reduced Form			
	Published Top	Full Professor	No. of Lifetime Citations	Positive Lifetime Citations
	(1)	(2)	(3)	(4)
Dismissal-induced fall in faculty quality	−.134**	−.090**	−6.137**	−.164**
	(.017)	(.021)	(2.218)	(.019)
Dismissal-induced increase in student/faculty ratio	.002	.000	−.042	.002
	(.001)	(.001)	(.114)	(.002)
Female	.004	−.119*	−10.723*	−.067
	(.048)	(.045)	(4.459)	(.058)
Foreigner	.031	−.147*	.942	−.033
	(.048)	(.065)	(6.151)	(.075)
Father's occupation	Yes	Yes	Yes	Yes
Cohort dummies	Yes	Yes	Yes	Yes
Department fixed effects	Yes	Yes	Yes	Yes
Observations	690	690	690	690
R^2	.221	.208	.185	.208

Common Robustness Check for Parallel Trend Assumption

Only Look at Pre-Period Data and Move Placebo Treatment some Years Back

- Here I move a placebo treatment to 1930.

	Placebo Moving Dismissal to 1930 (Only Pre-1933 Observations)			
	Published Top (5)	Full Professor (6)	No. of Lifetime Citations (7)	Positive Lifetime Citations (8)
Dismissal-induced fall in faculty quality	−.023	.053	3.434	−.037
	(.031)	(.037)	(5.597)	(.030)
Dismissal-induced increase in student/faculty ratio	−.004	−.002	−.462	−.002
	(.004)	(.005)	(.431)	(.003)
Female	.009	−.167*	−12.114**	−.104
	(.066)	(.068)	(4.228)	(.071)
Foreigner	−.017	−.136	−7.169	−.050
	(.103)	(.102)	(6.479)	(.134)
Father's occupation	Yes	Yes	Yes	Yes
Cohort dummies	Yes	Yes	Yes	Yes
Department fixed effects	Yes	Yes	Yes	Yes
Observations	403	403	403	403
R^2	.302	.291	.224	.260

Use Dismissal as IV

- OLS model to the effect of university quality on PhD student outcomes:

$$\text{Outcome}_{idt} = \beta_1 + \beta_2(\text{Avg. Faculty Quality})_{dt-1}$$
$$+ \beta_3(\text{Student/Faculty Ratio})_{dt-1}$$
$$+ \beta_4\text{Female}_{idt} + \beta_5\text{Foreign}_{idt} + \beta_6\ \text{CohortFE}_t + \beta_7\text{DepFE}_d + \varepsilon_{idt}$$

- University quality and student/faculty ratio are endogenous $\rightarrow$ use dismissal as IV.
- 2 Endogenous Variables $\rightarrow$ 2 First Stage Regressions:

1. $\text{Avg. Faculty Quality}_{idt} = \gamma_1$
   $+ \gamma_2(\text{Dismissal induced Reduction in Faculty Quality})_{dt}$
   $+ \gamma_3(\text{Dismissal induced increase in Student/Faculty Ratio})_{dt}$
   $+ \gamma_4\ \text{Female}_{idt} + \gamma_5\text{Foreign}_{idt} + \gamma_6\ \text{CohortFE}_t + \gamma_5\ \text{DepFE}_d + \varepsilon_{idt}$
2. $\text{Student/Faculty Ratio}_{idt} = \delta_1$
   $+ \delta_2(\text{Dismissal induced Reduction in Faculty Quality})_{dt}$
   $+ \delta_3(\text{Dismissal induced increase in Student/Faculty Ratio})_{dt}$
   $+ \delta_4\ \text{Female}_{idt} + \delta_5\text{Foreign}_{idt} + \delta_6\ \text{CohortFE}_t + \delta_5\ \text{DepFE}_d + \varepsilon_{idt}$

First Stages

	Dependent Variable	
	Average Quality (1)	Student/Faculty Ratio (2)
Dismissal-induced fall in faculty quality	−1.236**	−4.195
	(.074)	(2.058)
Dismissal-induced increase in student/faculty ratio	.014	.439**
	(.008)	(.116)
Female	.142*	1.165
	(.060)	(.705)
Foreigner	.046	−1.971
	(.097)	(1.183)
Cohort dummies	Yes	Yes
University fixed effects	Yes	Yes
Observations	690	690
R^2	.795	.757
Cragg-Donald eigenvalue statistic	25.2	

- To test for weak instruments one cannot simply look at the first stage F-statistics because here we have 2 endogenous regressors and 2 IVs. $\longrightarrow$ use Cragg-Donald EV statistic here critical value is 7.03.

OLS and IV

	Published Top		Full Professor	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Average faculty quality	.056**	.102**	.037	.076**
	(.018)	(.015)	(.021)	(.015)
Student/faculty ratio	.000	.003	.000	−.001
	(.001)	(.002)	(.001)	(.003)
Female	−.015	−.022	−.099*	−.103**
	(.059)	(.055)	(.041)	(.036)
Foreigner	.014	.022	−.134*	−.135*
	(.048)	(.045)	(.053)	(.053)
Cohort dummies	Yes	Yes	Yes	Yes
Department fixed effects	Yes	Yes	Yes	Yes
Observations	690	690	690	690
R^2	.163		.155	
Cragg-Donald eigenvalue statistic		25.2		25.2