Beyond the Gold Standard: Critiques, Alternatives & The Future

Final Lecture: Econometrics of Policy Evaluation (EPE)

Cristian Huse

Part I: The “Gold Standard” & Its Discontents

The Promise of the Randomistas

The Goal: Clean identification of Causal Effects.

The Method: Randomization eliminates selection bias

$$E(Y(0)|D = 1) = E(Y(0)|D = 0)$$

> *"RCTs shifted economics closer to medicine." — Banerjee & Duflo*

But... is "What Works" enough?

- **Internal Validity:** High (We know it worked *here*).
- **External Validity:** Unknown (Will it work *there*?).
- **Mechanism:** "Black Box" (We don't know *why*).

The Critique: Deaton & Cartwright (2018)

Main arguments: RCTs act as a "Black Box." They give us an unbiased estimate of an average effect, but they often fail to deliver the generalizability beyond the experimental sample – which is so needed for policy.

- RCTs require assumptions beyond "random assignment":
  - Sample representativeness
  - No interference between units
  - Treatment conditions mimic policy context
- Violations might lead to two problems

1. **The Scaling Problem:** A small NGO pilot $\neq$ a national government policy.

The Critique: Deaton & Cartwright (2018)

2. **The "Why" Problem:**

- Results tell what worked there and then, not why or if it works elsewhere
- Without a model of **behaviour**, we cannot predict how the policy interacts with local norms or institutions.

*"Randomization does not equalize everything... it only equalizes things **in expectation**."*

"RCTs help answer "what works" but can fail at answering "why it works" or "how it will work in a new environment."

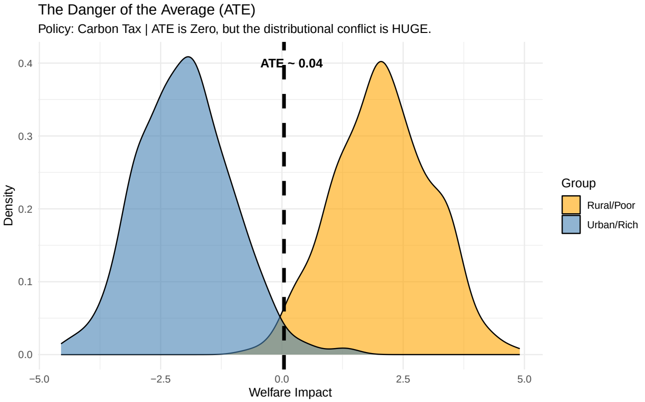
The Critique: Heckman & Smith (1995)

This is the foundational critique that every evaluator must know.

1. **Substitution Bias:** The control group isn't doing "nothing." They are finding the **next best alternative**. The RCT measures the **incremental** benefit, not the **total** benefit.
2. **General Equilibrium Effects:** If you train 100 people, they get jobs. If you train 1,000,000 people, you might just lower wages or displace other workers.
3. **Randomization Bias:** The sites that **agree** to be randomized are "special" (e.g., better managed, more innovative).

Visualizing the “Average” Trap

Why ATE (Average Treatment Effect) is dangerous for policy.



Part II: Theory & Structure

Heckman & Pinto (2022): Structural Causal Analysis

- Heckman & Pinto (2022) argue:
  - Policy questions often require models of behavior and mechanisms
  - Econometric approaches (structural models) embed theory into counterfactuals
  - This allows simulation of outcomes in new contexts – something RCTs alone cannot do
- Structural inference helps:
  - Analyze distributional effects
  - Consider equilibrium responses
  - Forecast impacts under new conditions

Unpacking the Mechanism: Heckman & Pinto (2022)

- How do we fix the "Black Box"? **Structural Causal Models.**
- Instead of just $D \rightarrow Y$, we model the **agent's choice**:

$$Y = f(X, \theta) \quad \text{(Outcome Function)}$$

$$D = 1[V(Z) > C] \quad \text{(Choice Function)}$$

- **Why do this?** If we know the parameters θ (e.g., price elasticity, risk aversion), we can simulate the policy in a **new** country or context without running a new RCT.

