A multi-cell experimental design to recover policy relevant treatment effects, with an application to online advertising*

Caio Waisman Brett R. Gordon

Kellogg School of Management
Northwestern University

February 28, 2023

Abstract

Experiments are an important tool to measure the impacts of interventions. However, in experimental settings with one-sided noncompliance, extant empirical approaches may not produce the estimates a decision-maker needs to solve their problem. For example, these experimental designs are common in digital advertising settings, but they are uninformative of decisions regarding the intensive margin—how much should be spent or how many consumers should be reached with a campaign. We propose a solution that combines a novel multi-cell experimental design with modern estimation techniques that enables decision-makers to recover enough information to solve problems with an intensive margin. Our design is straightforward to implement. Using data from advertising experiments at Facebook, we demonstrate that our approach outperforms standard techniques in recovering treatment effect parameters. Through a simple advertising reach decision problem, we show that our approach generates better decisions relative to standard techniques.

Keywords: Marginal Treatment Effects, Field Experiments, Causal Inference, Digital Advertising, Advertising Measurement.

* We thank Ivan Canay, Gastón Illanes, Garrett Johnson, and Ilya Morozov for helpful comments. E-mail addresses for correspondence: caio.waisman@kellogg.northwestern.edu, b-gordon@kellogg.northwestern.edu.

1 Introduction

Experiments are considered a particularly reliable tool to measure the causal effects of interventions. However, decision-makers who conduct experiments may be unsatisfied solely with measuring these effects; they need more information to inform specific decisions they have to make. Yet in a variety of settings, a decision-maker's ability to experiment is limited because they cannot fully control treatment assignment. This can restrict the collection of the information necessary to make these decisions.

In such cases, the most common experimental route involves randomizing eligibility to receive treatment. For example, firms want to measure the effectiveness of their digital ad campaigns, but they cannot directly randomize advertising exposure. Instead, they randomly assign consumers to be *eligible* or *ineligible* to be exposed to ads. This experimental design, where noncompliance is one-sided, is not only popular for measuring online advertising effects (see Johnson (2022) for a recent review) but it is also used in economics, political science, and medicine.[1]

This design provides valuable information: it enables the researcher to recover the average treatment effect on the treated (ATT). This parameter is important because it quantifies the effect of the treatment on the observable subset of units that were eligible to receive treatment; alternatively, it quantifies the loss that would have been experienced had the experiment not been conducted. In addition, when compared with costs, it shows whether an existing policy, such as the current advertising policy used by firms, is beneficial. However, this treatment effect parameter may be of limited assistance when it comes to decisions with an intensive margin, such as when an advertiser must decide how many consumers to reach or how large a budget to set.

In this paper, we propose an approach that allows the researcher (and the decision-maker) to obtain the information necessary to make these types of decisions. Our approach combines a novel multi-cell experimental design with modern estimation techniques to recover the marginal treatment effect (MTE) function. This function allows us to inform these decisions and to recover the most common treatment effect parameters of interest, including the ATT. Without our experimental design, we explain why implementing these estimation techniques would require non-trivial additional restrictions. We illustrate our approach through a series of simulations that are calibrated using advertising experiments at Facebook. We show that our proposed experimental design yields more

[1] Experiments with one-sided noncompliance have also been used in the context of: online A/B tests (Deng et al., 2019); clinical trials (Sommer and Zeger, 1991); breast self-examination treatments (Mealli et al., 2004); interventions to incentivize voter turnout (Green et al., 2003); effects of job training (Schochet et al., 2008) and job assistance (Crépon et al., 2013) programs; the impacts of access to microcredit (Crépon et al., 2015); the effects of deworming drugs on children's health and education (Miguel and Kremer, 2004); and housing voucher policies (Chetty et al., 2016).

precise estimates of the MTE function relative to those that would be obtained from the typical experiment design. Using these estimates, we demonstrate that our method produces solutions to a decision-maker's problem that can differ materially from those that would be obtained using standard techniques.

To introduce our setting, we consider an advertiser deciding what fraction of users to reach with advertising from among a target audience.[2] Using this decision problem, we describe the typical experimental design with one-sided noncompliance and define the policy relevant treatment effect (PRTE), which is how we refer to the treatment effect parameter necessary to solve such a decision problem. In doing so, we explain why this experimental design does not deliver this object unless nontrivial restrictions are imposed.

Our empirical approach is inspired by the estimation method in Brinch et al. (2017), who show how to obtain an approximation to the MTE function using a discrete instrumental variable that generates variation in the probability of treatment. We point out that direct application of Brinch et al. (2017) to a single-cell experiment with one-sided noncompliance is infeasible because the data contain too few moments to fully recover the set of parameters, and explain why potential parameter restrictions would impose a nontrivial form of structure on the endogeneity of exposure and/or treatment. As we show in Section 4.3, we lack any empirical or theoretical guidance to justify these restrictions.

The multi-cell experimental design we propose remedies this underidentification problem and enables the researcher to obtain a credible approximation of the MTE function. Our design first randomly allocates units across C cells, and then within each cell, units are once again randomly split into test and control groups. Consistent with typical limitations around treatment assignment in practice, each cell features an experiment with one-sided noncompliance. We show how this multi-cell design resolves the estimation problem by generating a sufficient number of moments to approximate the MTE function using a polynomial of degree C.

Implementing our approach requires the researcher to generate sufficient variation in the propensity score—the probability of treatment given eligibility—across experimental cells. In Section 3.5, we provide some practical guidance on how to achieve this even under the common case of one-sided noncompliance found in digital advertising settings. Fortunately, even though researchers cannot directly control treatment assignment, they can influence the probability of treatment by adjusting the budget per user. This quantity depends on the relative sizes of the cells and how the budget is distributed across cells—decisions that are under the researcher's control. Critically, relative to a single-cell test, our method does not require the overall budget allocated to the experiment to be increased.

[2] In Section 2.3, we explain why this problem is equivalent to one where the advertiser sets a budget level.

We apply our method to data generating processes (DGPs) calibrated to an advertising experiment at Facebook.[3] We first consider simple cubic polynomials for the MTE function and show that our method can perform well in approximating it, and contrast these results to what could be obtained in a single-cell design using Brinch et al. (2017) through the imposition of additional parameter restrictions. Although these parameter restrictions remedy the underidentification problem, we find they produce unreliable estimates of the MTE function. Moreover, we are unaware of any way to assess which of these restrictions is most justifiable.

The differences across methods in their MTE function estimates lead to a natural question: how much do these differences matter for the decision-making problem? To shed light on this question, we revisit the firm's decision regarding the fraction of target consumers to reach with their advertising using a simple example. We find that our approach succeeds in virtually eliminating any losses in expected profits across different DGPs. In turn, under specific circumstances, while certain parametric restrictions combined with the direct application of Brinch et al. (2017) to data from a single-cell design can perform well in this regard, such an approach is more likely to yield high losses in expected profit.

We also consider an example in which the underlying MTE function is more complex. Our goal is use this more difficult case to assess the role the number of cells and values taken by the propensity score play in implementing our method. In particular, we seek to obtain evidence as to whether there can be concrete guidance to practitioners concerning how to choose these objects. We find this not to be the case both for approximating the MTE function and for minimizing expected profit losses. Furthermore, imposing common assumptions from the causal inference literature such as monotone treatment response does not remedy this lack of guidance. Nevertheless, we demonstrate that it is straightforward for practitioners to reallocate a budget designated for a single-cell experiment across many cells given expected values for the propensity scores. Consequently, it is always possible to convert a single-cell design into a multi-cell one given a number of cells and propensity score values without altering the original budget chosen for conducting the experiment.

Our paper makes three contributions. First, we contribute to the broad literature on estimating treatment effects in experiments with one-sided noncompliance. In particular, we develop an experimental design that is built to leverage modern estimation techniques when only eligibility to receive treatment can be randomized. These techniques have their origins in the work of Björklund and Moffitt (1987) and Heckman and Vytlacil (2005), who showed identification of the MTE function using a continuous instrumental variable with observational data. More recently, recognizing that instruments are often discrete, Brinch et al. (2017) showed how to recover polynomial MTE functions, or, equivalently, how to recover a polynomial approximation to the MTE function, whereas Mogstad et al. (2018)

[3] We could apply our method to data from a multi-cell experiment if we had access to such data.

showed how to obtain partial identification of the MTE.[4] Neither study considers how these methods can be used in combination with experimental data specifically, and in particular when the design of the experiment can be altered to enhance estimation. This is our primary contribution: to tailor the experimental design to exploit these methods.

Second, we add to the expanding literature on estimating online advertising effects. Much of this work focuses on recovering the intent-to-treat (ITT) or ATT parameters using experiments with one-sided noncompliance.[5] Obtaining such estimates is useful to document advertising effects and to inform an advertiser's extensive margin decision of whether to advertise (a "go/no go" decision). However, an advertiser is unable to apply these estimates to choose the intensive margin of how many consumers to reach with advertising. Our paper makes a contribution by starting to fill this gap in the literature using an approach that embeds a causal inference setup in the advertiser's optimization problem. To the best of our knowledge, the only other paper that uses the MTE framework in marketing is Daljord et al. (2022), who apply Mogstad et al. (2018) to data from a promotion targeting experiment with two-sided noncompliance conducted with a hotel chain.

Third, our paper is related to work that examines an advertiser's decision problem. Early work in this area sought to determine the optimal budget allocation given an aggregate advertising response model (Sethi, 1977; Holthausen Jr. and Assmus, 1982; Simon, 1982; Basu and Batra, 1988). More recent work studies this problem in online advertising settings (Pani et al., 2017; Baardman et al., 2019; Zhao et al., 2019; Geng et al., 2021). However, none of these papers have causal inference in mind. Waisman et al. (2022b) provides a framework to recover treatment effect parameters that account for parallel experimentation by competitors to inform an advertiser's extensive margin decision. A different strand of this literature connects causal inference with advertising decisions, specifically a firm's optimal bidding strategy in real-time bidding (RTB) environments (Lewis and Wong, 2018; Waisman et al., 2022a). Neither of these papers obtain the MTE, which is unnecessary for the decision problems they study. With the MTE, we can solve a broader set of advertising decision problems, though our method does not account for other experimentation costs that would be relevant in RTB settings.

