

Empirical Tools of Public Finance

131 Undergraduate Public Economics
Emmanuel Saez
UC Berkeley

DEFINITIONS

Empirical public finance: The use of data and statistical methods to measure the impact of government policy on individuals and markets (example: how an increase of taxes affects work behavior)

Correlation: Two economic variables are correlated if they move together (example: height and weight across individuals)

Causality: Two economic variables are causally related if the movement of one causes movement of the other (example: good nutrition as an infant increases adult height)

THE IMPORTANT DISTINCTION BETWEEN CORRELATION AND CAUSALITY

There are many examples where causation and correlation can get confused.

In statistics, this is called the *identification problem*: given that two series are correlated, how do you identify whether one series is causing another?

THE IDENTIFICATION PROBLEM

The attempt to interpret a correlation as a causal relationship without sufficient thought to the underlying process generating the data is a common problem.

For any correlation between two variables A and B, there are three possible explanations, one or more of which could result in the correlation:

- 1) A is causing B
- 2) B is causing A
- 3) Some third factor is causing both

The general problem that empirical economists face in trying to use existing data to assess the causal influence of one factor on another is that one cannot immediately go from correlation to causation.

RANDOMIZED TRIALS AS A SOLUTION

Randomized trial: The ideal type of experiment designed to test causality, whereby a group of individuals is randomly divided into a treatment group, which receives the treatment of interest, and a control group, which does not.

Treatment group: The set of individuals who are subject to an intervention being studied.

Control group: The set of individuals comparable to the treatment group who are not subject to the intervention being studied.

Randomized trials have been used in medicine for many decades and have become very popular in economics, especially development economics in the last 15 years

THE PROBLEMS OF BIAS

Bias: Any source of difference between treatment and control groups that is correlated with the treatment but is not due to the treatment.

Having large sample sizes allows researchers to eliminate any consistent differences between groups by relying on the statistical principle called *the law of large numbers*: the odds of getting the wrong answer approaches zero as the sample size grows.

Statisticians develop methods to evaluate the precision of estimates and create confidence intervals around estimates

EXAMPLES OF RANDOMIZED TRIALS

Randomized Trials of ERT (estrogen replacement therapy)

The randomized trial of ERT tracked over 16,000 women ages 50–79 who were recruited to participate in the trial by 40 clinical centers in the United States. The study was supposed to last 8.5 years but was stopped after 5.2 years because its conclusion was already clear: ERT did in fact raise the risk of heart disease.

Randomized Trials in the Welfare Reform (TANF) Context

Randomized trials are equally useful in the context of public policy. A number of states implemented randomized trials to test various aspects of proposed welfare reform.

WHY WE NEED TO GO BEYOND RANDOMIZED TRIALS

Even the gold standard of randomized trials has some potential problems.

1) External validity: The results are only valid for the sample of individuals who volunteer to be either treatments or controls, and this sample may be different from the population at large (e.g., randomized experiment in Sweden or US would not necessarily generate the same results)

2) Attrition: Individuals may leave the experiment before it is complete. Reduction in the size of samples over time, which, if not random, can lead to bias estimates.

Outside randomized experiments, bias is a pervasive problem that is not easily remedied. There are, however, methods available that can allow us to approach the gold standard of randomized trials.

