Econometric Causality

(2008) *International Statistical Review*, 76(1):1-27

James J. Heckman
University of Chicago

Econ 312, Spring 2023

THE UNIVERSITY OF CHICAGO

Introduction

- The econometric approach develops explicit models of outcomes where the causes of effects are investigated and the mechanisms governing the choice of treatment are analyzed.
- The relationship between treatment outcomes and treatment choice mechanisms is studied.
- A careful accounting of the unobservables in outcome and treatment choice equations facilitates the design of estimators.

- Both objective and subjective evaluations are considered, where subjective valuations are those of the person receiving treatment as well as the persons assigning it.
- Differences between anticipated and realized objective and subjective outcomes are analyzed.
- Models for simultaneous treatment effects are developed.
- A careful distinction is made between models for potential outcomes and empirical methods for identifying treatment effects.

The Econometric Approach

- Counterfactuals are possible outcomes in different hypothetical states of the world.
- An example would be the health outcomes for a person associated with taking or not taking a drug.
- Causal comparisons entail contrasts between outcomes in possible states defined so that only the presence or absence of the drug varies across the states.
- The person receiving the drug is the same as the person who does not, except for treatment status and, possibly, the outcome associated with treatment status.

- The problem of causal inference is to assess whether manipulation of the treatment, holding all other factors constant, affects outcomes.
- The concept of causality developed in this paper and in the statistical treatment effect literature is based on the notion of controlled variation—variation in treatment holding other factors constant.
- It is distinct from other notions of causality based on prediction (e.g., Granger, 1969, and Sims, 1972). Holland (1986) makes useful distinctions among commonly invoked definitions of causality.
- Cartwright (2004) discusses a variety of definitions of causality from a philosopher's perspective.

- The econometric approach to causal inference carefully distinguishes three problems: (a) Defining counterfactuals, (b) Identifying causal models from idealized data of population distributions (infinite samples without any sampling variation), and (c) Identifying causal models from actual data, where sampling variability is an issue.
- The contrast between (b) and (c) arises from the difference between empirical distributions based on sampled data and population distributions generating the data.
- Table 1 delineates the three distinct problems.

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 Causal Parameters from Hypothetical Population Data	Mathematical Analysis of Point or Set Identification
3	Identifying Parameters from Real Data	Estimation and Testing Theory

- The first problem entails the application of science, logic and imagination.
- It is also partly a matter of convention.
- A model of counterfactuals is more widely accepted the more widely accepted are its ingredients, which are the rules used to derive a model, including whether or not the rules of logic and mathematics are followed, and its agreement with established theories.

- Models are descriptions of hypothetical worlds obtained by varying — hypothetically — the factors determining outcomes.
- Models are not empirical statements or descriptions of actual worlds.
- However, they are often used to make predictions about actual worlds and they are often abstract representations of empirical descriptions.

- The second problem (b) is one of inference in very large samples.
- Can one recover counterfactuals (or means or distributions of counterfactuals) from data that are free of any sampling variation?
- This is the identification problem.

- The third problem is one of inference in practice.
- Can one recover a given model or a desired counterfactual from a given set of data?
- Solutions to this problem entail issues of inference and testing in real world samples.
- This is the problem most familiar to statisticians and empirical social scientists.
- The boundary between problems (b) and (c) is permeable depending on how "the data" are defined.

- Some of the controversy surrounding construction of counterfactuals and causal models is partly a consequence of analysts being unclear about these three distinct problems and often confusing them.
- Particular methods of estimation (e.g., matching or instrumental variable estimation) have become associated with "causal inference" in some circles, and even the definition of certain "causal parameters", because issues of definition, identification and estimation have sometimes been conflated.

- The econometric approach to policy evaluation separates these problems and emphasizes the provisional nature of causal knowledge.
- Some statisticians reject the notion of the provisional nature of causal knowledge and seek an assumption-free approach to causal inference (see, e.g., Tukey, 1986). However, human knowledge advances by developing theoretical models and testing them against data.
- The models used are inevitably provisional and depend on *a priori* assumptions.
- Even randomization, properly executed, cannot answer all of the relevant causal questions.

- Many "causal models" in statistics are incomplete guides to interpreting data or for suggesting answers to particular policy questions.
- They are motivated by the experiment as an ideal.
- They do not clearly specify the mechanisms determining how hypothetical counterfactuals are realized or how hypothetical interventions are implemented except to compare "randomized" with "nonrandomized" interventions.
- They focus only on outcomes, leaving the model for selecting outcomes only implicitly specified.
- The construction of counterfactual outcomes is based on appeals to intuition and not on formal models.

- Because the mechanisms determining outcome selection are not modeled in the statistical approach, the metaphor of "random assignment" is often adopted.
- This emphasis on randomization or its surrogates, like matching or instrumental variables, rules out a variety of alternative channels of identification of counterfactuals from population or sample data.

- The focus on randomization has practical consequences because of the conflation of Task 1 with Tasks 2 and three in Table 1. Since randomization is used to define the parameters of interest, this practice sometimes leads to the confusion that randomization is the only way — or at least the best way — to identify causal parameters from real data.
- Extreme versions of this approach deny causal status to any intervention that cannot in principle be implemented by a practical, real world experiment.

- One reason why many statistical models are incomplete is that they do not specify the sources of randomness generating variability among agents, i.e., they do not specify why otherwise observationally identical people make different choices and have different outcomes given the same choice.
- They do not distinguish what is in the agent's information set from what is in the observing statistician's information set, although the distinction is fundamental in justifying the properties of any estimator for solving selection and evaluation problems.

- They do not distinguish uncertainty from the point of view of the agent whose behavior is being analyzed from variability as analyzed by the observing analyst.
- They are also incomplete because they are recursive.
- They do not allow for simultaneity in choices of outcomes of treatment that are at the heart of game theory and models of social interactions and contagion (see, e.g., Brock and Durlauf, 2001; Tamer, 2003).

- Economists since Haavelmo (1943, 1944) have recognized the value of precise models for constructing counterfactuals, for answering "causal" questions and addressing more general policy evaluation questions.
- The econometric framework is explicit about how models of counterfactuals are generated, the sources of the interventions (the rules of assigning "treatment"), and the sources of unobservables in treatment allocations and outcomes and their relationship.
- Rather than leaving the rules governing selection of treatment implicit, the econometric approach uses explicit relationships between the unobservables in outcome and selection mechanisms to identify causal models from data and to clarify the nature of identifying assumptions.

- The goal of the econometric literature, like the goal of all science, is to understand the causes producing effects so that one can use empirical versions of the models to forecast the effects of interventions never previously experienced, to calculate a variety of policy counterfactuals and to use scientific theory to guide the choices of estimators and the interpretation of the evidence.
- These activities require development of a more elaborate theory than is envisioned in the current literature on causal inference in statistics.

- Many causal models in statistics are black-box devices designed to investigate the impact of "treatments" —often complex packages of interventions— on observed outcomes in a given environment.
- Unbundling the components of complex treatments is rarely done.
- Explicit scientific models go into the black box to explore the mechanism(s) producing the effects.
- In the terminology of Holland (1986), the distinction is between understanding the "effects of causes" (the goal of the treatment effect literature as a large group of statisticians define it) or understanding the "causes of effects" (the goal of the econometric literature building explicit models).

- By focusing on one narrow black-box question, the treatment effect literature avoids many of the problems confronted in the econometrics literature that builds explicit models of counterfactuals and assignment mechanisms.
- This is its great virtue.
- At the same time, it produces parameters that are more limited in application.
- Without further assumptions, these parameters do not lend themselves to extrapolation out of sample or to accurate forecasts of impacts of other policies besides the ones being empirically investigated.

- By not being explicit about the contents of the black-box (understanding the causes of effects), the treatment effect literature ties its hands in using information about basic behavioral parameters obtained from other studies as well as scientific intuition to supplement available information in the data in hand.
- It lacks the ability to provide explanations for estimated "effects" grounded in theory.
- When the components of treatments vary across studies, knowledge does not accumulate across treatment effect studies, whereas it does accumulate across studies estimating models generated from common parameters that are featured in the econometric approach.

Policy Evaluation Questions and Criteria of Interest

- This section presents three central policy evaluation questions.
- Individual level treatment effects are defined and the evaluation problem is discussed in general terms.

Three Policy Evaluation Problems Considered in This Paper

- Three broad classes of policy evaluation problems are considered in economics.
- Policy evaluation problem one is:

P1 *Evaluating the Impact of Historical Interventions on Outcomes Including Their Impact in Terms of the Well-Being of the Treated and Society at Large*

- By historical, I mean documented interventions.
- A variety of outcomes and criteria are used to form these evaluations depending on the question at hand.
- Economists distinguish objective or public outcomes that can in principle be measured by all external observers from "subjective" outcomes that are the evaluations of the agents experiencing treatment (e.g., patients) or the agents prescribing treatment (e.g., physicians). Objective outcomes are intrinsically *ex post* ("after the fact") in nature.
- The statistical literature on causal inference focuses exclusively on *ex post* objective outcomes.
- Subjective outcomes can be *ex ante* ("anticipated") or *ex post*. Thus the outcome of a medical trial produces both a cure rate and the pain and suffering of the patient.

- *Ex ante* anticipated pain and suffering may be different from *ex post* realized pain and suffering.
- Agents may also have *ex ante* evaluations of the objective outcomes that may differ from their *ex post* evaluations.
- By impact I mean constructing either individual level or population level counterfactuals and their valuations.
- By well-being, I mean the valuations of the outcomes obtained from the intervention of the agents being analyzed or some other party (e.g., the parents of the agent or "society" at large). They may be *ex ante* or *ex post*. P1 is the problem of *internal validity*. It is the problem of identifying a given treatment parameter or a set of treatment parameters in a given environment.

- Most policy evaluation is designed with an eye toward the future and toward informing decisions about new policies and application of old policies to new environments.
- It is helpful to distinguish a second problem encountered in policy analysis.

P2 *Forecasting the Impacts (Constructing Counterfactual States) of Interventions Implemented in one Environment in Other Environments, Including Their Impacts In Terms of Well-Being.*

- Included in these interventions are policies described by generic characteristics (e.g., tax or benefit rates or therapy used, including intensity) that are applied to different groups of people or in different time periods from those studied in implementations of the policies on which data are available.
- This is the problem of *external validity*: taking a treatment parameter or a set of parameters estimated in one environment to another environment (see, e.g., **?**). The environment includes the characteristics of individuals and of the treatments.

- Finally, the most ambitious problem is forecasting the effect of a new policy, never previously experienced.

P3 *Forecasting the Impacts of Interventions (Constructing Counterfactual States Associated with Interventions) Never Historically Experienced to Various Environments, Including Their Impacts in Terms of Well-Being.*

- This problem requires that one use past history to forecast the consequences of new policies.
- It is a fundamental problem in knowledge.
- P3 is a problem that economic policy analysts have to solve daily.
- I now present a framework within which analysts can address these problems in a systematic fashion.
- It is also a framework that can be used for causal inference.