The rest of this paper proceeds as follows. Section 2 introduces the typical experimental design through the advertiser's decision problem and shows this design does not provide the information needed to solve this problem. Section 3 presents our empirical strategy that consists of a novel multi-cell experimental design, an estimation technique to recover an approximation to the MTE function, and discusses practical issues in implementing this design. Section 4 uses data from Facebook advertising experiments to illustrate the benefits of our methodology relative to a direct application of Brinch et al. (2017). Section

[4] We are working to incorporate our multi-cell design into the methods in Mogstad et al. (2018).

[5] Examples include Lewis and Reiley (2014); Brodersen et al. (2015); Johnson et al. (2016, 2017a,b); Gordon et al. (2019); Sahni et al. (2019); Barajas and Bhamidipati (2021); Gui et al. (2021); Gordon et al. (2022).

5 concludes.

2 Setting

In this section, we consider a specific decision problem faced by a firm, that of what fraction of a target audience they advertise to in order to maximize expected profits. We focus on this problem because it enables us to introduce our model, to describe the typical experimental design with one-sided noncompliance, and to explain why the treatment effect obtained from this design (the ATT) does not solve the decision-maker's problem.

2.1 Firm's advertising problem

We introduce our model by considering a specific decision problem. Suppose a firm wishes to choose the *fraction* of consumer from a target segment to reach with advertising to maximize expected profit. As we show below, this decision is equivalent to choosing an advertising budget to reach a given proportion of consumers.

Let D be an indicator for whether a unit (user) is treated (exposed to advertising), Y_1 be the outcome when $D = 1$, and Y_0 be the outcome when $D = 0$. The observed outcome can be written as:

$$Y = DY_1 + (1 - D)Y_0. \tag{1}$$

Let ν be the fraction of units exposed to the treatment. Assume that the firm can convert outcomes into monetary amounts by multiplying them by a known constant, δ. Let the cost of treating a fraction ν of units be given by a known cost function, $\kappa(\nu)$. Then the firm's expected profit maximization problem is:

$$\max_{\nu \in [0,1]} \left(\delta \times \{ \nu \mathbb{E}\left[Y|D = 1\right] + (1 - \nu)\mathbb{E}\left[Y|D = 0\right] \} - \kappa(\nu) \right) \tag{2}$$

To solve this problem, the firm needs to compute the conditional expectations above, which are unknown. The specific object the firm needs to know to solve this problem is what we refer to as the policy relevant treatment effect (PRTE). We now discuss how this parameter can be estimated.

2.2 Experimental design

There are several methods to estimate the conditional expectations in expression (2) from data; arguably, one preferred way to collect these data is by running an experiment, ideally one in which treatment itself is randomly assigned to the experimental units. However, it is often the case, such as the one we consider, that the experimenter, in this instance the firm, does not fully control treatment assignment and therefore cannot randomize it. The most common solution in these situations is to randomize *eligibility* to receive treatment instead, which is the experimental design we address.[6]

Let Z be an indicator for whether the unit is *eligible* to receive treatment, which we assume is randomly assigned. Following Heckman and Vytlacil (2005), let treatment be given by:

$$D = \mathbb{1}\{\nu(Z) \geq U\}, \tag{3}$$

where $\nu(\cdot)$ governs the process of selection into treatment and which could potentially depend on a vector of observable characteristics X, as could the distribution of treatment effects.[7] This functional form is either known or estimable from data generated from the experiment by both the firm and the researcher. The error term U represents selection into treatment that the firm and the researcher do not observe.

Following Mogstad et al. (2018), we maintain the following standard assumption.

Assumption 1.
(i) $U \perp\!\!\!\perp Z$, *where* $\perp\!\!\!\perp$ *denotes statistical independence.*
(ii) $\mathbb{E}[Y_d|Z,U] = \mathbb{E}[Y_d|U]$ *and* $\mathbb{E}\left[Y_d^2\right] < \infty$ *for* $d \in \{0,1\}$.
(iii) U *is continuously distributed.*

Assumptions 1(*i*) and 1(*ii*) require Z to be exogenous with respect to the selection and outcome processes, thereby characterizing it as a valid instrumental variable for the treatment indicator, D. This is guaranteed in our setting because Z is randomly assigned. Given Assumption 1(*i*), Vytlacil (2002) showed that the assumption that the index of the selection is additively separable as in equation (3) is equivalent to the monotonicity condition from Imbens and Angrist (1994). Finally, Assumption 1(*iii*) is a weak regularity condition that allows us to normalize $U \sim U(0,1)$. Under these conditions, this model is equivalent to that of Imbens and Angrist (1994). Assumption 1 allows us to define the propensity score

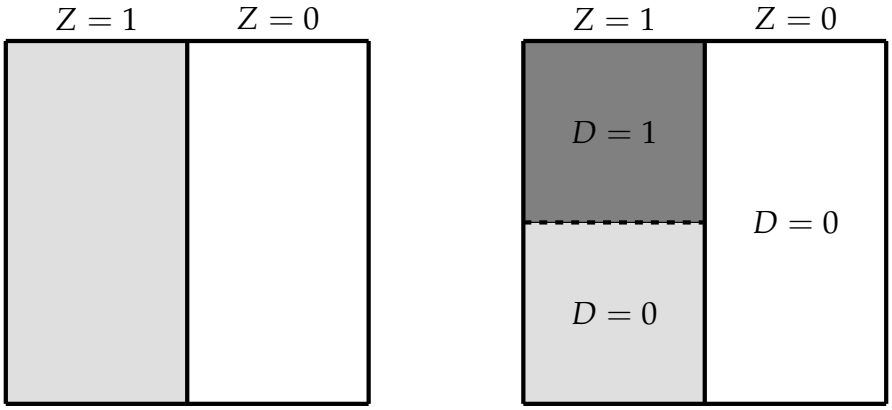
[6] If an experiment is infeasible, the firm could apply a model to observational data. However, lacking an exogenous source of variation on treatment, it may be difficult to reliably estimate treatment effect parameters due to unobservable confounds that are correlated with both treatment and outcomes (Gordon et al., 2019, 2022).

[7] For simplicity, we omit these observable characteristics from our model, but they can be added in a straightforward manner. We demonstrate this in Appendix A.

as

$$\nu(Z) = \Pr\left(D = 1|Z\right). \tag{4}$$

Figure 1 illustrates this experimental design. It can be seen as a single-cell design. Within this cell, which we refer to as Cell 1, units are randomly assigned to $Z = 1$ or $Z = 0$, corresponding to Figure 1a. Since $\nu(1) \in (0, 1)$, some, but not all units that are eligible to receive treatment are actually treated ($D = 1$)—left column of Figure 1b—, and because $\nu(0) = 0$, none of the units that are ineligible to receive treatment are treated ($D = 0$)—right column of Figure 1b.



(a) Cell 1: Setup (b) Cell 1: Treatment

Figure 1: Single-cell experiment with one-sided noncompliance

This setup corresponds to an experiment with *one-sided noncompliance*, a typical experimental design in many online advertising settings. Under standard conditions, any experimental design that features a binary treatment and a valid binary instrument can identify a local average treatment effect (LATE) parameter. However, one-sided noncompliance gives us the ability to estimate another important treatment effect parameter, the average treatment effect on the treated (ATT), defined as $\text{ATT} \equiv \mathbb{E}\left[Y_1 - Y_0|D = 1\right]$, because it implies that ATT = LATE.

Much of the recent literature on advertising measurement stops once a focal treatment effect(s) has been recovered. However, work in this area has been less focused on connecting those estimates to advertising decisions. This motivates our interest in the advertiser's decision problem, which we return to next.

2.3 Revisiting the firm's advertising problem

Using equations (1) and (3) and the normalization that $U \sim U(0,1)$, we can rewrite the firm's optimization problem as:

$$\max_{\nu \in [0,1]} \left(\delta \times \{ \nu \mathbb{E}\left[Y_1 | U \leq \nu \right] + (1-\nu) \mathbb{E}\left[Y_0 | U \geq \nu \right] \} - \kappa(\nu) \right). \tag{5}$$

As we show in Appendix B, it follows that:

$$\mathbb{E}\left[Y_1 | U \leq \nu \right] = \int_0^{\nu} m_1(u) \frac{1}{\nu} du \quad \text{and} \quad \mathbb{E}\left[Y_0 | U > \nu \right] = \int_{\nu}^{1} m_0(u) \frac{1}{1-\nu} du, \tag{6}$$

where we used that $f(u) = 1$ since U follows a standard uniform distribution. The functions $m_d(u)$, where $d \in \{0,1\}$, are defined as $\mathbb{E}\left[Y_d | = u \right]$. These functions are known as the marginal treatment response (MTR) functions.

As we also show in Appendix B, plugging the expressions in equation (6) back into equation (5) allows us to rewrite the firm's decision problem as:

$$\max_{\nu \in [0,1]} \left(\delta \times \int_0^{\nu} \text{MTE}(u) du - \kappa(\nu) \right), \tag{7}$$

where we defined the marginal treatment effect (MTE) function as:

$$\text{MTE} \equiv \mathbb{E}\left[Y_1 - Y_0 | U = u \right] = m_1(u) - m_0(u). \tag{8}$$

The MTE can be interpreted as the expected treatment effect at a particular (marginal) realization of the unobservable $U = u$. One of the benefits of this function is that, as shown, for example, in Heckman and Vytlacil (2005), it can be used to obtain most treatment effect parameters of interest, such as the average treatment effect (ATE). We present the mapping between the MTE function and some of these parameters in Appendix C.

For simplicity, assume that the MTE function is decreasing and that the cost function $\kappa(\cdot)$ is convex and differentiable. Then, the optimal fraction of units to be treated, ν^*, is determined by the first-order condition:

$$\delta \times \text{MTE}(\nu^*) = \kappa'(\nu^*). \tag{9}$$

The solution is for the monetary amount of the marginal treatment effect to be equal to the marginal cost of treatment. In other words, the solution is to equalize expected marginal revenue with marginal cost.

