Moving Beyond Mean Treatment Effects

RCTs and Distributional Questions

Research Master Thesis

Marie-Aimée Salopiata

13/07/21

Number of ECTs: 30

Supervisor: Dr. Conrad Heilmann

Advisor: Dr. William Peden

Wordcount: 30000

Erasmus Institute for Philosophy and Economics (EIPE)
Erasmus University Rotterdam

Table of Contents

List of Abbreviations

Evidence-Based Policy – EBP

Mean Treatment Effects Argument – MTE Argument

Perfect Positive Dependence Assumption – PPDA

Perfect Negative Dependence Assumption – PNDA

Randomized Controlled Trials – RCTs

Quantile Treatment Effects – QTEs

List of Figures

Introduction

Within the area of policy evaluation, the main goal is to identify the effect of a policy or program (cf. Imbens and Wooldrige 2009, p. 5). And while researchers have developed countless methods for this, probably none of them has received as much attention over the last years as the *randomized controlled trial* (in the following: RCT) (de Souza Leao et al. 2019, p. 383).

Simply put, RCTs allow us to learn about the effects of policies in an experimental way (thus "trial"). To this end, experimental subjects partaking in an RCT are *randomly allocated* into two groups (thus "randomized"). One group of experimental subjects, the treatment group, is then exposed to the policy of interest. For instance, if the policy of interest was a class-size reduction policy, then the treatment group would be taught in smaller classes. The other group of experimental subjects, the control group, is not exposed to the policy of interest. In the class-size reduction example, this group would be taught in standard-sized classes (Heckman and Smith 1995, p. 85).

Ideally, the random allocation of experimental subjects to control and treatment group makes the exposure to the policy the only systematic difference in characteristics between the two groups (thus "controlled"). If this holds, then the effect of the policy examined in the RCT can be identified by comparing average outcomes in the treatment group and the control group. In the class-size reduction RCT, we would compare the average grades of the treatment group, which was put in smaller classes, with the average grades of the control group, which was not. The resulting parameter is called the *mean treatment effect* (Heckman 2020, p. 8).

The attention that RCTs have received as a method for evaluating policies is not confined to academic policy evaluation. In addition, the influential *Evidence-Based Policy Movement*, which urges policymakers to base their implementation decisions on evidence, promotes RCTs as the privileged source of evidence for policymaking. That is, proponents of evidence-based policy (in the following: EBP) argue that the evidence upon which policymakers should base their policy implementation decisions should ideally come from RCTs. This judgment is, firstly, based on the reliability of the causal estimates that RCTs provide. Secondly, it is defended by arguing that RCTs only require minimal assumptions (Deaton and Cartwright 2018a, p. 2).

Project

In this thesis, I defend an argument that questions whether this defence of the special status of RCTs for the EBP movement holds up. The EBP movements' defence of the special status of RCTs based on the reliable causal estimates that RCTs produce and the minimal assumptions RCTs require has been criticized both within the philosophical and the economic literature. And while philosophers have mainly criticized this defence by pointing to difficulties with generalizing RCT results to other policy contexts of interest ("external validity") (Deaton and Cartwright 2018a, p. 3), this thesis focuses on another argument, from the economic literature.

The argument in the focus of this thesis is concerned with the usefulness of *mean treatment effects.* As we saw above, this parameter is usually obtained from RCT data and then presented as the effect of the policy. However, authors in the economic literature have pointed to the *insufficient informativeness* of mean treatment effects for policymaking. For instance, a mean treatment effect of zero might indicate that none of the experimental subjects experienced any effect from the policy. However, it might also suggest that some experimental subjects benefitted a lot while others did not or even got harmed. In light of the insufficient information conveyed by mean treatment effects, authors from the economic literature stress that *distributional parameters,* such as the proportion of people who benefitted from a policy, convey additional policy-relevant information (Deaton 2010, p. 439; Na et al. 2015, p. 292).

When the necessity of obtaining distributional parameters from RCT data to properly inform policymakers is defended in the literature, the following problem with obtaining these parameters is always emphasized simultaneously. Namely, in contrast to mean treatment effects, many of these distributional parameters that are argued to convey policy-relevant information cannot be obtained from RCT data alone. Instead, *additional assumptions* are required for this (Bedoya et al. 2017, p. 6).

Since these additional assumptions figure prominently in the discussion in this thesis, let us briefly discuss what these additional assumptions are and why they are necessary to obtain distributional parameters from RCT data. For this, recall that in an RCT, we can only observe an experimental subject in *either* the state of treatment or the state of control. Thus, as I discuss in more detail later, pure RCT data only provides us with the two *marginal distributions of outcomes* (short: Marginal distributions): One for the control and one for the treatment group (Heckman and Smith 1995, p. 96).

However, from pure experimental data, we do not learn the outcomes for any experimental subject in *both* the state of treatment and the state of control. For obtaining these outcomes, the additional assumptions in the focus of this thesis are necessary. That is, only if we supplement the minimal RCT assumptions that allow for the identification of mean treatment effects with additional assumptions about the relationship between the outcomes in the state of treatment and control, we obtain the outcomes for all experimental subjects in both states. For instance, we could assume that the policy has the same effect on every experimental subject. Then, we could recover the outcome in the case of treatment for a subject in the control group by adding this constant effect to her control outcome, which we observe, and vice versa. By obtaining the outcomes for all experimental subjects for both states, we obtain the *joint distribution of outcomes* (short: Joint distribution), which we will discuss more in-depth later (Heckman et al. 1997, p. 488).

The joint distribution is necessary for obtaining many of the distributional parameters argued to be of interest to policymakers from RCT data, such as the proportion of people benefitting from a policy (cf. Bedoya et al. 2017, p. 6). Thus, additional assumptions that go beyond the minimal RCT assumptions that allow for the identification of mean treatment effects are necessary to obtain the policy-relevant distributional information conveyed by these parameters from RCT data.

From the above, the following argument emerges, which I refer to as the “Mean Treatment Effects Argument” (in the following: MTE argument):

> ***P1:*** Mean treatment effects, which are obtainable from RCT data based on minimal assumptions, are insufficiently informative for policymaking.
>
> ***P2:*** Distributional parameters obtained from RCT data convey additional, policy-relevant information.
>
> ***P3:*** The joint distribution is needed for obtaining policy-relevant distributional information conveyed by distributional parameters from RCT data.
>
> ***P4:*** To obtain the policy-relevant distributional information conveyed by distributional parameters that depend on the joint distribution from RCT data, (1) additional assumptions that (2) go beyond the minimal RCT assumptions are necessary.
>
> ***C:*** Therefore, for obtaining (some types of) policy-relevant evidence from RCT data, we need to move beyond the minimal assumptions that allow for the identification of mean treatment effects.

Reconstructed as above, the MTE argument mirrors an argument brought forward by Angus Deaton (Deaton 2010, pp. 439f.), even though Deaton does not use the term "MTE argument". More implicit endorsements of the MTE argument in this form can be found in Heckman 2020 (Heckman 2020, p. 9 and pp. 16ff.) or Heckman and Smith 1995 (Heckman and Smith 1995, p. 87ff. and p. 96). Other authors, such as Subramanian et al. (Subramanian et al. 2018, pp. 1ff.) or Djebbari and Smith (Djebbari and Smith 2008, p. 65), focus on defending certain premises of the MTE argument, usually the first or the second.

Research Question and Relevance

In this thesis, I discuss the persuasiveness of the MTE argument just sketched. In Chapter 2, I establish the relevance of such a discussion for the philosophical discussion on RCTs and their role in the EBP movement. To this end, I argue that the MTE argument, if shown to be persuasive, could support the philosophical case against the special role of RCTs for EBP. As I discuss in Chapter 1, this case needs support because the external validity critique is not well accepted beyond the philosophical literature. From the discussion in Chapters 1 and 2, the following research question that I answer in the remainder of the thesis emerges:

> *Research Question: Is the MTE argument a persuasive argument that can support the philosophical case against the special role of RCTs for EBP?*

Argument

In my discussion of this research question, I focus on the *third and fourth premise* of the MTE argument. This focus is due to two reasons: Firstly, as we shall see, these two premises crucially determine the persuasiveness of the MTE argument as an argument in support of the philosophical case against the special role of RCTs for EBP. Secondly, methodological advancements in the RCT literature, such as quantile treatment effects or bounding, shed doubts on the claims made by these two premises. Defending the third and the fourth premise of the MTE argument in light of these methodological advancements, I argue that:

> (i) Distributional information that can only be obtained from RCT data by obtaining the joint distribution *(Chapter 3)* is of relevance to policymakers (*Chapter 4*).
>
> (ii) Additional assumptions are necessary to obtain this policy-relevant information from RCT data in a way that is useful for policymakers *(Chapter 5*).
>
> (iii) The additional assumptions which are necessary for obtaining this policy-relevant information go beyond what EBP proponents think counts as a minimal assumption *(Chapter 6).*

By arguing for (i)-(iii), I defend the persuasiveness of the MTE argument as an argument in support of the philosophical case against the special role of RCTs for EBP.

Chapter Summaries

To establish this conclusion, I proceed as follows:

Chapter 1: The first chapter provides the basis for establishing the relevance of a discussion of the persuasiveness of the MTE argument for the philosophical discussion on the special role of RCTs for EBP. To this end, I first introduce the EBP movement and their argument for ascribing a special role to RCTs. As I discuss, the minimality of assumptions that RCTs are argued to require figures prominently into this argument. I subsequently present the philosophical external validity critique, which has focused on this alleged minimality of assumptions. I then establish that the external validity critique needs support because it is not well accepted beyond the philosophical discussion.

Chapter 2: In the second chapter, I argue that the MTE argument – if shown to be persuasive – could provide the needed support for the philosophical case against the special role of RCTs for EBP. By showing this, I establish the relevance of the discussion in the remainder of the thesis. To this end, I introduce the MTE argument in-depth and point to decisive parallels between the external validity and the MTE argument: Both arguments focus on the minimal assumptions that RCTs are argued to require for producing evidence for policymaking. I conclude the chapter by pointing out that the persuasiveness of the MTE argument hinges on establishing its third and fourth premise, which are endangered by methodological advancements in the RCT literature.

Chapter 3: In this chapter, I lay the foundation for defending the third premise of the MTE argument. As I discuss, this premise is threatened: The host of distributional information that can be obtained from RCT data solely based on the marginal distributions renders the necessity of distributional information solely obtainable from the joint distribution doubtful. I discuss subgroup means and quantile treatment effects as the primary examples of distributional parameters that solely depend on the marginal distributions. I then point to distributional information that cannot be obtained from these two distributional parameters, such as the effect of a policy on individuals at given quantiles of the initial outcome distribution.

Chapter 4: In this chapter, I argue that distributional information solely obtainable from distributional parameters that depend on the joint distribution is relevant for policymakers. I do so by employing an argument from the economic literature, according to which this distributional information is more nuanced. I then spell out one way to defend the policy-relevance of this more nuanced information. According to this defence, nuanced distributional information on the effect of a policy on individuals at given quantiles of the initial outcome distribution is relevant for all policymakers, assuming that policymakers care about the electorate's support. Thereby, the third premise of the MTE argument can be defended.

Chapter 5: In this chapter, I turn to the fourth premise of the MTE argument. To this end, I first discuss (1) whether we need additional assumptions to obtain the joint distribution, and thus distributional information of relevance to policymakers, from RCT data. I defend assumption-based approaches to obtaining the joint distribution against bounding as a possibility to learn about distributional parameters that depend on the joint distribution without making additional assumptions.

Chapter 6: In this final chapter, I discuss part (2) of the fourth premise. According to (2), the assumptions that allow us to obtain the joint distribution and thus policy-relevant distributional information go beyond the minimal RCT assumptions. I establish that all the assumptions that provide policy-makers with the distributional information they require are strong and thus go beyond the minimal RCT assumptions. To this end, I use criteria for what counts as a weak and thus minimal assumption employed by EBP proponents themselves.

All this considered, we can say that proponents of the MTE argument are correct in claiming that (i) the joint distribution and thus (ii) additional assumptions that (iii) go beyond the minimal RCT assumptions are necessary for obtaining policy-relevant distributional information from RCT data. Having defended the third and the fourth premise of the MTE argument in light of methodological advancements in the RCT literature, I conclude that the MTE argument is persuasive and can thus strengthen the philosophical case against the special role of RCTs for EBP.

Chapter 1

The Philosophical Case Against the Special Role of RCTs for EBP Needs Support

Introduction

In this first chapter, I introduce the EBP movement and its influence on policymaking nowadays. I shall do so in *section one.* In the *second section*, I discuss the important role that RCTs play in the EBP movement and present the "RCT argument" used to defend this special status of RCTs. The RCT argument claims that RCTs generate reliable estimates of causal effects and only require minimal substantive assumptions.

In the *third section*, I present the external validity argument that philosophers have brought forward to attack this defence of the special status of RCTs for the EBP movement. However, as I discuss in *section four,* this philosophical case against the EBP movements' defence for ascribing a special role to RCTs needs support. This need arises because the central claim behind the external validity critique, i.e. that evidence from methods other than RCTs is required for establishing the external validity of RCT results, is not well accepted beyond the philosophical literature. This will motivate the consideration of the MTE argument in the next chapter.

1. *The EBP Movement*

The EBP movement defends the view that policymaking should be based on rigorous scientific evidence. According to EBP proponents, policymaking has relied on expert opinion and "dogmas" too much in the past. Doing so turned out to provide unreliable guidance for policymaking. This, in turn, prevented much-needed progress in key policy areas. To reverse this negative trend, EBP advocates urge policymakers to focus on "what works" and to base their policy decisions on evidence regarding policy effectiveness. This new approach to policymaking has been promoted successfully by the EBP movement since the early 2000s. Especially in the US and the UK, but also in other countries, their ideas are already put into practice (Reiss 2013, pp. 197f.; Sanderson 2003, pp. 332ff.).

A policy that illustrates how evidence-based policy works in practice is the "No Child Left Behind" Act which was introduced in the US in 2001:

> The recent enactment of *No Child Left Behind,* and its central principle that federal funds should support educational activities backed up by "scientific research", offers an opportunity to bring rapid, evidence-driven progress – for the first time – to U.S. elementary and secondary education. Education is a field in which a vast number of interventions […] have gone in or out

> of fashion over time with little regard to rigorous evidence. As a result, over the past 30 years the United States has made almost no progress in raising the achievement of elementary and secondary school students, […] despite a 90 percent increase in real public spending per student. Our nation's extraordinary inability to raise educational achievement stands in stark contrast to our remarkable progress in improving human health over the same time period – progress which […] is largely the result of evidence-based government policies in the field of medicine (CEBP 2002, p. iii).

This quote nicely illustrates the disappointment of EBP proponents with past approaches to policymaking and the progress they expect from their approach.

As already became clear, EBP proponents do not judge all types of evidence to be "rigorous" enough to inform policymaking reliably. Expert opinion, for instance, would not fall into this category. When describing what types of evidence they have in mind for reliably informing policymaking, EBP advocates focus on the *methods* used to generate the evidence. To their mind, some methods produce better-suited evidence for policymaking than others, independently of the policy context (Reiss 2013, p. 200).

These judgements led the movement to formulate so-called evidence hierarchies that rank evidence produced by different methods according to their potential to inform policymaking reliably. That is, policy predictions based on types of evidence that rank high in the evidence hierarchy are judged to be more reliable than policy predictions based on low-ranked evidence (Cartwright and Hardie 2012, p. 136; Cartwright 2012, p. 303).

A classic example for an evidence hierarchy looks as follows (NICE 2006, p. 47, discussed in Cartwright 2012, p. 303 and Clarke et al. 2014, p. 340):

1 ++	High quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1 +	Well conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias
1 -	Meta-analyses, systematic reviews, or RCTs with a high risk of bias
2 ++	High quality systematic reviews of case control or cohort studies
	High-quality case-control or cohort studies with a very low risk of confounding, bias or chance and a high probability that the relationship is causal
2 +	Well-conducted case-control or cohort studies with a low risk of confounding, bias or chance and a moderate probability that the relationship is causal
2 -	Case-control or cohort studies with a high risk of confounding, bias, or chance and a significant risk that the relationship is not causal
3	Non-analytic studies (for example, case reports, case series)
4	Expert opinion, formal consensus

Table 1: Example for an Evidence Hierarchy

Based on these evidence hierarchies, EBP proponents have launched so-called "warehouses" for policies that have been shown to work based on evidence that ranks high in the evidence hierarchies. The idea of these warehouses is that policymakers interested in bringing about a particular policy outcome can pick the policy supported by the strongest evidence (Cartwright and Hardie 2012, p. 136).

From the example of an evidence hierarchy above, the special role that RCTs play for the EBP movement becomes apparent: According to EBP proponents, RCTs are the preferred method for generating rigorous evidence for policy-making purposes. This means, in turn, that the policies that can be found in the EBP movement's warehouses have usually been shown to "work" by RCTs. Via this channel, RCT results thus heavily influence the policy-making decisions of policymakers who adopt the EBP vision (*Ibid.*, p. 122; Cowen 2019, p. 4f.; Reiss 2013, p. 198ff.).

What justifies this privileged role of RCTs that led scholars to refer to RCTs as the EBP movement's "gold standard" (Cartwright 2012, p. 298) for generating evidence for policymaking? This shall be the topic of the next section. There, I introduce what I call the "RCT argument," which EBP proponents employ to defend the special role that they ascribe to RCTs.

2. How EBP Proponents Defend the Special Role for RCTs

The "RCT argument" that EBP proponents bring forward when defending the privileged role they ascribe to RCTs when it comes to producing evidence for policymaking is based on two propositions[1] (cf. Reiss 2013, pp. 201f.).

Firstly, it is argued that RCTs are a *highly reliable source of evidence for causal claims.* Evidence for causal claims, in turn, is argued to be crucial for policymaking. Being confident in whether a particular policy "works", or is effective in producing the desired outcome, requires evidence on whether the policy in question has a causal relation to the desired outcome. Thus, the most reliable source for evidence about causal claims – RCTs according to EBP proponents – deserves the top place in evidence hierarchies that are meant to inform policymaking (*Ibid.;* DoE 2003, pp. 1ff.; Pearce et al. 2014, p. 388 and p. 396).

Secondly, it is argued that RCTs only require *"minimal substantive assumptions"* (Deaton and Cartwright 2018a, p. 2). This has been judged to be a significant advantage of RCTs over other econometric methods, such as regression analysis. These methods rely on a lot of background assumptions which are often strong and thus appear incredible. On the other hand, RCTs are argued to produce quantities of interest based on a very limited number of general, more credible assumptions. I discuss these assumptions in more detail below, but an example would be that randomization has been done properly (Banerjee and Duflo 2008, p. 26; Deaton 2010, p. 438; DoE 2003, p. 2; Muller 2019, p. 1; Reiss 2013; p. 202).

In the following, let us examine how these two propositions that ground the EBP movement's case that policymaking should ideally be based on evidence from RCTs can be defended.

2.1. RCTs are a Highly Reliable Source of Evidence for Causal Claims

Let us start with the first proposition of the EBP proponents' RCT argument. This proposition states that RCTs are a highly reliable source of evidence for causal claims. To understand the details behind the claim reflected in this proposition, let us have a closer look at how RCTs work.

[1] In the literature, one can also find other propositions figuring into the defence of a special status for RCTs in the production of evidence for policymaking. For instance, it is often argued that RCTs provide the researcher with the best control over the assignment mechanism (cf. Imbens 2010, p. 408). However, most of these propositions boil down to the two propositions discussed here: For instance, the control over the assignment mechanism is the reason for the reliability of the causal estimates that RCTs provide.

As we saw, RCTs aim at identifying the causal effect of a policy by *randomly* allocating experimental subjects into two groups. The subjects allocated to the *treatment group* are exposed to the policy of interest while the subjects in the *control group* are not. Thereby, RCTs allow for the identification of causal effects despite what is called the *fundamental problem of causal inference* (Heckman and Smith 1995, p. 87; Imbens and Wooldrige 2009, p. 6).

The fundamental problem of causal inference arises because we can never observe an experimental subject both in the state of treatment and the state of control simultaneously. This is a problem because it prevents us from identifying the causal effect of a policy by simply comparing the outcome of each subject in the state of treatment and the state of control (Heckman and Smith 1995, p. 87).

For instance, imagine that we are interested in the effect of attending a private university, in contrast to a public university, on later earnings in the context of the US[2]. To obtain this effect, we would ideally want to compare the later earnings of each university graduate for the case in which she has attended a private and in which she has attended a public university. This, however, is impossible because every student can only attend *either* a private or a public university at the same time.

