STATISTICAL HORIZONS

Propensity Score Analysis

Shenyang Guo, Ph.D.

Upcoming Seminar:

September 10-12, 2020, Remote Seminar

Propensity Score Analysis

1. Overview

1.1 Observational studies and challenges
1.2 Why and when PSA is needed?
1.3 Overview of corrective methods

Recommended Textbooks

- Guo, S. & Fraser, W.M. (2014). *Propensity Score Analysis: Statistical Methods and Applications, Second Edition*. Thousand Oaks, CA: Sage Publications.
- Morgan, S.L, & Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.
- Rosenbaum, P. R. (2010). *Design of Observational Studies*. New York: Springer.

Observational Studies

- An observational study is an empirical investigation whose objective is to elucidate causal relationships when it is infeasible to use randomized controlled trials (Cochran, 1965).
- Observational data: survey, census, administrative, or any data that were not generated by RCT.
- Observational studies ~ evaluations with a quasi-experimental design (Shadish, Cook, & Campbell, 2002).

Association $\neq$ Causation

Three criteria for a causal relation (Lazarsfeld, 1959):

1. A causal relationship between two variables must have temporal order, in which the cause must precede the effect in time
2. The two variables should be empirically correlated with one another
3. Most important, the correlation is not spurious

Purpose of Evaluation

The field of program evaluation is distinguished principally by cause-effect studies that aim to answer a key question:

To what extent can the net difference observed in outcomes between treated and nontreated groups be attributed to an intervention, given that all other things are held constant?

Note. The term "intervention research" refers to the design and evaluation of programs.

Internal Validity and Threats

- Internal validity – the validity of inferences about whether the relationship between two variables is causal (Shadish, Cook, & Campbell, 2002).
- In program evaluation and observational studies in general, researchers are concerned about threats to internal validity. These threats are factors affecting outcomes other than intervention or the focal stimuli. There are nine types of threats.*
- Selection bias is the most problematic one!

*These include differential attrition, maturation, regression to the mean, instrumentation, and testing effects.

Why and when propensity score analysis is needed? (1)

Need 1: Remove Selection Bias

The randomized clinical trial is the "gold standard" in outcome evaluation. However, in social and health research, RCTs are not always practical, ethical, or even desirable. Under such conditions, evaluators often use quasi-experimental designs, which – in most instances – are vulnerable to selection. Propensity score models help to remove selection bias.

Example: In an evaluation of the effect of Catholic versus public school on learning, Morgan (2001) found that the Catholic school effect is strongest among Catholic school students who are less likely to attend Catholic schools.

Why and when propensity score analysis is needed? (2)

Need 2: Analyze causal effects in observational studies

- Observational data - those that are not generated by mechanisms of randomized experiments, such as surveys, administrative records, and census data.
- To analyze such data, an ordinary least square (OLS) regression model using a dichotomous indicator of treatment does not work, because in such model the error term is correlated with explanatory variables. The violation of OLS assumption will cause an inflated and asymptotically biased estimate of treatment effect.

The Problem of Contemporaneous Correlation in Regression Analysis

Consider a routine regression equation for the outcome, Y_i:

$$Y_i = \alpha + \tau W_i + \beta X_i + e_i$$

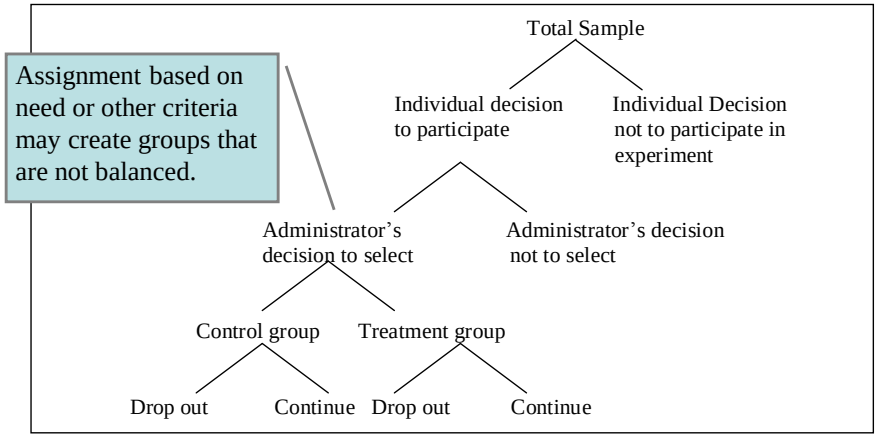
where W_i is a dichotomous variable indicating intervention, and X_i is the vector of covariates for case i.

In this approach, we wish to estimate the effect (τ) of treatment (W) on Y_i by controlling for observed confounding variables (X_i).

When randomization is compromised or not used, the correlation between W and e may not be equal to zero. As a result, the ordinary least square estimator of the effect of intervention (τ) may be biased and inconsistent. W is not exogenous.

