Chapter 70

ECONOMETRIC EVALUATION OF SOCIAL PROGRAMS, PART I: CAUSAL MODELS, STRUCTURAL MODELS AND ECONOMETRIC POLICY EVALUATION*

JAMES J. HECKMAN
The University of Chicago, USA
American Bar Foundation, USA
University College Dublin, Ireland

EDWARD J. VYTLACIL
Columbia University, USA

Contents

* This research was supported by NSF: 9709873, 0099195, and SES-0241858 and NICHD: R01-HD32058, and the American Bar Foundation. The views expressed in this chapter are those of the authors and not necessarily those of the funders listed here. Handbook conferences in London and Chicago were supported by NSF grant SBR-9601142. We have benefited from comments received from Thierry Magnac and Costas Meghir at the UCL Handbook of Econometrics Conference, December 1998; general comments at the 2001 Chicago Handbook Conference and discussions with Jaap Abbring, Pedro Carneiro, Steve Durlauf, Hugo Garduño-Arredondo, Seong Moon, Salvador Navarro, Rodrigo Pinto, Peter Savelyev, G. Adam Savvas, Mohan Singh, John Trujillo, Semih Tumen, and Yu Xie. Jaap Abbring and T.N. Srinivasan made especially detailed and very helpful comments on the first draft of this chapter that greatly improved its content. Portions of this chapter first appeared in Heckman (2005) but are substantially revised here.

Handbook of Econometrics, Volume 6B

DOI: 10.1016/S1573-4412(07)06070-9

Abstract

This chapter relates the literature on the econometric evaluation of social programs to the literature in statistics on "causal inference". In it, we develop a general evaluation framework that addresses well-posed economic questions and analyzes agent choice rules and subjective evaluations of outcomes as well as the standard objective evaluations of outcomes. The framework recognizes uncertainty faced by agents and *ex ante*

and *ex post* evaluations of programs. It also considers distributions of treatment effects. These features are absent from the statistical literature on causal inference. A prototypical model of agent choice and outcomes is used to illustrate the main ideas.

We formally develop models for counterfactuals and causality that build on Cowles Commission econometrics. These models anticipate and extend the literature on causal inference in statistics. The distinction between fixing and conditioning that has recently entered the statistical literature was first developed by Cowles economists. Models of simultaneous causality were also developed by the Cowles group, as were notions of invariance to policy interventions. These basic notions are updated to nonlinear and nonparametric frameworks for policy evaluation more general than anything in the current statistical literature on "causal inference". A formal discussion of identification is presented and applied to clearly formulated choice models used to evaluate social programs.

Keywords

causal models, counterfactuals, policy evaluation, policy invariance, structural models, identification

JEL classification: C10, C50

1. Introduction

Evaluating policy is a central problem in economics.[1] Evaluations entail comparisons of outcomes produced from alternative policies using different valuation criteria. Such comparisons often require constructing estimates of outcomes for policies that have never been implemented. They require that the economist construct counterfactuals.[2] Counterfactuals are required to forecast the effects of policies that have been tried in one environment but are proposed to be applied in new environments and to forecast the effects of new policies.

This chapter surveys recent approaches to the empirical construction of economic counterfactuals. The traditional approach to constructing policy counterfactuals in econometrics, first developed in the 1930s, builds econometric models using data, economic theory and statistical methods. The early econometric pioneers developed macroeconomic general equilibrium models and estimated them on aggregate time series data. Later on, economists used newly available microdata on families, individuals and firms to build microstructural models. This approach unites economics, statistics and microdata to build models to evaluate policies, to forecast the effects of extending the policies to new environments and to forecast the effects of new policies. It is exemplified in the chapters by Reiss and Wolak (Chapter 64); Ackerberg, Benkard, Berry and Pakes (Chapter 63); Athey and Haile (Chapter 60); Bond and Van Reenen (Chapter 65); Blundell, MaCurdy and Meghir (Chapter 69); and Blundell and Stoker (Chapter 68) of this Handbook.

More recently, some economists have adapted statistical "treatment effect" approaches that apply methods developed in statistics, educational research, epidemiology and biostatistics to the problem of evaluating economic policy. This approach takes the randomized trial as an ideal. It is much less explicit about the role of economic theory (or any theory) in interpreting evidence or in guiding empirical analyses. The goal of this chapter is to exposit, interpret and unite the best features of these two approaches.

The topics of econometric policy evaluation and policy forecasting are vast, and no chapter within the page limits of a Handbook chapter can cover all aspects of it. In this chapter we focus on microeconomic policy evaluation and policy forecasting.

We focus our discussion on the analysis of a class of latent variable (or "index") models that form the core of modern microeconometrics. Discrete choice theory [McFadden (1974, 1981, 1984, 1985, 2001)] and models of joint discrete and continuous variables [Heckman (1974, 1979, 2001), Heckman and MaCurdy (1986)] are based on latent variable models.[3] Such models provide a framework for integrating economic theory and statistical analysis. They are also frameworks for constructing policy counterfactuals.

[1] We use the term "policy" in a very general sense. It includes alternative actions which might be undertaken by organizations such as private businesses, governments or by family members.

[2] Counterfactuals are not necessarily contrary to fact. They are not directly observed.

[3] These models have their origins in mathematical psychology [Thurstone (1927), Bock and Jones (1968)].

Useful surveys of the econometrics of these models include Maddala (1983), Amemiya (1985), Ruud (2000) and Wooldridge (2002).

Microstructural models can be used to construct a wide variety of policy counterfactuals. They can also be used to evaluate existing policies and to forecast the effects of new policies. Embedded in general equilibrium models, they can also be used to evaluate the effects of changing the scale of existing policies or introducing new policies with substantial coverage [see, e.g., Heckman, Lochner and Taber (1998), Blundell et al. (2004)].

Applications of these models are legion. So are criticisms of this approach. Critics grant the interpretability of the economic frameworks and the parameters derived from them. At the same time, they question the strong functional form, exogeneity, support and exclusion assumptions used in classical versions of this literature, and the lack of robustness of empirical results obtained from them [see Goldberger (1983), Arabmazar and Schmidt (1982), Ruud (1981), Lewis (1986), Angrist and Krueger (1999) among many others].[4] While there have been substantial theoretical advances in weakening the parametric structure used to secure identification of the models used in the early work [see, e.g., Manski (1975, 1988), Heckman and Honoré (1990), Matzkin (1992, 1993, 1994, 2003, 2007), Powell (1994), and Chen (1999)], progress in implementing these procedures in practical empirical problems has been slow and empirical applications of semi-parametric methods have been plagued by issues of sensitivity of estimates to choices of smoothing parameters, trimming parameters, bandwidths and the like [see Chapter 74 (Ichimura and Todd); Chapter 76 (Chen); and Chapter 77 (Carrasco, Florens and Renault) of this Handbook]. The arbitrariness in the choice of parametric models that motivates recent work in semiparametric and nonparametric econometrics has its counterpart in the choice of nonparametric and semiparametric estimation parameters. Often, parametric structural models are computationally cumbersome [see Geweke and Keane (2001)] and identification in dynamic recursive models is often difficult to establish [see Rust (1994), Magnac and Thesmar (2002)], although progress has been made [see Taber (2001), Aguirregabiria (2004), Heckman and Navarro (2007)]. The curse of dimensionality and the complexity of computational methods plague high dimensional parametric models and nonparametric models alike. These considerations motivate pursuit of simpler, more transparent and more easily computed and replicable methods for analyzing economic data and for econometric policy analysis.

The recent literature on treatment effects emphasizes nonparametric identification of certain parameters, robustness, and simplicity (or transparency of identification) as its

[4] We note that most of this literature is based on Monte Carlo analysis or worst case analyses on artificial samples. The empirical evidence on nonrobustness of conventional parametric models is mixed. [See Heckman (2001)]. It remains to be established on a systematic basis that classical normality assumptions invariably produce biased estimates. The evidence in Heckman and Sedlacek (1985) and Blundell, Reed and Stoker (2003) shows that normality is an accurate approximation to log earnings data in economic models of self-selection. The analysis of Todd (1996) shows that parametric probit analysis is accurate for even extreme departures from normality.

main goals. In addition, it recognizes certain forms of heterogeneity in responses to treatment. These are major advances over the traditional structural literature. By focusing on one parameter instead of many, this approach can identify that parameter under weaker conditions than are required for structural parameters that answer many questions. At the same time, this literature is often unclear in stating what economic question the estimated parameters answer. Simplicity in estimation is often accompanied by obscurity in interpretation. The literature also ignores the problems of applying estimated "effects" to new environments or estimating the "effects" of new programs never previously implemented. A new language of counterfactuals and causality has been created. This chapter exposits the treatment effect models and relates them to more explicitly formulated structural econometric models.

Estimators for "causal effects" in the recent treatment effect literature make implicit behavioral assumptions that are rarely exposited. Many papers in the modern treatment effect literature, especially those advocating instrumental variables or natural experiments, proceed by picking an instrument or a natural experiment and defining the parameter of interest as the estimand corresponding to the instrument.[5] Economists using matching make the strong implicit assumption that the information acted on by the agents being studied is as good as that available to the analyst-economist. The literature is often unclear as to what variables to include in conditioning sets and what variables to exclude and the conditions under which an estimator identifies an economically interesting parameter.

The goal of this chapter and Chapter 71 of this Handbook is to integrate the treatment effect literature with the literature on micro-structural econometrics based on index models and latent variable models to create an economically interpretable econometric framework for policy evaluation and cost-benefit analysis that possesses the best features of the modern treatment effect literature: a clear statement of conditions required to secure identification, as well as robustness and transparency. "Causal effects" or "treatment parameters" are defined in terms of economically interpretable parameters. Counterfactuals and causality are interpreted within the framework of choice-theoretic economic models.

1.1. The relationship of this chapter to the literature on causal inference in statistics

The existing literature on "causal inference" in statistics is the source of inspiration for the recent econometric treatment effect literature and we examine it in detail. The literature in statistics on causal inference confuses three distinct problems that are carefully distinguished in this chapter and in the literature in economics:

[5] An estimand is the parameter defined by the estimator. It is the large sample limit of the estimator, assuming it exists.

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

Task	Description	Requirements
1	Defining the set of hypotheticals or counterfactuals	A scientific theory
2	Identifying parameters (causal or otherwise) from hypothetical population data	Mathematical analysis of point or set identification
3	Identifying parameters from real data	Estimation and testing theory

- Definitions of counterfactuals.
- Identification of causal models from idealized data of population distributions (infinite samples without any sampling variation). The hypothetical populations may be subject to selection bias, attrition and the like. However, all issues of sampling variability are irrelevant for this problem.
- Identification of causal models from actual data, where sampling variability is an issue. This analysis recognizes the difference between empirical distributions based on sampled data and population distributions generating the data.

Table 1 delineates the three distinct problems.

The first problem is a matter 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:

- the rules used to derive a model including whether or not the rules of logic and mathematics are followed;
- its agreement with other theories; and
- its agreement with the evidence.

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.

The second problem 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 problems? This is the identification problem. Two distinct issues that are central to policy evaluation are (1) solving the problem of selection bias and (2) constructing counterfactual states from large samples of data.

The third problem is one of inference in practice. Can one recover a given model or the 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.[6] The boundary between problems two and three is permeable depending on how "the data" are defined.

This chapter focuses on the first two problems. Many applied economists would be unwilling to stop at step 2 and would seek estimators with desirable small sample properties. For a valuable guide to methods of estimation, we direct readers to Chapter 74 (Ichimura and Todd) of this Handbook.

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" and even the definition of certain "causal parameters" because issues of definition, identification and estimation have been confused in the recent literature.

The econometric approach to policy evaluation separates these problems and emphasizes the conditional nature of causal knowledge. Human knowledge advances by developing counterfactuals and theoretical models and testing them against data. The models used are inevitably provisional and conditional on *a priori* assumptions.[7] Blind empiricism leads nowhere. Economists have economic theory to draw on but recent developments in the econometric treatment effect literature often ignore it.

Current widely used "causal models" in epidemiology and statistics are incomplete guides to interpreting data or for suggesting estimators for particular problems. Rooted in biostatistics, 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. Extreme versions of this approach deny causal status to any intervention that cannot in principle be implemented by a practical, real world experiment.

Because the mechanisms determining outcome selection are not modeled in the statistical approach, the metaphor of "random selection" is often adopted. This emphasis

[6] Identification in small samples requires establishing the sampling distribution of estimators, and adopting bias as the criterion for identifiability. This approach is conventional in classical statistics but has fallen out of favor in semiparametric and nonparametric econometrics [see, e.g., Manski (2003)].

[7] See Quine (1951). Thus to quote Quine, "The totality of our so-called knowledge or beliefs, from the most casual matters of geography or history to the profoundest laws of atomic physics ... is a man made fabric which impinges on experience only at the edges ... total science is like a field of force whose boundary conditions are experience ... A conflict with experience on the periphery occasions readjustments in the interior of the field. Reevaluation of some statements require reevaluation of others, because of their logical interconnections ... But the total field is so underdetermined by its boundary conditions, experience, that there is much latitude of choice as to what statements to re-evaluate in the light of any single contrary experience." [Quine (1951)]. We thank Steve Durlauf for suggesting this quote which suggests the awareness of the conditional nature of all knowledge, including causal knowledge, by a leading philosopher.

on randomization – or its surrogates like matching – rules out a variety of alternative channels of identification of counterfactuals from population or sample data. This emphasis has practical consequences because of the conflation of step one with steps two 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. In truth, this is not always so, as we show in this chapter.

One reason why epidemiological and 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 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 economist.

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 [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 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 model phenomena at a deeper level, to understand the causes producing the 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 economic 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 epidemiology and statistics.

The recent literature sometimes contrasts structural and causal models.[8] The contrast is not sharp because the term "structural model" is often not precisely defined. There are multiple meanings for this term, which we clarify in this chapter. The essential contrast between causal models and explicit economic models as currently formulated is in the range of questions that they are designed to answer. Causal models as formulated

[8] See, e.g., Angrist and Imbens (1995) and Angrist, Imbens and Rubin (1996).

in statistics and in the econometric treatment effect literature are typically black-box devices designed to investigate the impact of "treatments" – which are often complex packages of interventions – on some observed set of outcomes in a given environment. Unbundling the components of complex treatments is rarely done. Explicit economic 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) and understanding the "causes of effects" (the goal of the literature building explicit economic models).

By focusing on one narrow black-box question, the treatment effect and natural experiment literatures can avoid many of the problems confronted in the econometrics literature that builds explicit economic models. This is its great virtue. At the same time, it produces parameters that are more limited in application. The parameters defined by instruments or "natural experiments" are often hard to interpret within any economic model. 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 blackbox (understanding the causes of effects), it ties its hands in using information about basic behavioral parameters obtained from other studies, as well as economic intuition to supplement available information in the data in hand. It lacks the ability to provide explanations for estimated "effects" grounded in economics or to conduct welfare economics. When the components of treatments vary across studies, knowledge does not cumulate across treatment effect studies whereas it accumulates across studies estimating common behavioral or technological parameters [see, e.g., the studies of labor supply in Killingsworth (1985), or the parameters of labor demand in Hamermesh (1993), or basic preference and income variability parameters as in Browning, Hansen and Heckman (1999)] which use explicit economic models to collate and synthesize evidence across apparently disparate studies. When the treatment effect literature is modified to address such problems, it becomes a nonparametric version of the literature that builds explicit economic models.

1.2. The plan of this chapter and our other contributions

Our contribution to this Handbook is presented in three chapters. Part I, Section 2, discusses core policy evaluation questions as a backdrop against which to compare alternative approaches to causal inference. A notation is developed and both individual level and population level causal effects are defined. Uncertainty at the individual level is introduced to account for one source of variation across agents in terms of outcomes and choices. We consider alternative criteria used to evaluate policies. We consider a wide variety of parameters of interest that arise in cost benefit analyses and more general analyses of the distribution of policy impacts. This section sets the stage for the rest of the chapter by defining the objects of interest that we study in this chapter.

Section 3 presents some prototypical econometric models that serve as benchmarks and reference points for the discussion throughout all three parts of this chapter. We

review the normal theory model because it is familiar and still widely used and is the point of departure for both the treatment effect and "structural" literatures.

Section 4 defines and discusses causal models, treatment effects, structural models and policy invariant parameters, and analyzes both subjective and objective evaluations of interventions. We also discuss the Neyman (1923)–Rubin (1978) model of causal effects that is influential in statistics and epidemiology.

We review the conventional "structural" (i.e., explicit economic modelling) approach based on latent variable models and recent nonparametric extensions. We define "structural" models and policy-invariant structural parameters using the framework of Hurwicz (1962). A definition of causal models with simultaneous outcomes is presented. The Neyman (1923)–Rubin (1978) model advocated in statistics is compared to explicit econometric models. We discuss how econometric models can be used to construct counterfactuals and answer the range of policy questions discussed in Section 2. We discuss the strengths and limitations of this approach and review recent semiparametric advances in this literature that are relevant to constructing robust policy counterfactuals.

We introduce Marschak's Maxim, implicitly presented in his seminal 1953 paper on policy evaluation.[9] The goal of explicitly formulated econometric models is to identify *policy-invariant* or *intervention-invariant* parameters that can be used to answer classes of policy evaluation questions [see Marschak (1953), Hurwicz (1962), Hansen and Sargent (1980), Lucas and Sargent (1981)].[10] Policy invariance is defined for a class of policy interventions. Policy invariant economic parameters may or may not be interpretable economic parameters. The treatment-effect literature also seeks to identify intervention-invariant parameters for a class of interventions. In this sense the structural and treatment effect literatures share common objectives.

Marschak implicitly invoked a decision-theoretic approach to policy evaluation in noting that for many decisions (policy problems), only *combinations* of explicit economic parameters are required – no single economic parameter need be identified. Hurwicz (1962) refined this idea by noting that to be useful in forecasting policy, the combinations must be invariant to policy variation with respect to the policies being evaluated.

Following Marschak's Maxim, we postulate specific economic questions that are interesting to address and ask what *combinations* of underlying economic parameters or functionals are required to answer them. Answering one question well usually requires fewer assumptions, and places less demands on the data, than answering a wide array of questions – the original goal of structural econometrics. Our approach differs from the approach commonly pursued in the treatment effect and natural experiment literatures

[9] Marschak was a member of the Cowles Commission that developed the first econometric models of policy evaluation. The Cowles Commission approached the policy evaluation problem by constructing models of the economy and then using them to forecast and evaluate policies. This approach is still used today.

[10] The terms "policy invariant" and "structural" are defined precisely in Section 4.8.

by defining a parameter of interest in terms of what *economic question* it answers rather than as the estimand of a favored estimator or instrument.

Section 5 discusses the problem of identification, i.e., the problem of determining models from data. This is task 2 in Table 1. Section 6 exposits identification conditions for the normal model as presented in Section 3.3. It also discusses the recent literature that generalizes the normal model to address concerns raised about nonrobustness and functional form dependence yet preserves the benefits of a structural approach.

Part II of our contribution (Chapter 71 of this Handbook) extends the index function framework, which underlies the modern theory of microeconometrics, to unify the literature on instrumental variables, regression discontinuity methods, matching, control functions and more general selection estimators. Our approach is explicitly nonparametric. We present identifying conditions for each estimator relative to a well-defined set of economic parameters. We initially focus on a two outcome model and then present results for models with multiple outcomes. Bounds are developed for models that are not point identified. We show how these models can be used to address a range of policy problems. We also discuss randomized social experiments. Randomization is an instrumental variable. The focus of Chapter 71 is on mean treatment effects.

Part III, coauthored by Abbring and Heckman (Chapter 72 of this Handbook), considers recent analyses for identifying the distributions of treatment effects. It also discusses new issues that arise in dynamic frameworks when agents are making choices under various information sets that are revealed over time. This takes us into the analysis of dynamic discrete choice models and models for dynamic treatment effects. This section also discusses recent micro-based general equilibrium evaluation frameworks and deals with the important problems raised by social interactions among agents in both market and nonmarket settings.

2. Economic policy evaluation questions and criteria of interest

This section first presents the three central policy evaluation questions discussed in this chapter. We then introduce our notation and define individual level treatment effects. The evaluation problem is discussed in general terms. Population level mean treatment parameters are then defined. Criteria for evaluating distributions of outcomes are presented along with option values. We explicitly account for private and social uncertainty. We discuss, in general terms, the type of data needed to construct the evaluation criteria. Throughout this section we present concrete examples of general points.

2.1. Policy evaluation problems considered in this chapter

Three broad classes of policy evaluation questions are considered in this chapter. Policy evaluation question one is:

P-1 *Evaluating the impact of historical interventions on outcomes including their impact in terms of welfare.*

By historical, we mean interventions actually experienced and documented. A variety of outcomes and welfare criteria might be used to form these evaluations. It is useful to distinguish objective or public outcomes from "subjective" outcomes. Objective outcomes are intrinsically *ex post* in nature. Subjective outcomes can be *ex ante* 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* expected pain and suffering may be different from *ex post* pain and suffering. Agents may also have *ex ante* evaluations of the objective outcomes that may differ from their *ex post* evaluations. By impact, we mean constructing either individual level or population level counterfactuals and their valuations. By welfare, we 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). The welfare evaluations may be *ex ante* or *ex post*.

P-1 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.[11] Focusing exclusively on objective outcomes, this is the problem addressed in the epidemiological and statistical literature on causal inference. A drug trial for a particular patient population is a prototypical problem in the literature. The econometric approach emphasizes valuation of the objective outcome of the trial (e.g., health status) as well as subjective evaluation of outcomes (patient's welfare), and the latter may be *ex post* or *ex ante*.

Most policy evaluation is designed with an eye toward the future and towards informing decisions about new policies and application of old policies to new environments. We distinguish a second task of policy analysis.

P-2 *Forecasting the impacts* (*constructing counterfactual states*) *of interventions implemented in one environment in other environments, including their impacts in terms of welfare*.

Included in these interventions are policies described by generic characteristics (e.g., tax or benefit rates, etc.) 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.[12] 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.

P-3 *Forecasting the impacts of interventions* (*constructing counterfactual states associated with interventions*) *never historically experienced to various environments, including their impacts in terms of welfare*.

[11] The terminology originates with Campbell and Stanley (1963).

[12] Again, this term is due to Campbell and Stanley (1963).

This problem requires that we use past history to forecast the consequences of new policies. It is a fundamental problem in knowledge. Knight (1921, p. 313) succinctly states the problem:

> *"The existence of a problem in knowledge depends on the future being different from the past, while the possibility of a solution of the problem depends on the future being like the past."*

P-3 is a problem that economic policy analysts have to solve daily. Appendix A shows the value of precisely formulated economic models in addressing problems P-2 and P-3. We 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.

2.2. Notation and definition of individual level treatment effects[13]

To evaluate is to value and to compare values among possible outcomes. These are two distinct tasks, which we distinguish in this chapter. We define outcomes corresponding to state (policy, treatment) s for agent ω as $Y(s, \omega)$, $\omega \in \Omega$. The agent can be a household, a firm, or a country. One can think of Ω as a universe of agents with element ω.[14] The ω encompasses all features of agents that affect Y outcomes. $Y(\cdot, \cdot)$ may be generated from a scientific or economic theory. It may be vector valued. The components of $Y(s, \omega)$ may be discrete, continuous or mixed discrete-continuous random variables.

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, we assume that this set is the same for all ω.[15] For each ω, we obtain a collection of possible outcomes given by $\{Y(s, \omega)\}_{s \in \mathcal{S}}$. The set $\mathcal{S}$ may be finite (e.g., there may be J states), countable, or may be defined on the continuum (e.g., $\mathcal{S} = [0, 1]$). 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

[13] Comments from Jaap Abbring were especially helpful in revising this section.

[14] Assume that $\Omega = [0, 1]$. We define random vectors $Y(\omega)$ for $\omega \in \Omega$. We can break out observed and unobserved values $X(\omega)$ and $U(\omega)$, for example.

[15] At the cost of more cumbersome notation, the $\mathcal{S}$ sets can be ω specific. This creates some measure-theoretic problems, and we do not take this more general approach in this chapter. Abbring and Heckman (Chapter 72) relax this assumption when they consider dynamic models and allow for person- and time-period-specific information sets.

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 job training program typically consists of a package of treatments. We might be interested in the package of one (or more) of its components. Thus s_1 may be months of vocational education, s_2 the quality of training and so forth.

The outcomes may be time subscripted as well, $Y_t(s, \omega)$ 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), Abbring and Van den Berg (2003), Abbring and Heckman (2007, Chapter 72), Lechner (2004) and Heckman and Navarro (2007) develop models for dynamic counterfactuals, where time-subscripted and ω-subscripted $\mathcal{S}$ arise as information accrues.

Under this assumption, the **individual treatment effect** for agent ω comparing objective outcomes of treatment s with objective outcomes of treatment s' is

$$Y(s, \omega) - Y(s', \omega), \quad s \neq s', \tag{2.1}$$

where we pick two elements $s, s' \in \mathcal{S}$. This is also called an **individual level causal effect**. This may be a nondegenerate random variable or a degenerate random variable. 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 welfare of participants as well as the objective outcomes [see Heckman and Smith (1998)]. Although statisticians 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, $R(Y(s, \omega), \omega)$, or in terms of planner preferences, R_G, or both types of comparisons might be made for the same outcome and their agreement or conflict evaluated). To simplify the notation, and at the same time allow for more general possibilities for arguments of the valuation function, we usually write $R(Y(s, \omega), \omega)$ as $R(s, \omega)$, suppressing the explicit dependence of R on $Y(s, \omega)$. 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 another possible treatment effect. Holding ω fixed holds all features of the agent fixed except the treatment assigned, s. Since the units of $R(s, \omega)$ 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.

These definitions of treatment effects embody Marshall's (1890) notion of *ceteris paribus* comparisons but now in utility space. A central feature of the econometric approach to program evaluation is the evaluation of subjective evaluations as perceived by decision makers and not just the objective evaluations focused on by statisticians.

The term "treatment" is used in multiple ways in this literature and this ambiguity is sometimes a source of confusion. In its most common usage, a treatment assignment mechanism is a rule $\tau : \Omega \to \mathcal{S}$ which assigns treatment to each ω. The consequences of the assignment are the outcomes $Y(s, \omega)$, $s \in \mathcal{S}$, $\omega \in \Omega$. The collection of these possible assignment rules is $\mathcal{T}$ where $\tau \in \mathcal{T}$. There are two aspects of a policy under this definition. The policy selects who gets what. More precisely, it selects individuals $\omega \in \Omega$ and specifies the treatment $s \in \mathcal{S}$ received.

In this chapter, we offer 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, adoption of a particular technology, and the like. Participation in treatment is usually a choice made by agents. 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, aggregate production technologies, 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 Progressa program in Mexico, and the treatment in that case is choice of schooling with its consequences for earnings.

Our description of treatment assignment recognizes individual choices and constraints and is more suitable for policy evaluation by economists. We specify assignment rules $a \in \mathcal{A}$ which map individuals $\omega \in \Omega$ 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 the like that affect agent choices of treatment.[16] The map a defines the rule used to assign $b \in \mathcal{B}$. It can include deterministic rules which 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

[16] Elements of b can be parameters of tax and benefit schedules that affect individual incentives. A more general setup is possible where ω-specific schedules are assigned to person ω. The cost of this generality is more complicated notation. For simplicity we confine attention to a fixed – but possibly very large – set of parameters defined for all agents.

mechanisms add additional elements of randomness to the environment.[17] Abusing notation, when randomization is used, we will redefine Ω to include this new source of randomness.

Some policies may have the same overall effect on the aggregate distribution of b, but may treat given individuals differently. Under an anonymity postulate, some would judge such policies as equivalent in terms of the constraints (benefits) offered, even though associated outcomes for individuals may be different. Another definition of equivalent policies is in terms of the distribution of aggregate outcomes associated with the treatments. In this chapter, we characterize policies at the individual level, recognizing that sets of $\mathcal{A}$ that are characterized by some aggregate distribution over elements of $b \in \mathcal{B}$ may be what others mean by a policy.[18]

Given $b \in \mathcal{B}$ allocated by constraint assignment mechanism $a \in \mathcal{A}$, agents pick treatments. We 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 $s \in \mathcal{S}$.[19] 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 our 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, as in the econometric approach.[20]

Policy invariance is a key assumption for the study of policy evaluation. It allows analysts to characterize outcomes without specifying how those outcomes are obtained. In our notation, 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. Second, for a given s for agent ω, the mechanism τ by which s is assigned to the agent

[17] Formally, the probability system for the model without randomization is $(\Omega, \sigma(\Omega), \mathcal{F})$ where Ω is the probability space, $\sigma(\Omega)$ is the σ-algebra associated with Ω and $\mathcal{F}$ is the measure on the space. When we account for randomization we need to extend Ω to $\Omega' = \Omega \times \Psi$, where Ψ is the new probability space induced by the randomization, and we define a system $(\Omega', \sigma(\Omega'), \mathcal{F}')$.

