**Treatment Evaluation**

1. **Introduction**

* The topic of treatment evaluation concerns measuring the impact of interventions on outcomes of interest, with the type of intervention and outcome.
* The treatment evaluation approach comes from medical sciences where intervention frequently means adopting a treatment regime.
* In economic applications treatment and interventions usually mean the same thing.
* Examples of treatments in the economic context are enrollment into a labor training program, being a member of a trade union, receipt of a transfer payment from a social program.
* If the treatment that is applied can vary in intensity or type, we use the term **multiple treatments**.
* A leading case is one in which the outcome of interest is a continuous variable, say , whereas the treatment variable is discrete and of on/off variety, say *,* where is dummy.
* An example of an intervention is labor market training.
* We will take the case of a continuous outcome and a binary-valued treatment as our leading case.
* Policy relevance of treatment evaluation is direct because “successful” treatments can be linked to desirable social programs.
* In a canonical single-treatment example we observe .
* The impact of a hypothetical change in *D* on *,* holding constant, is of interest. (**potential outcome model)**
* However, no individual is simultaneously observed in both states.
* It can be tackled by methods of causal inference carried out in terms of **counterfactuals**.
* *Ceteris paribus assumption* is held.
* We focus on a family of measures of treatment effectiveness.
* These measures are functions of parameters and data, and they enable comparisons with policy-relevant counterfactuals.
* Another emphasis in the literature on treatment evaluation is on the advantages of identiﬁcation secured using minimal functional form and exclusion restrictions.
* The feasibility of semiparametric identiﬁcation is relatively easier to establish for treatment effect estimation in linear models.

1. Setup and Assumptions

* We detail the assumptions that permit use of the key matching and propensity score estimators that are presented later.

1. Treatment Effects Framework

* Let there be a target population for the treatment of interest.
* Let denote the number of randomly selected individuals who are eligible for treatment.
* Let denote the number of being treated.

Let denote the number of being nontreated.

* Random assignment implies that the treatment assignment ignores the possible impact of the treatment on the outcomes.
* Let be the vector of observations on the scalar-valued outcome variable *y*, a vector of observable variables x, and a binary indicator of a treatment variable .
* Measuring the effect of the treatment would be to construct a measure that compares the average outcomes of the treated and nontreated groups.
* The problem is that there is no random assignment mechanism for treatment.
* Most treatment evaluation studies have a partial equilibrium character.
* This assumption will not do if one were considering a treatment program that affected an entire sector that was a signiﬁcant part of the national economy.

1. Conditional Independence Assumption

* An important assumption is the ***conditional independence assumption*** .
* Participation in the treatment program does not depend on outcomes, after controlling for the variation in outcomes included by difference in .
* A stronger assumption
* The more commonly used assumption is
* The conditional independence assumption is broad and implies the following:
* Participation decision does not affect the ***distribution of potential outcomes***.
* To see the impact of this assumption, let be linear.

where*.*

* An assumption that is weaker is .

This assumption is used in establishing identiﬁability of a population -- average treatment effect on the treated (ATET). (unconfoundedness assumption or ignorability assumption)

* The assumption is tantamount to treatment assignment that ignores outcomes; hence it is appropriate to refer to it as the ignorability assumption.

1. **Matching Assumption**

* This assumption is necessary if the treatment variable is to be treated as exogenous, which is essential for simplicity in estimation.
* It states that .
* This assumption ensures that for each value of there are both treated and nontreated cases.
* If the assumption were to fail, then we could potentially have individuals with vectors who are all treated and those with a different x who are all untreated.
* The condition is sufﬁcient.

1. Conditional Mean Assumption

which implies that does not determine participation.

