Cowles Commission Structural Models, Causal Effects and Treatment Effects: A Synthesis

James Heckman
University of Chicago
University College Dublin

Econometric Policy Evaluation, Lecture I
Koopmans Memorial Lectures
Cowles Foundation
Yale University
September 26, 2006

Staff List (University of Chicago)
Report of Research Activities, July 1, 1952-June 30, 1954

President	Alfred Cowles		
Executive Director	Rosson L. Cardwell		
Research Director	Tjalling C. Koopmans		
Research Associates	Stephen G. Allen Martin J. Beckmann Rosson L. Cardwell Gerard Debreu John Gurland Arnold C. Harberger	I.N. Herstein Clifford Hildreth H.S. Houthakker Tjalling C. Koopmans Jacob Marschak C. Bartlett McGuire	Richard F. Muth Roy Radner Leo Törnqvist Daniel Waterman Christopher Winstein
Research Assistants	Gary Becker Francis Bobkoski William L. Dunaway	Thomas A. Goldman Edwin Goldstein Mark Nerlove	Lester G. Telser Alan L. Tritter Jagna Zahl
Research Consultants	Stephen G. Allen Theodore W. Anderson Kenneth J. Arrow Herman Chernoff Carl F. Christ	Harold T. Davis Trygve Haavelmo Clifford Hildreth William C. Hood H.S. Houthakker	Leonid Hurwicz Lawrence R. Klein Theodore S. Motzkin Herman Rubin Herbert A. Simon
Guests	Pierre F.J. Baichere Karl Henrik Borch Jacques Dreze Atle Harald Elsas Masao Fukuoka	Jose Gil-Pelaez William Hamburger Herman F. Karreman William E. Krelle Giovanni Mancini	Rene F. Montjoie Sigbert J. Prais Bertram E. Rifas Ciro Tognetti

Monograph 10

STATISTICAL INFERENCE IN DYNAMIC ECONOMIC MODELS

by

COWLES COMMISSION RESEARCH STAFF MEMBERS AND GUESTS

Edited by
TJALLING C. KOOPMANS

With Introduction by
JACOB MARSCHAK

John Wiley & Sons, Inc., New York
Chapman & Hall, Limited, London
1950

Monograph 10 table of contents

Monograph 10 table of contents

Monograph 14

STUDIES IN
ECONOMETRIC METHOD

by

COWLES COMMISSION RESEARCH STAFF MEMBERS

Edited by

WILLIAM C. HOOD AND TJALLING C. KOOPMANS

John Wiley & Sons, Inc., New York
Chapman & Hall, Limited, London
1953

Monograph 14 table of contents

Cowles Commission motto:

For 20 years, the motto of the Cowles Commission, printed on its monographs and reports, was based on Lord Kelvin's dictum paraphrased as,

> "*Science is measurement*"

Cowles Commission motto:

By 1965 the importance of theory for interpreting evidence had become so apparent that the motto was changed to

> *"Theory and measurement"*

It is fitting

For many years at the University of Chicago, Cowles researchers worked in a building carved with the quotation by Lord Kelvin,

> "*When you cannot measure, your knowledge is meager and unsatisfactory.*"

It is fitting

For many years at the University of Chicago, Cowles researchers worked in a building carved with the quotation by Lord Kelvin,

> "*When you cannot measure, your knowledge is meager and unsatisfactory.*"

My lectures build on these works and these themes.

Introduction

- To focus ideas, analyze a prototypical policy evaluation problem.

Introduction

- To focus ideas, analyze a prototypical policy evaluation problem.
- Country can adopt a policy (e.g., democracy).

Introduction

- To focus ideas, analyze a prototypical policy evaluation problem.
- Country can adopt a policy (e.g., democracy).
- Choice Indicator:

Introduction

- To focus ideas, analyze a prototypical policy evaluation problem.
- Country can adopt a policy (e.g., democracy).
- Choice Indicator:
    - $D = 1$ if it adopts.

Introduction

- To focus ideas, analyze a prototypical policy evaluation problem.
- Country can adopt a policy (e.g., democracy).
- Choice Indicator:
  - $D = 1$ if it adopts.
  - $D = 0$ if not.

- Two outcomes $(Y_0(\omega), Y_1(\omega)),\ \omega \in \Omega$

- Two outcomes $(Y_0(\omega), Y_1(\omega)),\ \omega \in \Omega$
  - $Y_0(\omega)$ if country does not adopt

- Two outcomes $(Y_0(\omega), Y_1(\omega)),\ \omega \in \Omega$
    - $Y_0(\omega)$ if country does not adopt
    - $Y_1(\omega)$ if country adopts

- Two outcomes $(Y_0(\omega), Y_1(\omega)),\ \omega \in \Omega$
    - $Y_0(\omega)$ if country does not adopt
    - $Y_1(\omega)$ if country adopts

- Causal effect on observed outcomes

- Two outcomes $(Y_0(\omega), Y_1(\omega)),\ \omega \in \Omega$
  - $Y_0(\omega)$ if country does not adopt
  - $Y_1(\omega)$ if country adopts
- Causal effect on observed outcomes
- Marshallian *ceteris paribus* causal effect:

$$Y_1(\omega) - Y_0(\omega)$$

Figure 1: Extended Roy economy for policy adoption
Distribution of gains and treatment parameters

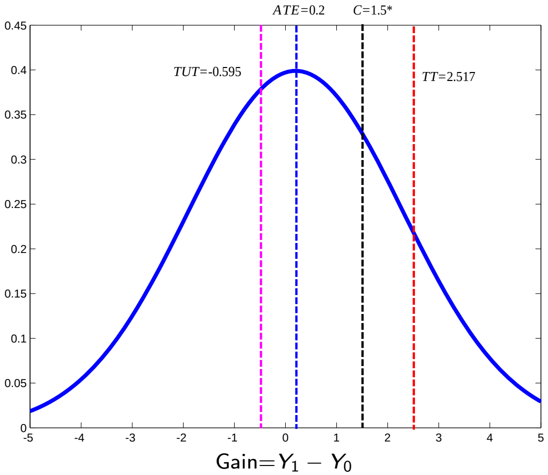


Figure 1 Legend

Suppose that a country has to choose whether to implement a policy. Under the policy, the GDP would be Y_1. Without the policy, the GDP of the country would be Y_0. For the sake of simplicity, suppose that

$$\begin{aligned} Y_1 &= \mu_1 + U_1 \\ Y_0 &= \mu_0 + U_0 \end{aligned}$$

where U_0 and U_1 are unobserved components of the aggregate output. The error terms (U_0, U_1) are dependent in a general way. Let δ denote the additional GDP due to the policy, i.e. $\delta = \mu_1 - \mu_0$. We assume $\delta > 0$. Let C denote the cost of implementing the policy. We assume that the cost is a fixed parameter C.

Figure 1 Legend

We relax this assumption below. The country's decision can be represented as:

$$D = \begin{cases} 1 & \text{if} \quad Y_1 - Y_0 - C > 0 \\ 0 & \text{if} \quad Y_1 - Y_0 - C \leq 0, \end{cases}$$

so the country decides to implement the policy $(D = 1)$ if the net gains coming from it are positive. Therefore, we can define the probability of adopting the policy in terms of the propensity score

$$\Pr(D = 1) = P(Y_1 - Y_0 - C > 0).$$

We assume that $(U_1, U_0) \sim N(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$, $\mu_0 = 0.67$, $\delta = 0.2$ and $C = 1.5$.

- More generally, define outcomes corresponding to state (policy, treatment) s for an "agent" characterized by ω as $Y(s,\omega)$, $\omega \in \Omega = [0,1]$, $s \in \mathcal{S}$, set of possible treatments.

- More generally, define outcomes corresponding to state (policy, treatment) s for an "agent" characterized by ω as $Y(s, \omega)$, $\omega \in \Omega = [0, 1]$, $s \in \mathcal{S}$, set of possible treatments.

- The agent can be any economic agent such as a household, a firm, or a country.

- More generally, define outcomes corresponding to state (policy, treatment) s for an "agent" characterized by ω as $Y(s, \omega)$, $\omega \in \Omega = [0, 1]$, $s \in \mathcal{S}$, set of possible treatments.
- The agent can be any economic agent such as a household, a firm, or a country.
- The $Y(s, \omega)$ are *ex post* outcomes realized after treatments are chosen.

- More generally, define outcomes corresponding to state (policy, treatment) s for an "agent" characterized by ω as $Y(s, \omega)$, $\omega \in \Omega = [0, 1]$, $s \in \mathcal{S}$, set of possible treatments.
- The agent can be any economic agent such as a household, a firm, or a country.
- The $Y(s, \omega)$ are *ex post* outcomes realized after treatments are chosen.
- Consider uncertainty and related *ex ante* and *ex post* evaluations in the Friday lecture.

- The **individual treatment effect** for agent ω.

$$Y(s,\omega) - Y(s',\omega), \qquad s \neq s', \quad s, s' \in S, \tag{1.1}$$

Individual level causal effect.

- The **individual treatment effect** for agent ω.

$$Y(s,\omega) - Y(s',\omega), \qquad s \neq s', \quad s, s' \in S, \qquad (1.1)$$

Individual level causal effect.

- Comparisons can also be made in terms of utilities $R(Y(s,\omega))$.

- The **individual treatment effect** for agent ω.

$$Y(s,\omega) - Y(s',\omega), \qquad s \neq s', \quad s,s' \in S, \qquad (1.1)$$

Individual level causal effect.

- Comparisons can also be made in terms of utilities $R(Y(s,\omega))$.
- $R(Y(s,\omega),\omega) > R(Y(s',\omega),\omega)$ if s is preferred to s'.

- The **individual treatment effect** for agent ω.

$$Y(s,\omega) - Y(s',\omega), \qquad s \neq s', \quad s, s' \in S, \qquad (1.1)$$

Individual level causal effect.

- Comparisons can also be made in terms of utilities $R(Y(s,\omega))$.
- $R(Y(s,\omega),\omega) > R(Y(s',\omega),\omega)$ if s is preferred to s'.
- The difference in subjective outcomes is $[R(Y(s,\omega),\omega) - R(Y(s',\omega),\omega)]$, and is another possible definition of a treatment effect. Holding ω fixed holds all features of the person fixed except the treatment assigned, s.

- The question,

  "What question is the analysis supposed to answer?"

  is the big unanswered question in the recent policy evaluation literature.

- The question,

  "What question is the analysis supposed to answer?"

  is the big unanswered question in the recent policy evaluation literature.

- The question is usually unanswered because it is unasked in much of the modern treatment effect literature which seeks to estimate "an effect" without telling you which effect or why it is interesting to know it.

- The question,

  "What question is the analysis supposed to answer?"

  is the big unanswered question in the recent policy evaluation literature.

- The question is usually unanswered because it is unasked in much of the modern treatment effect literature which seeks to estimate "an effect" without telling you which effect or why it is interesting to know it.

- The answer to the question shapes the way we go about policy evaluation analysis.

- The question,

  "What question is the analysis supposed to answer?"

  is the big unanswered question in the recent policy evaluation literature.

- The question is usually unanswered because it is unasked in much of the modern treatment effect literature which seeks to estimate "an effect" without telling you which effect or why it is interesting to know it.

- The answer to the question shapes the way we go about policy evaluation analysis.

- A central point in the Cowles research program (Marschak, 1949, 1953).

P-1

Evaluating the Impact of Interventions on Outcomes Including Their Impact in Terms of Welfare

- "Internal validity": Campbell and Stanley, 1963: looking at a program in place.
- Consider both objective or public outcomes Y and "subjective" outcomes R.
- Objective outcomes are intrinsically *ex post* in nature. Subjective outcomes can be *ex ante* or *ex post*.
- *Ex ante* expected pain and suffering may be different from *ex post* pain and suffering. Agents may also have *ex ante* evaluations of the objective outcomes that may differ from their *ex post* evaluations.

P-2

Forecasting the Impacts (Constructing Counterfactual States) of Interventions Implemented in one Environment in Other Environments, Including Their Impacts In Terms of Welfare.

"External validity": This is the problem of projecting evaluations in one environment to another environment.

P-3

Forecasting the Impacts of Interventions (Constructing Counterfactual States Associated with Interventions) Never Historically Experienced to Various Environments, Including Their Impacts in Terms of Welfare.

- The problem of forecasting the effect of a new policy never tried in any environment.
- All three problems entail identification of counterfactuals.
- But they place different demands on models and the data.

- In answering these questions it is important to separate three tasks.

- In answering these questions it is important to separate three tasks.
- In applied work and in statistical analyses of “causality” these tasks are often confused.

Table 1: Three distinct tasks arising in the analysis of causal models

Task	Description	Requirements
1	Defining the Set of Hypotheticals or Counterfactuals	A Scientific Theory
2	Identifying Parameters (Causal or Otherwise) from Hypothetical Population Data	Mathematical Analysis of Point or Set Identification
3	Identifying Parameters from Data	Estimation and Testing Theory

- When is $Y(s, \omega)$ an adequate description of the outcome of a policy?

- When is $Y(s,\omega)$ an adequate description of the outcome of a policy?
- Standard approach in the treatment effect literature assumes that there is a mechanism $\tau \in \mathcal{T}$ allocating "agents" $\omega \in \Omega$ to treatment $s \in \mathcal{S}$.