[18] Anonymity is a central assumption in the modern income inequality literature. See Foster and Sen (1997).

[19] Note that including $\mathcal{B}$ in the domain of definition of τ is redundant since the map $a : \Omega \to B$ selects an element $b \in \mathcal{B}$. We make b explicit to remind the reader that agents are making choices under constraints.

[20] Thus, under full compliance, $a : \Omega \to \mathcal{S}$ and $a = \tau$, where $\mathcal{B} = \mathcal{S}$.

under assignment mechanism $a \in \mathcal{A}$ is irrelevant for the values assumed by realized outcomes. Both assumptions define what we mean by policy invariance.

Policy invariance allows us to describe outcomes by $Y(s, \omega)$ and ignore features of the policy and choice environment in defining outcomes. If we have to account for the effects of incentives and assignment mechanisms on outcomes, we 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. The following policy invariance assumptions justify collapsing these arguments of $Y(\cdot)$ down to $Y(s, \omega)$.

(PI-1) *For any two constraint assignment mechanisms* $a, a' \in \mathcal{A}$ *and incentives* $b, b' \in \mathcal{B}$, *with* $a(\omega) = b$ *and* $a'(\omega) = b'$, *and for all* $\omega \in \Omega$, $Y(s, \omega, a, b, \tau) = Y(s, \omega, a', b', \tau)$, *for all* $s \in \mathcal{S}_{\tau(a,b)}(\omega) \cap \mathcal{S}_{\tau(a',b')}(\omega)$ *for assignment rule* τ *where* $\mathcal{S}_{\tau(a,b)}(\omega)$ *is the image set for* $\tau(a, b)$. *For simplicity we assume* $\mathcal{S}_{\tau(a,b)}(\omega) = \mathcal{S}_{\tau(a,b)}$ *for all* $\omega \in \Omega$.[21]

This assumption says that for the same treatment s and agent ω, different constraint assignment mechanisms a and a' and associated constraint assignments b and b' produce the same outcome. For example, this assumption rules out the possibility that the act of randomization or the act of pointing a gun at the agent to secure cooperation with planner intentions has an effect on outcomes, given that the agent ends up in s. (PI-1) is a strong assumption and we discuss evidence against it in Chapter 71.

A second 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:

(PI-2) *For each constraint assignment* $a \in \mathcal{A}$, $b \in \mathcal{B}$ *and all* $\omega \in \Omega$, $Y(s, \omega, a, b, \tau) = Y(s, \omega, a, b, \tau')$ *for all* τ *and* $\tau' \in \mathcal{T}$ *with* $s \in \mathcal{S}_{\tau(a,b)} \cap \mathcal{S}_{\tau'(a,b)}$, *where* $\mathcal{S}_{\tau(a,b)}$ *is the image set of* τ *for a given pair* (a, b).

Again, we exclude the possibility of ω-specific image sets $\mathcal{S}_{\tau(a,b)}$ and $\mathcal{S}_{\tau'(a,b)}$. In principle, not all agents ω may be able to attain s for all (a, b) pairs. We invoke this assumption to simplify the analysis and to avoid excess notational and mathematical complexity. Assumption (PI-2) states that the actual mechanism used to assign treatment does not affect the outcomes. It rules out, among other things, social interactions and general equilibrium effects. Abbring and Heckman (Chapter 72) discuss evidence against this assumption.

These invariance postulates are best discussed in the context of specific economic models. We restate these conditions, which are closely related to the invariance conditions of Hurwicz (1962), when we discuss his treatment of policy invariance in Section 4.6 below, after we have specific economic models in hand.

[21] This final assumption can be easily relaxed, but at a major notational cost.

If treatment effects based on subjective evaluations are also considered, we need to broaden invariance assumptions (PI-1) and (PI-2) to produce invariance in rewards for certain policies and assignment mechanisms. It would be unreasonable to claim that utilities $R(\cdot)$ do not respond to incentives. Suppose, instead, that we examine subsets of constraint assignment mechanisms $a \in \mathcal{A}$ that give the same incentives (elements $b \in \mathcal{B}$) to agents, but are conferred by different delivery systems, a. For each $\omega \in \Omega$, define the set of mechanisms delivering the same incentive or constraint b as $\mathcal{A}_b(\omega)$:

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

We allow for the possibility that the set of delivery mechanisms that deliver b may vary among the ω. Let $R(s, \omega, a, b, \tau)$ represent the reward to agent ω from a treatment s with incentive b allocated by mechanism a with an assignment to treatment mechanism τ. To account for invariance with respect to the delivery system, we assume (PI-1) and additional conditions:

(PI-3) *For any two constraint assignment mechanisms $a, a' \in \mathcal{A}$ and incentives $b, b' \in \mathcal{B}$ with $a(\omega) = b$ and $a'(\omega) = b'$, and for all $\omega \in \Omega$, $Y(s, \omega, a, b, \tau) = Y(s, \omega, a', b', \tau)$ for all $s \in \mathcal{S}_{\tau(a,b)}(\omega) \cap \mathcal{S}_{\tau(a',b')}(\omega)$ for assignment rule τ, where $\mathcal{S}_{\tau(a,b)}(\omega)$ is the image set of $\tau(a, b)$ and for simplicity we assume that $\mathcal{S}_{\tau(a,b)}(\omega) = \mathcal{S}_{\tau(a,b)}$ for all $\omega \in \Omega$. In addition, for any mechanisms $a, a' \in \mathcal{A}_b(\omega)$, producing the same $b \in \mathcal{B}$ under the same conditions postulated in the preceding sentence, and for all ω, $R(s, \omega, a, b, \tau) = R(s, \omega, a', b, \tau)$.*

This assumption says, for example, that utilities are not affected by randomization or the mechanism of assignment of constraints. We present evidence against this assumption in Chapter 71.

Corresponding to (PI-2) we have a policy invariance assumption for the utilities with respect to the mechanism of assignment:

(PI-4) *For each pair (a, b) and all $\omega \in \Omega$,*

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

$$R(s, \omega, a, b, \tau) = R(s, \omega, a, b, \tau')$$

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

This assumption rules out general equilibrium effects, social externalities in consumption, etc. in both subjective and objective outcomes. Observe that it is possible to satisfy (PI-1) and (PI-2) but not (PI-3) and (PI-4). For example, randomization may affect subjective evaluations through its effect of adding uncertainty into the decision process but it may not affect objective valuations. We discuss this possibility in Chapter 71 and show that it is empirically important.[22]

[22] We do not develop the third possible case when the roles of R and Y are reversed so that R is invariant and Y is not.

2.2.1. More general criteria

One might compare outcomes in different sets that are ordered. Thus if $Y(s, \omega)$ is scalar income and we compare outcomes for $s \in \mathcal{S}_A$ with outcomes for $s' \in \mathcal{S}_B$, where $\mathcal{S}_A \cap \mathcal{S}_B = \emptyset$, then one might compare Y_{s_A} to Y_{s_B}, where

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

where we 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.[23] Another contrast compares the best choice with the next best choice. To do so, define $s' = \operatorname{argmax}_{s \in \mathcal{S}} \{ Y(s, \omega) \}$ 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. In principle, many different 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.

Social welfare theory constructs aggregates over Ω or nonempty, nonsingleton subsets of Ω [see Sen (1999)]. Let $s_p(\omega)$ denote the $s \in \mathcal{S}_p$ that ω receives under policy p. This is a shorthand notation for the element in $\mathcal{S}_\tau$ determined by the map $p = (a, \tau)$ assigned to agent ω under policy p. A comparison of two policy outcomes $\{s_p(\omega)\}_{\omega \in \Omega}$ and $\{s_{p'}(\omega)\}_{\omega \in \Omega}$, where $p \neq p'$ for some $\omega \in \Omega$, using the social welfare function defined over outcomes $R_G(\{Y(s, \omega), \omega\}_{\omega \in \Omega})$ can be expressed as

$$R_G\big(\big\{Y(s_p(\omega), \omega)\big\}_{\omega \in \Omega}\big) - R_G\big(\big\{Y(s_{p'}(\omega), \omega)\big\}_{\omega \in \Omega}\big).$$

A special case of this analysis is cost-benefit analysis where willingness to pay measures $W(s_p(\omega), \omega)$ are associated with each agent using some compensating or equivalent variation measure for general preferences. The cost-benefit comparison of two policies p and p' is

Cost Benefit:

$$\mathbf{CB}_{p,p'} = \int_\Omega W\big(Y(s_p(\omega), \omega)\big)\, \mathrm{d}\mu(\omega) - \int_\Omega W\big(Y(s_{p'}(\omega), \omega)\big)\, \mathrm{d}\mu(\omega),$$

where p, p' are two different policies and p' may correspond to a benchmark of no policy and $\mu(\omega)$ is the distribution of ω.[24] The Benthamite criterion replaces $W(Y(s(\omega), \omega))$ with $R(Y(s(\omega), \omega))$ in the preceding expressions and integrates utilities across agents.

Benthamite:

$$\mathbf{B}_{p,p'} = \int_\Omega R\big(Y(s_p(\omega), \omega)\big)\, \mathrm{d}\mu(\omega) - \int_\Omega R\big(Y(s_{p'}(\omega), \omega)\big)\, \mathrm{d}\mu(\omega).$$

[23] This analysis could be done for vector $Y(s, \omega)$ provided that $\{Y(s, \omega)\}_{s \in \mathcal{S}}$ is an ordered set.

[24] These willingness-to-pay measures are standard in the social welfare evaluation literature. See, e.g., Boadway and Bruce (1984).

We now discuss the problems that arise in constructing these and other evaluation criteria. This takes us into the problem of causal inference, the second problem delineated in Table 1. We are discussing inference in a population and not in a sample so no issues of sampling variability arise.

2.3. The evaluation problem

Operating purely within the domain of theory, we have assumed a well defined set of individuals $\omega \in \Omega$ and a universe of counterfactuals or hypotheticals for each agent $Y(s, \omega)$, $s \in \mathcal{S}$. Different policies $p \in \mathcal{P}$ give different incentives by assignment mechanism a to agents who are allocated to treatment by a rule $\tau \in \mathcal{T}$. In the absence of a theory, there are no well defined rules for constructing counterfactual or hypothetical states or constructing the assignment to treatment rules.[25] Economic 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. No empirical problem arises in constructing these theoretically possible worlds. Indeed, in forecasting new policies, or projecting the effects of old policies to new environments, some of the $Y(s, \omega)$ may have never been observed for anyone. Different theories produce different $Y(s, \omega)$ and different assignment mechanisms.

The evaluation problem, in contrast with the model construction 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 subfields in economics assume access to different types of data. They also make different assumptions about the underlying models generating the counterfactuals and mechanisms for selecting which counterfactuals are actually observed.

For each policy regime, at any point in time we observe agent ω in some state but not in any of the other states. Thus we do not observe $Y(s', \omega)$ for agent ω if we observe $Y(s, \omega)$, $s \neq s'$. Let $D_p(s, \omega) = 1$ if we observe agent ω in state s under policy regime p. Keeping the policy regime p implicit simplifies the notation so henceforth we work with $D(s, \omega)$ recognizing that it should always be understood as implicitly p subscripted with a constraint assignment mechanism (a) and a treatment assignment mechanism (τ). In this notation, $D(s, \omega) = 1$ implies that $D(s', \omega) = 0$ for $s \neq s'$.

[25] Efforts like those of Lewis (1974) to define admissible counterfactual states without an articulated theory as "closest possible worlds" founder on the lack of any meaningful metric or topology to measure "closeness" among possible worlds.

We observe $Y(s,\omega)$ if $D(s,\omega)=1$ but we do not observe $Y(s',\omega)$, for $s\neq s'$. We keep the p implicit. We can define observed $Y(\omega)$ for a finite or countable $\mathcal{S}$ as

$$Y(\omega)=\sum_{s\in\mathcal{S}} D(s,\omega)Y(s,\omega).^{26} \tag{2.2}$$

Without further assumptions, constructing an empirical counterpart to the individual level causal effect (2.1) is impossible from the data on $(Y(\omega), D(\omega))$, $\omega\in\Omega$. This formulation of the evaluation problem is known as Quandt's switching regression model [Quandt (1958, 1974)] 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 $D(1,\omega)=\mathbf{1}[Y(1,\omega)\geqslant Y(0,\omega)]$.[27] 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.

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 as in a Benthamite criterion) to evaluate policies.[28]

Social experiments attempt to create treatment assignment rules so that $D(s,\omega)$ is random with respect to $\{Y(s,\omega)\}_{s\in\mathcal{S}}$ (i.e., so that receipt of treatment is independent of the outcome of treatment). When agents self-select into treatment, rather than are randomly assigned to it, in general the $D(s,\omega)$ are not independent of $\{Y(s,\omega)\}_{s\in\mathcal{S}}$. Such selection arises in the Roy model example. This selection rule creates the potential for self-selection bias in inference.

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 $Y(s,\omega)$.

Observe that in the 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

[26] In the general case, $Y(\omega)=\int_{\mathcal{S}} D(s,\omega)Y(s,\omega)\,\mathrm{d}s$ where $D(s,\omega)$ is a Dirac function.

[27] Thus $\tau(\omega)=1$ for ω satisfying $Y(1,\omega)\geqslant Y(0,\omega)$ and $\tau(\omega)=0$ for ω satisfying $Y(1,\omega)<Y(0,\omega)$.

[28] One might argue that even a universal policy p like social security has different benefits $b\in\mathcal{B}$ (tax-benefit rates) for persons with different characteristics, so that there is not universal coverage in the sense that we have used it here.

point is more general and receives considerable emphasis in the econometrics literature.[29] Choices by agents provide information on subjective evaluations which are of independent interest.

A central problem analyzed in this chapter is 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 and we investigate both in this chapter. The first avenue, featured in explicitly formulated econometric models, often called "structural econometric analysis," is to model $Y(s,\omega)$ 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 framework models this dependence.[30] Like all scientific models, this approach stresses understanding the factors underlying outcomes and the choice of outcome equations and their dependence. Empirical models based on economic theory pursue this avenue of investigation.[31] 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 (2.1), 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 treatment effect literature focuses exclusively on policy problem P-1 for the subset of outcomes that is observed. It ignores the problem of forecasting a new policy in a new environment (problem P-2), or a policy never previously experienced (problem P-3). Forecasting the effects of new policies is a central task of science, ignored in the treatment effect literature.

2.4. *Population level treatment parameters*

Constructing (2.1) or any of the other individual level parameters defined in Section 2.2 for a given agent is a difficult task because we rarely observe the same agent ω in distinct

[29] See, e.g., Heckman and Smith (1998).

[30] See Heckman and Honoré (1990) and Heckman (2001) for a discussion of this model.

[31] We include in this approach methods based on panel data or more generally the method of paired comparisons as applications of the scientific approach. Under special conditions discussed in Heckman and Smith (1998), we can observe the same agent in states s and s' in different time periods, and can construct (2.1) for all ω.

states s. In addition, some of the states in $\mathcal{S}$ may not be experienced by anyone. The conventional approach in the treatment effect literature is to reformulate the parameter of interest to be some summary measure of the population distribution of treatment effects like a mean or the distribution itself rather than attempting to identify individual treatment effects. It confines attention to subsets of $\mathcal{S}$ that are observed in a particular data set. Thus, the objects of interest are redefined to be distributions of $(Y(j,\omega) - Y(k,\omega))$ over ω or certain means (or quantiles) of the distribution of $(Y(j,\omega) - Y(k,\omega))$ over ω conditional on ω lying in a set $\{\omega\colon\ X(\omega) = x\}$, i.e., conditioning on $X(\omega)$ [Heckman, Smith and Clements (1997)]. They may instead consist of distributions of $Y(j,\omega)$ and $Y(k,\omega)$ separately [Abadie, Angrist and Imbens (2002), Chernozhukov and Hansen (2005)]. Depending on the conditioning sets used, different summary measures of the population distribution of treatment effects are produced. In addition, the standard implicit assumption in the treatment literature is that all states in $\mathcal{S}$ are observed and that assumptions (PI-1) and (PI-2) hold [Holland (1986), Rubin (1986)].

The conventional parameter of interest, and the focus of many investigations in economics and statistics is the average treatment effect or ATE. For program (state, treatment) j compared to program (state, treatment) k, it is

$$\text{ATE}(j,k) = E\big(Y(j,\omega) - Y(k,\omega)\big), \tag{2.3a}$$

where expectations are taken with respect to the distribution of ω. Conditioning on covariates X, which are associated with the observed components of ω, this parameter is

$$\text{ATE}(j,k \mid x) = E\big(Y(j,\omega) - Y(k,\omega) \mid X = x\big). \tag{2.3b}$$

It is the effect of assigning an agent to a treatment – taking someone from the overall population (2.3a) or a subpopulation conditional on X (2.3b) – and determining the mean gain of the move from base state k, averaging over the factors that determine Y but are not captured by X. This parameter is also the effect of moving the economy from a universal policy (characterized by policy k) and moving to a universal policy of j (e.g., from no social security to full population coverage). Such a policy would likely induce social interactions and general equilibrium effects which are assumed away in the treatment effect literature and which, if present, fundamentally alter the economic interpretation placed on the parameter.

A second conventional parameter in this literature is the average effect of treatment on the treated. Letting $D(j,\omega) = 1$ denote receipt of treatment j, the conventional parameter is

$$\text{TT}(j,k) = E\big(Y(j,\omega) - Y(k,\omega) \mid D(j,\omega) = 1\big). \tag{2.4a}$$

For a population conditional on $X = x$ it is

$$\text{TT}(j,k \mid x) = E\big(Y(j,\omega) - Y(k,\omega) \mid D(j,\omega) = 1, X(\omega) = x\big). \tag{2.4b}$$

We present precise models for decision rules below.

These parameters are the mean impact of moving agents from k to j for those people who get treatment, unconditional and conditional on X. It is the benefit part of the

information needed to conduct a cost-benefit evaluation for an existing program. Under certain conditions, it is useful in making "up or out" decisions about an existing program – whether or not the program should be kept or terminated.[32]

A parallel pair of parameters for nonparticipants is treatment on the untreated, where $D(j,\omega)=0$ denotes no treatment at level j:

$$\mathrm{TUT}(j,k)=E\big(Y(j,\omega)-Y(k,\omega)\mid D(j,\omega)=0\big), \tag{2.5a}$$

$$\mathrm{TUT}(j,k\mid x)=E\big(Y(j,\omega)-Y(k,\omega)\mid D(j,\omega)=0, X(\omega)=x\big). \tag{2.5b}$$

These parameters answer the question of how extension of a given program to nonparticipants as a group would affect their outcomes (unconditional and conditional on X, respectively).

The ATE parameter does not condition on a choice. It is policy invariant under conditions (PI-1) and (PI-2). The TT and TUT parameters condition on individual choices and are policy invariant only under the stronger conditions (PI-3) and (PI-4).

Analogous to the pairwise comparisons, we can define setwise comparisons for ordered sets. Thus, in the notation of Section 2.2, we can define the population mean version of the best in $\mathcal{S}_A$ compared with the best in $\mathcal{S}_B$ by

$$E\big(Y_{s_A}(\omega)-Y_{s_B}(\omega)\big),$$

where

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

or we can compare the mean best in the choice set with the mean second best, $E(Y_{s'}(\omega)-Y_{s_B}(\omega))$, where $s'=\operatorname{argmax}_{s\in\mathcal{S}}\{Y(s,\omega)\}$ and $\mathcal{S}_B=\mathcal{S}\setminus\{s'\}$. These parameters can be defined conditional on X.

The population treatment parameters just discussed are average effects: how the average in one treatment group compares to the average in another. The distinction between the marginal and average return is a central concept in economics. It is often of interest to evaluate the impact of marginal extensions (or contractions) of a program. Incremental cost-benefit analysis is conducted in terms of marginal gains and benefits. Let $R(Y(k,\omega),C(k,\omega),\omega)$ be the utility of person ω with outcome $Y(k,\omega)$ and cost $C(k,\omega)$. The **effect of treatment for people at the margin of indifference** (EOTM) between j and k, given that these are the best two choices available is, with respect to personal preferences, and with respect to choice-specific costs $C(j,\omega)$,

$$\begin{aligned}&\mathrm{EOTM}^R(j,k)\\&\quad=E\left(Y(j,\omega)-Y(k,\omega)\;\middle|\;\begin{array}{l}R(Y(j,\omega),C(j,\omega),\omega)=R(Y(k,\omega),C(k,\omega),\omega);\\ R(Y(j,\omega),C(j,\omega),\omega)\\ R(Y(k,\omega),C(k,\omega),\omega)\end{array}\right\}\geqslant R(Y(l,\omega),C(l,\omega),\omega)\\&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad l\neq j,k\Bigg).\end{aligned} \tag{2.6}$$

[32] See, e.g., Heckman and Smith (1998).

This is the mean gain to agents indifferent between j and k, given that these are the best two options available. In a parallel fashion, we can define $\text{EOTM}^{R_G}(Y(j,\omega) - Y(k,\omega))$ using the preferences of another agent (e.g., the parent of a child; a paternalistic bureaucrat, etc.).

An analogous parameter can be defined for mean setwise comparisons. Thus we can define two versions of EOTM:

$$\text{EOTM}^{R}(s_A, s_B) = E\left(Y_{s_A} - Y_{s_B} \;\middle|\; \begin{array}{l} R(Y(s_A,\omega), C(s_A,\omega), \omega) \\ \quad = R(Y(s_B,\omega), C(s_B,\omega), \omega) \end{array}\right),$$

where s_A and s_B are distinct elements and $A \cap B = \emptyset$, and

$$\text{EOTM}^{R}\big(\{s'\}, \mathcal{S} \setminus \{s'\}\big) = E\left(Y_{s'} - Y_{s_B} \;\middle|\; \begin{array}{l} R(Y(s',\omega), C(s',\omega), \omega) \\ \quad = R(Y(s_B,\omega), C(s_B,\omega), \omega) \end{array}\right),$$

where s_B is the optimal choice in the set of $\mathcal{S} \setminus \{s'\}$. Again, these parameters can be defined conditional on $X = x$. Other setwise comparisons can be constructed. A generalization of this parameter called the **marginal treatment effect**, introduced into the evaluation literature by Björklund and Moffitt (1987), further developed in Heckman and Vytlacil (1999, 2000, 2005) and defined precisely in Chapter 71 of this Handbook, plays a central role in organizing and interpreting a wide variety of econometric estimators in this chapter.[33]

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. We consider distributional parameters in the next subsection.

Of special interest in policy analysis is the **policy relevant treatment effect**. It is the effect on aggregate outcomes of one policy regime $p \in \mathcal{P}$ compared to the effect of another policy regime. For it to be an interesting parameter, we assume (PI-1) and (PI-2) but not necessarily (PI-3) and (PI-4).

$$\textbf{PRTE:} \quad E_p\big(Y(s,\omega)\big) - E_{p'}\big(Y(s,\omega)\big), \quad \text{where } p, p' \in \mathcal{P},$$

where the expectations are taken over different spaces of policy assignment rules. This parameter is a version of a Benthamite policy criterion.

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 we can identify the mean of $Y(j,\omega)$ and the mean of $Y(k,\omega)$ from an experiment

[33] There are technical measure theoretic issues regarding whether EOTM is uniquely defined. They are discussed in Chapter 71.

where j is the treatment and k is the baseline, we can form the average treatment effect for j compared to k (2.3a). These can be formed over two different groups of agents. By a similar argument, we can form the treatment on the treated parameter (TT) (2.4a) or (TUT) (2.5a) 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 Hawthorne effects or randomization (disruption) bias. See Heckman (1992) and the discussion in Chapter 71.

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 economic parameters. The elevation of population means as the central population level "causal" parameters promotes randomization as an ideal estimation method. This focus on means converts a metaphor for outcome selection – randomization – into an ideal. We next turn to a discussion of distributions of counterfactuals.

2.5. *Criteria of interest besides the mean: 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 modern political economy [Persson and Tabellini (2000)] seeks to know the proportion of agents who benefit from policy regime p compared with p'. 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

$$\Pr\big(Y\big(s_p(\omega),\omega\big) \geqslant Y\big(s_{p'}(\omega),\omega\big)\big).$$

This is called the **voting criterion**. 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

$$\Pr\big(Y(j,\omega) \geqslant Y(k,\omega)\big).$$

Under (PI-1) and (PI-2) this is the same across all policy regimes.[34] We might be interested in the quantiles of $Y(s_p(\omega),\omega) - Y(s_{p'}(\omega),\omega)$ or of $Y(j,\omega) - Y(k,\omega)$ 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.

The traditional literature on program evaluation focuses its attention on mean impacts. When the outcomes are in value units, these can be used to measure the effect of

[34] See Abbring and Heckman (Chapter 72). General equilibrium effects invalidate assumptions (PI-1) and (PI-2).

a program on total social output and are the basis of efficiency analyses. The implicit assumption of the traditional cost-benefit literature is that "a dollar is a dollar," regardless of who receives it.[35]

An emphasis on efficiency to the exclusion of distribution is not universally accepted.[36] An emphasis on efficiency is premised on the assumption that distributional issues are either irrelevant or that they are settled by some external redistribution mechanism using a family or a social welfare function.

Outcomes from many activities like health programs, educational subsidies and training programs are not transferrable. Moreover, even if all program outputs can be monetized, the assumption that a family or social welfare function automatically settles distributional questions in an optimal way is questionable. Many programs designed to supply publicly provided goods are properly evaluated by considering the incidence of their receipt and not the aggregate of the receipts. Hence counterfactual distributions are required. Distributions of counterfactuals are also required in computing option values of social programs, which we discuss next.

2.6. Option values

Voluntary social programs confer options, and these options can change threat points and bargaining power, even if they are not exercised.[37] It is, therefore, of interest to assess these option values. The most interesting versions of option values require knowledge of the joint distribution of potential outcomes. We consider the analysis of treatments offered within a policy regime. Persons offered a subsidized job may take it or opt for their best unsubsidized alternative. The option of having a subsidized alternative job will in general convey value. The option may be conferred simply by eligibility for a program or it may be conferred only on participants. The program creates an option for participants, if prior to participating in it, their only available option comes from the distribution of $Y(k, \omega)$, say F_k. Following or during participation in the program, the individual has a second option $Z(\omega)$ drawn from distribution F_Z. If both options are known prior to choosing between them, and agents are outcome maximizers, then the observed outcome $Y(j, \omega)$ is the maximum of the two options, $Y(j, \omega) = \max(Y(k, \omega), Z(\omega))$. The option $Z(\omega)$ may be available only during the period of program participation, as in a wage subsidy program, or it may become a permanent feature of the choice set as when a marketable skill is acquired. It is useful to distinguish the case where the program offers a distribution F_Z from which new offers are received each period from the case where a permanent $Z(\omega)$ value is created. Much of the literature on program evaluation implicitly equates $Z(\omega)$ with $Y(j, \omega)$. This is valid only if treatment is an irreversible condition that supplants $Y(k, \omega)$ or else

[35] See Harberger (1971).
[36] See Little and Mirrlees (1974).
[37] See, e.g., Osborne (2004).

$Z(\omega) \geqslant Y(k, \omega)$ for all ω so that agents who take the treatment use the skills conferred by it. In either case, agents offered $Z(\omega)$ always choose $Z(\omega)$ over $Y(k, \omega)$ or are indifferent, so $Y(j, \omega) \equiv Z(\omega)$ and the estimated distribution of $Y(j, \omega)$ is equivalent to the estimated distribution of $Z(\omega)$. In general it is useful to determine what a program offers to potential participants, what the offer is worth to them, and to distinguish the offered option from the realized choice.

The expected value of having a new option $Z(\omega)$ in addition to $Y(k, \omega)$ is

$$\text{(OP-1)} \quad E(\max(Y(k, \omega), Z(\omega))) - E(Y(k, \omega)),$$

assuming that potential participants in a program can choose freely between $Y(k, \omega)$ and $Z(\omega)$. This is the difference in expected outcomes between a two-option world and a one-option world, assuming that both are known at the time the choice between them is made. It is useful to distinguish the opportunities created from the program, $Z(\omega)$, from the options selected. The program extends opportunities to potential participants. Providing a new opportunity that may be rejected may improve the average outcome among agents who choose $Y(k, \omega)$ over $Z(\omega)$ through affecting the distribution of the $Y(k, \omega)$ offered to the agents.

For example, the outside option can improve bargaining power. If a housewife receives an outside job offer, her bargaining power at home may increase. If a program gives participants a second distribution from which they receive a new draw each period, and if realizations of the pair $(Y(k, \omega), Z(\omega))$ in each future period are independently and identically distributed, then the addition to future wealth of having access to a second option in every period is

$$\frac{1}{r}\big[E\big(\max\big(Y(k, \omega), Z(\omega)\big)\big) - E\big(Y(k, \omega)\big)\big],$$

where r is the interest rate. If Z is available only for a limited time period, as would be the case for a job subsidy, (OP-1) is discounted over that period and the expression should be appropriately modified to adjust for the finite life.