OBSERVATIONAL DATA

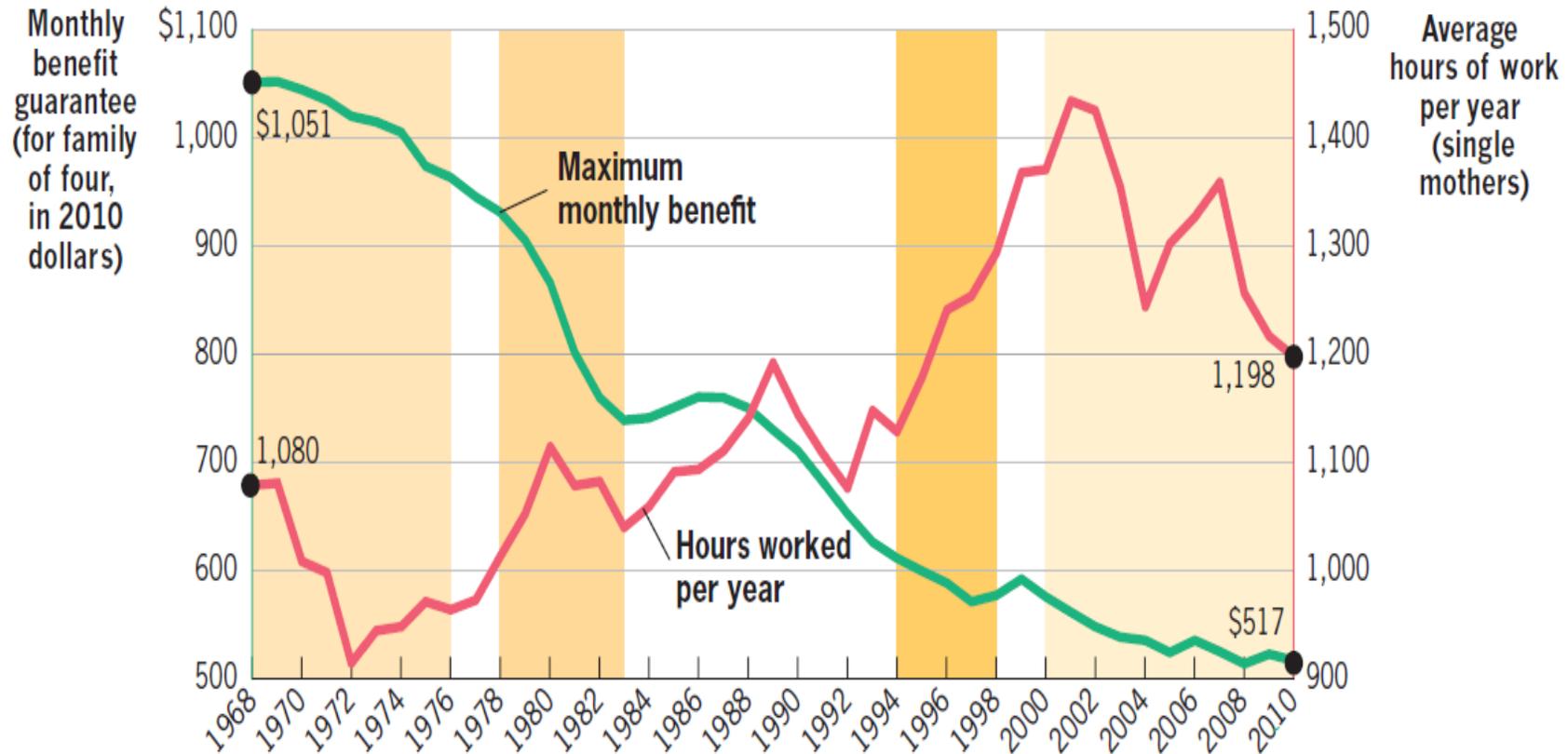
Observational data: Data generated by individual behavior observed in the real world, not in the context of deliberately designed experiments.

Time series analysis: Analysis of the co-movement of two series over time.

Cross-sectional regression analysis: Statistical analysis of the relationship between two or more variables exhibited by many individuals at one point in time.

3.3

Time Series Analysis: Cash Welfare Guarantee and Hours Worked Among Single Mothers



Time series analysis: Analysis of two series over time.