The solution to this optimization problem can be used to determine the firm's optimal

budget for advertising, which is then $\kappa(\nu^*)$. It can also accommodate an exogenous budget by adding a constraint that $\kappa(\nu)$ must not exceed it. Importantly, notice that the object the firm requires to solve their decision problem is the MTE function itself—this function *is* the PRTE. In turn, the ATT, which we can recover from data collected from the experimental design outlined earlier, is insufficient for the firm to make this decision because it lacks information relevant to solve the first-order condition shown in equation (9). In addition, notice that knowing the MTE function allows the firm to set the optimal budget for *any* cost function $\kappa(\cdot)$. This can make the approach we suggest, that of recovering the MTE function, preferable to one where the firm directly optimizes budget conditional on a cost function, even when this is the sole object the firm cares about.

3 Empirical approach

Our goal is to recover credible estimates of the MTE function because it can be used to obtain multiple treatment effect parameters, including the ATT, and because it is an input to solve multiple decision problems, such as the one we presented above.

In this section, we first present our proposed multi-cell experimental design. Second, we discuss how to connect the data generated from this design to the MTR functions, which allow us to recover the MTE function. Third, we explain our approximation strategy, which is motivated by and leverages the techniques in Brinch et al. (2017)—henceforth "BMW". As we show in Section 4.3, a direct application of BMW to a single-cell experiment with one-sided noncompliance does not yield sufficient information to obtain credible estimates of the MTE function. Fourth, we show how to use the approximations to solve a Bayesian version of the decision-maker's advertising problem from Section 2.3. Fifth, we discuss two important practical elements of implementing our proposed experimental design: allocating the budget across cells and the number of cells to use. This includes discussion of the variation needed to recover the approximation to the MTE function and explains how our design provides this variation.

3.1 A multi-cell experimental design

In our multi-cell design, first units are randomly divided across C cells and then, given assignment to cell c, are randomly split into test and control groups within each cell. We define $\mathcal{C} = 1, \ldots, c, \ldots, C$ to indicate assignment to cell c and Z_c as the indicator for treatment eligibility of an experimental unit from cell c. All these within-cell experiments feature one-sided noncompliance, so $\Pr(D = 1|Z_c = 0) = 0$ for all c.

We maintain the following assumption.

Assumption 2.
(i) $\Pr(Z_c = z|\mathcal{C} = c) \in (0,1)$ *for all* c *and all* z.
(ii) $\nu(Z_c = 1) \equiv \Pr(D = 1|Z_c = 1) \in (0,1)$ *for all* c.
(iii) $\nu(Z_c = 1) \neq \nu(Z_{c'} = 1)$ *for all* $c \neq c'$.

Assumptions 2(*i*) and 2(*ii*) are innocuous. First, remember that the experimenter has full control over the test/control split for each cell, and so can guarantee that $\Pr(Z_c = z|\mathcal{C} = c)$ is always strictly between 0 and 1. Second, consider cases in which the probability of treatment conditional on eligibility is either 0 or 1. If $\nu(Z_c = 1) = 1$, the endogeneity problem is resolved because eligibility to receive treatment becomes equivalent to exposure to treatment itself. In turn, if $\nu(Z_c = 1) = 0$, this estimation exercise becomes meaningless because it implies that it is impossible for units to receive the treatment under consideration.

Assumption 2(*iii*) requires that the probability of treatment conditional on eligibility varies across cells. The extent to which the experimenter is able to induce this variation is context-specific.[8] For instance, in online advertising, treatment is exposure to ads, which is determined through auctions. Hence, the advertiser, as the experimenter, can influence treatment compliance by changing the average budget per user. The higher it is, the more likely the user is to be exposed to the ad. With a multi-cell experiment, this variation can be obtained by simply allocating the budget across cells appropriately.

As we show in Section 3.3, Assumption 2 is crucial for BMW's method to be implementable in the context of our multi-cell design. On the other hand, with a single-cell experiment with one-sided noncompliance, the application of BMW's method requires the imposition of an additional constraint to alleviate an underidentification problem, which we show in Section 4.3.

3.2 Data generated from multi-cell design

All the information obtained from the multi-cell design about the MTE function is captured by the following moments:

$$\psi_{dzc} \equiv \mathbb{E}\left[Y|D = d, Z_c = z_c, \mathcal{C} = c\right], \tag{10}$$

[8] In settings where treatment is solely an active choice by the experimental unit, this might be more difficult to achieve. For example, when treatment is a job training program, the decision of whether to enroll in the program is entirely the individual's choice. The experimenter can vary incentives for the individual to take the program, but the effectiveness of these incentives is unknown ex ante.

where $d \in \{0,1\}$, $z_c \in \{0,1\}$ and $c = 1,\ldots,C$. These moments are nonparametrically identified in the data. To see how they provide information about the MTE function, we rely on the definition of treatment in equation (3) and the expressions in equation (6) to obtain:

$$\begin{aligned}\psi_{1zc} &= \mathbb{E}\left[Y|D=1,Z_c=z_c,\mathcal{C}=c\right] \\ &= \mathbb{E}\left[Y_1|U\leq\nu(z_c),Z_c=z_c\right] \\ &= \frac{1}{\nu(z_c)}\int_0^{\nu(z_c)} m_1(u)du \end{aligned} \tag{11}$$

and

$$\begin{aligned}\psi_{0zc} &= \mathbb{E}\left[Y|D=0,Z_c=z_c,\mathcal{C}=c\right] \\ &= \mathbb{E}\left[Y_0|U>\nu(z_c),Z_c=z_c\right] \\ &= \frac{1}{1-\nu(z_c)}\int_{\nu(z_c)}^1 m_0(u)du. \end{aligned} \tag{12}$$

Hence, we have a known relationship between identified moments and the underlying MTR functions, $m_0(u)$ and $m_1(u)$, which we can then leverage to obtain information about the MTE function.

At first, it might seem like the multi-cell design generates $3C$ different moments because $d \in \{0,1\}$, $z_c \in \{0,1\}$ and $d = 1$ only if $z_c = 1$ would imply three moments per cell. However, note that $\nu(Z_c = 0) = 0$ for all $c = 1,\ldots,C$. From equation (12), this implies that

$$\begin{aligned}\psi_{00c} &= \mathbb{E}\left[Y|D=0,Z_c=0,\mathcal{C}=c\right] \\ &= \mathbb{E}\left[Y_0|U>0,Z_c=0\right] \\ &= \int_0^1 m_0(u)du \\ &\equiv \psi_{00} \quad \text{for all } c. \end{aligned} \tag{13}$$

Hence, the multi-cell design generates $2C+1$ different moments. Next, we show that these moments are sufficient to construct an approximation to the MTE function.

3.3 Approximation Method

BMW show that if an instrument Z takes C different values, each associated with a propensity score that is strictly between 0 and 1, then we can approximate the MTR functions with a polynomial of degree $C-1$ provided that the propensity scores are also different from

one another.[9]

We adapt this approach to our multi-cell experimental design. When $d = 1$, we observe C different values for ψ_{1zc} from equation (11). When $d = 0$, we observe $C + 1$ different values for ψ_{0zc}, with C values from equation (12) and one value from equation (13).

Given the variation in the observed moments and in the propensity score, we consider the following polynomial approximations of the MTR functions:

$$\tilde{m}_1(u;\lambda_1) = \sum_{c=0}^{C-1} \lambda_{1c} u^c \quad \text{and} \quad \tilde{m}_0(u;\lambda_0) = \sum_{c=0}^{C} \lambda_{0c} u^c, \tag{14}$$

where it should be noted that the approximation when $d = 0$ is of one higher degree compared to $d = 1$. Plugging (14) back into the right-hand side of equations (11) and (12), we obtain the following approximations to the moments:

$$\begin{aligned}
\tilde{\psi}_{1zc} &\equiv \frac{1}{\nu(z_c)} \int_0^{\nu(z_c)} \sum_{c'=0}^{C-1} \lambda_{1c'} u^{c'} du \\
&= \sum_{c'=0}^{C-1} \lambda_{1c'} \frac{1}{\nu(z_c)} \int_0^{\nu(z_c)} u^{c'} du \\
&= \sum_{c'=0}^{C-1} \lambda_{1c'} \left(\frac{\nu(z_c)^{c'}}{c'+1} \right)
\end{aligned} \tag{15}$$

and

$$\begin{aligned}
\tilde{\psi}_{0zc} &\equiv \frac{1}{1-\nu(z_c)} \int_{\nu(z_c)}^{1} \sum_{c'=0}^{C} \lambda_{0c'} u^{c'} du \\
&= \sum_{c'=0}^{C} \lambda_{0c'} \frac{1}{1-\nu(z_c)} \int_{\nu(z_c)}^{1} u^{c'} du \\
&= \sum_{c'=0}^{C} \lambda_{0c'} \left(\frac{\sum_{s=0}^{c'} \nu(z_c)^s}{c'+1} \right)
\end{aligned} \tag{16}$$

[9] Alternatively, if the MTR functions *are* polynomials of degree $C - 1$ or less, then they are point identified, and, consequently, so is the MTE function.

for all $c \in C$. We can stack these terms and represent (15) and (16) in matrix form:

$$\underbrace{\begin{bmatrix} \tilde{\psi}_{111} \\ \tilde{\psi}_{112} \\ \vdots \\ \tilde{\psi}_{11C} \end{bmatrix}}_{\tilde{\psi}_1} = \underbrace{\begin{bmatrix} 1 & \frac{\nu(z_1)}{2} & \dots & \frac{\nu(z_1)^{C-1}}{C} \\ 1 & \frac{\nu(z_2)}{2} & \dots & \frac{\nu(z_2)^{C-1}}{C} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \frac{\nu(z_C)}{2} & \dots & \frac{\nu(z_C)^{C-1}}{C} \end{bmatrix}}_{\nu_1} \underbrace{\begin{bmatrix} \lambda_{10} \\ \lambda_{11} \\ \vdots \\ \lambda_{1,C-1} \end{bmatrix}}_{\lambda_1} \tag{17}$$

and

$$\underbrace{\begin{bmatrix} \tilde{\psi}_{00} \\ \tilde{\psi}_{011} \\ \tilde{\psi}_{012} \\ \vdots \\ \tilde{\psi}_{01C} \end{bmatrix}}_{\tilde{\psi}_0} = \underbrace{\begin{bmatrix} 1 & \frac{1}{2} & \dots & \frac{1}{C+1} \\ 1 & \frac{1+\nu(z_1)}{2} & \dots & \frac{1+\nu(z_1)+\dots+\nu(z_1)^C}{C+1} \\ 1 & \frac{1+\nu(z_2)}{2} & \dots & \frac{1+\nu(z_2)+\dots+\nu(z_2)^C}{C+1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \frac{1+\nu(z_C)}{2} & \dots & \frac{1+\nu(z_C)+\dots+\nu(z_C)^C}{C+1} \end{bmatrix}}_{\nu_0} \underbrace{\begin{bmatrix} \lambda_{00} \\ \lambda_{01} \\ \vdots \\ \lambda_{0,C} \end{bmatrix}}_{\lambda_0}. \tag{18}$$

Provided that the matrices ν_1 and ν_0 from equations (17) and (18) are invertible, we can compute λ_1 and λ_0 by replacing $\tilde{\psi}_1$ and $\tilde{\psi}_0$ with their observed counterparts from equation (10): $\lambda_1 = \nu_1^{-1}\tilde{\psi}_1$ and $\lambda_0 = \nu_0^{-1}\tilde{\psi}_0$. The invertibility of ν_1 and ν_0 is ensured by Assumption 2. Having recovered the λs that parameterize the approximation to the MTR functions, we can obtain an approximation to the MTE function by equation (8) and compute approximations to other treatment effect parameters of interest.

