

8: Kausal inferens

Videregående kvantitative metoder i studiet af politisk adfærd

Frederik Hjorth
fh@ifs.ku.dk
fghjorth.github.io
[@fghjorth](https://twitter.com/fghjorth)

Institut for Statskundskab
Københavns Universitet

7. november 2018

Opsamling
○○○

Motivation
○○

Angrist & Pischke (2010)
○○○○○○○○

Samii (2016)
○○○○○○○○○○

Case: Caroll (2018)
○○○○

Kig fremad
○○

1 Opsamling

2 Motivation

3 Angrist & Pischke (2010)

4 Samii (2016)

5 Case: Caroll (2018)

6 Kig fremad

Sidste gang:

- tidyng
- visualisering: ggplot2
- tips & tricks
- øvelse: hvorfor Brexit?

Opsamling

○●○

Motivation

○○

Angrist & Pischke (2010)

○○○○○○○○

Samii (2016)

○○○○○○○○○○

Case: Caroll (2018)

○○○○

Kig fremad

○○

<https://calendly.com/fghjorth/15min>

Opsamling

○○●

Motivation

○○

Angrist & Pischke (2010)

○○○○○○○○

Samii (2016)

○○○○○○○○○○

Case: Caroll (2018)

○○○○

Kig fremad

○○

<https://fghjorth.typeform.com/to/nen7Xb>

Opsamling
○○○

Motivation
●●

Angrist & Pischke (2010)
○○○○○○○○

Samii (2016)
○○○○○○○○

Case: Caroll (2018)
○○○○

Kig fremad
○○

Case: arbejdspladswellness-programmer



A yoga class at a New York wellness center. Randomized controlled trials can reverse the conclusions of observational studies. Chad Rhym for The New York Times

Traditionel tilgang: TEST-kriterierne

- **Tidsrækkefølge**
- **Empirisk sammenhæng**
- Fravær af **Spuriøsitet**
- **Teoretisk forklaring**

→ opfylder en sammenligning af wellness-deltagere og ikke-deltagere TEST-kriterierne?

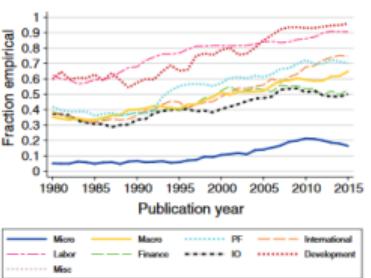


FIGURE 4. WEIGHTED FRACTION EMPIRICAL BY FIELD

Note: Five-year moving averages of the weighted fraction of publications in each field that are empirical.

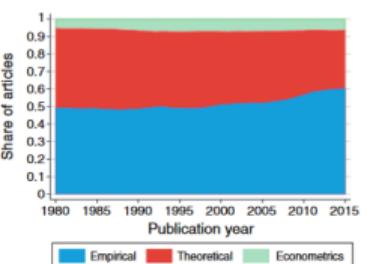


FIGURE 5. PUBLICATIONS BY STYLE

Note: Five-year moving averages of unweighted publication shares in each style.

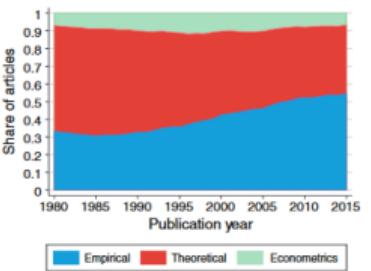


FIGURE 6. WEIGHTED PUBLICATIONS BY STYLE

Note: Five-year moving averages of weighted publication shares in each style.

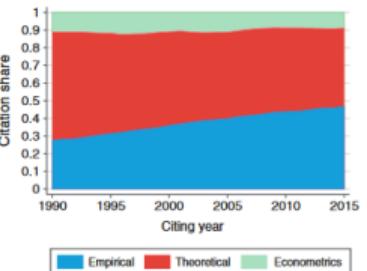


FIGURE 7. WEIGHTED CITATIONS BY STYLE

Note: Five-year moving averages of weighted citation shares in each style.

Angrist, J., Azoulay, P., Ellison, G., Hill, R., & Lu, S. F. (2017). Economic research evolves: Fields and styles. *American Economic Review*, 107(5), 293-97.