If the realizations $(Y(k, \omega), Z(\omega))$ are not known at the time when decisions to exercise the option are made, (OP-1) is modified to

$$\text{(OP-2)} \quad \max(E(Y(k, \omega) \mid \mathcal{I}_\omega), E(Y(j, \omega) \mid \mathcal{I}_\omega)) - E(Y(k, \omega) \mid \mathcal{I}_\omega),$$

where these expectations are computed against agent ω's information set $\mathcal{I}_\omega$.[38] Constructing these option values in general requires knowing the joint distribution of $Z(\omega)$ and $Y(k, \omega)$, and cannot be obtained from means or from social experiments which only identify marginal distributions. We now turn to a systematic accounting of uncertainty.

[38] A third definition of option value recognizes the value of having uncertainty resolved at the time decisions to choose between $Z(\omega)$ and $Y(k, \omega)$ are made. That definition is

$$\text{(OP-3)} \quad E(\max(Z(\omega), Y(k, \omega))) - \max(E(Z(\omega) \mid \mathcal{I}_\omega), E(Y(k, \omega) \mid \mathcal{I}_\omega)) = \text{(OP-1)–(OP-2)}.$$

2.7. Accounting for private and social uncertainty

Systematically accounting for uncertainty introduces additional considerations that are central to economic analysis but that are ignored in the 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 treatment effect literature does not.

In the literature on welfare economics and social choice, one form of decision-making under uncertainty plays a central role. The "Veil of Ignorance" of Vickrey (1945, 1961) and Harsanyi (1955, 1975) postulates that agents are completely uncertain about the positions of individuals in the distribution of outcomes under each policy, or should act as if they are completely uncertain, and they should use expected utility criteria (Vickrey–Harsanyi) or a maximin strategy [Rawls (1971)] to evaluate welfare under alternative policies. Central to this viewpoint is the anonymity postulate that claims the irrelevance of any particular agent's outcome to the overall evaluation of social welfare. This form of ignorance is sometimes justified as an ethically correct position that captures how an objectively detached observer should evaluate alternative policies, even if actual participants in the political process use other criteria. An approach based on the Veil of Ignorance is widely used in applied work in evaluating different income distributions [see Foster and Sen (1997)]. It is empirically easy to implement because it only requires information about the marginal distributions of outcomes produced under different policies. If the outcome is income, policy j is preferred to policy k if the income distribution under j stochastically dominates the income distribution under k.[39]

An alternative criterion is required if agents act in their own self-interest, or in the interest of certain other groups (e.g., the poor, the less able) and have at least partial knowledge about how they (or the groups they are interested in) will fare under different policies. The outcomes in different regimes may be dependent, so that agents who benefit under one policy may also benefit under another [see Carneiro, Hansen and Heckman (2001, 2003), Cunha, Heckman and Navarro (2005, 2006)].

Because agents typically do not possess perfect information, the simple voting criterion that assumes perfect foresight over policy outcomes that is discussed in Section 2.5 may not accurately characterize choices. It requires modification. Let $\mathcal{I}_\omega$ denote the information set available to agent ω. He or she evaluates policy j against k using that

[39] See Foster and Sen (1997) for a definition of stochastic dominance.

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

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

The proportion of people who prefer j is

$$PB(j \mid j,k) = \int \mathbf{1}\big[E\big[R\big(Y(j,\omega),\omega\big) \mid \mathcal{I}_\omega\big] \geqslant E\big[R\big(Y(k,\omega),\omega\big) \mid \mathcal{I}_\omega\big]\big]\,\mathrm{d}\mu(\mathcal{I}_\omega), \tag{2.7}$$

where $\mu(\omega)$ is the distribution of ω in the population whose preferences over outcomes are being studied.[40,41] The voting criterion presented in Section 2.5 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 $Y(k)$. Abbring and Heckman (Chapter 72) offer an example of the application of this criterion.

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). It is useful to index the information set $\mathcal{I}_\omega$ by t, $\mathcal{I}_{\omega,t}$, to recognize that information about the outcomes of policies may accrue over time. *Ex ante* and *ex post* assessments of a voluntary program need not agree.

Ex post assessments of a program through surveys administered to agents who have completed it [Katz et al. (1975), Hensher, Louviere and Swait (1999)], may disagree with *ex ante* assessments of the program. Both may reflect honest valuations of the program. 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 is 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,

$$R\big(Y(j,\omega),\omega\big) \geqslant E\big(R\big(Y(k,\omega),\omega\big) \mid \mathcal{I}_\omega\big), \tag{2.8}$$

where the information is post-treatment. Survey questionnaires about "client" satisfaction with a program may capture subjective elements of program experience not captured by "objective" measures of outcomes that usually exclude psychic costs and benefits. Heckman, Smith and Clements (1997) present evidence on this question. Carneiro,

[40] Agents would not necessarily vote "honestly", although in a binary choice setting they do and there is no scope for strategic manipulation of votes. See Moulin (1983). *PB* is simply a measure of relative satisfaction and need not describe a voting outcome when other factors come into play.

[41] See Cunha, Heckman and Navarro (2006) for computations regarding both types of joint distributions.

Hansen and Heckman (2001, 2003), Cunha, Heckman and Navarro (2005, 2006) and Heckman and Navarro (2007) develop econometric methods for distinguishing *ex ante* from *ex post* evaluations of social programs, which are surveyed in Abbring and Heckman (Chapter 72).

2.8. The data needed to construct the criteria

Four ingredients are required to implement the criteria discussed in this section: (a) private preferences, including preferences over outcomes by the decision maker; (b) social preferences, as exemplified by the social welfare function; (c) distributions of outcomes in alternative states, and for some criteria, such as the voting criterion, *joint* distributions of outcomes *across* policy states; and (d) *ex ante* and *ex post* information about outcomes. Cost benefit analysis only requires information about means of measured outcomes and for that reason is easier to implement. The statistical treatment effect literature largely focuses on *ex post* means, but recent work in econometrics focuses on both *ex ante* and *ex post* distributions [see Carneiro, Hansen and Heckman (2001, 2003), Cunha, Heckman and Navarro (2005, 2006), Heckman, Smith and Clements (1997)]. This chapter focuses on methods for producing ingredients (c) and (d). There is a large literature on recovering private preferences [see, e.g., Chapter 67 (Blundell, MaCurdy and Meghir) of this Handbook] and on recovering technology parameters [see, e.g., Chapter 62 (Reiss and Wolak); and Chapter 61 (Ackerberg, Benkard, Berry and Pakes) of this Handbook]. The rich set of questions addressed in this section contrasts sharply with the focus on mean outcomes in epidemiology and statistics which ignores private and social preferences and distributions of outcomes. Carneiro, Hansen and Heckman (2001, 2003), Cunha, Heckman and Navarro (2005, 2006) and Heckman and Navarro (2007) present methods for extracting private information on outcomes and their evolution over time. We now present some examples of explicit economic models drawing on core elements of modern econometrics. We build on these examples throughout our chapter.

3. Roy and generalized Roy examples

To make the discussion more specific and to introduce a parametric version of the framework for discrete choice with associated outcomes that motivates the analysis in this chapter, we introduce versions of the Roy (1951) and generalized Roy models, define various treatment effects and introduce uncertainty into the analysis. We show how the Roy model and its extensions solve policy problems P-1–P-3 that are the focus of this chapter. We first develop the generalized Roy framework for a setting of perfect certainty, specialize it to the two-outcome case, and then introduce uncertainty. We produce some normal theory examples because normality is conventional and easy to link to standard regression theory. The analyses reported in Section 6, Appendix B and Chapter 71 relax the normality assumption.

3.1. A generalized Roy model under perfect certainty

Suppose that there are $\bar{S}$ states associated with different levels of schooling, or some other outcome such as residence in a region, or choice of technology. Associated with each choice s is a valuation of the outcome of the choice $R(s)$, where R is the valuation function and s is the state. Define Z as individual variables that affect choices. Each state may be characterized by a bundle of attributes, characteristics or qualities $Q(s)$ that fully characterize the state. If $Q(s)$ fully describes the state, $R(s) = R(Q(s))$. To simplify the notation, we do not use the ω notation in this section, but keep it implicit.

Suppose that $R(s)$ can be written in additively separable form in terms of deterministic and random components. We assume that the Z is observed. Let ν denote unobserved components as perceived by the econometrician. In this notation,

$$R(s) = \mu_R(s, Z) + \eta(s, Z, \nu), \tag{3.1}$$

where $\mu_R(s, Z)$ is the deterministic component of the utility function expressed in terms of observed variables Z and $\eta(s, Z, \nu)$ represents unobservables from the point of view of the econometrician (recall that we assume that there is no uncertainty facing the agent).[42] McFadden (1981) describes a large class of discrete choice models that can be represented in this form. Additive separability is convenient but not essential [Matzkin (1992, 1993, 1994)]. An example of these models is a random coefficient choice model where $R(s) = \gamma_s' z = \bar{\gamma}_s' z + \nu_s' z$, where $\bar{\gamma}_s$ is the mean of γ_s and ν_s is the deviation of γ_s from its mean. In the McFadden (1974) model, $\mu_R(s, z) = \bar{\gamma}_s' z + \nu_s$, where ν_s is independent of Z and also independent of s. In this abstract notation, the characteristics of choice s are embedded in the definition of γ_s. A more explicit version would write $\gamma_s = \gamma(Q(s))$, where $Q(s)$ are the characteristics of choice s. To simplify notation we write $\eta(s, Z, \nu)$ as $\eta(s)$.

Associated with each choice is outcome $Y(s)$ which may be vector valued. These outcomes can depend on X. For simplicity and familiarity we work with the scalar case. Following Carneiro, Hansen and Heckman (2003) and Heckman and Navarro (2007), we can accommodate the case where $Y(s)$ is a vector of continuous, discrete and mixed discrete-continuous outcomes. Again, for simplicity we drop "ω" and assume an additively separable case where $\mu_Y(s, X)$ is a deterministic function expressed in terms of observables and $U(s, X, \varepsilon)$, $s = 1, \ldots, \bar{S}$, are unobservables:

$$Y(s) = \mu_Y(s, X) + U(s, X, \varepsilon).$$

We leave the details of constructing the random variables $\eta(s, Z, \gamma)$ and $U(s, X, \varepsilon)$ for a later section of this chapter. For now one could work with the shorthand notation $U(s, X, \varepsilon) = U(s)$ and $\eta(s, Z, \gamma) = \eta(s)$.

[42] One definition of $\mu_R(s, Z)$ is $\mu_R(s, Z) = E\big[R(s) \mid Z\big]$, but other definitions are possible. The "structural" approach derives $\mu_R(s, Z)$ from economic theory.

This framework serves as a benchmark index model against which we can measure the recent contributions and limitations of the treatment effect literature. The chapters in the Handbook series by McFadden (1984), Heckman and MaCurdy (1986), Matzkin (1994), Blundell, MaCurdy and Meghir (2007), Reiss and Wolak (2007), Ackerberg et al. (2007), and Athey and Haile (2007) exposit detailed econometric analyses of specific economic models that are based on versions of this structure and extensions of it. Economically well-posed econometric 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. These explicit features make these models, like all scientific models, useful vehicles for interpreting empirical evidence using economic theory, for collating and synthesizing evidence across studies using economic theory, for measuring the welfare effects of policies, and for forecasting the welfare and direct effects of previously implemented policies in new environments and the effects of new policies.

The set of possible treatments $\mathcal{S}$ is $\{1, \ldots, \bar{S}\}$, the set of state labels. The set of counterfactual outcomes is $\{Y(s, X)\}_{s\in\mathcal{S}}$. The treatment assignment mechanism is produced by utility maximization:

$$D(j) = 1 \quad \text{if } \operatorname{argmax}_{s\in\mathcal{S}}\big\{R(s)\big\} = j, \tag{3.2}$$

where in the event of ties, choices are made by a flip of a coin. Thus agents *self select* into treatment and the probabilities of selection which are defined at the individual level are either zero or one for each agent (agents choose outcomes with certainty). Appendix B presents a proof of nonparametric identification of this generic model.

Other mechanisms for selection into sector s could be entertained. In the background, policy "p", under which choices are made, is kept implicit. Policies can operate to change Z, X, and the distributions $\eta(s, Z, \nu)$, $U(s, X, \varepsilon)$. Section 5 presents a more detailed analysis of policy regimes. Operating within a policy regime, and a particular treatment selection rule, we do not have to take a position on assumptions (PI-3) and (PI-4), which are assumptions about outcomes across policy regimes and across assignment rules within policy regimes. We next present examples of these models. We also introduce examples of models with uncertainty.

3.1.1. Examples of models that can be fit into this framework

Scalar income The original static Roy model (1951) writes $Y(j)$ as scalar income in sector j. For instance, sectors can be regions, industries [Heckman and Sedlacek (1985)], schooling levels [Willis and Rosen (1979), Carneiro, Hansen and Heckman (2003)] or union status [Lee (1978)]. See Heckman (2001) for a survey of these applications.

In the original setup, $R(j) \equiv Y(j)$, $Z = X$ and $Y(j)$ is scalar income in sector j so agents are income maximizers. In extensions of this model, there are sector-specific costs $C(j)$ which may depend on $Z = (X, W)$, $R(j) = Y(j) - C(j)$. This allows for nonpecuniary components as in Heckman and Sedlacek (1985), Carneiro, Hansen

and Heckman (2003), Cunha, Heckman and Navarro (2005, 2006) and others, or tuition costs as in Willis and Rosen (1979). Policies may operate on costs or returns. Agents may be uncertain about future income when they make their choices so the decision rule is to go to the sector if $E(Y(1) - C(1) - Y(0) \mid \mathcal{I}) \geqslant 0$. *Ex post* returns are $(Y(1)-C(1)-Y(0))$. See Carneiro, Hansen and Heckman (2003), and Cunha, Heckman and Navarro (2005, 2006).

Choice of technology In this application, the profit-maximizing firm faces J technologies. $Y(j)$ is output. $F_j : X \to Y(j)$ maps inputs into outputs for technology j, assumed to be strictly concave and twice differentiable. There is a cost of inputs, $C(j)$, possibly including fixed cost components. As before, let $Z = (X, W)$. Assume that profit, $R(j)$, is maximized for each technology so $R(j) = \max_X\{F_j(X) - C(j, X, W)\}$, and

$$D(j) = 1 \quad \text{if } \operatorname{argmax}_\ell \big\{R(\ell)\big\} = j.$$

The potential outcome vector is $(Y(j), R(j), X(j), C(j))$ where $X(j)$ is the input vector chosen if j is chosen. In this example, utility, $R(j)$, is profit and firms are assumed to pick the technology with the highest profit. Policies operate on costs, profit taxes, and on returns [see Pitt and Rosenzweig (1989)].

Dynamic education choices Following Eckstein and Wolpin (1989, 1999), Keane and Wolpin (1997) and Heckman and Navarro (2007), we may explicitly account for information updating at attained schooling level s. We introduce uncertainty. Let $E(R(s, s+1) \mid \mathcal{I}_s)$ be the value of continuing on to the next schooling level given that an agent has already attained s and possesses information set $\mathcal{I}_s$. This value includes the options opened up by taking s. $D_{s,s+1}(\mathcal{I}_s) = 1$ if an agent continues from level s to level $s+1$. $D_{s,s+1}(\mathcal{I}_s) = \mathbf{1}[E(R(s, s+1) \mid \mathcal{I}_s) \geqslant 0]$ and equals 0 otherwise. Associated with each outcome is a payoff stream of future income and option values associated with the choice Y_{s+1}. Abbring and Heckman (Chapter 72) discuss dynamic counterfactuals and dynamic discrete choice.

Many other examples could be given. The literature on estimation and identification in structural models is active [see Rust (1994), Geweke and Keane (2001), Aguirregabiria (2004), Heckman and Navarro (2007)]. The unifying theme underlying all of these models is that latent variables (the utilities or value functions) generate observed outcomes. Since outcomes (or agent-predicted outcomes) affect choices, there is selection bias. To make the discussion specific and have a model in hand, we exposit a normal theory generalized Roy model in Section 3.3. First we use this framework to define treatment effects.

3.2. Treatment effects and evaluation parameters

The **individual level treatment effect** (2.1) for objective outcomes is

$$Y(s) - Y(s') = \mu_Y(s, X) - \mu_Y(s', X) + U(s) - U(s'). \tag{3.3}$$

The **subjective evaluation individual treatment effect** of program s compared to program s' is

$$R(s) - R(s') = \mu_R(s, Z) - \mu_R(s', Z) + \eta(s) - \eta(s')$$

in the metric of the valuation function. An alternative measure of the *relative* subjective evaluation of the program is

$$D(s, s', Z) = \mathbf{1}\big[R(s) \geqslant R(s')\big].$$

If $D(s, s') = 1$, the agent (weakly) prefers s over s'.

As in Section 2, one can define set-wise comparisons of treatment effects. Thus one can compare the outcome of the best with the outcome of the next best as in Dahl (2002), defining

$$s' = \operatorname{argmax}_{s \in \mathcal{S}}\big\{Y(s)\big\} \quad \text{and} \quad \mathcal{S}_B = \mathcal{S} \setminus \{s'\}$$

so that the treatment effect comparing the best to the next best is

$$Y(s') - Y(s_B).$$

Other comparisons can be made. Instead of private preferences, there may be social preferences of the "planner" defined over the choices of the individuals. Cost benefit criteria would be defined in a corresponding fashion.

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

The **selection problem** arises because we only observe certain agents in any state. Thus we observe $Y(s)$ only for agents for whom $D(s) = 1$. 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.

We now define the population treatment parameters using this framework. Comparing s with s', $\mathrm{ATE}(s, s' \mid X) = \mu_Y(s, X) - \mu_Y(s', X)$. Treatment on the treated for those choosing between s and s' given X, Z is

$$\begin{aligned} &E\big(Y(s) - Y(s') \mid X, Z, D(s) = 1\big) \\ &\quad = \mathrm{TT}(s, s' \mid X, Z) \\ &\quad = \mu_Y(s, X) - \mu_Y(s', X) + E\big[U(s) - U(s') \mid X, Z, D(s) = 1\big], \end{aligned}$$

where the final term is the sorting gain that arises from agents selecting into the treatment. ATE and TT can be defined for the best compared to the next best.

$$\begin{aligned} \mathrm{ATE}(s, s_B \mid X, Z) &= \mu_Y(s, X) - E\Big[\max_{j \in \mathcal{S} \setminus \{s\}} \big\{Y(j)\big\} \mid X, Z\Big], \\ \mathrm{TT}(s, s_B \mid X, Z) &= \mu_Y(s, X) + E\big(U(s) \mid D(s) = 1, X, Z\big) \\ &\quad - E\Big[\max_{j \in \mathcal{S} \setminus \{s\}} \big\{Y(j)\big\} \mid D(s) = 1, X, Z\Big]. \end{aligned}$$

The effect of treatment given X for agents at the margin of participation between s and s' (EOTM) using the analysis of Section 2.4 is

$$\text{EOTM}(s, s') = \mu_Y(s, X) - \mu_Y(s', X) + E\big[U(s) - U(s') \mid R(s) = R(s')\big],$$

where $R(s), R(s') \geqslant R(k)$, $s, s' \neq k$. We can define setwise versions of this parameter as well. Using the model, we can also compute the distributional criteria introduced in Section 2.5, e.g., the proportion of people who benefit from being in s compared to s':

$$\Pr\big(R(s) \geqslant R(s') \mid Z = z\big).$$

We can form quantiles of the outcome distribution and evaluate the quantile treatment effects [e.g., Chernozhukov and Hansen (2005)]. Letting $q^s(\nu)$ be the νth quantile of the $Y(s)$ distribution, the quantile treatment effects for a lth quantile are $q^s(l) - q^{s'}(l)$. From the agent preferences, and the outcome distributions we can form all of the treatment effects discussed in Section 2 for environments of perfect certainty.

For a known model, we can answer policy question P-1 within the sample used to fit the model. Thus we can solve the problem of internal validity by fitting the model (3.1) and (3.2). Policy question P-2 involves extrapolating the model to new regions of X, Z. This can be solved using parametric functional forms (e.g., $\mu_Y(s, X) = X\beta_s$ and $\mu_R(s, Z) = Z\gamma_s$). If $U(s)$ and $\eta(s)$ are independent of X, Z, the task is simplified. If they are not independent, then it is necessary to model the dependence of $U(s), \eta(s)$ on (X, Z) over the new support of (X, Z).

Policy problem P-3 entails the evaluation of new outcome states never previously experienced, for example a new element s. As suggested by the quotation from Frank Knight cited in Section 2, one avenue of solution is to characterize β_s and γ_s as functions of baseline characteristics that describe all programs $\beta_s = \beta(Q(s))$, $\gamma_s = \gamma(Q(s))$ and to characterize the dependence of $U(s), \eta(s)$ on $Q(s)$. Provided that we can define a new program s' as a combination of the characteristics of previous programs, and $\beta(Q(s)), \gamma(Q(s))$ (and the distributions of $U(s), \eta(s)$) are defined over supports that include $Q(s)$, we can solve P-3. We provide a specific example of this approach in the next subsection.

3.3. A two-outcome normal example under perfect certainty

To make the discussion concrete, it is helpful to exposit a prototypical model of choice and associated outcomes. The Roy model (1951) and its extensions [Gronau (1974), Heckman (1974), Willis and Rosen (1979), Heckman (1990), Carneiro, Hansen and Heckman (2003)] are at the core of microeconometrics.

Consider the following simple version of the Roy model. Persons start off in sector "0" (e.g., primary schooling). To simplify expressions, we write $Y(s)$ as Y_s in this section, and in other places where it is convenient to do so. We create parallel notation for $U(s) = U_s$. The variables Y_1 and Y_0 can be interpreted as the outcomes from being

in sectors 1 and 0, respectively. We model these as

$$Y_1 = X\beta_1 + U_1, \tag{3.4a}$$

$$Y_0 = X\beta_0 + U_0, \tag{3.4b}$$

and associated costs (prices) as a function of W,

$$C = W\beta_C + U_C. \tag{3.4c}$$

In a schooling example, tuition and distance to school would be candidates for inclusion in W. The valuation of "1" relative to "0" is $R = Y_1 - Y_0 - C$. Substituting from (3.4a)–(3.4c) into the expression for R, we obtain the relative evaluation of outcome "1" versus outcome "0" as

$$R = X(\beta_1 - \beta_0) - W\beta_C + U_1 - U_0 - U_C.^{43}$$

Sectoral choice is indicated by D, where $D = 1$ if the agent selects 1, $D = 0$ otherwise:

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

We define $\upsilon = (U_1 - U_0 - U_C)$, $Z = (X, W)$ and $\gamma = (\beta_1' - \beta_0', -\beta_C')$ so we can write $R = Z\gamma + \upsilon$. The generalized Roy model assumes that (recalling $Z = (X, W)$)

(i) $Z \perp\!\!\!\perp (U_0, U_1, U_C)$ (independence),
(ii) $(U_0, U_1, U_C) \sim \mathcal{N}(0, \Sigma_{\text{GR}})$ (normality),

where $\mathcal{N}(0, \Sigma_{\text{GR}})$ is normal with mean zero and variance–covariance matrix Σ_{GR} and "GR" stands for the generalized Roy model.

From its definition, $E(\upsilon) = 0$. The **Roy model** is the special case where $\beta_C = 0$ and $U_C = 0$, so choices are made solely on the basis of income, $R = Y_1 - Y_0$. The **extended Roy model** sets $\beta_C \neq 0$, but $U_C = 0$ so choices are made on net income subtracting costs but the determinants of the cost components (W) are observed by the analysts.

For the **generalized Roy model**, the probability of selecting treatment (outcome) 1 is

$$\Pr(R \geqslant 0 \mid Z = z) = \Pr(\upsilon \geqslant -z\gamma) = \Pr\left(\frac{\upsilon}{\sigma_\upsilon} \geqslant \frac{-z\gamma}{\sigma_\upsilon}\right) = \Phi\left(\frac{z\gamma}{\sigma_\upsilon}\right),$$

where Φ is the cumulative distribution function of the standard normal distribution and the last result follows from the symmetry of standard normal variables around zero. The choice probability is sometimes called the "propensity score" by statisticians. Higher values of the index lead to higher values of the probability of participation; $z\gamma$ is the mean scale utility function. Higher values of $z\gamma$ correspond to higher values of net utility from choosing treatment 1 over treatment 0.

[43] This use of R as a relative evaluation is a slight abuse of notation. Before we used R as absolute level of utility. However, choice valuations are always relative to some benchmark, so there is little possibility of confusion in this usage.

The variance–covariance matrix of (U_0, U_1, υ) is

$$\Sigma_\upsilon = \begin{pmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{0\upsilon} \\ \sigma_{01} & \sigma_1^2 & \sigma_{1\upsilon} \\ \sigma_{0\upsilon} & \sigma_{1\upsilon} & \sigma_\upsilon^2 \end{pmatrix},$$

where σ_{ij} is the covariance between outcomes i and j.

In this model, the average treatment effect given $X = x$ is

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

Treatment on the treated is

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

where the third equality follows from independence assumption (i). The **local average treatment effect** (LATE) of Imbens and Angrist (1994) is the average gain to program participation for those induced to receive treatment through a change in $Z[= (X, W)]$ by a component of W not in X. Such a change affects choices but not potential outcomes. Let $D(z)$ be the random variable D when we fix $W = w$ and let $D(z')$ be the random variable when we fix $W = w'$. The LATE parameter as defined by Heckman and Vytlacil (1999) is

$$\begin{aligned} \mathrm{LATE}(z, z', x) &= E(Y_1 - Y_0 \mid D(z) = 0, D(z') = 1, X = x) \\ &= x(\beta_1 - \beta_0) + E\big(U_1 - U_0 \mid R(z) < 0, R(z') \geqslant 0, X = x\big) \\ &= x(\beta_1 - \beta_0) + E(U_1 - U_0 \mid -z'\gamma \leqslant \upsilon < -z\gamma), \end{aligned}$$

using independence assumption (i) and the index structure to obtain the final result.

A definition of LATE introduced by Heckman and Vytlacil (1999, 2000, 2005) can be made independent of the existence of any instrument. Imbens and Angrist (1994) define LATE by invoking an instrument and thereby apparently conflate tasks 1 and 2 in Table 1 (the tasks of definition and identification). We can define LATE as the mean return for agents with values of $\upsilon \in [\underline{\upsilon}, \bar{\upsilon}]$. Instruments W may not exist, yet LATE can still be defined as

$$\mathrm{LATE}\big(x, \upsilon \in [\underline{\upsilon}, \bar{\upsilon}]\big) = x(\beta_1 - \beta_0) + E(U_1 - U_0 \mid \underline{\upsilon} \leqslant \upsilon < \bar{\upsilon}).$$

With this definition, we can separate task 1 of Table 1 from task 2. If $\underline{\upsilon} = -z'\gamma$ and $\bar{\upsilon} = -z\gamma$, we obtain the instrument-dependent version of LATE in the Roy model.

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

$$E(Y_1 - Y_0 \mid \upsilon = \upsilon^*,\ X = x,\ Z = z) = x(\beta_1 - \beta_0) + E(U_1 - U_0 \mid \upsilon = \upsilon^*).$$

This parameter is a generalization of a parameter introduced into the evaluation literature by Björklund and Moffitt (1987). It is the mean return for agents for whom $X = x$, $Z = z$, and $\upsilon = \upsilon^*$. It is defined independently of any instrument. At a special point of evaluation where $R = 0$ (i.e. $z\gamma + \upsilon = 0$), the MTE is a willingness to pay measure that informs us how much an agent at the margin of participation (in the indifference set) would be willing to pay to move from "0" to "1". This particular point of evaluation for the marginal treatment effect is what we called "EOTM" (the effect of treatment for agents at the margin of indifference) in Section 2.4.

Under regularity conditions satisfied by the normal distribution and expressing it in instrument-dependent form, EOTM can be defined as the limit form of LATE,

$$\begin{aligned}
&\lim_{z\gamma \to z'\gamma} \text{LATE}(z, z', x) \\
&\quad = x(\beta_1 - \beta_0) + \lim_{z\gamma \to z'\gamma} E(U_1 - U_0 \mid -z'\gamma \leqslant \upsilon < -z\gamma) \\
&\quad = x(\beta_1 - \beta_0) + E(U_1 - U_0 \mid \upsilon = -z'\gamma).^{44}
\end{aligned}$$

LATE, as interpreted by Heckman and Vytlacil (1999, 2000, 2005), is the average return for agents with $\upsilon \in [-z'\gamma, -z\gamma]$. This parameter expresses the outcome of manipulating the values at which we set υ by manipulation of the mean scale utility $z\gamma$, but holding X fixed. The relative preferences for state 1 compared to state 0, but not the outcomes Y_1, Y_0, are affected by such changes because we fix X. An example of such a change in Z is a change in tuition but not a change in variables directly affecting Y_1, Y_0 (the X).