Definition of Individual Level Treatment Effects

- To evaluate is to value and to compare values among possible outcomes.
- These are two distinct tasks.
- Define outcomes corresponding to state (policy, treatment) s for an agent ω as $Y(s, \omega)$, $\omega \in \Omega$.
- The agent can be a household, a patient, a firm, or a country.

- One can think of Ω as a universe of agents.
- Assume that $\Omega = [0, 1]$.
- $Y(\cdot, \cdot)$ may be vector valued, but to simplify the exposition, I work with scalar outcomes.
- (See Heckman and Vytlacil, 2007a, for an analysis with vector outcomes.)

- The $Y(s, \omega)$ are outcomes realized after treatments are chosen.
- In advance of treatment, agents may not know the $Y(s, \omega)$ but may make forecasts about them.
- These forecasts may influence their decisions to participate in the program or may influence the agents who make decisions about whether or not an individual participates in the program.
- Selection into the program based on actual or anticipated components of outcomes gives rise to the selection problem in the evaluation literature.

- Let $\mathcal{S}$ be the set of possible treatments with elements denoted by s.
- For simplicity of exposition, assume that this set is the same for all ω.
- For each ω, one obtains a collection of possible outcomes given by $\{Y(s,\omega)\}_{s\in\mathcal{S}}$.
- For simplicity, I assume that the set $\mathcal{S}$ is finite (Heckman and Vytlacil, 2007a, consider more general cases). For example, if $\mathcal{S}=\{0,1\}$, there are two treatments, one of which may be a no-treatment state (e.g., $Y(0,\omega)$ is the outcome for an agent ω not getting a treatment like a drug, schooling or access to a new technology, while $Y(1,\omega)$ is the outcome in treatment state 1 for agent ω getting the drug, schooling or access). A two treatment environment receives the most attention in the theoretical literature, but the multiple treatment environment is the one most frequently encountered in practice.

- Each "state" (treatment) may consist of a compound of subcomponent states.
- In this case, one can define s itself as a vector (e.g., $s = (s_1, s_2, \ldots, s_K)$ for K components) corresponding to the different components that comprise treatment.
- Thus a medical protocol typically consists of a package of treatments.
- One might be interested in the package of one (or more) of its components.
- Thus s_1 might be months of treatment with one drug, s_2 the quality of physicians, and so forth.
- No generality is lost by assuming that s is a scalar, since each distinct treatment can be given a distinct label.

- The outcomes may be time subscripted, with $Y_t(s, \omega)$ corresponding to outcomes of treatment measured at different times.
- The index set for t may be the integers, corresponding to discrete time, or an interval, corresponding to continuous time.
- In principle, one could index $\mathcal{S}$ by t, which may be defined on the integers, corresponding to discrete time, or an interval corresponding to continuous time.

- The $Y_t(s, \omega)$ are realized or *ex post* (after treatment) outcomes.
- When choosing treatment, these values may not be known.
- Gill and Robins (2001), Van der Laan and Robins (2003), Abbring and Van den Berg (2003) , Abbring and Heckman (2007); **?**, and Heckman and Navarro (2007) develop models for dynamic counterfactuals, where time-subscripted and ω-subscripted $\mathcal{S}$ arise as information accrues.
- Throughout this essay I keep the time subscript implicit.

- The **individual treatment effect** for agent ω comparing objective outcomes of treatment s with objective outcomes of treatment s' is

$$Y(s,\omega) - Y(s',\omega), \qquad s \neq s', \tag{1}$$

  for two elements $s, s' \in \mathcal{S}$.
- This is also called an **individual level causal effect**. The causal effect is the Marshallian (1890) *ceteris paribus* change of outcomes for an agent across states s and s'.
- Only s and s' are varied.

- Other comparisons are of interest in assessing a program.
- Economists are interested in the well-being of participants as well as the objective outcomes (see Heckman and Smith, 1998). Although statisticians often reason in terms of assignment mechanisms, economists recognize that agent preferences often govern actual choices.
- Comparisons across outcomes can be made in terms of utilities (personal, $R\left(Y\left(s,\omega\right),\omega\right)$, or in terms of planner preferences or physician preferences, R_G, or both types of comparisons might be made for the same outcome and their agreement or conflict evaluated). Utility functions produce subjective valuations of outcomes by the agents being treated or the planner.

- To simplify the notation, and at the same time allow for more general possibilities for arguments of the valuation function, write $R\left(Y\left(s,\omega\right),\omega\right)$ as $R(s,\omega)$, suppressing the explicit dependence of R on $Y\left(s,\omega\right)$.
- In this notation, one can ask if $R(s,\omega) > R(s',\omega)$, or not (is the agent better off as a result of treatment s compared to treatment s'?). The difference in subjective outcomes is $R(s,\omega) - R(s',\omega)$, and is a type of treatment effect.
- Holding ω fixed holds all features of the agent fixed except the treatment assigned, s.

- Since the units of utility, $R(s, \omega)$, are arbitrary, one could instead record for each s and ω an indicator if the outcome in s is greater or less than the outcome in s', i.e., $R(s, \omega) > R(s', \omega)$, or not.
- This is also a type of treatment effect.
- Agents making decisions about treatment may be only partially informed about realized payoffs at the time they make decisions.
- Modeling the distinction between anticipated and realized outcomes is an integral part of the econometric approach to causality and policy evaluation.
- A central feature of the econometric approach to program evaluation is the evaluation of subjective valuations as perceived by decision makers and not just objective valuations.

- The term "treatment" is used in multiple ways in various literatures.
- In its most common usage, a treatment assignment mechanism is a rule $\tau : \Omega \to \mathcal{S}$ which assigns treatment to each individual ω.
- The consequences of the assignment are the outcomes $Y(s, \omega)$, $s \in \mathcal{S}$, $\omega \in \Omega$.
- Denote the collection of possible assignment rules by $\mathcal{T}$ where $\tau \in \mathcal{T}$.
- The policy is very simple.
- The policy selects who gets what.
- More precisely, it selects individuals ω and specifies the treatment $s \in \mathcal{S}$ received.

A More Behavioral Approach to Treatment Assignment

- The econometric literature offers a more nuanced definition of treatment assignment that explicitly recognizes the element of choice by agent ω in producing the treatment assignment rule.
- Treatment can include participation in activities such as schooling, training, a medical therapy, adoption of a particular technology, and the like.
- Participation in treatment is often a choice made by agents.
- Modeling this choice process is a distinctive feature of the econometric approach.

- Under a more comprehensive definition of treatment, agents are assigned incentives like taxes, subsidies, endowments and eligibility that affect their choices, but the agent chooses the treatment selected.
- Agent preferences, program delivery systems, market structures, and the like might all affect the choice of treatment.
- The treatment choice mechanism may involve multiple actors and multiple decisions that result in an assignment of ω to s.
- For example, s can be schooling while $Y(s, \omega)$ is earnings given schooling for agent ω.
- A policy may be a set of payments that encourage schooling, as in the PROGRESA program in Mexico, and the treatment in that case is choice of schooling with its consequences for earnings.
- The s can also be a medical protocol that requires compliance by the patient (choice behavior) to be effective.

- The following description of treatment assignment recognizes individual choices and constraints and is more suitable to policy evaluation that recognizes the role of choice by agents.
- Specify assignment rules $a \in \mathcal{A}$ that map individuals ω into constraints (benefits) $b \in \mathcal{B}$ under different mechanisms.
- In this notation, a constraint assignment mechanism a is a map $a : \Omega \to \mathcal{B}$ defined over the space of agents.
- The constraints may include endowments, eligibility, taxes, subsidies, and other incentives that affect agent choices of treatment.
- Elements of b can be parameters of tax and benefit schedules that affect individual incentives.

- While a more general setup is possible, where ω-specific schedules are assigned to person ω, the cost of such generality is more complicated notation.
- For simplicity, I confine attention to a fixed — but possibly very large — set of parameters defined for all agents.
- The map a defines the rule used to assign $b \in \mathcal{B}$.
- It can include deterministic rules that give schedules mapping ω into $\mathcal{B}$, such as tax schedules or eligibility schedules.
- It can also include random assignment mechanisms that assign ω to an element of $\mathcal{B}$.
- Random assignment mechanisms add additional elements of randomness to the environment.
- Abusing notation, when randomization is used, redefine Ω to include this new source of randomness.

- Given $b \in \mathcal{B}$ allocated by constraint assignment mechanism $a \in \mathcal{A}$, agents pick treatments.
- Define treatment assignment mechanism $\tau : \Omega \times \mathcal{A} \times \mathcal{B} \to \mathcal{S}$ as a map taking agent $\omega \in \Omega$ facing constraints $b \in \mathcal{B}$ assigned by mechanism $a \in \mathcal{A}$ into a treatment $s \in \mathcal{S}$.
- In settings with choice, τ is the choice rule used by agents where $\tau \in \mathcal{T}$, a set of possible choice rules.

- It is conventional to assume a unique $\tau \in \mathcal{T}$ is selected by the relevant decision makers, although that is not required in this definition.
- A policy regime $p \in \mathcal{P}$ is a pair $(a, \tau) \in \mathcal{A} \times \mathcal{T}$ that maps agents denoted by ω into elements of s.
- In this notation, $\mathcal{P} = \mathcal{A} \times \mathcal{T}$.

- Incorporating choice into the analysis of treatment effects is an essential and distinctive ingredient of the econometric approach to the evaluation of social programs.
- The traditional treatment-control analysis in statistics equates mechanisms a and τ.
- An assignment in that literature is an assignment to treatment, not an assignment of incentives and eligibility for treatment with the agent making treatment choices.
- In this notation, the traditional approach has only one assignment mechanism and treats noncompliance with it as a problem rather than as a source of information on agent preferences, which is a central feature of the econometric approach (Heckman and Smith, 1998). Thus, under full compliance, $a : \Omega \rightarrow \mathcal{S}$ and $a = \tau$, where $\mathcal{B} = \mathcal{S}$.

- Policy invariance is a key assumption for any study of policy evaluation.
- It allows analysts to characterize outcomes without specifying how those outcomes are obtained.
- Policy invariance has two aspects.
- The first aspect is that, for a given $b \in \mathcal{B}$ (incentive schedule), the mechanism $a \in \mathcal{A}$ by which ω is assigned a b (e.g., random assignment, coercion at the point of a gun, etc.) and the incentive $b \in \mathcal{B}$ are assumed to be irrelevant for the values of realized outcomes for each s that is selected.
- (Note this may be confusing if b also a factor determining outcomes in s — but s already includes this effect.)

- Second, for a given s for agent ω, the mechanism τ by which s is assigned to the agent under assignment mechanism $a \in \mathcal{A}$ is irrelevant for the values assumed by realized outcomes.
- Both assumptions define what economists mean by policy invariance.
- Policy invariance was first defined and formalized by Marschak (1953) and Hurwicz (1962).

