Journal of Econometrics 198 (2017) 41–64

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Evidence of randomisation bias in a large-scale social experiment: The case of ERA

Barbara Sianesi *

Institute for Fiscal Studies, 7 Ridgmount Street, WC1E 7AE London, UK

ARTICLE INFO

Article history:
Received 2 June 2014
Received in revised form
28 November 2015
Accepted 15 January 2017
Available online 25 January 2017

JEL classification:
C14
C21
J18
J38

Keywords:
Social experiments
Randomisation bias
Sample selection
Treatment effects
Matching methods
Reweighting estimators
Partial identification
External validity

ABSTRACT

We set out a theoretical framework for the systematic consideration of 'randomisation bias', estimate the causal impact of randomisation on participation patterns in an actual trial, and propose a non-experimental way of assessing the extent to which the experimental impacts are representative of the impacts that would have been experienced by the study sample that would have been obtained in the absence of random assignment. We also extend our estimator to deal with binary outcomes and to account for selective survey non-response, and explore partial and point identification of the parameter of interest under alternative assumptions on the selection process.

1. Introduction

Widely hailed as the gold standard in program evaluation, social experiments are the most reliable partial-equilibrium method for evaluating whether a program works, on average, for its participants[1]—provided certain conditions are met. An overarching label for such identifying conditions is the "no randomisation bias" assumption (a term coined by Heckman, 1992), which rules out that random assignment *per se* has affected potential outcomes, as well as the program participation process. In the presence of randomisation bias, the evaluation device of random assignment effectively prevents the study from recovering the causal parameter it was set up to obtain. One of the most powerful critiques of the use of randomised experiments in the social sciences is thus the possibility that individuals might react to the randomisation itself, thereby potentially affecting the validity and policy-relevance of the causal inference emerging from the trial. To our knowledge, however, there is to date no robust empirical evidence on the existence and scope of randomisation bias in actual social experiments.

After setting out a theoretical framework for the systematic consideration of randomisation bias, this paper documents for the first time how the process of randomly allocating individuals has modified who has participated in an actual large-scale social experiment and estimates the extent to which this kind of randomisation bias has altered the treatment effect parameter being recovered.

The issue which motivated the paper arose in the Employment Retention and Advancement (ERA) study, which ran in six districts across the UK between 2003 and 2007 and randomly assigned over 16,000 individuals. The largest randomised trial of a social program in the UK at the time, it was set up to test the effectiveness of offering time-limited support once in work, in the form of advisory

* Fax: +44 0 2073234780.
E-mail address: barbara_s@ifs.org.uk.

1 For a discussion and appraisal of social experiments, see e.g. Burtless (1995) and Heckman and Smith (1995).

http://dx.doi.org/10.1016/j.jeconom.2017.01.003

services and a new set of financial incentives rewarding sustained full-time work and the completion of training whilst employed. Eligible for this initiative were long-term unemployed over the age of 25 mandated to enter the New Deal 25 Plus (ND25+) program, and lone parents who volunteered for the New Deal for Lone Parents (NDLP) program.[2] In the first follow-up year, the employment chances of both intake groups remained largely unaffected, while a sizeable experimental impact was found in terms of earnings, especially for the NDLP group (see Hendra et al., 2011, for the final appraisal of ERA).

It has however emerged that participants in the ERA study were a selected subsample of the population who would receive the treatment in routine mode: some eligibles actively refused to be randomly assigned (the "formal refusers"), while some were diverted from the treatment due to caseworkers' incentives that would not have played out absent randomisation (the "diverted customers"). A sizeable fraction of the eligibles – 23% of ND25+ and 30% of NDLP – were thus not represented in the experiment.

While the policymaker would be interested in the average treatment effect of offering ERA services and incentives for all those who would have been eligible to receive such an offer in the absence of randomisation (the *ATE*), the experimental evaluation can only provide unbiased impact estimates only for those who reached the randomisation stage and agreed to be randomly assigned. It is important to stress that it was the experimental set-up *per se* which gave rise to diverted customers and formal refusers, as these eligible individuals were denied or refused participation in something which in normal circumstances one could not be denied or one could not refuse: *becoming eligible* for financial incentives and personal advice. Randomisation can thus be viewed as having affected the process of participation in ERA, resulting in an adviser-selected and self-selected subgroup which is potentially different from the sample of New Deal entrants who would have been exposed to the offer of ERA had it not been evaluated via random assignment.

