Quantitative Methods in Economics
Introduction to observational studies

Maximilian Kasy

Harvard University, fall 2016

2) Estimation of Treatment Effects in Observational Studies

Observational Studies

- Randomization is called the “gold standard” for causal inference because it balances observed and unobserved confounders
- Cannot always randomize so we do observational studies, where we can directly adjust for the observed variables and use indirect methods to adjust for unobserved variables
- We want to **design** observational studies that **approximate experiments**:
  - “The planner of an observational study should always ask himself: How would the study be conducted if it were possible to do it by controlled experimentation” (Cochran 1965)

Approximating Experiments

Treatments, Covariates, Outcomes

It is important to distinguish between:

Covariates: Pre-treatment variables, potential confounders
Outcomes: Variables potentially affected by the treatment

- **Randomized Experiment**: Well-defined treatment, clear distinction between covariates and outcomes
- **Better Observational Study**: Well-defined treatment, clear distinction between covariates and outcomes
- **Poorer Observational Study**: Hard to say when treatment began or what the treatment really is. Distinction between covariates and outcomes is blurred. No baseline survey.

Approximating Experiments

How were treatments assigned?

- **Randomized Experiment**: Random assignment
- **Better Observational Study**: Assignment is not random, but assignment mechanism is clearly described. Try to find "natural experiments", where assignment is "as good as random"
- **Poorer Observational Study**: No attention given to the assignment mechanism

Approximating Experiments

Were treated and controls comparable?

- **Randomized Experiment**: Balance table for observables.
- **Better Observational Study**: Balance table for observables. Ideally sensitivity analysis for unobservables.
- **Poorer Observational Study**: No direct assessment of comparability is presented.

Approximating Experiments

Eliminating plausible alternatives to treatment effects?

- **Randomized Experiment**: List plausible alternatives and experimental design includes features that shed light on these alternatives (e.g. placebos). Report on potential attrition and non-compliance.
- **Better Observational Study**: List plausible alternatives and study design includes features that shed light on these alternatives (e.g. multiple control groups, data on potential confounders, etc.)
- **Poorer Observational Study**: Alternatives are mentioned in the discussion section

Approximating Experiments

Study Protocol

- **Randomized Experiment**: Before the experiment starts, a protocol describes the design, outcomes, type of analysis, etc
- **Better Observational Study**: Before the analysis of the data starts, a protocol describes the design, outcomes, type of analysis, etc
- **Poorer Observational Study**: If we run many regressions, something publishable will turn up sooner or later ...