Kausal inferens i 'the Age of Heavy Metal': modelbaseret inferens

»The 1970s and early 1980s saw rapid growth in mainframe computer size and power. Stata had yet to appear, but magnetic tape jockeys managed to crunch more and more numbers in increasingly elaborate ways. For the most part, however, increased computing power did not produce more credible estimates.« (9)

1983: Leamer-kritikken

Table 2. A linear formulation of the VP-function

Level version

$$VP_t = [\alpha_1 \cdot Lp_t + \alpha_2 \cdot Lu_t + \dots]_e + [\beta_{1i} \cdot B_i + \beta_2 \cdot T_t + \{\beta_3 \cdot Lv_t + \dots\}]_p + r_t$$

Change version (in first differences)^a

$$\Delta VP_t = [\alpha_1 \cdot L\Delta p_t + \alpha_2 \cdot L\Delta u_t + \dots]_e + [\beta_{1i} \cdot B'_i + \beta_2 \cdot T'_t + \{\beta_3 \cdot L\Delta v_t + \dots\}]_p + e_t$$

General features/Conventions

VP the Vote or the Poll series used. VP_t is expressed in percent of all voters.

α, β coefficients to be estimated.

e_t, r_t residuals.

Δ first difference operator. The next variable is in change form: $\Delta z_t = z_t - z_{t-1}$.

L backshift/lag operator. The next variable may be lagged, often by less than one year.

... indicate that there may be more variables.

Economic variables. The e-part in the \ll_e -brackets

p_t first economic series. Taken to be inflation.

u_t second economic series. Taken to be the rate of unemployment.

Political variables. The p-part in the \ll_p -brackets^b

B_i constant(s). Typically broken up to be government specific.^c

T_t the trend. It is normally constrained to be the same under all governments, d
be government specific like the constant.

v_t political variable. It may be another dummy variable to take care of a special interest group. It may be a genuine variable like the Rally-Around-the-Flag-variable.

Oprindeligt svar til Leamer-kritikken: flere, vildere robusthedstests

I Just Ran Two Million Regressions

By XAVIER X. SALA-I-MARTIN*

Following the seminal work of Robert Barro (1991), the recent empirical literature on economic growth has identified a substantial number of variables that are partially correlated with the rate of economic growth. The basic methodology consists of running cross-sectional regressions of the form

$$(1) \quad \boldsymbol{\gamma} = \boldsymbol{\alpha} + \boldsymbol{\beta}_1 \mathbf{x}_1 + \boldsymbol{\beta}_2 \mathbf{x}_2 + \cdots + \boldsymbol{\beta}_n \mathbf{x}_n + \varepsilon$$

where $\boldsymbol{\gamma}$ is the vector of rates of economic growth, and $\mathbf{x}_1, \dots, \mathbf{x}_n$ are vectors of explanatory variables, which vary across researchers and across papers. Each paper typically reports a (possibly nonrandom) sample of the regressions actually run by the researcher. Variables like the initial level of income, the investment rate, various measures of education, some poli-

An initial answer to this question was given by Ross Levine and David Renelt (1992).¹ They applied Edward Leamer's (1985) *extreme-bounds test* to identify 'robust' empirical relations in the economic growth literature. In short, the extreme-bounds test works as follows. Imagine that there is a pool of N variables that previously have been identified to be related to growth and one is interested in knowing whether variable z is "robust." One would estimate regressions of the form

$$(2) \quad \boldsymbol{\gamma} = \boldsymbol{\alpha}_j + \boldsymbol{\beta}_{yz} \mathbf{y} + \boldsymbol{\beta}_{zj} z + \boldsymbol{\beta}_{xj} \mathbf{x}_j + \varepsilon$$

where \mathbf{y} is a vector of variables that always appear in the regressions (in the Levine and Renelt paper, these variables are the initial level of income, the investment rate, the secondary school enrollment rate, and the rate

Kritik af Sala-i-Martin:

»Sala-i-Martin's (1997) investigation of extreme bounds analysis must have been fun. Happily, however, this kind of agnostic specification search has not emerged as a central feature of contemporary empirical work. Although Sala-i-Martin succeeds in uncovering some robustly significant relations (the “fraction of the population Confucian” is a wonderfully robust predictor of economic growth), we don't see why this result should be taken more seriously than the naive capital punishment specifications criticized by Leamer. (...) Wide-net searches of this kind offer little basis for a causal interpretation.« (17)

Meanwhile, i statskundskaben... Putnam (2007)