In the context of RCTs, the fundamental problem of causal inference is circumvented. This is possible because randomization makes it likely that the control group is a suitable *counterfactual* for the treatment group. That is, the control group indicates what would have happened to the treatment group had it not been exposed to the policy of interest. For instance, in an RCT in which private college attendance is randomly determined, the control group would indicate what wages the treatment group would have gotten had it not attended a private but a public university (*Ibid.*, p. 87f.; Reiss 2013, p. 201). In other words, the random assignment of experimental subjects to treatment and control group makes it likely that the exposure to the policy of interest is the only systematic difference between treatment and control group.

Given this, we can see how the fundamental problem of causal inference is circumvented in RCTs: If control and treatment group only differ systematically regarding the exposure to the policy of interest, we can attribute the difference in average outcomes between treatment and control group to the policy of interest. Then, the average causal effect of a policy can be

[2] See Dale and Krueger 2002 for a discussion of this example.

obtained by comparing the average outcome of the control group with that of the treatment group (Duflo and Kremer 2003, p. 4).

This contrasts RCTs with other study designs, which are often plagued with what is called "selection bias". In the absence of randomized assignment to treatment and control group, it will, at least partly, depend on the choices of individuals whether they get exposed to the policy. This "self-selection" leads to differences between control and treatment group on top of the fact that one group is exposed to the policy and the other is not (Imbens 2010, p. 404).

For instance, if one would compare the wages of students who attended private universities with that of students who attended public universities, one might conclude that having attended a private university has a positive causal effect on wages. This conclusion, however, might be mistaken since having attended a private university is not the only systematic difference between graduates of private and public universities. Instead, characteristics such as a high motivation probably led students to attend private universities (i.e. to "select into treatment").

In this example, the motivation of the private university graduates would thus be an additional difference between control and treatment group that arises from self-selection into the treatment. This difference – which is difficult to control for and determines the outcome variable of interest – prevents us from identifying the average causal effect of attending a private university by comparing the average wages of private university graduates with that of public university graduates. This is because this comparison would not only reflect the effect of having attended a private university but also of being more motivated, i.e. the selection bias (Angrist and Pischke 2014, pp. 47ff.).

Disentangling the "real" effect from the selection bias is impossible, making the causal estimates resulting from studies like the one sketched above hardly reliable. RCT results, on the other hand, are argued not to be plagued with selection bias. This is because the random assignment of the treatment makes it likely that the differences between the two groups that induced selection bias in the example above are balanced out[3]. For instance, in the case of

[3] This is, however, *not guaranteed* by randomization. In finite samples, treatment and control group might still differ regarding characteristics other than the exposure to the policy by pure chance. Only in the limit, properly conducted randomization guarantees that there are no systematic differences between treatment and control group apart from the exposure to the treatment (Deaton and Cartwright 2018a, p. 5).

random assignment to private universities, private and public university attendees are likely similar regarding motivation (Athey and Imbens 2018, p. 87; Heckman and Smith 1995, p. 89).

Therefore, RCTs are argued to produce estimates of average causal effects without being plagued by selection bias[4]. This claim forms the justification for the first proposition of the RCT argument, i.e. that RCTs are highly reliable sources of evidence for causal claims. Let us now turn to the second proposition, which states that RCTs only require minimal substantive assumptions for producing quantities of interest for policymaking.

2.2. RCTs Only Require Minimal Substantive Assumptions

To understand the idea behind the claim made by the second proposition of the RCT argument, let us look at the assumptions required for identifying the mean treatment effect as one quantity of interest[5]. For obtaining the mean treatment effect from RCT data, average outcomes in treatment and control group are compared. Introducing some notation, we can write this as follows:

$$E(Y_i\,|T_i = 1) - E(Y_i|T_i = 0)$$

In this formula, $E[Y_i]$ denotes the expected value (or population average) of the variable Y_i. In turn, $E[Y_i|T_i = 1]$ denotes the average value of Y_i for the part of the population for which $T_i = 1$ (Angrist and Pischke 2015, pp. 19f.; Heckman and Smith 1995, p. 87).

As outlined, the term for the mean treatment effect contains both the "real" average causal effect and a potential selection bias. To see this, we can decompose this expression as follows:

$$\begin{aligned} &E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 0) \\ &= [E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 1)] \\ &+ [E(Y_{i0}|T_i = 1) - E(Y_{i0}|T_i = 0)] \end{aligned}$$

In this decomposed expression, the second expression in square brackets is the selection bias. As we have seen, letting exposure to the policy be determined by random assignment makes it likely that the selection bias term equals zero, leaving only the first term in square brackets. This term, in turn, can be written as the average treatment effect on the treated:

---

[4] This argument hinges on the empirical importance of selection bias (Worrall 2010, p. 292) which is a topic of debate in the literature (see e. g. *Ibid.,* pp. 292f. or Heckman and Smith 1995, p. 90f.).

[5] The discussion in this section follows Deaton 2010, p. 439.

$$E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 1)$$
$$= E(Y_{i1} - Y_{i0}|T_i = 1)$$

This last step works because the mean is a linear operator. Therefore, the difference in averages equals the average of the difference. Thus, we have arrived at the average treatment effect on the treated by comparing the average outcomes of treatment and control group. For this, we only had to assume random assignment of the treatment and that the mean is a linear operator. The average treatment effect on the treated, in turn, can be seen as the average effect for all given that treatment and control group only differ by randomization (Deaton 2010, p. 439; Deaton and Cartwright 2018a, p. 4).

This grounds the case for the second proposition of the RCT argument brought forward by EBP proponents in favour of the "gold standard" role of RCTs[6]. Now that we have looked at the defence of EBP proponents for RCTs as the privileged source of evidence for policymaking, let us discuss why this defence has been seen critically in the philosophical literature.

3. *The Philosophical Debate About the EBP Movements' Defence of the Special Role of RCTs*

As already mentioned, the philosophical discussion on the EBP movement has mostly evolved around concerns regarding the *external validity* of RCT results. This section, firstly, introduces these philosophical concerns. Secondly, I discuss why these concerns provide an argument against the EBP movements' defence of the special role of RCTs discussed above.

3.1. Philosophical Concerns About External Validity

The influx of RCTs into policymaking via the EBP movement has raised many points of criticism among philosophers. One of the most frequently discussed points of critique concerns the external validity of RCT results (e.g. Clarke et al. 2014, pp. 346ff.; Favereau and Nagatsu 2020, p. 192; Teira and Reiss 2013, p. 211).

[6] In practice, more assumptions are required when calculating quantities like the average treatment effect: For instance, it must be assumed that experimental subjects comply with their assignment to treatment or control group. Confronted with this, RCT proponents often point to the many ways in which we can deal with these practical problems: For instance, the assumption of perfect compliance to the treatment assignment is argued not to be problematic, since the literature has developed ways to obtain the average causal effect in the presence of non-compliance (Duflo and Kremer 2003, pp. 21f.; Imbens 2010, p. 408).

In the context of policymaking, external validity means that a policy that has been shown – e. g. by an RCT – to work in one context is also likely to yield the desired effect in another context of interest (Clarke et al. 2014, pp. 346f.). How confident we can be in the result of the initial RCT for the context that it was conducted in is, in contrast, a question of internal validity (Imbens 2018, p. 52).

The issue that many philosophers have emphasized regarding the external validity of RCT results is that the causal evidence from an – internally valid – RCT is, on its own, not enough to establish the external validity of the effect found in the RCT. Instead, it is argued, additional evidence from sources other than RCTs is needed for establishing this (Cartwright and Hardie 2012, p. 8; Deaton and Cartwright 2018a, p. 14; Worrall 2010, pp. 295f.).

Let us make this critique more concrete before establishing how it endangers the RCT argument of EBP proponents discussed above. For this, let us discuss what kind of evidence philosophers deem essential for establishing the external validity of an RCT result and why they argue that this evidence cannot be obtained from RCTs alone.

As philosophers have frequently emphasized, understanding *why* and *how* a policy measure produces the effect found in an RCT is essential for establishing the external validity of an RCT result (Cartwright and Hardie 2012, p. 8; Favereau and Nagatsu 2020, p. 194; Muller 2019, p. 1).

To understand why this is important, consider the following part of the answer to the question of why and how a policy measure produces the desired effect: In many cases, philosophers argue, a policy measure only yields the desired outcome in conjunction with the presence of certain *support factors.* For instance, a policy measure that reduces class sizes might only improve student outcomes in the presence of support factors such as qualified teachers and enough suitable facilities (Cartwright and Hardie 2012, pp. 61ff.).

Without these support factors, the policy of interest will not produce the desired effect. Therefore, knowing the relevant support factors and knowing whether they are present in the policy context of interest is essential for establishing an RCT result's external validity. For instance, arguing that the result of the class-size reduction RCT is likely to apply to other policy contexts requires the following evidence: Evidence on qualified teachers and suitable facilities being relevant support factors and evidence on the presence of these two factors in the policy context of interest (Deaton and Cartwright 2018a, p. 12).

Importantly, philosophers argue that evidence regarding how and why a policy intervention produces the desired effect *cannot be obtained from RCTs alone.* Instead, evidence from other sources and methods is necessary for obtaining this evidence and thus for establishing the external validity of an RCT result (Cartwright 2007, p. 16).

For illustrating this, re-consider the example of support factors and class-size reductions. From an RCT that has established that reducing class sizes indeed improves test scores, we only learn that the required support factors were in place in the setting in which the RCT was conducted. However, we do not learn what the relevant support factors are and whether they are present in other policy contexts of interest. For establishing this, philosophers argue that we need theory or structured thinking to identify support factors. For finding out about their presence in the context of interest, observational studies are claimed to be helpful (Cartwright and Hardie 2012, pp. 124ff.).

In sum, the philosophical external validity argument thus rests on the assertion that establishing the external validity of RCT results requires (1) evidence on how and why a policy intervention produces an effect and that this evidence (2) cannot be obtained from RCTs alone.

3.2. Attacking the EBP Movements' Defence of the Special Role of RCTs

How does this philosophical argument criticize the EBP proponents' case for RCTs as the privileged source of evidence for policymaking? As discussed, this case rests on claiming that (1) RCTs provide reliable estimates for causal claims and (2) only require minimal substantive assumptions. The philosophical external validity argument attacks the second proposition.

That is, it is argued that for knowing whether the reliable causal estimates obtained from RCTs generalize to other policy contexts, evidence from sources or methods other than RCTs is necessary. These other methods, however, rest on different and usually more substantive assumptions than RCTs do themselves. Thus, for collecting the additional evidence required for establishing the external validity of RCT results, it is necessary to go beyond the minimal RCT assumptions on which the RCT argument of EBP proponents rests. Therefore, the second proposition of the EBP proponents' RCT argument regarding the minimality of RCT assumptions does not hold up when RCT results' external validity is concerned (Deaton and Cartwright 2018a, p. 3; Muller 2019).

Establishing that initial RCT results are externally valid is, in turn, argued to be essential for the aim of the EBP movement to base policymaking solely on evidence that ideally comes from RCTs. If the results of – internally valid – RCTs do not generalize to other contexts, basing policy decisions on evidence from RCTs only makes sense if the RCT has been conducted in the context of interest. This threatens core ideas behind EBP practices, such as telling policymakers to look for policies that have been shown to be effective in other contexts in the EBP proponents' warehouses (Deaton and Cartwright 2018a, p. 11).

To summarize, the philosophical external validity critique questions the EBP movements' defence for ascribing a special status to RCTs. It does so by pointing out that assumptions beyond the minimal RCT assumptions are required to establish the external validity of RCT results, which is argued to be essential for policymaking. By attacking the RCT argument made by EBP proponents to defend the special role they ascribe to RCTs, the external validity argument questions this special role[7]. Thereby, this argument made many philosophers sceptical regarding the placement of RCTs at the top of the evidence hierarchies of EBP proponents (Cartwright 2012, p. 299; Clarke et al. 2014, p. 340; Stegenga 2013, pp. 318f.).

4. *The Philosophical Case Against the Special Role of RCTs for EBP Needs Support*

While philosophers see the external validity argument as a reason to doubt the special status of RCTs for EBP, this argument is not widely acknowledged as a reason to reject the privileged role of RCTs for EBP beyond the philosophical discussion. In this section, I argue that this means that the philosophical case against the special role of RCTs for EBP, which mainly rests on the external validity argument, needs support.

While the importance of establishing the external validity of RCT results is widely accepted, many authors, especially within the economic literature, have argued against the claim that evidence from methods other than RCTs is necessary for this. Instead, it is argued that evidence obtainable *from RCTs alone* is sufficient for establishing the external validity of RCT

[7] In the remainder of the thesis, I shall thus, for brevity reasons, frequently refer to the external validity argument as an argument that questions the special role of RCTs for the EBP movement. However, it would be more accurate to say that the external validity argument questions whether this role can be defended by pointing to the reliable causal estimates that RCTs produce and the minimal assumptions that RCTs require.

results. For instance, authors such as Athey and Imbens have argued that the external validity of initial RCT results can be established by replicating the result in additional RCTs conducted in different settings (Athey and Imbens 2018, p. 80). Or, it has been proposed to estimate the effect of a policy for a different setting based on the initial RCT result by calculating subgroup effects based on the initial RCT data and reweighting these according to the composition of the population of interest (Deaton and Cartwright 2018a, p. 13).

This discussion shows that the essential claim of the philosophical external validity critique – that the external validity of RCT results can only be established with evidence from other methods which rely on additional, substantive assumptions – is contested. However, this claim is necessary for using the external validity critique as an argument against the special role of RCTs for EBP. That is, without evidence from methods other than RCTs being necessary for establishing the external validity of RCT results, no assumptions beyond the minimal RCT assumptions are required for producing useful evidence for policymaking. Then, the external validity critique cannot make a case against the special role of RCTs for EBP by questioning the minimality of assumptions that RCTs require. Thus, the philosophical case against the special role of RCTs for EBP, which mainly relies on the external validity critique, needs support.

Conclusion

In this chapter, I have introduced the EBP movement and the special role that RCTs play in this movement. As discussed, this special role is grounded in what I call the "RCT argument". This argument defends the special status of RCTs by pointing to the reliable causal estimates that RCTs produce and the minimal assumptions that RCTs require. However, we have seen that the RCT argument is questioned in the philosophical literature, mainly based on concerns about external validity.

The philosophical external validity argument attacks the RCT argument by arguing that RCT results must hold up in other policy contexts of interest for using evidence from RCTs for policymaking. For establishing this, it is argued, evidence from methods other than RCTs is needed. These methods, in turn, rely on assumptions beyond the minimal RCT assumptions.

However, we have also seen that the philosophical case against the special role of RCTs for EBP needs support. This is because the essential premise of the external validity argument – i.e. that evidence from methods other than RCTs is necessary for establishing the external validity of RCT results - is not widely accepted beyond the philosophical literature.

In the next chapter, I point to an argument that has not yet received much attention from philosophers but could, as I shall argue, fill the gap created by the absence of acceptance of the external validity argument: The MTE argument.

Chapter 2

The MTE Argument Can Provide the Needed Support

Introduction

In this chapter, I argue that the MTE argument can provide the needed support for the philosophical case against the special role of RCTs for EBP if it can be shown to be persuasive. To this end, I first present the MTE argument in more detail in sections *one, two, three and four*.

I then analyse parallels in the argumentative strategy of the external validity and the MTE argument in *section five.* As I show, both arguments point to assumptions beyond the minimal RCT assumptions, which are argued to be necessary for obtaining policy-relevant evidence from RCT data. Based on this, I *conclude* that the MTE argument can provide the required support for the philosophical case against the special role of RCTs for EBP if it can be shown to be persuasive. As I outline, this persuasiveness hinges on establishing the third and fourth premise of the MTE argument, which I turn to in the remainder of the thesis.

1. The Structure of the MTE Argument

As discussed, the MTE argument can be depicted as follows:

> ***P1:*** Mean treatment effects, which are obtainable from RCT data based on minimal assumptions, are insufficiently informative for policymaking.
>
> ***P2:*** Distributional parameters obtained from RCT data convey additional, policy-relevant information.
>
> ***P3:*** The joint distribution is needed for obtaining policy-relevant distributional information conveyed by distributional parameters from RCT data.
>
> ***P4***: To obtain policy-relevant distributional information conveyed by distributional parameters that depend on the joint distribution from RCT data, (1) additional assumptions that (2) go beyond the minimal RCT assumptions are necessary.
>
> ***C:*** Therefore, for obtaining (some types of) policy-relevant evidence from RCT data, we need to move beyond the minimal assumptions that allow for the identification of mean treatment effects.

In the following sections, I provide a detailed introduction of the MTE argument by discussing its four premises, starting with the first one.

2. The First Premise of the MTE Argument

By stating that mean treatment effects obtained from RCT data are insufficiently informative for policy-making purposes, the first premise of the MTE argument reflects frequently voiced criticisms regarding the emphasis on mean treatment effects in the analysis of RCT data.

These criticisms, which are often voiced in the economic literature, aim at the practice of thinking of mean treatment effects as the main parameter of interest in the context of RCTs. And indeed, as we saw, the mean treatment effect is usually reported as the main result in the analysis of RCT data (Bedoya et al. 2017, p. 1; Imbens and Wooldrige 2009, p. 15).

The rise of the EBP movement has strengthened the focus on mean treatment effects. As discussed, the EBP movement focuses on "what works" in policymaking. That is, the focus of the EBP movement is on evidence for policy effectiveness. And of course, for providing evidence that a policy measure was effective in producing a given outcome in the studied situation, mean treatment effects obtained from RCT data are sufficient (Heckman 2020, p. 33; Na et al. 2015, FN 2).

To attack the emphasis of mean treatment effects in the analysis of RCT data, the criticisms reflected in the first premise of the MTE argument start from concerns about *heterogeneity in treatment effects.* That is, the individual-level effects of the policies studied in RCTs often vary substantially within the studied population[8]. For instance, it is likely that not all students in a trial school profit the same way from a class-size reduction measure: For some, the increased attention from their teachers might make all the difference. Other students might already have performed well without more attention (Banerjee and Duflo 2008, p. 13; Na et al. 2015, p. 291; Subramanian et al. 2018, p. 1).

Why does treatment effect heterogeneity render mean treatment effects insufficiently informative for policymaking, as the first premise of the MTE argument suggests? In the literature, two theoretical arguments for why that is can be found: Firstly, it is argued that mean treatment effects might apply to *no one or a limited number of individuals* in the studied population if there is treatment effect heterogeneity. Secondly, it is argued that mean treatment effects might suggest *false conclusions* regarding the desirability of a policy in the presence of

[8] For empirical evidence regarding treatment effect heterogeneity see e.g. Bitler et al. 2003 or Djebbari and Smith 2008.

treatment effect heterogeneity. To grasp these arguments and thus the support for the claim reflected by the first premise of the MTE argument, consider the following example:

Imagine an RCT conducted in different schools in a country, aiming to determine whether smaller class-sizes improve student test scores[9]. Also, imagine that the mean treatment effect obtained from the RCT is zero, and thus indicates that the policy is ineffective (discussed in Deaton and Cartwright 2018a, p. 16 and Subramanian et al. 2018, p. 1).

Does this result mean that a specific school among the schools studied in the RCT should not reduce class-sizes? According to some authors, this is not the correct conclusion in the presence of heterogeneity in policy responses. As they argue, heterogeneity in policy responses might mean that the calculated mean treatment effect for the population applies to none or very little of the individuals in the population (Subramanian et al. 2018, pp. 2f.).

Thus, while the mean treatment effect in our class-size example equals zero, heterogeneity in policy responses might mean that none of the schools experienced an effect of zero. Instead, the mean of zero might reflect that some schools benefitted a lot from the policy while other schools got harmed by the intervention. Unless we know about the underlying heterogeneity in policy responses in the population, there is no way for us to distinguish between the two stories (Deaton 2010, p. 439; Na et al. 2015, p. 290).

For policymakers, it seems problematic that the average policy response reflected by mean treatment effects might apply to no one or a limited number of individuals in their population. After all, policymakers want to introduce policies that benefit individuals. However, if the averages obtained from RCTs do not apply to any of the individuals the policymaker is interested in, then RCT results in the form of means are insufficiently informative for policymakers (Deaton 2010, p. 441).

This problem implies another argument for why mean treatment effects obtained from RCTs are insufficiently informative for policymakers, as suggested by the first premise of the MTE argument. That is, mean treatment effects might give a wrong impression regarding the desirability of a policy measure.

[9] In this case, the experimental population is also the target population of the policy. This is useful for separating external validity concerns and concerns about the informativeness of mean treatment effects (see section 5).