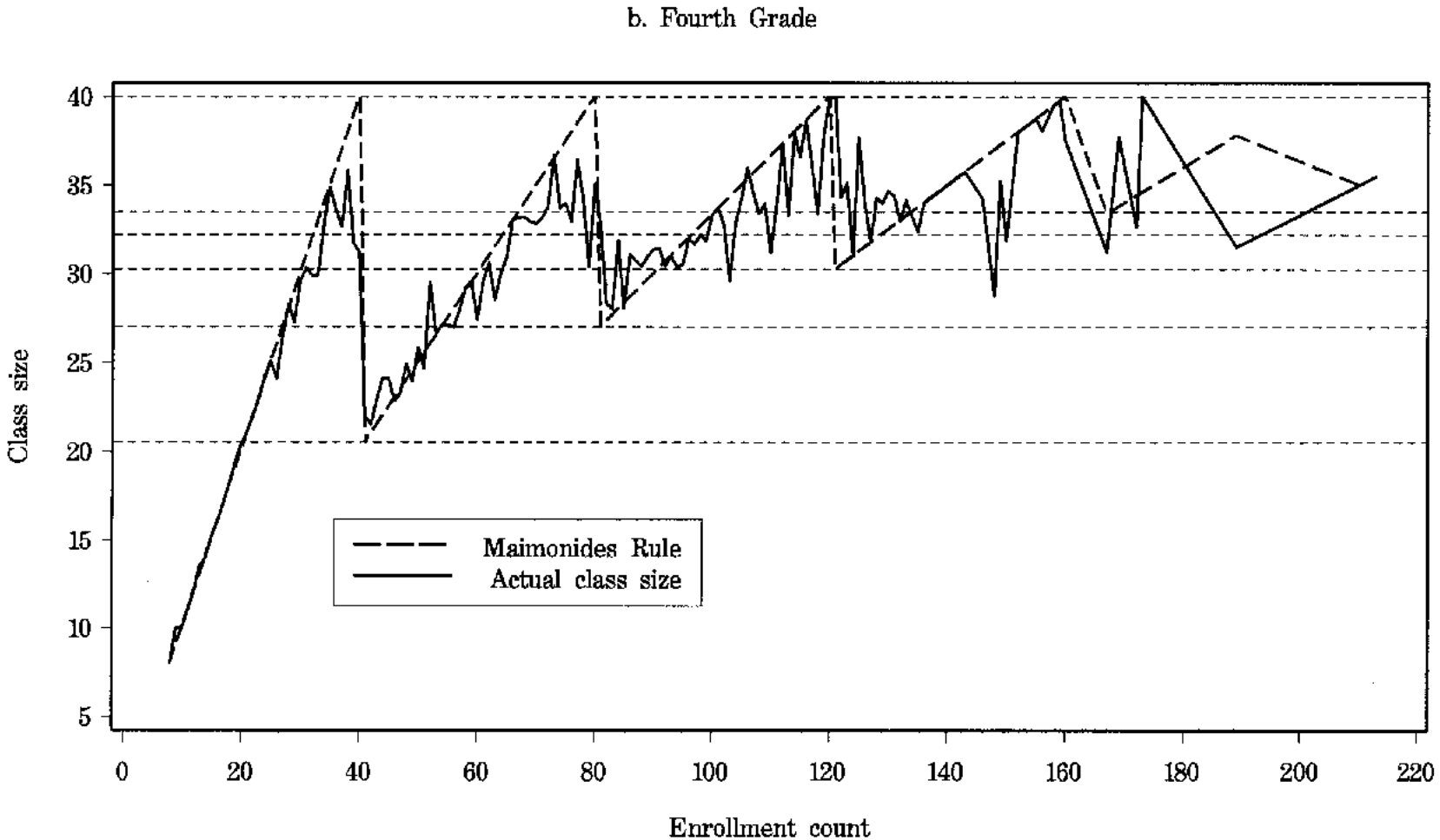
Observational Studies

Design features we can use to handle unobservables:

- Design comparisons so that unobservables are likely to be balanced (e.g. “homogeneous” sub-samples, groups where treatment assignment was “accidental”)
- Difference-in-differences: unobservables may differ, but their effect may not change much in time
- Instrumental variables: find variables that “randomize” some people into treatment
- Sensitivity analysis and bounds

Class Size on Student Achievement

Angrist and Lavy (1999): Maimonides' rule



Health Effects of Smoking

Kaprio and Koskenvuo (1989): Monozygotic twins

TABLE 4. The Finnish twin study. First death by smoking status among smoking-discordant twin pairs. Kaprio and Koskenvuo (1989).

	Smokers	Non-smokers
All causes	17	5
Coronary heart disease	9	0
Lung cancer	2	0

Seat Belts on Fatality Rates

Evans (1986): Two front seat passengers

Table 1.1 Crashes in FARS 1975–1983 in which the front seat had two occupants, a driver and a passenger, with one belted, the other unbelted, and one died and one survived.

	Driver Passenger	Not Belted Belted	Belted Not Belted
Driver Died	Passenger Survived	189	153
Driver Survived	Passenger Died	111	363

Fertility and Labor Market Participation

Angrist and Evans (1988): Gender composition of previous children

TABLE 5—WALD ESTIMATES OF LABOR-SUPPLY MODELS

	1980 PUMS			1990 PUMS			1980 PUMS		
		Wald estimate using as covariate:			Wald estimate using as covariate:			Wald estimate using as covariate:	
Variable	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Same sex*	*More than 2 children*	*Number of children*	Mean difference by *Twins-2*	*More than 2 children*	*Number of children*
More than 2 children	0.0600 (0.0016)	—	—	0.0628 (0.0016)	—	—	0.6031 (0.0084)	—	—
Number of children	0.0765 (0.0026)	—	—	0.0836 (0.0025)	—	—	0.8094 (0.0139)	—	—
Worked for pay	−0.0080 (0.0016)	−0.133 (0.026)	−0.104 (0.021)	−0.0053 (0.0015)	−0.084 (0.024)	−0.063 (0.018)	−0.0459 (0.0086)	−0.076 (0.014)	−0.057 (0.011)
Weeks worked	−0.3826 (0.0709)	−6.38 (1.17)	−5.00 (0.92)	−0.3233 (0.0743)	−5.15 (1.17)	−3.87 (0.88)	−1.982 (0.386)	−3.28 (0.63)	−2.45 (0.47)
Hours/week	−0.3110 (0.0602)	−5.18 (1.00)	−4.07 (0.78)	−0.2363 (0.0620)	−3.76 (0.98)	−2.83 (0.73)	−1.979 (0.327)	−3.28 (0.54)	−2.44 (0.40)
Labor income	−132.5 (34.4)	−2208.8 (569.2)	−1732.4 (446.3)	−119.4 (42.4)	−1901.4 (670.3)	−1428.0 (502.6)	−570.8 (186.9)	−946.4 (308.6)	−705.2 (229.8)
ln(*Family income*)	−0.0018 (0.0041)	−0.029 (0.068)	−0.023 (0.054)	−0.0085 (0.0047)	−0.136 (0.074)	−0.102 (0.056)	−0.0341 (0.0223)	−0.057 (0.037)	−0.042 (0.027)

Notes: The samples are the same as in Table 2. Standard errors are reported in parentheses.