Table 3. Predicting Trust in Neighbours from Individual and Contextual Variables

	B	S. E.	Beta	t	Sig.
(Constant)	0.79	0.11		7.0	0.0000
R's age	0.01	0.00	0.15	21.4	0.0000
R owns home (v. rent)	0.25	0.01	0.13	19.7	0.0000
R's education (years)	0.04	0.00	0.13	19.1	0.0000
R's ethnicity: black	-0.31	0.02	-0.12	-18.6	0.0000
Census tract poverty rate	-0.66	0.09	-0.08	-7.1	0.0000
R's satisfaction with current finances	0.10	0.01	0.08	12.4	0.0000
R's ethnicity: Latino	-0.24	0.02	-0.07	-9.8	0.0000
R's household income (\$100,000)	0.14	0.02	0.05	7.5	0.0000
County: Non-violent Crimes per Capita	-2.57	0.41	-0.05	-6.2	0.0000
Census tract Herfindahl Index of Ethnic Homogeneity	0.18	0.04	0.04	5.1	0.0000
Census Tract Population Density (100,000 per sq. mi)	-0.39	0.08	-0.04	-4.8	0.0000
Census Tract Percent Living Same Town as Five Years Earlier	-0.24	0.04	-0.04	-5.4	0.0000
R's decades in this community	.020	.004	0.04	5.3	0.0000
Census Tract Percent Renters	-0.14	0.04	-0.04	-3.5	0.0006
Census Tract Percent Bachelor's Degree	0.29	0.07	0.03	4.3	0.0000
R is Spanish-speaker	-0.13	0.03	-0.03	-4.1	0.0001
R is female	0.05	0.01	0.03	4.7	0.0000
Census Tract Gini Coefficient for Household Income	0.39	0.15	0.02	2.7	0.0069
Census Tract Average Commute Time (hours)	-0.21	-0.06	-0.02	-3.4	0.0006
R's ethnicity: Asian	-0.09	0.03	-0.02	-3.3	0.0011
Census Tract Percent United States Citizens	0.21	0.09	0.02	2.2	0.0264
County: Violent Crimes per Capita	6.59	3.35	0.02	2.0	0.0489
Census Tract Percent Over 65	0.21	0.10	0.01	2.1	0.0364
R is a citizen	0.06	0.03	0.01	2.1	0.0356
R's average monthly work hours	.002	.001	0.01	1.8	0.0732
R is resident of South	-0.02	0.02	-0.01	-1.2	0.2182
R is resident of Midwest	-0.02	0.02	-0.01	-1.0	0.3296
R is resident of West	0.01	0.02	0.01	0.8	0.4238
R's commuting time (hours)	-0.00	0.01	0.00	-0.2	0.8069

Notes: Question was 'How much can you trust people in your neighbourhood?' N = 23,260.
Adj. R² = 0.26.

Drejning sidenhen: 'credibility-revolutionen'

»Empirical microeconomics has experienced a **credibility revolution**, with a consequent increase in policy relevance and scientific impact. Sensitivity analysis played a role in this, but as we see it, the primary engine driving improvement has been a focus on the quality of empirical research designs. This emphasis on research design is in the spirit of Leamer's critique, but it did not feature in his remedy.« (4)

modelbaseret inferens → designbaseret inferens

Hvorfor skete credibility-revolutionen?

- bedre og mere data
- færre distraktioner sfa. bedre forståelse for robusthed
- reorientering mod bedre designs
- mere transparente diskussioner af identifikationsstrategier
- frustration med uendelige robusthedstests (jf. Sala-i-Martin)

Opsamling

○○○

Credibility-revolutionen

Motivation

○○

Angrist & Pischke (2010)

○○○○○○○●

Samii (2016)

○○○○○○○○

Case: Caroll (2018)

○○○

Kig fremad

○○

Kritikker af credibility-revolutionen

- ekstern validitetskritik: fokus på alt for 'lokale' effekter
- substanskritik: 'freakonomics'-forskning uden 'econ'

Den klassiske tilgang: masseproduktion af 'pseudo-general pseudo-facts'

»At the turn of the millennium, the modal quantitative research design was one in which researchers assembled data on theoretically interesting dependent and independent variables (...) Researchers then **assessed the presumably causal relationships in these data using regressions** with informally motivated sets of control variables to reduce the potential for confounding.«

Sidenhen: credibility-revolutionen

»This convention in quantitative causal research appears to be breaking down, and more quantitative causal research is moving toward causal empiricism. This (...) represents a major change in what researchers believe are **credible ways of doing causal inference**.«