- Policy invariance allows one to describe outcomes by $Y(s, \omega)$ and ignore features of the policy and choice environment in defining outcomes.
- If one has to account for the effects of incentives and assignment mechanisms on outcomes, one must work with $Y(s, \omega, a, b, \tau)$ instead of $Y(s, \omega)$.
- The more complex description is the outcome associated with treatment state s for person ω, assigned incentive package b by mechanism a which are arguments of assignment rule τ.
- See Heckman and Vytlacil (2007a) for precise definitions of invariance.

- The invariance assumptions state that for the same treatment s and agent ω, different constraint assignment mechanisms a and a' and associated constraint state assignments b and b' produce the same outcome.
- For example, they rule out the possibility that the act of randomization or the act of pointing a gun at an agent to secure cooperation with planner intentions has an effect on outcomes, given that the agent ends up in s.
- This is a strong assumption.

- Another invariance assumption invoked in the literature is that for a fixed a and b, the outcomes are the same, independent of the treatment assignment mechanism.
- This assumption states that the actual mechanism used to assign treatment does not affect the outcomes.
- It rules out, among other things, social interactions, contagion and general equilibrium effects.
- Heckman (1992), Heckman and Smith (1998), Heckman et al. (1999) and Heckman and Vytlacil (2007b) discuss evidence against this assumption, and Heckman, Lochner, and Taber (1998a,b,c) show how to relax it.

- If treatment effects based on subjective evaluations are also considered, as is distinctive of the econometric approach, it is necessary to broaden invariance assumptions to produce invariance in rewards for certain policies and assignment mechanisms (see Heckman and Vytlacil, 2007a). The required invariance assumptions state, for example, that utilities are not affected by randomization or the mechanism of assignment of constraints.
- Heckman (1992), Heckman, LaLonde, and Smith (1999) and Heckman and Vytlacil (2007b) present evidence against this assumption.

- Another invariance assumption rules out social interactions in both subjective and objective outcomes.
- It is useful to distinguish invariance of objective outcomes from invariance of subjective outcomes.
- Randomization may affect subjective evaluations through its effect of adding uncertainty into the decision process but it may not affect objective valuations.
- The econometric approach models how assignment mechanisms and social interactions affect choice and outcome equations rather than postulating *a priori* that invariance postulates for outcomes are always satisfied for outcomes.

More General Criteria

- There are many comparisons the analyst might make (see, e.g., Heckman, Smith, and Clements, 1997). One might compare outcomes in different sets that are ordered.
- Define $\text{argmax}_s\{Y(s,\omega)\}_{s\in\mathcal{S}}$ as the value of s that produces the maximal $Y(s,\omega)$ for $s \in \mathcal{S}$.

- Thus if $Y(s,\omega)$ is scalar income and one compares outcomes for $s \in \mathcal{S}_A$ with outcomes for $s' \in \mathcal{S}_B$, where $\mathcal{S}_A \cap \mathcal{S}_B = \varnothing$, then one might compare Y_{s_A} to Y_{s_B}, where

$$s_A = \text{argmax}_{s \in \mathcal{S}_A}\{Y(s,\omega)\} \quad \text{and} \quad s_B = \text{argmax}_{s \in \mathcal{S}_B}\{Y(s,\omega)\},$$

where I suppress the dependence of s_A and s_B on ω.

- This compares the best in one choice set with the best in the other.
- Another contrast compares the best choice with the next best choice.
- To do so, define $s' = \text{argmax}_{s \in \mathcal{S}} \{Y(s, \omega)\}$ and $\mathcal{S}_B = \mathcal{S} \setminus \{s'\}$ and define the treatment effect as $Y_{s'} - Y_{s_B}$.

- This is the comparison of the highest outcome over $\mathcal{S}$ with the next best outcome.
- Many other individual-level comparisons might be constructed, and they may be computed using personal preferences, $R(\omega)$, using the preferences of the planner, R_G, or using the preferences of the planner over the preferences of agents.
- Heckman (2005) and Heckman and Vytlacil (2007a,b) present a comprehensive discussion of alternative criteria.

The Evaluation Problem

- In the absence of a theory, there are no well defined rules for constructing counterfactual or hypothetical states or constructing the rules for assignment to treatment.
- Lewis (1974) defines admissible counterfactual states without an articulated theory as "closest possible worlds". His definition founders on the lack of any meaningful metric or topology to measure "closeness" among possible worlds.

- Articulated scientific theories provide algorithms for generating the universe of internally consistent, theory-consistent counterfactual states.
- These hypothetical states are possible worlds.
- They are products of a purely mental activity.
- Different theories produce different $Y(s, \omega)$ and different assignment mechanisms.

- The **evaluation problem** is that the analyst observes each agent in one of $\bar{S}$ possible states.
- One does not know the outcome of the agent in other states that are not realized, and hence cannot directly form individual level treatment effects.
- The **selection problem** arises because one only observes certain agents in any state.
- Thus one observes $Y(s, \omega)$ only for agents who choose (or are chosen) to be in that state.
- In general, the outcomes of agents found in $S = s$ are not representative of what the outcomes of agents would be if they were randomly assigned to s.

- The evaluation problem is an identification problem that arises in constructing the counterfactual states and treatment assignment rules produced by these abstract models using data.
- This is the second problem presented in Table 1. This problem is not precisely stated until the data available to the analyst are precisely defined.
- Different areas of knowledge assume access to different types of data.

- For each policy regime, at any point in time one observes agent ω in some state but not in any of the other states.
- Thus one does not observe $Y(s', \omega)$ for agent ω if one observes $Y(s, \omega)$, $s \neq s'$.
- Let $D(s, \omega) = 1$ if one observes agent ω in state s under policy regime p, where I keep the policy regime p implicit to simplify the notation.
- In this notation, $D(s, \omega) = 1$ implies that $D(s', \omega) = 0$ for $s \neq s'$.

- $Y(s,\omega)$ is observed if $D(s,\omega)=1$ but not $Y(s',\omega)$, for $s \neq s'$.
- One can define observed $Y(\omega)$ as

$$Y(\omega)=\sum_{s\in\mathcal{S}} D(s,\omega)\, Y(s,\omega)\,. \tag{2}$$

Without further assumptions, constructing an empirical counterpart to the individual level causal effect (1) is impossible from the data on $(Y(\omega), D(\omega))$, $\omega \in \Omega$, where $D(\omega)=\{(D(s,\omega)\}_{s\in\mathcal{S}}$.

- This formulation of the evaluation problem is known as Quandt's switching regression model (Quandt, 1958) and is attributed in statistics to Neyman (1923), Cox (1958) and Rubin (1978). A version of it is formulated in a linear equations context for a continuum of treatments by Haavelmo (1943). The Roy model (1951) is another version of this framework with two possible treatment outcomes ($\mathcal{S} = \{0, 1\}$) and a scalar outcome measure and a particular assignment mechanism τ which is that $D(1, \omega) = \mathbf{1}[Y(1, \omega) > Y(0, \omega)]$, where $\mathbf{1}[\cdot]$ means $\mathbf{1}[\cdot] \geq 1$ if the argument "$\cdot$" is true and $= 0$ otherwise.
- Thus $\tau_p(\omega) = 1$ for ω satisfying $Y(1, \omega) \geq Y(0, \omega)$ and $\tau(\omega) = 0$ for ω satisfying $Y(1, \omega) < Y(0, \omega)$.

- The mechanism of selection depends on the potential outcomes.
- Agents choose the sector with the highest income so the actual selection mechanism is not a randomization.
- Versions of this model with more general self-selection mechanisms are surveyed in Heckman (1990), Heckman and Smith (1998), Heckman and Vytlacil (2007a,b), and Abbring and Heckman (2007).

- The evaluation literature in macroeconomics analyzes policies with universal coverage at a point in time (e.g., a tax policy or social security) so that $D(s, \omega) = 1$ for some s and all ω.
- It uses time series data to evaluate the impacts of policies in different periods and typically uses mean outcomes (or mean utilities) to evaluate policies.

- The problem of self selection is an essential aspect of the evaluation problem when data are generated by the choices of agents.
- The agents making choices may be different from the agents receiving treatment (e.g., parents making choices for children). Such choices can include compliance with the protocols of a social experiment as well as ordinary choices about outcomes that people make in everyday life.
- As a consequence of self-selection, the distribution of the $Y(s, \omega)$ observed are not the population distribution of randomly sampled $Y(s, \omega)$.

- In the prototypical Roy model, the choice of treatment (including the decisions not to attrite from the program) is informative on the relative evaluation of $Y(s, \omega)$.
- This point is more general and receives considerable emphasis in the econometrics literature (e.g., Heckman and Smith, 1998; Heckman and Vytlacil, 2007a). Choices by agents provide information on subjective evaluations which are of independent interest.

- The evaluation problem arises from the absence of information on outcomes for agent ω other than the outcome that is observed.
- Even a perfectly implemented social experiment does not solve this problem (Heckman, 1992). Randomization identifies only one component of $\{Y(s,\omega)\}_{s\in\mathcal{S}}$ for any agent.
- In addition, even with large samples and a valid randomization, some of the $s \in \mathcal{S}$ may not be observed if one is seeking to evaluate new policies never experienced.

- There are two main avenues of escape from this problem.
- The first avenue, featured in explicitly formulated econometric models, often called "structural econometric analysis," is to model $Y(s, \omega)$ 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.
- 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 $D(s, \omega)$.

- The Roy model explicitly models this dependence.
- See Heckman and Honoré (1990) and Heckman (2001) for a discussion of this model.
- Heckman (1990), Heckman and Smith (1998), Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2005) extend the Roy model.
- This approach stresses understanding the factors underlying outcomes and the choice of outcome equations and their dependence.

- Empirical models explicitly based on scientific theory pursue this avenue of investigation.
- Some statisticians call this the "scientific approach" and are surprisingly hostile to it.
- See Holland (1986).

- A second avenue of escape, and the one pursued in the recent treatment effect literature, redirects attention away from estimating the determinants of $Y(s, \omega)$ toward estimating some population version of (1), most often a mean, 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 ignored.
- The statistical treatment effect literature focuses exclusively on policy problem P1 for the subset of outcomes that is observed.
- It ignores the problem of forecasting a new policy in a new environment (problem P2), or a policy never previously experienced (problem P3). Forecasting the effects of new policies is a central task of science.

New Population Level Treatment Parameters Introduced in Economics

- Economists and statisticians often draw on the same set of population level treatment parameters such as the average treatment effect (ATE), $(E(Y(s) - Y(s')))$, treatment on the treated, (TT) $E((Y(s) - Y(s')) \mid D(s) = 1)$, and treatment on the untreated (TUT) $E((Y(s) - Y(s')) \mid D(s') = 1)$.
- In this subsection, we keep the "ω" implicit to simplify the notation.
- In some discussions in statistics, ATE is elevated to primacy as *the* causal parameter.
- Economists use different causal parameters for different policy problems.

