The Important Role of Heterogeneity in Social and Biological Models

Professor Jeffrey Smith
Department of Economics
Gerald R. Ford School of Public Policy Studies
and Institute for Social Research
University of Michigan
econjeff@umich.edu

RCGD / IHPI Seminar
January 12, 2015

Who the heck are you?

Ph.D. from Chicago in 1996, supervised by Jim Heckman

Western Ontario, 1994-2001
Maryland, 2001-2005
Michigan, 2005-present

Main affiliation in economics, plus Ford School, plus QMP and PRC at ISR

An interdisciplinary guy

My research

Social experiments

Non-experimental methods (especially matching and regression discontinuity)

Performance management

Active labor market programs

University quality / mismatch and related issues in higher education

Participant evaluation

Introduction

Different disciplines approach these issues differently

My goal: provide a brief, accessible introduction to how economists think about these issues

Notation

The Frost-Fisher-Neyman-Roy-Quandt-Rubin potential outcomes framework

Let Y_1 denote the outcome in the treated state

Let Y_0 denote the outcome in the untreated state

The fundamental evaluation problem is that we observe at most one of these two outcomes for each person

Let D be an indicator for participation in the program

Let R be an indicator for randomization into the treatment group in an experiment

Usual parameters of interest

The usual parameter of interest is the impact of treatment on the treated given by

$$E(Y_1 - Y_0 | D=1) = E(Y_1 | D=1) - E(Y_0 | D=1).$$

This is sometimes called the ATET or TOT or ATT or just TT

In an experiment, this parameter is estimated by

$$E(Y_1 \mid D=1, R=1) - E(Y_0 \mid D=1, R=0).$$

In non-experimental evaluations, the unobserved counterfactual is obtained through econometric manipulation of the outcomes of non-participants: persons with $(D=0)$

Other parameters of interest

Parameters requiring the joint distribution of outcomes

Experiments and standard non-experimental methods provide only the marginal outcome distributions $f(Y_1 \mid D=1, R=1)$ and $f(Y_0 \mid D=1, R=1)$

Some parameters require the joint distribution $f(Y_1, Y_0 \mid D=1, R=1)$

Example: Fraction gaining or losing from the program

Example: Percentiles (e.g. median) of the impact distribution.

Example: Impact variance

Example: Outcome correlation.

Other parameters of interest (continued)

Heckman, Smith and Clements (1997) discuss how to use bounds and other methods to obtain information on parameters that depend on the joint distribution

Recent application and survey: Djebbari and Smith (2008)

Bounds / set identification / partial identification

Simple example: 2 x 2
Rows: employment status in control state: NE: 0.4, E: 0.6
Columns: employment status in treatment state: NE: 0.2, E: 0.8
These are the marginal distributions provided by an experiment

The formula for the Frechét-Höffding bounds is:

$$\max[F_1(Y_1)+F_0(Y_0)-1,0)]\leq F(Y_1,Y_0)\leq\min[F_1(Y_1),F_0(Y_0)].$$

Apply to the case of the (NE, NE) or (0, 0) cell in the example

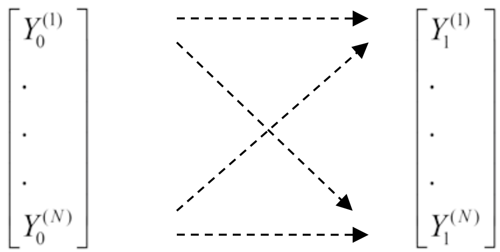
$F_1(0)=0.2$ and $F_0(0)=0.4$

$$\max[0.2+0.4-1,0]\leq F(0,0)\leq\min[0.2,0.4].$$

Thus, the probability of (NE, NE) is between 0.0 and 0.2. Cool!

Bounds (continued)

The bounds correspond to the cases of rank correlations of -1 and 1.



Rank preservation minimizes the impact variance while rank inversion maximizes it.