In the special case of the Roy model, $C = 0$, $R = Y_1 - Y_0$ and $\upsilon = U_1 - U_0$, the MTE is

$$E(Y_1 - Y_0 \mid U_1 - U_0 = u_1 - u_0,\ X = x) = x(\beta_1 - \beta_0) + (u_1 - u_0).$$

In the special case where $R = 0$, $x(\beta_1 - \beta_0) = -(u_1 - u_0)$ and MTE at this point of evaluation is zero (i.e. EOTM is zero).

We can work with $Z\gamma$ or with the propensity score $P(Z)$ interchangeably. Under our normality assumptions, ATE is defined as before. Treatment on the Treated can be defined using the standard selection formulae. We have already defined Φ as the distribution function for a standard unit normal random variable; $\phi(\psi) = \Phi'(\psi)$ is the density of this variable evaluated at ψ. Using results on the truncated normal surveyed in Heckman and Honoré (1990), and summarized in Appendix C, we can express treatment on the treated given Z, normalizing the variance of υ to 1 to simplify the notation,

$$\begin{aligned}
\text{TT}(x, z) &= E(Y_1 - Y_0 \mid X = x, Z = z, \upsilon \geqslant -z\gamma) \\
&= x(\beta_1 - \beta_0) + \text{Cov}(U_1 - U_0, \upsilon)\lambda(z\gamma),
\end{aligned}$$

[44] The regularity conditions apply to families of distributions that are more general than the normal ones. [These are discussed further in Chapter 71.]

where

$$\lambda(z\gamma) = \frac{\phi(z\gamma)}{\Phi(z\gamma)}.$$

λ is monotone decreasing in $z\gamma$ and $\lim_{z\gamma\to\infty}\lambda(z\gamma) = 0$ and $\lim_{z\gamma\to-\infty}\lambda(z\gamma) = \infty$. These and other properties of truncated normal random variables are presented in Appendix C.[45]

As noted by Heckman (1980) and Heckman and Robb (1985), because $\Phi(\psi)$ is monotone increasing in ψ, $z\gamma = \Phi^{-1}(\Pr(D(Z) = 1 \mid Z = z))$, and we can substitute everywhere for $z\gamma$ by $P(z) = \Pr(D(Z) = 1 \mid Z = z)$, the propensity score, to reach

$$\begin{aligned}\mathrm{TT}(x, z) &= \mathrm{TT}\big(x, P(z)\big)\\ &= x(\beta_1 - \beta_0) + \mathrm{Cov}(U_1 - U_0, \upsilon)K\big(P(z)\big).\end{aligned}$$[46]

Observe that if $\mathrm{Cov}(U_1 - U_0, \upsilon) = 0$, ATE = TT. If $\mathrm{Cov}(U_1 - U_0, \upsilon) > 0$, TT > ATE because of purposive sorting into sector 1. A positive covariance is guaranteed by the Roy model because $\upsilon = U_1 - U_0$. As $z\gamma$ increases, more agents with low values of υ are drawn in to sector 1. If υ is positively correlated with $U_1 - U_0$, we lower the average quality of participants (agents for whom $R > 0$) as we increase $z\gamma$.

As $z\gamma \to \infty$, $P(z) \to 1$, and the distance between ATE and TT goes to zero. Agents with high values of the probability of participation are a random sample of the U_1 but obviously not a random sample of the $z\gamma$. Limit set arguments of the type that set $P(z)$ to one or zero play a crucial role in versions of semiparametric identification of economic choice models and in the entire treatment effect literature that seeks to identify ATE by the method of instrumental variables.

The LATE parameter for the generalized Roy model can be derived using the fact that if $(y, r) \sim N(\mu_y, \mu_r, \sigma_y, \sigma_r, \rho)$ and $b > a$, then

$$E(y \mid a \leqslant r < b) = \mu_y + \rho\sigma_y\left(\frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)}\right),$$

where $\alpha = (a - \mu_r)/\sigma_r$, $\beta = (b - \mu_r)/\sigma_r$. Using an instrument dependent definition of LATE and normalizing $\mathrm{Var}(\upsilon) = 1$,

$$\begin{aligned}\mathrm{LATE}(z, z', x) &= E(Y_1 - Y_0 \mid x, -z'\gamma \leqslant \upsilon < -z\gamma)\\ &= x(\beta_1 - \beta_0) + \mathrm{Cov}(U_1 - U_0, \upsilon)\left[\frac{\phi(z\gamma) - \phi(z'\gamma)}{\Phi(z'\gamma) - \Phi(z\gamma)}\right],\end{aligned} \tag{3.5}$$

where the final result uses the symmetry of the normal density. The Marginal Treatment Effect (MTE) corresponds to the expected outcome gain for those agents who are just indifferent to the receipt of treatment at the given value of the unobservable υ. Formally, recalling that we normalize $\mathrm{Var}(\upsilon) = 1$,

[45] Notice that $d = -z\gamma$ in the notation of Appendix C.

[46] $K(P(Z)) = \frac{\phi(\Phi^{-1}(P(z)))}{P(z)}$.

$$\begin{aligned}\text{MTE}(x, -z\gamma) &= x(\beta_1 - \beta_0) + E(U_1 - U_0 | \upsilon = -z\gamma) \\ &= x(\beta_1 - \beta_0) + \text{Cov}(U_1 - U_0, \upsilon)[-z\gamma].^{47}\end{aligned} \tag{3.6}$$

In terms of the propensity score, we can write $\text{MTE}(x, 1 - P(z)) = x(\beta_1 - \beta_0) + (\rho_1\sigma_1 - \rho_0\sigma_0)\Phi^{-1}(1 - P(z))$. As long as $\text{Cov}(U_1 - U_0, \upsilon) > 0$, those with high values of $P(z)$ (high values of $z\gamma$) have the *lowest* mean returns to participation. Evaluating MTE when $z\gamma$ is large corresponds to the case where the average outcome gain is evaluated for those agents with unobservables making them on average less likely to participate. Higher mean scale utilities draw in those agents with unobservables that make them less likely to participate. When $\upsilon = 0$, MTE = ATE as a consequence of the symmetry of the normal distribution.

The other evaluation criteria discussed in Section 2 can be formed using the normal model. The proportion of agents who benefit from the program in subjective terms is the propensity score $P(Z)$. In the special case of the Roy model where $C \equiv 0$, this is also the proportion of agents who benefit in "objective" terms ($\Pr(Y_1 \geqslant Y_0)$). The policy relevant treatment effect depends on the exact specification of policies. We develop versions of the policy relevant treatment effect in Chapter 71. Given the ingredients of the discrete choice model (3.1) with associated outcomes (3.3), we can generate all of the treatment effects and counterfactual distributions discussed in Section 2.

The linearity, exogeneity, separability and normality assumptions invoked in this section make it possible to solve policy problems P-1–P-3. We can solve policy problem P-2 (the extrapolation problem) using this model evaluated at new values of (X, Z). By construction the (U_1, U_0, υ) are independent of (X, Z), and given the functional forms all the mean treatment parameters can be generated for all (X, Z).

By parameterizing the β_i to depend only on measured characteristics, it is possible to forecast the demand for new goods and solve policy problem P-3. For example, suppose that β_1, β_0 and γ only depend on the characteristics of the policies. A special case would be

$$\beta_1(Q_1) = \Lambda Q_1', \tag{3.7a}$$

$$\beta_0(Q_0) = \Lambda Q_0', \tag{3.7b}$$

where Q_1 and Q_0 are $1 \times J$ vectors of characteristics of programs, and X is a $1 \times K$ vector of agent-specific characteristics, and Λ is a $K \times J$ matrix. Z is a $1 \times M$ vector

[47] Note that using L'Hôpital's Rule, MTE can be regarded as the limit form of LATE. Setting $\sigma_\upsilon = 1$, we obtain

$$\begin{aligned}\text{MTE}(x, -z\gamma) &= x(\beta_1 - \beta_0) + \text{Cov}(U_1 - U_0, \upsilon) \lim_{t \to -z\gamma} \left[\frac{\phi(-z\gamma) - \phi(t)}{\Phi(t) - \Phi(-z\gamma)} \right] \\ &= x(\beta_1 - \beta_0) + \text{Cov}(U_1 - U_0, \upsilon) \lim_{t \to -z\gamma} \left[\frac{(\phi(-z\gamma) - \phi(t))/(-z\gamma - t)}{(\Phi(t) - \Phi(-z\gamma))/(-z\gamma - t)} \right] \\ &= x(\beta_1 - \beta_0) + \text{Cov}(U_1 - U_0, \upsilon)[-z\gamma].\end{aligned}$$

and Γ is a $M \times J$ matrix of characteristics such that

$$\gamma(Q_1) - \gamma(Q_0) = \Gamma\big[Q_1' - Q_0'\big].$$

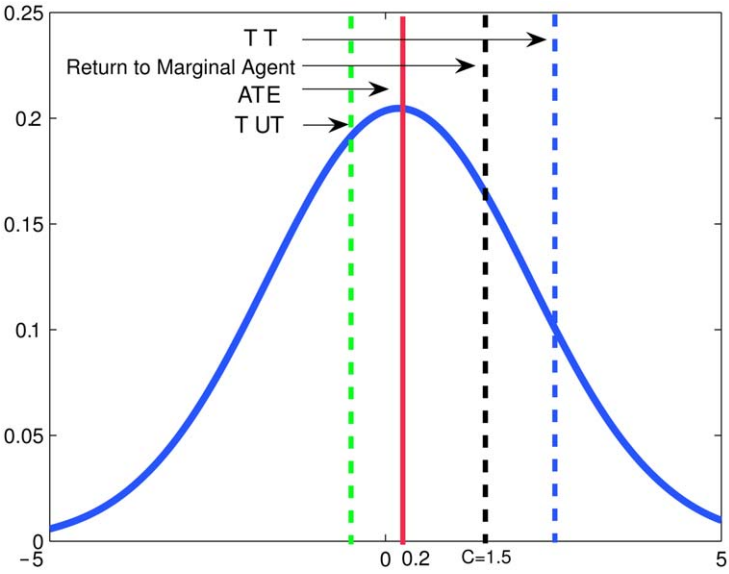
Under this assumption, all programs can be put on a common basis in terms of the characteristics they offer. The characteristics of agents are (X, Z). For a new program, generated by a new bundle of fixed characteristics, we can solve P-3 if we can also characterize the distributions of $\upsilon(s)$ and $U(s)$ in terms of the $Q(s)$. One special case is where the $\upsilon(s)$ and $U(s)$ do not depend on s, as in Quandt and Baumol (1966) or McFadden (1974). Then all effects of the new program come through the β and γ. We now consider some examples of the Roy model. It defines the economic choice framework used throughout this Handbook chapter, so it is useful to gain intuition about it.

3.3.1. Examples of Roy models

Figure 1, adapted from Heckman, Urzua and Vytlacil (2006), displays the distribution of gross gains $(Y_1 - Y_0)$ from adopting a treatment. The generating model is an extended Roy model with parameters given at the base of the table. The model builds in positive sorting on unobservables because $\upsilon = U_1 - U_0$ so $\text{Cov}(U_1 - U_0, \upsilon) > 0$. All agents face the same cost of treatment adoption C. The return to the treatment for the randomly selected agent is ATE $(= 0.2)$. Given $C = 1.5$, the return to the agent at the margin is 1.5. The average return for the adopting agents is TT $(= 2.666)$. Thus the agents adopting the treatment are the ones who benefit from it. This is a source of evaluation bias in evaluating programs.

Figure 2 plots the parameters ATE(p), TT(p), MTE(p) and TUT(p) (treatment on the untreated) that underlie the model used to generate Figure 1. Table 2 presents the formulae for the treatment parameters as a function of p. Here "p" denotes a value of $P(Z)$ and not a policy as in the previous sections. The declining MTE(p) is the prototypical pattern of diminishing returns that accompanies an expansion of treatment (MTE declines in $U_D = u_D = p$). Agents with low levels of $Z\gamma$ $(P(Z))$ that adopt the treatment must do so because their unobservables make them more likely to. They have high values of υ $(R = Z\gamma + \upsilon)$ that compensate for the low values of $Z\gamma$. Since υ is positively correlated with $U_1 - U_0$ and Z does not enter $\mu_1(X) - \mu_0(X)$, the MTE is high for the low p agents at the margin of indifference. As cost C falls, more agents are drawn in to adopt treatment and the return falls. The pattern for treatment on the treated (TT(p)) is explained by similar considerations. As participation becomes less selective, the selected agent outcomes converge to the population average. As more agents participate, the stragglers are, on average, less effective adopters of the treatment. This explains the pattern for TUT(p). Observe that the slopes of these curves would reverse if there is negative sorting on unobservables $(\text{Cov}(U_1 - U_0, \upsilon) < 0)$. In this case, participants in the program would be those with below-average unobservables. Figure 3 plots the trade-off in $Z\gamma$ and υ that make agents indifferent and the two regions demarcated by the line of indifference. Agents with $(Z\gamma, \upsilon)$ traits to the right of the line have $D = 1$. Agents with traits below the line have $D = 0$.

$$U_1 - U_0 \not\perp D$$



$$\varphi = Y_1 - Y_0$$

$$\text{TT} = 2.666,\ \text{TUT} = -0.632$$
$$\text{Return to marginal agent} = C = 1.5$$
$$\text{ATE} = \mu_1 - \mu_0 = \bar{\varphi} = 0.2$$

The model

Outcomes	Choice model
$Y_1 = \mu_1 + U_1 = \alpha + \bar{\varphi} + U_1$ $Y_0 = \mu_0 + U_0 = \alpha + U_0$	$D = \begin{cases} 1 & \text{if } R \geqslant 0 \\ 0 & \text{if } R < 0 \end{cases}$
General case	
$(U_1 - U_0) \not\perp D$ $\text{ATE} \neq \text{TT} \neq \text{TUT}$	

The researcher observes (Y, D, C)
$Y = \alpha + \varphi D + U_0$ where $\varphi = Y_1 - Y_0$

Parameterization

$\alpha = 0.67$ $\quad (U_1, U_0) \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ $\quad \mu_Z = (2, -2)$ $\quad R = Y_1 - Y_0 - C$

$\bar{\varphi} = 0.2$ $\quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}$ $\quad \boldsymbol{\Sigma}_Z = \begin{bmatrix} 9 & -2 \\ -2 & 9 \end{bmatrix}$ $\quad C = 1.5$

Figure 1. Distribution of gains. The extended Roy economy. Adapted from Heckman, Urzua and Vytlacil (2006).

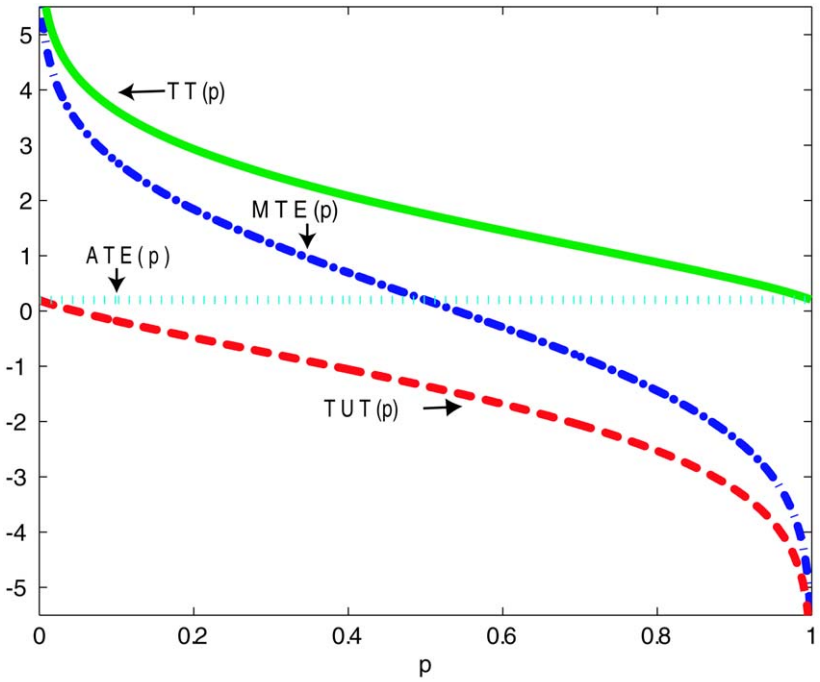


Figure 2. Treatment parameters as a function of $P(Z) = p$. Adapted from Heckman, Urzua and Vytlacil (2006).

Table 2
Treatment parameters evaluated at $P(Z) = p$

Parameter	Definition	Under assumptions (model below)
Marginal treatment effect	$E[Y_1 - Y_0 \mid R = 0, U_D = p]$	$\bar{\varphi} + \sigma_{U_1 - U_0} \Phi^{-1}(1-p)$
Average treatment effect	$E[Y_1 - Y_0 \mid P(Z) = p]$	$\bar{\varphi}$
Treatment on the treated	$E[Y_1 - Y_0 \mid R > 0, P(Z) = p]$	$\bar{\varphi} + \sigma_{U_1 - U_0} \frac{\phi(\Phi^{-1}(p))}{p}$
Treatment on the untreated	$E[Y_1 - Y_0 \mid R \leqslant 0, P(Z) = p]$	$\bar{\varphi} - \sigma_{U_1 - U_0} \frac{\phi(\Phi^{-1}(p))}{p}$

Note. $\Phi(\cdot)$ and $\phi(\cdot)$ represent the cdf and pdf of a standard normal distribution, respectively. $\Phi^{-1}(\cdot)$ represents the inverse of $\Phi(\cdot)$.

This example shows how the extended Roy model can be used to define the distribution of treatment effects. Mean treatment parameters are derived from it. The Roy model and its extensions are examples of economic models that can be used to define counterfactuals (in this case Y_0 and Y_1). They are purely theoretical constructs. We discuss iden-

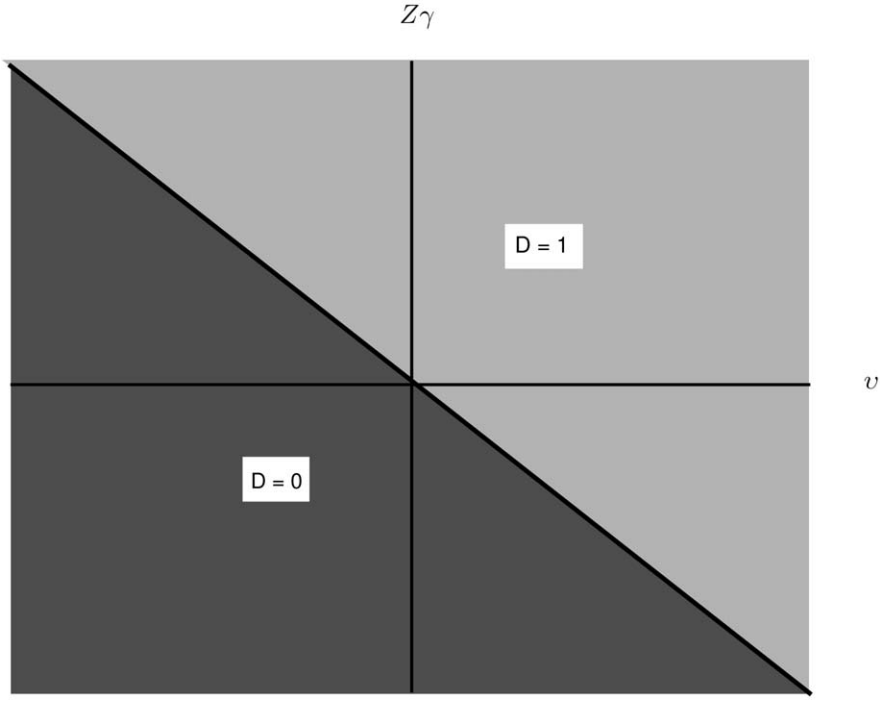


Figure 3. Partitions of $Z\gamma$ and υ into $D = 0$ and $D = 1$. The boundary ($Z\gamma + \upsilon = 0$) is the margin of indifference.

tification of this model and its extensions in Section 6.1. In Chapter 71 and in Abbring and Heckman (Chapter 72), we consider how alternative evaluation estimators identify, or do not identify, the parameters of this basic economic model, and its extensions.

3.4. Adding uncertainty

Because it does not rely on explicitly formulated economic models, the treatment effect literature is not clear about the sources of variability and uncertainty that characterize choices and outcomes and their relationship. The econometric approach to program evaluation is very clear about the sources of uncertainty and variability in the econometric model.

In devising estimators and interpreting estimated parameters, it is helpful to distinguish the information available to the agent from the information available to the observing econometrician. 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 bear. In general the agent's information is not the same as the econometrician's, and they may not be nested. The agent may know things in advance that the

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

Let $\mathcal{I}_{ea}$ 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. In place of (3.1), we can write, using somewhat nonstandard notation,

$$R(s, \mathcal{I}_{ea}) = \mu_R(s, \mathcal{I}_{ea}) + \upsilon(s, \mathcal{I}_{ea})$$

reflecting that *ex ante* valuations are made on the basis of *ex ante* information where $\mu_R(s, \mathcal{I}_{ea})$ is determined by variables that are known to the econometrician and $\upsilon(s, \mathcal{I}_{ea})$ 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

$$\operatorname{argmax}_{s\in\mathcal{S}}\{R(s, \mathcal{I}_{ea})\} \neq \operatorname{argmax}_{s\in\mathcal{S}}\{R(s, \mathcal{I}_{ep})\}$$

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

Determining agent information sets is a major research topic in structural econometrics [see Abbring and Campbell (2005), Miller (1984), Pakes (1986), Chan and Hamilton (2003), Carneiro, Hansen and Heckman (2003), Cunha, Heckman and Navarro (2005)]. 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.[48] In this chapter, we analyze both *ex ante* and *ex post* objective outcomes and subjective valuations. Abbring and Heckman (Chapter 72) show how to implement these distinctions.

In the context of the simple two-outcome model developed in Section 3.3, we can define $R(\mathcal{I}_{ea})$ as

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

Under perfect foresight, the agent knows Y_1, Y_0 and C as in the classical generalized Roy model; $\mathcal{I}_{ea} \supseteq \{Y_1, Y_0, C\}$. More generally, the choice equation is generated by $D(\mathcal{I}_{ea}) = \mathbf{1}[R(\mathcal{I}_{ea}) \geqslant 0]$. *Ex post*, different choices might be made. *Ex ante*, agents may be uncertain about aspects of the choices that they made. For different specifications of the information set we obtain different choices.

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. Abbring and Heckman (Chapter 72) survey models with uncertainty.

[48] As Hicks (1946, p. 179) puts it, "*Ex post calculations of capital accumulation have their place in economic and statistical history; they are useful measure for economic progress; but they are of no use to theoretical economists who are trying to find out how the system works, because they have no significance for conduct.*"

We consider identifiability of the generalized Roy model under certainty in Section 6. The recent literature on semiparametric econometric models surveyed in Chapter 73 (Matzkin) of this Handbook enables economists to relax the normality, separability and functional form assumptions developed in the early literature on structural estimation while at the same time preserving the economic content of the structural literature.

Before developing this topic, we clarify the distinction between structural models and causal models and we relate the statistical treatment effect literature to the literature on structural economic models.

4. Counterfactuals, causality and structural econometric models

The literature on policy evaluation in economics sometimes compares "structural" approaches with "treatment effect" or "causal" models.[49] These terms are used loosely. This section formally defines "structural" models and uses them as devices for generating counterfactuals. We consider both outcome and treatment choice equations. We compare the econometric model for generating counterfactuals and causal effects with the Neyman (1923)–Rubin (1978) model of causality and compare "causal" parameters with "structural" parameters. We compare and evaluate the structural equations approach and the treatment effects approach. We restore the "ω" notation introduced in Section 2 because it clarifies our discussion.

4.1. Generating counterfactuals

The treatment effect approach and the explicitly economic approach differ in the detail with which they specify both observed and counterfactual outcomes $Y(s, \omega)$, for different treatments denoted by "s". The econometric approach models counterfactuals much more explicitly than is common in the application of the treatment effect approach. This difference in detail corresponds to the differing objectives of the two approaches. This greater attention to detail in the structural approach facilitates the application of theory to provide interpretation of counterfactuals and comparison of counterfactuals across data sets using the basic parameters of economic theory. These models also suggest strategies for identifying parameters (task 2 in Table 1). Models for counterfactuals are the basis for extending historically experienced policies to new environments and for forecasting the effects of new policies never previously experienced. These are policy questions P-2 and P-3 stated in Section 2. Comparisons are made across treatments to define the individual level (ω) causal effect of s relative to s' as in (2.1).

Models for counterfactuals are in the mind. They are internally consistent frameworks derived from theory. Verification and identification of these models are logically distinct tasks that should be carefully distinguished from the purely theoretical act of

[49] See, e.g., Angrist and Imbens (1995) and Angrist, Imbens and Rubin (1996).

constructing internally consistent models. No issue of sampling, inference or selection bias is entailed in constructing theoretical models for counterfactuals.

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

$$y(s) = g_s(x, u_s), \tag{4.1}$$

where x and u_s are fixed variables specified by the relevant economic theory. This notation allows for different unobservables u_s to affect different outcomes.[50] $\mathcal{D}$ is the domain of the mapping $g_s : \mathcal{D} \rightarrow \mathcal{R}^y$, 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 (4.1). If we model the *ex post* realizations of outcomes, it is entirely reasonable to invoke an all causes model. *Ex post*, 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 (4.1) is a production function relating inputs (factors) to outputs. The notation x and u_s anticipates the practical econometric problem that some arguments of functional relationship (4.1) are observed while other arguments may be unobserved by the econometrician. In the analysis of this section, their roles are symmetric. 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.[51] Thus, Equation (4.1) maps admissible inputs into possible *ex post* outcomes. Our notation allows for different unobservables from a common list u to appear in different outcome equations.

A "deep structural" version of (4.1), discussed in Sections 3.2 and 3.3, 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:

$$y(s) = g(q_s, x, u_s), \tag{4.2}$$

where now the domain of g, $\mathcal{D}$, is defined for q_s, x, u_s so that we have $g : \mathcal{D} \rightarrow \mathcal{R}^y$.[52] 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, $g : (q_s, x, u_s) \rightarrow \mathcal{R}^y$, 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 P-3 since new policies (treatments) are generated from common characteristics, and all policies are put on a common basis.

[50] An alternative notation would use a common u and lets g_s select out s-specific components.

[51] The support is the region of the domain of definition where we have data on the function. Thus if $\mathcal{D}_x$ is the domain of x, the support of x is the region $\text{Supp}(x) \subset \mathcal{D}_x$ such that the data density $f(x)$ satisfies the condition $f(x) > 0$ for $x \in \text{Supp}(x)$.

[52] An example is given by Equations (3.7a) and (3.7b).

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 P-3 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 the statistical approach that focuses on problem P-1 is that it works solely with outcomes rather than inputs. However, it is silent on how to solve problems P-2 and P-3 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. u_s are the unobserved characteristics of both the schools and the pupil. If we 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 we can identify (4.2) over the domain of the function defined by the new package, we can solve problem P-3. If the same schooling input (same q_s) is applied to different students (those with different x) and we can identify (4.1) or (4.2) over the new domain of definition, we solve problem P-2. By digging deeper into the "causes of the effects" we can do more than just compare the effects of treatments in place with each other. In addition, as we show in Chapter 71, 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.

Another example from the theory of labor supply writes hours of work h as a function of the before tax wage w, where s is the tax rate that is assumed common across all agents, and other characteristics are denoted u_s. Treatment in this example is the proportional tax rate s. We may write hours of work in tax regime s, for a person with wage w and characteristics x as

$$h_s = h\big(w(1-s), x, u_s\big)$$

as the labor supply for proportional tax rate s for an agent with characteristics (x, u_s).[53] This may be a factual (observed) quantity or a counterfactual quantity. Different tax rates (policies) produce different counterfactuals which are generated by a common function. We return to this example on several occasions throughout this chapter.

Our analysis in Section 3.3 provides a deep structural generalized Roy model example of causal functions. The outcome equations parameterized by (3.7a) and (3.7b) are examples of models with deep structural parameters that can be used to solve P-2 and P-3.

[53] This notation permits the unobservable to differ across tax regimes.

Equations (4.1) and (4.2) are sometimes called Marshallian causal functions [see Heckman (2000)]. Assuming that the components of (x, u_s) or (q_s, x, u_s) are variation-free,[54] a feature that may or may not be produced by the relevant theory, we may vary each argument of these functions to get a *ceteris paribus* causal effect of the argument on the outcome. 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. Equations (3.4a)–(3.4b) are examples of Marshallian causal functions where (X, U) are the observed and unobserved variables.

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 we vary the package from q_s to $q_{s'}$ we get different treatment effects.

