Empirical Methods, Fall 2025

Wojciech Kopczuk, adapted by Kyle Coombs

Vassar College

September 11, 2025

Evaluate EITC effects

- Now Gov. Hochul asks you to estimate the effects of the 2025 EITC increase on labor force participation and hours worked
- An aide says NY EITC recipients worked the same hours on average in 2025 as they did in 2024, so the EITC had no effect
- Any issues with this estimation strategy?
- What alternatives are there?
- What data would you want to answer this question?

Empirical Methods

The key problem: correlation is not causality.

Variables are **correlated** if they move together.

The relationship between variables is **causal** if one of the variables is causing movement in the other.

Examples:

- roosters and sunrise
- per capita cheese consumption and deaths by bedsheet entanglement
- education and income
- tax rates and income

More at https://www.tylervigen.com/spurious-correlations

Possible explanations of a correlation

Suppose that variables A and B are correlated. What are the possibilities?

- A is causing B
- B is causing A
- Some other factor is causing both A and B
- Accident — there is no true relationship (in small samples)

Identification problem: if variables are correlated, how can we establish whether one is causing the other?

Furthermore, we want to know the direction of causality **and** the strength of the effect (there may be *both* a causal relationship and correlation)

Extra challenge in economics: people optimize, which can offset or overstate a causal relationship

Randomization

- Ideal, infeasible experiment: apply different treatments (more education, different tax system etc.) to the same population in parallel universes.
- Randomly assigning treatment attempts to gets close to ideal
- Treatment and Control groups

  Endogeneity bias[1]: Differences between treatment and control that is *correlated* with but not due to the treatment.

  Exogeneity: Treatment is independent of the potential outcomes.
- Randomization means treatment and control differ only due to treatment
- The difference in outcomes is then the causal effect of the treatment

[1] Other forms of bias include sample selection, multicollinearity, misspecification, autocorrelation, heteroskedasticity, aggregation bias, publication bias, etc.

Potential problems with randomization

- Do it wrong
- Attrition (leaving the study)
- External validity (volunteers special, experiments stylized)
- Cost (expensive to enforce)
- Ethical problems (See IRB)

Examples of randomized studies in Public Economics

- Randomized tax enforcement experiments — info provision, audits
- Effect of explaining EITC incentives on income/labor supply
- Randomizing various aspects of 1996 welfare reform (job training, work requirements, case worker assistance)
- Public health insurance (Medicaid) assigned by lottery in Oregon
- Universal Basic Income experiments

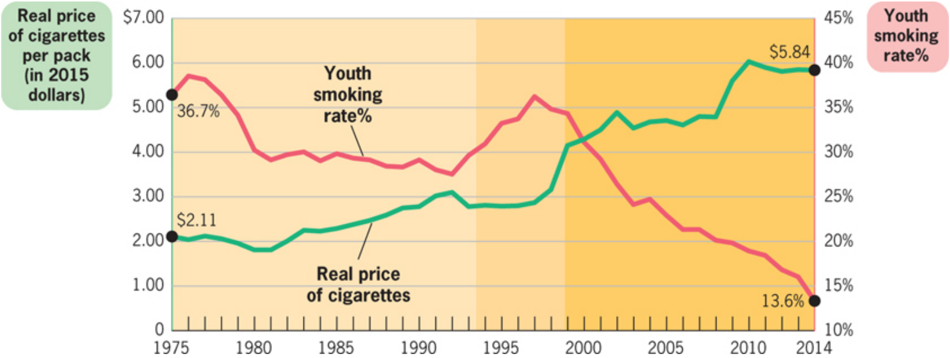
Observational data

- Data based on observation and measurement of actual behavior in the real world and not generated by an experiment
- **Time series**: observing (multiple) series over time
- **Cross-sectional**: observing many units (e.g., individuals, firms) once
- **Repeated cross-section**: a lot of units at different points in time (but potentially different ones at different points)
- **Panel data:** a lot of units that can be tracked over time

Time-series analysis

- Comparison of movement of variables over time
- Problem: too many things change over time, is 2003 a good control for 2004?
- Useful when there are sharp, repeated, and "isolated" changes in the treatment variable of interest