PROBLEMS WITH TIME SERIES ANALYSIS

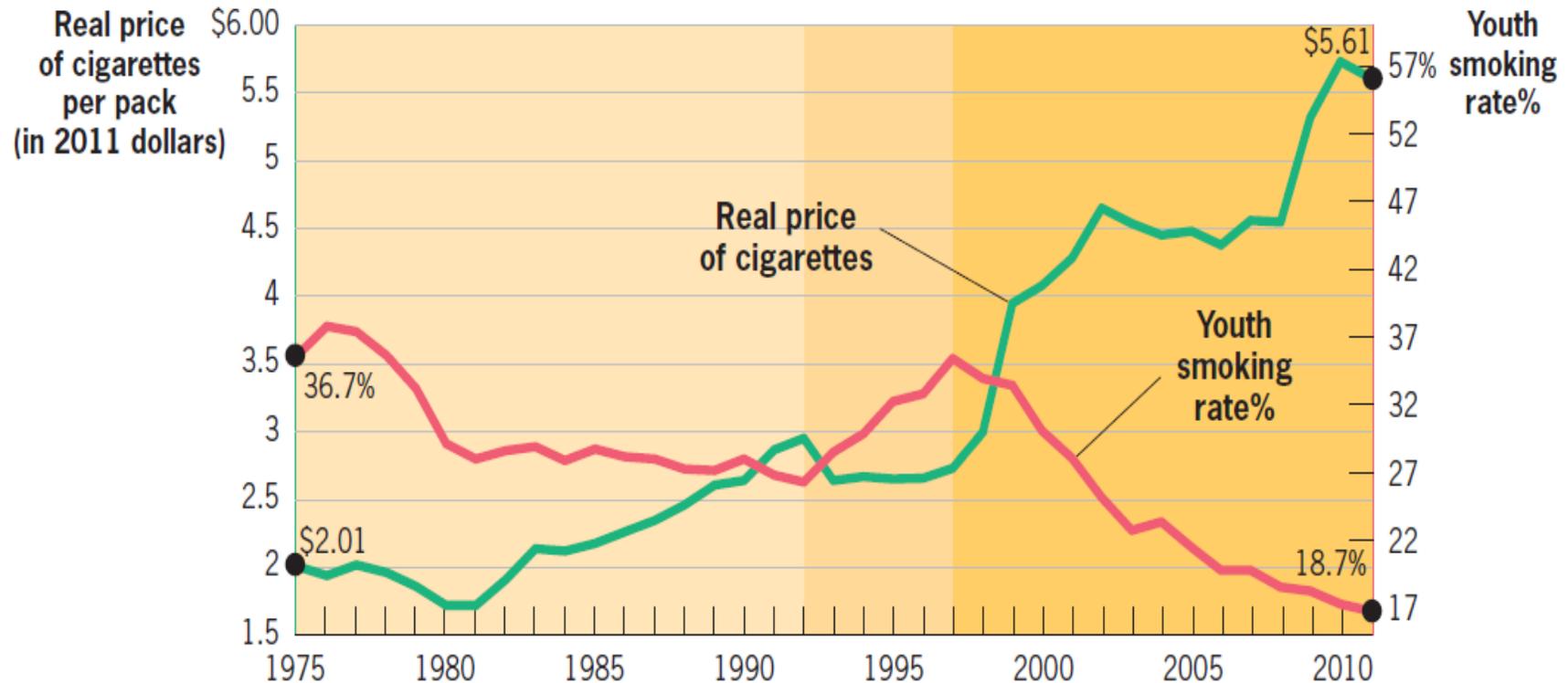
1) Although this time series correlation is striking, it does not necessarily demonstrate a causal effect of TANF benefits on labor supply

When there is a slow-moving trend in one variable through time, as is true for the general decline in income guarantees over this period, it is very difficult to infer its causal effects on another variable.

2) Other factors get in the way of a causal interpretation of this correlation over time; factors such as economic growth and a more generous Earned Income Tax Credit (EITC) can cause bias in this time series analysis because they are also correlated with the outcome of interest.