The Marginal Treatment Effect (MTE)

Reference: Heckman & Vytlacil (2005)

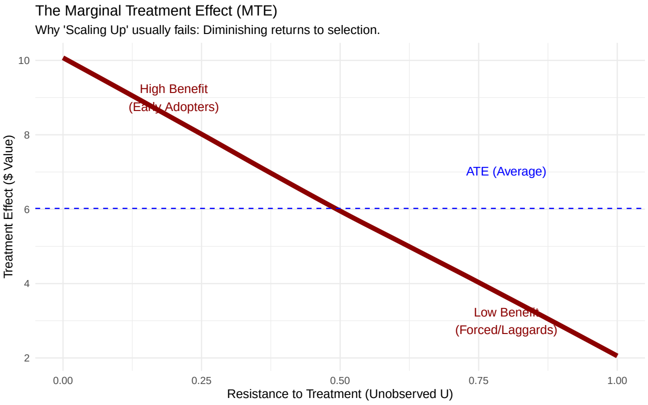
The effect of a policy depends on **who selects into it**.

- **MTE:** The treatment effect for people **indifferent** between participating and not participating.
- **The Insight:** People who gain the most often participate first. As you expand the program (scale up), you start pulling in people with lower gains (diminishing returns).

Interactive Concept: The MTE Curve

Imagine a **Home Insulation Subsidy**.

1. **High Gainers:** People with drafty houses (join immediately).
2. **Low Gainers:** People with new houses (only join if subsidy is huge).



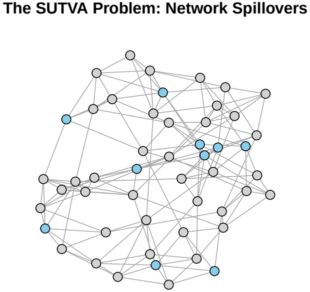
Part III: Labour Application

When “Control” is contaminated

- A key assumption of RCTs is **SUTVA** (Stable Unit Treatment Value Assumption): **My outcome depends only on my treatment.**
- Where this fails:
  - **Vaccines:** If I get vaccinated, you (control) are safer. (Positive Spillover).
  - **Job Training:** If I get trained, I take the job you (control) wanted. (Negative Displacement).
- **Example:** Crépon et al. (2013) on Labour Markets in France.
  - Found that job placement assistance helped **participants** but hurt **non-participants** by stealing their slots. Net effect was zero.

Simulating Spillovers in R

Imagine a village where treating one person affects their neighbors.



A control unit (Grey) connected to Blue units gets a 'Spillover Boost'

Part IV: Energy & Environmental Applications

Why Are RCTs Rare in Energy Policy?

- Energy policy often involves:
  - Large-scale regulatory changes
  - Infrastructure and macroeconomic feedback
  - Equilibrium effects
- **Example:**
  - Policy impacts are negotiated, contingent, and interact with market institutions – not neatly randomized

The External Validity Puzzle: Energy Efficiency

Case Study: Allcott (2015) – Site Selection Bias.

- **The RCT:** Opower Home Energy Reports (nudges to save energy).
- **The Result:** Initial RCTs showed huge savings.
- **The Reality:** The first utility companies to sign up were "Green Superstars."
- **The Drop:** As the program expanded to normal utilities, the effect size dropped by $\sim$50%.

Lesson: We need a model of *site selection* to predict the national impact.

Structural Models in Action: Transport

Case Study: Bento et al. (2009) – Gas Taxes.

- **Problem:** You cannot randomize a national gas tax.
- **Solution:** Estimate a structural model of vehicle choice (demand elasticities).
- **Findings:**
  - Flat gas tax = Regressive (hurts poor).
  - **Revenue Recycling:** If you return the revenue via lump-sum transfers, the policy becomes **Progressive**.
- **Only a structural model with general equilibrium could find this.**

Part V: The Future (Cheer up!)

Where do we go from here?