For instance, imagine that our class-size reduction RCT revealed a positive mean treatment effect. Such a result seems to indicate that the policy in question is desirable. However, it could be that this result merely reflects that a few schools benefited very much from the smaller classes while the students in most schools experienced a slight decline in test scores. This information would shed another light on the desirability of the policy. However, as we saw, it cannot be obtained from mean treatment effects (Deaton 2010, pp. 439ff.; Bitler et al. 2003, p. 1).

In sum, the first premise of the MTE argument reflects frequently voiced concerns about the focus on mean treatment effects in the analysis of RCT data for policy-making purposes. Two theoretical arguments back up these concerns. According to these arguments, heterogeneity in treatment effects might mean that mean effects apply to none or very little of the individuals in the population of interest or might indicate false conclusions about the desirability of a policy.

Having introduced the first premise of the MTE argument and its support, let us now turn to its second premise, which holds that distributional parameters convey additional policy-relevant information.

3. The Second Premise of the MTE Argument

Given the insufficient informativeness of mean treatment effects for policymaking, economists have come up with several parameters that are argued to convey additional policy-relevant information in the presence of heterogeneity in policy responses (Heckman et al. 1999, p. 1868; Na et al. 2015, p. 292).

These parameters include metrics like the *proportion of people who benefitted or got harmed by an intervention* (Carneiro et al. 2003, p. 1; Na et al. 2015, p. 292), *the median treatment effect* (Deaton 2010, p. 439; Heckman and Smith 1995, p. 87), the *gains of a particular subgroup* (Bedoya et al. 2017, p. 1), *quantile treatment effects* (Bitler et al. 2003, pp. 1ff.; Jackson et al. 2013, pp. 92f.), or the *effect of the policy on the dispersion of the outcome* (Firpo 2010, p. 1). In line with Heckman (Heckman 2008, p. 20) and Bedoya et al. (Bedoya et al. 2017, p. 1), I shall refer to these metrics as “distributional parameters” and the questions that they answer as “distributional questions” in the following.

The often-voiced claim that these distributional parameters convey additional, policy-relevant information compared to the insufficient information conveyed by mean treatment effects is reflected in the second premise of the MTE argument. Importantly, this claim is also backed up by empirical results regarding the information policymakers deem relevant when making implementation decisions based on evidence. To grasp the empirical support for the second premise of the MTE argument, let us look at the results from such a study. This study was conducted by Petticrew et al., who analysed what evidence policymakers engaged in evidence-based policymaking deem relevant.

One finding that emerges from their study is that policymakers think that the evidence presented to them "often had little to say about inequalities, as it commonly reported on average rather than distributional effects" (Petticrew et al. 2004, p. 813). Based on such quotes, the authors conclude that their study identified "specific gaps, in particular the need for information on the distributional effects of interventions" (*Ibid.*, p. 814). Furthermore, information on "distributional effects of interventions was also seen as crucial, but absent" (*Ibid.*) by policymakers. In a follow-up paper, the authors describe the findings of their previous study by saying that "there was a perceived need for evaluations of the differential impacts of policies on different socioeconomic groups" (Whitehead et al. 2004, p. 817).

Based on these results, it seems plausible that policymakers indeed think that what I call "distributional parameters" convey additional, policy-relevant information compared to the insufficient information conveyed by mean treatment effects. Thus, the claim reflected by the second premise of the MTE argument is backed up by empirical evidence. Having introduced the first two premises of the MTE argument and the theoretical arguments and empirical results backing them up, let us turn to the third and the fourth premise.

4. The Third and the Fourth Premise of the MTE Argument

According to the third premise of the MTE argument, the joint distribution is needed for obtaining policy-relevant distributional information conveyed by distributional parameters from RCT data. To obtain this policy-relevant information by obtaining the joint distribution, as reflected by the fourth premise, (1) additional assumptions that (2) go beyond the minimal RCT assumptions are required. In the following, let us grasp the ideas behind these two premises, starting with the third one.

4.1. The Third Premise of the MTE Argument

For understanding the idea behind the third premise of the MTE argument, let us first recall the joint distribution, which figures prominently in this premise. As we briefly discussed in the introduction, the joint distribution gives us the outcomes of each experimental subject for both the state of treatment and the state of control. Formally, we can write this as $F(Y_0,Y_1)$. For instance, in a class-size reduction RCT, the joint distribution gives us the test scores of each experimental subject for the state in which she was allocated to a small and a large class. From knowing the outcomes of each experimental subject for both states, i.e. $F(Y_0,Y_1)$, we can also infer how the *impact* is distributed, i.e. $F(Y_1\text{-}Y_0)$[10] (Heckman and Smith 1995, p. 96; Heckman et al. 1999, p. 1879).

In contrast, the marginal distributions give us the control and treatment group outcomes, respectively. Formally, we can write this as $F(Y_0)$ and $F(Y_1)$. In the class-size reduction RCT, the marginal distributions thus give us the test scores for the individuals in the smaller classes and those in the larger classes, respectively (Heckman and Smith 1995, p. 96).

The marginal distributions are sufficient for obtaining the mean treatment effect, which can, given that the mean is a linear operator, be identified by comparing the mean outcomes in treatment and control group. These, in turn, are features of the marginal distributions. This contrasts the mean treatment effect with other parameters, such as some of the distributional parameters discussed above. For instance, comparing the median outcomes of treatment and control group in an RCT does not, in general, give us the median treatment effect. This is because these distributional parameters depend on the joint distribution, as reflected by the third premise of the MTE argument (Heckman and Smith 1995, p. 89; Imbens and Wooldrige 2009, p. 17).

4.2. The Fourth Premise of the MTE Argument

For understanding the idea behind the fourth premise of the MTE argument, let us recall the fundamental problem of causal inference discussed in the previous chapter. As we saw, this problem arises because we can only observe every experimental subject in *either* the control or the treatment group simultaneously. Therefore, for each experimental subject, we only learn the outcome for either the state of treatment or the state of control from RCT data.

[10] Since the impact distribution depends on the joint distribution, I will stick to the latter in the following.

Formally, we either learn Y_{i0} (outcome for individual i for the state of control) or Y_{i1} (outcome for individual i for the state of treatment) but never (Y_{i0}, Y_{i1}) (Heckman and Smith 1995, p. 96).

Therefore, pure RCT data only provides us with the two marginal distributions, one for the control ($F(Y_0)$) and one for the treatment group ($F(Y_1)$). However, since we cannot observe any experimental subject in both treatment and control group simultaneously and thus only know one coordinate of (Y_0, Y_1), pure experimental data does not give us the joint distribution ($F(Y_0, Y_1)$) (*Ibid.,* p. 96; Bitler et al. 2003, p. 16).

Why is it impossible to obtain the joint distribution from RCT data alone? To see this, consider the graph below (taken from Djebbari and Smith 2008, p. 69) and imagine that it was taken from a class-size reduction RCT. In the graph, you can see the two marginal distributions that we obtain from pure RCT data: The one for the control group, on the left and the one for the treatment group, on the right.

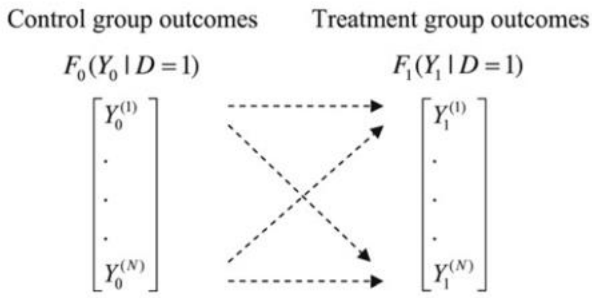

Figure 1: Illustration Marginal Distributions

The problem with obtaining the joint distribution based on pure RCT data arises as follows: From knowing the outcome of a given experimental subject in the control group, we can infer nothing about where she would have appeared in the distribution of outcomes in the treatment group. For instance, an individual in the control wing of the class-size reduction RCT might be located at the lower end of the respective marginal distribution. However, we do not know where she would appear in the marginal distribution of the treatment group. As the arrows indicate, she might, for instance, appear at the lower end of the distribution again, but she might as well also move up to the upper end of the distribution. Without knowing this, we cannot obtain the outcomes for any experimental subject in both states, which would give us the joint distribution (Heckman et al. 1997, p. 495).

For obtaining the joint distribution based on the two marginal distributions, we would need information about the relationship between the outcomes in the state of treatment and control, that is, between the two marginal distributions. This information is, however, not available from pure RCT data. Thus, we need to *assume* something about this relationship to obtain the joint distribution from RCT data (Heckman and Smith 1995, p. 89 and p. 96).

For instance, we could assume that the class-size policy had the same impact on everyone, i.e. that the two marginal distributions only differ by a constant. For example, for an individual in the control wing of the class-size reduction RCT, we could then recover her test score in the case of treatment by simply adding the constant effect to her control test score, which we observe (i.e. $Y_{i0} + C = Y_{i1}$). This would give us both coordinates of (Y_{i0}, Y_{i1}) for each experimental subject and thus the joint distribution. The necessity of additional assumptions for recovering the joint distribution grounds the claim made by the fourth premise of the MTE argument.

This concludes our discussion on the claims reflected by the third and the fourth premise of the MTE argument. Importantly, we have not yet discussed arguments or empirical evidence that support these claims. As we shall see below, this support is fading in light of methodological advancements in the RCT literature. Motivated by this fading support, I defend the third and fourth premise, and thus the persuasiveness of the MTE argument, in the following chapters. Before being able to do so, let us first establish the relevance of such a discussion for the philosophical discussion on the special role of RCTs for the EBP movement in the next section.

5. *The Relevance of the MTE Argument for the Philosophical Discussion on the EBP Movement*

Based on the preceding analysis of the MTE argument, I now point to decisive parallels between the MTE and the external validity argument. In doing so, I want to establish the relevance of a discussion on the persuasiveness of the MTE argument for the philosophical discussion on the special role of RCTs for the EBP movement.

As discussed, the external validity argument is the main argument employed by philosophers when questioning the special role of RCTs for the EBP movement. The external validity argument attacks this special role by pointing out that RCT results must hold up in other policy contexts of interest ("external validity") for using evidence from RCTs for policymaking. For establishing the external validity of an RCT result, it is argued, evidence from methods other than RCTs is needed. These methods, in turn, rely on assumptions beyond the minimal RCT assumptions.

In the following, I argue that the external validity and the MTE argument have a similar argumentative strategy. Firstly, both arguments can be read as acknowledging the first proposition of the RCT argument, that RCTs provide reliable evidence for causal claims.

Secondly, reliable evidence on policy effectiveness is argued not to be enough for making policy decisions based on evidence. In addition, RCTs would have to 1) provide evidence that the policy in question is also effective in producing the desired outcome in other contexts of interest and 2) deliver policy-relevant information conveyed by distributional parameters.

As outlined, 1) and 2) are then argued only to be possible if one is willing to go beyond the minimal substantive RCT assumptions due to which RCTs are taken to be a privileged source of evidence for policymaking. According to the external validity argument, the methods required for generating the evidence necessary for establishing the external validity of RCT results require additional, substantive assumptions. According to the MTE argument, additional, substantive assumptions are necessary for obtaining the joint distribution and, thus, policy-relevant distributional information from RCT data. Thereby, both arguments attack the second proposition of the RCT argument of EBP proponents.

However, as we saw in the last chapter, the key premise of the external validity argument that makes it possible to use this argument to question the second proposition of the RCT argument is not broadly acknowledged. Therefore, I argued, the philosophical case against the special role of RCTs for EBP, which mainly relies on the external validity argument, needs support. Given that, as we just saw, the MTE argument and the external validity argument have the same argumentative strategy, this support could come from the MTE argument. However, to lend this support, it is necessary to show that the MTE argument is more persuasive than the external validity argument has been so far.

Additionally, the MTE argument can be used to attack the special role of RCTs for EBP in cases in which the external validity argument cannot. For instance, in cases in which the trial sample is known to be a representative sample of the target population of interest, external validity concerns do not arise (Subramanian et al. 2018, p. 1). Thus, even if the external validity argument was broadly acknowledged to be persuasive, it could not be used to question the special role of RCTs for EBP in these cases. However, the problem that mean treatment effects are insufficiently informative for policymakers also arises in cases in which the trial sample is known to be a representative sample of the target population of interest. Thus, the MTE argument, if shown to be persuasive, can question the special role of RCTs for EBP in these cases, while the external validity argument cannot.

These two considerations – i.e. that the MTE argument, if shown to be persuasive, could support the philosophical case against the special role of RCTs for EBP in cases in which (a) the external validity argument holds up but is not broadly accepted and in cases in which (b) the external validity argument does not hold up – call for a philosophical discussion on the persuasiveness of the MTE argument to find out if it can fulfil these two purposes.

Conclusion

In this chapter, we have established that the MTE argument – if shown to be persuasive - could strengthen the philosophical case against the special role of RCTs for the EBP movement. As established in the last chapter, this case needs support because the main premise of the external validity critique is not well accepted beyond the philosophical literature. By analysing parallels in the argumentative strategy of the MTE and the external validity argument, we have seen that the MTE argument could fill the gap created by this absence of acceptance. We also discussed that it could even strengthen the philosophical case against the special role of RCTs for EBP in cases in which the external validity argument cannot do so.

To see if the MTE argument can fulfil these two functions and thus strengthen the philosophical case against the special role of RCTs for EBP, the remainder of the thesis presents a philosophical discussion of the persuasiveness of the MTE argument. Above, we have already seen that theoretical arguments and empirical evidence back up the first premise and the second premise of the MTE argument. However, we have not discussed the support for the third and the fourth premise.

Should these two premises turn out to be unconvincing, then the MTE argument faces similar problems as the external validity argument does. That is, it might be that, contrary to what the third premise of the MTE argument suggests, the joint distribution is not necessary for obtaining distributional parameters that convey distributional information of relevance to policy-makers from RCT data. Then, the assumptions that allow us to recover the joint distribution, as discussed above, would not be required for obtaining policy-relevant information from

RCT data. In the same vein, it might be, contrary to what is suggested by the fourth premise, that there are ways of obtaining the policy-relevant distributional information from the joint distribution without having to make additional assumptions.

In either case, no additional, substantive assumptions would be required for obtaining distributional information of relevance for policymakers from RCT data. Then, the MTE argument could not support the philosophical case against the special role of RCTs for the EBP movement. As we just saw, using the MTE argument for this purpose is only possible because the MTE argument questions the alleged minimality of assumptions that RCTs require for producing evidence of relevance for policymakers. Thus, the third and fourth premise of the MTE argument are essential for using the MTE argument to argue against the special status of RCTs for EBP.

However, methodological advancements in the RCT literature cast doubt on these two essential premises. Possibilities such as subgroup analysis, quantile treatment effects and bounding seem to render either the joint distribution or additional assumptions unnecessary for obtaining policy-relevant distributional information from RCT data. This would weaken the potential of the MTE argument to support the philosophical case against the special role of RCTs for the EBP movement by questioning the second proposition of the RCT argument. Thus, the following chapters discuss the persuasiveness of the MTE argument with a focus on whether its third and fourth premise hold up in light of these methodological advancements, starting with the third premise.

Chapter 3

The Threat from Distributional Parameters that Solely Depend on the Marginal Distributions

Introduction

This chapter presents the first step towards defending the claim advanced by the third premise of the MTE argument, i.e. that the *joint distribution* is *necessary* for obtaining distributional information of relevance to policy-makers from RCT data.

As I show in the *first section*, this premise is threatened: The host of distributional information available from distributional parameters that do not depend on the joint distribution questions the alleged necessity of distributional information solely available from the joint distribution for informing policymakers. I introduce subgroup means as the classic case for such a parameter and quantile treatment effects as the, as I argue, strongest case.

Having established that especially quantile treatment effects provide policymakers with a lot of useful distributional information, I discuss the distributional information that policy-makers cannot obtain from these two distributional parameters in *section two*. There, I discuss that these parameters do not provide policymakers with information on the distribution of individual-level impacts, such as the impact of a policy on individuals located at a given quantile of the initial outcome distribution. This result provides the basis for the next chapter, where I argue that this is information that is relevant for policymakers, thereby defending the third premise of the MTE argument.

1. The Threat from Distributional Parameters Obtainable from RCT Data Alone

As discussed in the previous chapter, the third premise of the MTE argument embodies the idea that many distributional parameters which convey policy-relevant distributional information depend on the joint distribution. Thus, it is argued, the joint distribution is required for obtaining policy-relevant distributional information from RCT data. In conjunction with the fourth premise of the MTE argument, the third premise thus attacks the second proposition of the RCT argument discussed above, i.e. that RCTs only require minimal substantive assumptions for producing evidence of relevance to policy-makers.

However, methodological advancements in the RCT literature cast doubts on this essential step in the MTE argument embodied in its third premise: Distributional parameters such as subgroup means or quantile treatment effects provide ways of obtaining a host of policy-relevant distributional information from RCT data without requiring the joint distribution. This is possible because these distributional parameters only depend on the marginal distributions, which we can obtain from RCT data alone. Thus, should it turn out that these parameters already provide policymakers with all relevant distributional information, then the joint distribution would not be necessary for obtaining policy-relevant distributional information from RCT data. This would undermine the third premise and thus the persuasiveness of the MTE argument as an argument in support of the philosophical case against the special role of RCTs for EBP.

In the following section, I analyse subgroup means and quantile treatment effects, highlighting the host of distributional information that these two parameters can provide policymakers with. Doing so will clarify how this host of distributional information questions the alleged necessity of distributional information solely available from the joint distribution for informing policymaking. Thereby, the threat for the third premise of the MTE argument just sketched will become more concrete.

1.1. The Classic Example: Subgroup Means

I start the analysis of distributional parameters which can be calculated from RCT data without obtaining the joint distribution with *subgroup means*. Subgroup means are classically presented as a powerful way of obtaining distributional information from RCT data (Bitler et al. 2003, p. 2; Jackson et al. 2013, p. 92). Consequently, subgroup means are often reported next to mean treatment effects in practice, making it a distributional parameter that has already gained widespread popularity (Djebbari and Smith 2008, p. 72).

In the following, I introduce subgroup means and explain how subgroup means can be calculated from RCT data without obtaining the joint distribution. Then, I discuss which distributional information subgroup means can provide policymakers with.

As the name indicates, subgroup means are informative about the average effect of a policy on a specific *subgroup* of experimental subjects. These subgroups are defined according to observable characteristics that their members share, such as gender or educational attainment. In other words, subgroup means are mean treatment effects that are not calculated for the whole experimental population but only for a subsample with specific observable characteristics (Bedoya et al. 2017, p. 8 and pp. 32f.).

For instance, imagine that we were interested in whether having free computer access in schools affects girls' digital literacy differently than it affects the digital literacy of their male peers. For finding out about this, we could obtain the subgroup mean for the subgroup of boys and for that of the girls. For this, we would have to compare the average digital literacy of treated girls with that of girls in the control group and proceed analogously for boys[11].

This example shows that subgroup means can be similarly obtained from RCT data as the mean treatment effects discussed in the previous chapter: Just as mean treatment effects, subgroup means are obtained by comparing the average outcomes in control and treatment group, the only difference being that the two groups share a given observable characteristic (*Ibid.,* pp. 32f.; Athey and Imbens 2018, p. 123).

From this comparison, it becomes clear why subgroup means do not depend on the joint distribution: The mean outcomes in control and treatment group, which are compared when obtaining subgroup means, are features of the two marginal distributions of treatment and control group. Now, which distributional information do subgroup means as the classic example for a distributional parameter that solely depends on the marginal distributions provide policymakers with?

As indicated in the examples above, subgroup means provide information regarding whether specific groups, that share given observable characteristics, are more or less likely to benefit from a policy measure of interest. This is very valuable distributional information for policymakers: One can imagine that policymakers might be especially interested in helping certain groups and would not want to implement policy measures if these groups would not benefit. For instance, a policymaker might not be willing to provide the free computer access discussed above if this does not improve girls' digital literacy.

And even though subgroup means already provide policymakers with a lot of information concerning the distributional effects of policies, a second distributional parameter is argued to provide even more detailed insights concerning this (Bitler et al. 2003 and Jackson et al. 2013).

[11] Obtaining subgroup means requires more conditions compared to obtaining the mean treatment effect for the whole sample. For instance, it is necessary that the subgroup of interest has observations in both the control and the treatment group (Bedoya et al. 2017, p. 32f.). However, these conditions are usually not seen as overly demanding because they can often be dealt with in a straightforward manner (Banerjee and Duflo 2008, pp. 26f.).

This parameter is called the *quantile treatment effect* and can, just as subgroup means, be obtained from RCT data without needing the joint distribution.