- The distinction between the marginal and average return is a central concept in economics.
- It is also of interest in medicine where the effect of a treatment on the marginal patient is an important question.
- It is often of interest to evaluate the impact of marginal extensions (or contractions) of a program or treatment regime.
- Incremental cost-benefit analysis is conducted in terms of marginal gains and benefits.

- 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 defined with respect to personal preferences and with respect to choice-specific costs $C(j)$.
- Formally, making the dependence of the reward on $Y(s), C(s)$ explicit, i.e., writing $R(Y(s), C(s))$ as the reward in state s,

$$\text{EOTM}^R(j,k) = E\left(Y(j) - Y(k) \,\middle|\, \begin{array}{l} R(Y(j), C(j)) = R(Y(k), C(k)); \\ \left.\begin{array}{l} R(Y(j), C(j)) \\ R(Y(k), C(k)) \end{array}\right\} \geq R(Y(l), C(l)) \\ \qquad\qquad\qquad\qquad\qquad\qquad l \neq j, k \end{array}\right) \quad (3)$$

This is the mean gain to agents indifferent between treatments j and k, given that these are the best two options available.

- In a parallel fashion, one can define $\text{EOTM}^{R_G}\left(Y(j) - Y(k)\right)$ using the preferences of another agent (e.g., the parent of a child, a paternalistic bureaucrat, etc.). This could be a subjective evaluation made by a physician, for example.
- Analogous parameters can be defined for mean setwise comparisons (see Heckman and Vytlacil, 2005, 2007a,b). A generalization of this parameter called the *marginal treatment effect* (MTE), introduced into the evaluation literature by Björklund and Moffitt (1987) and further developed in Heckman and Vytlacil (1999, 2005, 2007b) and Heckman, Urzúa, and Vytlacil (2006), plays a central role in organizing and interpreting a wide variety of econometric and statistical estimators as weighted averages of marginal treatment effects.

- Many other mean treatment parameters can be defined depending on the choice of the conditioning set.
- Analogous definitions can be given for median and other quantile versions of these parameters (see Heckman, Smith, and Clements, 1997; Abadie, Angrist, and Imbens, 2002). Although means are conventional, distributions of treatment parameters are also of considerable interest.
- I discuss distributional parameters in the next subsection.

- Of special interest in policy analysis is the **policy relevant treatment effect** (PRTE). It is the effect on aggregate outcomes of one policy regime $p \in \mathcal{P}$ compared to the effect of another policy regime.
- Under invariance assumptions, suppressing the ω, defining

$$Y_p = \sum_{s \in \mathcal{S}} D_p(s) Y(s), \qquad Y_{p'} = \sum_{s \in \mathcal{S}} D_{p'}(s) Y(s),$$

$$\text{PRTE:} \qquad E\left(Y_p\right) - E\left(Y_{p'}\right), \qquad \text{where } p, p' \in \mathcal{P},$$

where the expectations are taken over different spaces of policy assignment rules.

- Notice that you do not need invariance so you can get by with more general results.

$$Y_p = \sum D_p(s) Y_p(s) \qquad Y_{p'} = \sum D_{p'}(s) Y_{p'}(s)$$

- Carneiro, Heckman, and Vytlacil (2006) show how to identify this parameter.
- Mean treatment effects play a special role in the statistical approach to causality.
- They are the centerpiece of the Holland (1986)–Rubin (1978) model and in many other studies in statistics and epidemiology.
- Social experiments with full compliance and no disruption can identify these means because of a special mathematical property of means.

- If one can identify the mean of $Y(j)$ and the mean of $Y(k)$ from an experiment where j is the treatment and k is the baseline, one can form the average treatment effect for j compared to k.
- These can be formed over two different groups of agents.
- By a similar argument, TT or TUT can be formed by randomizing over particular subsets of the population (those who would select treatment and those who would not select treatment respectively), assuming full compliance and no bias arising from the randomization.
- See Heckman (1992), Heckman and Vytlacil (2007b) and Abbring and Heckman (2007).

- The case for randomization is weaker if the analyst is interested in other summary measures of the distribution or the distribution itself.
- In general, randomization is not an effective procedure for identifying median gains, or the distribution of gains or many other key parameters.
- The elevation of population means as the primary "causal" parameters promotes randomization as an ideal estimation method.

Distributions of Counterfactuals

- Although means are traditional, the answers to many interesting evaluation questions require knowledge of features of the distribution of program gains other than some mean.
- Thus it is of interest to know if some fraction of the population benefits from a treatment even if on average there is zero benefit or a negative mean outcome.

- Let s_p be shorthand notation for assignment of ω to outcome s under policy p and the associated set of treatment assignment mechanisms.
- For any two regimes p and p', the proportion who benefit is $\Pr\left(Y\left(s_p(\omega),\omega\right) \geq Y\left(s_{p'}(\omega),\omega\right)\right)$.
- This is called the **voting criterion** (Heckman, Smith, and Clements, 1997). It requires knowledge of the joint distribution of the two arguments in the inequality.
- Experiments, without further assumptions, can only identify marginal distributions and not the joint distributions required to identify the voting criterion (Heckman, 1992).

- 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 $\Pr\left(Y\left(j,\omega\right) \geq Y\left(k,\omega\right)\right)$.
- One might be interested in the quantiles of $Y\left(s_p(\omega),\omega\right) - Y\left(s_{p'}(\omega),\omega\right)$ or of $Y\left(j,\omega\right) - Y\left(k,\omega\right)$ for $s_p(\omega) = j$ and $s_p(\omega) = k$ or the percentage who gain from participating in j (compared to k) under policy p.
- More comprehensive analyses would include costs and benefits.
- Distributional criteria are especially salient if program benefits are not transferrable or if restrictions on feasible social redistributions prevent distributional objectives from being attained.
- Abbring and Heckman (2007) present a comprehensive survey of approaches to identifying joint distributions of counterfactual outcomes.

Accounting for Uncertainty

- Systematically accounting for uncertainty introduces additional considerations that are central to economic analysis but that are largely ignored in the statistical treatment effect literature as currently formulated.
- Persons do not know the outcomes associated with possible states not yet experienced.
- If some potential outcomes are not known at the time treatment decisions are made, the best that agents can do is to forecast them with some rule.

- Even if, *ex post*, agents know their outcome in a benchmark state, they may not know it *ex ante*, and they may always be uncertain about what they would have experienced in an alternative state.
- This creates a further distinction: that between *ex post* and *ex ante* evaluations of both subjective and objective outcomes.
- The economically motivated literature on policy evaluation makes this distinction.
- The statistical treatment effect literature does not.

- Because agents typically do not possess perfect information, a simple voting criterion that assumes perfect foresight over policy outcomes may not accurately predict choices and requires modification.
- Let $\mathcal{I}_\omega$ denote the information set available to agent ω.
- He or she evaluates policy j against k using that information.

- Under an expected utility criterion, agent ω prefers policy j over policy k if

$$E(R(Y(j,\omega),\omega) \mid \mathcal{I}_\omega) \geq E(R(Y(k,\omega),\omega) \mid \mathcal{I}_\omega).$$

The proportion of people who prefer j is

$$PB\left(j \mid j,k\right) = \int \mathbf{1}\left(E[R\left(Y(j,\omega),\omega\right) \mid \mathcal{I}_\omega]\right) \geq E\left(R\left(Y(k,\omega),\omega\right) \mid \mathcal{I}_\omega\right) d\mu\left(\mathcal{I}_\omega\right), \quad (4)$$

where $\mu(\omega)$ is the distribution of ω in the population whose preferences over outcomes are being studied.

- The voting criterion presented in the previous section is the special case where the information set $\mathcal{I}_\omega$ contains $(Y(j,\omega), Y(k,\omega))$, so there is no uncertainty about $Y(j)$ and $Y(k)$.
- Cunha, Heckman, and Navarro (2005), Cunha, Heckman, and Navarro (2006) and Abbring and Heckman (2007) offer examples of the application of this criterion.
- See Cunha, Heckman, and Navarro (2005, 2006) for computations regarding both types of joint distributions.

- Accounting for uncertainty in the analysis makes it essential to distinguish between *ex ante* and *ex post* evaluations.
- *Ex post*, part of the uncertainty about policy outcomes is resolved although agents do not, in general, have full information about what their potential outcomes would have been in policy regimes they have not experienced and may have only incomplete information about the policy they have experienced (e.g., the policy may have long-run consequences extending after the point of evaluation).

- In advance of choosing an activity, agents may be uncertain about the outcomes that will actually occur.
- They may also be uncertain about the full costs they will bear.
- In general the agent's information is not the same as the analyst's, and they may not be nested.
- The agent may know things in advance that the analyst may never discover.
- On the other hand, the analyst, 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.
- One can write the evaluation of s by an agent, using somewhat nonstandard notation, as

$$R\left(s, \mathcal{I}_a\right) = \mu_R\left(s, \mathcal{I}_a\right) + \upsilon\left(s, \mathcal{I}_a\right),$$

reflecting that *ex ante* valuations are made on the basis of *ex ante* information where $\mu_R\left(s, \mathcal{I}_a\right)$ is determined by variables that are known to the econometrician and $\upsilon\left(s, \mathcal{I}_a\right)$ are components known to the agent but not the econometrician.

- *Ex post* evaluations can also be made using a different information set $\mathcal{I}_{ep}$ reflecting the arrival of information after the choice is realized.
- It is possible that

$$\text{argmax}_{s\in\mathcal{S}}\left\{R\left(s,\mathcal{I}_a\right)\right\} \neq \text{argmax}_{s\in\mathcal{S}}\left\{R\left(s,\mathcal{I}_{ep}\right)\right\},$$

in which case there may be *ex post* regret or elation about the choice made.

- 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.

- The econometrician may possess yet a different information set, $\mathcal{I}_e$.
- Choice probabilities computed against one information set are not generally the same as those computed against another information set.
- Operating with hindsight, the econometrician may be privy to some information not available to agents when they make their choices.

- *Ex post* assessments of a program through surveys administered to agents who have completed it may disagree with *ex ante* assessments of the program.
- Both may reflect honest valuations of the program (Hensher, Louviere, and Swait, 1999; Katz, Gutek, Kahn, and Barton, 1975). They are reported when agents have different information about it or have their preferences altered by participating in the program.
- Before participating in a program, agents may be uncertain about the consequences of participation.

- An agent who has completed program j may know $Y(j,\omega)$ but can only guess at the alternative outcome $Y(k,\omega)$ which they have not experienced.
- In this case, *ex post* "satisfaction" with j relative to k for agent ω who only participates in k is synonymous with the following inequality,

$$R(Y(j,\omega),\omega) \geq E(R(Y(k,\omega),\omega) \mid \mathcal{I}_\omega), \tag{5}$$

where the information is post-treatment.