We use the convention that lower case values are used to define fixed values and upper case notation denotes random variables. In defining (4.1) and (4.2), we 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 act on these nonstochastic variables. If uncertainty is a feature of the environment, (4.1) and (4.2) can be interpreted as *ex post* realizations of the counterfactual. 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, (4.1) and (4.2) 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

$$E\big(Y(s,\omega) \mid \mathcal{I}_\omega\big) = E\big(g\big(Q(s,\omega), X(\omega), U(s,\omega)\big) \mid \mathcal{I}_\omega\big),^{55} \tag{4.3}$$

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 $\mathcal{I}_\omega$. This distribution may differ from the distribution produced by "reality" or nature if agent expectations are different from objective reality.[56] In the presence of intrinsic uncertainty, the relevant decision maker

[54] More precisely, if $\mathcal{X}, \mathcal{U}$ or $\mathcal{Q}, \mathcal{X}, \mathcal{U}$ are the domains of (4.1) and (4.2), $\mathcal{D} = (\mathcal{X}, \mathcal{U}) = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \times \mathcal{U}_1 \times \cdots \times \mathcal{U}_M$ or $(\mathcal{Q}, \mathcal{X}, \mathcal{U}) = \mathcal{Q}_1 \times \cdots \times \mathcal{Q}_K \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \times \mathcal{U}_1 \times \cdots \times \mathcal{U}_M$ where we assume K components in $\mathcal{Q}$, N components in $\mathcal{X}$, and M components in $\mathcal{U}$.

[55] The expectation might be computed using the information sets 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.

[56] Thus agents do not necessarily use rational expectations, so the distribution used by the agent to make decisions need not be the same as the distribution generating the data.

acts on (4.3) but the *ex post* counterfactual is

$$Y(s,\omega) = E\big(Y(s,\omega) \mid \mathcal{I}_\omega\big) + \nu(s,\omega), \tag{4.4}$$

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 (Chapter 72) present Roy model examples of models accounting for uncertainty.

This statement of the basic deterministic model reconciles the all causes model (4.1) and (4.2) 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, so no assumption of rational expectations is necessarily imposed. 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 we 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

$$E\big(Y(s,\omega) \mid \mathcal{I}_\omega\big) - E\big(Y(s',\omega) \mid \mathcal{I}_\omega\big). \tag{4.5}$$

However, agents may not act in terms of these *ex ante* effects if they have decision criteria (utility functions) that are not linear in the outcomes but may form expectations of nonlinear functions of $Y(s,\omega), s = 1, \ldots, \bar{S}$. We discuss *ex ante* valuations of outcomes in the next section.

The value of the scientific (or explicitly structural) approach to the construction of counterfactuals is that it models the unobservables and the sources of variability among observationally identical people. Since it is the unobservables that give rise to selection bias and problems of inference that are central to empirically rigorous causal analysis, economists using the scientific approach can draw on economic theory to design and justify methods to control for selection bias. This avenue is not available to adherents of the statistical approach. Statistical approaches that are not explicit about the sources of the unobservables make strong implicit assumptions which, when carefully exposited, are often unattractive. We exposit these assumptions in Chapter 71 when we discuss specific policy evaluation estimators.

The models for counterfactuals (4.1) and (4.2) are based on theory. The arguments of these functions are varied by hypothetical manipulations. These are thought experiments. When analysts attempt to construct counterfactuals empirically, they must carefully distinguish between these theoretical relationships and the empirical relationships determined by conditioning only on the observables.

The data used to determine these functions may be limited in its support. In this 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. A good example is the problem of multicollinearity. If the X in a sample are functionally dependent, it is not possible to identify the Marshallian causal function with respect to all 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 (4.1) or (4.2).

We 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.

4.2. 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.[57] Pearl suggests a particular physical mechanism for fixing variables and operationalizing causality, but it is not central to his or any other definition of causality.

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). We thus abstract from the problem of limited support that is discussed in the preceding section. We may write (dropping the "ω" notation for random variables, as we did in Section 3)

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

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

[57] Florens and Heckman (2003) distinguish conditioning from fixing, and generalize Pearl's analysis to both static and dynamic settings.

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, we can write this model formulated at the population level as a conditional expectation,

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

Since we condition on both X and U, there is no further source of variation in Y. 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, we only condition on X, we obtain

$$E(Y \mid X = x) = x\beta + E(U \mid X = x).^{58} \tag{4.6}$$

This relationship does not generate U-constant (Y, X) relationships. It generates only an X-constant relationship. Unless we condition on all of the "causes" (the right-hand side variables), the empirical relationship (4.6) 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 $y = g(q, x, u)$. A model specified in terms of random variables Q, X, U with the same support as q, x, u has as its conditional expectation $g(Q, X, U)$ under general conditions. Conditioning only on Q, X does not in principle identify $g(q, x, u)$.

Conditioning and fixing on the arguments of g or g_s are the same operations in an "all causes" model if all causes are accounted for. In general, they are not the same. This analysis can be generalized to account for the temporal resolution of uncertainty if we include $\nu(s, \omega)$ as an argument in the *ex post* causal model. The outcomes can include both objective outcomes $Y(s, \omega)$ and subjective outcomes $R(Y(s, \omega), \omega)$.

Statisticians and epidemiologists often do not distinguish between fixing and conditioning because they typically define the models that they analyze in terms of some type of conditioning on observed random variables. However, thought experiments in models of hypotheticals that vary factors are distinct from variations in conditioning variables. The latter conflate the effects of variation in X, holding U fixed, with the effects of X in predicting the unobserved factors (the U) in the outcome equations. This is the crucial distinction introduced in Haavelmo's fundamental 1943 paper.

[58] We assume that the mean of U is finite.

4.3. Modeling the choice of treatment

Parallel to causal models for outcomes are causal models for the choice of treatment. Consider *ex ante* personal valuations of outcomes based on expectations of gains from receiving treatment s:

$$E\big(R\big(Y(s,\omega), C(s,\omega), Q(s,\omega), \omega\big) \mid \mathcal{I}_\omega\big), \quad s \in \mathcal{S},$$

where, as before, $C(s,\omega)$ is the price or cost agent ω must pay for participation in treatment s. We decompose $C(s,\omega)$ into observables and unobservables. We thus write $C(s,\omega) = K(W(s,\omega), \eta(s,\omega))$. We allow utility R to be defined over the characteristics that generate the treatment outcome (e.g., quality of teachers in a schooling choice model) as well as attributes of the agent. In parallel with the g_s function generating the $Y(s,\omega)$, we write

$$R\big(Y(s,\omega), C(s,\omega), Q(s,\omega), \omega\big) = f\big(Y(s,\omega), W(s,\omega), Q(s,\omega), \eta(s,\omega), \omega\big).$$

Parallel to the analysis of outcomes, we may keep $Q(s,\omega)$ implicit and use f_s functions instead of f. In the Roy model of Section 3.3, $R = Y_1 - Y_0 - C$ is the agent's subjective evaluation of treatment.

Our analysis includes both measured and unmeasured attributes as perceived by the econometrician. The agent computes expectations against his/her subjective distribution of information. We allow for imperfect information by postulating an ω-specific information set. If agents know all components of future outcomes, the upper case letters become lower case variables which are known constants. The $\mathcal{I}_\omega$ are the causal factors for agent ω. In a utility maximizing framework, choice $\hat{s}$ is made if $\hat{s}$ is maximal in the set of valuations of potential outcomes

$$\big\{E\big(R\big(Y(s,\omega), C(s,\omega), Q(s,\omega), \omega\big) \mid \mathcal{I}_\omega\big),\ s \in \mathcal{S}\big\}.$$

In this interpretation, the information set plays a key role in specifying agent preferences. Actual realizations may not be known at the time decisions are made. 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 (2003), 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 as we demonstrate in Chapter 71. A parallel analysis can be made if the decision maker is not the same as the agent whose objective outcomes are being evaluated.

4.4. The econometric model vs. the Neyman–Rubin model

Many statisticians and social scientists invoke a model of counterfactuals and causality attributed to Donald Rubin by Paul Holland (1986) but which is actually due to Neyman

(1923).[59] This model arises from the statistical literature on the design of experiments.[60] 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 we have done in Equations (4.1)–(4.4), using the econometric or "structural" approach. Rubin and Neyman offer no model of the choice of which outcome is selected. Thus there is 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 our notation, Rubin assumes (PI-1) and (PI-2) as presented in Section 2.[61] Since he does not develop choice equations or subjective evaluations, he does not consider the more general invariance conditions (PI-3) and (PI-4) for both objective and subjective evaluations developed in Section 2.2. Assumptions (PI-1) and (PI-2) are versions of familiar invariance assumptions developed in Cowles Commission econometrics and formalized in Hurwicz (1962) but applied only to outcome equations and not to treatment choice equations. Assumption (PI-1) says that the objective outcomes are the same irrespective of the policy or assignment mechanism that implements it within a policy regime. (PI-2) assumes no general equilibrium effects or social interactions among agents for objective outcomes. Thus the outcomes for an agent are the same whether one agent receives treatment or many receive treatment.

More formally, the Rubin model assumes

(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) (PI-1): *Invariance of counterfactuals for objective outcomes to the mechanism of assignment within a policy regime*;

(R-3) (PI-2): *No social interactions or general equilibrium effects for objective outcomes*;

and

(R-4) *There is no simultaneity in causal effects, i.e., outcomes cannot cause each other reciprocally*.

[59] The framework attributed to Rubin was developed in statistics by Neyman (1923), Cox (1958) and others. Parallel frameworks were independently developed in psychometrics [Thurstone (1927)] and economics [Haavelmo (1943), Roy (1951), Quandt (1958, 1972)].

[60] See Cox (1958) for a classic treatment of this subject.

[61] Rubin (1986) calls these assumptions "SUTVA" for Stable Unit Treatment Value Assumption.

Two further implicit assumptions in the application of the model are that P-1 is the only evaluation problem of interest and that mean causal effects are the only objects of interest.

The econometric approach is richer and deeper than the statistical treatment effect approach. Its signature features are:

1. 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.
2. The analysis of subjective evaluations of outcomes $R(s, \omega)$, $s \in \mathcal{S}$, and the use of choice data to infer them.
3. The analysis of *ex ante* and *ex post* realizations and evaluations of treatments. This analysis enables analysts to model and identify regret and anticipation by agents. Points 2 and 3 introduce agent decision making into the treatment effect literature.
4. Development of models for identifying entire distributions of treatment effects (*ex ante* and *ex post*) rather than just the traditional mean parameters focused on by many statisticians. These distributions enable analysts to determine the proportion of people who benefit from treatment, something not attempted in the statistical literature on treatment effects.
5. Development and identification of distributional criteria allowing for analysis of alternative social welfare functions for outcome distributions comparing different treatment states.
6. Models for simultaneous causality.
7. Definitions of parameters made without appeals to hypothetical experimental manipulations.
8. Clarification of the need for invariance of parameters with respect to classes of manipulations to answer classes of questions.[62]

We 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)]. Unlike the Neyman–Rubin model, these models do not start with the experiment as an ideal but 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

[62] This notion is featured in the early Cowles Commission work. See Marschak (1953) and Koopmans, Rubin and Leipnik (1950). It is formalized in Hurwicz (1962) as discussed below in Section 4.6. Rubin's "SUTVA" as embodied in (R-2) and (R-3) is a special case of the invariance condition formalized by Hurwicz and discussed in Section 4.6 below.

discussed in Section 3 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 nonrandomization, and not an explicit treatment of particular selection mechanisms in the nonrandomized case as developed in the econometrics literature. Even under ideal conditions, randomization cannot answer some very basic questions such as what proportion of a population benefits from a program.[63] And in practice, contamination and cross-over effects make randomization a far from sure-fire solution even for constructing ATE.[64]

Statisticians sometimes conflate the three tasks delineated in Table 1. This problem is especially acute among the "causal analysts." The analysis of Holland (1986, 1988) 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.[65] 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, Li and Hickman (2005) echo Rubin and Holland in insisting that if an experiment cannot "in principle" be performed, a causal effect cannot be defined.[66]

The act of definition is logically separate 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 glorifies randomization as the "gold standard" of causal inference.

[63] This point is made in Heckman (1992). See also Carneiro, Hansen and Heckman (2001, 2003), where this proportion is identified using choice data and/or supplementary proxy measures. See also Cunha, Heckman and Navarro (2005, 2006). Abbring and Heckman (Chapter 72) discuss this work.

[64] See the evidence on disruption bias and contamination bias arising in randomized trials that is presented in Heckman, LaLonde and Smith (1999), Heckman et al. (2000) and the discussion in Section 9 of Chapter 71.

[65] "Without treatment definitions that specify actions to be performed on experimental units, we cannot unambiguously discuss causal effects of treatments" [Rubin (1978, p. 39)].

[66] The LATE parameter of Imbens and Angrist (1994) is defined by an instrument and conflates task 1 and 2 (definition and identification). In Section 3.3 and in Chapter 71, we define the LATE parameter abstractly and separate issues of definition of parameters from issues of identification. Imbens and Angrist (1994) use instrumental variables as surrogates for randomization.

In the Neyman–Rubin model, the sources of variability generating $Y(s, \omega)$ as a random variable are not specified. The "causal effect" of s compared to s' is defined as the treatment effect (2.1). Holland (1986, 1988) argues that it is an advantage of the Rubin model that it is not explicit about the sources of variability among observationally identical agents, or about the factors that generate $Y(s, \omega)$. Holland and Rubin focus on mean treatment effects as the interesting causal parameters.

The econometric approach to causal inference supplements the 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. It moves beyond the dichotomy "missing at random" or "not missing at random". 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 vector of characteristics of treatment s for agent ω, $C(s, \omega)$ are costs and $\{Y(s, \omega)\}_{s \in \mathcal{S}}$ are the outcomes and the "|" denotes that these variables are defined conditional on $\mathcal{I}_\omega$ (the agent's information set).[67] The variables determining choices are analyzed. Along with the *ex ante* valuations that generate $D(s, \omega)$ are the *ex post* valuations discussed in Section 2.6.[68,69]

Knowledge of the relationship between choices and counterfactuals suggests appropriate methods for solving selection problems. By analyzing the relationship of the unobservables in the outcome equation, and the unobservables in the treatment choice equation, the analyst can use *a priori* theory to devise appropriate estimators to identify causal effects.

The econometric approach, unlike the Neyman–Rubin model, emphasizes the welfare of the agents being studied (through R_G or $R(Y(s, \omega), \omega)$ or $R = Y_1 - Y_0 - C$ in the Roy model) – the "subjective evaluations" – as well as the objective evaluations. The econometric approach also distinguishes *ex ante* from *ex post* subjective evaluations, so it can measure both agent satisfaction and regret.[70]

In addition, 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

[67] If other agents make the treatment assignment decisions, then the determinants of $D(s, \omega)$ are modified according to what is in their information set.

[68] Corresponding to these random variables are the deterministic all causes counterparts $d(s), q_s, c(s), \{y(s)\}, i$, where the $(\{c(s)\}_{s \in \mathcal{S}}, \{q_s\}_{s \in \mathcal{S}}, \{y(s)\}_{s \in \mathcal{S}}, i)$ generate the $d(s) = 1$ if $(\{c(s)\}_{s \in \mathcal{S}}, \{q_s\}_{s \in \mathcal{S}}, \{y(s)\}_{s \in \mathcal{S}}) \in \Psi$, a subset of the domain of the generators of $d(s)$. Again the domain of definition of $d(s)$ is not necessarily the support of $c(s, \omega), q_s(\omega), \{Y(s, \omega)\}_{s \in \mathcal{S}}$ and $\mathcal{I}_\omega$.

[69] Random utility models generating $D(s, \omega)$ originate in the work of Thurstone (1927) and McFadden (1974, 1981).

[70] See Cunha, Heckman and Navarro (2005, 2006) for estimates of subjective evaluations and regret in schooling choices. Abbring and Heckman (Chapter 72) review their work.

or for forecasting new policies (problems P-2 and P-3). Its focus on population mean treatment effects elevates randomization and matching to the status of preferred estimators. Such methods cannot identify distributions of treatment effects or general quantiles of treatment effects.[71]

One major limitation of the Neyman–Rubin model is that it is recursive. It does not model causal effects of outcomes that occur simultaneously. We now present a model of simultaneous causality based on conventional simultaneous equations techniques that illustrate the power of the econometric approach. This analysis also illustrates one version of a "structural" economic model – the Cowles Commission model.

4.5. Nonrecursive (simultaneous) models of causality

A system of linear simultaneous equations captures interdependence among outcomes Y. For simplicity, we focus on *ex post* outcomes so in this subsection, we ignore revelation of information over time and we keep "ω" implicit. To focus the issue 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.[72] 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

$$\Gamma Y + BX = U, \quad E(U) = 0, \tag{4.7}$$

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 (4.7) 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.[73] 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.[74] When the model is

[71] Angrist, Imbens and Rubin (1996) contrast structural models with causal models. The structural models they consider are the linear structural simultaneous equations models which we discuss as a special case of our analysis of nonrecursive models in Section 4.5. The appropriate comparison would be with nonseparable structural outcome models with correlated coefficients which is discussed in Heckman and Vytlacil (2001, 2005) and in Chapter 71. Angrist, Imbens and Rubin fail to note the recursive nature of Rubin model and the fundamentally nonrecursive nature of general structural models.

[72] This approach merges tasks 1 and 2 in Table 1. We do this in this section because the familiarity of the simultaneous equations model as a statistical model makes the all causes, fixed variable, *ex post* version confusing to many readers familiar with this model.

[73] 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 $\Gamma Y_t + BX_t = U_t$, $E(U_t) = 0$.

[74] In a time series model, the internal variables are Y_t determined in period t.

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 (4.7) is multiplied by Γ^{-1}. The reduced form is $Y = \Pi X + \mathcal{E}$ where $\Pi = -\Gamma^{-1}B$ and $\mathcal{E} = \Gamma^{-1}U$.[75] 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. We discuss invariance of structural parameters further in the next subsection.

Π is assumed to be invariant. 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 $\mathcal{E}$.[76] In the population generating (4.7), 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.[77] Nonlinear versions are available [Fisher (1966), Matzkin (2004, Chapter 73)]. For simplicity, we stick to the linear version, developing the nonlinear version in footnotes.[78]

The structural form (4.7) 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

$$Y_1 = \alpha_1 + \gamma_{12}Y_2 + \beta_{11}X_1 + \beta_{12}X_2 + U_1, \tag{4.8a}$$

[75] In this section only, Π refers to the reduced form coefficient matrix and not the family of probabilities of treatment assignment Π_p, as in earlier sections.

[76] The original formulations of this model assumed normality so that only means and variances were needed to describe the joint distributions of (Y, X).

[77] 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.

[78] 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 the contribution by Matzkin (Chapter 73) in this Handbook.

$$Y_2 = \alpha_2 + \gamma_{21} Y_1 + \beta_{21} X_1 + \beta_{22} X_2 + U_2. \tag{4.8b}$$

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

$$E(U_1 \mid X_1, X_2) = 0 \tag{4.9a}$$

and

$$E(U_2 \mid X_1, X_2) = 0. \tag{4.9b}$$

Completeness guarantees that (4.8a) and (4.8b) have a determinate solution for (Y_1, Y_2).

Applying Haavelmo's (1943) analysis to (4.8a) and (4.8b), 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

$$Y_1 = \pi_{10} + \pi_{11} X_1 + \pi_{12} X_2 + \mathcal{E}_1, \tag{4.10a}$$

$$Y_2 = \pi_{20} + \pi_{21} X_1 + \pi_{22} X_2 + \mathcal{E}_2. \tag{4.10b}$$

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

$$\begin{aligned} \pi_{11} &= \frac{\beta_{11} + \gamma_{12}\beta_{21}}{1 - \gamma_{12}\gamma_{21}}, \quad \pi_{12} = \frac{\beta_{12} + \gamma_{12}\beta_{22}}{1 - \gamma_{12}\gamma_{21}}, \\ \pi_{21} &= \frac{\gamma_{21}\beta_{11} + \beta_{21}}{1 - \gamma_{12}\gamma_{21}}, \quad \pi_{22} = \frac{\gamma_{21}\beta_{12} + \beta_{22}}{1 - \gamma_{12}\gamma_{21}}, \end{aligned} \tag{4.11}$$

and

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

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 (4.8a) or (4.8b), 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.[79] In this case, which is thought to be the most favorable to the application of least squares to (4.8a) and (4.8b), it is evident from (4.10a) and (4.10b) 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 (4.8a) and (4.8b).[80] 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

$$\frac{\pi_{21}}{\pi_{11}} = \gamma_{21}.$$[81]

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.

[79] See Fisher (1966).

[80] Some readers of an earlier draft of this chapter suggested that the mere fact that we can write (4.8a) and (4.8b) 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.

[81] In a general nonlinear model,

$$Y_1 = g_1(Y_2, X_1, X_2, U_1),$$

$$Y_2 = g_2(Y_1, X_1, X_2, U_2),$$

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

$$Y_1 = \varphi_1(X_1, X_2, U_1, U_2),$$

$$Y_2 = \varphi_2(X_1, X_2, U_1, U_2)$$

(which requires satisfaction of a local implicit function theorem). 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.

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).[82]

The intuition for these results is that if $\beta_{12} = 0$, we can vary Y_2 in Equation (4.8a) 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 (4.10b) if $\beta_{22} \neq 0$.[83] Symmetrically, by excluding X_1 from (4.8b), 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 (4.8a) and (4.8b) for causal analysis is invariance of the parameters $(\Gamma, \beta, \Sigma_U)$ to manipulations of the external variables. This invariance embodies the key idea in assumptions (PI-1)–(PI-4), which are versions of Hurwicz's invariance condition discussed in Section 4.6. 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.[84] We can use the population simultaneous equations model to define the class of admissible variations and address problems of definitions (task 1 of Table 1). If for a given model, the parameters of (4.8a) or (4.8b) 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 (4.8a) were not affected by the randomization, then Y_2 would be ex-

[82] 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. See Heckman (1978).

[83] Notice that we could also use U_2 as a source of variation in (4.10b) 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.

[84] Indeed matching or, equivalently, OLS in this context, using the right-hand side variables of (4.8a) and (4.8b), does not identify causal effects as Haavelmo (1943) established long ago.

ogenous in Equation (4.8b) and one could identify causal effects by least squares.[85] 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.

4.5.1. Relationship to Pearl's analysis

Controlled variation in external forcing variables is the key to defining causal effects in nonrecursive models. It is of some interest to readers of Pearl's influential book on causality (2000) to compare our use of the standard simultaneous equations model of econometrics in defining causal parameters to his. In the context of Equations (4.8a) and (4.8b), Pearl defines a causal effect by "shutting one equation down" or performing "surgery".

He implicitly assumes that "surgery", or shutting down an equation in a system of simultaneous equations, uniquely fixes one outcome or internal variable (the consumption of the other agent in our example). In general, it does not. Putting a constraint on one equation places a restriction on the entire set of internal variables. In general, no single equation in a system of simultaneous equations uniquely determines any single outcome variable. Shutting down one equation might also affect the parameters of the other equations in the system and violate the requirements of parameter stability.

A clearer manipulation that can justify Pearl's approach but shows its special character is to assume that it is possible to fix Y_2 by assuming that it is possible to set $\gamma_{21} = 0$. Assume that U_1 and U_2 are uncorrelated.[86] This together with $\gamma_{21} = 0$ makes the model recursive.[87] It assumes that agent 1 is unaffected by the consumption of agent 2. Under these assumptions, one can regress Y_1 on Y_2, X_1, and X_2 in the population and recover all of the causal parameters of (4.8a). Variation in U_2 breaks the perfect collinearity among Y_2, X_1, and X_2. In general, as we discuss in the next subsection, it is often not possible to freely set some parameters without affecting the rest of the parameters of a model.

Shutting down an equation or fiddling with the parameters in Γ is not required to *define* causality in an interdependent, nonrecursive system or to identify causal parameters. The more basic idea is *exclusion* of different external variables from different equations which, when manipulated, allow the analyst to construct the desired causal quantities.

One can move from the problem of definition (task 1 of Table 1) to identification (task 2) by using population analog estimation methods – in this case the method of

[85] Note that we are breaking the rules we set out in Section 2 in this example and elsewhere in this section by discussing tasks 1 and tasks 2 interchangeably.

[86] Alternatively, one can assume that it is possible to measure U_1 and control for it.

[87] For a discussion of recursive systems as devices for defining causality, see Wold (1956).

indirect least squares.[88] There are many ways other than through exclusions of variables to identify this and more general systems. Fisher (1966) presents a general analysis of identification in both linear and nonlinear simultaneous equations systems. Matzkin (2004, Chapter 73) substantially extends this literature.

4.5.2. *The multiplicity of causal effects that can be defined from a simultaneous equations system*

In the context of the basic nonrecursive model, there are many possible causal variations, richer than what can be obtained from the reduced form. Using the reduced form ($Y = X\Pi + \mathcal{E}$), one can define causal effects as *ceteris paribus* effects of variables in X or $\mathcal{E}$ on Y. This definition solves out for all of the intermediate effects of the internal variables on each other. Using the structure (4.7), one can define the effect of one internal variable on another holding constant the remaining internal variables and (X, U). We have established that such causal effects may not be defined within the rules specified for a particular structural model. Exclusions and other restrictions discussed in Fisher (1966) make definitions of causal effects possible under certain conditions.

One can, in general, solve out from the general system of equations for a subset of the Y (e.g., Y^* where $Y = (Y^*, Y^{**})$), using the reduced form of the model, and use *quasi-structural* models to define a variety of causal effects that solve out for some but not all of the possible causal effects of Y on each other. These quasi-structural models may be written as

$$\Gamma^{**}Y^{**} = \Pi^{**}X + U^{**}.$$

This expression is obtained by using the reduced form for component Y^*: $Y^* = \Pi^*X + \mathcal{E}^*$ and substituting for Y^* in (4.7). U^{**} is the error term associated with this representation. There are many possible quasi-structural models. Causal effects of internal variables may or may not be defined within them, depending on the assumed *a priori* information.

The causal effect of one component of Y^{**} on another does not fix Y^* but allows the Y^* components to adjust as the components of Y^{**} and the X are varied. Thus the Y^* are not being held fixed when X and/or components of the Y^{**} are varied. Viewed in this way, the reduced form and the entire class of quasi-structural models do not define any *ceteris paribus* causal effect relative to all of the variables (internal and external) in the original system since they do not fix the levels of the other Y (in the case of reduced forms) or Y^* (in the case of the quasi-structural models). Nonetheless, the reduced form may provide a good guide to predicting the effects of certain interventions that affect the external variables. The quasi-structural models may also provide a useful guide for predicting certain interventions, where components of Y^{**} are fixed by policy. The reduced

[88] Two-stage least squares would work as well.

form defines a net causal effect of variations in X as they affect the internal variables. There are many quasi-structural models and corresponding thought experiments.

This discussion demonstrates another reason why causal knowledge is provisional in addition to the *a priori* specification of the internal and external variables in this system. Different analysts may choose different subsystems of equations derived from (4.7) to work with and define different causal effects within the different possible subsystems. Some of these causal effects may not be identified, while others may be. Systems smaller or larger than (4.7) can be imagined. The role of *a priori* theory is to limit the class of models and the resulting class of counterfactuals and to define which ones are interesting. *Ceteris paribus* manipulations of one variable are meaningfully defined only if we specify the variables being manipulated and the variables being held constant. This is the position we have taken in Section 4.1.

In this section, we have exposited the Cowles Commission definition of structure. We now present a basic definition of structure in terms of invariance of equations to classes of interventions. Invariance is a central idea in causal analysis and policy analysis.

4.6. Structure as invariance to a class of modifications

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

The mechanisms generating counterfactuals and the choices of counterfactuals have already been characterized in Sections 4.1 and 4.3. Policies can act on preferences and the arguments of preferences (and hence choices), on outcomes $Y(s, \omega)$ and the determinants affecting outcomes or on the information facing agents. Recall that g_s, $s \in \mathcal{S}$, generates outcomes while f_s, $s \in \mathcal{S}$, generates subjective evaluations.[89] Specifically,

(i) *Policies can shift the distributions of the determinants of outcomes and choices* (Q, Z, X, U, η), *where* $Q = \{Q(s, \omega)\}_{s\in\mathcal{S}}$, $Z = \{Z(s, \omega)\}_{s\in\mathcal{S}}$, $X = \{X(s, \omega)\}_{s\in\mathcal{S}}$, $\eta = \{\eta(s, \omega)\}_{s\in\mathcal{S}}$ *and* $U = \{U_s(\omega)\}_{s\in\mathcal{S}}$ *in the population. This may entail defining the* g_s *and* f_s *over new domains. Let* $\mathcal{X} = (Q, Z, X, U, \eta)$ *be sets of arguments of the determinants of outcomes. Policies shifting the distributions of these variables are characterized by maps* $T_\chi : \chi \mapsto \chi'$.

(ii) *Policies can select new* f, g *or* $\{f_s, g_s\}_{s\in\mathcal{S}}$ *functions. In particular, new arguments (e.g., amenities or characteristics of programs) may be introduced as a result of policy actions creating new attributes. Policies shifting functions map*

[89] By f_s, we mean s-specific valuation functions.

f, g or $\{f_s, g_s\}_{s\in\mathcal{S}}$ into new functions $T_f : f_s \mapsto f'_s$; $T_g : g_s \mapsto g'_s$. This may entail changes in functional forms with a stable set of arguments as well as changes in arguments of functions.