Price of cigarettes and youth smoking rate

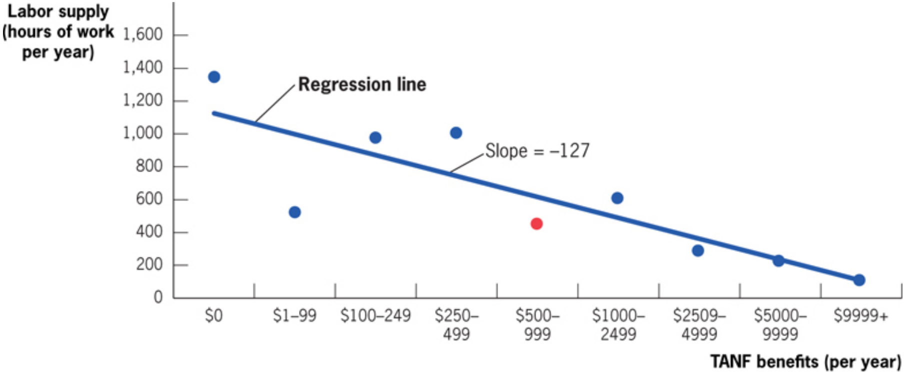


Gruber, *Public Finance and Public Policy*, Figure 3.1

Cross-sectional analysis

- Comparison of many individuals at one point in time
- Regression analysis: finding the best fitting relationship between the dependent variable (e.g., labor supply) and independent variables (e.g., welfare benefits, education, age)

Welfare benefits and labor supply



Gruber, *Public Finance and Public Policy*, Figure 3.4
What does the line capture?

Comments on regression analysis

- Econometric method to find the best fitting relationship: regression

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

- Results that it yields
  - coefficient estimate $\hat{\beta}$ — slope of the relationship (127 in the example)
  - standard error often in parentheses (e.g. 127 (25)), confidence interval, significance level of β — the precision of the estimate.
  - In the TANF example, 95% confidence interval is approximately $(78, 176)$ from $(\hat{\beta} - 1.96 \cdot \text{SE}, \hat{\beta} + 1.96 \cdot \text{SE})$

Problems with regression analysis

- Regression describes a relationship: $X \uparrow 1 \Leftrightarrow Y \uparrow \beta$ (on average)
- Causality is *ceteris paribus*, “all else equal” $X \uparrow 1 \Rightarrow Y \uparrow \beta$ (on average)
- Interpretation of β depends on the research design and assumptions
- Observations may differ by Z, which affects $Y \Rightarrow$ not “all else equal”
- Do you have non-causal explanations for the TANF result?
- For example: single mothers who work less (regardless of benefits) may also be the ones receiving higher benefits $\Rightarrow$ correlation (**endogeneity**)

Potential "solutions" to identify causality with regressions

- Potential solution: control for relevant characteristics Z (marital status, num. of children, education, potential wage etc.) — "control variables"

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

- Problem: hard to control for *everything* that's relevant
- Imperfect solution: check robustness to many potential controls
- Better solution: understand why X may vary for reasons unrelated to ε and focus on exploiting this source of variation ("research design")
- This is the goal of the "causal inference" toolkit

Causal inference toolkit

What are some ways to do causal inference?

- **Randomized experiments** – the gold standard
- **Instrumental variables** – a variable that is correlated with the treatment but not the outcome (except through the treatment)
- **First differences** – comparing the same unit before and after a treatment
- **Difference-in-difference** – comparing the difference between treatment and control before and after a treatment
- **Regression discontinuity** – comparing units just above and below a threshold that are otherwise similar

Natural experiments

- Treatment and control groups created by nature (or, rather, policy)
- Examples: tax cut in New Jersey but not in New York; $\uparrow$ EITC benefits for single parents, but not married parents
- With repeated cross-section or panel data, you can observe changes before and after treatment in the treatment group:

$$\Delta\text{treated} = Y^{\text{treat}}_{\text{Post}} - Y^{\text{treat}}_{\text{Pre}} = \text{treatment} + \text{other things}$$

- and control group:

$$\Delta\text{controls} = Y^{\text{control}}_{\text{Post}} - Y^{\text{control}}_{\text{Pre}} = \text{other things}$$

- treatment $= \Delta$treated $- \Delta$controls
- This is called "difference in difference"
- We can never be 100% certain that all sources of bias are dealt with

Difference-in-difference — example

Using Quasi-Experimental Variation

Arkansas

	1996	1998	Difference
Benefit guarantee	\$5,000	\$4,000	–\$1,000
Hours of work per year	1,000	1,200	200