- Survey questionnaires about "client" satisfaction with a program capture subjective elements of program experience not captured by "objective" measures of outcomes that usually exclude psychic costs and benefits.
- Heckman et al. (1997), Heckman and Smith (1998), and Heckman, LaLonde, and Smith (1999) present evidence on this question.
- Carneiro et al. (2001, 2003), Cunha et al. (2005, 2006) and Heckman and Navarro (2007) develop econometric methods for distinguishing *ex ante* from *ex post* evaluations of social programs.

- See Abbring and Heckman (2007) for an extensive survey of this literature.
- Heckman and Vytlacil (2007a) discuss the data needed to identify these criteria, and present examples of Roy models and their extensions that allow for more general decision rules and imperfect information by agents.
- They show how to use economic models to form treatment parameters.

A Specific Model

- Y_1 is an index of well being of the patient if treated; Y_0 if untreated.
- At any point in time, a person can be either treated or untreated.
- The decision to treat may be made on the basis of the expected outcomes $E(Y_1 \mid \mathcal{I})$ and $E(Y_0 \mid \mathcal{I})$ and costs $E(C \mid \mathcal{I})$ where the expectations are those of the relevant decision maker – the patient, the doctor or possibly the parent if the patient is a child.
- The costs might be the pain and suffering of the patient and/or the direct medical costs of the patient.
- For any problem, the costs C and expectations $\mathcal{I}$ are for the relevant decision maker who decides who gets treatment.

- From the point of view of the patient the expected utility or value of treatment is $E(Y_1 \mid \mathcal{I}) - E(C \mid \mathcal{I})$.
- The value of no treatment is $E(Y_0 \mid \mathcal{I})$.

- The expected net value is

$$E(Y_1 \mid \mathcal{I}) - E(C \mid \mathcal{I}) - E(Y_0 \mid \mathcal{I}). \tag{6}$$

Then for patients who pick a treatment based on maximum gain

$$D = \begin{cases} 1, & \text{if } [E(Y_1 \mid \mathcal{I}) - E(C \mid \mathcal{I}) - E(Y_0 \mid \mathcal{I})] \geq 0; \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

or, more succinctly,
$D = \mathbf{1}[(E(Y_1 \mid \mathcal{I}) - E(C \mid \mathcal{I}) - E(Y_0 \mid \mathcal{I})) \geq 0]$.

- This is the generalized Roy model developed in Cunha et al. (2005). See Heckman and Vytlacil (2007a) for a survey of such models.

- If the doctor makes the decision to treat, then the relevant C and $\mathcal{I}$ are those of the doctor.
- Instead of a drug, the treatment can be schooling, migration, installation of a technology and the potential outcomes are the counterfactuals with or without treatment.
- The *ex post* treatment effect is $Y_1 - Y_0$.
- The *ex ante* effect is $E(Y_1 \mid \mathcal{I}) - E(Y_0 \mid \mathcal{I})$.

- Behavioral or scientific theory motivates the construction of (Y_0, Y_1) and the decision to assign treatment.
- The most basic model in economics is the Roy model previously mentioned.
- The decision maker's information is perfect.
- There are no direct costs of treatment $(C = 0)$ and the decision rule is

$$D = \mathbf{1}(Y_1 \geq Y_0). \tag{8}$$

  Those who get treatment are the ones who benefit from it.
- Thus the treated are a non-random sample of the general population, and there is selection bias in using the treated sample to infer what the average person would experience if selected at random.

- The econometric approach models the dependence between observed $Y = DY_1 + (1 - D)Y_0$ and D to suggest alternative estimators to identify causal parameters.
- Recent work identifies various mean treatment effects, distributions of treatment effects and the cost of treatment including the pain and suffering of the patients.

- Commonly used specifications are

$$\begin{aligned} Y_1 &= X\beta_1 + U_1 \\ Y_0 &= X\beta_0 + U_0 \\ C &= Z\gamma + U_C, \end{aligned} \tag{9}$$

  where (X, Z) are observed by the analyst and (U_1, U_0, U_C) are unobserved.

- The patient may know more or less than the analyst.
- Econometric models allow for the patient to know more (observe more) than the analyst and analyze patient selection into treatment accounting for the asymmetry in knowledge between the economist and the the person being observed.

- (Matching assumes that, conditional on X and Z, D is independent of Y_0, Y_1 and so assumes a lot of information is available to the analyst.) The Roy model sets $\gamma = 0, U_C = 0$ and assumes normality for (U_0, U_1).
- These distribution and parametric assumptions are relaxed in the recent econometric literature (See Heckman and Vytlacil, 2007a, for a review).

- The statistical approach does not model the treatment assignment rule or its relationship to potential outcomes.
- The econometric approach makes the treatment assignment equation the centerpiece of its focus and considers both objective and subjective valuations as well as *ex ante* $(E(Y_1 \mid \mathcal{I}), E(Y_0 \mid \mathcal{I}), E(C \mid \mathcal{I}))$ and *ex post* outcomes (Y_1, Y_0, C).
- For this model, EOTM is $E(Y_1 - Y_0 \mid E(Y_1 \mid \mathcal{I}) - E(Y_0 \mid \mathcal{I}) - E(C \mid \mathcal{I}) = 0)$, i.e., the gain to people just indifferent between treatment and no treatment.

- The literature on policy evaluation in statistics sometimes compares econometric "structural" approaches with "treatment effect" or "causal" models (see, e.g., Angrist and Imbens, 1995; Angrist, Imbens, and Rubin, 1996). The comparison is not clear because the terms are not precisely defined.
- Heckman (2005) and Heckman and Vytlacil (2007a) formally define "structural" models and use them as devices for generating counterfactuals.
- They consider both outcome and treatment choice equations.
- This section presents a brief introduction to the econometric approach and compares it with models for causal inference in statistics.

Generating Counterfactuals

- The traditional model of econometrics is the "all causes" model.
- It writes outcomes as a deterministic mapping of inputs to outputs:

$$y(s) = g_s(x, u_s), \tag{10}$$

  where x and u_s are fixed variables specified by the relevant economic theory.
- The notation anticipates the distinction between observable (x) and unobservable (u_s) that is important in empirical implementation.
- The role of the two types of variables in (10) is symmetric.

- This notation allows for different unobservables u_s to affect different outcomes.
- $\mathcal{D}$ is the domain of the mapping $g_s : \mathcal{D} \longrightarrow \mathcal{R}^y$, where $\mathcal{R}^y$ is the range of y.
- There may be multiple outcome variables.
- All outcomes are explained in a functional sense by the arguments of g_s in (10). If one models the *ex post* realizations of outcomes, it is entirely reasonable to invoke an all causes model since the realizations are known (*ex post*) and all uncertainty has been resolved.
- Implicit in the definition of a function is the requirement that g_s be "stable" or "invariant" to changes in x and u_s.
- The g_s function remains stable as its arguments are varied.
- Invariance is a key property of a causal model.

- Equation (10) is a production function relating inputs (factors) to outputs.
- g_s maps (x, u_s) into the range of y or image of $\mathcal{D}$ under g_s, where the domain of definition $\mathcal{D}$ may differ from the empirical support.
- Thus, equation (10) maps admissible inputs into possible *ex post* outcomes.
- This notation allows for different unobservables from a common list u to appear in different outcome equations.

- A "deep structural" version of (10) models the variation of the g_s in terms of s as a map constructed from generating characteristics q_s, x and u_s into outcomes:

$$y(s) = g\left(q_s, x, u_s\right), \tag{11}$$

where now the domain of g, $\mathcal{D}$, is defined for q_s, x, u_s so that $g : \mathcal{D} \longrightarrow \mathcal{R}^y$.

- The components q_s provide the basis for generating the counterfactuals across treatments from a base set of characteristics.
- g maps (q_s, s, u_s) into the range of y, $g : (q_s, x, u_s) \longrightarrow \mathcal{R}^y$, where the domain of definition $\mathcal{D}$ of g may differ from the empirical support.

- In this specification, different treatments s are characterized by different bundles of a set of characteristics common across all treatments.
- This framework provides the basis for solving policy problem P3 since new policies (treatments) are generated from common characteristics, and all policies are put on a common basis.
- If a new policy is characterized by known transformations of (q_s, x, u_s) that lie in the domain of definition of g, policy forecasting problem P3 can be solved.
- The argument of the maps g_s and g are part of the *a priori* specification of a causal model.
- Analysts may disagree about appropriate arguments to include in these maps.

- One benefit of a treatment effect approach that focuses on problem P1 is that it works solely with outcomes rather than inputs.
- However, it is silent on how to solve problems P2 and P3 and provides no basis for interpreting the population level treatment effects.

- Consider alternative models of schooling outcomes of pupils where s indexes the schooling type (e.g., regular public, charter public, private secular and private parochial). The q_s are the observed characteristics of schools of type s.
- The x are the observed characteristics of the pupil.
- The u_s are the unobserved characteristics of both the schools and the pupil.
- If one can characterize a proposed new type of school as a new package of different levels of the same ingredients x, q_s, and u_s and one can identify (11) over the domain of the function defined by the new package, one can solve problem P3.

- If the same schooling input (same q_s) is applied to different students (those with different x) and one can identify (10) or (11) over the new domain of definition, one can solve problem P2.
- By digging deeper into the "causes of the effects" one can do more than just compare the effects of treatments in place with each other.
- In addition, modeling the u_s and its relationship with the corresponding unobservables in the treatment choice equation is highly informative on the choice of appropriate identification strategies.

- Equations (10) and (11) are sometimes called Marshallian causal functions.
- Assuming that the components of (x, u_s) or (q_s, x, u_s) are variation-free, a feature that may or may not be produced by the relevant theory, one may vary each argument of these functions to get a *ceteris paribus* causal effect of the argument on the outcome.
- (See Heckman and Vytlacil, 2007a, for a precise definition of variation-free.) Some components may be variation-free while others are not.
- These thought experiments are conducted for hypothetical variations.
- Recall that the *a priori* theory specifies the arguments in the causal functions and the list of things held fixed when a variable is manipulated.

- Changing one coordinate while fixing the others produces a Marshallian *ceteris paribus* causal effect of a change in that coordinate on the outcome variables.
- Varying q_s fixes different treatment levels.
- Variations in u_s among agents explain why people with the same x characteristics respond differently to the same treatment s.
- The *ceteris paribus* variation need not be for a single variable of the function.
- A treatment generally consists of a package of characteristics and if one varies the package from q_s to $q_{s'}$ one gets different treatment effects.

- I use the convention that lower-case values are used to define fixed values and upper-case notation denotes random variables.
- In defining (10) and (11), I have explicitly worked with fixed variables that are manipulated in a hypothetical way as in the algebra of elementary physics.
- In a purely deterministic world, agents respond to these non-stochastic variables.
- Even if the world is uncertain, *ex post*, after the realization of uncertainty, the outcomes of uncertain inputs are deterministic.
- Some components of u_s may be random shocks realized after decisions about treatment are made.