(iii) *Policies may affect individual information sets $(\mathcal{I}_\omega)_{\omega\in\Omega}$. $T_{\mathcal{I}_\omega} : \mathcal{I}_\omega \mapsto \mathcal{I}'_\omega$.*

Clearly, any particular policy may incorporate elements of all three types of policy shifts.

Parameters of a model or parameters derived from a model are said to be policy invariant with respect to a class of policies if they are not changed (are invariant) when policies within the class are implemented. We have explicitly introduced such invariance in our discussion of the Cowles version of the structural model with respect to policies that change X, but not for policies that change the distribution of U. This notion is partially embodied in assumptions (PI-1) and (PI-2), which are defined solely in terms of *ex post* outcomes. More generally, policy invariance for f, g or $\{f_s, g_s\}_{s\in\mathcal{S}}$ requires for a class of policies $\mathcal{P}_A \subseteq \mathcal{P}$,

(PI-5) *The functions f, g or $\{f_s, g_s\}_{s\in\mathcal{S}}$ are the same for all values of the arguments in their domain of definition no matter how their arguments are determined, for all policies in $\mathcal{P}_A$.*

This definition is a version of (PI-3) and (PI-4) for the specific notation of the choice model developed in this chapter and for specific types of policies. This definition can be made separately for f, g, f_s, g_s or any function derived from them. It requires that when we change an argument of a function its value is the same for the same change of input irrespective of how we change it. It is defined relative to a class of policies and not necessarily for all policies.

In the econometric approach to policy evaluation, the analyst attempts to model how a policy shift affects outcomes without reestimating any model. Thus, for the tax and labor supply example presented in Section 4.1, with labor supply function $h_s = h(w(1-s), x, u_s)$, it is assumed that we can shift tax rate s without affecting the functional relationship mapping $(w(1-s), x, u_s)$ into h_s. If, in addition, the support of $w(1-s)$ under one policy is the same as the support determined by the available economic history, for a class of policy modifications (tax changes), the labor supply function can be used to accurately predict the outcomes for that class of tax policies. It would not be able to accurately forecast policies that extend the support of h_s to a new domain or if it shifts preferences in a way never previously experienced (e.g., by appealing to patriotism in time of war). In such cases, the domains of f and g would have to be extended to accurately forecast policy changes, and additional assumptions would have to be made. We discuss such assumptions in Chater 71 of our contribution to this Handbook.

In the simultaneous equations model analyzed in the last subsection, invariance requires stability of Γ, B and Σ_U to interventions. Such models can be used to accurately forecast the effects of policies that can be cast as variations in the inputs to that model

that keep the parameters invariant. Policy invariant parameters are not necessarily causal parameters as we noted in our analysis of reduced forms in the preceding section. Thus, in the simultaneous equations model, depending on the *a priori* information available, it may happen that no causal effect of one internal variable on another may be defined but if Π is invariant to modifications in X, the reduced form is policy invariant for those modifications. The class of policy invariant parameters is thus distinct from the class of causal parameters, but invariance is an essential attribute of a causal model. For counterfactuals $Y(s, \omega)$, if assumption (PI-1) is not postulated for a class of policies $\mathcal{P}_A$, all of the treatment effects defined in Section 2 would be affected by policy shifts.

Rubin's SUTVA assumptions (R-2) and (R-3) are versions of Hurwicz's (1962) invariance assumptions for the functions generating objective outcomes. Thus Rubin's assumption (R-3) postulates that $Y(s, \omega)$ is invariant to all policies that change f but does not cover policies that change g or the support of Q. Within the treatment effects framework, a policy that adds a new treatment to $\mathcal{S}$ is not policy invariant for treatment parameters comparing the new treatment to any other treatment unless the analyst can model all policies in terms of a generating set of common characteristics specified at different levels, as in formulation (4.2) or our example in Section 3.3. The lack of policy invariance makes it potentially misleading to forecast the effects of new policies using treatment effect models.

"Deep structural" parameters generating the f and g are invariant to policy modifications that affect technology, constraints and information sets except when the policies extend the historical supports. Invariance can only be defined relative to a class of modifications and a postulated set of preferences, technology, constraints and information sets. Thus causal parameters can only be precisely identified within a class of modifications.

4.7. Alternative definitions of "structure"

The terms "structural equation" or "structure" are used differently by different analysts and are a major source of confusion in the policy analysis literature. In this section, we briefly distinguish three other definitions of structure besides our version of Hurwicz (1962). The traditional Cowles Commission structural model of econometrics was presented in Section 4.5. It is a nonrecursive model for defining and estimating causal parameters. It is a useful vehicle for distinguishing effects that can be defined in principle (through *a priori* theory) from effects that are identifiable from data. This is the contrast between tasks 1 and 2 of Table 1. The framework arose as a model to analyze the economic phenomenon of supply and demand in markets, and to analyze policies that affected price and quantity determination.

A second definition of structure, currently the most popular in the applied economics literature, defines an equation as structural if it is derived from an explicitly formulated economic theory. Consider a consumer demand problem where a consumer ω chooses among goods $X(\omega)$ given money income $M(\omega)$ and prices P, $P'X(\omega) \leqslant M(\omega)$. Pref-

erences of ω, $R(X(\omega), \omega)$, are quasiconcave in $X(\omega)$ and twice differentiable. Many economists would say that $R(X(\omega), \omega)$ is structural because it describes the preferences of agent ω. There would be similar agreement that technology parameters are structural parameters.

When we solve for the demand functions, under standard conditions, we obtain $X(\omega) = X(\frac{P}{M(\omega)}, \omega)$. These are sometimes called "reduced form" expressions by analogy with the Cowles Commission simultaneous equations literature exposited in Section 4.5, assuming that prices normalized by income are exogenous. While any convention is admissible, this one is confusing since we can recover the preferences (up to a monotonic function) given the demand function under standard regularity conditions [see, e.g., Varian (1978)]. Is the indirect utility function $\widetilde{R}^*(\omega, \frac{P}{M(\omega)}) = R(X(\frac{P}{M(\omega)}), \omega) = R^*(\frac{P}{M(\omega)}, \omega)$ structural or reduced form?

While the notion of structure in this widely applied usage is intuitively clear, it is not the same notion of structure as used in Cowles Commission econometrics as defined in Section 4.5. It is structural in the sense that the internal variables (the X in this example) are substituted out for externally specified (to the consumer) P and M. At the market level, this distinction is not clear cut since X and P are jointly determined. The notion of a "reduced form" is not clearly specified until the statistical properties of X, P or M have been specified. Recall that the Cowles Commission definition of reduced form (a) solves out the X in terms of P and M and (b) assumes that P and M are "exogenous" relative to the unobserved variables. In current popular usage, a reduced form makes both assumptions.

A third definition of a structural model is as a finite parameter model. Structural in this sense means low dimensional and is not related to the endogeneity of any variable or the economic interpretation placed on the equations. Clearly the Cowles Commission model is finite dimensional if the dimensions of Y and X are finite. Nonlinear finite parameter versions of the Cowles Commission models as in Fisher (1966) are also structural in these systems. Systems that are structural in this sense are useful for extrapolation of functions out of their empirical supports.

A more basic definition of a system of structural equations, and the one featured in this chapter, is a system of equations invariant to a class of modifications. Without such invariance one cannot trust the models to forecast policies or make causal inferences. Invariance to modifications requires a precise definition of a policy, a class of policy modifications and specification of a mechanism through which policy operates. It makes clear that "structure" is a concept that is relative to the potential policy changes studied by the analyst. A system structural for one class of policy modifications may not be structural for another.

4.8. Marschak's Maxim and the relationship between the structural literature and the statistical treatment effect literature

The absence of explicit models of outcomes and choice is a prominent feature of the statistical treatment effect literature. A major goal of this chapter and our other chap-

ter in this Handbook is to infuse economics into the treatment effect literature and to understand its achievements and implicit identifying assumptions in economic terms. Economically 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 these models, like all scientific 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 welfare effects of policies; (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, this literature makes fewer statistical assumptions in terms of exogeneity, functional form, exclusion and distributional assumptions than the standard structural estimation literature in econometrics. These are the attractive features of this approach.

In reconciling these two literatures, we reach back to a neglected but important paper by Marschak (1953). Marschak noted that for many questions of policy analysis, it is not necessary to identify fully specified economic 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). Thus in the simultaneous equations system example presented in Section 4.5, policies that only affect X may be forecasted using reduced forms, not knowing the full structure, provided that the reduced forms are invariant to the modifications being considered.[90] Forecasting other policies may only require partial knowledge of the full simultaneous equations system.

We 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 (problem P-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 economic question being addressed by the particular treatment effect being identified. When the treatment effect literature does not clearly specify the economic question being addressed, it does not implement Marschak's Maxim.

Population mean treatment parameters are often identified under weaker conditions than are traditionally assumed in structural econometric analysis. Thus to identify the average treatment effect for s and s' we only require $E(Y(s,\omega) \mid X = x) - E(Y(s',\omega) \mid X = x)$. Under (PI-1) and (PI-2), this parameter answers the policy question of determining the average effect on outcomes of moving an agent from s' to s. The parameter

[90] Thus we require that the reduced form Π defined in Section 4.5 does not change when we change the X.

is not designed to evaluate a whole host of other policies. We do not have to know the functional form of the generating g_s functions nor does X have to be exogenous. We do not have to invoke the stronger conditions (PI-3) and (PI-4) about invariance of the choice equations.

However, if we seek to identify $E(Y(s,\omega) \mid X = x,\ D(s,\omega) = 1) - E(Y(s',\omega) \mid X = x, D(s,\omega) = 1)$, we need to invoke versions of (PI-3) and (PI-4) because we condition on a choice. We do not condition on a choice in defining the average treatment effects.

Explicitly formulated economic models or low dimensional economic or statistical models may or may not be structural in the sense defined in this chapter. They may be invariant to some policy modifications but not to others.

Causal models are defined independently of any particular policy manipulation. But if the variations in the arguments of the causal (Marshallian) functions correspond to variations in some policy, causal models as we have defined them, are structural since by definition, causal functions are invariant to variations in the arguments of the functions that generate them.

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 answer policy question P-1.

Invariant, explicitly formulated, economic models are useful for addressing policy problems P-2 and P-3: extrapolation and predicting the effects of new policies, respectively. Invariant low dimensional models are sometimes useful for solving extrapolation problem P-2.

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 discussed in Section 4.2. Recall that 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 an 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.[91] We do not need to separate out the causal effect of X on Y, β, from the effect of X on the unobservables operating through $E(U \mid X)$.

Viewed in this light, the treatment effect literature that compares the outcome associated with $s \in \mathcal{S}$ with the outcome associated with $s' \in \mathcal{S}$ seeks to recover a causal effect of s relative to s'. It is a particular causal effect for a particular set of policy interventions. It seeks effects that hold all other factors, observed and unobserved, constant.

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

[91] See, e.g., Goldberger (1964).

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.

Marschak's Maxim is an application of Occam's Razor to policy evaluation. For certain classes of policy interventions, designed to answer problem P-1, the treatment effect approach may be very powerful and more convincing than explicit economic models which require more assumptions.

Considerable progress has been made in relaxing the parametric structure assumed in the early explicitly economic models [see Matzkin (1994), and Chapter 73 of this Handbook]. As the treatment effect literature is extended to address the more general set of policy forecasting problems entertained in the explicitly economic literature, the distinction between the two approaches will vanish although it is currently very sharp. This chapter, Heckman and Vytlacil (2005) and Heckman (2007) are attempts to bridge this gulf.

Up to this point in the chapter, everything that has been discussed precisely is purely conceptual although we have alluded to empirical problems and problems of identification going from data of various forms to conceptual models. We now discuss the identification problem, which must be solved if causal models are to be empirically operational.

5. Identification problems: Determining models from data

Unobserved counterfactuals are the source of the problems considered in this chapter. For an agent ω in state s, we observe $Y(s, \omega)$ but not $Y(s', \omega)$, $s' \neq s$. A central problem in the literature on causal inference is how to identify counterfactuals and the derived treatment parameters. Unobservables, including missing data, are at the heart of the identification problem for causal inference. As we have seen, counterfactuals play a key role in structural policy analysis.

Different evaluation estimators differ in the amount of knowledge they assume that the analyst has relative to what the agents being studied have when making their program enrollment decisions (or their decisions are made for them as a parent for a child). This distinction is a matter of the quality of the available data. Unless the analyst has access to all of the relevant information that produces the dependence between outcomes and treatment rules (i.e., that produces selection bias), he/she must devise methods to control for the unobserved components of relevant information. We define relevant information precisely in Chapter 71. Loosely speaking, relevant information is the information which, if available to the analyst and conditioned on, would eliminate selection bias.

There may be information known to the agent but not known to the observing analyst that does not give rise to the dependence between outcomes and choices. It is the infor-

mation that gives rise to the dependence between outcomes and treatment choices that matters for eliminating selection bias, and this is the relevant information.

A priori one might think that the analyst knows a lot less than the agent whose behavior is being analyzed. At issue is whether the analyst knows less *relevant* information, which is not so obvious, if only because the analyst can observe the outcomes of decisions in a way that agents making decisions cannot. This access to *ex post* information can sometimes give the analyst a leg up on the information available to the agent.

Policy forecasting problems P-2 and P-3 raise the additional issue that the support over which treatment parameters and counterfactuals are identified may not correspond to the support that is required to construct a particular policy counterfactual. Common to all scientific models, there is the additional issue of how to select (X, Z), the conditioning variables, and how to deal with them if they are endogenous. Finally, there is the problem of lack of knowledge of functional forms of the models. Different econometric methods solve these problems in different ways. We first present a precise discussion of identification before we turn to a discussion of these issues and how they affect the properties of different evaluation estimators.

5.1. The identification problem

The identification problem asks whether theoretical constructs have any empirical content in a hypothetical population or in real samples. By empirical content, we mean whether the model is uniquely determined by the available data. This formulation considers tasks two and three in Table 1 together, although some analysts like to separate these issues, focusing solely on task two (identification in large samples). The identification problem considers what particular models within a broader class of models are consistent with a given set of data or facts. Specifically, consider a model space M. This is the set of admissible models that are produced by some theory for generating counterfactuals. Elements $m \in M$ are admissible theoretical models.

We may only be interested in some features of a model. For example, we may have a rich model of counterfactuals $\{Y(s, \omega)\}_{s \in \mathcal{S}}$, but we may only be interested in the average treatment effect $E[Y(s, \omega) - Y(s', \omega)]$. Let the objects of interest be $t \in T$, where "t" stands for the target – the goal of the analysis. The target space T may be the whole model space M or something derived from it, a more limited objective.

Define map $g : M \to T$. This maps an element $m \in M$ into an element $t \in T$. In the example in the preceding paragraph, T is the space of all average treatment effects produced by the models of counterfactuals. We assume that g is onto.[92] Associated with each model is an element t derived from the model, which could be the entire model itself. Many models may map into the same t so the inverse map (g^{-1}), mapping T

[92] By this, we mean that for every $t \in T$, there is an element $m \in M$ such that g sends m to t, i.e., the image of M by g is the entire set T. Of course, g may send many elements of M to a single element of T. Note that g as used here is not necessarily the same g as used in Section 4.

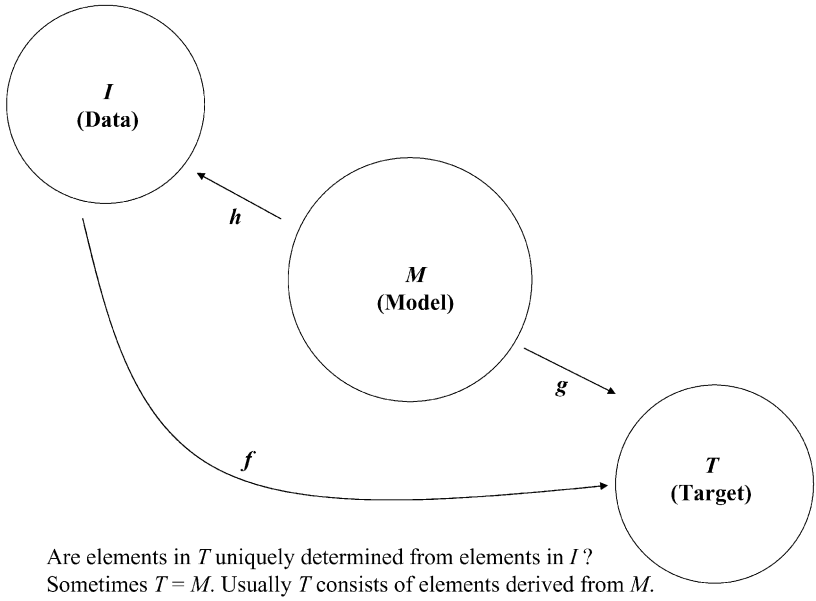


Figure 4. Schematic of model (M), data (I) and target (T) parameter spaces.

to M, may not be well defined. Thus many different models may produce the same average treatment effect.

Let the class of possible information or data be I. Define a map $h: M \to I$. For an element $i \in I$, which is a given set of data, there may be one or more models m consistent with i. If i can only be mapped into a single m, the model is exactly identified.[93] If there are multiple m's, consistent with i, these models are not identified. Thus, in Figure 4, many models (elements of M) may be consistent with the same data (single element of I).

Let $M_h(i)$ be the set of models consistent with i. $M_h(i) = h^{-1}(\{i\}) = \{m \in M: h(m) = i\}$. The data i reject the other models $M \setminus M_h(i)$, but are consistent with all models in $M_h(i)$. If $M_h(i)$ contains more than one element, the data produce set-valued instead of point-valued identification. If $M_h(i) = \emptyset$, the empty set, no model is

[93] Associated with each data set i is a collection of random variables $Q(i)$, which may be a vector. Let $F_Q(q \mid m)$ be the distribution of Q under model m. To establish identification on nonnegligible sets, one needs that, for some true model m^*,

$$\Pr\bigl(\bigl|F_Q(q \mid m^*) - F_Q(q \mid m)\bigr| > \varepsilon\bigr) > 0$$

for some $\varepsilon > 0$ for all $m \neq m^*$. This guarantees that there are observable differences between the data generating process for Q given m and for Q given m^*. We can also define this for $F_Q(q \mid t^*)$ and $F_Q(q \mid t)$. Note that Q is an abstract random variable and not necessarily the specific attributes defined in Section 4.

consistent with the data. By placing restrictions on models, we can sometimes reduce the number of elements in $M_h(i)$ if it has multiple members. Let $RE \subset M$ be a set of restricted models. Thus it is sometimes possible by imposing restrictions to reduce the number of models consistent with the data. Recall that in the two-agent model of social interactions, if $\beta_{12} = 0$ and $\beta_{21} = 0$, we could uniquely identify the remaining parameters under the other conditions maintained in Section 4.5. Thus $RE \cap M_h(i)$ may contain only a single element. Another way to solve this identification problem is to pick another data source $i' \in I$, which may produce more restrictions on the class of admissible models. More information provides more hoops for the model to jump through.

Going after a more limited class of objects such as features of a model ($t \in T$) rather than the full model ($m \in M$) is another way to secure unique identification. Let $M_g(t) = g^{-1}(\{t\}) = \{m \in M\colon\ g(m) = t\}$. Necessary and sufficient conditions for the existence of a unique map $f : I \to T$ with the property $f \circ h = g$ are (a) h must map M *onto* I and (b) for all $i \in I$, there exists $t \in T$ such that $M_h(i) \subseteq M_g(t)$. Condition (b) means that even though one element $i \in I$ may be consistent with many elements in M, so that $M_h(i)$ consists of more than one element, it may be that all elements in $M_h(i)$ are mapped by g into a single element of T. The map f is onto since $g = f \circ h$ and g is onto by assumption. In order for the map f to be one-to-one, it is necessary and sufficient to have equality of $M_h(i)$ and $M_g(t)$ instead of simply inclusion.

If we follow Marschak's Maxim and focus on a smaller target space T, it is possible that g maps the admissible models into a smaller space. Thus the map f described above may produce a single element even if there are multiple models m consistent with the data source i that would be required to answer broader questions. This could arise, for example, if for a given set of data i, we could only estimate the mean μ_1 of Y_1 up to a constant q and the mean μ_2 of Y_2 up to the same constant q. But we could uniquely identify the element $\mu_1 - \mu_2 \in T$.[94] In general, identifying elements of T is easier than identifying elements of M. Thus, in Figure 4, even though many models (elements of M) may be consistent with the same $i \in I$, only one element of T may be consistent with that i. We now turn to empirical causal inference and illustrate the provisional nature of causal inference.

5.2. *The sources of nonidentifiability*

The principle source of identification problems for policy problems P-1–P-3 is the absence of data on outcomes other than the one observed for the agent. Thus if agent ω is observed in state s we observe $Y(s, \omega)$ but not $Y(s', \omega)$, $s' \in \mathcal{S}$, $s \neq s'$. If we had data

[94] Most modern analyses of identification assume that sample sizes are infinite, so that enlarging the sample size is not informative. However, in any applied problem this distinction is not helpful. Having a small sample (e.g., fewer observations than regressors) can produce an identification problem. Our definition of identification addresses task two and task three together if we assume that samples are finite.

on the outcomes for agents in all states in $\mathcal{S}$, we could form *ex post* counterfactuals and solve P-1. We still need to value these counterfactuals (i.e., construct $R(Y(s, \omega))$).

Even with such ideal data, it is necessary to extend $\{Y(s, \omega)\}_{s \in \mathcal{S}}$ and the appropriate valuation functions to new supports to answer policy questions P-2 and P-3. For many econometric estimators, it is necessary to account for the limited supports available in many empirical samples. One can only meaningfully compare comparable agents. A nonparametric approach to estimation guarantees that this condition is satisfied. Respecting empirical support conditions restricts the class of identified parameters, even considering only problem P-1. As we will discuss below, failure of support conditions plagues different estimators and estimation strategies.

Another source of identification problems is the uncertainty regarding the choice of the conditioning variables (the X, W and Z) in any application. This problem is intrinsic to all estimation problems. It affects some estimators more than others, as we note in Chapter 71. For some estimators and for some policy problems, the endogeneity of the regressors is a major concern. We delineate these problems for each estimator and each policy problem. Closely related is the asymmetry in the information available to analysts and the agents they study which we previously discussed. This entails the problem of specifying the information on which agents condition their actions, distinguishing them from the information available to the econometrician and accounting for any information shortfalls. For example, the method of matching makes strong assumptions about the information available to analysts which cannot be verified but which drive the interpretation of the results.

There is also the problem of functional forms. Many traditional approaches to the construction of structural models and econometric counterfactuals make assumptions about the functional forms of outcome equations and choice equations and the distributions of the unobservables. Methods differ in their reliance on these functional forms. Lack of knowledge of the required functional forms is a source of identification problems.

Table 3
Sources of identification problems considered in this chapter

(i) Absence of data on $Y(s', \omega)$ for $s' \in S \backslash \{s\}$ where s is the state selected (the evaluation problem).
(ii) Nonrandom selection of observations on states (the selection problem).
(iii) Support conditions may fail (outcome distributions for $F(Y_s \mid X = x)$ may be defined on only a limited support of X so $F(X \mid D_s = 1)$ and $F(X \mid D_{s'} = 1)$ have different X supports or limited overlap in their supports).
(iv) Functional forms of outcome equations and distributions of unobservables may be unknown. To extend some function $Y = G(X)$ to a new support requires functional structure: It cannot be extended outside of sample support by a purely nonparametric procedure.
(v) Determining the (X, Z, W) conditioning variables.
(vi) Different information sets for the agent making selection $\mathcal{I}_a$ and the econometrician trying to identify the model $\mathcal{I}_e$ where $\mathcal{I}_a \neq \mathcal{I}_e$.

Table 3 lists the major sources of identification problems. We discuss the sensitivity of alternative evaluation methods to this array of problems in Chapter 71. We next present an identification analysis of our prototypical economic model of choice and outcomes which serves as a benchmark model against which we can formulate the implicit assumptions made in alternative econometric approaches to policy evaluation.

6. Identification of explicit economic models

For the Roy model developed in Section 3, Heckman and Honoré (1990), show that under the conditions they specify it is possible to identify the distribution of treatment outcomes $(Y_1 - Y_0)$ without invoking functional form assumptions. Randomization can only identify the marginal distributions of Y_0 and of Y_1 and not the joint distribution of $(Y_1 - Y_0)$ or the quantiles of $(Y_1 - Y_0)$ [see Heckman (1992)]. Thus, under its assumptions, the Roy model is more powerful than randomization in producing the distributional counterfactuals discussed in Abbring and Heckman (Chapter 72).[95] The role of the choice equation is to motivate and justify the choice of an evaluation method.[96] This is a central feature of the econometric approach that is missing from the statistical and epidemiological literature on treatment effects.

Considerable progress has been made in relaxing the parametric structure assumed in the early structural models. As the treatment effect literature is extended to address the more general set of policy forecasting problems entertained in the structural literature (especially problems P-2 and P-3), the distinction between the two literatures will vanish. This section presents some examples of traditional structural models, how they can be used to construct treatment effects, and how treatment effects can be generated under much weaker conditions.

6.1. Using parametric assumptions to generate population level treatment parameters

We now present a brief analysis of identification of the extended Roy model and the generalized Roy model analyzed in Section 3.3. This framework provides a convenient platform from which to summarize the power and limitations of the current literature in structural economics. Matzkin (Chapter 73 of this Handbook) provides a comprehensive discussion of identification. Write a two-sector model with outcomes Y_1, Y_0 under perfect certainty as

$$Y_1 = \mu_1(X, U_1), \tag{6.1a}$$

$$Y_0 = \mu_0(X, U_0) \tag{6.1b}$$

[95] The same analysis applies to matching, which cannot identify the distribution of $(Y_1 - Y_0)$ or derived quantiles.

[96] See Heckman and Robb (1985, 1986).

and costs

$$C = \mu_C(W, U_C). \tag{6.1c}$$

Agents choose sector 1 if $R = Y_1 - Y_0 - C \geqslant 0$. Otherwise they choose sector 0. We have shown in Section 3 how this model can be used to generate the common treatment effects discussed in Section 2. At issue in this section is how to identify the parameters of Equations (6.1a)–(6.1c) from data where only one outcome (Y_1 or Y_0) is observed. Recent advances in microeconometrics allow nonparametric identification of these equations and the distributions of (U_0, U_1, U_C) under conditions we specify below.

First consider identification of the two-outcome generalized Roy model for normal error terms developed in Section 3.3. Suppose that we observe Y_1 when $D = 1$ and Y_0 when $D = 0$. Observed Y may be written in switching regression form as in Quandt (1958, 1972):

$$Y = DY_1 + (1 - D)Y_0.$$

We assume that the analyst observes (Z, X, Y, D), where $Z = (X, W)$. In addition to assumptions (i)–(ii) given in Section 3.3 and Equations (3.4a)–(3.4c), we assume that the model is of full rank.

One traditional approach to econometric identification [see Heckman and Robb (1985, 1986)] is to solve the selection problem for Y_1 and Y_0 and then to use the parameters of the model to solve the evaluation problem. Solutions to the selection problem are developed in Heckman (1976, 1979, 1990), Heckman and Honoré (1990) and are popularized in numerous surveys [see, e.g., Maddala (1983)]. Summarizing known results, assuming $Y_1 = \mu_1(x) + U_1$ and $Y_0 = \mu_0(x) + U_0$, $C = W\varphi + U_C$, and defining $\upsilon = U_1 - U_0 - U_C$, and normalizing $\mathrm{Var}(\upsilon) = 1$,

$$E(Y_1 \mid D = 1,\ X = x,\ Z = z) = x\beta_1 + \mathrm{Cov}(U_1, \upsilon)\lambda(z\gamma),$$
$$E(Y_0 \mid D = 0,\ X = x,\ Z = z) = x\beta_0 + \mathrm{Cov}(U_0, \upsilon)\tilde{\lambda}(z\gamma),$$

where $\lambda(z\gamma) = \varphi(z\gamma)/\Phi(z\gamma)$ and $\tilde{\lambda}(z\gamma) = -\varphi(z\gamma)/\Phi(-z\gamma)$. We can identify γ from a first stage discrete choice analysis (a probit analysis with D as the dependent variable and Z as the regressor) if the Z are of full rank. Under additional rank conditions on the X, we can form $\lambda(z\gamma)$ and $\tilde{\lambda}(z\gamma)$ and use linear regression to recover β_1, $\mathrm{Cov}(U_1, \upsilon)$, β_0, $\mathrm{Cov}(U_0, \upsilon)$ from the conditional means of Y_1 and Y_0. As first proved by Heckman (1976, 1979), we can use the residuals from the regression equations to identify σ_0^2 and σ_1^2. We can also identify the covariances $\sigma_{1\upsilon}$ and $\sigma_{0\upsilon}$ from the coefficients on $\lambda(z\gamma)$ and $\tilde{\lambda}(z\gamma)$ respectively. Without further information, we cannot recover σ_{01} and hence the joint distribution of (Y_0, Y_1). Thus the model is not fully identified, although the marginal distributions are uniquely identified.[97]

[97] Vijverberg (1993) uses a sensitivity or bounding analysis to determine what classes of joint distributions are consistent with the data.