Heckman, Smith and Clements (1997) show how to use the estimated distribution of impacts in the rank preservation case to test the null of the common effect model, i.e. of a zero impact variance.

HSC (1997) strongly reject the null in their empirical example

Quantile treatment effects

Compare quantiles of the treated and untreated outcome distributions

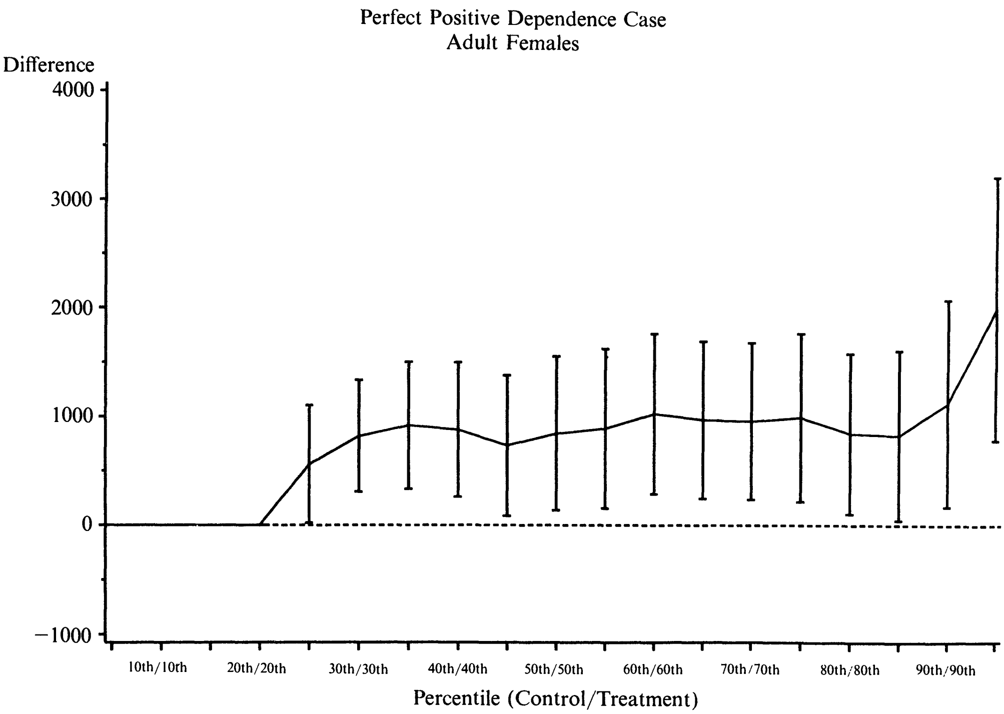
Interpretation 1: impacts on quantiles of the outcome distribution

Interpretation 2: impacts at quantiles of the outcome distribution

The latter interpretation requires an assumption about the joint distribution of outcomes, namely rank preservation.

Testing rank preservation: see Bitler, Gelbach and Hoynes (2005)
Intuition: covariate balance at percentiles of the outcome distribution

The impacts under rank preservation provide the lower bound on the impact variance



1 National JTPA Study 18 month impact sample
2. Standard errors for the quantiles are obtained using methods described in Csorgo (1993)

FIGURE 1

Treatment–control differences at percentiles of the 18 month earnings distribution

In an Appendix available on request, we explore the sensitivity of these estimates to measurement error in earnings. Our basic inferences are not altered, including our major inference bounding the variability in programme impacts away from zero.

(*c*) *The discrete case*

The Fréchet–Hoeffding bounds apply to all bivariate outcome distributions.[20] Variables may be discrete, continuous or both discrete and continuous. In this section, we use the bounding distributions to establish the variability in the distribution of impacts on employment status. The latent distribution underlying this situation is multinomial.[21] Let (E, E) denote the event "employed with treatment" and "employed without treatment" and let (E, N) be the event "employed with treatment, not employed without treatment." Similarly, (N, E) and (N, N) refer respectively to cases where a person would not be employed if treated but would be employed if not treated, and where a person would not be employed in either state. The probabilities associated with these events are P_{EE}, P_{EN}, P_{NE} and P_{NN}, respectively. This model can be written in the form of a contingency table. The columns refer to employment and non-employment in the untreated state. The rows refer to employment and non-employment in the treated state.