I shall thus introduce quantile treatment effects (in the following: QTEs) in the next subsection. I then present empirical evidence for the claim that QTEs provide policymakers with even more detailed information concerning the distributional effects of policies than subgroup means do. For this, I use a case study from education policy. Finally, I present theoretical results concerning the high informativeness of QTEs for policymaking. Thereby, it becomes clear why QTEs provide the strongest case against the claim reflected in the third premise of the MTE argument, i.e. that distributional information solely obtainable from the joint distribution is relevant for policy-makers.

1.2. The Strongest Case: QTEs

In the case study that I want to discuss, Jackson et al. re-analyse data from one of the most famous RCTs within education policy: The Tennessee STAR experiment. In this RCT, the impact of class-size reductions on student outcomes was examined. For this, students were randomly assigned to either a small class containing 13-17 students or a regular class with 22-25 students[12]. Then, student achievement, measured by how the students scored in a standardized test, was compared between the two groups. The RCT revealed a strikingly positive effect of smaller class-sizes on student outcomes (Jackson et al. 2013, pp. 94f.).

Inspired by this striking result, many researchers have examined the distributional effects of class-size reductions using the STAR data. For this, researchers have mostly looked at subgroup means. In doing so, they found that subgroups characterized by coming from low-income households or belonging to an ethnic minority benefited more from class-size reductions compared to more advantaged students. For instance, researchers found a higher subgroup mean for black students than for white students (*Ibid.,* pp. 99ff.).

In contrast to these subgroup analyses, Jackson et al. use QTEs for analysing the distributional effects of class-size reductions based on the STAR data. QTEs are estimated by looking at the difference in quantiles of the outcome distributions of treatment and control group (Athey and Imbens 2018, p. 91; Bitler et al. 2003, p. 16).

[12] A third group was assigned to a regular class with a full-time teacher's aid (Jackson et al. 2013, p. 94).

For instance, obtaining the QTE at the 0.5 quantile in the case of the STAR experiment would work as follows: We would take the difference between the median test score of the treatment group, which was allocated to smaller classes, and the median test score of the control group. Doing so would be informative about the effect of a class-size reduction measure at the median of the achievement distribution (Jackson et al. 2013, p. 95).

For this example, a formal expression of QTEs would look as follows: $QTE_{0.5} = q_{Y(1)}(0.5) - q_{Y(0)}(0.5)$. From this formal expression, it becomes clear that QTEs can also be obtained from RCT data solely based on the marginal distributions and thus without requiring the joint distribution: As the formula shows, QTEs are, as mean treatment effects, calculated by comparing features of the marginal distributions. The only difference is that QTEs are calculated by taking the difference between quantiles of the marginal distributions and not the means (Athey and Imbens 2018, p. 91).

Having introduced QTEs, let us go back to our case study by Jackson et al., who employ QTEs for analysing the distributional effects of class-size reductions using the STAR data. For this, Jackson et al. estimate these QTEs for all 99 centiles of the test score distribution. Thereby, they obtain the effect of class-size reductions at each of the 99 centiles. In doing so, they find that the effects of smaller classes are positive over all of the achievement distribution and that the highest gains are at the top of the achievement distribution (Jackson et al. 2013, pp. 96ff.).

Interestingly, this result paints a different picture concerning the distributional effects of the STAR experiment compared to the results obtained by authors who have used subgroup means instead of QTEs for analysing the STAR data. As pointed out, these authors found that the subgroup mean for students from disadvantaged socio-economic backgrounds was higher than the mean for students from advantaged backgrounds. However, as Jackson et al. explain, students from disadvantaged socio-economic backgrounds usually, and also in the case of the STAR experiment, have lower test scores compared to their peers. Therefore, the finding from Jackson et al.'s QTE analysis (i.e. that the biggest gains are at the top of the achievement distribution) sheds another light on the distributional effects of class-size reductions than the finding from the subgroup analyses (i.e. that students from disadvantaged socio-economic backgrounds benefitted most) (*Ibid.,* p. 100).

As Jackson et al. argue, this seeming contradiction between the findings from their QTE analysis and the subgroup analyses indicates that QTEs provide us with a more detailed picture of the distributional effects of policies. While the subgroup analyses implied that class-size reductions are especially beneficial for students from disadvantaged socio-economic backgrounds, the analysis from QTEs qualifies this result. It does so by revealing that the result is mainly driven by students from disadvantaged backgrounds who are at the top of the achievement distribution (Jackson et al. 2013, p. 93 and p. 101).

From this example, we can see that QTEs provide policymakers with an even more detailed and rich picture of the distributional effects of policies, thereby enriching the information obtainable from subgroup analyses. More generally, in cases in which the intra-group variation is larger than the inter-group variation in mean impacts, like in the present case, QTEs reveal information regarding the distributional effects of interventions that subgroup means do not reveal. For instance, in the case of the STAR experiment, the QTE analysis revealed more information regarding the distributional effects of the policy because the variation in impacts *within* the group of students from disadvantaged backgrounds was larger than the variation in impacts *between* the group of students from disadvantaged and advantaged backgrounds (*Ibid.*, p. 93).

These empirical results regarding the detailed distributional information that QTEs provide policymakers with are also supported by theoretical research on the informativeness of QTEs regarding the distribution of impacts. For example, it has been shown that if any obtained QTE is negative (positive), we know that the treatment effect must be negative (positive) for some parts of the outcome distribution. For instance, given the positive QTEs obtained by Jackson et al., we can conclude that at least some students benefitted from the class-size reductions. Given these properties, QTEs provide policymakers with a detailed picture of the distribution of treatment effects (*Ibid.,* pp. 95f.; Bitler et al. 2003, pp. 17).

Furthermore, QTEs can be used for obtaining many other distributional parameters of interest, such as the Gini coefficient. This is because many distributional parameters are inequality measures, which are, in turn, functions of quantiles. Since many of these inequality measures are important in policymaking discourses on the distributional effects of interventions, QTEs provide valuable input for these discourses (Atkinson 1970, p. 244; Firpo 2007, p. 259; Frölich and Molly 2013, pp. 347f.).

Thus, there are empirical and theoretical arguments for a high informativeness of QTEs regarding the distributional effects of interventions. Thereby, QTEs provide policymakers with a host of distributional information. This severely threatens the third premise of the MTE argument. It does so by casting doubts on the relevance of distributional information conveyed by distributional parameters that depend on the joint distribution for policymaking.

To refute this threat in the next chapter, the following section highlights which distributional information is *not* obtainable from QTEs as the strongest case for a distributional parameter that can be obtained solely from the marginal distributions. This provides the basis for my later argument that this distributional information is relevant for policymaking.

2. What Quantile Treatment Effects Miss

To see which information *cannot* be obtained from QTEs, it is helpful to distinguish between QTEs and quantiles of the difference Y_1-Y_0.

As described above, QTEs are differences in quantiles of the two marginal distributions and can be written as follows: $QTE_s = q_{Y(1)}(s) - q_{Y(0)}(s)$. They thus must be distinguished from the quantiles of the difference Y_1-Y_0, which can be written as follows: $QD_s = q_{Y(1) - Y(0)}(s)$. For instance, it is thus not the case that the difference in median test scores between the treatment and the control group in the STAR experiment generally equals the median of the difference of the two test score distributions, i.e. the median treatment effect[13]. (Athey and Imbens 2018, p. 91; Imbens and Wooldrige 2009, p. 17).

This general difference arises in the presence of what Bedoya et al. call "individual mobility across the distribution" (Bedoya et al. 2017, p. 7). With that, they mean that a policy can induce re-orderings regarding the ranks individuals occupy in the distribution of interest. For instance, it might be that a student in the STAR experiment who would have had the lowest score without treatment scores highest when being placed in a smaller class (*Ibid.*; Jackson et al. 2013, p. 95). In the literature, it is generally accepted that there are instances of mobility effects and, thus, that QTEs generally differ from quantiles of the difference Y_1-Y_0 (Athey and Imbens 2018, p. 91; Imbens and Wooldrige 2009, p. 17).

This general difference between QTEs and quantiles of the difference Y_1-Y_0 in the presence of mobility effects is important because it prevents us from generally interpreting QTEs as quantiles of the difference Y_1-Y_0. In the presence of mobility effects, QTEs only tell us about the

[13] This differentiates quantiles from the mean treatment effect, for which this applies because the mean is a linear operator.

effect of a policy on a given quantile, say, the median, of the outcome distribution. For instance, in the study of Jackson et al., the obtained QTEs would be informative about whether class-size reductions caused the median of the achievement distribution to move up. However, QTEs are not informative about the policy's impact on students who were originally, that is, in the absence of the policy, at the median of the achievement distribution. In contrast, this is what the quantile of the difference Y_1-Y_0 is informative about (Abadie et al. 1998, p. 10).

That mobility effects prevent us from generally interpreting QTEs as quantiles of the difference Y_1-Y_0 heavily limits the informativeness of QTEs, as the strongest case for a distributional parameter that does not depend on the joint distribution, for policymaking. To illustrate this, consider the following example study discussed by Bedoya et al. (2017):

In this study (Bruhn et al. 2016), Bruhn et al. conduct an RCT to find out about the effect of financial education in high school on the financial proficiency of students. For this, schools were randomly allocated into two groups: A treatment group, which received classes in basic finance, and a control group, which did not. When comparing the financial proficiency of students in control and treatment schools, the authors found a significant positive effect of financial education on financial proficiency (*Ibid.*, pp. 257ff.).

Bedoya et al. re-analysed the data of this RCT to gain policy-relevant insights beyond Bruhn et al.'s result that students' financial literacy in treatment schools was higher on average. To also learn about the distributional impacts of this policy measure, Bedoya et al. calculated QTEs. Their results can be depicted as follows (Bedoya et al. 2017, p. 46):

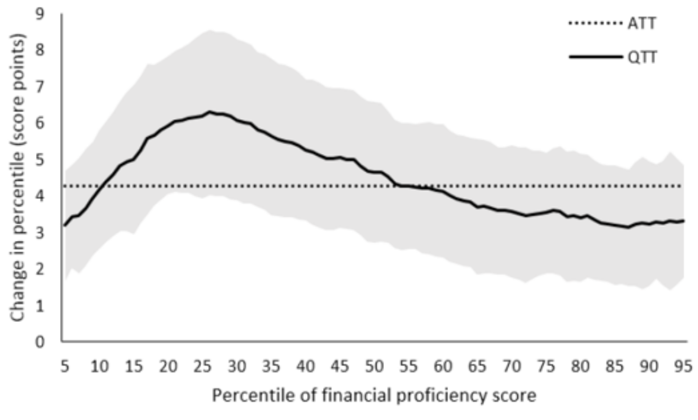

Figure 2: QTEs based on the Data from Bruhn et al. (2016)

In this graph, the dashed line depicts the mean treatment effect and the solid line the QTEs. Now, which policy-relevant information can be derived from this? At first sight, it seems that students with a low financial proficiency score benefitted more than the average, judging from the fact that the change in the bottom percentiles is, relatively speaking, larger (Bedoya et al. 2017, p. 47). This would be valuable information for policymakers especially interested in the financial literacy of this group.

However, it seems likely that some students at the bottom of the original financial proficiency distribution would move up the distribution in the case of financial education. These mobility effects prevent us from interpreting the depicted QTEs as quantiles of the difference Y_1-Y_0. Therefore, the QTEs in the graph tell us nothing about *individuals* at given quantiles of the original financial proficiency distribution. In the presence of mobility effects, the only distributional information that this graph provides us with is that some students benefitted from the intervention since the QTEs are positive. From this, we could not even exclude the possibility that some pupils were harmed by the policy (*Ibid.*, p. 14 and p. 47).

This illustrates how mobility effects generally limit the informativeness of QTEs for policymaking by making QTEs uninformative about effects on individuals located at given quantiles of the original outcome distribution. We could overcome this limitation and, for instance, find out if low-scoring students benefit more from financial education by obtaining the quantiles of the difference Y_1-Y_0 directly. However, as becomes apparent from the formalization of this distributional parameter ($QD_s = q_{Y(1) - Y(0)}(s)$), doing so requires obtaining the joint distribution.

This discussion on the limited informativeness of QTEs in the presence of mobility effects illustrates a general result regarding information obtainable from distributional parameters that only depend on the marginal distributions. Namely, these parameters are only informative about the impact of a policy *on outcome distributions*. For instance, QTEs are generally only informative about policy effects on given quantiles of the outcome distributions. In contrast, distributional parameters that depend on the joint distribution are informative about the *distribution of individual treatment impacts*. For instance, the quantiles of the difference Y_1-Y_0 tell us about the effect of a policy on *individuals* originally located at given quantiles of the outcome distribution (*Ibid.*, p. 3; Firpo 2010, p. 1).

This shows that even though especially QTEs as the strongest case for a distributional parameter that solely depends on the marginal distributions provide us with a host of distributional information, information about the distribution of individual-level impacts is missed by them. Information on the distribution of individual-level impacts is, importantly, not limited to information about policy effects on individuals originally located at given quantiles of the outcome distribution. Two other often mentioned examples in the literature are the proportion of people who benefit from, or are harmed by, a policy and the variance of treatment effects (Bedoya et al. 2017, pp. 15ff.).

Conclusion

In this chapter, we looked at subgroup means and QTEs as the two main examples for distributional parameters which are obtainable from RCT data without requiring the joint distribution. As discussed, especially QTEs provide us with a lot of distributional information. Thereby, the necessity of the joint distribution for obtaining policy-relevant distributional information from RCT data is called into question. However, we have also seen that distributional parameters that solely depend on the marginal distributions are not informative about the distribution of individual-level impacts. For instance, they do not inform us about the impact of a policy on individuals located in given quantiles of the initial distribution.

This result forms the basis for the discussion in the next chapter. There, I discuss whether the threat for the third premise of the MTE argument raised in this chapter can be refuted. As already indicated, this depends on whether distributional parameters that solely depend on the marginal distributions provide policymakers with *all* policy-relevant distributional information. Then, the information about which we have established that it can only be obtained based on distributional parameters that depend on the joint distribution would become *irrelevant*. This would imply that the third premise of the MTE argument, which is crucial for the MTE argument to be a persuasive argument against the special status of RCTs for the EBP movement, does not hold up.

Chapter 4

Defending the Necessity of Distributional Information that is Solely Obtainable from the Joint Distribution for Policy Making

Introduction

In the last chapter, we focused on the third premise of the MTE argument. This premise holds that the joint distribution is needed to obtain policy-relevant distributional information conveyed by distributional parameters from RCT data. As discussed, the policy-relevance of this distributional information is called into question by the host of distributional information available from distributional parameters that do not depend on the joint distribution, especially from QTEs. This threatens the third premise of the MTE argument.

However, we also discussed that the strength of this threat depends on the following: Whether the host of distributional information obtainable from these parameters is already *sufficient* for policymakers or whether distributional information that is solely obtainable from RCT data by obtaining the joint distribution is also *necessary*. This is the question that we shall discuss in this chapter, thereby defending the third premise of the MTE argument.

In the economic literature, the question of whether distributional information that is solely obtainable from the joint distribution is a necessary input for policy-making decisions or if the information obtainable from the marginal distributions is sufficient is heavily debated:

On one side of this debate, scholars argue that the only distributional information that policymakers require is information on how a policy affects outcome distributions (e.g. Manski 1996 or Imbens 2010). That is, they argue that *only* information about how a policy impacts the distributions of interest for the scenario with and without the policy is policy-relevant (Firpo 2010, p. 1). As discussed, this information is obtainable from the marginal distributions alone.

On the other side of this debate, researchers argue that information about the distribution of individual treatment effects is *also* policy-relevant, over and above the information on how a policy affects outcome distributions (*Ibid.*). As we saw, this type of distributional information is only obtainable from the joint distribution. Inspired by a formulation in Bitler et al. (Bitler et al. 2003, p. 28), I shall refer to this view as the "Names-Matter-view".

By defending the relevance of distributional information solely obtainable from the joint distribution for policymaking, the Names-Matter-view backs up the third premise of the MTE argument. Thus, we can turn to defences of the Names-Matter-view for finding an argument to defend the third premise of the MTE argument. In doing so, we can refute the challenge of a potential sufficiency of distributional information solely obtainable from the marginal distributions for policymaking.

In the *first section* of this chapter, I thus discuss a frequently employed argument from the Names-Matter-view literature. As I argue, this argument relies on an implicit assumption that would have to be defended for using it to strengthen the third premise of the MTE argument. In *section two*, I extract a potential defence for this implicit assumption from other writings of Names-Matter proponents. In doing so, I defend the third premise of the MTE argument.

1. A Problematic Implicit Assumption

In the frequently employed argument for the Names-Matter-view that I discuss in this section, Names-Matter proponents focus on the conditions under which the claim of the contrary view, i.e., that such information is not relevant for policymaking, would hold up. As Names-Matter proponents argue, two such conditions, out of which one *needs to hold* to defend the contrary view, can be identified. Then, they argue that neither of these two conditions is plausible. Thus, so the argument goes, distributional information solely obtainable from the joint distribution is relevant for policymaking (e.g. Heckman et al. 1997, p. 491).

The two conditions of which one has to hold up to defend the contrary view according to Names-Matter proponents are the following: Either policymakers would have to think of distributional issues that can solely be illuminated by distributional parameters that depend on the joint distribution as *irrelevant*. That would be the case if policymakers deemed, for instance, the effect of a policy on individuals who were originally located at a given quantile of the outcome distribution irrelevant for their implementation decisions (cf. *Ibid.*).

Alternatively, the opposite view is argued to be defendable if distributional issues about which we can only learn based on the joint distribution can be *offset* by transferring policy benefits. For instance, it might be that a policymaker would deem the fraction of people who would be harmed by a policy option relevant for her implementation decision. However, even if a policy shown to have a positive mean treatment effect in an RCT harms most people while only benefiting a few, it might be possible to eliminate this undesirable feature and redistribute benefits from those who benefit to those who do not. In this case, the information on the positive mean treatment effect would be sufficient for the policymaker even though she cares about the proportion of people who would be harmed (cf. Bedoya et al. 2017, p. 2; Heckman et al. 1997, p. 491).

Let us now see how Names-Matter proponents argue that neither of these two conditions holds up, starting with the second one. According to Names-Matter-view proponents, the condition that transfers can offset undesirable distributional consequences of policies does not hold. This is because, firstly, not all policy benefits can be redistributed. Take the following example from education policy: The achievement benefits of students profiting from class-size reductions cannot just be conferred upon those who do not benefit or are even harmed. Just as many other goods, such as health, education benefits are not transferable. Thus, calling for a redistribution of benefits to offset undesirable distributional consequences of policies is argued not to be a viable option in many cases (Heckman and Vytlacil 2007, p. 4806).

Secondly, even in cases where this is possible because the policy outcomes in question, such as those from our class-size reduction policy, can be translated in monetary terms and thus be redistributed, it is argued that practical problems emerge. For instance, transfers might be costly and thus efficiency-reducing. Or, there might be political restrictions on feasible redistributions. Most importantly, Bedoya et al. point out that "implementing the optimal transfer scheme requires some knowledge of the distribution of gains and losses" (Bedoya et al. 2017, p. 2). However, this knowledge is only obtainable from the joint distribution, as we have seen in the last chapter. Thus, according to Names-Matter proponents, the possibility to redistribute cannot be counted as a reason for accepting the contrary view, which sees no role for information that is solely obtainable from the joint distribution (Heckman et al. 1997, p. 490).

This discussion shows that Names-Matter proponents have formulated many arguments to back up their claim that one of the two conditions conditional on which one could accept the opposite view does not hold. Now, what about the other condition? Interestingly, arguments for why policymakers deem information solely obtainable from the joint distribution relevant for their policy-making decisions are much scarcer. This is, presumably, because Names-Matter-view proponents take this relevance for granted.

However, relying on this as an implicit assumption becomes problematic if we want to use this argument to defend the third premise of the MTE argument. These problems would arise if it turned out that only a limited number of policymakers comply with the intuition of the Names-Matter-view. That is, most policymakers might not deem distributional information such as the proportion of people who got harmed by a policy relevant for their implementation decisions. To answer the distributional questions from most policymakers based on RCT data, we would then not need the joint distribution and thus no additional assumptions.

If that was true, then the argument for the Names-Matter-view discussed in this section would be unhelpful for defending the third premise of the MTE argument and thus its persuasiveness. This is because these considerations would imply that there are *very few cases* in which we would need assumptions beyond the minimal RCT assumptions to obtain relevant evidence for policymakers from RCT data. This would severely limit the scope of the MTE argument as an argument against the special role of RCTs for EBP.