3.3

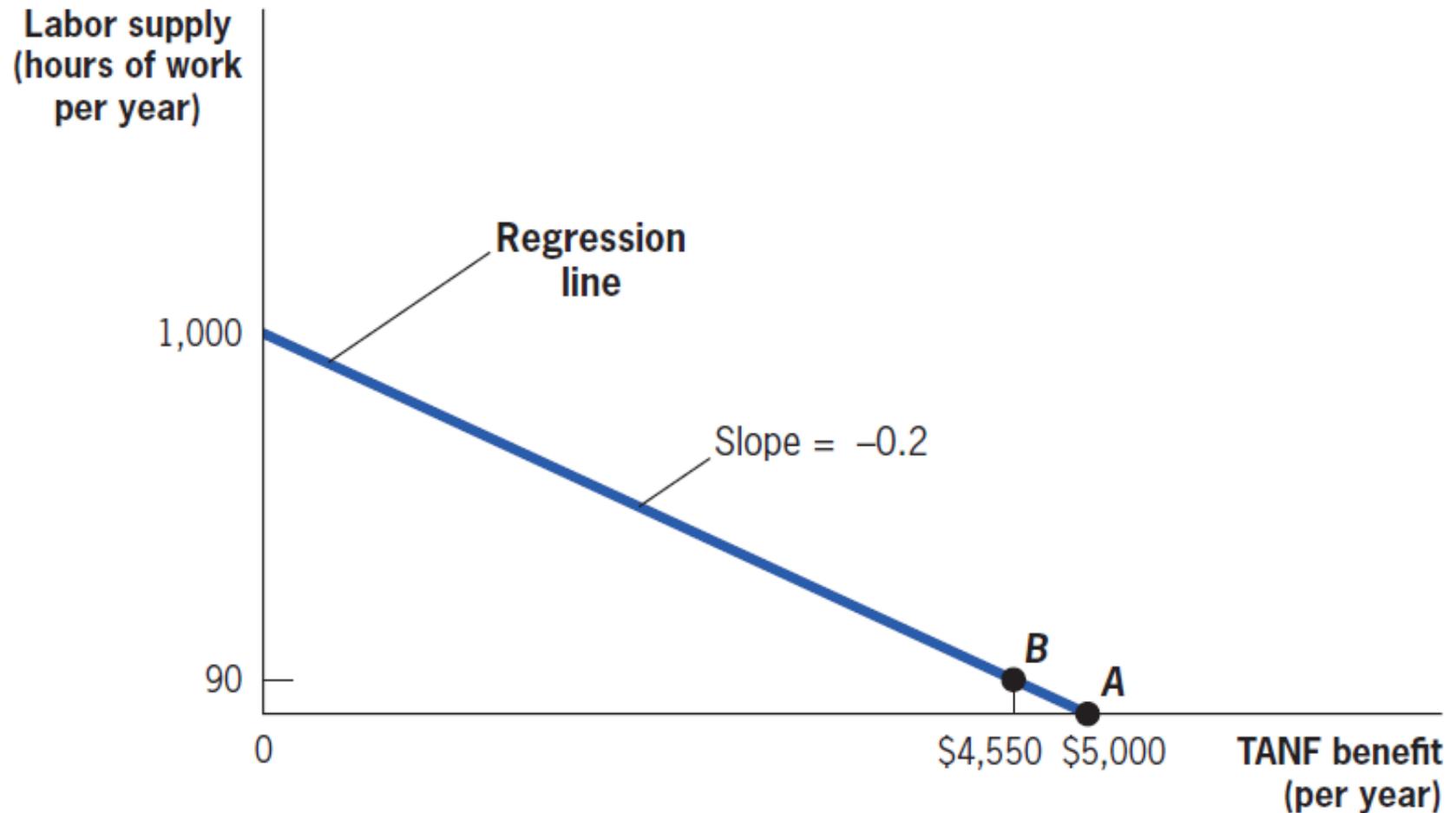
When Is Time Series Analysis Useful? Cigarette Prices and Youth Smoking



- Sharp, simultaneous changes in prices and smoking rates in 1993 and 1998–onward
- Known causes: price war, tobacco settlements

3.3

Cross-Sectional Regression Analysis: Labor Supply and TANF Benefit



REGRESSION

Regression line: The line that measures the best linear approximation to the relationship between any two variables.

$$Y = X\beta + \varepsilon$$

X is the independent variable data (TANF benefit guarantee)

Y is the dependent variable data (labor supply)

β is the coefficient that measures the effect of X on Y

ε is the error term (captures variations in Y not related to X).

Ordinary least square regression (OLS) estimates β without bias if ε is not correlated with X

REGRESSION ESTIMATES

The estimated coefficient $\hat{\beta}$ is reported with standard errors in parentheses