Sources of Selection

Example of Selection Bias: Decision Tree for Evaluation of Social Experiments



Source: Maddala, 1983, p. 266

Overview of Corrective Methods:
Four Models Described by Guo & Fraser (2014)

1. Heckman's sample selection model (Heckman, 1976, 1978, 1979) and its revised version estimating treatment effects (Maddala, 1983)

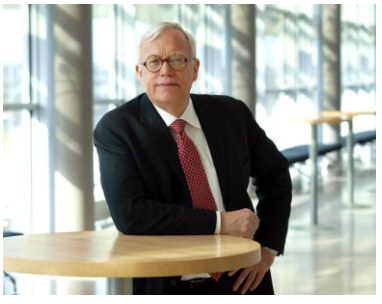
Overview of Corrective Methods:
Four Models Described by Guo & Fraser (2014)

2. Propensity score matching (Rosenbaum & Rubin, 1983), optimal matching (Rosenbaum, 2002), propensity score weighting, modeling treatment dosage, and related models

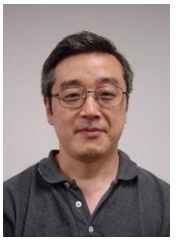
Overview of Corrective Methods:
Four Models Described by Guo & Fraser (2014)

3. Matching estimators (Abadie & Imbens, 2002, 2006)

Overview of Corrective Methods:
Four Models Described by Guo & Fraser (2014)



4. Propensity score analysis with nonparametric regression (Heckman, Ichimura, & Todd, 1997, 1998)



Other Corrective Models

- Instrumental variables approaches (Guo & Fraser [2010] reviews this method)
- Regression discontinuity design
- Interrupted time series design
- Bayesian approaches to inference for average treatment effects
- Marginal structural models (Robins, 1999a, 1999b)
- Directed acyclic graphs (Pearl, 2000)

List of Programs Conducting Propensity Score Analysis (Elizabeth Stuart)

http://www.biostat.jhsph.edu/~estuart/propensityscoressoftware.html

Propensity Score Analysis

2. Conceptual Frameworks & Assumptions

2.1 The Neyman-Rubin counterfactual framework
2.2 The assumption of strongly ignorable treatment assignment
2.3 The stable unit treatment value assumption
2.4 Heckman's Scientific Model of Causality
2.5 Two Traditions

Readings for Session 2

Guo & Fraser, chapter 2.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66,* 688-701.

Heckman, J. J. (2005). The scientific model of causality. *Sociological Methodology, 35*, 1-97.

Holland, P. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association, 81,* 945-970.

Rubin, D.B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics,2,* 808-840.

Sobel, M. E. (2005). Discussion: "The scientific model of causality." *Sociological Methodology*, 35, 99–133.

Background Information (1)

- A factual account (regularity): “A has caused B”. A counterfactual account (i.e., abandoning the regularity account): “B would not have occurred if it were not for A”.
- The logic here is analogous to hypothesis testing (H_a and H_0).
- Counterfactuals are at the heart of science. Much of the discussion about counterfactuals is philosophical (e.g., ancient Greek philosophers such as Aristotle; David Hume 1748; John Stuart Mill 1843; David Lewis, 1973; Galileo Galilei 1564-1642).

Background Information (2)

- The formal use of the counterfactual framework to define unit-level causal effects is due to Neyman in 1923 in the context of randomized experiments, and was marvelously clarifying contribution.
- Important studies using counterfactual framework for randomized experiment are: Fisher (1935), Kempthorne (1952), Cochran and Cox (1950), and Cox (1958).
- Rubin (1974) was the first to define causal effects in both randomized experiments and observational studies. This is a milestone. It signifies that the same underlying principles can be used to design both types of studies.
- Observational studies guided by this framework can be thought as a correction of, or balancing, data to make crucial assumptions embedded in the randomized experiment tenable.

The Neyman-Rubin Counterfactual Framework (1)

- **Counterfactual**: what would have happened to the treated subjects, had they not received treatment?
- The Neyman–Rubin **counterfactual framework** (CF) states that individuals selected into treatment and nontreatment groups have potential outcomes in both states: the one in which they are observed and the one in which they are not observed. This framework is expressed as:

$$Y_i = W_i Y_{1i} + (1 - W_i) Y_{0i}$$

- The key message conveyed in this equation is that to infer a causal relationship between W_i (the cause) and Y_i (the outcome) the analyst cannot directly link Y_{1i} to W_i under the condition $W_i = 1$; instead, the analyst must check the outcome of Y_{0i} under the condition of $W_i = 0$, and compare Y_{0i} with Y_{1i}.