- Thus if uncertainty is a feature of the environment, (10) and (11) can be interpreted as *ex post* realizations of the counterfactual as uncertainty is resolved.
- *Ex ante* versions may be different.
- From the point of view of agent ω with information set $\mathcal{I}_\omega$, the *ex ante* expected value of $Y(s,\omega)$ is

$$E\left(Y(s,\omega) \mid \mathcal{I}_\omega\right) = E\left(g\left(Q(s,\omega), X(\omega), U(s,\omega)\right) \mid \mathcal{I}_\omega\right), \quad (12)$$

where $Q(s,\omega)$, $X(\omega)$, $U(s,\omega)$ are random variables generated from a distribution that depends on the agent's information set indexed by $\mathcal{I}_\omega$.

- The expectation might be computed using the information set of the relevant decision maker (e.g., the parents in the case of the outcomes of the child) who might not be the agent whose outcomes are measured.
- These random variables are drawn from agent ω's subjective distribution.
- This distribution may differ from the distribution produced by "reality" or nature if agent expectations are different from objective reality.

- In the presence of intrinsic uncertainty, the relevant decision maker acts on (12) but the *ex post* counterfactual is

$$Y(s,\omega) = E(Y(s,\omega) \mid \mathcal{I}_\omega) + \nu(s,\omega), \tag{13}$$

where $\nu(s,\omega)$ satisfies $E(\nu(s,\omega) \mid \mathcal{I}_\omega) = 0$.

- In this interpretation, the information set of agent ω is part of the model specification but the realizations come from a probability distribution, and the information set includes the technology g.
- This representation clarifies the distinction between deterministic *ex post* outcomes and intrinsically random *ex ante* outcomes.
- Abbring and Heckman (2007) survey econometric evaluation models accounting for uncertainty.

- This restatement of the basic deterministic model reconciles the "all causes" model (10) and (11) with the intrinsic uncertainty model favored by some statisticians (see, e.g., Dawid, 2000, and the discussion following his paper). *Ex ante*, there is uncertainty at the agent (ω) level but *ex post* there is not.
- The realizations of $\nu(s, \omega)$ are ingredients of the *ex post* "all causes" model, but not part of the subjective *ex ante* "all causes" model.
- The probability law used by the agent to compute the expectations of $g(Q(s, \omega), X(\omega), U_s(\omega))$ may differ from the objective distribution that generates the observed data.
- In the *ex ante* "all causes" model, manipulations of $\mathcal{I}_\omega$ define the *ex ante* causal effects.

- Thus, from the point of view of the agent, one can vary elements in $\mathcal{I}_\omega$ to produce Marshallian *ex ante* causal response functions.
- The *ex ante* treatment effect from the point of view of the agent for treatment s and s' is

$$E\left(Y\left(s,\omega\right)\mid\mathcal{I}_\omega\right)-E\left(Y\left(s',\omega\right)\mid\mathcal{I}_\omega\right). \tag{14}$$

- The data used to determine these functions may be limited in their support.
- In that case analysts cannot fully identify the theoretical relationships over hypothetical domains of definition.
- In addition, in the support, the components of X, $U(s)$ and $\mathcal{I}_\omega$ may not be variation free even if they are variation free in the hypothetical domain of definition of the function.
- If the X in a sample are functionally dependent, it is not possible to identify the Marshallian causal function with respect to variations in x over the available support even if one can imagine hypothetically varying the components of x over the domains of definition of the functions (10) or (11).

- I next turn to an important distinction between fixing and conditioning on factors that gets to the heart of the distinction between causal models and correlational relationships.
- This point is independent of any problem with the supports of the samples compared to the domains of definition of the functions.

Fixing vs. Conditioning

- The distinction between *fixing* and *conditioning* on inputs is central to distinguishing true causal effects from spurious causal effects.
- In an important paper, 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.
- 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.
- If the conditioning set is sufficiently rich, fixing and conditioning are the same in an *ex post* all causes model.
- Pearl suggests a particular physical mechanism for fixing

- The distinction between fixing and conditioning is most easily illustrated in the linear regression model analyzed by Haavelmo (1943). Let $y = x\beta + u$.
- While y and u are scalars, x may be a vector.
- The linear equation maps every pair (x, u) into a scalar $y \in \mathbb{R}$.
- Suppose that the support of random variable (X, U) in the data is the same as the domain of (x, u) that are fixed in the hypothetical thought experiment and that the (x, u) are variation-free (i.e., can be independently varied coordinate by coordinate). Abstract from the problem of limited support that is discussed in the preceding section.

- Dropping the "ω" notation for random variables, write

$$Y = X\beta + U.$$

U is assumed to have a finite mean.

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

- Formally, one can write this model defined at the population level as a conditional expectation,

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

  Since conditioning is on both X and U, there is no further source of variation in Y.

- This is a deterministic model that coincides with the "all causes" model.
- Thus on the support, which is also assumed to be the domain of definition of the function, this model is the same model as the deterministic, hypothetical model, $y = x\beta + u$.
- 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, one only conditions on X, one obtains

$$E(Y \mid X = x) = x\beta + E(U \mid X = x). \tag{15}$$

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

- It generates only an $X-$constant relationship.
- Unless one conditions on all of the "causes" (the right hand side variables), the empirical relationship (15) does not identify causal effects of X on Y.
- The variation in X also moves the conditional mean of U given X.
- This analysis can be generalized to a nonlinear model $y = g(q, x, u)$ (See Pearl, 2000). It can be generalized to account for the temporal resolution of uncertainty if one includes $\nu(s, \omega)$ as an argument in the *ex post* causal model.
- The outcomes can include both objective outcomes $Y(s, \omega)$

- Parallel to causal models for outcomes are causal models for the choice of treatment (see Heckman and Vytlacil, 2007a). Accounting for uncertainty and subjective valuations of outcomes (e.g., pain and suffering for a medical treatment) is a major contribution of the econometric approach (see, e.g., Carneiro, Hansen, and Heckman, 2003; Chan and Hamilton, 2006; Cunha, Heckman, and Navarro, 2005, 2006; Cunha and Heckman, 2007; and Heckman and Navarro, 2007). The factors that lead an agent to participate in treatment s may be dependent on the factors affecting outcomes.
- Modeling this dependence is a major source of information used in the econometric approach to construct counterfactuals from real data.
- A parallel analysis can be made if the decision maker is not the same as the agent whose objective outcomes are being evaluated.

The Econometric Model vs. the Neyman–Rubin Model

- Many statisticians and social scientists use a model of counterfactuals and causality attributed to Donald Rubin by Paul Holland (1986). The framework was developed in statistics by Neyman (1923), Cox (1958) and others.

- Parallel frameworks were independently developed in psychometrics (Thurstone, 1927) and economics (Haavelmo, 1943; Quandt, 1958, 1972; Roy, 1951). The statistical treatment effect literature originates in the statistical literature on the design of experiments.
- It draws on hypothetical experiments to define causality and thereby creates the impression in the minds of many of its users that random assignment is the most convincing way to identify causal models.
- Some would say it is the only way to identify causal models.

- 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 econometric approach (see equations (10)–(13)). Rubin and Neyman offer no model of the choice of which outcome is selected.
- No lower case, "all causes" model explicitly specified in this approach nor is there any discussion of the social science or theory producing the outcomes studied.

- In this notation, Rubin (1986) invokes versions of traditional econometric invariance assumptions called "SUTVA" for Stable Unit Treatment Value Assumption.
- Since he does not develop choice equations or subjective evaluations, he does not consider the more general invariance conditions for both objective and subjective evaluations briefly discussed below. Invariance assumptions were developed in Cowles Commission econometrics and formalized in Hurwicz (1962). They are surveyed in Heckman and Vytlacil (2007a).

The Rubin model assumes

(R-1) (R-1)

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

(R-2) (R-2)

Invariance of counterfactuals for objective outcomes to the mechanism of assignment within a policy regime;

(R-3) (R-3)

No social interactions or general equilibrium effects for objective outcomes; and

(R-4) (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 that P1 is the only evaluation problem of interest and that mean causal effects are the only objects of interest.

- The econometric approach considers a wider array of policy problems than the statistical treatment effect approach.
- Its signature features are:

- Points 2 and 3 introduce agent decision making into the treatment effect literature.
- Development of an explicit framework for outcomes $Y(s, \omega)$, $s \in \mathcal{S}$, measurements and the choice of outcomes where the role of unobservables ("missing variables") in creating selection problems and justifying estimators is explicitly developed.
- The analysis of subjective evaluations of outcomes $R(s, \omega)$, $s \in \mathcal{S}$, and the use of choice and compliance data to infer them.
- 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.

- Development of models for identifying and evaluating entire distributions of treatment effects (*ex ante* and *ex post*) rather than just the traditional mean parameters.
- These distributions enable analysts to determine the proportion of people who benefit from treatment, a causal parameter not considered in the statistical literature on treatment effects.

- Models for simultaneous causality.
- Definitions of parameters made without appeals to hypothetical experimental manipulations.

- Clarification of the need for invariance of parameters with respect to different classes of manipulations to answer different classes of questions.
- I now amplify these points.

- Selection models defined for potential outcomes with explicit treatment assignment mechanisms were developed by Gronau (1974) and Heckman (1974, 1976, 1978, 1979) in the economics literature before the Neyman–Rubin model was popularized in statistics.
- The econometric discrete choice literature (McFadden, 1974, 1981) uses counterfactual utilities or subjective evaluations as did its parent literature in mathematical psychology (Thurstone, 1927, 1959). The model sketched in section 111 considers both choices and outcomes of choices.

- Unlike the Neyman–Rubin model, these models do not start with the experiment as an ideal but they start with well-posed, clearly articulated models for outcomes and treatment choice where the unobservables that underlie the selection and evaluation problem are made explicit.
- The hypothetical manipulations define the causal parameters of the model.
- Randomization is a metaphor and not an ideal or "gold standard".

- In contrast to the econometric model, the Holland (1986) – Rubin (1978) definition of causal effects is based on randomization.
- The analysis in Rubin's 1976 and 1978 papers is a dichotomy between randomization ("ignorability") and non-randomization, and not an explicit treatment of particular selection mechanisms in the non-randomized case as developed in the econometrics literature.
- There is no explicit discussion of treatment selection rules like (8) and their relationship with realized outcomes.

- Even under ideal conditions, randomization cannot answer some very basic questions, such as what proportion of a population benefits from a program (Heckman, 1992). See Carneiro et al. (2001, 2003), where this proportion is identified using choice data and/or supplementary proxy measures.
- See also Cunha et al. (2005, 2006) and Cunha and Heckman (2007). Abbring and Heckman (2007) discuss this work.
- In practice, contamination and cross-over effects make randomization a far from sure-fire solution even for constructing $E(Y_1 - Y_0)$.
- See the evidence on disruption bias and contamination bias arising in randomized trials that is presented in Heckman et al. (1999) and Heckman et al. (2000).