The "War" (Randomistas vs. Structuralists) is over. The "Synthesis" has begun.

1. **Theory-Driven Experiments:** Design RCTs specifically to test a parameter (e.g., "What is the price elasticity of solar adoption?"), not just a program.
2. **Transportability:** Using formal logic (Pearl, Bareinboim) to re-weight RCT results for new populations.
3. **Machine Learning:** Not just for prediction, but for **Heterogeneity**.

The Synthesis (RCT + Structural Models)

Reference: Duflo et al (2012) – Getting Teachers to Come to School.

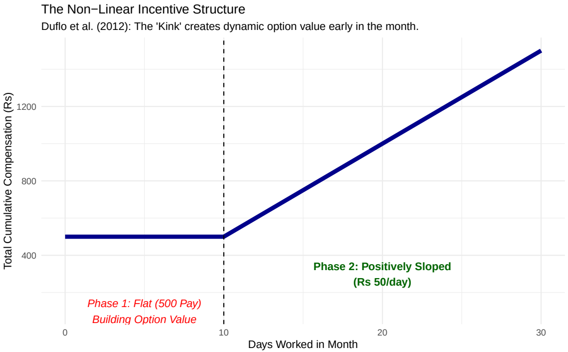
- This paper represents the "Best of Both Worlds."
  - **The Problem:** Teacher absenteeism in India (42% rate).
  - **The RCT:** Randomly assigned schools to have cameras + financial incentives.
    - Teachers were required to take pictures with class at the start and end of school day to document attendance.
    - **Pay Structure:** Rs 500 base. If you work $>$ 10 days, you get Rs 50 per extra day.
    - **The Result:** Absenteeism fell by 21 percentage points. Test scores rose by 0.17 SD.

The Synthesis

- **So why did they need a Structural Model?**
  - Because the pay scheme was **non-linear.**
  - Working on Day 1 pays Rs 0 directly... **but** it moves you closer to the “Bonus Zone” (Day 11+) $\rightarrow$ latent variable.
  - A “Reduced Form” regression cannot capture this **Dynamic Option Value.**

Visualizing Total Compensation

- **The Structural Insight (Option Value):** Teachers are forward-looking. They work on Days 1-10 not for immediate cash, but to "unlock" the upward-sloping part of the compensation curve later in the month. A simple regression misses this dynamic mechanism.



The Payoff

- **The Policy Payoff:**
  - Using the model, they simulated **counterfactual** payment schemes (e.g., Linear Pay vs. Convex Pay).
  - They found the RCT scheme was efficient, but could be tweaked to save costs while maintaining attendance.
  - **An RCT alone could never tell you that.**

The New Frontier: Causal Machine Learning

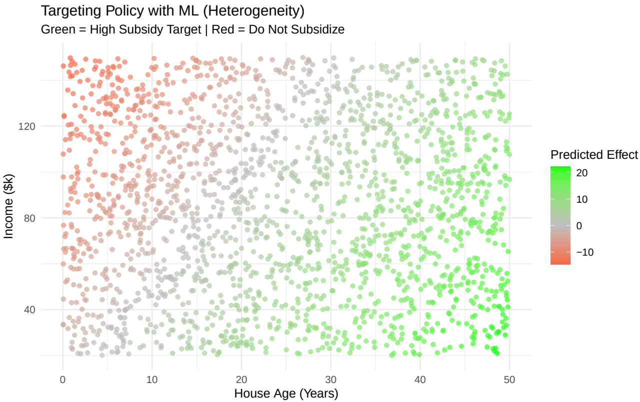
- We used to run one regression for the ATE.
- Now, we use **Causal Forests** to find $\tau(i)$ (The effect for person i).

Why? Precision.

- Instead of subsidizing everyone, subsidize only those with high **MTE**.
- Instead of a blanket training program, target those with the highest predicted gain.

A Simulation of the Future

Imagine we have 2,000 households. Who should get the solar subsidy?