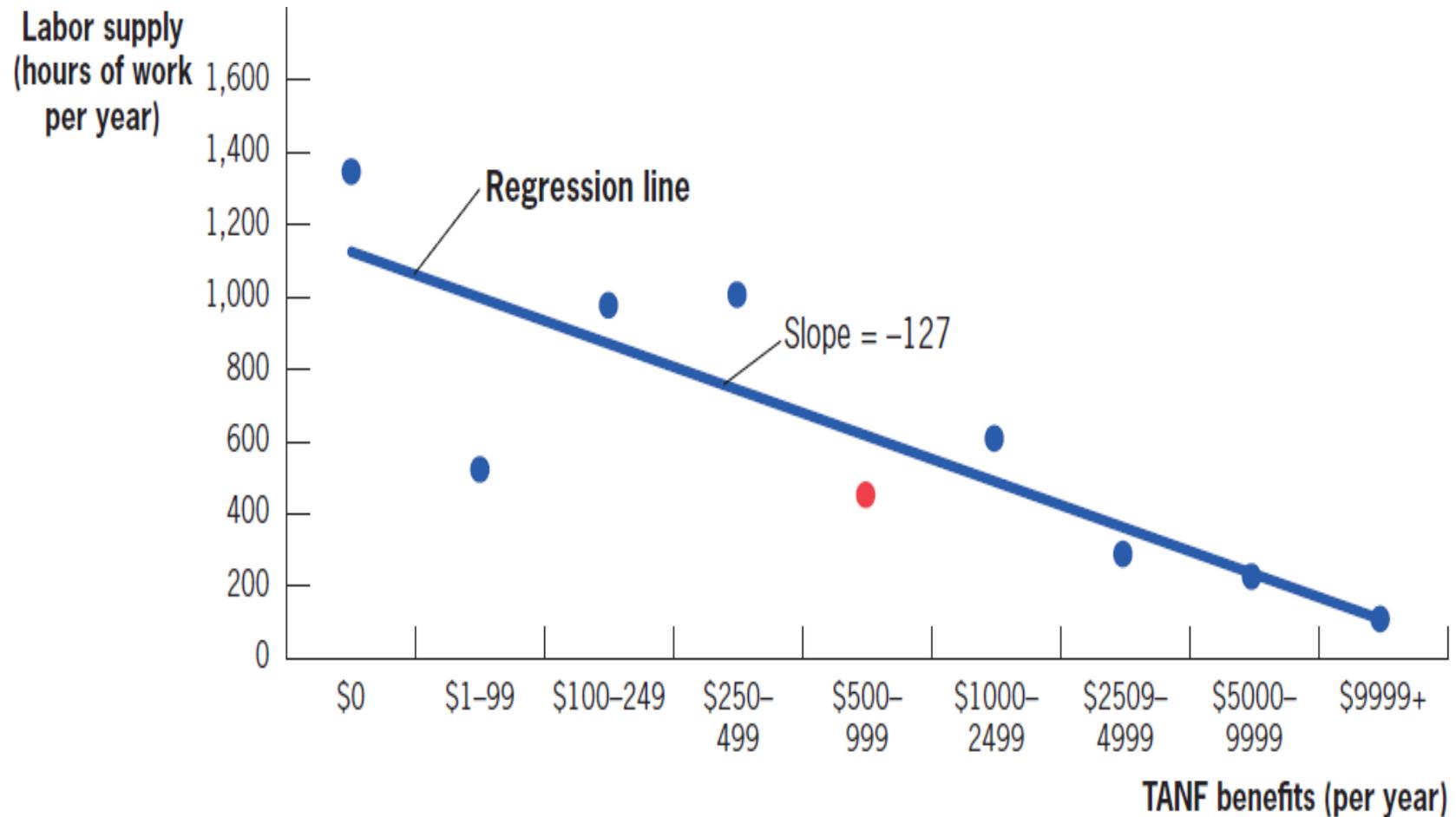
Example: $\hat{\beta} = .5(.1)$ should be understood as β is in confidence interval $(.5 - 2 \cdot .1, .5 + 2 \cdot .1) = (.3, .7)$ with probability 95%.

We have standard errors because we do not know the exact value of β

When estimated coefficient is more than twice the standard error, we can conclude that it is significantly positive (i.e., is above zero with probability 95%).

3.3

Example with Real-World Data: Labor Supply and TANF Benefits



PROBLEMS WITH CROSS-SECTIONAL REGRESSION ANALYSIS

The result summarized in Figure 3-4 seems to indicate strongly that mothers who receive the largest TANF benefits work the fewest hours. Once again, however, there are several possible interpretations of this correlation.

One interpretation is that higher TANF benefits are causing an increase in leisure.