The possibility that the implicit assumption made by Names-Matter proponents only holds up for a limited number of policymakers is especially salient in light of our discussion in the last chapter. There, we have seen that much information about which we intuitively think that it would be relevant for policymakers is readily available from distributional parameters that only depend on the marginal distributions. For instance, if policymakers wished to cater to the interest of specific groups such as children or women, information from subgroup means is all they need. Even if questions on inequality were on their minds, the inequality measures obtainable from QTEs would provide the necessary insights.

This discussion shows the need for defending the implicit assumption on which the argument for the Names-Matter-view discussed in this section relies. For this, we need an argument showing why policymakers would deem distributional information solely obtainable from distributional parameters that depend on the joint distribution relevant. Especially, this argument needs to hold even in light of the intuition that most information that policymakers deem relevant is obtainable from the marginal distributions. Without such an argument, the relevance of distributional information solely obtainable from the joint distribution for policymaking would be limited. This would limit the MTE argument's scope.

Ultimately, it will remain an empirical question whether distributional information solely obtainable from the joint distribution is deemed relevant by policymakers. However, I argue in the next section that we can read one of the rare defences of Names-Matter proponents for why this information, which is argued to be more nuanced, would be policy-relevant as suggesting the following: That we only need to make *one general and plausible assumption* about the interests of policymakers to render this information relevant for them. Based on this defence of the implicit assumption made in the argument for the Names-Matter view discussed in this section, the third premise of the MTE argument can be defended.

2. Defending the Names-Matter-View

2.1. A Defence of the Implicit Assumption

For arriving at this plausible and general assumption, let us turn to one of the scarce defences of the implicit assumption that distributional information that is solely obtainable from the joint distribution is relevant for policymakers from the Names-Matter-view literature:

> Using the joint distribution […], it is possible to develop a *more nuanced understanding* of the distributional impacts of public policies, and to move beyond comparisons of aggregate overall distributions induced by different policies to consider how people in different portions of an initial distribution are affected by public policy (Carneiro et al. 2003, p. 1, emphasis M.S.).

As Carneiro et al. suggest in this quote, the understanding that policy makers gain about the impact of their policies based on the joint distribution is *more nuanced* compared to the understanding that the marginal distributions convey. As the authors seem to indicate, this nuanced information conveyed by the joint distribution is thus also more informative for policymakers. Thereby, the Carneiro et al. quote provides grounds for defending the policy-relevance of this distributional information. Importantly, it would do so even in light of the intuition discussed above, i.e. that most information that policymakers deem relevant can be obtained from the marginal distributions.

This defence thus provides a starting point for strengthening the third premise of the MTE argument. In the following, I want to make the support that this defence could lend to the claim made by the third premise of the MTE argument more concrete. For this, I want to present *one way* of spelling out why precisely the nuanced information on "how people in different portions of an initial distribution are affected by public policy" might be policy-relevant.

For presenting one way in which this nuanced information could be policy-relevant, I will explicate a quote by Heckman et al.[14], which is often discussed in the literature (e.g. Wu and Perloff 2007, p. 2). As I argue, Heckman et al.'s proposal can be read as suggesting that the nuanced information which is solely obtainable from the joint distribution becomes policy-

[14] Moving from the nuanced information that the joint distribution provides for policymakers to asking why this information is policy-relevant requires several steps and assumptions that would be worth discussing in-depth. However, the purpose of explicating the proposal by Heckman et al. is merely to sketch one way in which one could spell out the policy-relevance of the nuanced distributional information that Carneiro et al. emphasize. Thus, I will not provide a detailed discussion on the assumptions and steps figuring into Heckman et al.'s account.

relevant under the general and plausible assumption that policymakers care about the support of voters. This quote reads as follows:

> The distribution of the benefits (and costs) from a programme determines the support for a programme if voters are self-interested or if they are altruistic. In median voter models, the mean is irrelevant unless it coincides with the median. An altruistic voter may wish to see the lot of the worst-off advanced if he adopts Rawls' (1972) maximin criterion for social justice (Heckman et al. 1997, p. 488).

To structure the following discussion, notice the taxonomy of voters and policymakers implied in this quote: Firstly, according to Heckman et al., there are two types of voters: Self-interested and altruistic voters. Secondly, it is suggested that we can divide policymakers into two groups: They either deem distributional issues that can solely be illuminated by distributional parameters that depend on the joint distribution relevant or they do not. Let us say that in the first case, policymakers are "concerned" about these distributional issues, and in the second case, they are "not concerned". From this, the following taxonomy emerges:

	Policy Makers	
Voters	altruistic, concerned	altruistic, not concerned
	self-interested, concerned	self-interested, not concerned

Table 2: Taxonomy Voter and Policy Maker Types

For the two states depicted on the left side of the table, the implicit assumption on which the argument of Names-Matter-view proponents discussed above relies does not need to be defended. In these states, policymakers already deem the nuanced distributional information that can solely be illuminated by the joint distribution relevant. For the two cases on the right-hand side, however, this implicit assumption needs to be defended. In these cases, policymakers do not deem the nuanced distributional information that can solely be obtained from the joint distribution relevant.

However, I argue that we can read Heckman et al. as suggesting that such information would become relevant for these policymakers as well under the assumption that *policymakers are interested in the support of the electorate for their policies*. This is because, according to the quote, the distribution of benefits, which we can only obtain from the joint distribution, determines whether the electorate supports a given policy.

To my mind, a defence of the policy-relevance of the nuanced distributional information that is solely obtainable from the joint distribution based on the assumption that policymakers are interested in voters' support[15] is likely to be fruitful. Namely, as I see it, this assumption is both *plausible* and *general.* This assumption is arguably plausible because it is commonly employed in research about policy-makers' decision-making processes (e.g. Dur 2001, p. 221 or Gustafsson 2019, p. 283). This assumption is general because it can be argued to hold for all or at least most policymakers[16] [17]. Thus, showing that the nuanced distributional information that is solely obtainable from the joint distribution becomes relevant for policymakers under the assumption that policymakers care about the support of voters plausibly strengthens the persuasiveness of the MTE argument without limiting its scope[18].

In the following, I spell out what I take to be the idea behind Heckman et al.'s indication that the nuanced distributional information that is solely obtainable from the joint distribution becomes relevant even for unconcerned policymakers, assuming that they are interested in the support of voters. In doing so, I aim to make Carneiro et al.'s defence for the implicit assumption made in the argument of the Names-Matter proponents discovered above more concrete. Backed up by this concrete defence, the Names-Matter-view argument can be used to strengthen the third premise of the MTE argument against a potential sufficiency of distributional information that is solely obtainable from the marginal distributions for policymaking.

[15] Heckman et al. focus on voters and policymakers, but this assumption can also be interpreted more broadly, and thereby encompasses even more cases: For instance, imagine a principal who is thinking about reducing class-sizes in her school. This case involves no policymakers or voters. Still, we can interpret the assumption made by Heckman et al. such that the principal is interested in the support of her students for the class-size reduction policy. Given this broader interpretation, the assumption that policymakers are interested in the support of the electorate also applies to usages of RCTs outside of the strict political-institutional realm.

[16] There is a difference between policymakers, who formulate policies, and politicians, who are elected and thus compete for voters. However, this distinction often becomes blurry nowadays (see e.g. Cairney et al. 2017 or National Co-Ordinating Centre for Public Engagement 2020). One reason for this is probably that concerns of politicians about the attraction of voters prominently figure into the decision-making process of policymakers. This is also why I will stick to the term "policy-makers" in the following.

[17] The generality of this assumption is limited in so far that it only holds for democracies: Only then it is important for policymakers to secure the support from voters. However, in the context of EBP, this does not impede the generality of this assumption because the success of the EBP movement is (so far) confined to democratic countries.

[18] Alternative assumptions based on which we could defend the contention of Names-Matter proponents that the nuanced distributional information that is solely obtainable from the joint distribution is relevant for policymakers would be that policymakers want to see no one harmed or that individuals are loss adverse (Firpo and Ridder 2008). However, it would have to be discussed whether these assumptions are as general and plausible as the assumption employed by Heckman et al.

For showing this, let us first examine the case in which policymakers are *unconcerned* and want to attract *self-interested* voters. From the Heckman et al. quote, we can infer that the median voter theorem might illuminate this case. This makes sense because the median voter theorem assumes that voters are self-interested. Let us thus discuss how the median voter theorem could explain why unconcerned policymakers need the nuanced information that is solely obtainable from the joint distribution to ensure the support of self-interested voters.

2.2. Unconcerned Policy Makers, Selfish Voters and The Median Voter Theorem

The median voter theorem is a widely used model employed to analyse policymakers' competition for voters. That is, the median voter theorem predicts, under certain assumptions, which policies policymakers will choose when they are competing for voters. According to this theorem, policymakers will decide to adopt the policy preferred by the *median voter* to maximize their votes. The median voter is the voter whose preferred policy is the median in the preference distribution of voters (Larcinese 2007, p. 2; Scervini 2012, p. 530).

Importantly, since this model assumes that voters are self-interested, the preference distribution of voters is determined by the relevant outcome distribution. That is, the median voter prefers a policy over another if the policy in question improves her position in the relevant outcome distribution more than another policy. For instance, let the policy in question be a proportional tax combined with a lump sum benefit for everyone. Then, the median voter will only be in favour of this policy if the median of the income distribution is below the mean: Only then she will be a net beneficiary of the policy and thus support it (Iversen and Goplerud 2018, p. 297; cf. Larcinese 2007, p. 2).

Now, how can this model be used to argue that unconcerned policy-makers need nuanced information that is solely obtainable from distributional parameters that depend on the joint distribution to attract self-interested voters, as suggested by Heckman et al.?

To see this, recall our discussion in the last chapter. There, we established that the joint distribution is necessary for obtaining the impact of a policy on individuals located at a given quantile in the initial outcome distribution. However, according to the median voter theorem, the impact of a policy on the voter(s) located at the median of the initial distribution is the crucial piece of information that policymakers need to find out whether self-interested voters support a policy implementation decision. If a policy is not beneficial for voters at the median of the initial distribution of interest, implementing the policy will not help the policymaker when competing for self-interested voters. In our public finance example, the policy would only be implemented if the median earner, i.e. the person at the median of the initial earnings distribution, benefits from it.

I have thus spelled out what I take to be the reasoning behind Heckman et al.'s contention that the nuanced information that is solely obtainable from the joint distribution is relevant for unconcerned policymakers interested in the support of self-interested voters. Now, let us turn to the case of *unconcerned* policymakers and *altruistic* voters.

2.3. Unconcerned Policy Makers, Altruistic Voters and The Worst-Off

From the quote above, it becomes clear that Heckman et al. suggest that altruistic voters want to see "the lot of the worst-off advanced". In this subsection, I spell out why this contention would mean that unconcerned policy makers interested in gaining the support of the altruistic electorate require the nuanced distributional information which is solely obtainable from the joint distribution[19].

From Heckman et al.'s contention regarding the interest of altruistic voters, it seems to follow that a policymaker interested in gaining the votes of an altruistic electorate needs information on how the policy in question affects the situation of the worst-off. That is, only if she had evidence that the policy measure would benefit the worst-off, she would be ensured that implementing the policy measure improves "the lot of the worst-off" and would thus secure her the support of the altruistic electorate.

In our tax policy example, this means that a policymaker interested in votes from an altruistic electorate would need information regarding the effect of the tax policy on the individuals at the lowest quantile of the initial income distribution. From this information, she could tell if the policy improves "the lot of the worst-off".

[19] The possibility of voting behaviour that is determined by factors other than individuals' self-interest, such as altruism, is explored by behavioural economists (Durante et al. 2014, p. 1067; Luebker 2012, p. 11). Within this literature, Heckman et al.'s contention that "an altruistic voter may wish to see the lot of the worst-off advanced" has its analogy in the desire to maximize the minimal payoff in a group. Such preferences are referred to as "maximin preferences" (Engelmann and Strobel 2004, p. 857). Interestingly, behavioural economists' experiments have produced evidence that indicates that maximin preferences play an important role when voters have no grounds for choosing selfishly. In one of these experiments a "median voter" must decide on the allocation of money between poor and rich voters. Since the median voter gets the same amount of money no matter which allocation she chooses, she has no grounds for making a self-interested choice (*Ibid.,* p. 859; Bolton and Ockenfels 2006, pp. 1906f.; Fehr et al. 2006, p. 1912). In this and similar experimental set-ups, the results indicate that maximin preferences decisively figure in the allocation decision of the altruistic voter (Engelmann and Strobel 2004, p. 861, *Ibid.* 2007, p. 293).

Now, is this information obtainable from the marginal distributions alone, that is, without needing the joint distribution? As we have seen in the last chapter, it is not: To obtain the effect of a policy on individuals at a given quantile of the initial outcome distribution (in this case, the lowest), we need information that is solely obtainable from the joint distribution.

2.4. Summary Argument

In the last sections, we have discussed a potential defence of the implicit assumption on which the argument for the Names-Matter-view discussed above relies. As we have seen, such a defence might point to the nuanced information about the effects of policies that the joint distribution provides policymakers with. To make this argument more concrete, I spelled out one way in which this nuanced information could be policy-relevant, using a quote by Heckman et al.. As I argued, this quote can be read as suggesting that this nuanced information becomes policy-relevant under the general and plausible assumption that policymakers care about the support of the voters.

I then turned to show how this plausible and general assumption allows for defending the policy-relevance of the nuanced distributional information obtainable from the joint distribution. To this end, I identified two cases from the Heckman et al. quote in which it is not clear why the nuanced distributional information that is only available from the joint distribution is relevant for policymakers: The cases in which policymakers are not concerned about the nuanced distributional information that is only available from the joint distribution and voters are either selfish or altruistic. However, I have argued that we can read Heckman et al. as suggesting that this nuanced distributional information becomes relevant for these policymakers as well under the assumption that policymakers are interested in attracting voters[20].

In the case of selfish voters, I have explicated Heckman et al. as suggesting that policymakers need information about the impact of the policy in question on voters at the median of the initial outcome distribution. In the case of altruistic voters, I have explicated the authors as

[20] It might be argued that this argument already assumes what it aims to show, as it starts out from the assumption that policymakers care about individuals (i.e. voters) for showing that information on individual-level impacts, which is solely obtainable from the joint distribution, is relevant for policymakers. However, an interest for the support of voters does not automatically entail the relevance of individual-level distributional information: For instance, one could argue that being interested in the support of voters means that policymakers need to know whether a policy is beneficial on average (i.e. has a positive MTE), instead of needing information on individual-level impacts. Still, the argument discussed in this chapter shows that information in the form of MTEs is not sufficient for policymakers who care about voters. Thus, the argument contributes to establishing the conclusion that information on individual-level impacts is policy-relevant instead of merely assuming it.

suggesting that policymakers need information about the impact of the policy in question on voters at the lowest quantile of the initial distribution. Each information is, as established in the last chapter, only obtainable from the joint distribution. Thereby, the argument developed in the preceding sections can be used to strengthen the third premise of the MTE argument.

Conclusion

In this chapter, I have refuted a major threat for the third premise of the MTE argument. This premise claims that the joint distribution is needed for obtaining policy-relevant distributional information conveyed by distributional parameters from RCT data. As I pointed out, the policy-relevance of this distributional information is called into question by a potential sufficiency of distributional information solely obtainable from the marginal distributions for policymaking. Should that be the case, then obtaining the joint distribution would not be required for obtaining policy-relevant distributional information from RCT data. This would contradict the claim made by the third premise of the MTE argument.

For refuting this threat against the third premise of the MTE argument, I have turned to what I call the “Names-Matter-view”. The Names-Matter-view holds that information on the distribution of individual-level impacts, which, as we have seen, is solely available from the joint distribution, is a necessary input for policy-making decisions. Thus, by defending the Names-Matter-view, we can defend the third premise of the MTE argument.

To this end, I have discussed a frequently made argument for the Names-Matter-view. Still, I pointed out that this argument relies on the implicit assumption that policymakers generally deem distributional information that can solely be illuminated by distributional parameters that depend on the joint distribution relevant. This is problematic because the salient possibility that many policymakers do not comply with this intuition would severely limit the scope of the MTE argument.

I have then turned to one of the scarce defences for this assumption from the Names-Matter-view literature. According to this defence, the distributional information that is solely obtainable from the joint distribution is policy-relevant because it provides policymakers with a more nuanced picture of the effects of their policies. To make the relevance of this nuanced information more concrete, I explicated a quote from Heckman et al.. As I argued, this quote suggests that this nuanced distributional information becomes relevant to all policymakers under the general and plausible assumption that they care about the electorate's support. That is, if policymakers want to find out whether their policy decisions would be supported by both a selfish and an altruistic electorate, they need the nuanced distributional information that is solely obtainable from the joint distribution. This explication yields a concrete defence that backs up the claim made by the third premise of the MTE argument.

We can thus conclude that even though many useful methods exist for obtaining distributional information of interest to policymakers solely based on the marginal distributions, information from distributional parameters that can solely be obtained from the joint distribution can be argued to become relevant for policymakers under the assumption that they care about the support of voters. Thereby, a major threat against the third premise of the MTE argument can be refuted.

However, as discussed above, the fourth premise of the MTE argument is equally important for establishing the persuasiveness of the MTE argument as an argument in support of the philosophical case against the special role of RCTs for EBP. In addition, this fourth premise is, just as the third premise, endangered by methodological advancements in the RCT literature. Thus, let us discuss this fourth premise in the next chapter.

Chapter 5

The Threat from Bounds

Introduction

In this chapter, we shall look at a first threat for the fourth premise of the MTE argument. This premise remains to be defended for establishing that the MTE argument is persuasive and can thus support the philosophical case against the special role of RCTs for EBP. To recap, the fourth premise of the MTE argument reads as follows:

> ***P4:*** To obtain policy-relevant distributional information conveyed by distributional parameters that depend on the joint distribution from RCT data, (1) additional assumptions that (2) go beyond the minimal RCT assumptions are necessary.

As becomes apparent from this formulation, the fourth premise of the MTE argument has two parts, (1) and (2), which both need to be established for defending this premise. This chapter focuses on establishing (1) the necessity of *additional assumptions* for obtaining the joint distribution and thus policy-relevant distributional information from RCT data.

As hinted above, this first part of the fourth premise of the MTE argument is endangered by methodological advancements in the RCT literature. Even though defenders of the MTE argument often claim that making additional assumptions is the only way to learn about distributional parameters that depend on the joint distribution, the possibility of *bounding* the joint distribution calls this into question. Bounding allows us to obtain a range within which the true value of distributional parameters that depend on the joint distribution falls. Importantly, bounds do so solely based on information obtainable from the marginal distributions and thus without additional assumptions. Thereby, bounding provides a way of learning about policy-relevant distributional information conveyed by distributional parameters that depend on the joint distribution without requiring additional assumptions.

I discuss this first threat for the fourth premise of the MTE argument in *section one*. In *section two,* I then refute this threat. I do so by pointing out that the ranges of distributional parameters that depend on the joint distribution implied by bounding are too wide to be informative for policymakers. Narrowing them down, however, requires assumptions akin to those that allow for the identification of the joint distribution. This establishes (1) the necessity of additional assumptions for obtaining useful, policy-relevant distributional information from RCT data by obtaining the joint distribution.

1. *The Threat from Bounds*

1.1. Introducing Bounds

Let us start with grasping the first threat for the fourth premise of the MTE argument. For this, let us first introduce bounds in this subsection and then discuss why bounds shed doubts on the claim that additional assumptions are necessary to obtain the joint distribution and parameters depending on it.

Bounding is a way of learning about parameters, such as the distributional parameters that we are discussing. However, instead of providing us with a single value for the parameter of interest implied by the data, bounds give us a *range of values* within which the true value of the parameter of interest lies. For example, consider the median treatment effect, i.e., the effect on the individuals located at the median of the initial distribution. Also, suppose that the true value of the median treatment effect is 4. Bounding the median treatment effect would thus mean that we obtain a range of numbers, e.g. from RCT data, that contains 4, the true number, but also other numbers (Manski 2010, p. 178).

In addition, to provide us with information about the respective parameter, this range of values must be smaller than the logically possible range of values within which the parameter of interest, here, the median treatment effect, lies (*Ibid.*). For instance, if student achievement in a class-size reduction RCT was measured by a standardized test on which students can score any number from 0 to 100, the range of logically possible values for the median treatment effect lies between -100 and 100. The range obtained via bounding would thus have to be a subset of [-100; 100] that includes 4, the true number, for instance [-10; 10].

Importantly, bounds are not informative regarding *where* the true value of the parameter of interest is likely to lie within this range. That is, every value within the identified range is equally likely to be the true value of the parameter of interest. For instance, in the example above, it is equally likely that the students located at the median of the initial test score distribution experienced a decline of 10 points in their test score by learning in a smaller class or an increase of 10 points. Thus, bounds only rule out values *outside* of the obtained range, but they do not bring us closer to finding the true value of the parameter of interest *within* this range (Bedoya et al. 2017, pp. 17f.).