- When is $Y(s, \omega)$ an adequate description of the outcome of a policy?
- Standard approach in the treatment effect literature assumes that there is a mechanism $\tau \in \mathcal{T}$ allocating "agents" $\omega \in \Omega$ to treatment $s \in \mathcal{S}$.
- Invariance says $Y(s, \omega, \tau) = Y(s, \omega) \; \forall \; \tau \in \mathcal{T}$.

- When is $Y(s,\omega)$ an adequate description of the outcome of a policy?
- Standard approach in the treatment effect literature assumes that there is a mechanism $\tau \in \mathcal{T}$ allocating "agents" $\omega \in \Omega$ to treatment $s \in \mathcal{S}$.
- Invariance says $Y(s,\omega,\tau) = Y(s,\omega)\ \forall\ \tau \in \mathcal{T}$.
- A policy is equated with an assignment mechanism s.

- When is $Y(s,\omega)$ an adequate description of the outcome of a policy?
- Standard approach in the treatment effect literature assumes that there is a mechanism $\tau \in \mathcal{T}$ allocating "agents" $\omega \in \Omega$ to treatment $s \in \mathcal{S}$.
- Invariance says $Y(s,\omega,\tau) = Y(s,\omega)\ \forall\ \tau \in \mathcal{T}$.
- A policy is equated with an assignment mechanism s.
- In econometric policy evaluation recognizing agent choices, we need a more general approach.

- Policies can only affect agent incentives. We cannot usually force people to choose treatments.

- Policies can only affect agent incentives. We cannot usually force people to choose treatments.
- Recognizing this is a distinctive feature of the economic approach.

- Policies can only affect agent incentives. We cannot usually force people to choose treatments.
- Recognizing this is a distinctive feature of the economic approach.
- A constraint assignment rule $a \in \mathcal{A}$.

- Policies can only affect agent incentives. We cannot usually force people to choose treatments.
- Recognizing this is a distinctive feature of the economic approach.
- A constraint assignment rule $a \in \mathcal{A}$.
- It maps $\omega \in \Omega$ into $\mathcal{B}$, a space of constraints or incentives (e.g., taxes, endowments, eligibility).

- Policies can only affect agent incentives. We cannot usually force people to choose treatments.
- Recognizing this is a distinctive feature of the economic approach.
- A constraint assignment rule $a \in \mathcal{A}$.
- It maps $\omega \in \Omega$ into $\mathcal{B}$, a space of constraints or incentives (e.g., taxes, endowments, eligibility).
- $a : \Omega \rightarrow \mathcal{B}$.

- For a given $b \in \mathcal{B}$, agents choose a particular treatment.

- For a given $b \in \mathcal{B}$, agents choose a particular treatment.
- $\tau : \Omega \times \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{S}$, $\tau \in \mathcal{T}$.

Notation and definitions of individual level treatment effects

- For a given $b \in \mathcal{B}$, agents choose a particular treatment.
- $\tau : \Omega \times \mathcal{A} \times \mathcal{B} \rightarrow \mathcal{S}$, $\tau \in \mathcal{T}$.
- A policy is a pair $p = (a, \tau)$.

- In the general case, outcomes depend on ω, s, a, b, τ

$$Y(\omega, s, a, b, \tau)$$

- In the general case, outcomes depend on ω, s, a, b, τ

$$Y(\omega, s, a, b, \tau)$$

- When can we write:

$$Y(\omega, s, a, b, \tau) = Y(s, \omega)?$$

Notation and definitions of individual level treatment effects

- In the general case, outcomes depend on ω, s, a, b, τ

$$Y(\omega, s, a, b, \tau)$$

- When can we write:

$$Y(\omega, s, a, b, \tau) = Y(s, \omega)?$$

- When can we ignore the mechanism $a \in \mathcal{A}$ and the treatment assignment rule $\tau \in \mathcal{T}$ in studying outcomes?

Notation and definitions of individual level treatment effects

- In the general case, outcomes depend on ω, s, a, b, τ

$$Y(\omega, s, a, b, \tau)$$

- When can we write:

$$Y(\omega, s, a, b, \tau) = Y(s, \omega)?$$

- When can we ignore the mechanism $a \in \mathcal{A}$ and the treatment assignment rule $\tau \in \mathcal{T}$ in studying outcomes?
- Need invariance postulates

- Policy invariance for objective outcomes:

PI-1

For any two constraint assignment mechanisms $a, a' \in \mathcal{A}$ *and incentives* $b, b' \in \mathcal{B}$, *with* $a(\omega) = b$ *and* $a'(\omega) = b'$, *and for all* $\omega \in \Omega$, $Y(s, \omega, a, b, \tau) = Y(s, \omega, a', b', \tau)$, *for all* $s \in \mathcal{S}_{\tau(a,b)}(\omega) \cap \mathcal{S}_{\tau(a',b')}(\omega)$ *for assignment rule* τ *where* $\mathcal{S}_{\tau(a,b)}(\omega)$ *is the image set for* $\tau(a, b)$. *For simplicity we assume* $\mathcal{S}_{\tau(a,b)}(\omega) = \mathcal{S}_{\tau(a,b)}$ *for all* $\omega \in \Omega$.

- Rules out effects of the constraint assignment mechanism and incentive schedules on realized outcomes.

PI-2

For each constraint assignment $a \in \mathcal{A}$ and $b \in \mathcal{B}$ and all $\omega \in \Omega$, $Y(s, \omega, a, b, \tau) = Y(s, \omega, a, b, \tau')$ for all τ and $\tau' \in \mathcal{T}$ with $s \in \mathcal{S}_{\tau'(a,b)} \cap \mathcal{S}_{\tau(a,b)}$, where $\mathcal{S}_{\tau(a,b)}$ is the image set of τ with assignment mechanism a and incentive b.

- For simplicity, we assume $\mathcal{S}_{\tau(a,b)}(\omega) = \mathcal{S}_{\tau(a,b)}, \; \forall\, \omega \in \Omega$.
- Rules out GE, peer effects, and social interactions.
- (PI-1) and (PI-2) say that it doesn't matter how the agent gets the incentives or what they are (PI-1), or who else gets the treatment or how it is chosen (PI-2).

- Given (PI-1) and (PI-2) we can write the outcome as

$$Y(s, \omega).$$

- Given (PI-1) and (PI-2) we can write the outcome as

$$Y(s, \omega).$$

- Develop a parallel set of invariance assumptions for utilities R.

- Given (PI-1) and (PI-2) we can write the outcome as

$$Y(s,\omega).$$

- Develop a parallel set of invariance assumptions for utilities R.
- First define

$$\mathcal{A}_b(\omega) = \{a \mid a \subseteq \mathcal{A}, a(\omega) = b\}, \; \omega \in \Omega.$$

PI-3

For any two constraint assignment mechanisms $a, a' \in \mathcal{A}$ *and incentives* $b, b' \in \mathcal{B}$ *with* $a(\omega) = b$ *and* $a'(\omega) = b'$*, and for all* $\omega \in \Omega$, $Y(s, \omega, a, b, \tau) = Y(s, \omega, a', b', \tau)$ *for all* $s \in \mathcal{S}_{\tau(a,b)}(\omega) \cap \mathcal{S}_{\tau(a',b')}(\omega)$ *for assignment rule* τ*, where* $\mathcal{S}_{\tau(a,b)}(\omega)$ *is the image set of* $\tau(a, b)$ *and for simplicity we assume that* $\mathcal{S}_{\tau(a,b)}(\omega) = \mathcal{S}_{\tau(a,b)}$ *for all* $\omega \in \Omega$*. In addition, for any mechanisms* $a, a' \in \mathcal{A}_b(\omega)$*, producing the same* $b \in \mathcal{B}$ *under the same conditions, and for all* ω*,*
$R(s, \omega, a, b, \tau) = R(s, \omega, a', b, \tau)$.

PI-4

For each pair (a, b) and all $\omega \in \Omega$,

$$\begin{aligned} Y(s, \omega, a, b, \tau) &= Y(s, \omega, a, b, \tau') \\ R(s, \omega, a, b, \tau) &= R(s, \omega, a, b, \tau') \end{aligned}$$

for all $\tau, \tau' \in \mathcal{T}$ and $s \in \mathcal{S}_{\tau(a,b)} \cap \mathcal{S}_{\tau'(a,b)}$.

How To Construct Counterfactuals?

- Central problem in the evaluation literature is the absence of information on outcomes for person ω other than the outcome that is observed.

How To Construct Counterfactuals?

- Central problem in the evaluation literature is the absence of information on outcomes for person ω other than the outcome that is observed.
- Even a perfectly implemented social experiment does not solve this problem.

How To Construct Counterfactuals?

- Central problem in the evaluation literature is the absence of information on outcomes for person ω other than the outcome that is observed.
- Even a perfectly implemented social experiment does not solve this problem.
- Randomization with full compliance identifies only one component of $\{Y(s,\omega)\}_{s\in\mathcal{S}}$ for any person.

How To Construct Counterfactuals?

- Central problem in the evaluation literature is the absence of information on outcomes for person ω other than the outcome that is observed.
- Even a perfectly implemented social experiment does not solve this problem.
- Randomization with full compliance identifies only one component of $\{Y(s,\omega)\}_{s\in\mathcal{S}}$ for any person.
- In addition, some of the $s \in \mathcal{S}$ may never be observed.

The evaluation problem

- For each policy regime, at any point in time we observe person ω in some state but not in any of the other states.

- For each policy regime, at any point in time we observe person ω in some state but not in any of the other states.

- Do not observe $Y(s', \omega)$ for person ω if we observe $Y(s, \omega)$, $s \neq s'$.

- For each policy regime, at any point in time we observe person ω in some state but not in any of the other states.

- Do not observe $Y(s', \omega)$ for person ω if we observe $Y(s, \omega)$, $s \neq s'$.

- Let $D(s, \omega) = 1$ if we observe person ω in state s under policy regime p.

- For each policy regime, at any point in time we observe person ω in some state but not in any of the other states.

- Do not observe $Y(s', \omega)$ for person ω if we observe $Y(s, \omega)$, $s \neq s'$.

- Let $D(s, \omega) = 1$ if we observe person ω in state s under policy regime p.

- Observed objective outcome

$$Y(\omega) = \sum_{s \in \mathcal{S}} D(s, \omega) Y(s, \omega) . \tag{2.1}$$

- The **evaluation problem** in this model is that we only observe each individual in one of $\bar{S}$ possible states.

- The **evaluation problem** in this model is that we only observe each individual in one of $\bar{S}$ possible states.
- We do not know the outcome of the individual in other states and hence cannot directly form individual level treatment effects.

- The **evaluation problem** in this model is that we only observe each individual in one of $\bar{S}$ possible states.
- We do not know the outcome of the individual in other states and hence cannot directly form individual level treatment effects.
- The **selection problem** arises because we only observe certain persons in any state.

- The **evaluation problem** in this model is that we only observe each individual in one of $\bar{S}$ possible states.
- We do not know the outcome of the individual in other states and hence cannot directly form individual level treatment effects.
- The **selection problem** arises because we only observe certain persons in any state.
- We observe $Y(s, \omega)$ only for persons for whom $D(s, \omega) = 1$.

- The **evaluation problem** in this model is that we only observe each individual in one of $\bar{S}$ possible states.
- We do not know the outcome of the individual in other states and hence cannot directly form individual level treatment effects.
- The **selection problem** arises because we only observe certain persons in any state.
- We observe $Y(s, \omega)$ only for persons for whom $D(s, \omega) = 1$.
- In general, the outcomes of persons found in $S = s$ are not representative of what the outcomes of people would be if they were randomly assigned to s.

- The Roy model (1951): Two possible treatment outcomes ($S = \{0, 1\}$) and a scalar outcome measure and a particular assignment mechanism
  $D(1, \omega) = \mathbf{1}[Y(1, \omega) > Y(0, \omega)]$
  (reveals $R(1, \omega) - R(0, \omega) \geq 0$).

- The Roy model (1951): Two possible treatment outcomes ($S = \{0, 1\}$) and a scalar outcome measure and a particular assignment mechanism
$D(1, \omega) = \mathbf{1}[Y(1, \omega) > Y(0, \omega)]$
(reveals $R(1, \omega) - R(0, \omega) \geq 0$).
- The economist's use of choice data distinguishes the econometric approach from the statistical approach.

How To Construct Counterfactuals?

- Two main avenues of escape from this problem.

How To Construct Counterfactuals?

- Two main avenues of escape from this problem.
- The first avenue, featured in explicitly formulated econometric models and often called "structural econometric analysis ", derives from the Cowles tradition.

How To Construct Counterfactuals?

- Two main avenues of escape from this problem.
- The first avenue, featured in explicitly formulated econometric models and often called "structural econometric analysis", derives from the Cowles tradition.
- Models $Y(s,\omega)$ explicitly in terms of its determinants as specified by theory.

How To Construct Counterfactuals?

- Two main avenues of escape from this problem.
- The first avenue, featured in explicitly formulated econometric models and often called "structural econometric analysis", derives from the Cowles tradition.
- Models $Y(s, \omega)$ explicitly in terms of its determinants as specified by theory.
- This entails describing the random variables characterizing ω and carefully distinguishing what agents know and what the analyst knows.

How To Construct Counterfactuals?

- This approach also models $D(s, \omega)$ and the dependence between $Y(s, \omega)$ and $D(s, \omega)$ produced from variables common to $Y(s, \omega)$ and $D(s, \omega)$.

How To Construct Counterfactuals?