Louisiana

	1996	1998	Difference
Benefit guarantee	\$5,000	\$5,000	\$0
Hours of work per year	1,050	1,100	50

Gruber, *Public Finance and Public Policy*, Table 3.1

By how much did the EITC increase labor supply? Results suggest that \$1,000 (\$1,000-\$0) reduction in benefits caused an increase in hours of work by 150 ($150 = 200 - 50$)

Difference-in-difference EITC (Eissa et al. 2006)

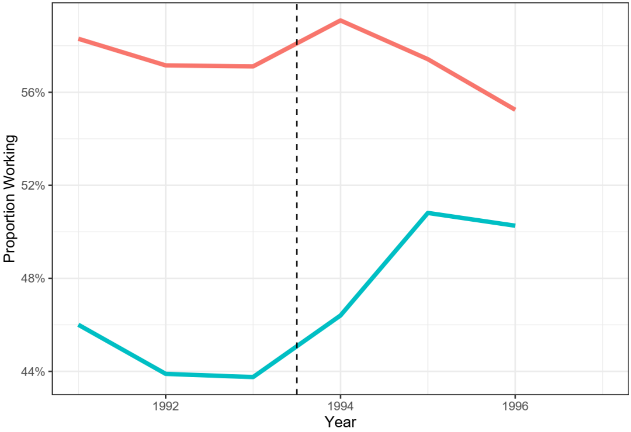
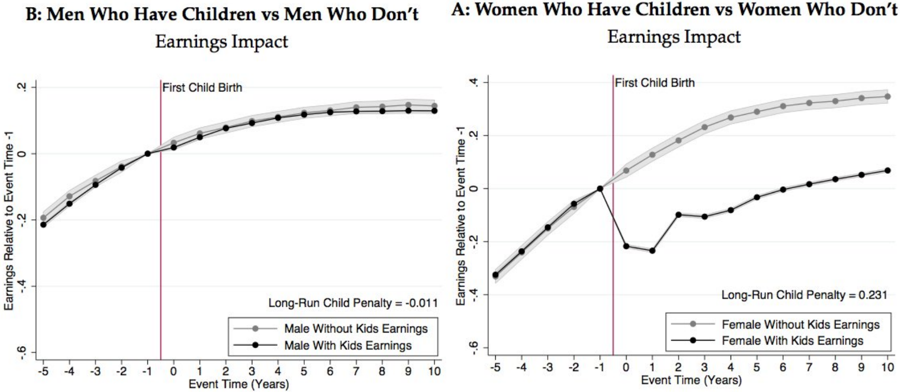
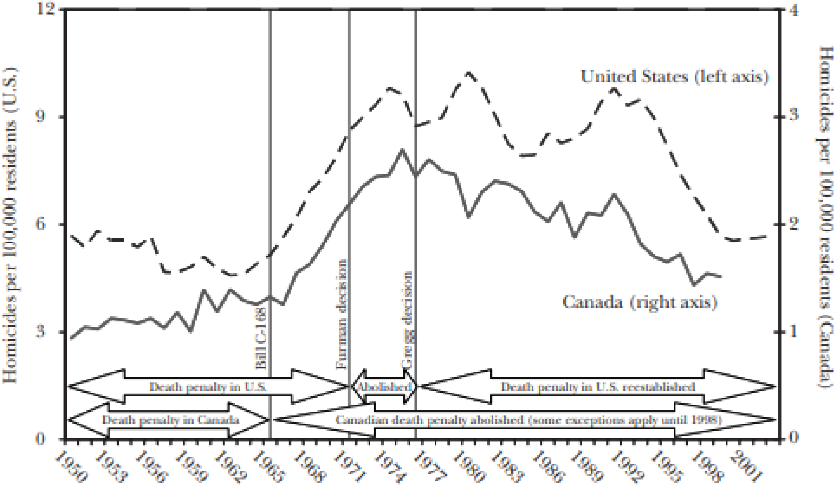


Figure 1: Difference-in-difference of the 1996 EITC increase on labor supply. The blue shows employment participation of single mothers, the red shows single women. Author's calculations using data compiled by Nick Huntington-Klein.



Source: Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard. 2019. "Children and Gender Inequality: Evidence from Denmark." American Economic Journal: Applied Economics, 11 (4): 181–209.
[The event: Having a child in Denmark for men and women.]