3.4 Utilization for decision-making: a Bayesian approach

The approximation method described above allows us to estimate the parameters λ_1 and λ_0 from data. These estimates can then be used for decision-making, for instance, through the optimization problem given in Section 2.3.

To see this more clearly, we plug (14) back into (7), which yields the following approximated version of the firm's optimization problem:

$$\max_{\nu \in [0,1]} \left(\delta \times \left[\sum_{c=0}^{C-1} \lambda_{1c} \frac{\nu^{c+1}}{c+1} - \sum_{c=0}^{C} \lambda_{0c} \frac{\nu^{c+1}}{c+1} \right] - \kappa(\nu) \right), \tag{19}$$

A naive approach would be to plug estimates of λ_1 and λ_0, say, $\hat{\lambda}_1$ and $\hat{\lambda}_0$ into (19) and

solve for the optimal ν^*. It is simple to obtain these estimates: it suffices to replace the $\psi = \{\psi_1, \psi_0\}$s and $\nu = \{\nu_1, \nu_0\}$ in (17) and (18) with their estimates and compute the implied λs. Notice that this is straightforward because, from expressions (4) and (10), we can estimate the ψs and νs through simple sample averages.

However, this plug-in approach ignores the uncertainty around the estimates $\hat{\lambda}_1$ and $\hat{\lambda}_0$, which should be accounted for when solving a statistical decision theory problem. Even though there are many different criteria to solve such problems, we adopt a Bayesian approach due to its convenience. This approach integrates the objective function with respect to the unknown parameters—in this case, λ_1 and λ_0—using their posterior distribution given the data.

To be precise, denote this distribution by $p\left(\lambda_1, \lambda_0 | \text{data}\right)$. By adopting a Bayesian approach we solve the following problem:

$$\max_{\nu \in [0,1]} \left(\delta \times \int_{\lambda_1, \lambda_0} \left[\sum_{c=0}^{C-1} \lambda_{1c} \frac{\nu^{c+1}}{c+1} - \sum_{c=0}^{C} \lambda_{0c} \frac{\nu^{c+1}}{c+1} \right] p\left(\lambda_1, \lambda_0 | \text{data}\right) d\lambda_1 d\lambda_0 - \kappa(\nu) \right) =$$
$$\max_{\nu \in [0,1]} \left(\delta \times \left[\sum_{c=0}^{C-1} \mathbb{E}\left[\lambda_{1c} | \text{data}\right] \frac{\nu^{c+1}}{c+1} - \sum_{c=0}^{C} \mathbb{E}\left[\lambda_{0c} | \text{data}\right] \frac{\nu^{c+1}}{c+1} \right] - \kappa(\nu) \right). \tag{20}$$

Hence, this new objective function depends solely on the posterior expected λs given the data, which is a consequence of our approximation being linear in these parameters.

Directly obtaining $p\left(\lambda_1, \lambda_0 | \text{data}\right)$, and thus $\mathbb{E}\left[\lambda_1 | \text{data}\right]$ and $\mathbb{E}\left[\lambda_0 | \text{data}\right]$, is challenging. Nevertheless, it is straightforward to: derive the posterior distribution of ψ and ν given the data; take draws from this distribution; apply (17) and (18) using these draws to obtain draws from $p\left(\lambda_1, \lambda_0 | \text{data}\right)$; use these new draws to compute $\mathbb{E}\left[\lambda_1 | \text{data}\right]$ and $\mathbb{E}\left[\lambda_0 | \text{data}\right]$; and then solve the decision problem in (20).

We now describe how to obtain draws from $p\left(\lambda_1, \lambda_0 | \text{data}\right)$. To this end, we need to set priors over ψ and ν, denoted by $q(\psi, \nu)$, and the likelihood function of the data, $\ell\left(Y, D | \mathcal{C}, Z; \psi, \nu\right)$. We condition on Z and $\mathcal{C}$ because they are randomly chosen.

We need to consider two cases. First, notice that because $D = 0$ when $Z_c = 0$ for all c, we can pool all observations such that $Z_c = 0$ from all cells and use them to obtain the posterior distribution of ψ_{00}, as given in equation (13), conditional on the data. More precisely, given a prior distribution $q(\psi_{00})$ and the likelihood $\ell(Y | Z = 0; \psi_{00})$, we can derive the posterior $p(\psi_{00} | Y, Z = 0)$. The form of this distribution clearly depends on what type of variable Y is. In our application below, Y is binary. Hence, for convenience

we set:

$$\begin{aligned} &Y|Z=0;\psi_{00} \sim \text{Bernoulli}\left(\psi_{00}\right) \\ &\psi_{00} \sim \text{Beta}(\alpha_0,\beta_0) \end{aligned} \tag{21}$$

which implies that

$$\psi_{00}|Y,Z=0 \sim \text{Beta}\left(\alpha_0+\sum_{i:Z_{ic}=0} Y_i, n_0-\sum_{i:Z_{ic}=0} Y_i+\beta_0\right), \tag{22}$$

where n_0 is the total number of ineligible users.

The second case conditions on $\mathcal{C}=c$ and $Z_c=1$. Denote the number of observations in this set by n_c. For convenience, we proceed in two steps, relying on the factorization:

$$\begin{aligned} \ell\left(Y,D|\mathcal{C}=c,Z_c=1;\psi_{11c},\psi_{01c},\nu(z_c)\right) = \ell\left(Y|D,\mathcal{C}=c,Z_c=1;\psi_{d1c}\right) \\ \times \ell\left(D|\mathcal{C}=c,Z_c=1;\nu(z_c)\right) \end{aligned} \tag{23}$$

First, recall that D is binary, so we proceed as above:

$$\begin{aligned} &D|Z_c=1,\mathcal{C}=c;\nu(z_c) \sim \text{Bernoulli}\left(\nu(z_c)\right) \\ &\nu(z_c) \sim \text{Beta}(\alpha_{Dc},\beta_{Dc}) \\ &\nu(z_c)|D,Z_c=1,\mathcal{C}=c \sim \text{Beta}\left(\alpha_{Dc}+\sum_{i:Z_{ic}=1,\mathcal{C}_i=c} D_i, n_c-\sum_{i:Z_{ic}=1,\mathcal{C}_i=c} D_i+\beta_{Dc}\right). \end{aligned} \tag{24}$$

The second step consists of obtaining the posterior distribution of ψ_{d1c}, as given in equations (11) and (12), conditional on the data. Once again, the form of this distribution clearly depends on what type of variable Y is. Given our application, we proceed as in (21) and (22). Denote the number of observations such that $D=d$, $Z_c=1$ and $\mathcal{C}=c$ by n_{d1c}. Then:

$$\begin{aligned} &Y|D=d,Z_c=1,\mathcal{C}=c;\psi_{d1c} \sim \text{Bernoulli}\left(\psi_{d1c}\right) \\ &\psi_{d1c} \sim \text{Beta}(\alpha_{d1c},\beta_{d1c}) \\ &\psi_{d1c}|Y,D=d,Z_c=1,\mathcal{C}=c \sim \text{Beta}\left(\alpha_{d1c}+\sum_{i:D_i=d,Z_{ic}=1,\mathcal{C}_i=c} Y_i, n_{d1c}-\sum_{i:D_i=d,Z_{ic}=1,\mathcal{C}_i=c} Y_i+\beta_{d1c}\right). \end{aligned} \tag{25}$$

3.5 Implementing the multi-cell experiment

The campaign budget is a critical lever for the advertiser because it impacts reach among the target audience. First, we describe how a budget designated to a single-cell experiment can be redistributed to a multi-cell experiment conditional on C and on the associated $\nu(Z_c = 1)$. Second, the experimenter's choice of the number of cells overall, C, directly relates to the budget designated for the experiment itself. Together, the budget and number of cells impact the realized propensity scores, $\nu(Z_c = 1)$. We discuss how the experimenter can choose C and attempt to influence $\nu(Z_c = 1)$ to generate the necessary variation for our approximation method.

It is possible that more precise prescriptions for some of these quantities could be achieved with certain assumptions on the underlying DGP. In Appendix D, we consider assumptions that are commonly made in the literature (e.g., monotone treatment response) and verify whether these lead to restrictions on the DGP such that the implied MTE function is necessarily "well-behaved." Through a series of examples we demonstrate that this is *not* the case.

3.5.1 Budget allocation for the experiment

From a practical perspective, it is straightforward to reallocate the budget designated to a single-cell experiment across several cells with different propensity scores.

To see this, first assume that all users are targeted (there is no experiment). Denote the advertising budget by B. Let the target audience consist of a continuum of users, which we normalize to one. This implies that the effective advertising budget, that is, the budget per user, is also B. Finally, suppose that the cost function $\kappa(\cdot)$ is strictly increasing. Then, the resulting fraction of treated users, $\check{\nu}$, is $\check{\nu} = \kappa^{-1}(B)$.

Now assume that the experiment with one-sided noncompliance is conducted and fix $\Pr(Z_1 = 1)$. The effective budget then becomes $B_1 = \frac{B}{\Pr(Z_1=1)}$, so that $\nu(Z_1 = 1) = \kappa^{-1}\left(\frac{B}{\Pr(Z_1=1)}\right)$. Notice that $\nu(Z_1 = 1) = \check{\nu}$ can be achieved by setting the advertising budget allocated to this experiment to be $\check{B}_1 = \Pr(Z_1 = 1) \times B$, so that the effective budget becomes B.