- This approach also models $D(s, \omega)$ and the dependence between $Y(s, \omega)$ and $D(s, \omega)$ produced from variables common to $Y(s, \omega)$ and $D(s, \omega)$.
- Specifies a full model and attempts to address problems (P-1)–(P-3).

How To Construct Counterfactuals?

- A second avenue, pursued in the recent treatment effect literature, redirects attention away from estimating the determinants of $Y(s, \omega)$ toward estimating some population version of individual "causal effects," without modeling what factors give rise to the outcome or the relationship between the outcomes and the mechanism selecting outcomes.

How To Construct Counterfactuals?

- A second avenue, pursued in the recent treatment effect literature, redirects attention away from estimating the determinants of $Y(s,\omega)$ toward estimating some population version of individual "causal effects," without modeling what factors give rise to the outcome or the relationship between the outcomes and the mechanism selecting outcomes.
- Agent valuations of outcomes are typically ignored.

How To Construct Counterfactuals?

- A second avenue, pursued in the recent treatment effect literature, redirects attention away from estimating the determinants of $Y(s, \omega)$ toward estimating some population version of individual "causal effects," without modeling what factors give rise to the outcome or the relationship between the outcomes and the mechanism selecting outcomes.
- Agent valuations of outcomes are typically ignored.
- The treatment effect literature focuses largely on policy problem (P-1) for the subset of outcomes that is observed.

How To Construct Counterfactuals?

- A second avenue, pursued in the recent treatment effect literature, redirects attention away from estimating the determinants of $Y(s,\omega)$ toward estimating some population version of individual "causal effects," without modeling what factors give rise to the outcome or the relationship between the outcomes and the mechanism selecting outcomes.
- Agent valuations of outcomes are typically ignored.
- The treatment effect literature focuses largely on policy problem (P-1) for the subset of outcomes that is observed.
- Seeks to answer a narrower problem.

- For program (state, treatment) j compared to program (state, treatment) k,

$$\text{ATE}(j,k) = E\left(Y(j,\omega) - Y(k,\omega)\right).$$

$$\text{TT}(j,k) = E\left(Y(j,\omega) - Y(k,\omega) \mid D(j,\omega) = 1\right). \quad (2.2)$$

- For program (state, treatment) j compared to program (state, treatment) k,

$$\mathrm{ATE}(j,k) = E\left(Y(j,\omega) - Y(k,\omega)\right).$$

$$\mathrm{TT}(j,k) = E\left(Y(j,\omega) - Y(k,\omega) \mid D(j,\omega) = 1\right). \quad (2.2)$$

- These are the traditional parameters for average returns.

- For program (state, treatment) j compared to program (state, treatment) k,

$$\mathrm{ATE}(j,k) = E\left(Y(j,\omega) - Y(k,\omega)\right).$$

$$\mathrm{TT}(j,k) = E\left(Y(j,\omega) - Y(k,\omega) \mid D(j,\omega) = 1\right). \quad (2.2)$$

- These are the traditional parameters for average returns.
- But for economic analysis, marginal returns are more important.

- The distinction between the marginal and average return is a central concept in economics.

- The distinction between the marginal and average return is a central concept in economics.
- The **Effect Of Treatment for People at the Margin of Indifference** (EOTM) between j and k, given that these are the best two choices available is, with respect to personal preferences, and with respect to choice-specific costs $C(j, \omega)$.

$$\mathrm{EOTM}^{R}(j,k) = \tag{2.3}$$

$$E\left(\begin{array}{c} Y(j,\omega) \\ -Y(k,\omega) \end{array} \middle| \begin{array}{l} R\left(Y\left(j,\omega\right), C\left(j,\omega\right), \omega\right) = R\left(Y\left(k,\omega\right), C\left(k,\omega\right), \omega\right); \\ \left.\begin{array}{l} R\left(Y\left(j,\omega\right), C\left(j,\omega\right), \omega\right) \\ R\left(Y\left(k,\omega\right), C\left(k,\omega\right), \omega\right) \end{array}\right\} \geq R\left(Y\left(\ell,\omega\right), C\left(\ell,\omega\right), \omega\right) \end{array} \right),$$

$$\ell \neq j, k.$$

- A generalization of this parameter called the **Marginal Treatment Effect**, introduced into the evaluation literature by Björklund and Moffitt (1987).

- A generalization of this parameter called the **Marginal Treatment Effect**, introduced into the evaluation literature by Björklund and Moffitt (1987).
- Return to people at the margin of choice.

- A generalization of this parameter called the **Marginal Treatment Effect**, introduced into the evaluation literature by Björklund and Moffitt (1987).
- Return to people at the margin of choice.
- Will discuss methods for identifying this return tomorrow.

Policy relevant treatment effect

- Effect on aggregate outcomes of one policy regime $p \in \mathcal{P}$ compared to the effect of another policy regime $p' \in \mathcal{P}$:

$$\textbf{PRTE}: \quad E(Y(s_p(\omega), \omega) - Y(s_{p'}(\omega), \omega)),$$
$$\text{where } p, p' \in \mathcal{P}.$$

$s_p(\omega)$ is treatment allocated under policy p.

Policy relevant treatment effect

- Effect on aggregate outcomes of one policy regime $p \in \mathcal{P}$ compared to the effect of another policy regime $p' \in \mathcal{P}$:

$$\textbf{PRTE}: \quad E(Y(s_p(\omega), \omega) - Y(s_{p'}(\omega), \omega)), \\ \text{where } p, p' \in \mathcal{P}.$$

$s_p(\omega)$ is treatment allocated under policy p.

- Corresponding to this objective outcome is the subjective counterpart:

$$\textbf{Subjective PRTE}: \quad E(R(s_p(\omega), \omega)) - E(R(s_{p'}(\omega), \omega)), \\ \text{where } p, p' \in \mathcal{P}.$$

- Modern political economy seeks to know the proportion of people who benefit from policy regime p compared with p'. Voting Criterion:

$$\Pr\left(Y\left(s_p(\omega),\omega\right) > Y\left(s_{p'}(\omega),\omega\right)\right).$$

- Modern political economy seeks to know the proportion of people who benefit from policy regime p compared with p'. Voting Criterion:

$$\Pr\left(Y\left(s_p(\omega),\omega\right) > Y\left(s_{p'}(\omega),\omega\right)\right).$$

- For particular treatments within a policy regime p, it is also of interest to determine the proportion who benefit from j compared to k as

$$\Pr\left(Y\left(j,\omega\right) > Y\left(k,\omega\right)\right).$$

- Modern political economy seeks to know the proportion of people who benefit from policy regime p compared with p'. Voting Criterion:

$$\Pr\left(Y\left(s_p(\omega),\omega\right) > Y\left(s_{p'}(\omega),\omega\right)\right).$$

- For particular treatments within a policy regime p, it is also of interest to determine the proportion who benefit from j compared to k as

$$\Pr\left(Y\left(j,\omega\right) > Y\left(k,\omega\right)\right).$$

- Option values also interesting: option of having access to a program.

- Modern political economy seeks to know the proportion of people who benefit from policy regime p compared with p'. Voting Criterion:

$$\Pr\left(Y\left(s_p(\omega),\omega\right) > Y\left(s_{p'}(\omega),\omega\right)\right).$$

- For particular treatments within a policy regime p, it is also of interest to determine the proportion who benefit from j compared to k as

$$\Pr\left(Y\left(j,\omega\right) > Y\left(k,\omega\right)\right).$$

- Option values also interesting: option of having access to a program.
- Uncertainty and regret (covered Friday).

A generalized Roy model under perfect certainty

- Given an economic model, we can trivially derive the treatment effects.

- Given an economic model, we can trivially derive the treatment effects.
- $\bar{S}$ states associated with different levels of schooling, or some other outcome such as residence in a region, or choice of technology.

A generalized Roy model under perfect certainty

- Given an economic model, we can trivially derive the treatment effects.
- $\bar{S}$ states associated with different levels of schooling, or some other outcome such as residence in a region, or choice of technology.
- Associated with each choice s is a valuation of the outcome of the choice $R(s)$ where R is the valuation function and s is the state. (We drop the ω argument here to simplify notation.)

- Given an economic model, we can trivially derive the treatment effects.
- $\bar{S}$ states associated with different levels of schooling, or some other outcome such as residence in a region, or choice of technology.
- Associated with each choice s is a valuation of the outcome of the choice $R(s)$ where R is the valuation function and s is the state. (We drop the ω argument here to simplify notation.)
- Z: observed individual variables that affect choices.

- Given an economic model, we can trivially derive the treatment effects.
- $\bar{S}$ states associated with different levels of schooling, or some other outcome such as residence in a region, or choice of technology.
- Associated with each choice s is a valuation of the outcome of the choice $R(s)$ where R is the valuation function and s is the state. (We drop the ω argument here to simplify notation.)
- Z: observed individual variables that affect choices.
- Each state may be characterized by a bundle of attributes, characteristics or qualities $Q(s)$ that fully characterize the state. If $Q(s)$ fully describes the state, $R(s) = R(Q(s))$.

$$R(s) = \mu_R(s, Z) + \upsilon(s, Z, \nu)$$

- Associated with each choice is outcome $Y(s)$ which may be vector valued.

- Associated with each choice is outcome $Y(s)$ which may be vector valued.
- The set of possible treatments $\mathcal{S}$ is $\{1, \ldots, \bar{S}\}$, the set of state labels.

- Associated with each choice is outcome $Y(s)$ which may be vector valued.
- The set of possible treatments $\mathcal{S}$ is $\{1,\ldots,\bar{S}\}$, the set of state labels.
- The assignment mechanism is specified by utility maximization:

$$D(j) = 1 \text{ if } \operatorname{argmax}_{s\in\mathcal{S}}\{R(s)\} = j, \tag{2.4}$$

where in the event of ties, choices are made by a flip of a coin.

- Associated with each choice is outcome $Y(s)$ which may be vector valued.
- The set of possible treatments $\mathcal{S}$ is $\{1, \ldots, \bar{S}\}$, the set of state labels.
- The assignment mechanism is specified by utility maximization:

$$D(j) = 1 \text{ if } \operatorname{argmax}_{s \in \mathcal{S}} \{R(s)\} = j, \tag{2.4}$$

where in the event of ties, choices are made by a flip of a coin.

- People *self-select* into treatment.

- The Roy model (1951) and its extensions (Gronau, 1974; Heckman, 1974; Willis and Rosen, 1979; Heckman, 1990; Carneiro, Hansen, and Heckman, 2003) are at the core of microeconometrics.

$$Y_1 = X\beta_1 + U_1 \tag{2.5a}$$
$$Y_0 = X\beta_0 + U_0, \tag{2.5b}$$

and associated costs (prices) as a function of W

$$C = W\beta_C + U_C. \tag{2.5c}$$

- The Roy model (1951) and its extensions (Gronau, 1974; Heckman, 1974; Willis and Rosen, 1979; Heckman, 1990; Carneiro, Hansen, and Heckman, 2003) are at the core of microeconometrics.

$$Y_1 = X\beta_1 + U_1 \tag{2.5a}$$

$$Y_0 = X\beta_0 + U_0, \tag{2.5b}$$

and associated costs (prices) as a function of W

$$C = W\beta_C + U_C. \tag{2.5c}$$

- Can embed into general equilibrium models (Heckman, Lochner and Taber, 1998; Wolpin and Lee, 2006)

The valuation of "1" relative to "0" is $R = Y_1 - Y_0 - C$. Substituting from (2.5a)–(2.5c) into the expression for R:

$$R = X(\beta_1 - \beta_0) - W\beta_C + U_1 - U_0 - U_C,$$

and sectoral choice is indicated by D where $D = 1$ if the agent selects 1; $= 0$ otherwise:

$$D = \mathbf{1}[R > 0].$$

A two outcome normal example under perfect certainty

- $\upsilon = (U_1 - U_0 - U_C)$, $Z = (X, W)$.

A two outcome normal example under perfect certainty

- $\upsilon = (U_1 - U_0 - U_C)$, $Z = (X, W)$.
- $\gamma = (\beta_1 - \beta_0, -\beta_C)$.

- $\upsilon = (U_1 - U_0 - U_C)$, $Z = (X, W)$.
- $\gamma = (\beta_1 - \beta_0, -\beta_C)$.
- Thus $R = Z\gamma + \upsilon$.

A two outcome normal example under perfect certainty

- $\upsilon = (U_1 - U_0 - U_C)$, $Z = (X, W)$.
- $\gamma = (\beta_1 - \beta_0, -\beta_C)$.
- Thus $R = Z\gamma + \upsilon$.
- Generalized Roy model:

A two outcome normal example under perfect certainty

- $\upsilon = (U_1 - U_0 - U_C)$, $Z = (X, W)$.
- $\gamma = (\beta_1 - \beta_0, -\beta_C)$.
- Thus $R = Z\gamma + \upsilon$.
- Generalized Roy model:
  - $Z \perp\!\!\!\perp (U_0, U_1, U_C)$ (independence),

A two outcome normal example under perfect certainty

- $\upsilon = (U_1 - U_0 - U_C)$, $Z = (X, W)$.
- $\gamma = (\beta_1 - \beta_0, -\beta_C)$.
- Thus $R = Z\gamma + \upsilon$.
- Generalized Roy model:
  - $Z \perp\!\!\!\perp (U_0, U_1, U_C)$ (independence),
  - $(U_0, U_1, U_C) \sim \mathcal{N}(0, \Sigma)$ (normality).