The Neyman-Rubin Counterfactual Framework (2)

- There is a crucial problem in the above formulation: Y_{0i} is not observed. Holland (1986, p. 947) called this issue the "fundamental problem of causal inference."
- The Neyman-Rubin CF holds that a researcher can estimate the counterfactual by examining the average outcome of the treatment participants (i.e., $E(Y_1|W=1)$]) and the average outcome of the nontreatment participants [i.e., $E(Y_0|W=0)$] in the population. Because both outcomes are observable, we can then define the treatment effect as a mean difference (the equation is known as "standard estimator for the average treatment effect"):

$$\tau = E(Y_1|W=1) - E(Y_0|W=0)$$

- With sample data, the estimator becomes:

$$\hat{\tau} = E(\hat{y}_1 \mid w = 1) - E(\hat{y}_0 \mid w = 0)$$

The Strongly Ignorable Treatment Assignment Assumption (1)

- The strongly ignorable treatment assignment (SITA) assumption (Rosenbaum & Rubin, 1983):

$$(Y_0, Y_1) \perp W \mid X.$$

- Different versions: "unconfoundedness" and "ignorable treatment assignment" (Rosenbaum & Robin, 1983), "selection on observables" (Barnow, Cain, & Goldberger, 1980), "conditional independence" (Lechner 1999), and "exogeneity" (Imbens, 2004)

The Strongly Ignorable Treatment Assignment Assumption (2)

The SITA assumption is the same assumption embedded in OLS regression

$$Y_i = \alpha + \tau W_i + \beta X_i + e_i$$

about the independence of the error term e_i from W_i (i.e., the "*contemporaneous independence" assumption or "exogeneity"*).

Comments about the SITA Assumption (1)

- When the treatment assignment is not ignorable, the use of the dummy variable *W* leads to endogeneity bias. Conceptualizing W as a dummy endogenous variable motivated Heckman (1978, 1979) to develop the sample selection model and Maddala (1983) to develop the treatment effect model.

Comments about the SITA Assumption (2)

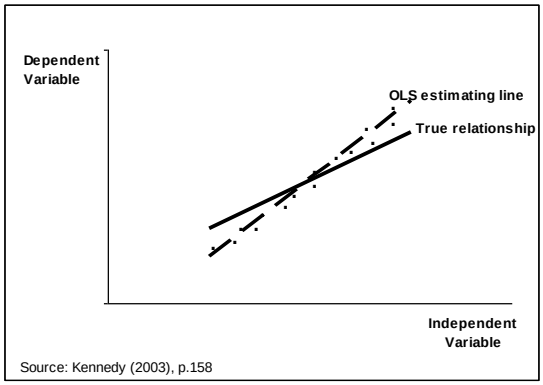
- The endogeneity problem leads to a biased and inconsistent estimation of the regression coefficient.

Assuming all variables are centered, we have $y|x = \beta_1 x + e$. The least squares estimate is $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$. This leads to $\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n} x_i e_i}{\sum_{i=1}^{n} x_i^2}$.

Hence, when *x* and *e* are correlated, the expected value for the far right-hand term will be nonzero, and the numerator will not go to zero as the sample size increases without limit. The least squares estimate then will be biased and inconsistent.

Comments about the SITA Assumption (3)

The problem is also known as “inflated slope” and “asymptotical bias” (Kennedy, 2003):



Source: Kennedy (2003), p.158

Comments about the SITA Assumption (4)

- In observational studies, the problem is often reflected as a “unmeasured variables” problem:

Suppose that a correctly specified regression model would be

$$\boldsymbol{y} = \boldsymbol{X_1}\beta_1 + \boldsymbol{X_2}\beta_2 + \varepsilon.$$

If we regress y on $\boldsymbol{X_1}$ without including $\boldsymbol{X_2}$, then the estimator becomes:

$$b_1 = (X_1'X_1)^{-1}X_1'y = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon.$$

Taking the expectation, we see that unless $X_1'X_2 = 0$ or $\beta_2 = 0$, b_1 is biased. The well-known result is the omitted variable formula

$E[b_1|X] = \beta_1 + P_{1.2}\beta_2$, where $P_{1.2}\beta_2$ is the bias.

Implications of the SITA Assumption

- Observational studies can be viewed as a process to reconstruct the data to correct for the violation of SITA.
- Recently, Rubin (2008) formally and explicitly defines this work (i.e., balance data to correct for violation of SITA) as the **design** stage of an observational study.
- Six essential steps for the design:
  1. Conceptualize the observational study as having arisen from a complex randomized experiment.
  2. What was the hypothetical randomized experiment that led to the observed dataset?
  3. Are sample sizes in the dataset adequate?
  4. Who are the decision makers for treatment assignment and what measurements were available to them?
  5. Are key covariates measured well?
  6. Can balance be achieved on key covariates?

The SUTVA Assumption (1)

- To evaluate program effects, statisticians also make the *Stable Unit Treatment Value Assumption*, or SUTVA (Rubin, 1980, 1986), which says that the potential outcomes for any unit do not vary with the treatments assigned to any other units, and there are no different versions of the treatment.
- Imbens (on his Web page) uses an aspirin example to interpret this assumption, that is, the first part of the assumption says that taking aspirin has no effect on your headache, and the second part of the assumption rules out differences on outcome due to different aspirin tablets.