1. **Propensity Scores**

* When treatment participation is not by random assignment but depends stochastically on a vector of observable variables , then the concept of propensity scores is useful.
* .
* An assumption that plays an important role in treatment evaluation is the balancing condition, which states that.
* This can be expressed alternatively by saying that for individuals with the same propensity score the assignment to treatment is random and should look identical in terms of their vector. The balancing condition is a testable hypothesis.
* .
* The intuition behind this result is that is a particular function of and, in a sense, contains less information than *.* Hence conditional independence given is implied for the same given

1. Treatment Effects and Selection Bias

* We begin by presenting two-widely used measures of treatment effect – one that averages over all individuals and one that averages over only the treated.

1. Two Key Parameters: ATE and ATET

* Deﬁne as the difference between the outcome in the treated and untreated states,

where we may condition on if desired.

* It is emphasized that is not directly observable because no individual can be observed in both states.
* Population values of the average treatment effect and average treatment effect on the treated are deﬁned as
* With sample analogues

where

* Each of these two cases, computation procedure is not direct.
* The ATE measure is relevant when the treatment has universal applicability.
* The ATET measure is relevant when we want to consider the average gain from treatment for the treated.
* Consider the average gain from participation given characteristics x.
* Given a sample of participants, can be estimated. But is not observable
* To make ATE operational we must ﬁnd an estimator for the second term
* ATE = E[|**x***, D* = 1] − E[|**x***, D* = 0]

= (**x**) − (**x**) + E[|**x***, D* = 1] − E[|**x***, D* = 0]

= (**x**) − (**x**) + E[|**x**] − E[|**x**]

= (**x**) − (**x**)*,*

1. Sampling and Selection Bias

* is unobservable.
* Use the observations in the treated group who didn’t accept the treatment to replace.
* **randomization bias** and **substitution bias**.
* Suppose that for the treated participants the outcome equation is

= E [|**x**] + *u*1

= (**x**) +

* and for the nonparticipants the equation is

= E [|**x**] +

= (**x**) + *.*

* Note that this speciﬁcation is of the switching regression type.
* We assume that E[|**x**] =E[|**x**] = 0, though E[|**x***, D*] and E[|**x***, D*] do not necessarily equal zero.
* A more common, but restrictive, speciﬁcation has

(**x**) = (**x**) + *αD.*

* The observed outcome *y* is written as

*y* = *D* + (1 − *D*) *.*

* Combining these equations we get

*y* = *D*((**x**) +)+ (1 − *D*)((**x**) + )

= (**x**) + *D* ((**x**) − (**x**) +−)+*.*

* Switching regression
* The second term measures the beneﬁt of participation.
* (**x**) − (**x**) measures the average gain to a participant with characteristics **x**
* (− ) is individual-speciﬁc beneﬁt.
* The second component may be observable by the participant, but not by the investigator.
* Average selection bias is the difference between program participants and nonparticipants in the base state.
* A special case is E[ − |**x***, D* = 1] = 0
* Selection bias arises when the treatment variable is correlated with the error in the outcome equation.
* This correlation could be induced by incorrectly omitted observable variables that partly determine *D* and *y*.
* Another source comprises unobserved factors that partly determine both *D* and *y.*

1. Selection on Observables

* In observational data, the problem of selection on observables is solved using regression and matching methods.
* The **control function estimator** is motivated by the possibility that a set of observable variables **z** that determine *D* may be correlated with the outcomes.
* Let us consider

=+ *α* +

* E [|*,* ] = E[|*, ,* ]
* In the case of **selection on observables** we may have .

Let us write

*.*

* This is a **control function estimator** based on the OLS/GLS estimation.
* Be careful of **selection on unobservables**.

1. Selection on Unobservables

* We now consider a special linear case in which the treatment participation decision is endogenous.
* “Endogenous dummy variable.”
* The breakdown of the conditional independence assumption implies that the simple least-squares regression cannot identify the ATE. The identification approach involves fairly strong identifying assumptions and is fully parametric.
* The conditional means in the outcome equations are takento be linear. The model is completed by adding a participation (binary) decision equation for .

where is a latent variable such that

* and it is assumed that E[|**x***,* **z**] = E[|**x***,* **z**] = 0.
* The variables **z** may overlap with **x***,* but it is assumed that at least one component of **z***,* denoted *,* is unique and is a nontrivial determinant of *D.*
* is an instrumental variable.
* It is assumed that the triple ( *, ,* ) is jointly multivariate normal distributed with zero mean and covariance matrix **Σ** given by
* The beneﬁt of participation, or the ATET, is given by

* In some sample situations this identiﬁcation strategy may be somewhat fragile.
* These estimators generally presume selection on observables only.