To see how bounds shed doubt on the claim made by the fourth premise of the MTE argument, let us establish another crucial feature of bounds obtained from RCT data. This feature has to do with *how* bounds narrow down the range of logically possible values for the parameter of interest to a more limited range. For this, the only information that is used is information that is obtainable from the *marginal distributions* (Djebbari and Smith 2008, p. 69).

To illustrate this, let us examine how a specific parameter that depends on the joint distribution can be bounded from RCT data only using information available from the marginal distributions and thus, without making additional assumptions: The variance of treatment effects[21]. Formally, this distributional parameter can be written as follows:

$$Var(TE) = Var(Y_1) + Var(Y_0) - 2\ Cov(Y_0, Y_1)$$

As this formal expression shows, the variance of treatment effects depends on the joint distribution because the covariance term depends on the joint distribution. Thus, to obtain this parameter from RCT data, we need to obtain the joint distribution, which requires additional assumptions. Alternatively, we can learn about this parameter by obtaining bounds for the covariance term. Thereby, we can determine bounds for the variance of treatment effects which depends on the covariance term.

For seeing how information obtainable from the marginal distributions allows us to do this, let us first decompose the parameter that we want to bound:

$$Cov(Y_0, Y_1) = \rho_{01}\sigma_0\sigma_1$$

In this formula, ρ_{01} denotes the correlation between Y_0 and Y_1. σ_0 and σ_1 denote the standard deviations of Y_0 and Y_1, respectively. Since σ_0 and σ_1 are features of the marginal distributions, they can be obtained from RCT data alone. Thus, the only remaining parameter that depends on the joint distribution is ρ_{01}, the correlation. However, we know that, per definition, any correlation must lie between -1 and 1. Therefore, without having to make any further assumptions, we know that:

$$-\sigma_1\sigma_0 \ \leq Cov(Y_0, Y_1) \ \leq \ \sigma_1\sigma_0$$

From having established that the true value of $Cov(Y_0, Y_1)$ must lie somewhere between $-\sigma_1\sigma_0$ and $\sigma_1\sigma_0$, we can obtain bounds on our distributional parameter of interest, the variance of treatment effects. That would look as follows:

[21] The discussion of this example is based on Bedoya et al. 2017, pp. 17ff. and Djebbari and Smith 2008, pp. 69f..

$$Var(Y_1) + Var(Y_0) - 2\sigma_1\sigma_0 \leq Var(TE) \leq Var(Y_1) + Var(Y_0) + 2\sigma_1\sigma_0$$

From this expression, we see that all parameters that restrict the range within which the true value of the variance of treatment effects falls are features of the marginal distributions, such as $Var(Y_1)$ or $2\sigma_1\sigma_0$. This example illustrates how the value of a distributional parameter that depends on the joint distribution can be bounded only based on information that is obtainable from the marginal distributions and thus without needing additional assumptions.

In a similar fashion, it is possible to bound other distributional parameters that depend on the joint distribution, such as the proportion of people who benefitted from or got harmed by a policy. In addition, we can use the marginal distributions to bound the entire joint distribution (Abadie and Cattaneo 2018, pp. 471ff.; Bedoya et al. 2017, p. 19; Djebbari and Smith 2008, p. 69; Fan and Park 2010, p. 933; Firpo and Ridder 2008, pp. 3f.).

We have thus seen that bounds enable us to restrict the range within which the true value of a distributional parameter that depends on the joint distribution falls solely based on the marginal distributions, and thus without requiring additional assumptions. Let us now discuss why this threatens the persuasiveness of the MTE argument.

1.2. How Bounds Question the Necessity of Additional Assumptions

As hinted, the possibility of bounding the joint distribution and parameters depending on it solely based on the marginal distributions raises doubts about whether we need additional assumptions to obtain distributional parameters that convey policy-relevant distributional information from RCT data. This possibility questions the persuasiveness of the MTE argument by shedding doubts on whether part (1) of its fourth premise holds up. Importantly, it does so even in light of our result from the last chapter. There, we established that policy-makers, under certain assumptions, require information from distributional parameters that depend on the joint distribution. However, if bounds sufficiently limit policy-makers' uncertainty about the information conveyed by these distributional parameters, the necessity of making additional assumptions for obtaining these parameters from RCT data is doubtful.

For instance, consider the effect of policies on the people at the median and the lowest quantile of the initial outcome distribution of interest. As established in the previous chapters, this distributional information can only be obtained from the joint distribution and is required by policymakers who care about the electorate's support for their policies. This is because policymakers can only attract altruistic and selfish voters if their policies benefit people at the median and the lowest quantile of the initial outcome distribution.

Now, if bounds on these two quantiles of the difference Y_1-Y_0 do not include any negative values, then the informational desire of these policymakers can be satisfied without having to make any additional assumptions. This is because bounds only require information from the marginal distributions, which are available from RCT data alone.

Thus, having shown that the third premise of the MTE argument holds up, i.e. that distributional parameters that depend on the joint distribution are necessary for obtaining policy-relevant information from RCT data, is not yet sufficient for establishing (1), i.e. that obtaining these parameters also requires additional assumptions. (1), however, figures into the fourth premise as one of the key premises of the MTE argument, which is thus threatened. In the following, let us see how this threat can be refuted.

2. *Refuting the Threat from Bounds*

To refute this threat, I will point to the empirically found wideness of bounds on distributional parameters that depend on the joint distribution. I will then argue that this wideness makes these bounds insufficiently informative for policymakers who require distributional information conveyed by distributional parameters that depend on the joint distribution. What is more, the only way in which we can make these bounds more informative is by making assumptions that are akin to those employed in assumption-based approaches to obtaining the joint distribution. Thereby, we can establish (1).

When bounding distributional parameters that depend on the joint distribution, a common finding by researchers is that the obtained bounds are *very wide.* That is, the information available from the marginal distributions alone does not seem to particularly limit the range within which the true value of the distributional parameter of interest falls (Bedoya et al. 2017, p. 17; Carneiro et al. 2003, p. 2; Djebbari and Smith 2008, p. 70; Heckman et al. 1997, p. 502).

For instance, when calculating bounds on the fraction of people who benefitted from the financial education program discussed in chapter 3, Bedoya et al. find that the pure RCT data is consistent with at least 12% of the individuals benefitting (lower bound) and at most 100% benefitting (upper bound) (Bedoya et al. 2017, p. 47). In other words, the data cannot rule out that a considerable fraction of students (i.e. 88%) experienced a decline or no change in their financial literacy due to the financial education policy. Nor can we exclude the possibility that all students improved their financial test scores[22].

In the following, I argue that this empirically found wideness of bounds on distributional parameters that depend on the joint distribution makes bounds *insufficiently informative* for policymakers who require distributional information conveyed by distributional parameters that depend on the joint distribution. For focusing the discussion, let us look at policymakers interested in voters' support. As discussed, this is a plausible and general assumption that makes the nuanced distributional information conveyed by distributional parameters that depend on the joint distribution relevant for policymakers. As we have seen, these policymakers need distributional information conveyed by the quantiles of the difference Y_1-Y_0. Thus, let us focus on bounds on this distributional parameter in the following.

As authors such as Abbring and Heckman point out, the phenomenon that bounds on distributional parameters that depend on the joint distribution are very wide also extends to bounds on the quantiles of the difference Y_1-Y_0 (Abbring and Heckman 2007, p. 5154). For an example that supports this claim, let us look at empirically obtained bounds on this distributional parameter, which can be found in a study conducted by Djebbari and Smith (2008).

In this study, the authors analyse data from the PROGRESA RCT. PROGRESA is a conditional-cash transfer program implemented to help Mexico's poor. That is, eligible households receive money conditional on e.g. sending their kids to school. The authors are interested in the effects of PROGRESA on consumption and particularly in the distributional impacts of the program on this outcome variable (*Ibid.*, p. 65).

To find out about the latter, the authors obtain bounds on several distributional parameters related to consumption, among those the quantiles of the difference Y_1-Y_0. For instance, bounding the effect on individuals at the 5th percentile of the initial consumption distribution, they find that the data is consistent with an increase in consumption of 2.27 and up to a

[22] The problem that bounds are often very wide is further aggravated due to sampling variation. That is, the marginal distributions which are used for obtaining these bounds are estimated from a sample of the population of interest instead of being the "true" distributions for this population. This creates sampling variation which makes the bounds obtained from the marginal distributions even wider (Bedoya et al. 2017, p. 20; Heckman et al. 1997, p. 502).

decrease in consumption of 413.64, measured in monetary units[23]. Other bounds obtained by the authors are similarly as wide (Djebbari and Smith 2008*,* pp. 70f.).

Now, this example illustrates how the empirically found wideness of bounds on distributional parameters that depend on the joint distribution makes these bounds insufficiently informative for policymakers interested in the voters' support for their policies. For this, recall that information regarding the effect of policies on the worst-off, represented by those at the lowest quantiles of the initial outcome distribution of interest, is of uttermost importance for policymakers interested in the support of the altruistic electorate: Should it turn out that the respective policy harms the worst-off, then policymakers cannot count on the support of the altruistic electorate.

Based on these considerations, we can see why the information obtainable from the bounds on the effect on individuals at the 5th quantile of the initial consumption distribution calculated by Djebbari and Smith is not helpful for policymakers who are interested in the support of voters and consider implementing PROGRESA: From these bounds, they only learn that the program might improve the consumption of the worst-off, considerably worsen it, or anything in between. Since all these possibilities arise with equal probability, the bounds provide no grounds for policymakers competing for the votes of the altruistic electorate to decide on the implementation of PROGRESA.

This discussion of bounds on the quantiles of the difference Y_1-Y_0 indicates that the empirically found wideness of bounds on distributional parameters that depend on the joint distribution makes these bounds insufficiently informative for policymakers who require distributional information conveyed by these distributional parameters because they are interested in the support of the voters.

One might object to this claim by arguing that it is based on the empirically found wideness of bounds, which is not given. Especially, there are ways of *tightening* bounds on parameters that depend on the joint distribution (Bedoya et al. 2017, p. 16). If these bounds could be

[23] To have some context for interpreting these results, the monthly average per capita consumption in the data of Djebbari and Smith was 204.47 (measured in monetary units) (Djebbari and Smith 2008, p. 67). Thus, the bounds obtained by the authors are even compatible with negative consumption resulting from PROGRESA, which would mean that households must sell off instead of buying items.

sufficiently tightened, they might reveal sufficient distributional information for policymakers, unlike the examples discussed above.

To see how the bounds on distributional parameters that depend on the joint distribution can be tightened beyond the ranges found in the studies discussed above, re-consider the example of bounding the variance of treatment effects[24]. As we have seen, this distributional parameter can be bounded by bounding the covariance between the outcome in the treated and the untreated state, which looks as follows:

$$-\sigma_1\sigma_0 \leq Cov(Y_0, Y_1) \leq \sigma_1\sigma_0$$

We also know that:

$$Cov\ (Y_0, Y_1) = \ \rho_{01}\sigma_1\sigma_0$$

In addition, recall that:

$$-1 \leq \rho_{01} \leq 1$$

The bounds on the covariance, and thus on the variance of treatment effects, could be tightened by assuming a smaller range for ρ_{01}. For instance, we could assume that potential outcomes are positively correlated, that is, that people who do well in the absence of the policy also do well when exposed to the policy. Formally, that would imply:

$$0 \leq \rho_{01} \leq 1$$

Then, the bounds on the covariance reduce to:

$$0 \leq Cov(Y_0, Y_1) \leq \sigma_0\sigma_1$$

That way, we can obtain tighter bounds on distributional parameters that depend on the joint distribution, such as the variance of treatment effects. Doing so would increase the informativeness of bounds for policymakers who require distributional information conveyed by distributional parameters that depend on the joint distribution. Potentially, these tightened bounds might then provide policymakers with all the information they require.

However, if we go down this road, we are using more than the information solely available from the marginal distributions. That is, in contrast to the wide bounds discussed above, the

[24] The discussion of this example is based on Bedoya et al. 2017, p. 18.

tighter bounds that are potentially more informative for policymakers are based on additional assumptions.

What is more, the assumptions based on which we can tighten the bounds on distributional parameters that depend on the joint distribution are also, in slightly different versions, those that allow us to obtain the joint distribution and thereby the exact value of these distributional parameters (cf. Bedoya et al. 2017, pp. 18ff.).

To see this, consider the example of the positive correlation assumption that allowed us to tighten the bounds on the variance of treatment effects in the example above. This assumption postulates a positive correlation between individuals' outcomes in the state of treatment and control. Thereby, this assumption establishes a relationship between individuals' outcomes in the state of treatment and control. As we have seen in chapter 2, establishing such a relationship is what assumptions that allow us to directly obtain the joint distribution and distributional parameters depending on it do. This makes the assumptions that allow us to recover the joint distribution and those that allow us to tighten bounds on distributional parameters that depend on the joint distribution akin to each other[25].

From this discussion, the following situation emerges: Empirically found bounds that solely employ information from the marginal distributions are so wide that they do not sufficiently inform policymakers who require distributional information conveyed by distributional parameters that depend on the joint distribution. However, tightening these bounds to increase their informativeness requires additional assumptions akin to the assumptions that give us the exact value of the respective distributional parameters. Thus, the only way to make bounds sufficiently informative for policymakers interested in voters' support is making assumptions akin to those employed in assumption-based approaches to obtaining the joint distribution.

Conclusion

In this chapter, we examined a threat for part (1) of the fourth premise of the MTE argument and thus for its persuasiveness. To this end, I introduced bounds as a method of learning about distributional parameters that depend on the joint distribution without making additional assumptions. To refute this threat, I have examined the informativeness of empirically obtained

[25] The assumption that allows us to recover the joint distribution and is akin to the positive correlation assumption is the Perfect Positive Dependence Assumption, which I will introduce in the next chapter.

bounds on distributional parameters that depend on the joint distribution for policymaking. As discussed, these bounds are too wide to be informative for policymakers who require distributional information solely obtainable from the joint distribution because they are interested in voters' support. Tightening them, however, is only possible by making assumptions akin to those that allow us to directly obtain the joint distribution and distributional parameters depending on it.

Given this, we can establish assumption-based approaches to obtaining the joint distribution as the only way to satisfactorily limit policy-makers' uncertainty about distributional information conveyed by distributional parameters that depend on the joint distribution based on RCT data. Thereby, a first threat for the fourth premise of the MTE argument can be refuted.

Chapter 6

The Threat from Weak Assumptions

Introduction

Having refuted the first threat for the fourth premise and thus for the persuasiveness of the MTE argument, I want to highlight a potential second and final threat for the fourth premise of the MTE argument. This threat thus remains to be refuted before being able to establish the persuasiveness of the MTE argument. I introduce this threat in *section one*, where I point to the possibility that the additional assumptions that allow us to obtain the joint distribution and distributional parameters depending on it might be *weak*. In this case, I argue, the MTE argument cannot refute the EBP proponents' claim that RCTs only require minimal substantive assumptions to produce policy-relevant evidence.

In *section three,* I then analyse the strength of this threat against the persuasiveness of the MTE argument. I do so by analysing whether the most frequently employed assumptions that allow us to recover the joint distribution – i.e. the Perfect Positive Dependence Assumption (short: PPDA), the Perfect Negative Dependence Assumption (short: PNDA) and the Independence Assumption – are weak. In this analysis, I use an operationalization of what counts as a weak assumption for EBP proponents themselves that I develop in *section two.* As I argue, one of these three assumptions – the PPDA – would indeed be seen as weak in the eyes of EBP proponents. This poses a threat to the fourth premise of the MTE argument.

In *section four,* I refute this final threat. I do so by arguing that the PPDA does not provide policymakers with the distributional information they require. In doing so, I defend the fourth premise of the MTE argument and thus its persuasiveness as an argument in support of the philosophical case against the special role of RCTs for EBP.

1. A Potential Final Threat for the Persuasiveness of the MTE Argument

In this section, I raise a potential final threat against the fourth premise of the MTE argument and thus its persuasiveness. To formulate this threat, let us recap the fourth premise:

> ***P4:*** To obtain policy-relevant distributional information conveyed by distributional parameters that depend on the joint distribution from RCT data, (1) additional assumptions that (2) go beyond the minimal RCT assumptions are necessary.

The potential final threat discussed in this chapter arises from doubts regarding (2), which is essential for the MTE argument's persuasiveness. To grasp this threat, recall that, as discussed in chapter 2, the MTE argument needs to refute the EBP proponents' claim that RCTs only require *minimal substantive assumptions* to produce evidence for policymaking. Only then, the MTE argument can support the philosophical case against the special role of RCTs for EBP, which hinges on attacking this proposition of the EBP proponents' RCT argument.

When spelling out the proposition of the EBP proponents' RCT argument that the MTE argument needs to refute, it becomes apparent how important (2) is for the persuasiveness of the MTE argument: As we saw, EBP proponents claim that, contrary to the strong and often incredible assumptions employed by other methods, the few assumptions required by RCTs to produce evidence for policymaking are *less strong*, but more of a general nature and thus also more credible (e.g. Reiss 2013, p. 202). From this formulation, it becomes apparent that the claim of EBP proponents that RCTs only require minimal substantive assumptions should not primarily be understood in a quantitative sense. Instead, a lot of emphasis is put on claiming that these assumptions are relatively weak and thus minimal in a more qualitative sense.

Thus, having shown (1), i.e. that additional assumptions are necessary for obtaining distributional information of interest to policymakers from RCT data, is not enough for establishing the fourth premise and thus the persuasiveness of the MTE argument. (1) only proves that RCTs require *more* assumptions to produce policy-relevant distributional information than to produce mean treatment effects. That is, we know that the quantity of assumptions required to obtain policy-relevant distributional information from RCT data is larger than the quantity of assumptions needed to obtain mean treatment effects from RCT data. However, it does not show that these additional assumptions also go beyond the minimal RCT assumptions in a qualitative sense, i.e. because they are *strong*.

That is, it might be that very weak assumptions allow us to recover the joint distribution. Then, (1) would not shed doubts on the EBP proponents' claim that RCTs only require minimal, in the sense of weak, assumptions for providing policy-relevant evidence. This would threaten the fourth premise and thus the persuasiveness of the MTE argument.

For refuting this potential final threat against the persuasiveness of the MTE argument, we thus need to show (2), i.e. that the assumptions required for obtaining the joint distribution are strong and thus go beyond the minimal RCT assumptions. To do so, however, we first need to operationalize what makes for a weak or strong assumption in the context of assumptions employed in RCTs that deliver results for EBP.

2. Operationalizing Strong and Weak Assumptions in the Context of EBP

For this operationalization, let us look at what EBP proponents themselves consider to be a strong or weak assumption. Doing so will allow us to judge whether the assumptions with which the joint distribution can be recovered from RCT data would count as strong and thus go beyond the minimal RCT assumptions, according to standards employed by EBP proponents themselves. Should that turn out to be the case, then the MTE argument would weaken the case of EBP proponents for a special role of RCTs for producing evidence for policymaking. It would do so by showing that even EBP proponents themselves would disagree that the assumptions necessary to obtain policy-relevant evidence conveyed by distributional parameters that depend on the joint distribution from RCT data are minimal, in the sense of being weak.

A tentative answer to the question of what counts as a strong or weak assumption for EBP proponents can be found in a paper co-authored by Esther Duflo, one of the most prominent EBP proponents. In this paper, Duflo and her co-authors discuss a new machine learning method for analysing data from RCTs. In doing so, the authors continuously emphasize that they are taking an "agnostic approach" that does not rely on "strong assumptions" (Chernozhukov et al. 2018, p. 5).

From their further specification of their agnostic approach, we can extract a tentative idea of what a strong or weak assumption means for EBP proponents:

> In this paper we take an agnostic view. We neither rely on any structured assumptions, which might be *difficult to verify or believe in practice* […] (*Ibid.*, emphasis M.S.).

From this quote, we can infer that EBP proponents would think of assumptions that are "difficult to verify" or "difficult to believe in practice" as assumptions that are too strong to fit their agnostic view. And while the authors do not specify what they mean with "difficult to verify" or "difficult to believe in practice", it seems plausible to operationalize an assumption that is "difficult to verify" as an assumption that is not testable. An assumption that is "difficult to believe in practice" most likely means an assumption that either clashes with common intuitions when applied or has so much empirical evidence against it that it cannot be believed. Assumptions that are both verifiable and believable would, in turn, be counted as weak[26].