Another possible interpretation is that in places with high TANF benefits, mothers have a high taste for leisure and wouldn't work much even if TANF benefits weren't available (this means exactly that ε is correlated with X)

CONTROL VARIABLES

It is essential in all empirical work to ensure that there are no factors that cause consistent differences in behavior across two groups (ε) and are also correlated with the independent variable X

Control variables: Additional variables Z that are included in cross-sectional regression models to account for differences between treatment and control groups that can lead to bias

$$Y = X\beta + Z\gamma + \varepsilon$$

In TANF case, Z would include race, education, number of children to control for demographic differences across states

Empirically, add Z variables and assess whether they change the estimate β . If estimate β varies a lot, we cannot be confident that identification assumption holds

QUASI-EXPERIMENTS: DEFINITION

Quasi-experiments (also called natural experiments)

Changes in the economic environment that create nearly identical treatment and control groups for studying the effect of that environmental change, allowing public finance economists to take advantage of quasi-randomization created by external forces

Example: one state (Arkansas) decreases generosity of welfare benefits while another comparable state (Louisiana) does not. Single mothers in Arkansas are the Treatment (T) group, Single mothers in Louisiana are the control (C) group.

QUASI-EXPERIMENTS: ESTIMATION

We consider a Treatment group (T) and a Control group (C) and outcome Y

Simple difference estimator: $D = Y^{T,After} - Y^{C,After}$ is the difference in average outcomes between treatment and control after the change

In randomized experiment, simple difference $D = Y^{T,After} - Y^{C,After}$ is sufficient because T and C are identical before the treatment

In quasi-experiment, T and C might not be comparable before treatment. You can compute $D^{before} = Y^{T,Before} - Y^{C,Before}$

If $D^{before} = 0$, you can be fairly confident that $D = Y^{T,After} - Y^{C,After}$ estimates the causal effect

Difference-in-Difference estimator

If simple difference $D^{before} = Y^{T,Before} - Y^{C,Before}$ is not zero, you can form the **Difference-in-Difference estimator**

$$DD = [Y^{T,After} - Y^{C,After}] - [Y^{T,Before} - Y^{C,Before}]$$

This measures whether the difference between treatment and control changes after the policy change

DD identifies the causal effect of the treatment if, absent the policy change, the difference between T and C would have stayed the same (this is called the parallel trend assumption)

3.3

Benefits and Labor Supply in Arkansas and Louisiana

Arkansas			
	1996	1998	Difference
Benefit guarantee (\$)	5,000	4,000	-1,000
Hours worked	1,000	1,200	200
Louisiana			
	1996	1998	Difference
Benefit guarantee (\$)	5,000	5,000	0
Hours worked	1,050	1,100	50

PROBLEMS WITH QUASI-EXPERIMENTS

With quasi-experimental studies, we can never be completely certain that we have purged all bias from the treatment–control comparison.

Quasi-experimental studies present various robustness checks to try to make the argument that they have obtained a causal estimate.

Examples: find alternative control groups, do a placebo comparing treatment and control DD when no policy change took place, etc.

Best way to check validity of DD estimator is to plot times series and assess whether a clear break between the two groups happens at the time of the reform

TWO GRAPHICAL EXAMPLES

1) Effects of lottery winnings on labor supply from Imbens, Rubin, Sacerdote AER'01

Ideal quasi-experiment to measure income effects as lottery generates random assignment conditional on playing \Rightarrow Very compelling graph, DD is convincing

2) Effects of the 1987 EITC expansion (tax credit for low income workers with kids) on labor supply from Eissa and Liebman QJE'96

Compares single mothers (Treatment) to single females with no kids (Control) \Rightarrow No compelling break in graph around 1987, DD is not convincing