Opsamling

○○○

'Pseudo-general pseudo-facts'

Motivation

○○

Angrist & Pischke (2010)

○○○○○○○○

Samii (2016)

○●○○○○○○○

Case: Caroll (2018)

○○○○

Kig fremad

○○

Problemer i klassiske regressionstilgange:

- ① mgl. ekstern validitet: nominel ctr. effektiv stikprøve
- ② mgl. intern validitet: misspecifikation

Opsamling

○○○

'Pseudo-general pseudo-facts'

Motivation

○○

Angrist & Pischke (2010)

○○○○○○○○

Samii (2016)

○○●○○○○○○

Case: Carroll (2018)

○○○○

Kig fremad

○○

Ad (1):



Figure 1. Nominal and effective samples from Jensen (2003), reproduced from Aronow and Samii (2016)

Opsamling

○○○

'Bad controls'

Motivation

○○

Angrist & Pischke (2010)

○○○○○○○○

Samii (2016)

○○○●○○○○○

Case: Carroll (2018)

○○○○

Kig fremad

○○

Ad (2):

»[W]hen thinking about controls, timing matters. Variables measured before the variable of interest was determined are generally good controls. (...) Because these variables were determined before the variable of interest, they cannot themselves be outcomes in the causal nexus.«

Opsamling

○○○

'Bad controls'

Motivation

○○

Angrist & Pischke (2010)

○○○○○○○○

Samii (2016)

○○○○●○○○○

Case: Carroll (2018)

○○○○

Kig fremad

○○

Table 1. Replication and Auxiliary Analyses for Laitin and Fearon (2003)

	Outcome						
	Civil War Onset						Per Capita Income (7)
	(1)	(2)	(3)	(4)	(5)	(6)	
Estimator	Logit	Logit	Logit	Logit	Logit	Logit	OLS
Prior war	-.95 ** (.31)				-.24 (.23)	-.38 (.25)	
Per capita income	-.34*** (.07)			-.29*** (.07)		-.29*** (.07)	
Ethnic fractionalization	.17 (.37)	1.12*** (.33)	1.12** (.42)	.35 (.39)	1.16** (.43)	.40 (.40)	-4.14*** (.90)
Observations	6,327	6,610	6,610	6,373	6,610	6,373	6,373
Country-clustered SEs			Y	Y	Y	Y	Y

Vi antager det gammelkendte setup m. privat uddannelse P_i , indkomst Y_i og confounder ability A_i :

$$Y_i = \alpha + \beta P_i + \gamma A_i + e_i \quad (1)$$

men vi kan kun observere 'late ability' A_{li} , som delvist er en konsekvens af privat uddannelse P_i :

$$A_{li} = \pi_0 + \pi_1 P_i + \pi_2 A_i \quad (2)$$

Opsamling

○○○

'Bad controls'

Motivation

○○

Angrist & Pischke (2010)

○○○○○○○○

Samii (2016)

○○○○○○●○○

Case: Caroll (2018)

○○○

Kig fremad

○○

Konsekvens: vores estimat af β afhænger af γ , π_1 og π_2 :

$$Y_i = \left(\alpha - \gamma \frac{\pi_0}{\pi_2} \right) + \left(\beta - \gamma \frac{\pi_1}{\pi_2} \right) P_i + \frac{\gamma}{\pi_2} A_{li} + e_i \quad (3)$$

→ A_{li} er en 'bad control', medfører *post-treatment adjustment bias*

Opsamling

○○○

'Bad controls'

Motivation

○○

Angrist & Pischke (2010)

○○○○○○○○

Samii (2016)

○○○○○○○●○

Case: Caroll (2018)

○○○

Kig fremad

○○

GOOGLE

Former Employees Are Suing Google Over Alleged Gender Discrimination



Melanie Ehrenkranz

9/14/17 4:00pm • Filed to: GOOGLE ▾

0 72 2



<https://gizmodo.com/>

Opsamling

○○○

'Bad controls'

Motivation

○○

Angrist & Pischke (2010)

○○○○○○○○

Samii (2016)

○○○○○○○○●

Case: Caroll (2018)

○○○○

Kig fremad

○○

»After the New York Times detailed the employee spreadsheets on Friday, Google spokesperson Gina Scigliano told Gizmodo that its own data shows, when you take “location, tenure, job role, level and performance” into account, that “women are paid 99.7% of what men are paid at Google.” Scigliano described the *Times* story as “extremely flawed.”«

hvad fortæller tallet 99.7 pct. os? hvad er problemet med den tolkning Scigliano lægger op til?