Having operationalized the notion of a weak assumption in the context of EBP, the next step is to analyse whether the assumptions that allow us to recover the joint distribution would be counted as weak given this operationalization. Should that be the case, then (2) would be false because the assumptions that allow us to obtain policy-relevant distributional information from RCT data do not go beyond the minimal RCT assumptions because they are minimal in the sense of being weak. This would threaten the persuasiveness of the MTE argument.

3. *Analysing the Assumptions that Allow us to Recover the Joint Distribution*

To analyse whether the assumptions that allow us to obtain the joint distribution would be counted as weak according to the criteria defined above, let us first introduce the most commonly employed assumptions for obtaining the joint distribution.

For introducing these assumptions, I use the PROGRESA study by Djebbari and Smith (2008), which we have discussed in the last chapter. Inspired by their result that bounds on distributional parameters that depend on the joint distribution are too wide, the two authors take an *assumption-based approach* to analysing the distributional impacts of PROGRESA. In doing so, they use all commonly employed assumptions for recovering the joint distribution: The PPDA, the PNDA and the Independence Assumption (Djebbari and Smith 2008, pp. 69f. and p. 75).

Having presented these assumptions using Djebbari and Smith's study, I then argue that only one of these assumptions – the PPDA – is weak according to the understanding of EBP proponents. That is, the PPDA is the only assumption that does not lack initial plausibility, is supported by empirical evidence and can be tested. This poses a threat to the fourth premise of the MTE argument.

[26] For another formulation of the same view on what makes for a strong assumption in the eyes of EBP proponents, consider this quote from the back cover summary of "Poor Economics", a book on EBP ideas co-authored by Duflo: "Billions of government dollars […] are dedicated to helping the world's poor. But much of the work they do is based on *assumptions that are untested generalizations at best, flat out harmful misperceptions at worst*" (Banerjee and Duflo 2011, summary back cover, emphasis M.S.).

3.1. *The Perfect Positive and the Perfect Negative Dependence Assumption*

Let us start by introducing the assumptions that allow us to recover the joint distribution using Djebbari and Smith's study. The first two assumptions discussed by the authors can be illustrated using the following graph, which we have already seen in chapter 2:

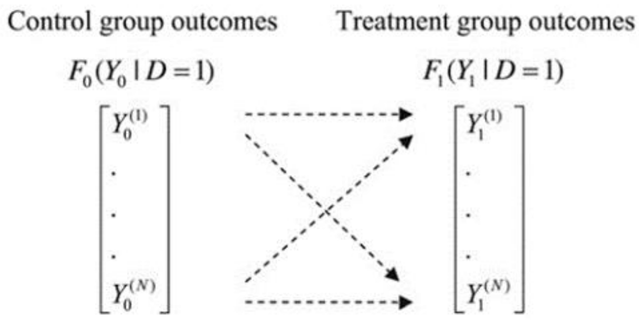

Figure 3: Illustration PPDA and PNDA

In chapter 2, we used this graph to explain why the joint distribution cannot be obtained from RCT data alone, which only gives us the marginal distributions. As discussed, this is because RCT data does not tell us where in the treatment distribution an experimental subject from the control group would appear, and vice versa.

From this discussion in chapter 2, we have also seen *which type* of assumptions would allow us to recover the joint distribution. As explained, these are assumptions that establish a relationship between individuals' outcomes in the treatment and the control group. Thereby, it can be determined where in the marginal distribution of the treatment group an experimental subject of the control group would appear and vice versa. Doing so gives us the outcome for each subject for both states and thus the joint distribution. Two assumptions of this type are the first two assumptions discussed by Djebbari and Smith. These assumptions can also be illustrated using the graph above and are depicted by the horizontal and the crossing arrows, respectively.

Let us start by discussing the assumption depicted by the horizontal arrows, which Djebbari and Smith refer to as the "Perfect Positive Dependence Assumption". The PPDA establishes the relationship between individuals' outcomes in the state of treatment and the state of control required for obtaining the joint distribution by assuming a perfect ranking across the quantiles of the two marginal distributions. That is, it is assumed that the quantile ranks of the two marginal distributions *correspond.* This would, for instance, imply that a person who appears in the highest percentile of one marginal distribution also appears in the highest percentile of the other marginal distribution (Djebbari and Smith 2008, p. 69; Abbring and Heckman 2007, p. 5159).

The two crossing arrows depict a second assumption for recovering the joint distribution, which works similarly to the PPDA and is referred to as the "Perfect Negative Dependence Assumption". While the PPDA assumes a perfect positive dependence of quantile ranks, such that the best in one distribution is also the best in the other distribution, the PNDA assumes the opposite to be true. As the crossing arrows indicate, according to the PNDA, the quantile ranks of the two marginal distributions are *inversely related*, such that those who do best when treated do worst when not treated and vice versa (Djebbari and Smith 2008, p. 69).

In the example of PROGRESA, this would mean that those appearing in the lowest quantile of the consumption distribution if not treated would be assumed to appear in the highest quantile of the consumption distribution if treated (*Ibid.*). However, just as the PPDA, the PNDA recovers the joint distribution by linking the quantile ranks of the two marginal distributions in a deterministic fashion.

So far, we have introduced the PPDA and the PNDA as two assumptions that recover the joint distribution by establishing a relationship between individuals' outcomes in the state of treatment and control. This is achieved by assuming a deterministic correlation between quantile ranks of the marginal distributions. Before turning to discuss whether these assumptions are weak, let us turn to a third assumption used by Djebbari and Smith in their PROGRESA study.

3.2. The Independence Assumption

According to this third assumption, the impact of a policy on the experimental subjects is *independent* of their untreated outcome level, i.e. Y_0. I shall thus refer to this assumption as the "independence assumption". In the PROGRESA example, and if consumption is the relevant outcome variable, assuming independence of the impact from the base outcome would mean that a participant's gains or losses from the cash transfer program are not related to her initial consumption level (*Ibid.*, p. 75).

Contrary to the assumptions discussed above, the independence assumption does not establish a link between the two marginal distributions. Instead, it assumes something about the *impact* and its relation to the untreated state. Thereby, the independence assumption directly identifies the distribution of impacts (i.e. $F(Y_1\text{-}Y_0)$)[27] on which parameters such as the quantiles of the difference $Y_1\text{-}Y_0$ depend. In contrast, the PPDA and the PNDA identify the joint distribution (i.e. $F(Y_0,Y_1)$) and thereby the distribution of impacts, which is implied by the joint distribution (cf. Bedoya et al. 2017, pp. 21ff.).

However, as Djebbari and Smith point out, the independence assumption *alone* does not give us the full distribution of impacts but only allows us to obtain the variance of treatment effects. To obtain the full distribution of impacts, additional assumptions about the form of the impact distribution would be necessary. For instance, one could assume that the impact is normally distributed (Djebbari and Smith 2008, p. 75; Heckman et al. 1997, p. 504). This sets the independence assumption apart from the PPDA and the PNDA, which identify the full joint distribution and thus also the full impact distribution.

As the discussion above clarifies, obtaining the full impact distribution is essential for obtaining policy-relevant distributional information from RCT data. This is because identifying the full impact distribution is required for obtaining the quantiles of the difference $Y_1\text{-}Y_0$ (Bedoya et al. 2017, p. 55; Heckman et al. 1997, p. 514). As established, this distributional parameter conveys distributional information relevant to all policymakers interested in voters' support. Thus, to obtain this generally policy-relevant distributional information from RCT data using the assumption-based approach, these assumptions must recover the whole impact distribution. However, the independence assumption only does so in conjecture with additional assumptions about the form of the impact distribution.

In the following, I argue that these additional assumptions needed to obtain the full impact distribution based on the independence assumption would be counted as strong by EBP proponents. This eliminates one potential threat to the persuasiveness of the MTE argument. In addition, I argue that also the PNDA would be counted as strong because it lacks initial plausibility. The PPDA, however, fulfils the criteria for a weak assumption employed by EBP proponents. Thus, it sheds doubts on (2), thereby posing a threat to the persuasiveness of the MTE argument.

[27] How the independence assumption recovers the impact distribution is not relevant for the discussion at hand, but I refer the interested reader to a clear discussion on this issue in Bedoya et al. (2017).

3.3. *The PPDA Qualifies as a Weak Assumption*

Let us start by discussing whether the independence assumption qualifies as weak according to the criteria employed by EBP proponents. To recall, according to these criteria, a weak assumption is testable and supported by empirical evidence as well as common intuitions.

At first glance, it seems that the independence assumption is a weak assumption according to these criteria: As for instance Bedoya et al. point out, a good case can be made for the plausibility of the independence assumption in many practical contexts. This is especially because the independence assumption only needs to hold conditional on observables. That is, the impact of a policy only has to be unrelated to the base outcome after having controlled for covariates such as age or gender (Bedoya et al. 2017, p. 21).

However, we have also seen that additional assumptions about the form of the impact distribution would be required for obtaining the full impact distribution based on the independence assumption. As outlined, these additional assumptions are crucial because the full impact distribution is required for obtaining the quantiles of the difference Y_1-Y_0, which give us distributional information that many policymakers plausibly require.

In contrast to the independence assumption, these additional assumptions seem to fare less well when applying the criteria for a strong assumption employed by EBP proponents. To see this, consider the following quote from Bedoya et al., who use the independence assumption for distributional analysis:

> [W]e can identify the entire distribution of treatment effects [i.e. the impact distribution, M.S.] if we are willing to make even stronger assumptions. In many applications, these assumptions are unjustified. We include this analysis here primarily for the purpose of illustration (*Ibid.*, p. 54).

While the authors grant that the independence assumption itself is plausible in many contexts, the quote indicates that this plausibility does not extend to the assumptions about the form of the impact distribution, which are necessary for recovering the whole distribution of impacts. As becomes apparent from the quote, these assumptions are judged to be unjustified in many applications. This judgment seems akin to claiming that these assumptions are difficult to believe in practice, which we operationalized as meaning that they lack initial plausibility or are not supported by empirical evidence. Thus, these additional assumptions, which are the only way in which the independence assumption can be used to obtain the quantiles of the difference Y_1-Y_0, would count as strong given the criteria employed by EBP proponents.

From this discussion, we can conclude that these additional assumptions make the independence assumption, which only does its job in conjecture with these assumptions, too strong to be weak given the criteria employed by EBP proponents. That is, recovering distributional parameters that convey distributional information that is plausibly relevant for many policymakers based on the independence assumption requires assumptions that go beyond the minimal RCT assumptions on which the RCT argument of EBP proponents rests.

The same can be argued about the PNDA, which, in contrast to the independence assumption, recovers the full joint distribution, and thereby the full impact distribution, directly. As explained, it does so by assuming that those who do best when treated do worst when not treated and vice versa.

Following the discussion of this assumption by Djebbari and Smith, we can conclude that the PNDA, just as the independence assumption, would be counted as strong according to the criteria used by EBP proponents. As established, one of these criteria for a strong assumption was the lack of initial plausibility. According to Djebbari and Smith, this holds for the PNDA. As the two authors put it, "this assumption about the joint distribution has little relevance to the real world" as it "lacks surface plausibility" (Djebbari and Smith 2008, p. 69).

To illustrate their strong intuition against the plausibility of the PNDA with an example, consider again the effect of PROGRESA as a conditional cash transfer program on consumption. According to the PNDA, those at the highest quantiles of the consumption distribution in the absence of PROGRESA are those at the lowest quantiles when participating in PROGRESA and vice versa. That is, receiving conditional cash transfers is assumed to move people from the highest to the lowest quantile of the consumption distribution, which indeed seems implausible.

So far, we have seen that both the independence assumption and the PNDA would be counted as strong assumptions according to criteria used by EBP proponents. Thus, these assumptions comply with part (2) of the fourth premise of the MTE argument, i.e. that assumptions beyond the minimal RCT assumptions are needed to obtain the joint distribution and thus policy-relevant distributional information from RCT data.

However, the case is different for the PPDA, which would be counted as weak no matter which of the criteria operationalized above is used: Firstly, the PPDA is, contrary to the PNDA, ascribed initial plausibility by Djebbari and Smith: As the authors argue, assuming that those who do well in the presence of the policy also do well without the policy is plausible for policies for which we do not expect a huge effect (Djebbari and Smith 2008, p. 69 and p. 71). For instance, it might indeed be plausible to assume that those who already had a lot of consumption in the absence of PROGRESA will continue to have a high consumption when receiving PROGRESA's conditional cash transfers.

Secondly, at least the implications of the PPDA are testable in different ways. From one of these indirect tests, we can, thirdly, derive empirical evidence for the PPDA that adds to its initial plausibility[28]: As Heckman et al. (1997) show, plausible results for distributional parameters that depend on the joint distribution require associating the quantiles of the two marginal distributions as assumed by the PPDA. That is, associating the quantiles in ways that venture too far from perfect positive dependence produces implausible values for distributional parameters such as the proportion of people who benefitted. This evidence thus supports the plausibility of the PPDA[29] (*Ibid.*, p. 506).

It thus follows that the PPDA qualifies as what I have called a "weak assumption" that would be seen as akin to the minimal RCT assumptions celebrated by EBP proponents. This is because the PPDA is testable, initially plausible and backed up by empirical evidence. Thus, obtaining policy-relevant distributional information based on RCT data and using the PPDA does not require assumptions that are not "minimal" in the eyes of EBP proponents. That is, adding the PPDA to the classical RCT assumptions does not refute the EBP proponents' claim that these assumptions, which give us all evidence of relevance to policymakers, are minimal.

This poses a threat to the persuasiveness of the MTE argument: Even though we have shown that the joint distribution is required for obtaining policy-relevant distributional information from RCT data and that (1) obtaining the joint distribution requires additional assumptions, the case for the PPDA indicates that one of these assumptions is weak in the eyes of EBP proponents. Therefore, the case of the PPDA suggests that (2) is false. Then, the MTE argument could not question the EBP proponents' claim that RCTs only require minimal assumptions for

[28] There is another source of empirical evidence for the PPDA which is evidence for decision rules which govern the decision to participate in a program. These decision rules induce the kind of dependence which the PPDA requires and thus, empirical evidence that these decision rules are followed (which exists) provides additional support for the PPDA (Heckman et al. 1997, p. 489).

[29] At the same time, this is empirical evidence against the PNDA (*Ibid.*, p. 506).

producing policy-relevant evidence. In the next section, I refute this threat and establish (2) and thus the fourth premise and the persuasiveness of the MTE argument.

4. Refuting the Final Threat for the Persuasiveness of the MTE Argument

To refute this threat, I argue that even though the PPDA is a minimal assumption in the eyes of EBP proponents, it does not provide policymakers with the distributional evidence that they require. As I explain, this is because an analysis of RCT data that involves the PPDA cannot satisfy what we have established as the most plausible and general informational desire of policymakers that can only be addressed based on the joint distribution, i.e. wanting to know whether a policy is beneficial for attracting voters. This is because an analysis of RCT data based on the PPDA will, per assumption, yield the answer that any policy has negligible effects on voters at the median and the lowest quantile of the initial outcome distribution and is thus not beneficial for attracting voters, instead of revealing the answer implied by the RCT data.

This result establishes the fourth premise of the MTE argument, which claims that (2) assumptions beyond the minimal RCT assumptions are required for obtaining the joint distribution and thus policy-relevant distributional information from RCT data. Having excluded the PPDA as an assumption that provides policymakers with the distributional information they require, only the PNDA and the independence assumption remain. However, we have seen that those two assumptions would be regarded as strong by EBP proponents and thus go beyond the minimal RCT assumptions, thereby confirming the central claim of the MTE argument.

4.1. The PPDA Does Not Provide Policy Makers with the Information they Need

In this subsection, I further specify the informational desire of policymakers that can only be satisfied based on assumption-based approaches to obtaining the joint distribution. Then, I establish that this desire cannot be satisfyingly addressed based on RCT data in conjunction with the PPDA.

To this end, let us recall our discussion in chapter 4. There, we established that under the condition that policymakers are interested in the support of the electorate, they require distributional information that can only be obtained from the joint distribution and thus, as discussed in the last chapter, based on additional assumptions like the PPDA. We also saw that this assumption about the interests of policymakers due to which they require distributional information that can solely be obtained from the joint distribution is both general and plausible. Thus, let us base the discussion on the informational desire of policymakers that can only be addressed by obtaining the joint distribution on this assumption in the following.

As we saw, the informational desire of policymakers who comply with this assumption consists in wanting to know about the effect of a policy on voters at the *median and the lowest quantile* of the initial outcome distribution of interest. Based on this information, they can tell whether implementing the policy will help them gain votes from altruistic and selfish voters.

Let us further specify this informational desire of policymakers before arguing that it cannot be sufficiently addressed based on an analysis of RCT data that involves the PPDA. For this, take the example of a policymaker interested in attracting altruistic voters. As we have seen, she needs evidence on the effect of the policy in question on people at the lowest quantile of the initial outcome distribution. Only if she learned from RCT data that this group benefits she would know that implementing the policy would be beneficial for attracting altruistic voters.

Importantly, these voters would only be attracted if the benefits for the people at the lowest quantile of the initial outcome distribution are *visible* for them. Should this group only experience minor changes, then the altruistic voters might miss that the policy helped this group. Then, implementing the policy would not help the policymaker to gain votes from the altruistic electorate. Thus, our policymaker is interested in learning whether her policy would have non-negligible effects on people at the lowest quantile of the initial outcome distribution.

We can thus specify the informational desire of policymakers who require distributional information that is solely obtainable from the joint distribution and thus based on additional assumptions because they are interested in the support of voters as follows: They want to know, based on RCT data, whether the policy in question has *non-negligible* effects on voters at the median or the lowest quantile of the initial outcome distribution.

In the following, I show that an analysis of RCT data that involves the PPDA cannot satisfy this specified informational desire of policymakers. This is because the PPDA already implies a specific answer regarding this informational desire by implying that policies only have negligible effects instead of revealing the answer implied by the RCT data.

As we have seen above, the PPDA assumes that policies only have a minor effect, such that those in a given quantile of the initial outcome distribution also appear in the same quantile in the post-policy distribution. For instance, the PPDA would imply that people at the lowest quantile of the initial outcome distribution are also in this quantile in the post-policy distribution. Thereby, the PPDA implies that policies have *negligible effects*: For instance, the lot of the worst-off would not be improved substantially by a policy if these people would still be in the lowest quantile of the relevant outcome distribution post-policy.

By implying negligible effects, the PPDA already implies a *specific answer* to the informational desire of policymakers specified above. While these policymakers want to know, based on RCT data, *if* the policy has non-negligible effects on voters at the median or the lowest quantile of the initial outcome distribution, obtaining this information from RCT data with the PPDA will, per assumption, yield the answer that this is not the case. Thereby, using the PPDA to address this specified informational desire of policymakers based on RCT data already implies a specific answer regarding this desire instead of providing a way to learn the answer implied by RCT data.

This result, in turn, implies that distributional analyses of RCT data based on the PPDA cannot satisfyingly address this informational desire of policymakers. To illustrate this point, consider our policymaker who hopes to attract the altruistic electorate. Motivated by this, she requests information on whether the effect of her policy on the worst-off is non-negligible from the experimental evaluation of the policy. As we have seen, this information can only be obtained from RCT data by obtaining the joint distribution. In addition, the only weak assumption for this purpose is the PPDA. However, as we just saw, addressing the informational desire of our policymaker based on the PPDA will, per assumption, yield the answer that the policy had a negligible effect on the worst-off. Thereby, analysing the distributional question from our policymaker using the PPDA already implies a specific answer to her question instead of providing a way to learn the answer implied by the RCT data, which is what the policymaker wanted to know.

We have thus established that even though the PPDA is a weak assumption, analyses of RCT data that employ the PPDA do not satisfyingly address the informational desire of policymakers who require distributional information that is solely obtainable from the joint distribution. In other words, the distributional information that policymakers obtain from such an analysis is not the information that they require.

In the following, I argue that this finding helps us to reject the threat against (2) and thus the fourth premise of the MTE argument discussed in this chapter. Having rejected this final threat, I establish the persuasiveness of the MTE argument.

4.2. Rejecting the Final Threat

In this subsection, I argue that, because the PPDA does not provide policymakers with the distributional information they require, we can exclude the PPDA from the set of assumptions that proponents of the MTE argument are concerned with. As we have seen, the MTE argument claims that (2) the assumptions required to obtain distributional information of relevance to policymakers based on RCT data by obtaining the joint distribution go beyond the minimal RCT assumptions. However, since the PPDA does not deliver the distributional information that policymakers require, as established above, we can exclude the PPDA from the assumptions that the MTE argument refers to.