For the Generalized Roy Model, the probability of selecting treatment 1 or "propensity score" is

$$\begin{aligned} \Pr(R > 0 \mid Z = z) &= \Pr(\upsilon > -z\gamma) \\ &= \Pr\left(\frac{\upsilon}{\sigma_\upsilon} > \frac{-z\gamma}{\sigma_\upsilon}\right) \\ &= \Phi\left(\frac{z\gamma}{\sigma_\upsilon}\right), \end{aligned}$$

where Φ is the cumulative distribution function of the standard normal distribution.

A two outcome normal example under perfect certainty

- The Average Treatment Effect given $X = x$ is

$$\begin{aligned} \text{ATE}(x) &= E\left(Y_1 - Y_0 \mid X = x\right) \\ &= x\left(\beta_1 - \beta_0\right). \end{aligned}$$

- The Average Treatment Effect given $X = x$ is

$$\begin{aligned} \mathrm{ATE}(x) &= E\left(Y_1 - Y_0 \mid X = x\right) \\ &= x\left(\beta_1 - \beta_0\right). \end{aligned}$$

- Treatment on the treated is

$$\begin{aligned} \mathrm{TT}\left(x, z\right) &= E\left(Y_1 - Y_0 \mid Z = z, D = 1\right) \\ &= x\left(\beta_1 - \beta_0\right) + E\left(U_1 - U_0 \mid \upsilon > -Z\gamma, Z = z\right) \\ &= x\left(\beta_1 - \beta_0\right) + E\left(U_1 - U_0 \mid \upsilon > -z\gamma\right). \end{aligned}$$

- The **local average treatment effect** (LATE) of Imbens and Angrist (1994) is the average gain to program participation for those induced to receive treatment through a change in $Z\,[=(X, W)]$ by a component of W not in X.

- The **local average treatment effect** (LATE) of Imbens and Angrist (1994) is the average gain to program participation for those induced to receive treatment through a change in $Z\,[=(X,W)]$ by a component of W not in X.
- The change affects choices but not potential outcomes $Y(s)$.

- The **local average treatment effect** (LATE) of Imbens and Angrist (1994) is the average gain to program participation for those induced to receive treatment through a change in $Z\,[= (X, W)]$ by a component of W not in X.
- The change affects choices but not potential outcomes $Y(s)$.
- Let $D(z)$ be the random variable D when we fix $W = w$ and let $D(z')$ be the random variable when we fix $W = w'$.

- The **local average treatment effect** (LATE) of Imbens and Angrist (1994) is the average gain to program participation for those induced to receive treatment through a change in $Z\,[=(X,W)]$ by a component of W not in X.

- The change affects choices but not potential outcomes $Y(s)$.

- Let $D(z)$ be the random variable D when we fix $W=w$ and let $D(z')$ be the random variable when we fix $W=w'$.

- This definition is instrument dependent.

A two outcome normal example under perfect certainty

- The **local average treatment effect** (LATE) of Imbens and Angrist (1994) is the average gain to program participation for those induced to receive treatment through a change in $Z\,[=(X,W)]$ by a component of W not in X.
- The change affects choices but not potential outcomes $Y(s)$.
- Let $D(z)$ be the random variable D when we fix $W=w$ and let $D(z')$ be the random variable when we fix $W=w'$.
- This definition is instrument dependent.
- There is a more general approach for defining this parameter (Heckman and Vytlacil, 1999, 2005).

- The LATE parameter is the mean return for people with values of $\upsilon \in [\underline{\upsilon}, \bar{\upsilon}]$.

$$
\begin{aligned}
\text{LATE}\,(z, z', x) \\
&= E\,(Y_1 - Y_0 \mid D(z) = 0, D(z') = 1, X = x) \\
&= x\,(\beta_1 - \beta_0) \\
&\qquad + E\,(U_1 - U_0 \mid R\,(z) \leq 0 \cap R\,(z') > 0, X = x) \\
&= x\,(\beta_1 - \beta_0) + E\,(U_1 - U_0 \mid -z'\gamma < \upsilon \leq -z\gamma)\,.
\end{aligned}
$$

- The LATE parameter is the mean return for people with values of $\upsilon \in [\underline{\upsilon}, \bar{\upsilon}]$.

$$
\begin{aligned}
&\text{LATE}(z, z', x) \\
&\quad = E\left(Y_1 - Y_0 \mid D(z) = 0, D(z') = 1, X = x\right) \\
&\quad = x\left(\beta_1 - \beta_0\right) \\
&\qquad + E\left(U_1 - U_0 \mid R(z) \leq 0 \cap R(z') > 0, X = x\right) \\
&\quad = x\left(\beta_1 - \beta_0\right) + E\left(U_1 - U_0 \mid -z'\gamma < \upsilon \leq -z\gamma\right).
\end{aligned}
$$

- Instruments W may not exist yet LATE can still be defined within the economic model as

$$
\begin{aligned}
&\text{LATE}\left(x, \upsilon \in [\underline{\upsilon}, \bar{\upsilon}]\right) \\
&\quad = x\left(\beta_1 - \beta_0\right) + E\left(U_1 - U_0 \mid \underline{\upsilon} < \upsilon \leq \bar{\upsilon}\right).
\end{aligned}
$$

- By Vytlacil's Theorem (2002), these two approaches are equivalent.

- By Vytlacil's Theorem (2002), these two approaches are equivalent.
- Will provide precise conditions for this equivalence in tomorrow's lecture. (See also Heckman and Vytlacil (2005))

- The **marginal treatment effect** (MTE) is defined conditional on X, Z, and $\upsilon = \upsilon^*$:

$$\begin{aligned} & E\left(Y_1 - Y_0 \mid \upsilon = \upsilon^*, X = x, Z = z\right) \\ = \ & x\left(\beta_1 - \beta_0\right) + E\left(U_1 - U_0 \mid \upsilon = \upsilon^*\right). \end{aligned}$$

It is the mean return for persons for whom $X = x$, $Z = z$, and $\upsilon = \upsilon^*$. It is defined independently of any instrument.

- The **marginal treatment effect** (MTE) is defined conditional on X, Z, and $\upsilon = \upsilon^*$:

$$\begin{aligned} & E\left(Y_1 - Y_0 \mid \upsilon = \upsilon^*, X = x, Z = z\right) \\ = \; & x\left(\beta_1 - \beta_0\right) + E\left(U_1 - U_0 \mid \upsilon = \upsilon^*\right). \end{aligned}$$

  It is the mean return for persons for whom $X = x$, $Z = z$, and $\upsilon = \upsilon^*$. It is defined independently of any instrument.

- At a special point of evaluation where $R = 0$ (i.e. $z\gamma + \upsilon = 0$), the MTE is a willingness to pay measure that informs us how much an agent at the margin of participation (in the indifference set) would be willing to pay to move from "0" to "1".

- Under regularity conditions, MTE is a limit form of LATE,

$$
\begin{aligned}
&\lim_{z\gamma \to z'\gamma} \text{LATE}(z, z', x) \\
&\qquad = x(\beta_1 - \beta_0) + \lim_{z\gamma \to z'\gamma} E(U_1 - U_0 \mid -z\gamma < v < -z'\gamma) \\
&\qquad = x(\beta_1 - \beta_0) + E(U_1 - U_0 \mid v = -z'\gamma)
\end{aligned}
$$

- Under regularity conditions, MTE is a limit form of LATE,

$$\begin{aligned}
&\lim_{z\gamma \to z'\gamma} \mathrm{LATE}\left(z, z', x\right) \\
&\quad = x\left(\beta_1 - \beta_0\right) + \lim_{z\gamma \to z'\gamma} E\left(U_1 - U_0 \mid -z\gamma < \upsilon < -z'\gamma\right) \\
&\quad = x\left(\beta_1 - \beta_0\right) + E\left(U_1 - U_0 \mid \upsilon = -z'\gamma\right)
\end{aligned}$$

- LATE is the average return for persons with $\upsilon \in \left[-z\gamma, -z'\gamma\right]$.

- Can work with $Z\gamma$ or with the propensity score $P(Z)$ interchangeably assuming V is absolutely continuous.

- Can work with $Z\gamma$ or with the propensity score $P(Z)$ interchangeably assuming V is absolutely continuous.
- $\Pr(Z\gamma > V) = \Pr(F_V(Z\gamma) > F_V(V))$.

A two outcome normal example under perfect certainty

$$\begin{aligned}
\mathrm{TT}(x,z) &= \mathrm{TT}(x,P(z)) \\
&= x(\beta_1-\beta_0) \\
&\qquad +\mathrm{Cov}(U_1-U_0,\upsilon)\underbrace{K(P(z))}_{\substack{\text{“control function”} \\ \text{Heckman and Robb (1985)}}}>0.
\end{aligned}$$

- Given the model, can build up these and other parameters.

$$
\begin{aligned}
\mathrm{TT}(x,z) &= \mathrm{TT}(x,P(z)) \\
&= x(\beta_1-\beta_0) \\
&\qquad +\mathrm{Cov}(U_1-U_0,\upsilon)\underbrace{K(P(z))}_{\text{“control function”}} > 0.
\end{aligned}
$$

Heckman and Robb (1985)

- Given the model, can build up these and other parameters.
- But for each of these parameters, we do not need to specify the full model to identify them.

$$
\begin{aligned}
\mathrm{TT}(x,z) &= \mathrm{TT}(x,P(z)) \\
&= x(\beta_1-\beta_0) \\
&\qquad +\mathrm{Cov}(U_1-U_0,\upsilon)\underbrace{K(P(z))}_{\substack{\text{“control function”}\\ \text{Heckman and Robb (1985)}}}>0.
\end{aligned}
$$

- Given the model, can build up these and other parameters.
- But for each of these parameters, we do not need to specify the full model to identify them.
- This is a main insight of the modern treatment effect literature.

- If we do specify and identify the full model, however, we can solve policy problem (P-2) (the extrapolation problem) using this model evaluated at new values of (X, Z).

- If we do specify and identify the full model, however, we can solve policy problem (P-2) (the extrapolation problem) using this model evaluated at new values of (X, Z).
- By construction the (U_1, U_0, υ) are independent of (X, Z), and given the functional forms all the mean treatment parameters can be generated for all (X, Z).

- If we do specify and identify the full model, however, we can solve policy problem (P-2) (the extrapolation problem) using this model evaluated at new values of (X, Z).
- By construction the (U_1, U_0, υ) are independent of (X, Z), and given the functional forms all the mean treatment parameters can be generated for all (X, Z).
- By parameterizing the β_i to depend only on measured characteristics, it is possible to forecast the demand for new goods and solve policy problem (P-3).

A two outcome normal example under perfect certainty

- Consider the following example.

A two outcome normal example under perfect certainty

- Consider the following example.
- Used throughout these lectures.

A two outcome normal example under perfect certainty

- Consider the following example.
- Used throughout these lectures.
- Distribution of gross gains to a country, $(Y_1 - Y_0)$, from adopting a policy in a Roy model.

Figure 1: Extended Roy economy for policy adoption
Distribution of gains and treatment parameters

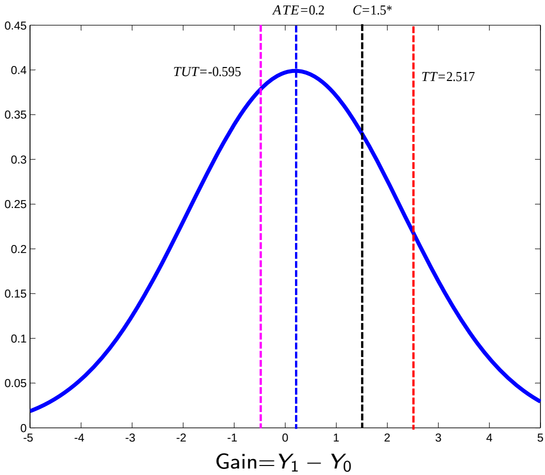


Figure 1 Legend

Suppose that a country has to choose whether to implement a policy. Under the policy, the GDP would be Y_1. Without the policy, the GDP of the country would be Y_0. For the sake of simplicity, suppose that

$$\begin{aligned} Y_1 &= \mu_1 + U_1 \\ Y_0 &= \mu_0 + U_0 \end{aligned}$$

where U_0 and U_1 are unobserved components of the aggregate output. The error terms (U_0, U_1) are dependent in a general way. Let δ denote the additional GDP due to the policy, i.e. $\delta = \mu_1 - \mu_0$. We assume $\delta > 0$. Let C denote the cost of implementing the policy. We assume that the cost is a fixed parameter C.

Figure 1 Legend

We relax this assumption below. The country's decision can be represented as:

$$D = \begin{cases} 1 & \text{if} \quad Y_1 - Y_0 - C > 0 \\ 0 & \text{if} \quad Y_1 - Y_0 - C \leq 0, \end{cases}$$

so the country decides to implement the policy $(D = 1)$ if the net gains coming from it are positive. Therefore, we can define the probability of adopting the policy in terms of the propensity score

$$\Pr(D = 1) = P(Y_1 - Y_0 - C > 0).$$

We assume that $(U_1, U_0) \sim N(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$,$\mu_0 = 0.67$, $\delta = 0.2$ and $C = 1.5$.

A two outcome normal example under perfect certainty

- The distribution of gains to adoption arises from the variability in policy effectiveness across countries.