Opsamling
○○○

Motivation
○○

Angrist & Pischke (2010)
○○○○○○○○

Samii (2016)
○○○○○○○○

Case: Caroll (2018)
●○○○

Kig fremad
○○

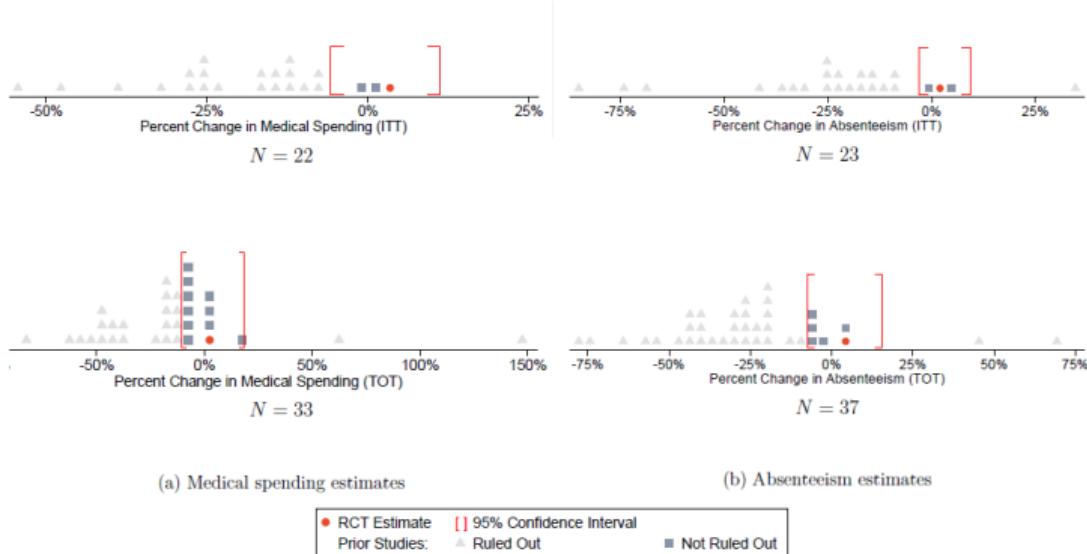
Case: arbejdspladswellness-programmer



A yoga class at a New York wellness center. Randomized controlled trials can reverse the conclusions of observational studies. Chad Rhym for The New York Times

»Here's the nerdy fun part, though. In addition to this analysis, the researchers also took the time to analyze the data as if it were an observational trial. In other words, they took the 3,300 who were offered the wellness program, then analyzed them the way a typical observational trial would, comparing those who participated with those who didn't.«

Figure 8: Comparison of experimental estimates to prior studies



Notes: Each figure shows the distribution of N point estimates from prior workplace wellness studies. Panel (a) plots intent-to-treat (ITT) and treatment-on-the-treated (TOT) estimates for medical spending. Panel (b) plots corresponding estimates for absenteeism. The point estimates from our own study ("RCT Estimate"), and their associated confidence intervals, are taken from Table 6, Column 3, for medical spending, and Table 4, Column 4 and Table 5, Column 3 for absenteeism. Our RCT estimates and confidence intervals are plotted in order to demonstrate the share of prior study point estimates we are able to rule out. Appendix Table B.1 provides the full details of this meta-analysis.

Jones, D., Molitor, D., & Reif, J. (2018). What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study (No. w24229). *National Bureau of Economic Research*.

Komplikationer ved eksperimenter:

»[Randomly assigning participation is] hard to do with a wellness program, especially since they are put into place for the entire company.« ← clustering

»Of course, not everyone given an opportunity to engage in such activities will take it, but more than half of those offered the program participated.« ← noncompliance

mere om begge i næste uge!

Eksperimenter I: potential outcomes framework og fordele ved randomisering

- Angrist & Pischke kap. 1
- Gerber & Green kap. 1+2
- fokus i begge: potential outcomes framework (i GG mest i kap. 2)
- case-tekst: GGL (2008) → kig på eksperimentbeskrivelse s. 36-39

Opsamling

○○○

Motivation

○○

Angrist & Pischke (2010)

○○○○○○○○

Samii (2016)

○○○○○○○○○○

Case: Caroll (2018)

○○○○

Kig fremad

○●

Tak for i dag!