Consider now a two-cell experiment, and fix $\Pr(\mathcal{C} = 1)$ and $\Pr(Z_1 = 1|\mathcal{C} = 1)$. Let σ_1 be the fraction of the original budget, B, allocated to Cell 1, so that $B_1 = \sigma_1 \times B$. The effective budget for Cell 1 is then $\frac{B_1}{\Pr(Z_1=1|\mathcal{C}=1)\times\Pr(\mathcal{C}=1)} = \frac{\sigma_1}{\Pr(Z_1=1|\mathcal{C}=1)\times\Pr(\mathcal{C}=1)} \times B$, so that $\nu(Z_1 = 1) = \kappa^{-1}\left(\frac{\sigma_1}{\Pr(Z_1=1|\mathcal{C}=1)\times\Pr(\mathcal{C}=1)} \times B\right)$. Equivalently, we then obtain $B_2 =$

$(1-\sigma_1)\times B$, so that the effective budget for Cell 2 becomes $\frac{B_2}{\Pr(Z_2=1|\mathcal{C}=2)\times(1-\Pr(\mathcal{C}=1))} = \frac{1-\sigma_1}{\Pr(Z_2=1|\mathcal{C}=2)\times(1-\Pr(\mathcal{C}=1))}\times B$ and $\nu(Z_2=1)=\kappa^{-1}\left(\frac{1-\sigma_1}{\Pr(Z_2=1|\mathcal{C}=2)\times(1-\Pr(\mathcal{C}=1))}\times B\right)$.

Hence, when running a two-cell design, given B and $\kappa(\cdot)$, the experimenter has four decision variables: $\Pr(\mathcal{C}=1)$, σ_1, $\Pr(Z_1=1|\mathcal{C}=1)$ and $\Pr(Z_2=1|\mathcal{C}=2)$. For the desired variation in the propensity score to be generated, it is necessary that $\frac{\sigma_1}{\Pr(Z_1=1|\mathcal{C}=1)\times\Pr(\mathcal{C}=1)} \neq \frac{1-\sigma_1}{\Pr(Z_2=1|\mathcal{C}=2)\times(1-\Pr(\mathcal{C}=1))}$. Thus, the experimenter can always guarantee that this constraint is satisfied.

This framework can be generalized to C cells in a straightforward manner. In this case, the experimenter has $3C-2$ variables: $\{\sigma_c, \Pr(\mathcal{C}=c)\}_{c=1}^{C-1}$ and $\{\Pr(Z_c=1|\mathcal{C}=c)\}_{c=1}^{C}$, under the constraints that, for all $c=1,\cdots,C$, $\sigma_c\in[0,1]$, $\Pr(Z_c=1|\mathcal{C}=c)\in[0,1]$ and $\Pr(\mathcal{C}=c)\in[0,1]$, plus $\sum_{c=1}^{C}\sigma_c=1$ and $\sum_{c=1}^{C}\Pr(\mathcal{C}=c)=1$. To ensure that the propensity scores differ from one another, it is then required that $\frac{\sigma_c}{\Pr(Z_c=1|\mathcal{C}=c)\times\Pr(\mathcal{C}=c)} \neq \frac{\sigma_{c'}}{\Pr(Z_{c'}=1|\mathcal{C}=c')\times\Pr(\mathcal{C}=c')}$ for all $c\neq c'$.

3.5.2 On the number of cells and propensity score values

We now discuss the choice of C and $\nu(Z_c=1)$ assuming that the budget for the experiment is not a concern.

Because a larger value for C translates into an approximating polynomial to the MTE function of higher order, it might be expected that more cells should yield a more credible approximation to the MTE function and, consequently, to the optimal decision the experimenter wishes to make. Nevertheless, as we show in Section 4.5, this need not be the case depending on the shape of the MTE function and on the values taken by the propensity score across cells.

Unfortunately, without strong assumptions, it is not possible to obtain sufficiently "well-behaved" MTE functions that allow for precise recommendations regarding what values the propensity score should take conditional on the number of cells. We found that commonly made assumptions in the literature such as positive correlations between the potential outcomes and the unobserved term that drives selection, monotonicity of the MTE function, or monotone treatment response (Manski, 2004), do not sufficiently discipline the MTE functions so as to enable precise guidance.

A sufficiently large budget and clear choices for the number of cells and propensity score values can enable the researcher to satisfactorily address identification of the MTE function, but in practice estimation can also be a concern to be addressed. With a finite number of observations, the number of cells in the experiment creates a type of bias-variance trade-

off. On the one hand, a higher number of cells generates more values of the propensity score, theoretically enabling a more flexible approximation of the MTE function, and thus decreasing bias. On the other hand, the number of observations per cell decreases, yielding noisier estimates of the approximating function, and thus increasing variance. Hence, the number of cells can be seen as somewhat akin to a bandwidth in nonparametric estimation. Once again, without strong assumptions on the underlying MTE function it is not possible to establish how to make progress on the task of choosing the number of cells vis-à-vis the available sample size.

4 Empirical application

We illustrate the value of our proposed multi-cell experimental design through a series of simulations calibrated to online advertising experiments at Facebook. We follow this simulation approach because we do not have data from a multi-cell experiment. Specifically, we use the results from a single-cell experiment with one-sided noncompliance to calibrate a set of data generating processes (DGPs). We then use these DGPs to simulate what our proposed multi-cell design would have produced had it been used instead of the typical single-cell design. The results confirm that our design enables the practitioner to approximate the underlying MTE function well.

We follow this analysis with results from a direct application of BMW's method. As we discussed above, the direct application of this method to data collected from a single-cell experiment with one-sided noncompliance requires the imposition of an additional restriction, for which there is little guidance. We demonstrate how sensitive the estimates can be to different possible restrictions. Using our simulated DGPs, we compare these estimates to the ones obtained from our approach and find that ours approximate the underlying MTE function better than all alternatives.

Then, we use the different approximations of the MTE function to derive the implied solutions to the optimization problem from equation (7), which was our original motivation behind this exercise instead of estimation of the MTE function. Once again, we find that our approach yields the solution that best approaches the true optimal solution, and, consequently, yields the lowest loss in expected profits.

Finally, we consider an example of a complex MTE function to illustrate how the performance of our method is impacted by the number of cells.

4.1 Data and Simulation Approach

Our simulation exercise is based on data from 15 large-scale online advertising experiments (or "studies") at Facebook used in Gordon et al. (2019), to which we direct the reader for more details on the experiments and underlying data. We display the key quantities in Table 1.

The number of observations per experiment ranges from 1,955,375 (Study 10) to 141,254,650 (Study 6). The fraction of units randomly assigned to be eligible to receive treatment, $\Pr(Z_c = 1)$, ranges from 0.17 (Study 7) to 0.85 (Study 2). There is substantial variation across studies in exposure rates conditional on eligibility, ranging from 0.066 (Study 9) to 0.81 (Study 15). This is relevant because the experimenter does not fully control this quantity.

Each study in Table 1 represents a single-cell experiment with one-sided noncompliance. As such, we observe the ATT and the expectations ψ_{11}, ψ_{01} and ψ_{00}, which correspond to the three regions in Figure 1b. These objects contain all the relevant information to estimate the MTE function.

In what follows, we focus solely on Study 4.[10] We use the data from this experiment to generate additional ψs, as defined in equation (10), that would have been generated with a multi-cell version of the experiment. We explain this calibration in detail below.

We proceed as follows. First, we specify functional forms for the MTR functions and choose their associated parameters to match the ψs we observe in the data. This allows us to generate the MTE function. Second, we consider the simplest version of our proposed design with only two cells, and choose the cell-specific eligibility probabilities and propensity scores. We also choose these values based on the quantities we observe in the data. Finally, using the MTE function and these probabilities we generate the additional ψs that would have been observed had this design been implemented through equations (11) and (12).

We start by specifying true MTR functions of the form:

$$\begin{aligned} m_1(u) &= m_{10} + m_{11}u + m_{12}u^2 \\ m_0(u) &= m_{00} + m_{01}u + m_{02}u^2 + m_{03}u^3. \end{aligned}$$

We chose these functional forms because they are the polynomials of lowest order that the simplest version of our design—with only two cells—cannot recover. With three or more cells, our approach can perfectly recover the true MTR functions. These forms imply that the MTE is a cubic polynomial.

[10] Future versions of this paper will include results based on simulations from the other experiments.

Table 1: Summary of data

Study	Vertical	Outcome	Observations	$\Pr(Z=1)$	$\nu(1)$	ψ_{11}	ψ_{01}	ψ_{00}	ATT
1	Retail	Checkout	2,427,494	0.50	0.76	0.115	0.017	0.105	0.035
2	Financial services	Checkout	86,183,523	0.85	0.48	0.026	0.007	0.033	0.001
3	E-commerce	Checkout	4,672,112	0.50	0.66	0.172	0.045	0.203	0.021
4	Retail	Checkout	25,553,093	0.70	0.37	0.029	0.016	0.033	0.033
5	E-commerce	Checkout	18,486,000	0.50	0.30	0.017	0.006	0.009	0.045
7	Retail	Checkout	67,398,350	0.17	0.51	0.145	0.106	0.247	0.007
8	E-commerce	Checkout	8,333,319	0.50	0.26	0.018	0.029	0.047	-0.002
9	E-commerce	Checkout	71,068,955	0.75	0.066	0.139	0.049	0.184	0.049
10	Tech	Checkout	1,955,375	0.60	0.65	0.008	0.032	0.113	0.003
11	E-commerce	Checkout	13,339,044	0.50	0.42	0.205	0.072	0.261	0.039
12	Retail	Checkout	5,566,367	0.50	0.77	4.93	0.646	5.487	0.078
13	E-commerce	Checkout	3,716,015	0.77	0.30	0.056	0.216	0.282	-0.033
14	E-commerce	Checkout	86,766,019	0.80	0.35	0.024	0.012	0.027	0.026
15	Retail	Checkout	9,753,847	0.50	0.81	1.19	0.223	1.385	0.034
1	Retail	Registration	2,427,494	0.50	0.76	0.551	0.015	0.078	0.643
5	E-commerce	Registration	18,486,000	0.50	0.30	0.298	0.048	0.078	0.893
8	E-commerce	Registration	8,333,319	0.50	0.26	0.007	0.006	0.01	0.010
10	Tech	Registration	1,955,375	0.60	0.65	0.275	0.110	0.363	0.033
14	E-commerce	Registration	86,766,019	0.80	0.35	0.225	0.077	0.165	0.393
2	Financial services	Page views	86,183,523	0.85	0.48	0.120	0.004	0.011	0.233
5	E-commerce	Page views	18,486,000	0.50	0.30	0.226	0.053	0.084	0.647
6	Telecom	Page views	141,254,650	0.75	0.61	0.340	0.059	0.356	0.069

For this exercise, we need to choose parameter values such that the implied moments match the observed ψ_{11}, ψ_{01} and ψ_{00}. Under these functional forms, this implies that we need to choose seven parameters to satisfy only three constraints subject to the additional constraint that the MTR functions have to be between 0 and 1, because the outcome of Study 4 is binary (checkout). We consider two sets of parameters to illustrate our proposed approach.