A two outcome normal example under perfect certainty

- The distribution of gains to adoption arises from the variability in policy effectiveness across countries.
- The model builds in positive sorting on unobservables because $\upsilon = U_1 - U_0$ so $\text{Cov}(U_1 - U_0, \upsilon) > 0$.

A two outcome normal example under perfect certainty

- The distribution of gains to adoption arises from the variability in policy effectiveness across countries.
- The model builds in positive sorting on unobservables because $\upsilon = U_1 - U_0$ so $\mathrm{Cov}(U_1 - U_0, \upsilon) > 0$.
- All countries face the same cost of policy adoption C.

A two outcome normal example under perfect certainty

- The distribution of gains to adoption arises from the variability in policy effectiveness across countries.
- The model builds in positive sorting on unobservables because $\upsilon = U_1 - U_0$ so $\mathrm{Cov}(U_1 - U_0, \upsilon) > 0$.
- All countries face the same cost of policy adoption C.
- The return to the policy in the randomly selected country is ATE $(= .2)$. Given $C = 1.5$, the return to the person at the margin is 1.5.

A two outcome normal example under perfect certainty

- The distribution of gains to adoption arises from the variability in policy effectiveness across countries.
- The model builds in positive sorting on unobservables because $\upsilon = U_1 - U_0$ so $\text{Cov}(U_1 - U_0, \upsilon) > 0$.
- All countries face the same cost of policy adoption C.
- The return to the policy in the randomly selected country is ATE $(= .2)$. Given $C = 1.5$, the return to the person at the margin is 1.5.
- The average return for the adopting countries is TT $(= 2.52)$.

- The distribution of gains to adoption arises from the variability in policy effectiveness across countries.
- The model builds in positive sorting on unobservables because $\upsilon = U_1 - U_0$ so $\text{Cov}(U_1 - U_0, \upsilon) > 0$.
- All countries face the same cost of policy adoption C.
- The return to the policy in the randomly selected country is ATE $(= .2)$. Given $C = 1.5$, the return to the person at the margin is 1.5.
- The average return for the adopting countries is TT $(= 2.52)$.
- Thus the countries adopting the policy are the ones who benefit from it. This is a source of evaluation bias in comparing policy effectiveness in different countries.

A two outcome normal example under perfect certainty

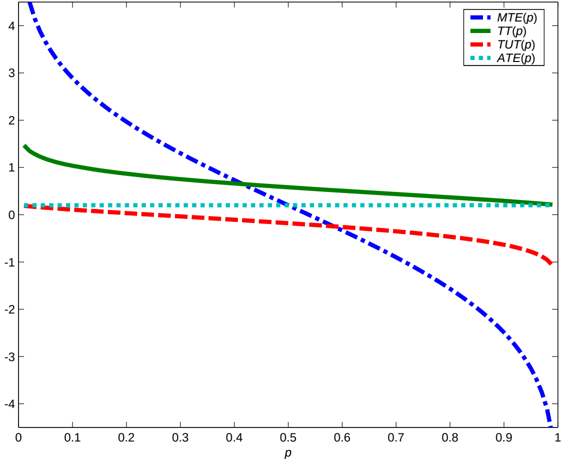
Parameter	Definition	Under Assumptions(*)
Marginal Treatment Effect	$E\left[Y_1 - Y_0 \mid R = 0, P(Z) = p\right]$	$\delta + \Phi^{-1}_{U_1 - U_0}(1-p)$
Average Treatment Effect	$E\left[Y_1 - Y_0 \mid P(Z) = p\right]$	φ
Treatment on the Treated	$E\left[Y_1 - Y_0 \mid R > 0, P(Z) = p\right]$	$\varphi + \dfrac{\phi_{U_1 - U_0}\left(\Phi^{-1}_{U_1 - U_0}(1-p)\right)}{p}$
Treatment on the Untreated	$E\left[Y_1 - Y_0 \mid R \leq 0, P(Z) = p\right]$	$\varphi - \dfrac{\phi_{U_1 - U_0}\left(\Phi^{-1}_{U_1 - U_0}(1-p)\right)}{1-p}$

Definitions of treatment parameters for the model given in figure 1.

Figure 2 plots the parameters $ATE(p)$, $\mathrm{TT}(p)$, $\mathrm{MTE}(p)$ and $\mathrm{TUT}(p)$ (treatment on the untreated) that underlie the model used to generate figure 1.

A two outcome normal example under perfect certainty

Figure 2: Extended Roy economy example, treatment parameters as a function of $\Pr(D = 1 \mid Z = z) = p$



Model generated by the parameters from the model at base of Figure 1.

A two outcome normal example under perfect certainty

- The declining MTE(p) is the prototypical pattern of diminishing returns that accompanies a policy expansion.

- The declining MTE(p) is the prototypical pattern of diminishing returns that accompanies a policy expansion.
- Countries with low levels of $Z\gamma$ ($P(Z)$) that adopt the policy must do so because their unobservables make them more likely to.

- The declining MTE(p) is the prototypical pattern of diminishing returns that accompanies a policy expansion.
- Countries with low levels of $Z\gamma$ ($P(Z)$) that adopt the policy must do so because their unobservables make them more likely to.
- As costs C fall, more countries are drawn in to adopt the policy, the return falls.

A two outcome normal example under perfect certainty

- The pattern for treatment on the treated (TT(p)) is explained by similar considerations. As participation becomes less selective, the selected country outcomes converge to the population average.

A two outcome normal example under perfect certainty

- The pattern for treatment on the treated (TT(p)) is explained by similar considerations. As participation becomes less selective, the selected country outcomes converge to the population average.
- As more countries participate, the stragglers are, on average, less effective adopters of the policy.

- The pattern for treatment on the treated $(\mathrm{TT}(p))$ is explained by similar considerations. As participation becomes less selective, the selected country outcomes converge to the population average.
- As more countries participate, the stragglers are, on average, less effective adopters of the policy.
- This explains the pattern for $\mathrm{TUT}(p)$.

- The pattern for treatment on the treated (TT(p)) is explained by similar considerations. As participation becomes less selective, the selected country outcomes converge to the population average.
- As more countries participate, the stragglers are, on average, less effective adopters of the policy.
- This explains the pattern for TUT(p).
- We get these parameters if we identify the full model.

A two outcome normal example under perfect certainty

- The pattern for treatment on the treated (TT(p)) is explained by similar considerations. As participation becomes less selective, the selected country outcomes converge to the population average.
- As more countries participate, the stragglers are, on average, less effective adopters of the policy.
- This explains the pattern for TUT(p).
- We get these parameters if we identify the full model.
- But do we need to?

- The pattern for treatment on the treated (TT(p)) is explained by similar considerations. As participation becomes less selective, the selected country outcomes converge to the population average.
- As more countries participate, the stragglers are, on average, less effective adopters of the policy.
- This explains the pattern for TUT(p).
- We get these parameters if we identify the full model.
- But do we need to?
- We consider this question but first consider a version of the analysis that allows for uncertainty.

- The agent may know things in advance that the econometrician may never discover.

- The agent may know things in advance that the econometrician may never discover.
- On the other hand, the econometrician, benefitting from hindsight, may know some information that the agent does not know when he is making his choices.

Adding uncertainty

- The agent may know things in advance that the econometrician may never discover.
- On the other hand, the econometrician, benefitting from hindsight, may know some information that the agent does not know when he is making his choices.
- Let $\mathcal{I}_a$ be the information set confronting the agent at the time choices are made and before outcomes are realized.

- The agent may know things in advance that the econometrician may never discover.
- On the other hand, the econometrician, benefitting from hindsight, may know some information that the agent does not know when he is making his choices.
- Let $\mathcal{I}_a$ be the information set confronting the agent at the time choices are made and before outcomes are realized.
- Agents may only imperfectly estimate consequences of their choices.

- The *ex ante* vs. *ex post* distinction is essential for understanding behavior.

- The *ex ante* vs. *ex post* distinction is essential for understanding behavior.
- In environments of uncertainty, agent choices are made in terms of *ex ante* calculations.

Adding uncertainty

- The *ex ante* vs. *ex post* distinction is essential for understanding behavior.
- In environments of uncertainty, agent choices are made in terms of *ex ante* calculations.
- Yet the treatment effect literature largely reports *ex post* returns.

As Hicks (1946, p. 179) puts it,

> "*Ex post calculations of capital accumulation have their place in economic and statistical history; they are useful measures for economic progress; but they are of no use to theoretical economists who are trying to find out how the system works, because they have no significance for conduct.*

- Define $R(\mathcal{I}_a)$ as

$$R(\mathcal{I}_a) = E(Y_1 - Y_0 - C \mid \mathcal{I}_a).$$

Adding uncertainty

- Define $R(\mathcal{I}_a)$ as

$$R(\mathcal{I}_a) = E(Y_1 - Y_0 - C \mid \mathcal{I}_a).$$

- Under perfect foresight, the agent knows Y_1, Y_0 and C as in the classical generalized Roy model.

- Define $R(\mathcal{I}_a)$ as

$$R(\mathcal{I}_a) = E(Y_1 - Y_0 - C \mid \mathcal{I}_a).$$

- Under perfect foresight, the agent knows Y_1, Y_0 and C as in the classical generalized Roy model.
- $\mathcal{I}_a \supseteq \{Y_1, Y_0, C\}$.

- Define $R\left(\mathcal{I}_a\right)$ as

$$R\left(\mathcal{I}_a\right) = E\left(Y_1 - Y_0 - C \mid \mathcal{I}_a\right).$$

- Under perfect foresight, the agent knows Y_1, Y_0 and C as in the classical generalized Roy model.
- $\mathcal{I}_a \supseteq \{Y_1, Y_0, C\}$.
- More generally, the choice equation is generated by $D\left(\mathcal{I}_a\right) = \mathbf{1}\left[R\left(\mathcal{I}_a\right) > 0\right]$.

- Define $R(\mathcal{I}_a)$ as
$$R(\mathcal{I}_a) = E(Y_1 - Y_0 - C \mid \mathcal{I}_a).$$
- Under perfect foresight, the agent knows Y_1, Y_0 and C as in the classical generalized Roy model.
- $\mathcal{I}_a \supseteq \{Y_1, Y_0, C\}$.
- More generally, the choice equation is generated by $D(\mathcal{I}_a) = \mathbf{1}[R(\mathcal{I}_a) > 0]$.
- *Ex post*, different choices might be made.

- Define $R(\mathcal{I}_a)$ as

$$R(\mathcal{I}_a) = E(Y_1 - Y_0 - C \mid \mathcal{I}_a).$$

- Under perfect foresight, the agent knows Y_1, Y_0 and C as in the classical generalized Roy model.
- $\mathcal{I}_a \supseteq \{Y_1, Y_0, C\}$.
- More generally, the choice equation is generated by $D(\mathcal{I}_a) = \mathbf{1}[R(\mathcal{I}_a) > 0]$.
- *Ex post*, different choices might be made.
- The econometrician may possess yet a different information set $\mathcal{I}_e$.

Adding uncertainty

- Define $R\left(\mathcal{I}_a\right)$ as
$$R\left(\mathcal{I}_a\right) = E\left(Y_1 - Y_0 - C \mid \mathcal{I}_a\right).$$
- Under perfect foresight, the agent knows Y_1, Y_0 and C as in the classical generalized Roy model.
- $\mathcal{I}_a \supseteq \{Y_1, Y_0, C\}$.
- More generally, the choice equation is generated by $D\left(\mathcal{I}_a\right) = \mathbf{1}\left[R\left(\mathcal{I}_a\right) > 0\right]$.
- *Ex post*, different choices might be made.
- The econometrician may possess yet a different information set $\mathcal{I}_e$.
- Stay tuned for the Friday lecture.

Counterfactuals, causality and structural econometric models

- The literature on policy evaluation in economics often contrasts "structural" approaches with "treatment effect" or "causal" models.

Counterfactuals, causality and structural econometric models

- The literature on policy evaluation in economics often contrasts "structural" approaches with "treatment effect" or "causal" models.
- Compare the econometric model for generating counterfactuals and causal effects with the Neyman (1923) – Rubin (1978) model of causality and compare "causal" parameters with "structural" parameters.

Counterfactuals, causality and structural econometric models

- The literature on policy evaluation in economics often contrasts "structural" approaches with "treatment effect" or "causal" models.
- Compare the econometric model for generating counterfactuals and causal effects with the Neyman (1923) – Rubin (1978) model of causality and compare "causal" parameters with "structural" parameters.
- This model is widely used in statistics and epidemiology.

Counterfactuals, causality and structural econometric models

- The literature on policy evaluation in economics often contrasts "structural" approaches with "treatment effect" or "causal" models.
- Compare the econometric model for generating counterfactuals and causal effects with the Neyman (1923) – Rubin (1978) model of causality and compare "causal" parameters with "structural" parameters.
- This model is widely used in statistics and epidemiology.
- It is advocated as a model for "causal analysis" by economists who don't know much economics.

- The treatment effect approach and the explicitly economic approach differ in the detail with which they specify both observed and counterfactual outcomes $Y(s, \omega)$, for different treatments denoted by "s."

- The treatment effect approach and the explicitly economic approach differ in the detail with which they specify both observed and counterfactual outcomes $Y(s, \omega)$, for different treatments denoted by "s."
- The econometric approach models counterfactuals much more explicitly than is common in the application of the treatment effect approach.