20. Formulae for multivariate bounds are given in Tchen (1980) and Rüschendorf (1982).
21. The following formulation owes a lot to the missing cell literature in contingency table analysis. See, e.g. Bishop, Fienberg and Holland (1975).

Random coefficient models

An alternative to the rank preservation assumption for identifying the joint distribution of outcomes (and thereby the distribution of impacts)

Assumes impacts uncorrelated with untreated outcome: $(Y_1 - Y_0) \perp D$

In the simple case, the variance of the treatment effect is the difference in the outcome variance between the treated and untreated units:

$$\text{var}(Y_1) = \text{var}(Y_0 + \Delta) = \text{var}(Y_0) + \text{var}(\Delta) \Rightarrow \text{var}(\Delta) = \text{var}(Y_1) - \text{var}(Y_0)$$

Note the implicit test based on the sign of the estimated impact variance.

Random coefficient models (continued)

Can assume a normal distribution (the classic case, often invoked in HLM / MLM style analyses)

Can assume a flexible parametric form

Can estimate non-parametrically via deconvolution as in HSC (1997)

When is this model *ever* economically plausible for voluntary treatments?

Subgroup effects

What works for whom?

Useful for targeting / statistical treatment rules

Example: US Worker Profiling and Reemployment Services System
Example: Response to Intervention (RtI) in education
Example: SMART (Sequential, Multiple Assignment, Randomized Trial)

See e.g. Smith and Staghøj (2008) and Manski (2004)

Systemic versus idiosyncratic variation in treatment effects
See e.g. Djebbari and Smith (2008), Bitler, Gelbach and Hoynes (2014)

Are subgroup effects structural?

In economics, structural means “policy invariant”

Structural subgroup effects are often assumed in the literature, particularly the literature on statistical treatment rules

What if effects are heterogeneous within subgroups? Consider an example:

Half of men have impact 10 and half have impact 4
Half of women have impact 12 and half have impact 1

Assume that the cost of participation is five, so top half of both groups participate if potential participants know their impacts

Evaluation finds program “works better for women” so gender-specific subsidies are provided to induce the remaining women to participate ….

Are subgroup effects structural (continued)

General points:

(1) Conditional mean impacts on the treated in general do not equal impact on marginal untreated person

(2) This relationship may vary across subgroups

Models of heterogeneous treatment effects

May or may not be models of subgroup effects

Relates to the question of whether subgroup effects are "structural"

Important for understanding mechanisms
Ex: Pitt, Rosenzweig and Hassan (2010) on gender differences in the impact of education

Can provide testable predictions
Ex: Bitler, Gelbach and Hoynes (2006) on Connecticut Jobs First

Can provide restrictions on the joint distribution of outcomes
Ex: Kline and Tartari (2014) also on Connecticut Jobs First

Huge opportunities for research here

Site / context effects and external validity

Suppose that individual impacts depend on both unit characteristics and site / context characteristics, as in:

$\Delta_{is} = g(X_i, X_s)$, where not all characteristics may be observed

Example of unit characteristics: general health, age
Example of context characteristics: family members who smoke

How to generalize the results of an experiment implemented in a particular population with particular unit and context characteristics

This is a problem of extrapolation. It requires, implicitly or explicitly, a model.

External validity (continued)

Key issues identified in Hotz, Imbens and Mortimer (2005):

1. Selection into the study from the population of possible sites

2. Common support

Usual approach: reweighting based on observed unit and context characteristics, paying attention to support issues

External validity concerns also have important implications for the design of the experiment (i.e. for initial site selection)

See also: Gechter (2014) and Muller (2014) and the Imbens (2013) review of the Manski (2013) book.