The lack of identification of the joint distribution does not preclude identification of the mean treatment parameters introduced in Sections 2 and 3. Note further that it is possible that there is selection bias for Y_1 ($\mathrm{Cov}(U_1, \upsilon) \neq 0$) and selection bias for Y_0 ($\mathrm{Cov}(U_0, \upsilon) \neq 0$) but no selection on gains $\mathrm{Cov}(U_1 - U_0, \upsilon) = 0$.

Using the analysis of Section 3.3 from the parameters that are identified from selection models, we can identify $\mathrm{ATE}(x), \mathrm{TT}(x, z), \mathrm{MTE}(x)$ from cross section data. Without further information, we cannot identify the joint distribution of the counterfactuals $F(y_1 - y_0 \mid X)$ nor can we determine the proportion of agents who benefit from treatment not accounting for costs $\Pr(Y_1 \geqslant Y_0 \mid Z)$. We *can* identify the proportion of agents who benefit accounting for their costs using choice or revealed preference data:

$$\Pr(Y_1 - Y_0 - C \geqslant 0 \mid Z = z) = \Phi(z\gamma).$$

In the special case of the Roy model, where $\upsilon = U_1 - U_0$, because we can identify the variance of U_1 and U_0, from the coefficients on $\lambda(z\gamma)$ and $\tilde{\lambda}(z\gamma)$, we can identify $\mathrm{Cov}(U_1, U_1 - U_0)$ and $\mathrm{Cov}(U_0, U_1 - U_0)$ and hence we can identify σ_{01}. Thus we can identify the proportion of agents who benefit from treatment, not including costs, because there are no costs and it is the same as $\Phi(z\gamma) = \Pr(Y_1 - Y_0 \geqslant 0 \mid Z = z)$.[98] By using choice data, the Roy model, under its assumptions, produces more information than randomization which only identifies the marginal distributions of Y_0 and Y_1 and not the joint distribution.

Without additional information, one cannot surmount the fundamental evaluation problem that one does not observe both Y_0 and Y_1 for the same agents. The Roy model overcomes this problem using choice data assuming that there are no costs of participation. If it is assumed that $U_C = 0$ but there are observed costs, one can identify γ as before, and identify the covariance σ_{01} because no new random variable enters the cost equation that is not in the outcome equation. This framework is what we call the extended Roy model. For this version of the generalized Roy model one can form all of the distributional treatment effects using the preceding analysis. In general, however, one cannot identify the joint distribution of (Y_1, Y_0) but one can identify the distributions of (Y_1, R) in the notation of Section 3 (or (U_1, υ)) and (Y_0, R) (or (U_0, υ)).

Normality assumptions are traditional and convenient. The linearity, exogeneity, separability and normality assumptions make it possible to solve policy forecasting problems P-1–P-3. By parameterizing the β_i to depend on Q_i as in Equations (3.7a)–(3.7b), it is possible to forecast the demand for new goods. The support problems that plague nonparametric estimators are absent. Heckman, Tobias and Vytlacil (2001, 2003) extend the normal model using alternative distributional assumptions. The normal selection model extends standard normal regression theory intuitions in a natural way. But

[98] If we only observe Y_1 or Y_0 but not both in the same sample, we can identify the covariance of (U_1, U_0) provided we normalize a mean (e.g., the mean of the missing Y). Thus if Y_1 is the market wage and Y_0 is the reservation wage, we rarely directly observe Y_0 but we observe Y_1. See Heckman (1974) and Heckman and Honoré (1990).

they are controversial. A huge effort in econometrics in the past 20 years has gone into relaxing these assumptions.[99]

6.2. *Two paths toward relaxing distributional, functional form and exogeneity assumptions*

At issue in this Handbook is whether the strong exogeneity, linearity and normality assumptions in the conventional literature in econometrics are required to form treatment effects and to evaluate policy. They are not. After this point of agreement, the recent literature on policy evaluation divides. The literature in microeconometric structural estimation focuses on relaxing the linearity, separability, normality and exogeneity conditions invoked in the early literature in order to identify (6.1a)–(6.1c) under much weaker conditions.

Recent advances in econometric theory greatly weaken the distributional and functional form assumptions maintained in the early econometric literature on selection bias. For example, Cosslett (1983), Manski (1988), and Matzkin (1992, 1993, 1994, 2003) relax the distributional assumptions required to identify the discrete choice model. Matzkin (1993) develops multivariate extensions. She surveys this literature in her 1994 Handbook Chapter. Heckman (1980, 1990), Heckman and Robb (1985, 1986), Heckman and Honoré (1990), Ahn and Powell (1993), Heckman and Smith (1998) and Carneiro, Hansen and Heckman (2003) present conditions for nonparametric and semiparametric identification of the selection model. Powell (1994) presents a useful survey for developments up to the early 1990s. Developments by Chen (1999) extend this analysis. Heckman (1990), Heckman and Smith (1998) and Carneiro, Hansen and Heckman (2003) show how to identify all of the mean treatment parameters as well as the distributional treatment parameters. We review the work on estimating distributions of treatment effects in Abbring and Heckman (Chapter 72).

Appendix B presents a formal nonparametric analysis of identification of the prototypical model of choice and outcomes developed in Section 3.1. From this and other explicitly economic models, the mean treatment effects and many distributional treatment effects discussed in Section 2 can be identified. For reasons discussed in the preceding subsection, one cannot form the joint distribution of outcomes across treatment states without some additional information such as the special Roy structure. Abbring and Heckman (Chapter 72) show how restrictions on the dimensionality of the unobservables and extra information can also produce identification of the joint distribution of $Y_1 - Y_0$. Matzkin (Chapter 73) provides a guide to the recent literature on nonparametric identification in explicitly economic models. The goal of this line of work is to

[99] The motivation for this research is largely based on Monte Carlo examples by Goldberger (1983), Arabmazar and Schmidt (1982) and others. In the study of earnings models with truncation and censoring, log normality is a good assumption [see Heckman and Sedlacek (1985)]. In the study of labor supply, it is a very poor assumption [see Killingsworth (1983), and the articles in the special issue of the *Journal of Human Resources* on labor supply and taxation, 1990]. See the evidence summarized in Heckman (2001).

preserve the economic content of the original Roy and generalized Roy models to collate evidence across studies in order to interpret evidence using economics, as well as to forecast the effects of new policies.

The recent literature on treatment effects identifies population level treatment effects under weaker conditions than are invoked in the traditional normal model. It does not aim to recover the structural parameters generating (6.1a)–(6.1c) but rather just certain derived objects, such as the mean treatment effects. These are taken as the invariant structural parameters. The class of modifications considered is the set of treatments in place.

Consider identification of ATE. It is not necessary to assume that X is exogenous if one conditions policy analysis on X and does not seek to identify the effect of changing X. The model of outcomes does not have to be separable in observables and unobservables. We can nonetheless identify ATE under very general conditions.

One transparent way is by randomization, discussed in Chapter 71. If agents of given X are randomized into sectors 1 and 0, and there is compliance with the randomization protocols, we can identify ATE by comparing the mean outcomes of agents randomized into sector 1 with the outcomes of those randomized into sector 0:

$$\text{ATE}(x) = E(Y_1 \mid X) - E(Y_0 \mid X).$$

Matching, discussed in Chapter 71, also identifies ATE without making any assumptions about the distributions of (U_1, U_0, U_C) or the functional forms of the relationships generating outcomes and choices (6.1a)–(6.1c) but assuming that conditioning on X randomizes choices and produces the same data as are generated from an experiment. By focusing on one treatment parameter, in this case ATE, and the questions ATE answers, we can proceed under weaker conditions than were used to develop the selection model although finding a common support for X when $D = 1$ and X when $D = 0$ may be a serious practical issue [see Heckman, Ichimura and Todd (1998)]. In general, matching or randomization do not identify TT or MTE.

ATE answers only one of the many evaluation questions that are potentially interesting to answer. But we can identify ATE under weaker assumptions than are required to identify the full generalized Roy model. Our analysis of ATE is an application of Marschak's Maxim. Doing one thing well has both its advantages and disadvantages. Many of the estimators proposed in the evaluation literature identify some parameters, and not others.

Our strategy in Chapter 71 of this Handbook is to survey the existing literature that relaxes normality assumptions in conducting policy evaluation but that preserves the index structure motivated by economic theory that is at the core of the generalized Roy model and its extensions. The goal is to present a unified analysis of the available models of treatment choice and treatment outcomes, and to unify the analysis of alternative estimation strategies using a nonparametric index model framework. This limits the generality of our survey. At the same time, it links the treatment literature to economic choice theory and so bridges the structural and treatment effect approaches.

Thus, in Chapter 71, we present an economically motivated framework that allows us to integrate the treatment effect literature with the literature on "structural" (economically motivated) econometric methods. We organize the alternative estimators of instrumental variables, matching, regression discontinuity design methods and the like within a common framework developed by Heckman and Vytlacil (1999, 2000, 2005).

Appendix A: The value of precisely formulated economic models in making policy forecasts

Explicitly formulated economic models are useful for three different purposes. First, the derivatives of such functions or finite changes generate the comparative statics *ceteris paribus* variations produced by economic theory. For example, tests of economic theory and measurements of economic parameters (price elasticities, measurements of consumer surplus, etc.) are based on structural equations.

Second, under invariance assumptions, structural equations can be used to forecast the effects of policies evaluated in one population in other populations, provided that the parameters are invariant across populations, and support conditions are satisfied. However, a purely nonparametric structural equation determined on one support cannot be extrapolated to other populations with different supports. Third, Marshallian causal functions and structural equations are one ingredient required to forecast the effect of a new policy, never previously implemented.

The problem of forecasting the effects of a policy evaluated on one population but applied to another population can be formulated in the following way. Let $Y(\omega) = \varphi(X(\omega), U(\omega))$, where $\varphi : \mathcal{D} \to \mathcal{Y}$, $\mathcal{D} \subseteq \mathbb{R}^J$, where $\mathcal{D}$ is the domain of the function, and $\mathcal{Y} \subseteq \mathbb{R}$. φ is a structural equation determining outcome Y, and we assume that it is known only over $\text{Supp}(X(\omega), U(\omega)) = \mathcal{X} \times \mathcal{U}$. $X(\omega)$ and $U(\omega)$ are random input variables. The mean outcome conditional on $X(\omega) = x$ is

$$E_H(Y \mid X = x) = \int_{\mathcal{U}} \varphi(X = x, u)\, \mathrm{d}F_H(u \mid X = x),$$

where $F_H(u \mid X)$ is the distribution of U in the historical data. We seek to forecast the outcome in a target population which may have a different support. The average outcome in the target population (T) is

$$E_T(Y \mid X = x) = \int_{\mathcal{U}^T} \varphi(X = x, u)\, \mathrm{d}F_T(u \mid X = x),$$

where $\mathcal{U}^T$ is the support of U in the target population. Provided that the support of (X, U) is the same in the source and the target populations, from knowledge of F_T it is possible to produce a correct value of $E_T(Y \mid X = x)$ for the target population. Otherwise, it is possible to evaluate this expectation only over the intersection set $\text{Supp}_T(X) \cap \text{Supp}_H(X)$, where $\text{Supp}_A(X)$ is the support of X in the source population. In order to extrapolate over the whole set $\text{Supp}_T(X)$, it is necessary to adopt some form

of parametric or functional structure. Additive separability in φ simplifies the extrapolation problem. If φ is additively separable

$$Y = \varphi(X) + U,$$

$\varphi(X)$ applies to all populations for which we can condition on X. However, some structure may have to be imposed to extrapolate from $\text{Supp}_H(X)$ to $\text{Supp}_T(X)$ if $\varphi(X)$ on T is not determined nonparametrically from H.

The problem of forecasting the effect of a new policy, never previously experienced, is similar in character to the policy forecasting problem just discussed. It shares many elements in common with the problem of forecasting the demand for a new good, never previously consumed.[100] Without imposing some structure on this problem, it is impossible to solve. The literature in structural econometrics associated with the work of the Cowles Commission adopts the following five step approach to this problem.

1. Structural functions are determined (e.g., $\varphi(X)$).
2. The new policy is characterized by an invertible mapping from observed random variables to the characteristics associated with the policy: $Q = q(X)$, where Q is the set of characteristics associated with the policy and q, $q : R^J \to R^J$, is a *known* invertible mapping.
3. $X = q^{-1}(Q)$ is solved to associate characteristics that in principle can be observed with the policy. This places the characteristics of the new policy on the same footing as those of the old.
4. It is assumed that, in the historical data, $\text{Supp}(q^{-1}(Q)) \subseteq \text{Supp}(X)$. This ensures that the support of the new characteristics mapped into X space is contained in the support of X. If this condition is not met, some functional structure must be used to forecast the effects of the new policy, to extend it beyond the support of the source population.
5. The forecast effect of the policy on Y is $Y(Q) = \varphi(q^{-1}(Q))$.

The leading example of this approach is Lancaster's method for estimating the demand for a new good [Lancaster (1971)]. New goods are viewed as bundles of old characteristics. McFadden's conditional logit scheme [1974] is based on a similar idea.[101]

Marschak's analysis of the effect of a new commodity tax is another example. Let $P(\omega)$ be the random variable denoting the price facing consumer ω. The tax changes

[100] Quandt and Baumol (1966), Lancaster (1971), Gorman (1980), McFadden (1974) and Domencich and McFadden (1975) consider the problem of forecasting the demand for a new good. Marschak (1953) is the classic reference for evaluating the effect of a new policy. See Heckman (2001) for a survey and synthesis of this literature.

[101] McFadden's stochastic specification is different from Lancaster's specification. See Heckman and Snyder (1997) for a comparison of these two approaches. Lancaster assumes that the $U(\omega)$ are the same for each consumer in all choice settings (they are preference parameters in his setting). McFadden allows for $U(\omega)$ to be different for the same consumer across different choice settings but assumes that the $U(\omega)$ in each choice setting are draws from a common distribution that can be determined from the demand for old goods.

the product price from $P(\omega)$ to $P(\omega)(1+t)$, where t is the tax. With sufficient price variation so that the assumption in Step 4 is satisfied (so the support of the price after tax, $\text{Supp}_{\text{post tax}}(P(\omega)(1+t)) \subseteq \text{Supp}_{\text{pretax}}(P(\omega))$, it is possible to use reduced form demand functions fit on a pretax sample to forecast the effect of a tax never previously put in place. Marschak uses a linear structural equation to solve the problem of limited support. From linearity, determination of the structural equations over a small region determines it everywhere.

Marshallian or structural causal functions are an essential ingredient in constructing such forecasts because they explicitly model the relationship between U and X. The treatment effect approach does not explicitly model this relationship so that treatment parameters cannot be extrapolated in this fashion, unless the dependence of potential outcomes on U and X is specified, and the required support conditions are satisfied. The Rubin (1978)–Holland (1986) model does not specify the required relationships. We discuss a specific way to implement this program in Chapter 71 of this contribution.

Appendix B: Nonparametric identification of counterfactual outcomes for a multinomial discrete choice model with state-contingent outcomes

Let outcomes in s be $Y(s) = \mu_Y(s, X) + U(s)$, $s = 1, \ldots, \bar{S}$, where there are $\bar{S}$ discrete states. Let $R(s) = \mu_R(s, Z) - V(s)$. The $U(s)$ and $V(s)$, $s = 1, \ldots, \bar{S}$, are assumed to be absolutely continuous and variation free as a collection of random variables. Thus the realization of one random variable does not restrict the realizations of the other random variables. State s is selected if

$$s = \text{argmax}_j \{R(j)\}_{j=1}^{\bar{S}}$$

and $Y(s)$ is observed. If s is observed, $D(s) = 1$. Otherwise $D(s) = 0$. $\sum_{s=1}^{\bar{S}} D(s) = 1$.

Define

$$\begin{aligned}
\mu_R^s(Z) &= \big(\mu_R(s, Z) - \mu_R(1, Z), \ldots, \mu_R(s, Z) - \mu_R(\bar{S}, Z)\big), \\
V^s &= \big(V(s) - V(1), \ldots, V(s) - V(\bar{S})\big), \\
\mu_R(Z) &= \big(\mu_R(1, Z), \ldots, \mu_R(\bar{S}, Z)\big), \\
\mu_Y(X) &= \big(\mu_Y(1, X), \ldots, \mu_Y(\bar{S}, X)\big), \\
F_V &= (F_{V(1)}, \ldots, F_{V(\bar{S})}), \\
D(s) &= \mathbf{1}(\mu_R^s(Z) \geqslant V^s).
\end{aligned}$$

Let $F_{U(s),V^s}$ be a candidate joint distribution of $(U(s), V^s)$, $s = 1, \ldots, \bar{S}$, with the true distribution being $F^*_{U(s),V^s}$. The true marginal distribution of V^s is $F^*_{V^s}$. The true marginal distribution of $U(s)$ is $F^*_{U(s)}$. Let $\mu_Y^*(X)$ denote the true value of $\mu_Y(X)$; $\mu_R^*(Z)$ is the true value of $\mu_R(Z)$. Define $\mathcal{M}_Y$ as the space of candidate conditional mean functions for Y: $\mu_Y \in \mathcal{M}_Y$. Define $\mathcal{M}_R$ as the space of candidate conditional mean functions for the discrete indices: $\mu_R \in \mathcal{M}_R$. Let $\mathcal{M} = \mathcal{M}_Y \times \mathcal{M}_R$.

In this notation, $(\mu_Y, \mu_R) \in \mathcal{M}$. Define $\mathcal{H}_V$ as the space of candidate distribution functions for V, $F_V \in \mathcal{H}_V$; $\mathcal{H}_{U,V}$ is the space of candidate distribution functions for $((U(1), V(1)), \ldots, (U(\bar{S}), V(\bar{S})))$, $F_{U,V} \in \mathcal{H}_{U,V}$.

Let $\mathcal{M}_Y^s, \mathcal{M}_R^s$ denote the spaces in which μ_Y^s, μ_R^s reside, $(\mu_Y^s, \mu_R^s) \in \mathcal{M}_Y^s \times \mathcal{M}_R^s$. Let $\mathcal{H}_{U,V}^s \subseteq \mathcal{H}_{U,V}$ denote the space in which candidate distributions $F_{U(s),V^s}$ reside, $F_{U(s),V^s} \in \mathcal{H}_{U,V}^s$. $\mathcal{H}_U^s$ and $\mathcal{H}_V^s$ are defined in a corresponding fashion.

Matzkin (1993) considers identification of polychotomous discrete choice models under the conditions of Theorem 1 below. We extend her analysis to allow for counterfactual outcomes adjoined to each choice. We can identify $\mu_Y(s, X)$, $s = 1, \ldots, \bar{S}$, over the support of X; $\mu_R(s, Z)$, up to scale over the support of Z and the joint distributions of $(U(s), V(s) - V(1), \ldots, V(s) - V(s-1), V(s) - V(s+1), \ldots, V(s) - V(\bar{S}))$ with the contrasts $V(s) - V(\ell)$, $\ell \neq s$, up to a scale that we present below in our discussion of Theorem 1.

THEOREM 1. *Assume*

(i) $\mu_R : \mathrm{Supp}(Z) \to \mathbb{R}^{\bar{S}}$ *is continuous for all* $\mu_R \in \mathcal{M}_R$.

(ii) $(U(s), V^s)$, $s = 1, \ldots, \bar{S}$, *are absolutely continuous random variables so that* $F_{U(s),V^s} \in \mathcal{H}_{U,V}^s$ *is continuous.* $E(U(s)) = 0$.

(iii) $\mathrm{Supp}(V^s) = \mathbb{R}^{\bar{S}-1}$, $s = 1, \ldots, \bar{S}$.

(iv) $(U(s), V^s) \perp\!\!\!\perp (X, Z)$, $s = 1, \ldots, \bar{S}$.

(v) *There exists a* $\widetilde{Z} \subseteq \mathrm{Supp}(Z)$ *such that for all* $\mu_R, \hat{\mu}_R \in \mathcal{M}_R$

  (a) $\mu^1(\widetilde{Z}) = \mathbb{R}^{\bar{S}-1}$.

  (b) $\mu_R^1(z) = \hat{\mu}_R^1(z)$ *for all* $z \in \widetilde{Z}$.

(vi) $\mathrm{Supp}(\mu_R^s(Z), X) = \mathrm{Supp}(\mu_R^s(Z)) \times \mathrm{Supp}(X)$.

(vii) *For all* $\mu_R, \hat{\mu}_R \in \mathcal{M}$ *and* $z \in \mathrm{Supp}(Z)$, $\mu_R(1, z) = \hat{\mu}_R(1, z)$.

Then $\mu_Y^*(s, X)$, $\mu_R^{*,s}(Z)$ *and* $F_{U(s),V^s}^*$, $s = 1, \ldots, \bar{S}$, *are identified.*[102]

PROOF. This theorem is a straightforward extension of Matzkin (1993, Theorem 2). The proof of identifiability of the $\mu_R^{*,s}(Z)$ and $F_{V^s}^*$, $s = 1, \ldots, \bar{S}$, follows directly from her analysis.

Thus, suppose that (F_{V^s}, μ_R^s) are observationally identical to $(F_{V^s}^*, \mu_R^{*,s})$ where both reside in the space $\mathcal{H}_V^s \times \mathcal{M}_R^s$. For all s,

$$F_{V^s}\big(\mu_R^s(z)\big) = F_{V^s}^*\big(\mu_R^{*,s}(z)\big)$$

for all $z \in \mathrm{Supp}(Z)$. For arbitrary $v \in \mathbb{R}^{\bar{S}-1}$, there exists $z_v \in \bar{Z}$ such that $\mu_R^1(z_v) = \mu_R^{1,*}(z_v) = v$ so that

$$F_{V^1}(v) = F_{V^1}\big(\mu_R^1(z_v)\big) = F_{V^1}^*\big(\mu_R^{*,1}(z_v)\big) = F_{V^1}^*(v)$$

[102] Assuming that $\mu_R(s, Z) = Z\gamma_s$, $s = 1, \ldots, \bar{S}$, simplifies the proof greatly and relies on more familiar conditions. See Heckman (1990), Heckman and Smith (1998) or Carneiro, Hansen and Heckman (2003). Matzkin (1993) presents alternative sets of conditions for identifiability of the choice model, all of which apply here.

for $v \in \mathbb{R}^{\bar{S}-1}$. Because V^s is a known linear transformation of V^1, this identifies $F^*_{V^s}$, $s = 1, \ldots, \bar{S}$. Given this distribution, following Matzkin, we can invert the choice probabilities to obtain $\mu_R^{*,s}(z)$, $s = 1, \ldots, \bar{S}$.

Armed with these results, we can find limit set $\mathcal{Z}(x)$, such that

$$\lim_{Z \to \mathcal{Z}(x)} \Pr\big(D(s) = 1 \mid Z = z, X = x\big) = 1$$

and thus $\lim_{Z \to \mathcal{Z}(x)} E(Y \mid D(s) = 1, Z = z, X = x) = \mu_y^*(s, x) + E(U(s))$. Using $E(U(s)) = 0$, we can identify the $\mu_Y^*(s, X)$ in those limit sets. We can vary $y(s)$ and trace out the marginal distribution of $U(s)$, $s = 1, \ldots, \bar{S}$, since $\lim_{Z \to \mathcal{Z}(x)} \Pr(Y(s) - \mu_y^*(s, x) \leqslant t \mid D(s) = 1, Z = z, X = x) = \Pr[U(s) \leqslant t]$. From the joint distribution of $Y(s)$, $D(s)$ given X, Z, we can identify $F^*_{U(s),V^s}$, $s = 1, \ldots, \bar{S}$, by tracing out different values of $y(s)$, given $X = x$, and $\mu_R^{*,s}(z)$. □

From this model, we can identify the marginal treatment effect [Carneiro, Hansen and Heckman (2003, p. 368, equation (71))] and all pairwise average treatment effects by forming suitable limit sets. We can also identify all pairwise mean treatment on the treated and treatment on the untreated effects.

In the general case, we can identify the densities of $U(s)$, $V(s) - V(1), \ldots, V(s) - V(\bar{S})$, $s = 1, \ldots, \bar{S}$, where $U(s)$ may be a vector and the contrasts are identified. Set $V(\bar{S}) \equiv 0$ (this is only one possible normalization). Then from the choice equation for $\bar{S}$ ($\Pr(D(\bar{S}) = 1 \mid Z = z)$) we can identify the pairwise correlations $\rho_{i,j} = \text{Correl}(V(i), V(j))$, $i, j = 1, \ldots, \bar{S} - 1$. We assume $-1 \leqslant \rho_{i,j} < 1$. If $\rho_{i,j} = 1$ for some i, j, the choice of a normalization is not innocuous. Under our conditions we can identify $\text{Var}(V(s) - V(\ell)) = 2(1 - \rho_{s,\ell})$. This is the scale for contrast s, ℓ. Define $\tau_{s,\ell} = [\text{Var}(V(s) - V(\ell))]^{1/2}$ where positive square roots are used.

Consider constructing the distribution of $Y(\ell)$ given $D(s) = 1, X, Z$. If $\ell \neq s$, this is a counterfactual distribution. From this distribution we can construct, among many possible counterfactual parameters, $E(Y(s) - Y(\ell) \mid D(s) = 1,\ X = x,\ Z = z)$, a treatment on the treated parameter.

To form the distribution of $(U(\ell), \frac{V(s)-V(1)}{\tau_{s,1}}, \ldots, \frac{V(s)-V(\bar{S})}{\tau_{s,\bar{S}}})$ for any $\ell \neq s$ from the objects produced from Theorem 1, we use the normalized versions of $V(s) - V(1)$, $\ldots, V(s) - V(\bar{S})$: $\frac{V(s)-V(1)}{\tau_{s,1}}, \ldots, \frac{V(s)-V(\bar{S})}{\tau_{s,\bar{S}}}$. From the density of $U(\ell)$, $\frac{V(\ell)-V(1)}{\tau_{\ell,1}}$, $\ldots, \frac{V(\ell)-V(\bar{S})}{\tau_{\ell,\bar{S}}}$ which we identify from Theorem 1, we can transform the contrast variables in the following way.

Define $q(\ell, s) = \frac{V(\ell)-V(s)}{\tau_{\ell,s}}$. From the definitions, $q(s, j) = \frac{V(s)-V(j)}{\tau_{s,j}} = \frac{q(\ell,j)\tau_{\ell,j}-q(\ell,s)\tau_{\ell,s}}{\tau_{s,j}}$, for all $j = 1, 2, \ldots, \bar{S}$. Substitute $\frac{q(\ell,j)\tau_{\ell,j}-q(\ell,s)\tau_{\ell,s}}{\tau_{s,j}}$, for $\frac{V(s)-V(j)}{\tau_{s,j}}$, $j = 1, 2, \ldots, \bar{S}$, $j \neq \ell$, in the density of $(U(\ell), \frac{V(\ell)-V(s)}{\tau_{\ell,s}}, \ldots, \frac{V(\ell)-V(\bar{S})}{\tau_{\ell,\bar{S}}})$ and use the Jacobian of transformation $\prod_{j=1,\ldots,\bar{S},\ j\neq\ell} |\tau_{\ell,j}|$ to obtain the desired density where "| |" denotes determinant. This produces the desired counterfactual density for all

$s = 1, \ldots, \bar{S}$. Provided that the Jacobians are nonzero (which rules out perfect dependence), we preserve all of the information and can construct the marginal distribution of any $U(\ell)$ for any desired pattern of latent indices. Thus we can construct the desired counterfactuals.