- Many leading causal analysts conflate the three points of Table 1. The analysis of Holland (1986) illustrates this point and the central role of the randomized trial to the Holland-Rubin analysis.
- After explicating the "Rubin model," Holland gives a very revealing illustration that conflates the first two tasks of Table 1. He claims that there can be no causal effect of gender on earnings because analysts cannot randomly assign gender.
- This statement confuses the act of defining a causal effect (a purely mental act) with empirical difficulties in estimating it.
- These are tasks 1 and 2 in Table 1.

- As another example of the same point, (Rubin, 1978, p. 39) denies that it is possible to define a causal effect of sex on intelligence because a randomization cannot *in principle* be performed.
- "Without treatment definitions that specify actions to be performed on experimental units, I cannot unambiguously discuss causal effects of treatments" (Rubin, 1978, p. 39). In this and many other passages in the statistics literature, a causal effect is defined by a randomization.
- Issues of definition and identification are confused.

- This confusion continues to flourish in the literature in applied statistics.
- For example, Berk et al. (2005) echo Rubin and Holland in insisting that if an experiment cannot "in principle" be performed, a causal effect cannot be defined.
- The local average treatment effect "LATE" parameter of Imbens and Angrist (1994) is defined by an instrument and conflates tasks 1 and 2 (definition and identification). Imbens and Angrist (1994) use instrumental variables as surrogates for randomization.
- Heckman and Vytlacil (1999, 2005) and Heckman, Urzúa, and Vytlacil (2006) define the LATE parameter abstractly and separate issues of definition of parameters from issues of identification.

- The act of definition is logically distinct from the acts of identification and inference.
- A purely mental act can define a causal effect of gender.
- That is a separate task from identifying the causal effect.
- The claim that causality can only be determined by randomization reifies randomization as the "gold standard" of causal inference.

- The econometric approach to causal inference is more comprehensive than the Neyman–Rubin model of counterfactuals with models of the choice of counterfactuals $\{D(s,\omega)\}_{s\in\mathcal{S}}$ and the relationship between choice equations and the counterfactuals.
- The $D(s,\omega)$ are explicitly modeled as generated by the collection of random variables $(Q(s,\omega), C(s,\omega), Y(s,\omega) \mid \mathcal{I}_\omega)$, $s \in \mathcal{S}$, where $Q(s,\omega)$ is the characteristic of treatment s for agent ω, $C(s,\omega)$ are costs and $\{Y(s,\omega)\}_{s\in\mathcal{S}}$ are the outcomes and the "$\mid$" denotes that these variables are defined conditional on $\mathcal{I}_\omega$ (the agent's information set). If other agents make treatment assignment decisions, then the determinants of $D(s,\omega)$ are modified according to what is in their information set.
- The variables determining choices are analyzed.

- Modeling $Y(s, \omega)$ in terms of the characteristics of treatment, and of the treated, facilitates comparisons of counterfactuals and derived causal effects across studies where the composition of programs and treatment group members may vary.
- It also facilitates the construction of counterfactuals on new populations and the construction of counterfactuals for new policies.
- The Neyman–Rubin framework focuses exclusively on population level mean "causal effects" or treatment effects for policies actually experienced and provides no framework for extrapolation of findings to new environments or for forecasting new policies (problems P2 and P3).

Nonrecursive (Simultaneous) Models of Causality

- A system of linear simultaneous equations captures interdependence among outcomes Y.
- For simplicity, we ignore the revelation of information over time.
- To focus the issue of on nonrecursive causal models, in this subsection we also assume that the domain of definition of the model is the same as the support of the population data.
- Thus the model for values of upper-case variables has the same support as the domain of definition for the model in terms of lower-case variables.

- The model developed in this section is rich enough to model interactions among agents.
- For simplicity we work with linear equations.
- We write this model in terms of parameters (Γ, B), observables (Y, X) and unobservables U as

$$\Gamma Y + BX = U, \qquad E(U) = 0, \tag{16}$$

where Y is now 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 (16) are sometimes called "structural equations." A better nomenclature, suggested by Leamer (1985), is that the Y are internal variables determined by the model and the X are external variables specified outside the model.
- This formulation is static.
- In a dynamic framework, Y_t would be the internal variables and the lagged Y, $Y_{t-k}, k > 0$, would be external to period t and be included in the X_t. Thus we could work with lagged dependent variables.
- The system would be $\Gamma Y_t + BX_t = U_t,\ E(U_t) = 0.$

- This definition distinguishes two issues: (a) defining variables (Y) that are determined from inputs outside the model (the X) and (b) determining the relationship between observables and unobservables.
- In a time series model, the internal variables are Y_t determined in period t.
- When the model is of full rank (Γ^{-1} exists), it is said to be "complete."
- A complete model produces a unique Y from a given (X, U).
- A complete model is said to be in reduced form when structural Equation (16) is multiplied by Γ^{-1}.
- The reduced form is $Y = \Pi X + \mathcal{E}$ where $\Pi = -\Gamma^{-1}B$ and $\mathcal{E} = \Gamma^{-1}U$.

- Π is the reduced form coefficient.
- This is a linear-in-the-parameters "all causes" model for vector Y, where the causes are X and $\mathcal{E}$.
- The "structure" is (Γ, B), Σ_U, where Σ_U is the variance-covariance matrix of U.
- In the Cowles Commission analysis it is assumed that Γ, B, Σ_U are invariant to general changes in X and translations of U.

- This is implied by the invariance of the structure but is a weaker requirement.
- The reduced form slope coefficients are Π, and $\Sigma_{\mathcal{E}}$ is the variance-covariance matrix of $\mathcal{E}$.
- The original formulations of this model assumed normality so that only means and variances were needed to describe the joint distributions of (Y, X).
- In the population generating (16), least squares recovers Π provided Σ_X, the variance of X, is nonsingular (no multicollinearity).
- In this linear-in-parameters equation setting, the full rank condition for Σ_X is a variation-free condition on the external variables.

- The reduced form solves out the Y to produce the net effect of X on Y.
- The linear-in-parameters model is traditional.
- The underlying all causes model writes $\Gamma y + Bx = u$, $y = \Pi x + \varepsilon$, $\Pi = -\Gamma^{-1}B$, $\varepsilon = \Gamma^{-1}u$.
- Recall that we assume that the domain of the all causes model is the same as the support of (X, U).
- Thus there is a close correspondence between these two models.
- Nonlinear versions are available (Fisher, 1966; Matzkin, 2004, 2007).

- Thus we can postulate a system of equations $G(Y, X, U) = 0$ and develop conditions for unique solution of reduced forms $Y = K(X, U)$ requiring that certain Jacobian terms be nonvanishing. See Matzkin (2007).
- The structural form (16) is an all causes model that relates in a deterministic way outcomes (internal variables) to other outcomes (internal variables) and external variables (the X and U).
- Without some restrictions, *ceteris paribus* manipulations associated with the effect of some components of Y on other components of Y are not possible within the model.

- We now demonstrate this point.
- For specificity, consider a two-agent model of social interactions.
- Y_1 is the outcome for agent 1; Y_2 is the outcome for agent 2.
- This could be a model of interdependent consumption where the consumption of agent 1 depends on the consumption of agent 2 and other agent-1-specific variables (and possibly other agent-2-specific variables).
- It could also be a model of test scores.
- We can imagine populations of data generated from sampling the same two-agent interaction over time or sampling different two-agent couplings at a point in time.

- Assuming that the preferences are interdependent, we may write the equations in structural form as

$$Y_1 = \alpha_1 + \gamma_{12} Y_2 + \beta_{11} X_1 + \beta_{12} X_2 + U_1, \quad (17a)$$
$$Y_2 = \alpha_2 + \gamma_{21} Y_1 + \beta_{21} X_1 + \beta_{22} X_2 + U_2. \quad (17b)$$

- This model is sufficiently flexible to capture the notion that the consumption of 1 (Y_1) depends on the consumption of 2 if $\gamma_{12} \neq 0$, as well as 1's value of X if $\beta_{11} \neq 0$, X_1 (assumed to be observed), 2's value of X , X_2 if $\beta_{12} \neq 0$ and unobservable factors that affect 1 (U_1).
- The determinants of 2's consumption are defined symmetrically.
- We allow U_1 and U_2 to be freely correlated.

- We assume that U_1 and U_2 are mean independent of (X_1, X_2) so

$$E\left(U_1 \mid X_1, X_2\right) = 0 \tag{18a}$$

and

$$E\left(U_2 \mid X_1, X_2\right) = 0. \tag{18b}$$

- Completeness guarantees that (17a) and (17b) have a determinate solution for (Y_1, Y_2).
- Applying Haavelmo's (1943) analysis to (17a) and (17b), the causal effect of Y_2 on Y_1 is γ_{12}.
- This is the effect on Y_1 of fixing Y_2 at different values, holding constant the other variables in the equation.

- Symmetrically, the causal effect of Y_1 on Y_2 is γ_{21}.
- Conditioning, i.e., using least squares, in general, fails to identify these causal effects because U_1 and U_2 are correlated with Y_1 and Y_2. This is a traditional argument.
- It is based on the correlation between Y_2 and U_1.
- But even if $U_1 = 0$ and $U_2 = 0$, so that there are no unobservables, least squares breaks down because Y_2 is perfectly predictable by X_1 and X_2. We cannot simultaneously vary Y_2, X_1 and X_2.
- To see why, we derive the reduced form of this model.

- Assuming completeness, the reduced form outcomes of the model after social interactions are solved out can be written as

$$Y_1 = \pi_{10} + \pi_{11} X_1 + \pi_{12} X_2 + \mathcal{E}_1, \tag{19a}$$
$$Y_2 = \pi_{20} + \pi_{21} X_1 + \pi_{22} X_2 + \mathcal{E}_2. \tag{19b}$$

- Least squares can identify the *ceteris paribus* effects of X_1 and X_2 on Y_1 and Y_2 because $E(\mathcal{E}_1 \mid X_1, X_2) = 0$ and $E(\mathcal{E}_2 \mid X_1, X_2) = 0$. Simple algebra informs us that

$$\pi_{11} = \frac{\beta_{11} + \gamma_{12}\beta_{21}}{1 - \gamma_{12}\gamma_{21}}, \qquad \pi_{12} = \frac{\beta_{12} + \gamma_{12}\beta_{22}}{1 - \gamma_{12}\gamma_{21}}, \tag{20}$$
$$\pi_{21} = \frac{\gamma_{21}\beta_{11} + \beta_{21}}{1 - \gamma_{12}\gamma_{21}}, \qquad \pi_{22} = \frac{\gamma_{21}\beta_{12} + \beta_{22}}{1 - \gamma_{12}\gamma_{21}},$$

and

$$\begin{aligned} \mathcal{E}_1 &= \frac{U_1 + \gamma_{12}U_2}{1 - \gamma_{12}\gamma_{21}}, \\ \mathcal{E}_2 &= \frac{\gamma_{21}U_1 + U_2}{1 - \gamma_{12}\gamma_{21}}. \end{aligned}$$

- Observe that because $\mathcal{E}_2$ depends on both U_1 and U_2 in the general case, Y_2 is correlated with U_1 through the direct channel of U_1 and through the correlation between U_1 and U_2.
- Without any further information on the variances of (U_1, U_2) and their relationship to the causal parameters, we cannot isolate the causal effects γ_{12} and γ_{21} from the reduced form regression coefficients.
- This is so because holding X_1, X_2, U_1 and U_2 fixed in (17a) or (17b), it is not *in principle* possible to vary Y_2 or Y_1, respectively, because they are exact functions of X_1, X_2, U_1 and U_2.
- This exact dependence holds true even if $U_1 = 0$ and $U_2 = 0$ so that there are no unobservables.
- See Fisher (1966).