This only leaves the PNDA and the independence assumption as potential candidates for assumptions for obtaining the joint distribution and thus policy-relevant distributional information from RCT data. However, these two assumptions would be considered too strong by EBP proponents to qualify as minimal assumptions. Therefore, these assumptions indeed go beyond the minimal RCT assumptions. Thus, proponents of the MTE argument are correct in claiming that the assumptions required to obtain distributional information of relevance to policymakers from RCT data go beyond the minimal RCT assumptions. This establishes (2) and thereby the fourth premise of the MTE argument.

Conclusion

In this chapter, we have examined a final threat for the fourth premise of the MTE argument, and thus its persuasiveness: That the assumptions that recover the joint distribution might be weak. This possibility endangers the persuasiveness of the MTE argument because it would imply that no assumptions that go beyond the minimal RCT assumptions are required for obtaining the joint distribution and, thus, policy-relevant distributional evidence from RCT data.

As established, one assumption that recovers the joint distribution would count as weak according to criteria employed by EBP proponents, i.e. the PPDA. Thus, providing policy-relevant distributional information based on RCT data by obtaining the joint distribution with the PPDA does not require assumptions that go beyond the minimal RCT assumptions. However, we established that the PPDA does not provide policymakers with the distributional information they require.

Thus, we can exclude the PPDA from the set of assumptions that allow us to obtain distributional information required by policymakers, which is conveyed by distributional parameters that depend on the joint distribution, from RCT data. This only leaves the PNDA and the independence assumption, about which we have established that they would be considered to be strong by EBP proponents.

Thus, proponents of the MTE argument are right in making the following claim reflected in the fourth premise of the MTE argument: The additional assumptions required to obtain policy-relevant evidence conveyed by distributional parameters that depend on the joint distribution from RCT data go beyond the minimal RCT assumptions. Thereby, we can establish the fourth premise of the MTE argument and thus its persuasiveness.

Conclusion

In this thesis, I have defended the persuasiveness of the MTE argument as an argument that can strengthen the philosophical case against the special role of RCTs for EBP. As outlined, the MTE argument threatens the special role that EBP proponents ascribe to RCTs by questioning the minimality of assumptions that RCTs require for producing policy-relevant evidence.

As outlined, this alleged minimality of assumptions is also the focus of the external validity argument, which grounds the philosophical case against the special role of RCTs for EBP. Given this parallel, I have argued in the second chapter that the MTE argument could strengthen the philosophical case against the special status of RCTs for EBP, if shown to be persuasive.

As outlined, support from the potentially persuasive MTE argument would be helpful for the philosophical case against the special role of RCTs for EBP. As we saw, this case is endangered because the key premise of the external validity argument is not broadly accepted beyond the philosophical discussion. In addition, we discussed that the MTE argument could support this philosophical case in cases in which the external validity argument would not be able to, e.g. when the trial sample is known to be a representative sample of the population of interest.

To defend the MTE argument's persuasiveness and strengthen the philosophical case against the special role of RCTs for EBP, I have focused on defending the third and the fourth premise of the MTE argument. These premises are essential for using the MTE argument as an argument for the philosophical case against the special role of RCTs for EBP, which requires questioning the alleged minimality of assumptions that RCTs employ for producing policy-relevant evidence. At the same time, methodological advancements in the RCT literature threaten these essential premises.

To defend these premises, I have started by defending the third premise in chapter 4. That is, I have argued that obtaining the *joint distribution* is necessary for obtaining policy-relevant information conveyed by distributional parameters from RCT data, even in light of methodological advancements such as subgroup means and quantile treatment effects. To show this, I argued that even though subgroup means and quantile treatment effects provide policymakers with a host of distributional information, information on the distribution of individual-level impacts, such as the effect of a policy on individuals located at given quantiles of the initial outcome distribution, is missed by them. However, I have outlined that this distributional information can be argued to be relevant for policymakers under the general and plausible assumption that policymakers are interested in the support of voters.

I have then defended the fourth premise, i.e. that *additional assumptions beyond the minimal RCT assumptions* are necessary for obtaining this policy-relevant distributional information from the joint distribution. To this end, I have defended (1), that additional assumptions are necessary for obtaining the joint distribution against bounding as a possibility of learning about distributional parameters that depend on the joint distribution without making additional assumptions in chapter 5. Subsequently, I have defended (2), that the only assumptions that would allow us to obtain the distributional information that policymakers require from RCT data by obtaining the joint distribution would be seen as strong by EBP proponents, and thus go beyond the minimal RCT assumptions that allow for the identification of mean treatment effects, in chapter 6.

By defending the third and fourth premise of the MTE argument in light of methodological advancements in the RCT literature, I have argued for the MTE argument's persuasiveness as an argument supporting the philosophical case against the special role of RCTs for EBP. The third and fourth premise of the MTE argument establish that assumptions beyond the minimal RCT assumptions are necessary to obtain policy-relevant distributional information conveyed by distributional parameters that depend on the joint distribution from RCT data. From this result, in conjecture with the first and the second premise of the MTE argument, the conclusion that we need to move beyond the minimal RCT assumptions on which the EBP proponents' defence of a special role of RCTs is based for obtaining policy-relevant evidence from RCT data follows. Therefore, the MTE argument can strengthen the philosophical case against the special role of RCTs for EBP.

From this conclusion, the following questions for further research emerge: Firstly, it remains to be discussed whether other policy evaluation methods, such as those ranked lower in the evidence hierarchies of EBP proponents, fare better regarding the assumptions required for obtaining policy-relevant distributional information. So far, we have only established that assumptions that are so strong that EBP proponents would not be comfortable with assuming them are required for obtaining policy-relevant distributional evidence from RCT data. It would be interesting to discuss whether this is also true for the assumptions needed to obtain policy-relevant distributional information using other policy evaluation methods.

Secondly, we have established that the special status of RCTs for the EBP movement cannot be defended by pointing to the minimal substantive assumptions that RCTs require for producing quantities of interest to policymakers. This is because these assumptions do not suffice for producing policy-relevant evidence conveyed by distributional parameters. This line of argument strengthens philosophical doubts about whether the placement of RCTs on the top of the evidence hierarchies of EBP proponents can be defended by pointing to the reliable causal estimates that RCTs provide and the minimal substantive assumptions they employ. However, concluding that this special status for RCTs cannot be defended based on other considerations would require further discussion.

Acknowledgements

The story of this thesis starts at almost the same time at which I started my Research Master at EIPE: What you have read in the sixth chapter were ideas that I already (in a very rough form!) began to engage in for my second assignment for my first EIPE class, Methodology of Economics. Together with topics that I have discussed in chapter five, these ideas ended up in my final essay for the same class. Some months and the first Corona wave later, these thoughts came together with issues that I have analysed in chapters three and four for my final Milestone. And finally, ideas for chapters one and two joined shortly before I started writing this thesis in January. So truly, you can say that this thesis and I have gone a long way together (still, I am happy and puzzled that I managed to keep about the same level of excitement for the topic from day one until now).

In the following, I want to thank the people who supported me on this long way. First of all, thanks to the EIPE community for the stimulating and supportive environment during my time in Rotterdam. Especially, thanks to both my supervisor and my advisor, Conrad and William, for their detailed feedback and the very efficient thesis process. Thanks also to Malvina, Max and Quint for all of the very helpful comments. Also, as a tutor, thanks to Conrad for always supporting all the crazy ideas that I had (from taking an in-person(!) class in Groningen to doing another Master's). Most importantly, I would like to thank all the great fellow students that I got lucky to spend so much time with during the past two years. Especially, thanks to Chiara, Ina and Max for all the talks, laughs and beers. You are an amazing crew! Also, thanks to the non-EIPE people with whom I share many nice memories from my time in Rotterdam, especially my DCGO bandmates and my colleagues from the Sustainability Hub.

Finally, thanks to my loving and crazy family for always backing me and my path through university. And of course, thanks to my beloved boyfriend Daniel for his never-ending support from both near and afar. Without you, my time in Rotterdam would have lacked many cheering calls, adventurous trips, and of course, my best memory, our time together in Hatta - in other words, it would not have been the same.

Bibliography

1) Abadie, Alberto et al. 1998: "Instrumental Variable Estimate of Quantile Treatment Effects". *NBER Technical Working Paper, No. 229.*
2) Abadie, Alberto and Cattaneo, Matias 2018: "Econometric Methods in Program Evaluation", in: *Annual Review of Economics, Vol. 10,* pp. 465-503.
3) Abbring, Jaap and Heckman, James 2007: "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation", in: Heckman, James and Leamer, Edward (Eds): *Handbook of Econometrics, Volume 6B*, North-Holland: Elsevier, pp. 5148-5286.
4) Angrist, Joshua D. and Pischke, Jörn-Steffen 2015: *Mastering ´Metrics. The Path from Cause to Effect.* Princeton: Princeton University Press.
5) Athey, Susan and Imbens, Guido 2018: "The Econometrics of Randomized Experiments", in: Banerjee, Abhijit and Duflo, Esther (Ed.): *Handbook of Field Experiments Vol. 1.* North-Holland: Elsevier.
6) Atkinson, Anthony 1970: "On the Measurement of Inequality", in: *Journal of Economic Theory, Vol. 2*, pp. 244-263.
7) Banerjee, Abhijit V. and Duflo, Esther 2008: "The Experimental Approach to Development Economics", *NBER Working Paper 14467.*
8) Banerjee, Abhijit V. and Duflo, Esther 2011: *Poor Economics. A Radical Rethinking of the Way to Fight Global Poverty.* New York: PublicAffairs.
9) Bedoya, Guadalupe et al. 2017: "Distributional Impact Analysis. Toolkit and Illustrations of Impacts Beyond the Average Treatment Effect", *World Bank Policy Research Working Paper Vol. 8193.*
10) Bitler, Marianne et al. 2003: "What Mean Impacts Miss: Distributional Impacts of Welfare Reform Experiments". *NBER Working Paper No. 10121.*
11) Bolton, Gary and Ockenfels, Axel 2006: "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment", in: *The American Economic Review, Vol. 96, No. 5*, pp. 1906-1911.
12) Bruhn, Miriam et al. 2016: "The Impact of High School Financial Education: Evidence from a Large-Scale Evaluation in Brazil", in: *American Economic Journal: Applied Economics, Vol. 8, No. 4*, pp. 256-295.

13) Carneiro, Pedro et al. 2003: "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice", *NBER Working Paper No. 9546.*
14) Cairney, Paul et al. 2017: „How to Communicate Effectively with Policymakers: Combine Insights from Psychology and Policy Studies", in: *Palgrave Communications, 3:37,* pp. 1-8.
15) Cartwright, Nancy 2007: "Are RCTs the Gold Standard?", in: *BioSocieties, Vol. 2,* pp. 11-20.
16) Cartwright, Nancy 2012: "RCTs, Evidence and Predicting Policy Effectiveness", in: Kincaid, Harold (Ed.): *The Oxford Handbook of Philosophy of Social Sciences*, pp. 289-318.
17) Cartwright, Nancy and Hardie, Jeremy 2012: *Evidence-based Policy: A Practical Guide to Doing it Better.* Oxford: Oxford University Press.
18) Chernozhukov, Victor et al. 2018: "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments", *NBER Working Paper 24678.*
19) Clarke, Brendan et al. 2014: "Mechanisms and the Evidence Hierarchy", in: *Topoi, Vol. 33*, pp. 339-360.
20) Coalition for Evidence-Based Policy (CEBP) 2002: *Bringing Evidence-Driven Progress to Education: A Recommended Strategy for the U.S. Department of Education.* Washington, DC: CEBP.
21) Cowen, Nick 2019: "For whom does 'what works' work? The political economy of evidence-based education", *CHESS Working Paper No. 2019-04.*
22) Dale, Stacy Berg and Krueger, Alan B. 2002: "Estimating The Payoff To Attending A More Selective College: An Application Of Selection On Observables And Unobservables", in: *Quarterly Journal of Economics Vol. 117(4),* pp. 1491-1527.
23) De Souza Leão, Luciana and Eyal, Gil 2019: "The rise of randomized controlled trials (RCTs) in international development in historical perspective", in: *Theory and Society Vol 48*, pp. 383-418.
24) Deaton, Angus 2010: "Instruments, Randomization and Learning about Development", in: *Journal of Economic Literature Vol. 48*, pp. 424-455.
25) Deaton, Angus and Cartwright, Nancy 2018a: "Understanding and Misunderstanding Randomized Controlled Trials", in: *Social Science and Medicine, Vol. 210*, pp. 2-21.

26) Djebbari, Habiba and Smith, Jeffrey 2008: “Heterogeneous Impacts in PROGRESA”, in: *Journal of Econometrics, No. 145,* pp. 64-80.
27) Duflo, Esther and Kremer, Michael 2003: „Use of Randomization in the Evaluation of Development Effectiveness”, *Paper prepared for the World Bank Operations Evaluation Department (OED) Conference on Evaluation and Development Effectiveness in Washington, D.C.*
28) Dur, Robert 2001: “Why do Policy Makers Stick to Inefficient Decisions?”, in: *Public Choice, Vol. 107*, pp. 221-234.
29) Durante, Ruben et al. 2014: “Preferences for Redistribution and Perception of Fairness: An Experimental Study”, in: *Journal of the European Economic Association, Vol. 12(4),* pp. 1059–1086.
30) Engelmann, Dirk and Strobel, Martin 2004: „Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments”, in: *The American Economic Review, Vol. 94, No. 4,* pp. 857-869.
31) Engelmann, Dirk and Strobel, Martin 2007: “Preferences over Income Distributions: Experimental Evidence”, in: *Public Finance Review, Vol. 35, No. 2*, pp. 285-310.
32) Fan, Yanqin and Park, Sang Soo 2010: “Sharp Bounds on the Distribution of Treatment Effects and their Statistical Inference”, in: *Econometric Theory, Vol. 26, No. 3,* pp. 931-951.
33) Favereau, Judith and Nagatsu, Michiru 2020: “Holding back from theory: limits and methodological alternatives of randomized field experiments in development economics”, in: *Journal of Economic Methodology, 27:3*, pp. 191-211.
34) Fehr, Ernst et al. 2006: Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment, in: *The American Economic Review, Vol. 96, No. 5,* pp. 1912-1917.
35) Firpo, Sergio 2007: “Efficient Semiparametric Estimation of Quantile Treatment Effects”, in: *Econometrica, Vol. 75, No. 1*, pp. 259-276.
36) Firpo, Sergio 2010: “Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures”, *IZA Discussion Paper 4841.*
37) Firpo Sergio and Ridder, Geert 2008: “Bounds on Functionals of the Distribution of Treatment Effects”, *IEPR WORKING PAPER 08.09.*

38) Frölich, Markus and Melly, Blaise 2013: “Unconditional Quantile Treatment Effects Under Endogeneity”, in: *Journal of Business & Economic Statistics, 31:3*, pp. 346-357.
39) Gustafsson, Anders 2019: “Busy Doing Nothing: Why Politicians Implement Inefficient Policies”, in: *Constitutional Political Economy, Vol. 30,* pp. 282-299.
40) Heckman, James and Smith, Jeffrey 1995: "Assessing the Case for Social Experiments", in: *Journal of Economic Perspectives, 9*, pp. 85- 110.
41) Heckman, James; Smith, Jeffrey and Clements, Nancy 1997: “Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts” in: *The Review of Economic Studies, Vol. 64, No. 4*, pp. 487-535.
42) Heckman, James; LaLonde, Robert and Smith, Jeffrey 1999: “The Economics and Econometrics of Active Labour Market Programs”, in: Ashenfelter, A. and Card, D. (Eds).: *Handbook of Labour Economics, Volume 3.* North-Holland: Elsevier, pp. 1866-2097.
43) Heckman, James and Vytlacil, Edward 2007: “Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation”, in: Heckman, James and Leamer, Edward (Eds): *Handbook of Econometrics, Volume 6B*, North-Holland: Elsevier, pp. 4780-4867.
44) Heckman, James 2008: “Econometric Causality”, *NBER Working Paper 13934.*
45) Heckman, James 2020: “Randomization and Social Policy Evaluation Revisited”, *NBER Technical Working Paper No. 107.*
46) Imbens, Guido and Wooldridge, Jeffrey 2009: “Recent Developments in the Econometrics of Program Evaluation”, in: *Journal of Economic Literature, Vol. 47, No. 1*, pp. 5-86.
47) Imbens, Guido 2010: “Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)”, in: *Journal of Economic Literature Vol. 48*, pp. 399 – 423.
48) Imbens, Guido 2018: “Comments on: “Understanding and Misunderstanding Randomized Controlled Trials by Cartwright and Deaton”, in: *Social Science & Medicine, Vol. 210*, pp. 50-52.
49) Iversen, Torben and Goplerud, Max 2018: “Redistribution Without a Median Voter: Models of Multidimensional Politics”, in: *Annual Review of Political Science, Vol. 21,* pp. 295-317.
50) Jackson, Erika et al. 2013: “Estimating the Distributional Effects of Education Reform: A Look at Project STAR”, in: *Economics of Education Review Vol. 13*, pp. 92-103.

51) Lacinese, Valentino 2007: “Voting over Redistribution and the Size of the Welfare State: The Role of Turnout”, in: *Political Studies*, pp. 1-18.

52) Luebker, Malte 2012: “Income Inequality, Redistribution and Poverty. Contrasting Rational Choice and Behavioural Perspective”. *ILO Research Paper No.1.*

53) Manski, Charles 1996: “Learning About Treatment Effects from Experiments with Random Assignment of Treatments”, in: *The Journal of Human Resources Vol. 31, No. 4,* pp. 709-733.

54) Manski, Charles 2010: “Partial Identification in Econometrics”, in: Durlauf, Steven and Blume, Lawrence (Eds): *Microeconometrics.* Basingstoke: Palgrave Macmillan, pp. 178-188.

55) Muller, Seán M. 2020: “The implications of a fundamental contradiction in advocating randomized trials for policy”, in: *World Development Vol. 127*, pp. 1-3.

56) Na, Chongmin et al. 2015: “On the Importance of Treatment Effect Heterogeneity in Experimentally-Evaluated Criminal Justice Interventions”, in: *Journal of Quantitative Criminology, Vol 31*, pp. 289-310.

57) National Co-Ordinating Centre for Public Engagement 2020: “Policy Makers”. Available under: *https://www.publicengagement.ac.uk/do-engagement/understanding-audiences/policy-makers.*

58) NICE 2006: *The guidelines manual.* London: National Institute for Health and Clinical Excellence. Available from: http://www.nice.org.uk.

59) Pearce, Warren and Raman, Sujatha 2014: “The new randomised controlled trials (RCT) movement in public policy: challenges of epistemic governance”, in: *Policy Sciences, Vol. 47*, pp. 387-402.

60) Petticrew et al. 2004: “Evidence for Public Health Policy on Inequalities: 1: The Reality According to Policy Makers”, in: *Journal of Epidemiology and Community Health, Vol. 58*, pp. 811-816.

61) Reiss, Julian 2013: *Philosophy of Economics: A Contemporary Introduction*. New York: Routledge.

62) Sanderson, Ian 2003: “Is it ‘what works’ that matters? Evaluation and evidence-based policy-making”, in: *Research Papers in Education, 18:4*, pp. 331-345.

63) Scervini, Francesco 2012: “Empirics of the Median Voter: Democracy, Redistribution

and the Role of the Middle Class", in: *Journal of Economic Inequality, Vol. 10*, pp. 529-550.

64) Stegenga, Jacob 2013: "Down with the Hierarchies", in: *Topoi 33 (2),* pp. 313-322.

65) Subramanian, S.V. et al. 2018: "The "Average Treatment Effect": A Construct Ripe for Retirement. A Commentary on Deaton and Cartwright", in: *Social Science & Medicine.*

66) Teira, David and Reiss, Julian 2013: "Causality, Impartiality and Evidence-Based Policy", in: Chao, Hsiang-Ke et al. (Eds): *Mechanism and Causality in Biology and Economics*, pp. 207-224.

67) US Department of Education et al. 2003: *Educational Practices Supported by Rigorous Evidence: A User Friendly Guide.*

68) Whitehead, Margaret et al. 2004: "Evidence for public health policy on inequalities: 2: Assembling the evidence jigsaw", in: *Journal of Epidemiology and Community Health, Vol. 58*, pp. 817-821.

69) Worrall, John 2010: "Do we Need Some Large, Simple Randomized Trials in Medicine?", in: Suarez, Mauricio et al. (Eds): *EPSA Philosophical Issues in the Sciences, Vol. 2*. Dordrecht: Springer, pp. 289-302.

70) Wu, Ximing and Perloff, Jeffrey 2007: "Information-Theoretic Deconvolution Approximation of Treatment Effect Distribution", *IRLE Working Paper No. 149-07.*