- The treatment effect approach and the explicitly economic approach differ in the detail with which they specify both observed and counterfactual outcomes $Y(s, \omega)$, for different treatments denoted by "s."
- The econometric approach models counterfactuals much more explicitly than is common in the application of the treatment effect approach.
- This difference in detail corresponds to the differing objectives of the two approaches.

- This greater attention to detail in the structural approach facilitates the application of theory to provide interpretations of counterfactuals and comparison of counterfactuals across data sets using the basic parameters of economic theory.

- This greater attention to detail in the structural approach facilitates the application of theory to provide interpretations of counterfactuals and comparison of counterfactuals across data sets using the basic parameters of economic theory.
- Structural approach seeks to answer (P-1)-(P-3).

- This greater attention to detail in the structural approach facilitates the application of theory to provide interpretations of counterfactuals and comparison of counterfactuals across data sets using the basic parameters of economic theory.
- Structural approach seeks to answer (P-1)-(P-3).
- This was the goal of Monograph 10.

Causal Effects: In Economics and in Statistics

- Cowles was the first group to formalize the notion of causality in a probability model.

Causal Effects: In Economics and in Statistics

- Cowles was the first group to formalize the notion of causality in a probability model.
- Distinction between *fixing* and *conditioning* on inputs is central to distinguishing true causal effects from spurious causal effects.

Causal Effects: In Economics and in Statistics

- Cowles was the first group to formalize the notion of causality in a probability model.
- Distinction between *fixing* and *conditioning* on inputs is central to distinguishing true causal effects from spurious causal effects.
- Haavelmo (1943) made this distinction in linear equation models.

Causal Effects: In Economics and in Statistics

- Cowles was the first group to formalize the notion of causality in a probability model.
- Distinction between *fixing* and *conditioning* on inputs is central to distinguishing true causal effects from spurious causal effects.
- Haavelmo (1943) made this distinction in linear equation models.
- Haavelmo's distinction is the basis for Pearl's 2000 book on causality that generalizes Haavelmo's analysis to nonlinear settings.

Causal Effects: In Economics and in Statistics

- Pearl defines an operator "do" to represent the mental act of fixing a variable to distinguish it from the action of conditioning which is a statistical operation.

Causal Effects: In Economics and in Statistics

- Pearl defines an operator "do" to represent the mental act of fixing a variable to distinguish it from the action of conditioning which is a statistical operation.
- $$Y = X\beta + U.$$

  "Nature" or the "real world" picks (X, U) to determine Y.

Causal Effects: In Economics and in Statistics

- Pearl defines an operator "do" to represent the mental act of fixing a variable to distinguish it from the action of conditioning which is a statistical operation.
- $$Y = X\beta + U.$$
  "Nature" or the "real world" picks (X, U) to determine Y.
- X is observed by the analyst and U is not observed, and (X, U) are random variables.

Causal Effects: In Economics and in Statistics

- Pearl defines an operator "do" to represent the mental act of fixing a variable to distinguish it from the action of conditioning which is a statistical operation.
- $$Y = X\beta + U.$$
  "Nature" or the "real world" picks (X, U) to determine Y.
- X is observed by the analyst and U is not observed, and (X, U) are random variables.
- This is an "all causes" model in which (X, U) determine Y. The variation generated by the hypothetical model varies one coordinate of (X, U), fixing all other coordinates to produce the effect of the variation on the outcome Y.

Causal Effects: In Economics and in Statistics

- Nature (as opposed to the model) may not permit such variation.

Causal Effects: In Economics and in Statistics

- Nature (as opposed to the model) may not permit such variation.
- We can write this model formulated at the population level as a conditional expectation:

$$E(Y \mid X = x, U = u) = x\beta + u.$$

Causal Effects: In Economics and in Statistics

- Nature (as opposed to the model) may not permit such variation.
- We can write this model formulated at the population level as a conditional expectation:

$$E(Y \mid X = x, U = u) = x\beta + u.$$

- Since we condition on both X and U, there is no further source of variation in Y in an "all causes" model.

Causal Effects: In Economics and in Statistics

- Fixing X at different values corresponds to doing different thought experiments with the X.

Causal Effects: In Economics and in Statistics

- Fixing X at different values corresponds to doing different thought experiments with the X.
- Fixing and conditioning are the same in this case.
  If, however, we only condition on X, we obtain

$$E(Y \mid X = x) = x\beta + E(U \mid X = x). \tag{3.1}$$

Causal Effects: In Economics and in Statistics

- Fixing X at different values corresponds to doing different thought experiments with the X.
- Fixing and conditioning are the same in this case.
  If, however, we only condition on X, we obtain

$$E(Y \mid X = x) = x\beta + E(U \mid X = x). \tag{3.1}$$

- This relationship does not generate U-constant (Y, X) relationships.

The econometric model vs. the Neyman–Rubin ‘causal” model

- The causal model most popular in statistics and epidemiology draws on hypothetical experiments to define causality.

The econometric model vs. the Neyman–Rubin ‘causal” model

- The causal model most popular in statistics and epidemiology draws on hypothetical experiments to define causality.
- Sometimes zealots claim that causality can only be determined if randomization is actually possible.

- The causal model most popular in statistics and epidemiology draws on hypothetical experiments to define causality.
- Sometimes zealots claim that causality can only be determined if randomization is actually possible.
- Neyman and Rubin postulate counterfactuals $\{Y(s,\omega)\}_{s\in\mathcal{S}}$ without modeling the factors determining the $Y(s,\omega)$ as is done in the “structural” approach.

- The causal model most popular in statistics and epidemiology draws on hypothetical experiments to define causality.
- Sometimes zealots claim that causality can only be determined if randomization is actually possible.
- Neyman and Rubin postulate counterfactuals $\{Y(s,\omega)\}_{s\in\mathcal{S}}$ without modeling the factors determining the $Y(s,\omega)$ as is done in the "structural" approach.
- Rubin and Neyman offer no model of the choice of which outcome is selected.

The econometric model vs. the Neyman–Rubin ‘causal” model

- In our notation, Neyman (1923) and Rubin assume (PI-1) and (PI-2), but not (PI-3) or (PI-4), since choice is not modeled.

The "Rubin Model"

R-1

$\{Y(s,\omega)\}_{s\in S}$, *a set of counterfactuals defined for ex post outcomes. It does not analyze valuations of outcomes nor does it explicitly specify treatment selection rules, except for contrasting randomization with nonrandomization.*

R-2

(PI-1) Invariance of counterfactuals to the assignment mechanism of treatment.

The "Rubin Model"

R-3

No social interactions or general equilibrium effects (PI-2).

R-4

There is no simultaneity in causal effects, i.e., outcomes cannot cause each other reciprocally.

Two further implicit assumptions in the application of the model are:

- (P-1) is the only problem of interest.
- Mean causal effects are the only objects of interest.
- No analysis of choice behavior.

The econometric approach is richer than the statistical treatment effect approach

Its signature features are:

The econometric approach is richer than the statistical treatment effect approach

Its signature features are:

1. Development of an explicit framework for outcomes, measurements and choice of outcomes where the role of unobservables ("missing variables") in creating selection problems and justifying estimators is explicitly developed.

The econometric approach is richer than the statistical treatment effect approach

Its signature features are:

1. Development of an explicit framework for outcomes, measurements and choice of outcomes where the role of unobservables ("missing variables") in creating selection problems and justifying estimators is explicitly developed.
2. The analysis of subjective evaluations of outcomes and the use of choice data to infer them.

The econometric approach is richer than the statistical treatment effect approach

3. The analysis of *ex ante* and *ex post* realizations and evaluations of treatments. This analysis enables analysts to model and identify regret and anticipation by agents. Points 2 and 3 introduce human decision making into the treatment effect literature.

The econometric approach is richer than the statistical treatment effect approach

3. The analysis of *ex ante* and *ex post* realizations and evaluations of treatments. This analysis enables analysts to model and identify regret and anticipation by agents. Points 2 and 3 introduce human decision making into the treatment effect literature.

4. Development of models for identifying entire distributions of treatment effects (*ex ante* and *ex post*) rather than just the traditional mean parameters focused on by statisticians. These distributions enable analysts to determine the proportion of people who benefit from treatment, something not attempted in the statistical literature on treatment effects.

The econometric approach is richer than the statistical treatment effect approach

5. Development and identification of distributional criteria allowing for analysis of alternative social welfare functions for outcome distributions comparing different treatment states.

The econometric approach is richer than the statistical treatment effect approach

5. Development and identification of distributional criteria allowing for analysis of alternative social welfare functions for outcome distributions comparing different treatment states.
6. Models for simultaneous causality.

The econometric approach is richer than the statistical treatment effect approach

5. Development and identification of distributional criteria allowing for analysis of alternative social welfare functions for outcome distributions comparing different treatment states.
6. Models for simultaneous causality.
7. Definitions of parameters made without appeals to hypothetical experimental manipulations.

The econometric approach is richer than the statistical treatment effect approach

5. Development and identification of distributional criteria allowing for analysis of alternative social welfare functions for outcome distributions comparing different treatment states.
6. Models for simultaneous causality.
7. Definitions of parameters made without appeals to hypothetical experimental manipulations.
8. Clarification of the need for invariance of parameters with respect to classes of manipulations to answer classes of questions. This notion is featured in the early Cowles Commission work. See Marschak (1953), Koopmans et al. (1950) and Hurwicz (1962).

- Economists separate out the three tasks in table 1.

- Economists separate out the three tasks in table 1.
- Statisticians sometimes conflate them.

- Economists separate out the three tasks in table 1.
- Statisticians sometimes conflate them.
- These distinctions are very clear in Cowles Monograph 10.

Table 1: Three distinct tasks arising in the analysis of causal models

Task	Description	Requirements
1	Defining the Set of Hypotheticals or Counterfactuals	A Scientific Theory
2	Identifying Parameters (Causal or Otherwise) from Hypothetical Population Data	Mathematical Analysis of Point or Set Identification
3	Identifying Parameters from Data	Estimation and Testing Theory

- Holland claims that there can be no causal effect of gender on earnings because analysts cannot randomly assign gender.

- Holland claims that there can be no causal effect of gender on earnings because analysts cannot randomly assign gender.
- This statement confuses the act of definition of the causal effect (a purely mental act) with empirical difficulties in estimating it (Steps 1 and 2 in Table 1).

- Holland claims that there can be no causal effect of gender on earnings because analysts cannot randomly assign gender.
- This statement confuses the act of definition of the causal effect (a purely mental act) with empirical difficulties in estimating it (Steps 1 and 2 in Table 1).
- In the statistics literature, a causal effect is **defined** by a randomization.

- Holland claims that there can be no causal effect of gender on earnings because analysts cannot randomly assign gender.
- This statement confuses the act of definition of the causal effect (a purely mental act) with empirical difficulties in estimating it (Steps 1 and 2 in Table 1).
- In the statistics literature, a causal effect is **defined** by a randomization.
- Issues of definition and identification are confused.

- A major limitation of the Neyman–Rubin model is that it is recursive. It cannot model causal effects of outcomes that occur simultaneously.

- A major limitation of the Neyman–Rubin model is that it is recursive. It cannot model causal effects of outcomes that occur simultaneously.
- By Cowles Monograph 10 this problem had been solved in econometrics.

- A major limitation of the Neyman–Rubin model is that it is recursive. It cannot model causal effects of outcomes that occur simultaneously.
- By Cowles Monograph 10 this problem had been solved in econometrics.
- It remains an open problem in statistics.

- Write the standard model of simultaneous equations in terms of parameters (Γ, B), observables (Y, X) and unobservables U as

$$\Gamma Y + BX = U, \qquad E(U) = 0, \tag{3.2}$$

where Y is a vector of endogenous and interdependent variables, X is exogenous $(E(U \mid X) = 0)$, and Γ is a full rank matrix.

- Write the standard model of simultaneous equations in terms of parameters (Γ, B), observables (Y, X) and unobservables U as

$$\Gamma Y + BX = U, \qquad E(U) = 0, \tag{3.2}$$

where Y is a vector of endogenous and interdependent variables, X is exogenous $(E(U \mid X) = 0)$, and Γ is a full rank matrix.

- Equation systems like (3.2) are sometimes called "structural equations."

- The Y are "internal" variables determined by the model and the X are "external" variables specified outside the model.

- The Y are "internal" variables determined by the model and the X are "external" variables specified outside the model.
- Assume the model is complete (Γ^{-1} exists), gives unique Y.

- The Y are "internal" variables determined by the model and the X are "external" variables specified outside the model.
- Assume the model is complete (Γ^{-1} exists), gives unique Y.
- Reduced form is $Y = \Pi X + R$ where $\Pi = -\Gamma^{-1}B$ and $R = \Gamma^{-1}U$.

- The "structure" is (Γ, B), Σ_U, where Σ_U is the variance-covariance matrix of U.

- The "structure" is (Γ, B), Σ_U, where Σ_U is the variance-covariance matrix of U.
- Assume that Γ, B, Σ_U are invariant to general changes in X and translations of U.

- The "structure" is (Γ, B) , Σ_U, where Σ_U is the variance-covariance matrix of U.
- Assume that Γ, B, Σ_U are invariant to general changes in X and translations of U.
- Without restrictions, *ceteris paribus* manipulations associated with the effect of some components of Y on other components of Y are not possible within the model.

Nonrecursive (simultaneous) models of causality

- Consider a two person model of social interactions.

Nonrecursive (simultaneous) models of causality

- Consider a two person model of social interactions.
- Y_1 is the outcome for agent 1;

- Consider a two person model of social interactions.
- Y_1 is the outcome for agent 1;
- Y_2 is the outcome for agent 2.