The resulting MTE functions are shown in Figure 2. Both DGPs correspond to functions that are very close to quadratic, so we expect that a two-cell design will suffice to obtain good approximations.

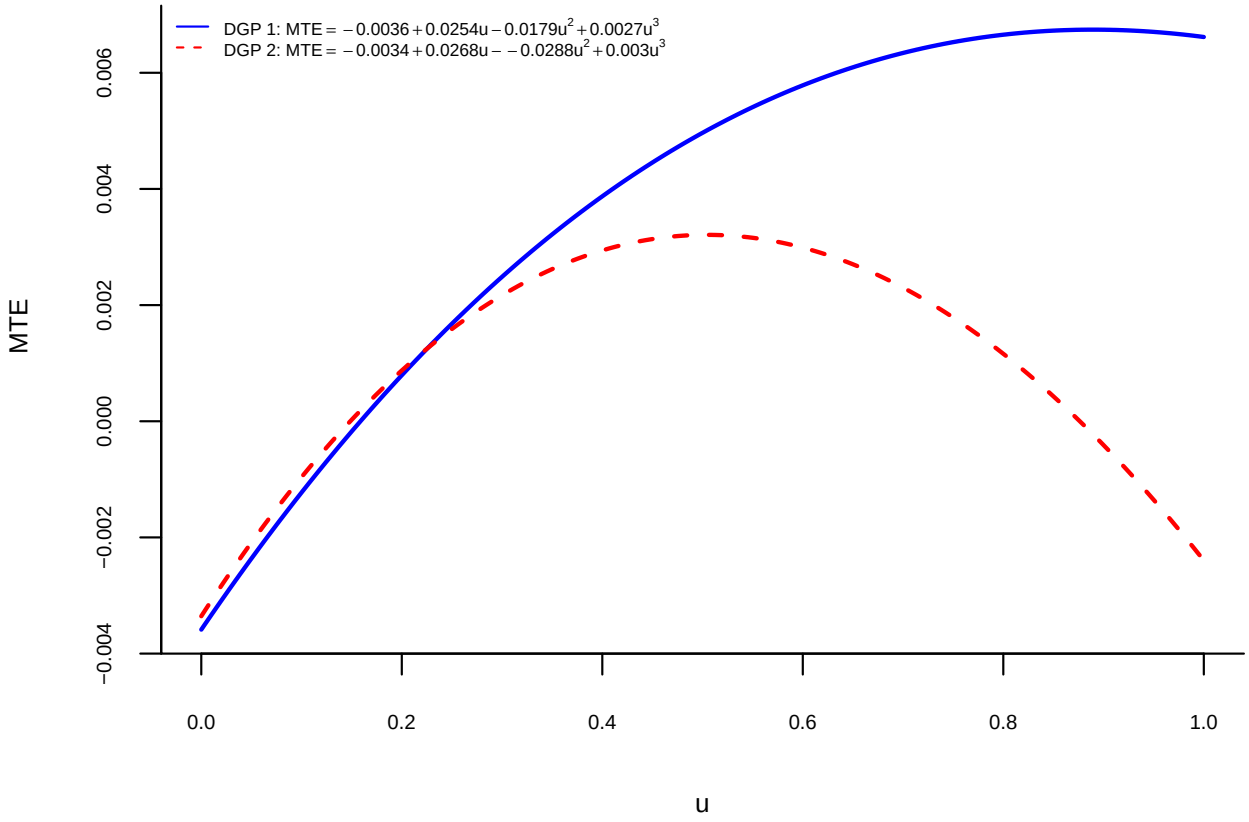


Figure 2: Simulated DGPs from Study 4

We use the MTE functions above to simulate data from our proposed experimental design with $C = 2$ cells. Notice that this design limits us to a linear approximation to $m_1(\cdot)$ and a quadratic approximation to $m_0(\cdot)$.

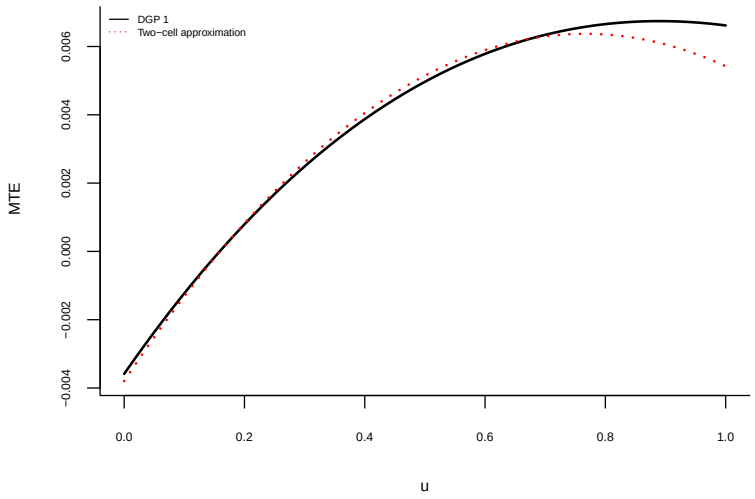
The first step it to set the propensity scores. Because Study 4 assigned users to be eligible to receive treatment with probability 0.7, we consider this to be one cell and add a second cell in which this probability equals 0.3. For the first cell we keep the propensity score at the original value, $\nu(Z_1 = 1) = 0.37$, and set the propensity score for the second cell so

that $\Pr(Z_1 = 1|C = 1) \times \nu(Z_1 = 1) = \Pr(Z_2 = 1|C = 2) \times \nu(Z_2 = 1)$, implying that $\nu(Z_2 = 1) \approx 0.86$.

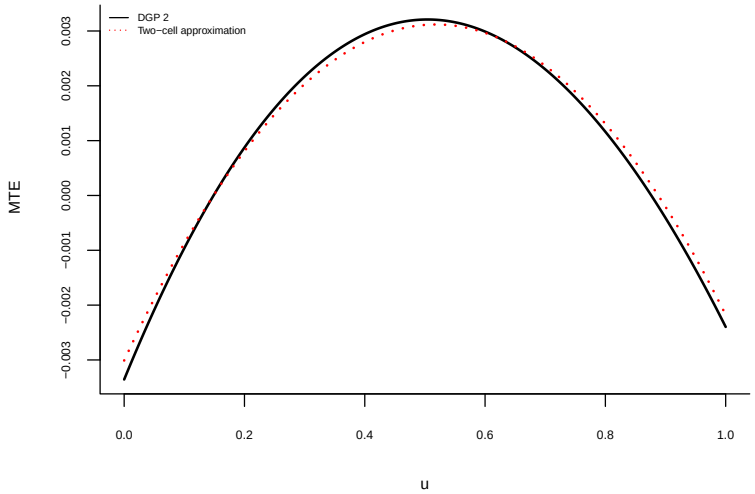
We use these propensity scores and the underlying MTE functions to compute the ψs from equations (11) and (12) that would have been observed had our design been implemented.

4.2 Results from our proposed approach

We use the ψs from above to obtain the approximated MTE functions following the procedure we described in Section 3.3. The resulting approximations to the DGPs we consider are shown in Figure 3.



(a) DGP 1: $\text{MTE}(u) = -0.44 + 2.89u - 2.22u^2 + 0.27u^3$



(b) DGP 2: $\text{MTE}(u) = -0.41 + 3.04u - 3.32u^2 + 0.3u^3$

Figure 3: Approximations to simulated DGPs from Study 4

Although the underlying MTE functions are cubic, the fact that their shapes are close to quadratic implies that a two-cell design yields good approximations. We quantify the quality of these approximations through three metrics. Denote the approximation to the true MTE function by $\text{MTE}_{\text{app}}(\cdot)$. The metrics we consider are the sup-norm, $\max_{u\in[0,1]} \left|\text{MTE}(u) - \text{MTE}_{\text{app}}(u)\right|$, and the L_2-norm, $\sqrt{\left(\int_0^1 \left[\text{MTE}(u) - \text{MTE}_{\text{app}}(u)\right]^2 du\right)}$. In addition, we consider the quality of the approximated ATE, ATE_{app}, relative to the true value, which we refer to as "ATE-norm:" $\frac{\text{ATE}_{\text{app}} - \text{ATE}}{\text{ATE}}$. We consider this metric as a different way of summarizing the discrepancy between the true and approximated MTE functions because often the ATE is the treatment effect parameter of original interest to the researcher.

The results are given in Table 2. Overall, our method generates a small difference between the approximations and their true values. These results will be more interpretable when made in comparison to others in the following sections.

Table 2: Distance between true and approximated MTE functions: two-cell design

Metric	DGP 1	DGP 2
sup-norm	0.00019	0.00015
L_2-norm	0.00036	0.00014
ATE-norm	-0.03009	0.02432

4.3 Direct application of BMW's method with only one cell

BMW's method *can* be applied to data obtained from the typical experimental design. However, because this design yields one-sided noncompliance, $\nu(0) = 0$ while $\nu(1) \in (0,1)$. Thus, equations (15) and (16) imply that the only moments identified from the data it provides are ψ_{11}, ψ_{01} and ψ_{00}, where we omit the subscript c to ease notation since there is only one cell. Based on the logic from equations (17) and (18), these moments allow us to approximate $m_1(\cdot)$ with a constant function and $m_0(\cdot)$ with a linear function. Consequently the MTE function itself can be approximated with a linear function.