Subgroup effects and fishing

Problem: if you estimate and report enough subgroup impacts, some of them will be statistically significant, even if they all really equal zero

Special case of what the literature calls the multiple comparisons problem, and a good illustration of the occasional oddness of classical statistical inference

Responses to fishing

1. Pre-commitment: confirmatory versus exploratory subgroup analyses

In US Department of Education evaluations, confirmatory outcomes / subgroups subject to adjustment for multiple comparisons while exploratory analyses are not.

2. Adjustment of p-values for multiple comparisons
See Schochet (2008) NCEE

3. Dimension reduction versus domain-specific indices
Ex: MTO studied by Kling, Liebman and Katz (2007)

See also Andersson (2008) for a Perry Preschool application.

Summary and conclusions

There are more parameters of interest than just ATET and ATE

Many methods exist for estimating those parameters, but more research remains to be done, especially to integrate the economics and the econometrics

Subgroup effects and site effects are not as simple as they might seem

External validity is important and economists have not thought much about it.

References / reading list

Anderson, Michael. 2008. “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of the American Statistical Association* 103(484): 1481-1495.

Bitler, Marianne, Jonah Gelbach, and Hilary Hoynes. 2005. “Distributional Impacts of the Self-Sufficiency Project.” NBER Working Paper No. 11626.

Bitler, Marianne, Jonah Gelbach, and Hilary Hoynes. 2006. “What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments.” *American Economic Review* 96: 988–1012.

Bitler, Marianne, Jonah Gelbach, and Hilary Hoynes. 2014. “Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment.” NBER Working Paper No. 20142.

Dehejia, Rajeev. 2003. “Was There a Riverside Miracle? A Framework for Evaluating Multi-Site Programs.” *Journal of Business and Economic Statistics* 21(1): 1-11.

Djebbari, Habiba and Jeffrey Smith. 2008. “Heterogeneous Impacts in PROGRESA.” *Journal of Econometrics* 145: 64-80.

Gechter, Michael. 2014. “Generalizing the Results from Social Experiments.” Unpublished manuscript, Boston University.

Heckman, James, Jeffrey Smith and Nancy Clements. 1997. “Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts.” *Review of Economic Studies* 64: 487-535.

Hotz, V. Joseph, Guido Imbens and Julie Mortimer. 2005. “Predicting the Efficacy of Future Training Programs Using Past Experiences at other Locations.” *Journal of Econometrics* 125(408): 241–270.

Imbens, Guido. 2013. “Book Review Feature: Public Policy in an Uncertain World.” *Economic Journal* 123: F401-F411.

Kling, Jeffrey, Jeffrey Liebman and Lawrence Katz. 2007. “Experimental Analysis of Neighborhood Effects.” *Econometrica* 75(1): 83-119.

Kline, Patrick and Melissa Tartari. 2014. “Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach.” Unpublished manuscript, University of California at Berkeley.

Manski, Charles. 2004. “Statistical Treatment Choice: An Application to Active Labour Market Programmes.” *Econometrica* 72 (4): 1221-1246.

Manski, Charles. 2013. *Public Policy in an Uncertain World*. Cambridge, MA: Harvard University Press.

Muller, Seán. 2014. “Interaction and External Validity: Obstacles to the Policy Relevance of Randomized Evaluations.” *World Bank Economic Review*, forthcoming.

Pitt, Mark, Mark Rosenzweig and Nazmul Hassan. 2010. “Human Capital Investment and the Gender Division of Labor in a Brawn-Based Economy.” Yale Growth Center Discussion Paper No. 989.

Schochet, Peter. 2008. *Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations*. National Center for Educational Evaluation.

Smith, Jeffrey and Staghøj, Jonas. 2008. “Using Statistical Treatment Rules for Assignment of Participants in Labor Market Programs.” Unpublished manuscript, University of Michigan.

Weiss, Michael, Howard Bloom and Thomas Brock. 2014. “A Conceptual Framework for Studying the Sources of Variation in Program Effects.” *Journal of Policy Analysis and*

Management 33(3): 778-808.