Appendix C: Normal selection model results

The properties of the normal selection model are generated by the properties of a truncated normal model which we now establish. See Heckman and Honoré (1990). Let Z be a standard normal random variable and let $\lambda(d) \overset{\text{def}}{\equiv} E[Z \mid Z \geqslant d]$. For all $d \in (-\infty, \infty)$, we prove the following results:

$$\text{(N-1)} \quad \lambda(d) = \frac{\frac{1}{\sqrt{2\pi}} \exp\{-\frac{d^2}{2}\}}{\Phi(-d)} > \max\{0, d\},$$

$$\text{(N-2)} \quad 0 < \frac{\partial \lambda(d)}{\partial d} = \lambda'(d) = \lambda(d)\big(\lambda(d) - d\big) < 1,$$

$$\text{(N-3)} \quad \frac{\partial^2 \lambda(d)}{\partial d^2} > 0,$$

$$\text{(N-4)} \quad 0 < \operatorname{Var}[Z \mid Z \geqslant d] = 1 + \lambda(d)d - [\lambda(d)]^2 < 1,$$

$$\text{(N-5)} \quad \frac{\partial \operatorname{Var}[Z \mid Z \geqslant d]}{\partial d} < 0,$$

$$\text{(N-6)} \quad E\big[\big(Z - \lambda(d)\big)^3 \mid Z \geqslant d\big] = \lambda(d)\big(2[\lambda(d)]^2 - 3d\lambda(d) + d^2 - 1\big) = \frac{\partial^2 \lambda(d)}{\partial d^2},$$

$$\text{(N-7)} \quad E[Z \mid Z \geqslant d] \geqslant \operatorname{mode}[Z \mid Z \geqslant d],$$

$$\text{(N-8)} \quad \lim_{d \to -\infty} \lambda(d) = 0, \quad \lim_{d \to \infty} \lambda(d) = \infty,$$

$$\text{(N-9)} \quad \lim_{d \to -\infty} \frac{\partial \lambda(d)}{\partial d} = 0, \quad \lim_{d \to \infty} \frac{\partial \lambda(d)}{\partial d} = 1,$$

$$\text{(N-10)} \quad \lim_{d \to -\infty} \operatorname{Var}[Z \mid Z \geqslant d] = 1, \quad \lim_{d \to \infty} \operatorname{Var}[Z \mid Z \geqslant d] = 0.$$

Results (N-2), (N-4) and (N-5) are implications of log concavity. (N-7) is an implication of symmetry and log concavity. (N-1) and (N-3) are consequences of normality. The left-hand side limits of (N-8) and (N-10) are true for any distribution with zero mean and unit variance. So is the right-hand limit of (N-8) provided that the support of Z is not bounded on the right. The right-hand limits of (N-9) and (N-10) are consequences of normality.

C.1. Proofs of results (N-1) to (N-10)

The moment generating function for a truncated normal distribution with truncation point d is:

$$\operatorname{mgf}(\beta) = e^{\beta/2}\frac{\int_{d-\beta}^{\infty}\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)\,\mathrm{d}u}{\int_{d}^{\infty}\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)\,\mathrm{d}u}.$$

The equality in (N-1) follows from:

$$\lambda(d) = E[Z \mid Z \geqslant d] = \left.\frac{\partial \operatorname{mgf}}{\partial\beta}\right|_{\beta=0}.$$

The inequality is obvious.

By direct calculation, $\lambda'(d) = \lambda(d)(\lambda(d) - d)$. Now note that

$$E\big[Z^2 \mid Z \geqslant d\big] = \left.\frac{\partial^2 \operatorname{mgf}}{\partial\beta^2}\right|_{\beta=0} = 1 + \lambda(d)d.$$

Therefore:

$$\operatorname{Var}[Z \mid Z \geqslant d] = 1 - \frac{\partial\lambda(d)}{\partial d}.$$

As $\operatorname{Var}[Z \mid Z \geqslant d] > 0$ and $\lambda(d)(\lambda(d) - d) > 0$ by (N-1), this proves (N-2) and (N-4). To prove (N-3) notice that $\operatorname{Var}[Z \mid Z \geqslant d] = 1 - \frac{\partial\lambda(d)}{\partial d}$, and therefore:

$$\frac{\partial^2\lambda(d)}{\partial d^2} = -\frac{\partial \operatorname{Var}[Z \mid Z \geqslant d]}{\partial d} > 0,$$

where the inequality follows from Proposition 1 in Heckman and Honoré (1990). (N-5) also follows from Proposition 1, whereas (N-6) follows by direct calculation from the expression for $E[(Z - \lambda(d))^3 \mid Z > d]$. (N-7) is trivial. (N-8) is obvious. The first part of (N-9) follows directly from L'Hôpital's rule. (N-2) and (N-3) imply that $\frac{\partial\lambda(d)}{\partial d}$ is increasing and bounded by 1. Therefore $\lim_{d\to\infty}\frac{\partial\lambda(d)}{\partial d}$ exists and does not exceed 1. If $\lim_{d\to\infty}\frac{\partial\lambda(d)}{\partial d} < 1$ then $\lambda(d)$ would eventually be less than d, contradicting (N-1). This proves the second part of (N-9). (N-9) and (N-4) imply (N-10).

References

Abadie, A., Angrist, J.D., Imbens, G. (2002). "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings". Econometrica 70 (1), 91–117 (January).

Abbring, J.H., Campbell, J.R. (2005). "A firm's first year". Technical Report TI 05-046/3, Tinbergen Institute Discussion Paper, May.

Abbring, J.H., Heckman, J.J. (2007). "Econometric evaluation of social programs, Part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6B. Elsevier. Chapter 72.

Abbring, J.H., Van den Berg, G.J. (2003). "The nonparametric identification of treatment effects in duration models". Econometrica 71 (5), 1491–1517 (September).
Ackerberg, D., Benkard, C.L., Berry, S., Pakes, A. (2007). "Econometric tools for analyzing market outcomes". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6A. Elsevier. Chapter 63.
Aguirregabiria, V. (2004). "Pseudo maximum likelihood estimation of structural models involving fixed-point problems". Economics Letters 84 (3), 335–340 (September).
Ahn, H., Powell, J. (1993). "Semiparametric estimation of censored selection models with a nonparametric selection mechanism". Journal of Econometrics 58 (1–2), 3–29 (July).
Amemiya, T. (1985). Advanced Econometrics. Harvard University Press, Cambridge, MA.
Angrist, J.D., Imbens, G.W. (1995). "Two-stage least squares estimation of average causal effects in models with variable treatment intensity". Journal of the American Statistical Association 90 (430), 431–442 (June).
Angrist, J.D., Krueger, A.B. (1999). "Empirical strategies in labor economics". In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, vol. 3A. Elsevier, New York, pp. 1277–1366.
Angrist, J.D., Imbens, G.W., Rubin, D. (1996). "Identification of causal effects using instrumental variables". Journal of the American Statistical Association 91 (434), 444–455.
Arabmazar, A., Schmidt, P. (1982). "An investigation of the robustness of the Tobit estimator to non-normality". Econometrica 50 (4), 1055–1063 (July).
Athey, S., Haile, P. (2007). "Nonparametric approaches to auctions". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6A. Elsevier, Amsterdam. Chapter 60.
Berk, R., Li, A., Hickman, L.J. (2005). "Statistical difficulties in determining the role of race in capital cases: A re-analysis of data from the state of Maryland". Journal of Quantitative Criminology 21 (4), 365–390 (December).
Björklund, A., Moffitt, R. (1987). "The estimation of wage gains and welfare gains in self-selection". Review of Economics and Statistics 69 (1), 42–49 (February).
Blundell, R., Stoker, T. (2007). "Models of aggregate economic relationships that account for heterogeneity". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6A. Elsevier, Amsterdam. Chapter 68.
Blundell, R., Reed, H., Stoker, T. (2003). "Interpreting aggregate wage growth: The role of labor market participation". American Economic Review 93 (4), 1114–1131 (September).
Blundell, R., Costa Dias, M., Meghir, C., Van Reenen, J. (2004). "Evaluating the employment effects of a mandatory job search program". Journal of the European Economic Association 2 (4), 569–606 (June).
Blundell, R., MaCurdy, T., Meghir, C. (2007). "Labor supply models: Unobserved heterogeneity, nonparticipation and dynamics". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6A. Elsevier, Amsterdam. Chapter 69.
Boadway, R.W., Bruce, N. (1984). Welfare Economics. B. Blackwell, New York.
Bock, R.D., Jones, L.V. (1968). The Measurement and Prediction of Judgment and Choice. Holden-Day, San Francisco.
Bond, S., Van Reenen, J. (2007). "Microeconometric models of investment and employment". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6A. Elsevier, Amsterdam. Chapter 65.
Brock, W.A., Durlauf, S.N. (2001). "Interactions-based models". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 5. North-Holland, New York, pp. 3463–3568.
Browning, M., Hansen, L.P., Heckman, J.J. (1999). "Micro data and general equilibrium models". In: Taylor, J.B., Woodford, M. (Eds.), Handbook of Macroeconomics, vol. 1A. Elsevier, pp. 543–633. Chapter 8.
Campbell, D.T., Stanley, J.C. (1963). Experimental and Quasi-Experimental Designs for Research. Rand McNally, Chicago. Originally appeared in Gage, N.L. (Ed.), Handbook of Research on Teaching.
Carneiro, P., Hansen, K., Heckman, J.J. (2001). "Removing the veil of ignorance in assessing the distributional impacts of social policies". Swedish Economic Policy Review 8 (2), 273–301 (Fall).
Carneiro, P., Hansen, K., Heckman, J.J. (2003). "Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice". International Economic Review 44 (2), 361–422 (May).

Carrasco, M., Florens, J.–P., Renault, E. (2007). "Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6B. Elsevier, Amsterdam. Chapter 77.

Chan, T.Y., Hamilton, B.H. (2003, July). "Learning, private information and the economic evaluation of randomized experiments". Working paper. Olin School of Business, Washington University, St. Louis.

Chen, S. (1999). "Distribution-free estimation of the random coefficient dummy endogenous variable model". Journal of Econometrics 91 (1), 171–199 (July).

Chen, X. (2007). "Large sample sieve estimation of semi-nonparametric models". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6B. Elsevier, Amsterdam. Chapter 76.

Chernozhukov, V., Hansen, C. (2005). "An IV model of quantile treatment effects". Econometrica 73 (1), 245–261 (January).

Cosslett, S.R. (1983). "Distribution-free maximum likelihood estimator of the binary choice model". Econometrica 51 (3), 765–782 (May).

Cox, D.R. (1958). Planning of Experiments. Wiley, New York.

Cunha, F., Heckman, J.J., Navarro, S. (2005). "Separating uncertainty from heterogeneity in life cycle earnings, the 2004 Hicks lecture". Oxford Economic Papers 57 (2), 191–261 (April).

Cunha, F., Heckman, J.J., Navarro, S. (2006). "Counterfactual analysis of inequality and social mobility". In: Morgan, S.L., Grusky, D.B., Fields, G.S. (Eds.), Mobility and Inequality: Frontiers of Research in Sociology and Economics. Stanford University Press, Stanford, CA, pp. 290–348. Chapter 4.

Dahl, G.B. (2002). "Mobility and the return to education: Testing a Roy model with multiple markets". Econometrica 70 (6), 2367–2420 (November).

Dawid, A. (2000). "Causal inference without counterfactuals". Journal of the American Statistical Association 95 (450), 407–424 (June).

Domencich, T., McFadden, D.L. (1975). Urban Travel Demand: A Behavioral Analysis. North-Holland, Amsterdam. Reprinted 1996.

Eckstein, Z., Wolpin, K.I. (1989). "The specification and estimation of dynamic stochastic discrete choice models: A survey". Journal of Human Resources 24 (4), 562–598 (Fall).

Eckstein, Z., Wolpin, K.I. (1999). "Why youths drop out of high school: The impact of preferences, opportunities and abilities". Econometrica 67 (6), 1295–1339 (November).

Fisher, R.A. (1966). The Design of Experiments. Hafner Publishing, New York.

Florens, J.-P., Heckman, J.J. (2003). "Causality and econometrics". Unpublished working paper, University of Chicago, Department of Economics.

Foster, J.E., Sen, A.K. (1997). On Economic Inequality. Oxford University Press, New York.

Geweke, J., Keane, M. (2001). "Computationally intensive methods for integration in econometrics". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 5. North-Holland, New York, pp. 3463–3568.

Gill, R.D., Robins, J.M. (2001). "Causal inference for complex longitudinal data: The continuous case". The Annals of Statistics 29 (6), 1785–1811 (December).

Goldberger, A.S. (1964). Econometric Theory. Wiley, New York.

Goldberger, A.S. (1983). "Abnormal selection bias". In: Karlin, S., Amemiya, T., Goodman, L.A. (Eds.), Studies in Econometrics, Time Series, and Multivariate Statistics. Academic Press, New York, pp. 67–84.

Gorman, W.M. (1980). "A possible procedure for analysing quality differentials in the egg market". Review of Economic Studies 47 (5), 843–856 (October).

Gronau, R. (1974). "Wage comparisons – a selectivity bias". Journal of Political Economy 82 (6), 1119–1143 (November–December).

Haavelmo, T. (1943). "The statistical implications of a system of simultaneous equations". Econometrica 11 (1), 1–12 (January).

Haavelmo, T. (1944). "The probability approach in econometrics". Econometrica 12 (Suppl.), iii–vi and 1–115.

Hamermesh, D.S. (1993). Labor Demand. Princeton University Press, Princeton, NJ.

Hansen, L.P., Sargent, T.J. (1980). "Formulating and estimating dynamic linear rational expectations models". Journal of Economic Dynamics and Control 2 (1), 7–46 (February).

Harberger, A.C. (1971). "Three basic postulates for applied welfare economics: An interpretive essay". Journal of Economic Literature 9 (3), 785–797 (September).
Harsanyi, J.C. (1955). "Cardinal welfare, individualistic ethics and interpersonal comparisons of utility". Journal of Political Economy 63 (4), 309–321 (August).
Harsanyi, J.C. (1975). "Can the maximin principle serve as a basis for morality? A critique of John Rawls's theory". American Political Science Review 69 (2), 594–606 (June).
Heckman, J.J. (1974). "Shadow prices, market wages and labor supply". Econometrica 42 (4), 679–694 (July).
Heckman, J.J. (1976). "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models". Annals of Economic and Social Measurement 5 (4), 475–492 (December).
Heckman, J.J. (1978). "Dummy endogenous variables in a simultaneous equation system". Econometrica 46 (4), 931–959 (July).
Heckman, J.J. (1979). "Sample selection bias as a specification error". Econometrica 47 (1), 153–162 (January).
Heckman, J.J. (1980). "Addendum to sample selection bias as a specification error". In: Stromsdorfer, E., Farkas, G. (Eds.), In: Evaluation Studies Review Annual, vol. 5. Sage Publications, Beverly Hills.
Heckman, J.J. (1990). "Varieties of selection bias". American Economic Review 80 (2), 313–318 (May).
Heckman, J.J. (1992). "Randomization and social policy evaluation". In: Manski, C., Garfinkel, I. (Eds.), Evaluating Welfare and Training Programs. Harvard University Press, Cambridge, MA, pp. 201–230.
Heckman, J.J. (2000). "Policies to foster human capital". Research in Economics 54 (1), 3–56 (March). With discussion.
Heckman, J.J. (2001). "Micro, data, heterogeneity and the evaluation of public policy: Nobel lecture". Journal of Political Economy 109 (4), 673–748 (August).
Heckman, J.J. (2005). "The scientific model of causality". Sociological Methodology 35 (1), 1–97 (August).
Heckman, J.J. (2007). Evaluating Economic Policy. Unpublished manuscript. University of Chicago.
Heckman, J.J., Honoré, B.E. (1990). "The empirical content of the Roy model". Econometrica 58 (5), 1121–1149 (September).
Heckman, J.J., MaCurdy, T.E. (1986). "Labor econometrics". In: Griliches, Z., Intriligator, M. (Eds.), Handbook of Econometrics, vol. 3. North-Holland, New York, pp. 1917–1977.
Heckman, J.J., Navarro, S. (2007). "Dynamic discrete choice and dynamic treatment effects". Journal of Econometrics 136 (2), 341–396 (February).
Heckman, J.J., Robb, R. (1985). "Alternative methods for evaluating the impact of interventions". In: Heckman, J.J., Singer, B. (Eds.), Longitudinal Analysis of Labor Market Data, vol. 10. Cambridge University Press, New York, pp. 156–245.
Heckman, J.J., Robb, R. (1986). "Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes". In: Wainer, H. (Ed.), Drawing Inferences from Self-Selected Samples. Springer-Verlag, New York, pp. 63–107. Reprinted in 2000, Lawrence Erlbaum Associates, Mahwah, NJ.
Heckman, J.J., Sedlacek, G.L. (1985). "Heterogeneity, aggregation and market wage functions: An empirical model of self-selection in the labor market". Journal of Political Economy 93 (6), 1077–1125 (December).
Heckman, J.J., Smith, J.A. (1998). "Evaluating the welfare state". In: Strom, S. (Ed.), Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium. Cambridge University Press, New York, pp. 241–318.
Heckman, J.J., Snyder, J.M. (1997). "Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators". RAND Journal of Economics 28, S142 (Special Issue).
Heckman, J.J., Vytlacil, E.J. (1999). "Local instrumental variables and latent variable models for identifying and bounding treatment effects". Proceedings of the National Academy of Sciences 96, 4730–4734 (April).
Heckman, J.J., Vytlacil, E.J. (2000). "The relationship between treatment parameters within a latent variable framework". Economics Letters 66 (1), 33–39 (January).

Heckman, J.J., Vytlacil, E.J. (2001). "Local instrumental variables". In: Hsiao, C., Morimune, K., Powell, J.L. (Eds.), Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya. Cambridge University Press, New York, pp. 1–46.

Heckman, J.J., Vytlacil, E.J. (2005). "Structural equations, treatment effects and econometric policy evaluation". Econometrica 73 (3), 669–738 (May).

Heckman, J.J., Smith, J.A., Clements, N. (1997). "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts". Review of Economic Studies 64 (221), 487–536 (October).

Heckman, J.J., Ichimura, H., Todd, P.E. (1998). "Matching as an econometric evaluation estimator". Review of Economic Studies 65 (223), 261–294 (April).

Heckman, J.J., Lochner, L.J., Taber, C. (1998). "General-equilibrium treatment effects: A study of tuition policy". American Economic Review 88 (2), 381–386 (May).

Heckman, J.J., LaLonde, R.J., Smith, J.A. (1999). "The economics and econometrics of active labor market programs". In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, vol. 3A. North-Holland, New York, pp. 1865–2097. Chapter 31.

Heckman, J.J., Hohmann, N., Smith, J., Khoo, M. (2000). "Substitution and dropout bias in social experiments: A study of an influential social experiment". Quarterly Journal of Economics 115 (2), 651–694 (May).

Heckman, J.J., Tobias, J.L., Vytlacil, E.J. (2001). "Four parameters of interest in the evaluation of social programs". Southern Economic Journal 68 (2), 210–223 (October).

Heckman, J.J., Tobias, J.L., Vytlacil, E.J. (2003). "Simple estimators for treatment parameters in a latent variable framework". Review of Economics and Statistics 85 (3), 748–754 (August).

Heckman, J.J., Urzua, S., Vytlacil, E.J. (2006). "Understanding instrumental variables in models with essential heterogeneity". Review of Economics Statistics 88 (3), 389–432.

Hensher, D., Louviere, J., Swait, J. (1999). "Combining sources of preference data". Journal of Econometrics 89 (1–2), 197–221 (March–April).

Hicks, J.R. (1946). Value and Capital: An Inquiry into Some Fundamental Principles of Economic Theory, second ed. Clarendon Press, Oxford.

Holland, P.W. (1986). "Statistics and causal inference". Journal of the American Statistical Association 81 (396), 945–960 (December).

Holland, P.W. (1988). "Causal inference, path analysis and recursive structural equation models". In: Clogg, C., Arminger, G. (Eds.), Sociological Methodology. American Sociological Association, Washington, DC, pp. 449–484.

Hurwicz, L. (1962). "On the structural form of interdependent systems". In: Nagel, E., Suppes, P., Tarski, A. (Eds.), Logic, Methodology and Philosophy of Science. Stanford University Press, pp. 232–239.

Ichimura, H., Todd, P.E. (2007). "Implementing nonparametric and semiparametric estimators". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6B. Elsevier, Amsterdam. Chapter 74.

Imbens, G.W., Angrist, J.D. (1994). "Identification and estimation of local average treatment effects". Econometrica 62 (2), 467–475 (March).

Journal of Human Resources (1990, Summer). Special Issue on Taxation and Labor Supply in Industrial Countries, Volume 25.

Katz, D., Gutek, A., Kahn, R., Barton, E. (1975). Bureaucratic Encounters: A Pilot Study in the Evaluation of Government Services. Survey Research Center Institute for Social Research, University of Michigan, Ann Arbor.

Keane, M.P., Wolpin, K.I. (1997). "The career decisions of young men". Journal of Political Economy 105 (3), 473–522 (June).

Killingsworth, M.R. (1983). Labor Supply. Cambridge University Press, Cambridge.

Killingsworth, M.R. (1985). "Substitution and output effects on labor demand: Theory and policy applications". Journal of Human Resources 20 (1), 142–152 (Winter).

Knight, F. (1921). Risk, Uncertainty and Profit. Houghton Mifflin Company, New York.

Koopmans, T.C., Rubin, H., Leipnik, R.B. (1950). "Measuring the equation systems of dynamic economics". In: Koopmans, T.C. (Ed.), Statistical Inference in Dynamic Economic Models. In: Cowles Commission Monograph, vol. 10. John Wiley & Sons, New York, pp. 53–237. Chapter 2.

Lancaster, K.J. (1971). Consumer Demand: A New Approach. Columbia University Press, New York.

Leamer, E.E. (1985). "Vector autoregressions for causal inference?". Carnegie–Rochester Conference Series on Public Policy 22, 255–303 (Spring).

Lechner, M. (2004). "Sequential matching estimation of dynamic causal models". Technical Report 2004, IZA Discussion Paper.

Lee, L.-F. (1978). "Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables". International Economic Review 19 (2), 415–433 (June).

Lewis, H.G. (1974). "Comments on selectivity biases in wage comparisons". Journal of Political Economy 82 (6), 1145–1155 (November–December).

Lewis, H.G. (1986). Union Relative Wage Effects: A Survey. University of Chicago Press, Chicago.

Little, I.M., Mirrlees, J.A. (1974). Project Appraisal and Planning for Developing Countries. Basic Books, New York.

Lucas, R.E., Sargent, T.J. (1981). Rational Expectations and Econometric Practice. University of Minnesota Press, Minneapolis.

Maddala, G. (1983). Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, New York.

Magnac, T., Thesmar, D. (2002). "Identifying dynamic discrete decision processes". Econometrica 70 (2), 801–816 (March).

Manski, C.F. (1975). "Maximum score estimation of the stochastic utility model of choice". Journal of Econometrics 3 (3), 205–228 (August).

Manski, C.F. (1988). "Identification of binary response models". Journal of the American Statistical Association 83 (403), 729–738 (September).

Manski, C.F. (2003). Partial Identification of Probability Distributions. Springer-Verlag, New York.

Marschak, J. (1953). "Economic measurements for policy and prediction". In: Hood, W., Koopmans, T. (Eds.), Studies in Econometric Method. Wiley, New York, pp. 1–26.

Marshall, D.A. (1890). Principles of Economics. Macmillan and Company, New York.

Matzkin, R.L. (1992). "Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models". Econometrica 60 (2), 239–270 (March).

Matzkin, R.L. (1993). "Nonparametric identification and estimation of polychotomous choice models". Journal of Econometrics 58 (1–2), 137–168 (July).

Matzkin, R.L. (1994). "Restrictions of economic theory in nonparametric methods". In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, New York, pp. 2523–2558.

Matzkin, R.L. (2003). "Nonparametric estimation of nonadditive random functions". Econometrica 71 (5), 1339–1375 (September).

Matzkin, R.L. (2004). "Unobserved instruments". Unpublished manuscript. Northwestern University, Department of Economics.

Matzkin, R.L. (2007). "Nonparametric identification". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6B. Elsevier Science. Chapter 73.

McFadden, D. (1974). "Conditional logit analysis of qualitative choice behavior". In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York.

McFadden, D. (1981). "Econometric models of probabilistic choice". In: Manski, C., McFadden, D. (Eds.), Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA.

McFadden, D. (1984). "Econometric analysis of qualitative response models". In: Griliches, Z., Intriligator, M. (Eds.), Handbook of Econometrics, vol. 2. North-Holland, New York, pp. 1396–1457.

McFadden, D. (1985). "Technical problems in social experimentation: Cost versus ease of analysis: Comment". In: Hausman, J.A., Wise, D.A. (Eds.), Social Experimentation, National Bureau of Economic Research Conference Report. University of Chicago Press, Chicago, pp. 214–218.

McFadden, D. (2001). "On selecting regression variables to maximize their significance". In: Hsiao, C., Morimune, K., Powell, J. (Eds.), Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya. Cambridge University Press, New York, pp. 259–280.

Miller, R.A. (1984). "Job matching and occupational choice". Journal of Political Economy 92 (6), 1086–1120 (December).

Moulin, H. (1983). The Strategy of Social Choice. North-Holland, New York.

Neyman, J. (1923). "Statistical problems in agricultural experiments". Journal of the Royal Statistical Society II (Suppl. (2)), 107–180.

Osborne, M.J. (2004). An Introduction to Game Theory. Oxford University Press, New York.

Pakes, A. (1986). "Patents as options: Some estimates of the value of holding European patent stocks". Econometrica 54 (4), 755–784 (July).

Pearl, J. (2000). Causality. Cambridge University Press, Cambridge, England.

Persson, T., Tabellini, G.E. (2000). Political Economics: Explaining Economic Policy. MIT Press, Cambridge, MA.

Pitt, M., Rosenzweig, M. (1989). "The selectivity of fertility and the determinants of human capital investments: Parametric and semi-parametric estimates". Living Standards Measurement 119, World Bank.

Powell, J.L. (1994). "Estimation of semiparametric models". In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, vol. 4. Elsevier, Amsterdam, pp. 2443–2521.

Quandt, R.E. (1958). "The estimation of the parameters of a linear regression system obeying two separate regimes". Journal of the American Statistical Association 53 (284), 873–880 (December).

Quandt, R.E. (1972). "A new approach to estimating switching regressions". Journal of the American Statistical Association 67 (338), 306–310 (June).

Quandt, R.E. (1974). "A comparison of methods for testing nonnested hypotheses". Review of Economics and Statistics 56 (1), 92–99 (February).

Quandt, R.E., Baumol, W. (1966). "The demand for abstract transport modes: Theory measurement". Journal of Regional Science 6, 13–26.

Quine, W.V.O. (1951). "Main trends in recent philosophy: Two dogmas of empiricism". The Philosophical Review 60 (1), 20–43 (January).

Rawls, J. (1971). "A Theory of Justice". Belknap Press of Harvard University Press, Cambridge, MA.

Reiss, P., Wolak, F. (2007). "Structural econometric modeling: Rationales and examples from industrial organization". In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 6A. Elsevier Science. Chapter 64.

Roy, A. (1951). " Some thoughts on the distribution of earnings". Oxford Economic Papers 3 (2), 135–146 (June).

Rubin, D.B. (1976). "Inference and missing data". Biometrika 63 (3), 581–592 (December).

Rubin, D.B. (1978). "Bayesian inference for causal effects: The role of randomization". Annals of Statistics 6 (1), 34–58 (January).

Rubin, D.B. (1986). "Statistics and causal inference: Comment: Which ifs have causal answers". Journal of the American Statistical Association 81 (396), 961–962.

Rust, J. (1994). "Structural estimation of Markov decision processes". In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics. North-Holland, New York, pp. 3081–3143.

Ruud, P.A. (1981). "Misspecification in Limited Dependent Variable Models". PhD thesis. Massachusetts Institute of Technology.

Ruud, P.A. (2000). An Introduction to Classical Econometric Theory. Oxford University Press, New York.

Sen, A.K. (1999). "The possibility of social choice". American Economic Review 89 (3), 349–378 (June).

Sims, C.A. (1977). "Exogeneity and causal orderings in macroeconomic models". In: New Methods in Business Cycle Research. Federal Reserve Bank of Minneapolis, Minneapolis, MN, pp. 23–43.

Taber, C.R. (2001). "The rising college premium in the eighties: Return to college or return to unobserved ability?". Review of Economic Studies 68 (3), 665–691 (July).

Tamer, E. (2003). "Incomplete simultaneous discrete response model with multiple equilibria". Review of Economic Studies 70 (1), 147–165 (January).

Thurstone, L.L. (1927). "A law of comparative judgement". Psychological Review 34, 273–286.
Thurstone, L.L. (1959). The Measurement of Values. University of Chicago Press, Chicago.
Tinbergen, J. (1930). "Bestimmung und Deutung von Angebotskurven". Zeitschrift für Nationalökonomie 1 (1), 669–679 (March).
Todd, P.E. (1996). "Essays on empirical methods for evaluating the impact of policy interventions in education and training". PhD dissertation. University of Chicago, Chicago.
Varian, H.R. (1978). Microeconomic Analysis. Norton, New York.
Vickrey, W. (1945). "Measuring marginal utility by reactions to risk". Econometrica 13 (4), 319–333 (October).
Vickrey, W. (1961). "Utility, strategy and social decision rules: Reply". Quarterly Journal of Economics 75 (3), 496–497 (August).
Vijverberg, W.P.M. (1993). "Measuring the unidentified parameter of the extended Roy model of selectivity". Journal of Econometrics 57 (1–3), 69–89 (May–June).
Willis, R.J., Rosen, S. (1979). "Education and self-selection". Journal of Political Economy 87 (5, Part 2), S7–S36 (October).
Wold, H.O.A. (1956). "Causal inference from observational data: A review of end and means". Journal of the Royal Statistical Society. Series A (General) 119 (1), 28–61.
Wooldridge, J.M. (2002). Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.