The ability to approximate the MTE function with a linear function might seem attractive, especially because it is not uncommon to maintain the assumption that the MTE function

is indeed linear.[11] Nevertheless, the missingness of ψ_{10} implies that this approximation inherently features a restriction that nontrivially impacts not only the quality of the approximation, but also the structure of the endogeneity of treatment.

To see this, suppose for simplicity that $m_d(u) = \lambda_{d0} + \lambda_{d1}u$ for $d \in \{0, 1\}$. It follows that:

$$\mathbb{E}[Y|D = 1, Z = z] = \lambda_{10} + \frac{\lambda_{11}}{2}\nu(z)$$
$$\mathbb{E}[Y|D = 0, Z = z] = \left(\lambda_{00} + \frac{\lambda_{01}}{2}\right) + \frac{\lambda_{01}}{2}\nu(z).$$

The data obtained from the typical experimental design thus enable us to recover λ_{00} and λ_{01}. Nevertheless, they do not allow us to recover λ_{10} and λ_{11} separately because we do not observe ψ_{10}. This is an underidentification problem: there are four parameters to be estimated (λ_{10}, λ_{11}, λ_{00} and λ_{01}) but only three moments available to estimate such parameters (ψ_{11}, ψ_{01} and ψ_{00}).

By approximating $m_1(\cdot)$ with a constant, the practitioner effectively imposes that $\lambda_{11} = 0$, enabling them to estimate λ_{10}. From a purely mechanical perspective, the higher $|\lambda_{11}|$ is, the lower the quality of the approximation. However, the constraint $\lambda_{11} = 0$ also has a deeper structural implication. It implies that all endogeneity stems from Y_0. This rules out certain forms of self-selection, such as the classic case of $D = \mathbb{1}\{Y_1 \geq Y_0\}$, where only units that benefit from treatment are treated. In our setting of online advertising, this precludes "perfect" ad exposure where units are exposed to the ad only if this benefits the advertiser.

Even though it is arguably more natural to set $\lambda_{11} = 0$ in accordance with the approximation method from Section 3.3, there are other constraints the practitioner might want to impose instead that are justifiable. A stronger restriction is to set $\lambda_{11} = \lambda_{01}$, thereby ruling out endogeneity altogether but also allowing us to recover λ_{10}. However, this is often considered implausible, including in the context of online advertising.

Alternatively, the practitioner can impose $\lambda_{10} = 0$, allowing them to recover λ_{11}. Given linearity, this restriction implies that $m_1(\cdot)$ is either always negative or always positive, which can be justified in cases where Y_1 is bounded either from above or below at 0. The same can be achieved for the MTE function by imposing that $\lambda_{10} = \lambda_{00}$, while also enabling the estimation of λ_{11}.

Unfortunately, we are unaware of any general theory or methodology that can provide clear guidance on which assumption above is most reasonable, although progress has

[11] Examples of studies that maintained this linearity assumption are Olsen (1980), Moffitt (2008), French and Song (2014), Brinch et al. (2017) and Kowalski (2021).

been recently made in this direction (e.g., Kowalski, 2023). Whichever restriction is chosen, a key point is that such additional constraint must *always* be imposed to implement BMW's estimator using data collected from an experiment with one-sided noncompliance, which, to our knowledge, has not been noted in the literature. Given the pervasiveness of such experimental designs, we hope our approach provides a valuable solution to recover the MTE function more credibly.

Below we present results from the direct application of BMW's method under the four aforementioned constraints to both of the DGPs. It is important to note that these four approximations will be the same for both DGPs because each was generated using the same values for $\{\psi_{11}, \psi_{01}, \psi_{00}\}$. Our multi-cell approach yields specific approximations to each DGP because it relies on different moments that depend on the underlying DGP. Figure 4 depicts the four approximated MTE functions.

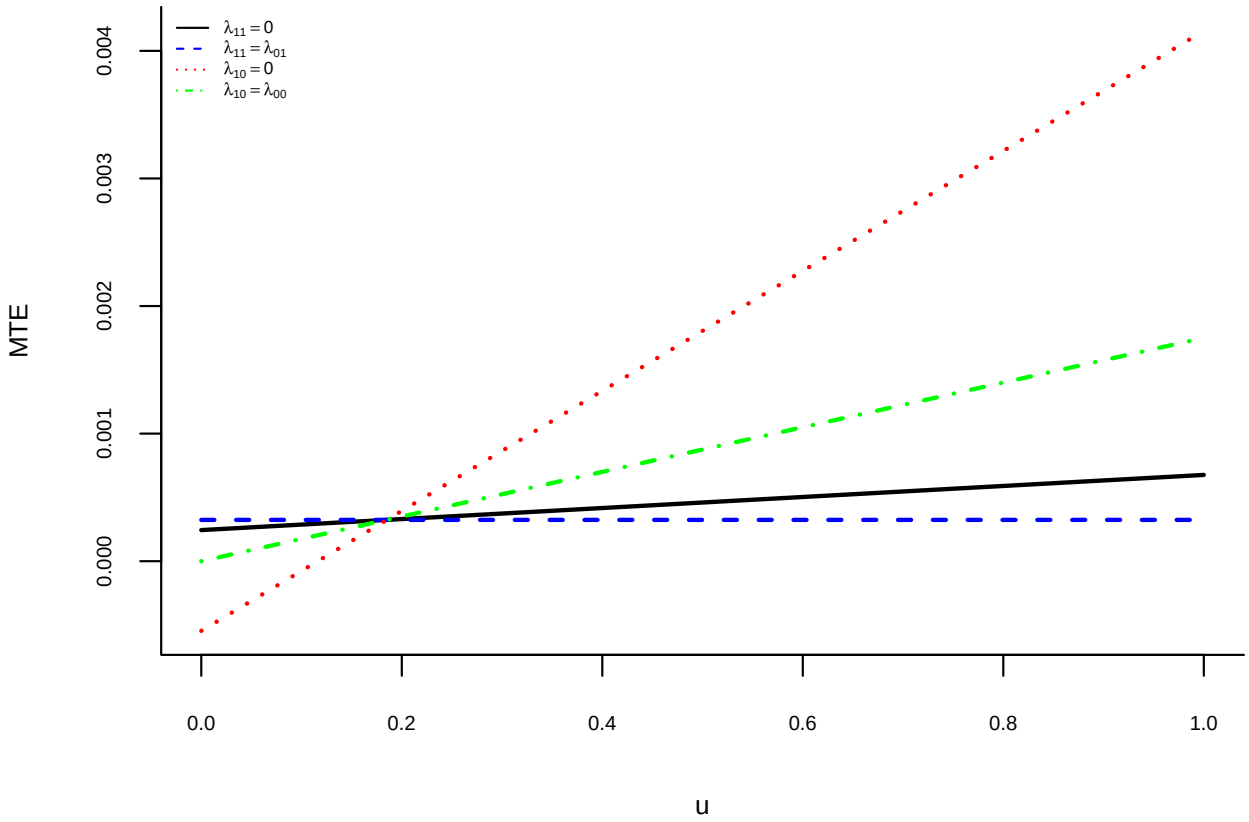


Figure 4: Approximations of MTE function from direct application of BMW's method

There is noticeable variation across the approximate MTE functions. The slopes range from 0 ($\lambda_{11} = \lambda_{01}$) to 0.0047 ($\lambda_{10} = 0$), which, as we show below, impacts decision-making considerably. Notice that setting $\lambda_{11} = \lambda_{01}$ imposes that the slope equals 0 because it implies that the MTE function does not depend on u, which is equivalent to ruling out

endogeneity in treatment assignment.

A different way of illustrating this variation is through the distance between the approximated MTE function and the true MTE, which are shown in Table 3. Once again, we highlight the restrictions that yield the best approximations for each DGP-metric pair.

Table 3: Distance between true and approximated MTE functions: one-cell design with additional restriction

Restriction	DGP 1	DGP 2	DGP 1	DGP 2	DGP 1	DGP 2
	sup-norm		L_2-norm		ATE-norm	
$\lambda_{11} = 0$	0.0061	0.0027	0.0045	0.002	-0.8794	-0.6103
$\lambda_{11} = \lambda_{01}$	0.0064	0.0029	0.0046	0.002	-0.9151	-0.7257
$\lambda_{10} = 0$	0.0036	0.0016	0.0027	0.0022	-0.5269	0.5294
$\lambda_{10} = \lambda_{00}$	0.0053	0.0024	0.0039	0.0019	-0.7707	-0.2586

The results from Tables 2 and 3 demonstrate that the two-cell design always gives a better approximation to the MTE function regardless of the DGP and of the metric used. This is unsurprising, especially in light of Figures 3 and 4.

Two results are particularly notable. First, no parameter restriction yields an approximation that always dominates the others. Second, depending on the metric used to measure the quality of the approximation, the best restriction can change. This can be seen for DGP 2: under the sup-norm, the restriction $\lambda_{10} = 0$ yields the best approximation, while under the L_2-norm, $\lambda_{10} = \lambda_{00}$.

Nevertheless, even though the direct applications of BMW provide relatively poor approximations of the MTE function, this does not necessarily imply that they are not helpful in informing decision-making. For the purposes of the latter, these approximations can work well if they imply an approximated expected profit function whose maximum is close to that of the true expected profit function. We investigate this matter below.

4.4 Implications for decision-making

The goal of experimentation often is to recover the PRTE given by a specific decision problem, which frequently does not correspond to typical treatment effect parameters such as the ATE or the ATT. Although the approaches in the previous subsection yield poor approximations of the MTE function, it is possible that they perform well when it comes to the implied optimal decisions.

We consider the firm's decision problem given in equation (7). For the sake of illustration, we set $\delta = 1$ and $\kappa(\nu) = 0.1\nu^4$. A firm can set δ based on their internal assessment of the value of a conversion event. We specify $\kappa(\nu)$ as convex to capture the notion that reaching the marginal consumer becomes more expensive as overall campaign reach increases. Most advertising platforms provide tools to advertisers to help them predict how reach is expected to vary as a function of their budget.[12] Based on the simulated DGPs we outlined above, the resulting expected profit functions are given in Figure 5.

The expected profit functions reflect the differences across the different DGPs shown in Figure 2. They demonstrate how different MTE functions can affect optimal decisions. In this case, the optimal decisions associated with DGPs 1 and 2 are to treat 100% and 75.5% of the population, respectively.