- Consider a two person model of social interactions.
- Y_1 is the outcome for agent 1;
- Y_2 is the outcome for agent 2.
-
$$\begin{aligned} Y_1 &= \alpha_1 + \gamma_{12} Y_2 + \beta_{11} X_1 + \beta_{12} X_2 + U_1 \quad (3.3a) \\ Y_2 &= \alpha_2 + \gamma_{21} Y_1 + \beta_{21} X_1 + \beta_{22} X_2 + U_2. \quad (3.3b) \end{aligned}$$

$$E\left(U_1 \mid X_1, X_2\right) = 0 \qquad (3.4a)$$

and

$$E\left(U_2 \mid X_1, X_2\right) = 0. \qquad (3.4b)$$

- Consider a two person model of social interactions.
- Y_1 is the outcome for agent 1;
- Y_2 is the outcome for agent 2.
- $$\begin{aligned} Y_1 &= \alpha_1 + \gamma_{12} Y_2 + \beta_{11} X_1 + \beta_{12} X_2 + U_1 \quad (3.3a) \\ Y_2 &= \alpha_2 + \gamma_{21} Y_1 + \beta_{21} X_1 + \beta_{22} X_2 + U_2. \quad (3.3b) \end{aligned}$$

$$E\left(U_1 \mid X_1, X_2\right) = 0 \tag{3.4a}$$

and

$$E\left(U_2 \mid X_1, X_2\right) = 0. \tag{3.4b}$$

- Causal effect of Y_2 on Y_1 is γ_{12}.

Nonrecursive (simultaneous) models of causality

- With no exclusions or *a priori* information cannot identify the causal effect.

- With no exclusions or *a priori* information cannot identify the causal effect.
- Cowles Monographs 10 and 14 showed how to solve this problem: Assume exclusion ($\beta_{12} = 0$).

- With no exclusions or *a priori* information cannot identify the causal effect.
- Cowles Monographs 10 and 14 showed how to solve this problem: Assume exclusion ($\beta_{12} = 0$).
- We can identify the *ceteris paribus* causal effect of Y_2 on Y_1.

- With no exclusions or *a priori* information cannot identify the causal effect.
- Cowles Monographs 10 and 14 showed how to solve this problem: Assume exclusion ($\beta_{12} = 0$).
- We can identify the *ceteris paribus* causal effect of Y_2 on Y_1.
- Thus if $\beta_{12} = 0$, from the reduced form

$$\frac{\pi_{12}}{\pi_{22}} = \gamma_{12}.$$

- With no exclusions or *a priori* information cannot identify the causal effect.
- Cowles Monographs 10 and 14 showed how to solve this problem: Assume exclusion ($\beta_{12} = 0$).
- We can identify the *ceteris paribus* causal effect of Y_2 on Y_1.
- Thus if $\beta_{12} = 0$, from the reduced form

$$\frac{\pi_{12}}{\pi_{22}} = \gamma_{12}.$$

- Other restrictions possible.

Alternative Definitions of Structure

- "Structural" is a term used loosely.

Alternative Definitions of Structure

- "Structural" is a term used loosely.
- In some places, it means "doing economics to analyze data."

Alternative Definitions of Structure

- "Structural" is a term used loosely.
- In some places, it means "doing economics to analyze data."
- In other places, it means "imposing arbitrary assumptions onto the data."

Alternative Definitions of Structure

- "Structural" is a term used loosely.
- In some places, it means "doing economics to analyze data."
- In other places, it means "imposing arbitrary assumptions onto the data."
- Four distinct meanings:

Alternative Definitions of Structure

- "Structural" is a term used loosely.
- In some places, it means "doing economics to analyze data."
- In other places, it means "imposing arbitrary assumptions onto the data."
- Four distinct meanings:
  1. Cowles definition just given.

Alternative Definitions of Structure

- "Structural" is a term used loosely.
- In some places, it means "doing economics to analyze data."
- In other places, it means "imposing arbitrary assumptions onto the data."
- Four distinct meanings:
  1. Cowles definition just given.
  2. Low-dimensional model.

Alternative Definitions of Structure

- "Structural" is a term used loosely.
- In some places, it means "doing economics to analyze data."
- In other places, it means "imposing arbitrary assumptions onto the data."
- Four distinct meanings:
  1. Cowles definition just given.
  2. Low-dimensional model.
  3. Derived from theory.

Alternative Definitions of Structure

- "Structural" is a term used loosely.
- In some places, it means "doing economics to analyze data."
- In other places, it means "imposing arbitrary assumptions onto the data."
- Four distinct meanings:
  1. Cowles definition just given.
  2. Low-dimensional model.
  3. Derived from theory.
  4. Invariant to a class of modifications.

Alternative Definitions of Structure

- "Structural" is a term used loosely.
- In some places, it means "doing economics to analyze data."
- In other places, it means "imposing arbitrary assumptions onto the data."
- Four distinct meanings:
  1. Cowles definition just given.
  2. Low-dimensional model.
  3. Derived from theory.
  4. Invariant to a class of modifications.

(Koopmans, Marschak and Hurwicz)

- A basic definition of a system of structural relationships is that it is a system of equations invariant to a class of modifications or interventions. (Hurwicz (1962))

Structure as invariance to a class of modifications

- A basic definition of a system of structural relationships is that it is a system of equations invariant to a class of modifications or interventions. (Hurwicz (1962))
- In the context of policy analysis, this means a class of policy modifications.

Structure as invariance to a class of modifications

- A basic definition of a system of structural relationships is that it is a system of equations invariant to a class of modifications or interventions. (Hurwicz (1962))
- In the context of policy analysis, this means a class of policy modifications.
- Implicit in Koopmans (1950) and Marschak (1953) and it is explicitly utilized by Sims (1977), Lucas and Sargent (1981) and Leamer (1985).

- A basic definition of a system of structural relationships is that it is a system of equations invariant to a class of modifications or interventions. (Hurwicz (1962))
- In the context of policy analysis, this means a class of policy modifications.
- Implicit in Koopmans (1950) and Marschak (1953) and it is explicitly utilized by Sims (1977), Lucas and Sargent (1981) and Leamer (1985).
- This definition requires a precise definition of a policy, a class of policy modifications and specification of a mechanism through which policy operates.

- A basic definition of a system of structural relationships is that it is a system of equations invariant to a class of modifications or interventions. (Hurwicz (1962))
- In the context of policy analysis, this means a class of policy modifications.
- Implicit in Koopmans (1950) and Marschak (1953) and it is explicitly utilized by Sims (1977), Lucas and Sargent (1981) and Leamer (1985).
- This definition requires a precise definition of a policy, a class of policy modifications and specification of a mechanism through which policy operates.
- Treatment effects can be structural for certain classes of modifications.

Marschak's maxim

- To reconcile the econometric and treatment effect literatures, go back to a neglected but important paper by Marschak (1953) and taught in his 1949 lectures at Chicago in the Cowles Commission.

- To reconcile the econometric and treatment effect literatures, go back to a neglected but important paper by Marschak (1953) and taught in his 1949 lectures at Chicago in the Cowles Commission.
- Marschak noted that for many specific questions of policy analysis, it is not necessary to identify fully specified economic models that are invariant to classes of policy modifications.

Marschak's maxim

- To reconcile the econometric and treatment effect literatures, go back to a neglected but important paper by Marschak (1953) and taught in his 1949 lectures at Chicago in the Cowles Commission.
- Marschak noted that for many specific questions of policy analysis, it is not necessary to identify fully specified economic models that are invariant to classes of policy modifications.
- Implicit was his use of what we would now call decision theory.

- All that may be required for certain policy analyses are combinations of subsets of the structural parameters, corresponding to the parameters required to forecast particular policy modifications, which are often much easier to identify (i.e., require fewer and weaker assumptions).

- All that may be required for certain policy analyses are combinations of subsets of the structural parameters, corresponding to the parameters required to forecast particular policy modifications, which are often much easier to identify (i.e., require fewer and weaker assumptions).
- Forecasting or evaluating policies may only require partial knowledge of the full simultaneous equations system.

- All that may be required for certain policy analyses are combinations of subsets of the structural parameters, corresponding to the parameters required to forecast particular policy modifications, which are often much easier to identify (i.e., require fewer and weaker assumptions).
- Forecasting or evaluating policies may only require partial knowledge of the full simultaneous equations system.
- This principle called **Marschak's maxim** in honor of this insight.

Marschak's maxim

- The modern statistical treatment effect literature implements Marschak's maxim where the policies analyzed are the treatments available under a particular policy regime $p \in \mathcal{P}$.

- The modern statistical treatment effect literature implements Marschak's maxim where the policies analyzed are the treatments available under a particular policy regime $p \in \mathcal{P}$.
- The goal of policy analysis under this approach is typically restricted to evaluating policies in place and not in forecasting the effects of new policies or the effects of old policies on new environments.

- What is often missing from the literature on treatment effects is a clear discussion of the economic question being addressed by the treatment effect being estimated.

Marschak's maxim

- What is often missing from the literature on treatment effects is a clear discussion of the economic question being addressed by the treatment effect being estimated.
- This is the unstated and hence the unanswered question in the literature.

- What is often missing from the literature on treatment effects is a clear discussion of the economic question being addressed by the treatment effect being estimated.
- This is the unstated and hence the unanswered question in the literature.
- When the treatment effect literature does not clearly specify the economic question being addressed, it does not implement Marschak's maxim.

- Population mean treatment parameters are often identified under weaker conditions than are traditionally assumed in structural econometric analysis.

- Population mean treatment parameters are often identified under weaker conditions than are traditionally assumed in structural econometric analysis.

- Thus to identify the average treatment effect for s and s' we only require

$$E\left(Y\left(s,\omega\right) \mid S=s, X=x\right) - E\left(Y\left(s',\omega\right) \mid S=s', X=x\right).$$

- Population mean treatment parameters are often identified under weaker conditions than are traditionally assumed in structural econometric analysis.
- Thus to identify the average treatment effect for s and s' we only require

$$E\left(Y\left(s,\omega\right)\mid S=s,X=x\right)-E\left(Y\left(s',\omega\right)\mid S=s',X=x\right).$$

- Do not need exogeneity of X.

- Population mean treatment parameters are often identified under weaker conditions than are traditionally assumed in structural econometric analysis.
- Thus to identify the average treatment effect for s and s' we only require

$$E\left(Y\left(s,\omega\right)\mid S=s,X=x\right)-E\left(Y\left(s',\omega\right)\mid S=s',X=x\right).$$

- Do not need exogeneity of X.
- Under (PI-1) and (PI-2), this parameter answers the policy question of determining the average effect on outcomes of moving a person from s' to s.

- Population mean treatment parameters are often identified under weaker conditions than are traditionally assumed in structural econometric analysis.

- Thus to identify the average treatment effect for s and s' we only require

$$E\left(Y\left(s,\omega\right)\mid S=s,X=x\right)-E\left(Y\left(s',\omega\right)\mid S=s',X=x\right).$$

- Do not need exogeneity of X.
- Under (PI-1) and (PI-2), this parameter answers the policy question of determining the average effect on outcomes of moving a person from s' to s.

- The parameter is not designed to evaluate a whole host of other policies.

- Viewed in this light, the treatment effect literature that compares the outcome associated with $s \in \mathcal{S}$ with the outcome associated with $s' \in \mathcal{S}$ seeks to recover a causal effect of s relative to s'.

- Viewed in this light, the treatment effect literature that compares the outcome associated with $s \in \mathcal{S}$ with the outcome associated with $s' \in \mathcal{S}$ seeks to recover a causal effect of s relative to s'.
- It is a particular causal effect for a particular set of policy interventions.

- Viewed in this light, the treatment effect literature that compares the outcome associated with $s \in \mathcal{S}$ with the outcome associated with $s' \in \mathcal{S}$ seeks to recover a causal effect of s relative to s'.
- It is a particular causal effect for a particular set of policy interventions.
- It is structural for this intervention.

Marschak's maxim

- Viewed in this light, the treatment effect literature that compares the outcome associated with $s \in \mathcal{S}$ with the outcome associated with $s' \in \mathcal{S}$ seeks to recover a causal effect of s relative to s'.
- It is a particular causal effect for a particular set of policy interventions.
- It is structural for this intervention.
- Marschak's maxim urges analysts to formulate the problem being addressed clearly and to use the minimal ingredients required to solve it.

Marschak's maxim

- The treatment effect literature addresses the problem of comparing treatments $s \in \mathcal{S}$ under policy regime $p \in \mathcal{P}$, for a particular environment.

- The treatment effect literature addresses the problem of comparing treatments $s \in \mathcal{S}$ under policy regime $p \in \mathcal{P}$, for a particular environment.
- As analysts ask more difficult questions, it is necessary to specify more features of the models being used to address the questions.

- The treatment effect literature addresses the problem of comparing treatments $s \in \mathcal{S}$ under policy regime $p \in \mathcal{P}$, for a particular environment.
- As analysts ask more difficult questions, it is necessary to specify more features of the models being used to address the questions.
- Marschak's maxim is an application of Occam's Razor to policy evaluation.

- For certain classes of policy interventions designed to answer problem (P-1), the treatment effect approached may be very powerful and more convincing than explicitly economically formulated models because they entail fewer assumptions.