- In this case, which is thought to be the most favorable to the application of least squares to (17a) and (17b), it is evident from (19a) and (19b) that when $\mathcal{E}_1 = 0$ and $\mathcal{E}_2 = 0$, Y_1 and Y_2 are exact functions of X_1 and X_2.
- There is no mechanism yet specified within the model to independently vary the right hand sides of Equations (17a) and (17b).
- Some economists suggest that the mere fact that we can write (17a) and (17b) means that we "can imagine" independent variation.

- By the same token, we "can imagine" a model

$$Y = \varphi_0 + \varphi_1 X_1 + \varphi_2 X_2,$$

but if part of the model is $(*)\, X_1 = X_2$, no causal effect of X_1 holding X_2 constant is possible in principle within the rules of the model.

- If we break restriction $(*)$ and permit independent variation in X_1 and X_2, we can define the causal effect of X_1 holding X_2 constant.
- The X effects on Y_1 and Y_2, identified through the reduced forms, combine the direct effects (through β_{ij}) and the indirect effects (as they operate through Y_1 and Y_2, respectively).
- If we assume exclusions $(\beta_{12} = 0)$ or $(\beta_{21} = 0)$ or both, we can identify the *ceteris paribus* causal effects of Y_2 on Y_1 and of Y_1 on Y_2, respectively, if $\beta_{22} \neq 0$ or $\beta_{11} \neq 0$, respectively.

- Thus if $\beta_{12} = 0$, from the reduced form

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

If $\beta_{21} = 0$, we obtain

$$\frac{\pi_{21}}{\pi_{11}} = \gamma_{21}.$$

- In a general nonlinear model,

$$\begin{aligned} Y_1 &= g_1(Y_2, X_1, X_2, U_1) \\ Y_2 &= g_2(Y_1, X_1, X_2, U_2), \end{aligned}$$

exclusion is defined as $\frac{\partial g_1}{\partial X_1} = 0$ for all (Y_2, X_1, X_2, U_1) and $\frac{\partial g_2}{\partial X_2} = 0$ for all (Y_1, X_1, X_2, U_2).

- Assuming the existence of local solutions, we can solve these equations to obtain

$$\begin{aligned} Y_1 &= \varphi_1(X_1, X_2, U_1, U_2) \\ Y_2 &= \varphi_2(X_1, X_2, U_1, U_2) \end{aligned}$$

- By the chain rule we can write

$$\frac{\partial g_1}{\partial Y_2} = \frac{\partial Y_1}{\partial X_1} \Big/ \frac{\partial Y_2}{\partial X_1} = \frac{\partial \varphi_1}{\partial X_1} \Big/ \frac{\partial \varphi_2}{\partial X_1}.$$

- We may define causal effects for Y_1 on Y_2 using partials with respect to X_2 in an analogous fashion.

- Alternatively, we could assume $\beta_{11} = \beta_{22} = 0$ and $\beta_{12} \neq 0$, $\beta_{21} \neq 0$ to identify γ_{12} and γ_{21}.
- These exclusions say that the social interactions only operate through the Y's.
- Agent 1's consumption depends only on agent 2's consumption and not on his value of X_2.
- Agent 2 is modeled symmetrically versus agent 1.
- Observe that we have *not* ruled out correlation between U_1 and U_2.

- When the procedure for identifying causal effects is applied to samples, it is called **indirect least squares**.
- The method traces back to Tinbergen (1930).
- The analysis for social interactions in this section is of independent interest.
- It can be generalized to the analysis of N person interactions if the outcomes are continuous variables.
- For binary outcomes variables, the same analysis goes through for the special case analyzed by Heckman and MaCurdy (1986).
- However, in the general case, for discrete outcomes generated by latent variables, it is necessary to modify the system to obtain a coherent probability model (Heckman, 1978).

- The intuition for these results is that if $\beta_{12} = 0$, we can vary Y_2 in Equation (17a) by varying the X_2.
- Since X_2 does not appear in the equation, under exclusion, we can keep U_1, X_1 fixed and vary Y_2 using X_2 in (19b) if $\beta_{22} \neq 0$.
- Notice that we could also use U_2 as a source of variation in (19b) to shift Y_2.
- The roles of U_2 and X_2 are symmetric.
- However, if U_1 and U_2 are correlated, shifting U_2 shifts U_1 unless we control for it.
- The component of U_2 uncorrelated with U_1 plays the role of X_2.

- Symmetrically, by excluding X_1 from (17b), we can vary Y_1, holding X_2 and U_2 constant.
- These results are more clearly seen when $U_1 = 0$ and $U_2 = 0$.
- Observe that in the model under consideration, where the domain of definition and the supports of the variables coincide, the causal effects of simultaneous interactions are defined if the parameters are identified in the sense of the traditional Cowles definition of identification (see, e.g., Ruud, 2000, for a modern discussion of these identification conditions).

- A hypothetical thought experiment justifies these exclusions.
- If agents do not know or act on the other agent's X, these exclusions are plausible.
- An implicit assumption in using (17a) and (17b) for causal analysis is invariance of the parameters $(\Gamma, \beta, \Sigma_U)$ to manipulations of the external variables.
- Invariance of the coefficients of equations to classes of manipulation of the variables is an essential part of the definition of structural models which we develop more formally below.

- This definition of causal effects in an interdependent system generalizes the recursive definitions of causality featured in the statistical treatment effect literature (Holland, 1988, and Pearl, 2000).
- The key to this definition is manipulation of external inputs and exclusion, not randomization or matching.
- Indeed matching or, equivalently, OLS in this context, using the right hand side variables of (17a) and (17b), does not identify causal effects as Haavelmo (1943) established long ago.
- We can use the population simultaneous equations model to define the class of admissible variations and address problems of definitions.

- If for a given model, the parameters of (17a) or (17b) shift when external variables are manipulated, or if external variables cannot be independently manipulated, causal effects of one internal variable on another cannot be defined *within that model*.
- If agents were randomly assigned to pair with their neighbors, and the parameters of (17a) were not affected by the randomization, then Y_2 would be exogenous in Equation (17b) and one could identify causal effects by least squares.

- At issue is whether such a randomization would recover γ_{12}.
- It might fundamentally alter agent 1's response to Y_2 if that agent is randomly assigned as opposed to being selected by the agent.
- Judging the suitability of an invariance assumption entails a thought experiment—a purely mental act.

Marschak's Maxim and the Relationship Between the Econometric Literature and the Statistical Treatment Effect Literature: A Synthesis

- The absence of explicit models of outcomes and choice is a prominent feature of the statistical treatment effect literature.
- Scientifically well-posed models make explicit the assumptions used by analysts regarding preferences, technology, the information available to agents, the constraints under which they operate, and the rules of interaction among agents in market and social settings and the sources of variability among agents.

- These explicit features make econometric models useful vehicles (a) for interpreting empirical evidence using theory; (b) for collating and synthesizing evidence across studies using economic theory; (c) for measuring the various effects of policies; and (d) for forecasting the welfare and direct effects of previously implemented policies in new environments and the effects of new policies.

- These features are absent from the modern treatment effect literature.
- At the same time, that literature makes fewer statistical assumptions in terms of independence, functional form, exclusion and distributional assumptions than the standard structural estimation literature in econometrics.
- These are the attractive features of this approach.

- However, the econometric literature has advanced greatly in recent years in terms of producing a robust version of its product.
- Major advances summarized in Powell (1994), Heckman and Vytlacil (2007b), and Matzkin (2007) have relaxed the strong parametric assumptions that characterized the early econometric literature.

- In reconciling these two literatures, I reach back to a neglected but important paper by Marschak (1953). Marschak noted that for many specific questions of policy analysis, it is not necessary to identify fully specified models that are invariant to classes of policy modifications.
- All that may be required for any policy analysis 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).

- I call this principle **Marschak's Maxim** in honor of this insight.
- The modern statistical treatment effect literature implements Marschak's Maxim where the policies analyzed are the treatments available under a particular policy regime and the goal of policy analysis is restricted to evaluating policies in place (task 1 in Table 1) 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 policy question being addressed by the particular treatment effect being identified and why it is interesting.

- 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 given X for s and s' one only requires

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

- Under invariance conditions about outcome equations, this parameter answers the policy question of determining the average effect on outcomes of moving an agent from s' to s when there are no social interaction or contagion effects.

- The parameter is not designed to evaluate a whole host of other policies.
- One does not have to know the functional form of the generating g_s functions nor does X have to be exogenous.
- One does not have to invoke strong conditions about invariance of the choice equations.
- However, if one seeks to identify

$$E(Y(s,\omega) \mid X = x, D(s,\omega) = 1) - E(Y(s',\omega) \mid X = x, D(s,\omega) = 1),$$

one needs to invoke invariance of choice equations recognizing the conditioning on a choice variable.

- No conditioning on a choice is required in defining average treatment effects.

- Treatment effects are causal effects for particular policies that move agents from $s \in \mathcal{S}$ to $s' \in \mathcal{S}$, $s' \neq s$, keeping all other features of the agent and environment the same.
- These effects are designed to address policy problem P1.
- Treatment effects and causal models can be generated from explicit economic models and are more easily interpreted.
- Invariant, explicitly formulated, economic models are useful for addressing policy problems P2 and P3: extrapolation and predicting the effects of new policies, respectively.

- If the goal of an analysis is to predict outcomes, and the environment is stable, then accurate predictions can be made without causal or structural parameters.
- Consider Haavelmo's analysis of fixing vs. conditioning. He analyzed the linear regression model $Y = X\beta + U$ and defined the causal effect of X on Y as the U-constant effect of variations in X.
- If the goal of the analysis is to predict the effect of X on Y, and if the environment is stable so that the historical data have the same distribution as the data in the forecast sample, least squares projections are optimal predictors under mean square error criteria (see, e.g., Goldberger, 1964). One does not need to separate out the causal effect of X on Y, β, from the effect of X on the unobservables operating through $E(U \mid X)$.

- Marschak's Maxim urges analysts to formulate the problem being addressed clearly and to use the minimal ingredients required to solve it.
- The treatment effect literature addresses the problem of comparing treatments under a particular policy regime for a particular environment.
- The original econometric pioneers considered treatments under different policy regimes and with different environments.
- As analysts ask more difficult questions, it is necessary to specify more features of the models being used to address the questions.