We show how the eligible individuals who did not participate in the experiment differ in systematic ways from the ones who did participate. These differences can be seen not only in the covariates but also in the outcomes in the absence of treatment by comparing randomised out participants to eligible non-participants. Formal refusers (mainly within the ND25+ group) are less employable and less inclined to accept government intervention. Diverted customers (mainly within the NDLP group) are more employable and seem to have been denied the offer of the treatment by caseworkers eager to keep them in the control treatment.

Non-participation in the ERA study would then introduce randomisation bias if the average effect for the experimental group is different from the average effect which would have arisen had ERA been run in routine mode,[3] the latter coinciding in our case with the average effect on the ERA eligible population. The actual severity of randomisation bias thus effectively boils down to how much the treatment effect on the ERA study participants differs from the treatment effect on the eligible non-participants.

The uncertainty introduced by non-participation is large, as testified by the width of the agnostic bounds we construct for the policy-relevant treatment effect, *ATE*. Clearly, any attempt to pin down a point estimate for the *ATE* will heavily rely on the identifying assumptions one is willing to make to extrapolate outside of the experimental sample. To somehow mitigate this concern we present evidence based on alternative assumptions.

As to point identification, we explore two non-experimental strategies.

The first one is based on the standard conditional independence assumption (CIA) that we observe all the ERA outcome-relevant characteristics that drive selection into the experiment.

While our data include demographics, information on the current unemployment spell, extremely detailed labour market histories over the previous three years and local factors, the CIA needed to identify the average treatment effect on the non-treated (the non-participants in our case) is admittedly strong. Under the assumption of no residual selection into the study based on unobserved idiosyncratic impact components, we can however formally test the validity of this standard CIA in terms of treatment outcomes by testing conditional independence with respect to no-treatment outcomes, and correct the standard non-experimental estimates from any selection bias arising from rejection of the latter condition. Of course, the corrected estimates are equivalent to those derived directly under the assumption of no selection on the gain. Our second, and preferred, strategy relies on this assumption, which is supported by the well-documented inability of both individuals and caseworkers to forecast treatment effects.

As advocated by Manski (1996) in the case of "experimentation on a context-specific subpopulation", we also explore bounds for the treatment effect of substantive interest based on the information available in the data and on standard non-parametric restrictions. Interestingly, our set-up allows us to assess whether these restrictions hold in terms of non-ERA outcomes.

We further extend our proposed estimators to deal with the non-linear case of binary outcomes and, for the case of survey-based outcome measures, to also account for selective non-response based on observed characteristics.

To summarise, the objective of the paper is twofold: to quantify the causal effect of randomisation on participation patterns and to assess the extent of randomisation bias this has introduced. While by its nature this second contribution is more tentative as it ultimately relies on untestable identifying assumptions, the first causal effect is directly identified in our data. Indeed the beauty of the ERA study is that it offers the rare chance to empirically measure the extent to which randomisation has affected the participation process. This is because (1) the treatment is the bestowing of an eligibility (to advisory services and financial incentives); (2) the parameter of interest is the average impact of offering this eligibility (an intention to treat effect); and (3) in the absence of randomisation, the offer of this eligibility would have covered a well-defined and observed population: all ND25+ and NDLP entrants in the six districts over the intake window.

We are not aware of substantive econometric research which has looked at the issue of randomisation bias. Indeed, non-participation in the ERA study, which takes place before random assignment, is a distinct problem from non- or partial-compliance (no-shows, drop-outs, non-take up), which takes place *after* treatments have been assigned.[4] This type of non-participation

[2] These two groups represent 83% of all ERA study participants. We do not consider the third target group due to its conceptually different set-up coupled with lack of data.

[3] An alternative but, as we discuss in Section 3.1, possibly less pertinent way to consider this issue is as a threat to the external validity of the experimental estimates.

[4] The set-up and aims of Dubin and Rivers (1993) are opposite to the ones in the current paper. In their set-up, refusal to participate in the wage subsidy experiment happened after random assignment (to the program group). While their experiment thus directly recovers the intention to treat, the authors aim to tease out the impact on the actual participants. Their formal refusers could be viewed as the program group "no-shows" of Bloom (1984), and indeed the approach followed by Dubin and Rivers builds upon the Bloom (i.e. instrumental variables) estimator. Note also that the non-participants in the ERA experiment were not exposed to ERA, and thus no link can be made to the literature on "dropouts" (see e.g. Heckman et al., 2000).