1. Matching and Propensity Score Estimators

* The counterfactual of the ATE is not identiﬁed. As a substitute we may obtain data from a set of potential comparison units that are not necessarily drawn from the same population as the treated units, but for whom the observable characteristics, **x**, match those of the treated units up to some selected degree of closeness.
* The average outcome for the untreated matched group identiﬁes the mean counter- factual outcome for the treated group in the absence of the treatment.
* Deﬁne the matching criteria

Treatment Effect Assumptions

* It is assumed the **overlap** (or **support) condition** (0<prob(D=1|x)<1) applies, which means that for every **x** there is a positive probability of nonparticipation.

* The control and treated populations have comparable observed characteristics.
* If *,* is imputed using the estimated conditional regression function *.*
* Matching estimators impute the missing value using the out- comes of the “nearest neighbors”.
* The matching estimator typically approximates the difference between the means.
* The variance of the estimator is estimated using many of the available results on variance of differences between the means.
* Matching is a persuasive and attractive methodology if (1) we can control for a rich set of **x** variables, (2) there are many potential controls, and (3) ATET is the parameter of interest. It also requires the “no general equilibrium effects” assumption.
* The initial step of establishing the nearest matches for each observation will also clarify whether comparable control observations are available.
* Suppose the treated cases are matched in terms of all observable covariates.

The average treatment effect is

* If the data involves some observed covariates and if treated and nontreated groups are matched on each combination of covariates.
* The average of the differential over all treated individuals and all measures the average treatment effect.

E[− |= 1] = E[{E [| *,* = 1] − E [| *,* = 0]}|= 1]

= E [|= 1]

* If the **x** variables are discrete, then the matching estimator is deﬁned

E[− |= 1] =

Exact Matching

* **Exact matching** is practicable when the vector of covariates is discrete and the sample contains many observations at each distinct value of .

* Inexact matching works by mapping x into a lower dimensional measure, continuous or discrete, usually a scalar that forms the basis of matching.

Propensity Scores

* The method of propensity scores is a popular inexact matching method.
* The comparison units are those whose propensity scores are sufﬁciently close to the treated unit.
* The **propensity score** isthe conditional probability of receiving treatment given **x**, denoted *p*(**x**).
* The propensity score is usually estimated using a parametric model such as a logit or probit but can in principle be estimated using nonparametric methods.
* 1.Matching Using Propensity Scores
* In the method of propensity scores one controls for the covariates by controlling for a particular function of the covariates, speciﬁcally the conditional probability of treatment, Pr [= 1|] *.*
* Matching is on the propensity score.
* This can be easily calculated by (for example) a logit regression.
* 2.Implementation Issues
* Our interest is in estimating consistently the participation probability rather than the estimates of parameters in the propensity score function.
* In implementing matching based on *p*() three issues are relevant: (1) whether to match with or without replacement, (2) the number of units to use in the comparison set, and (3) the choice of the matching method.
* Matching without replacement means that any observation in the comparison group is matched to no more than one treated observation, that which is the closest match.
* The issue of choosing the number of cases in the comparison set involves tradeoff between bias and variance.
* The variance is reduced whereas bias increases.
* A partial solution is to use the better matches—— “**caliper matching**.”
* Data quality plays a key role in robust estimation of treatment effects by matching methods.
* The issue of the sensitivity of the results to the chosen method is not amenable to a simple direct answer.
* The outcome may vary across different samples, depending on the extent of overlap between the treated and untreated observations.