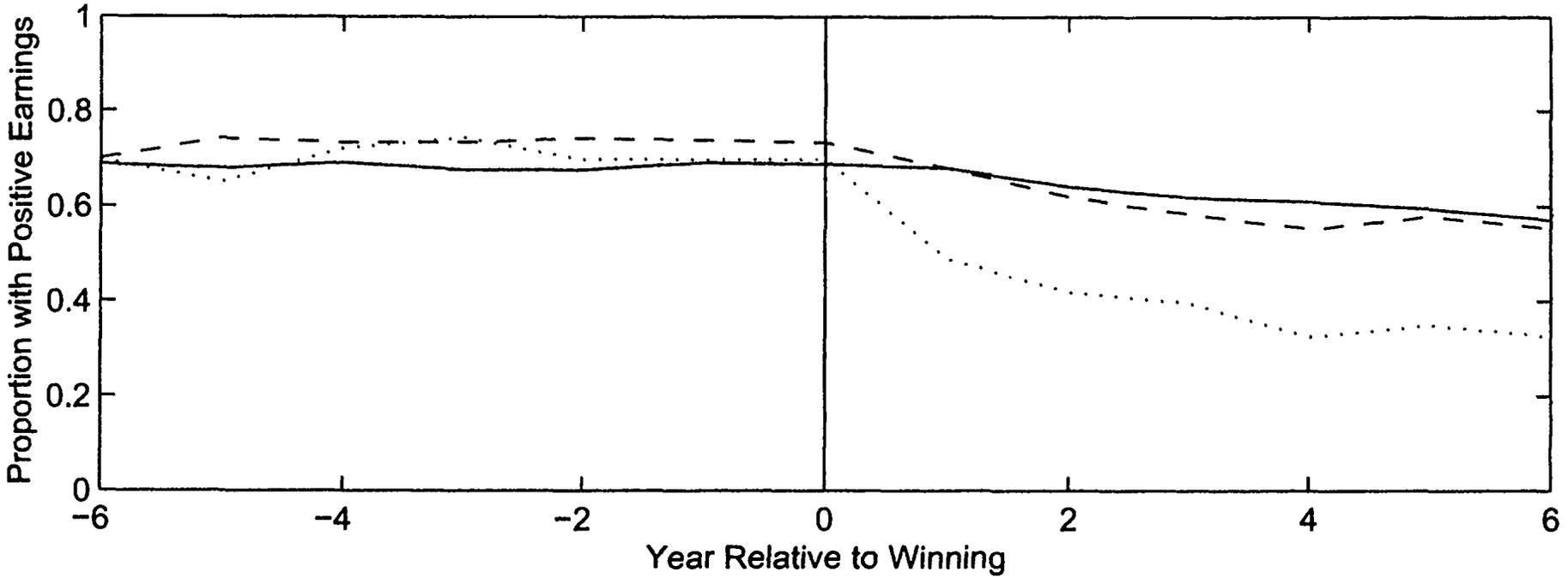
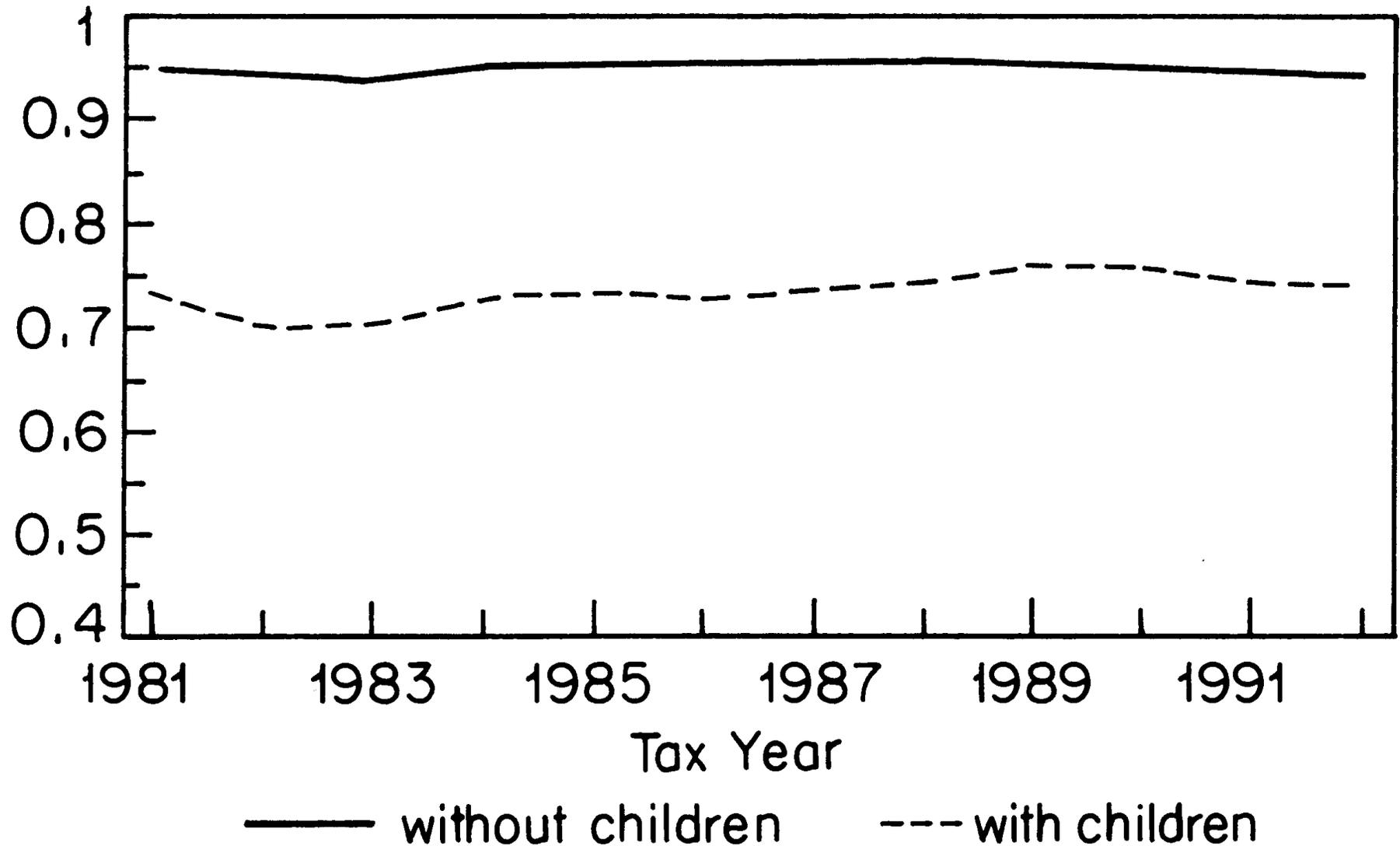