(U.S. and Canada rates on the left and right y-axes, respectively)



Source: Donohue and Wolfers (2005).

Source: Donohue and Wolfers (2005) via Angrist and Pischke (2010) shows the homicidal crime rate of US and Canada track similarly despite changes to death penalty – suggesting that the death penalty had little effect on crime.

Regression discontinuity

- Treatment and control separated by an arbitrary threshold:
  - Physical characteristics (weight, age, etc)
  - Policy thresholds (e.g. income, population, GPA etc.)
  - Political borders (e.g. county, state, etc.)

Within z units of a threshold z^* we see:

$$\Delta \text{treated} = \text{treatment} - \text{control} \quad \text{if} \quad |z| \leq z^*$$

Key assumptions:

- No manipulation at the threshold
- Nothing else changes at the threshold

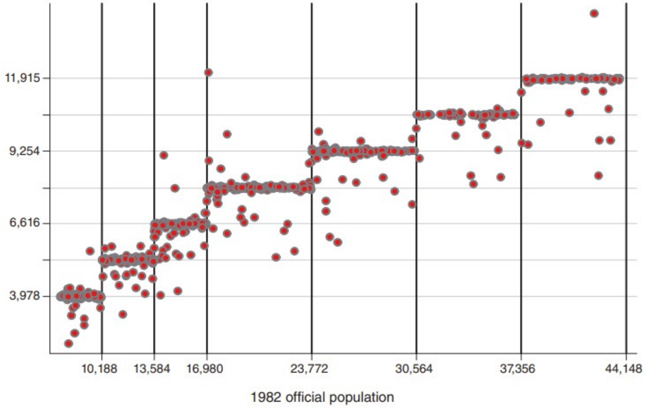


FIGURE 1. FPM TRANSFERS, 1982–1985 (*,000 2008 Reais*)

Source: Litschig, Stephan, and Kevin M. Morrison. 2013. "The Impact of Intergovernmental Transfers on Education Outcomes and Poverty Reduction." American Economic Journal: Applied Economics, 5 (4): 206–40.

Brazilian Municipality level data. X-axis is population binned by percentage points away from a threshold for receiving increased transfers due to a spending formula. Y-axis is amount of Fundo de Participação dos Municípios transfers received.

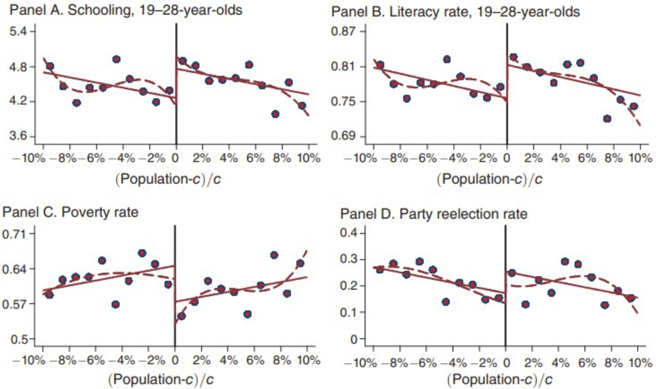


Figure 5. Impacts on Schooling, Literacy, Poverty, and Party Reelection

Source: Litschig, Stephan, and Kevin M. Morrison. 2013. "The Impact of Intergovernmental Transfers on Education Outcomes and Poverty Reduction." American Economic Journal: Applied Economics, 5 (4): 206–40.

Brazilian Municipality level data. X-axis is population binned by percentage points away from a threshold for receiving increased transfers due to a spending formula. Y-axis is the effect education, poverty, and political outcomes.

Structural Estimation

- We've covered "reduced form" methods.
- Structural estimation targets underlying utility or technology functions ("structural parameters").
- Imposes economic theory-based restrictions (e.g., negative substitution effect).
- Regression finds the best-fit line; structural estimation fits a model-based shape.
- Advantage: Explores more policy experiments.
  - Simulates untested policies.
  - Potentially more "externally" valid.
- Disadvantage: Imposes more assumptions on data.

Overview

- Correlation $\neq$ causation
- Multivariate regression with controls only goes so far
- Randomized experiments are the gold standard
- Causal inference toolkit uses natural experiments to identify causality
- Structural estimation uses economic theory to identify causal effects