- For certain classes of policy interventions designed to answer problem (P-1), the treatment effect approached may be very powerful and more convincing than explicitly economically formulated models because they entail fewer assumptions.
- However, considerable progress has been made in relaxing the parametric structure assumed in the early explicitly economic models.

Marschak's maxim

- As the treatment effect literature is extended to address the more general set of policy forecasting problems entertained in the explicitly economic literature, the distinction between the two approaches will vanish.

- As the treatment effect literature is extended to address the more general set of policy forecasting problems entertained in the explicitly economic literature, the distinction between the two approaches will vanish.
- To make these methods empirically operational, we need to investigate the identification problem.

- As the treatment effect literature is extended to address the more general set of policy forecasting problems entertained in the explicitly economic literature, the distinction between the two approaches will vanish.
- To make these methods empirically operational, we need to investigate the identification problem.
- This is task 2 in table 1.

Table 1: Three distinct tasks arising in the analysis of causal models

Task	Description	Requirements
1	Defining the Set of Hypotheticals or Counterfactuals	A Scientific Theory
2	Identifying Parameters (Causal or Otherwise) from Hypothetical Population Data	Mathematical Analysis of Point or Set Identification
3	Identifying Parameters from Data	Estimation and Testing Theory

Identification problems: determining models from data

- Consider model space M.

Identification problems: determining models from data

- Consider model space M.
- This is the set of admissible models that are produced by some theory for generating counterfactuals.

Identification problems: determining models from data

- Consider model space M.
- This is the set of admissible models that are produced by some theory for generating counterfactuals.
- Elements $m \in M$ are admissible theoretical models.

Identification problems: determining models from data

- Consider model space M.
- This is the set of admissible models that are produced by some theory for generating counterfactuals.
- Elements $m \in M$ are admissible theoretical models.
- Map $g : M \to T$ maps an element $m \in M$ into an element $t \in T$.

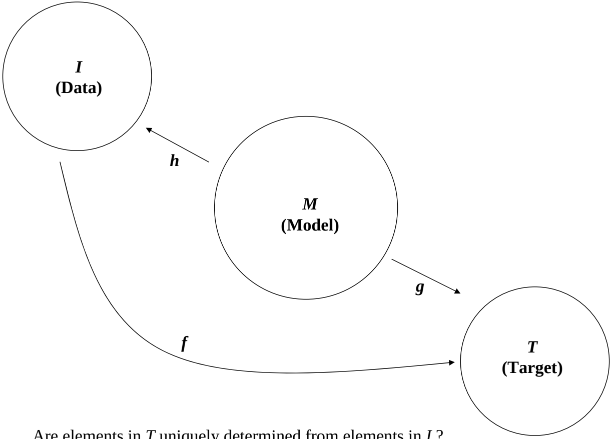
Identification problems: determining models from data

- Consider model space M.
- This is the set of admissible models that are produced by some theory for generating counterfactuals.
- Elements $m \in M$ are admissible theoretical models.
- Map $g : M \rightarrow T$ maps an element $m \in M$ into an element $t \in T$.
- Let the class of possible information or data be $\mathcal{I}$.

Identification problems: determining models from data

- Consider model space M.
- This is the set of admissible models that are produced by some theory for generating counterfactuals.
- Elements $m \in M$ are admissible theoretical models.
- Map $g : M \to T$ maps an element $m \in M$ into an element $t \in T$.
- Let the class of possible information or data be $\mathcal{I}$.
- Define a map $h : M \to i \in \mathcal{I}$.

Figure 4: Schematic of model (M), data ($\mathcal{I}$), and target (T) parameter spaces



Are elements in T uniquely determined from elements in I ?
Sometimes $T = M$. Usually T consists of elements derived from M.

- Let $M_h(i)$ be the set of models consistent with i.

- Let $M_h(i)$ be the set of models consistent with i.
- $M_h(i) = h^{-1}(\{i\}) = \{m \in M : h(m) = i\}$. The data i reject the other model $M \backslash M_h(i)$.

- Let $M_h(i)$ be the set of models consistent with i.
- $M_h(i) = h^{-1}(\{i\}) = \{m \in M : h(m) = i\}$. The data i reject the other model $M \backslash M_h(i)$.
- By placing restrictions on models, we can sometimes reduce the number of elements in $M_h(i)$ i fit has multiple members. $RE \subset M$

- Let $M_h(i)$ be the set of models consistent with i.
- $M_h(i) = h^{-1}(\{i\}) = \{m \in M : h(m) = i\}$. The data i reject the other model $M \setminus M_h(i)$.
- By placing restrictions on models, we can sometimes reduce the number of elements in $M_h(i)$ i fit has multiple members. $RE \subset M$
- Going after a more limited class of objects such as features of a model ($t \in T$) rather than the full model ($m \in M$).

- Let $M_h(i)$ be the set of models consistent with i.
- $M_h(i) = h^{-1}(\{i\}) = \{m \in M : h(m) = i\}$. The data i reject the other model $M \backslash M_h(i)$.
- By placing restrictions on models, we can sometimes reduce the number of elements in $M_h(i)$ i fit has multiple members. $RE \subset M$
- Going after a more limited class of objects such as features of a model ($t \in T$) rather than the full model ($m \in M$).
- Let $M_g(t) = g^{-1}(\{t\}) = \{m \in M : g(m) = t\}$

- Let $M_h(i)$ be the set of models consistent with i.
- $M_h(i) = h^{-1}(\{i\}) = \{m \in M : h(m) = i\}$. The data i reject the other model $M \backslash M_h(i)$.
- By placing restrictions on models, we can sometimes reduce the number of elements in $M_h(i)$ i fit has multiple members. $RE \subset M$
- Going after a more limited class of objects such as features of a model ($t \in T$) rather than the full model ($m \in M$).
- Let $M_g(t) = g^{-1}(\{t\}) = \{m \in M : g(m) = t\}$
- $f : \mathcal{I} \rightarrow T$ with the property $f \circ h = g$ are (a) h must map M onto $\mathcal{I}$ and (b) for all $i \in \mathcal{I}$, there exists $t \in T$ such that $M_h(i) \subseteq M_g(t)$.

Figure 5A: identified model

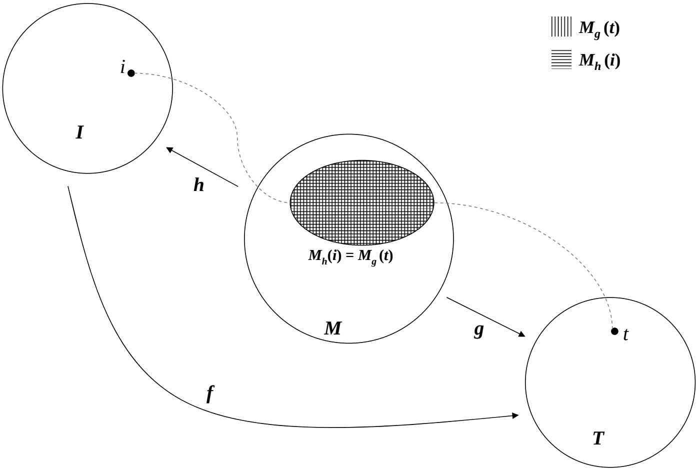


Figure 5B: nonidentified model

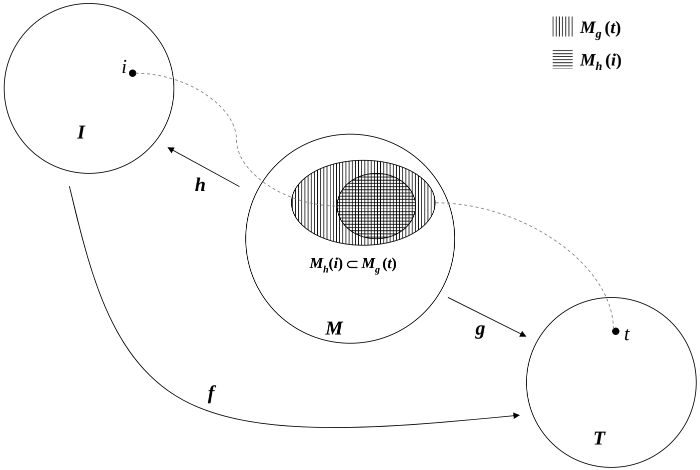


Table 3: sources of identification problems

(i) Absence of data on $Y(s', \omega)$ for $s' \in S \setminus \{s\}$ where s is the state selected (the evaluation problem).

(ii) Nonrandom selection of observations on states (the selection problem).

(iii) Support conditions may fail (outcome distributions for $F(Y_s \mid X = x)$ may be defined on only a limited support of X so $F(X \mid D_s = 1)$ and $F(X \mid D_{s'} = 1)$ have different X supports or limited overlap in their supports).

(iv) Functional forms of outcome equations and distributions of unobservables may be unknown. To extend some function $Y = G(X)$ to a new support requires functional structure: Cannot be extended outside of sample support by a purely nonparametric procedure.

(v) Determining the (X, Z, W) conditioning variables.

(vi) Different information sets for the agent making selection $\mathcal{I}_a$ and the econometrician trying to identify the model $\mathcal{I}_e$ where $\mathcal{I}_a \neq \mathcal{I}_e$.

- Thus recall our discussion of ATE. It is not necessary to assume that X is exogenous if one conditions policy analysis on X and does not seek to identify the effect of changing X.

- Thus recall our discussion of ATE. It is not necessary to assume that X is exogenous if one conditions policy analysis on X and does not seek to identify the effect of changing X.

- ATE answers only one of the many evaluation questions that are potentially interesting to answer. But we can identify ATE under weaker assumptions than are required to identify the full generalized Roy model.

- Thus recall our discussion of ATE. It is not necessary to assume that X is exogenous if one conditions policy analysis on X and does not seek to identify the effect of changing X.

- ATE answers only one of the many evaluation questions that are potentially interesting to answer. But we can identify ATE under weaker assumptions than are required to identify the full generalized Roy model.

- This is an application of Marschak's maxim.

- Thus recall our discussion of ATE. It is not necessary to assume that X is exogenous if one conditions policy analysis on X and does not seek to identify the effect of changing X.
- ATE answers only one of the many evaluation questions that are potentially interesting to answer. But we can identify ATE under weaker assumptions than are required to identify the full generalized Roy model.
- This is an application of Marschak's maxim.
- But if you focus on one parameter, you should justify why it is interesting.

Two paths toward relaxing distributional, functional form and exogeneity assumptions

- The strong exogeneity, linearity and normality assumptions in the conventional literature in econometrics used to form treatment effects and to evaluate policy are not required.

Two paths toward relaxing distributional, functional form and exogeneity assumptions

- The strong exogeneity, linearity and normality assumptions in the conventional literature in econometrics used to form treatment effects and to evaluate policy are not required.
- The literature in microeconometric structural estimation focuses on relaxing the linearity, separability, normality and exogeneity conditions invoked in the early literature in order to identify parameters under much weaker conditions.

Two paths toward relaxing distributional, functional form and exogeneity assumptions

- The recent literature on treatment effects identifies population level treatment effects under weaker conditions than are invoked in the traditional normal model.

- The recent literature on treatment effects identifies population level treatment effects under weaker conditions than are invoked in the traditional normal model.
- These variations in treatments are taken as the invariant structural parameters.

Two paths toward relaxing distributional, functional form and exogeneity assumptions

- The recent literature on treatment effects identifies population level treatment effects under weaker conditions than are invoked in the traditional normal model.
- These variations in treatments are taken as the invariant structural parameters.
- The class of modifications considered is the set of treatments in place.

Summary

- The vision of the Cowles Commission of using theory to guide measurement and conduct policy analysis is alive and well, and relevant to the modern policy evaluation literature.

Summary

- The vision of the Cowles Commission of using theory to guide measurement and conduct policy analysis is alive and well, and relevant to the modern policy evaluation literature.
- The assumptions of the founding fathers have been relaxed in many ways and its methods extended, but the vision is still relevant.

Summary

- The vision of the Cowles Commission of using theory to guide measurement and conduct policy analysis is alive and well, and relevant to the modern policy evaluation literature.
- The assumptions of the founding fathers have been relaxed in many ways and its methods extended, but the vision is still relevant.
- Econometrics is far ahead of statistics in the area of developing principles for constructing counterfactuals and performing causal inference.

Summary

- The recent treatment effect literature can be interpreted, under one view, as implementing the Marschak maxim.

Summary

- The recent treatment effect literature can be interpreted, under one view, as implementing the Marschak maxim.
- However, it often misses important features of the economic policy evaluation problem and puts estimators before the economics.

Summary

- The recent treatment effect literature can be interpreted, under one view, as implementing the Marschak maxim.
- However, it often misses important features of the economic policy evaluation problem and puts estimators before the economics.
- It often uses an estimator to define the parameter of interest.

Summary

- The recent treatment effect literature can be interpreted, under one view, as implementing the Marschak maxim.
- However, it often misses important features of the economic policy evaluation problem and puts estimators before the economics.
- It often uses an estimator to define the parameter of interest.
- The goal of these lectures is to unite these literatures in order to build an economic research tool effective for policy evaluation.

Summary

- The recent treatment effect literature can be interpreted, under one view, as implementing the Marschak maxim.
- However, it often misses important features of the economic policy evaluation problem and puts estimators before the economics.
- It often uses an estimator to define the parameter of interest.
- The goal of these lectures is to unite these literatures in order to build an economic research tool effective for policy evaluation.
- My next two lectures demonstrate one way to do this.