FIGURE 2. PROPORTION WITH POSITIVE EARNINGS FOR NONWINNERS, WINNERS, AND BIG WINNERS

Note: Solid line = nonwinners; dashed line = winners; dotted line = big winners.

Source: Imbens et al (2001), p. 784

All Unmarried Females



STRUCTURAL MODELING

Structural estimates: Builds a theoretical model of individual behavior and then estimates the parameters of the model. Estimates of the features that drive individual decisions, such as income and substitution effects or parameters of the utility function.

Reduced form estimates: Measures of the total impact of an independent variable on a dependent variable, without decomposing the source of that behavior response in terms of underlying parameters of the utility functions

Reduced form estimates are more transparent and convincing but structural estimates are more directly useful to make predictions for alternative policies

CONCLUSION

The central issue for any policy question is establishing a causal relationship between the policy in question and the outcome of interest.

We discussed several approaches to distinguish causality from correlation. The gold standard for doing so is the randomized trial, which removes bias through randomly assigning treatment and control groups.

Unfortunately, however, such trials are not available for every question we wish to address in empirical public finance. As a result, we turn to alternative methods such as time series analysis, cross-sectional regression analysis, and quasi-experimental analysis.

Each of these alternatives has weaknesses, but careful consideration of the problem at hand can often lead to a sensible solution to the bias problem that plagues empirical analysis.

REFERENCES

Jonathan Gruber, Public Finance and Public Policy, Fourth Edition, 2012
Worth Publishers, Chapter 3

Marquis, Grace S., et al. "Association of breastfeeding and stunting in Peruvian toddlers: an example of reverse causality." International Journal of Epidemiology 26.2 (1997): 349-356.(web)

Hotz, V. Joseph, Charles H. Mullin, and John Karl Scholz. "Welfare, employment, and income: evidence on the effects of benefit reductions from California." The American Economic Review 92.2 (2002): 380-384.(web)

Imbens, Guido W., Donald B. Rubin, and Bruce I. Sacerdote. "Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players." American Economic Review (2001): 778-794.(web)

Eissa, Nada, and Jeffrey B. Liebman. "Labor supply response to the earned income tax credit." The Quarterly Journal of Economics 111.2 (1996): 605-637.(web)