Take-aways

- **Respect the Gold Standard**, but don't worship it. RCTs are a tool, not a religion.
- **Embrace Theory.** Data without a model is just noise. Use structural thinking to understand mechanisms.
- **Data + Models is the Way.** As in other areas of Economics.
- **Think Distributionally.** ATE is dead. Long live heterogeneity.
- **Be Optimistic.** With big data, Causal ML, and structural integration, we are entering a golden age of evidence-based policy.

References

Key Debates & Foundations

- **Heckman, J. J., & Smith, J. A. (1995).** Assessing the Case for Social Experiments. *Journal of Economic Perspectives*, 9(2), 85–110.
  - *The classic critique on substitution bias and randomization bias.*
- **Heckman, J. J., & Vytlacil, E. (2005).** Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, 73(3), 669-738.
  - *The formal derivation of the Marginal Treatment Effect (MTE).*
- **Duflo, E., Hanna, R., & Ryan, S. P. (2012).** Incentives Work: Getting Teachers to Come to School. *American Economic Review*, 102(4), 1241–1278.
  - *The seminal paper combining RCTs with a structural model.*
- **Deaton, A., & Cartwright, N. (2018).** Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2-21.
  - *The philosopher/economist critique on mechanisms and external validity.*

References

Key Debates & Foundations cont'd

- **Heckman, J. J., & Pinto, R. (2022).** The Econometric Model for Causal Policy Analysis. *Annual Review of Economics*, 14, 893-923.
  - *A modern synthesis of structural modeling and causality.*

References

Applied Examples

- **Allcott, H. (2015).** Site Selection Bias in Program Evaluation. *The Quarterly Journal of Economics*, 130(3), 1117–1165.
  - *Demonstrates why "scaling up" fails using Opower energy data.*
- **Bento, A. M., Goulder, L. H., Jacobsen, M. R., & von Haefen, R. H. (2009).** Distributional and Efficiency Impacts of Increased US Gasoline Taxes. *American Economic Review*, 99(3), 667-99.
  - *Uses structural models to design revenue recycling mechanisms.*
- **Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., & Zamora, P. (2013).** Do Labor Market Policies have Displacement Effects? *The Quarterly Journal of Economics*, 128(2), 531–580.
  - *The landmark RCT that explicitly measured spillovers.*

References

Applied Examples cont'd

- **Athey, S., & Wager, S. (2018).** Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523), 1228-1242.
  - *A key paper on Causal Forests and ML in econometrics.*

Appendix

Decision Trees: The Building Block

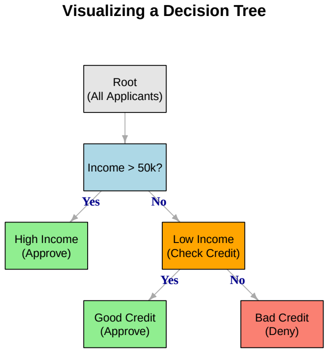
Definition: A **Decision Tree** is an algorithm that splits a population into smaller, more homogeneous groups based on input characteristics (covariates).

- **Root Node:** The starting point (the entire population).
- **Splitting Rule:** A question (e.g., "Is Income $>$ \$50k?") that divides data to best separate outcomes.
- **Leaf Node:** The final grouping where a prediction is made.

Decision Trees: Toy Example

Toy Example: Loan Approval Algorithm

- **Logic:** If Income is High, you are approved. If Income is Low, we check your Credit Score.



Causal Trees: Algorithm

The Logic: Standard trees split to minimize prediction error of Y. Causal trees split to **maximize the difference in treatment effects** (τ) between leaves.

Algorithm in Plain English:

1. **Honesty (Splitting):** Divide data into Set A (to build the tree) and Set B (to estimate effects).
2. **Search:** Find the variable (e.g., Age) that creates the biggest gap in treatment effects between the "Left" node and "Right" node.
3. **Estimate:** Calculate the effect in the final leaf using the held-out Set B.

Toy Simulation (R Code): Imagine a drug that *only* works for people under 40.

Causal Trees: Visualizing Heterogeneity