[12]For example, advertiser tools to estimate campaign audience size are offered by Google (https://support.google.com/google-ads/answer/2475441?hl=en) and Meta (https://www.facebookblueprint.com/student/activity/212722).

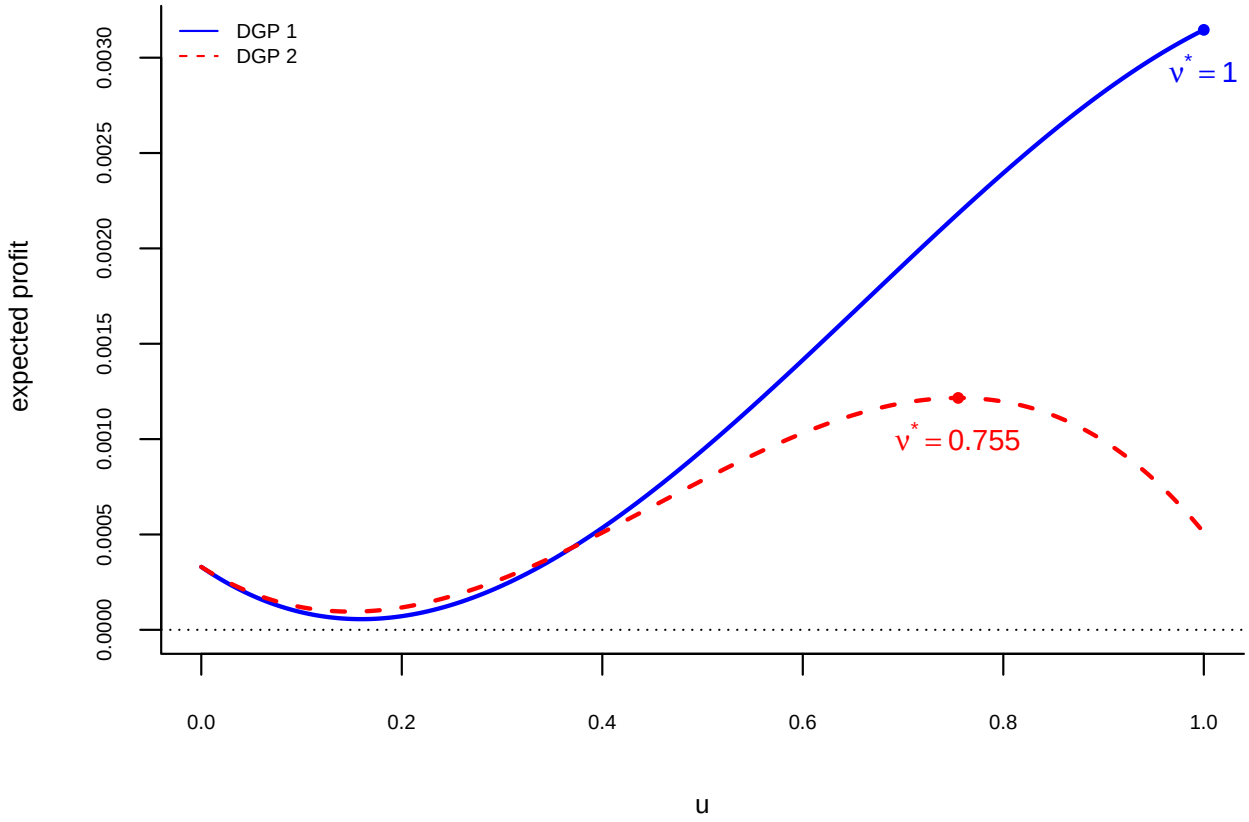


Figure 5: Expected profit functions from simulated DGPs of Study 4

We now compare the true optimal solutions to what the decision-maker would do if information obtained from our experimental design was available, following the procedure we presented in Section 3.4. As a comparison, we also consider results from using the four different results obtained from applying BMW's method directly under the different additional restrictions outlined above. Table 4 shows the expected profit losses from using these different approaches for each of the three DGPs.

Table 4: Decisions from approximated expected profit functions

DGP	True ν^*	$\lambda_{11} = 0$		$\lambda_{11} = \lambda_{01}$		$\lambda_{10} = 0$		$\lambda_{10} = \lambda_{00}$		Multi-cell	
		ν^*	Loss	ν^*	Loss	ν^*	Loss	ν^*	Loss	ν^*	Loss
1	1	0.484	0.723	0.432	0.792	1	0	0.661	0.454	1	0
2	0.755	0.484	0.391	0.432	0.509	1	0.58	0.661	0.061	0.762	0.0003

The multi-cell approach yields virtually no losses across all DGPs and thus dominates the

four direct applications of BMW. The magnitude of the expected profit losses is associated with the quality of the approximation to the underlying MTE function, reflecting the patterns observed in Figure 3.

In turn, the direct applications of BMW do not show a systematic pattern. This is perhaps to be expected: Figure 5 shows that the optimal solutions differ substantially across the different DGPs. However, direct applications of BMW's method yield the same solution for each possible additional constraint across all DGPs. Consequently, when one solution performs well under a given DGP, it will probably perform poorly under a different DGP.

The additional restriction that yields the lowest losses for DGPs 1, 2 and 3 are $\lambda_{10} = 0$, $\lambda_{10} = \lambda_{00}$ and $\lambda_{11} = \lambda_{01}$ (which interestingly rules out endogeneity of treatment), respectively. These results agreement with those from Table 3.

Notice that setting $\lambda_{10} = 0$ eliminates expected profit losses under DGP 1 and setting $\lambda_{10} = \lambda_{00}$ yields low losses under DGP 2. The contrast between these findings and those from Table 3 show that approximating the MTE function and minimizing expected profit losses are connected, but not fully aligned, tasks.

4.5 A more complex DGP

So far we have focused on cubic MTE functions. Given that this is a simple functional form, it might be considered unsurprising that our approach can perform satisfactorily. We now consider a more complex MTE function to dig deeper into the ability of our proposed multi-cell experimental design to provide a good approximation of this function and to inform decision-making. In particular, we assess how the performance of our approach changes as the number of cells increases.

4.5.1 New MTE and expected profit functions

We choose $m_1(u) = \urcorner \frac{1}{1+u}$ and $m_0(u) = \lrcorner \frac{1}{(1+u)^2} + \sqsupset \sin^2(2\pi u)$. As before, the parameters $\urcorner$, and $\sqsupset$ and $\lrcorner$ are computed to match the observed ψ_{11}, ψ_{01} and ψ_{00}. Figure 6 depicts the resulting MTE function, and we refer to this as DGP 4. We consider this function to be "complex" because is not monotonic, concave or convex over the entire domain.

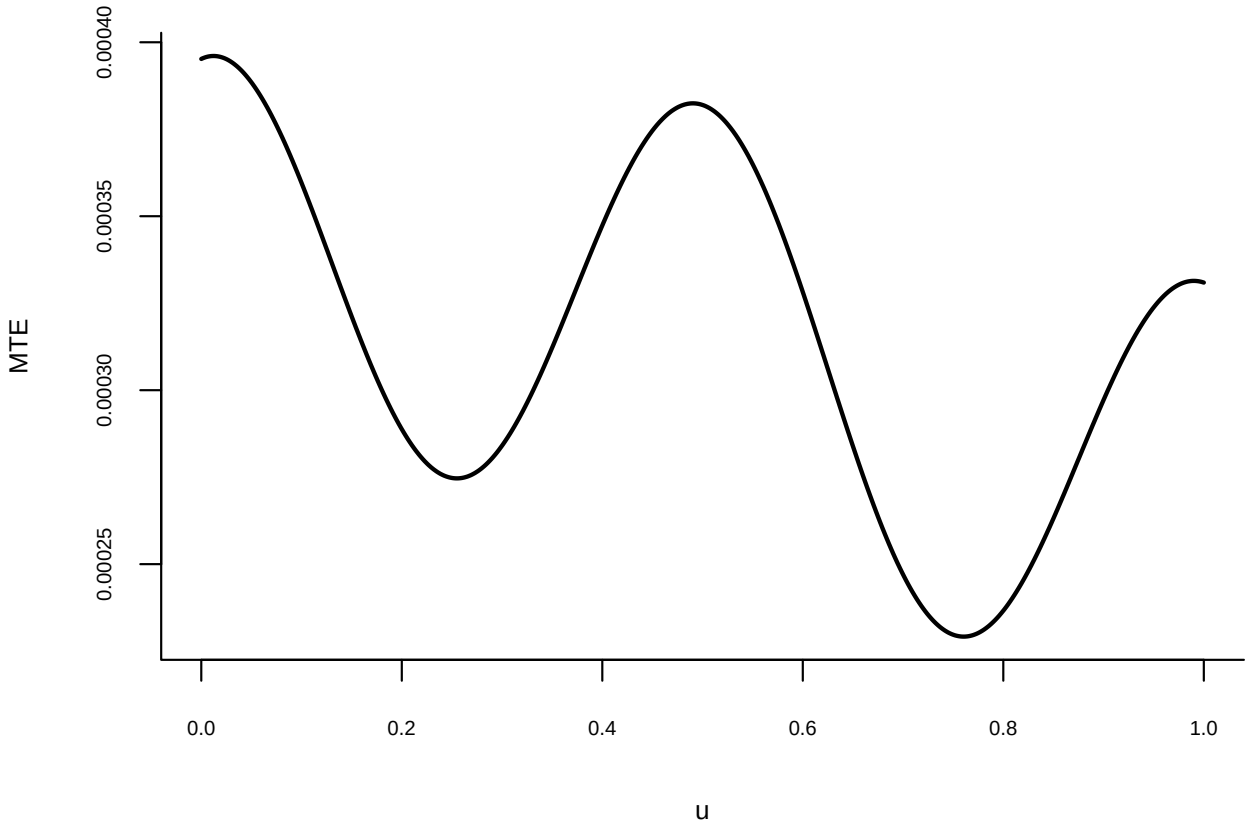


Figure 6: $\text{MTE}(u) = \beth_{\frac{1}{1+u}} - \beth_{\frac{1}{(1+u)^2}} - \beth \sin^2(2\pi u)$

Figure 7 shows the implied expected profit function. Under DGP 4, the optimal decision is to treat 91.54% of the population.

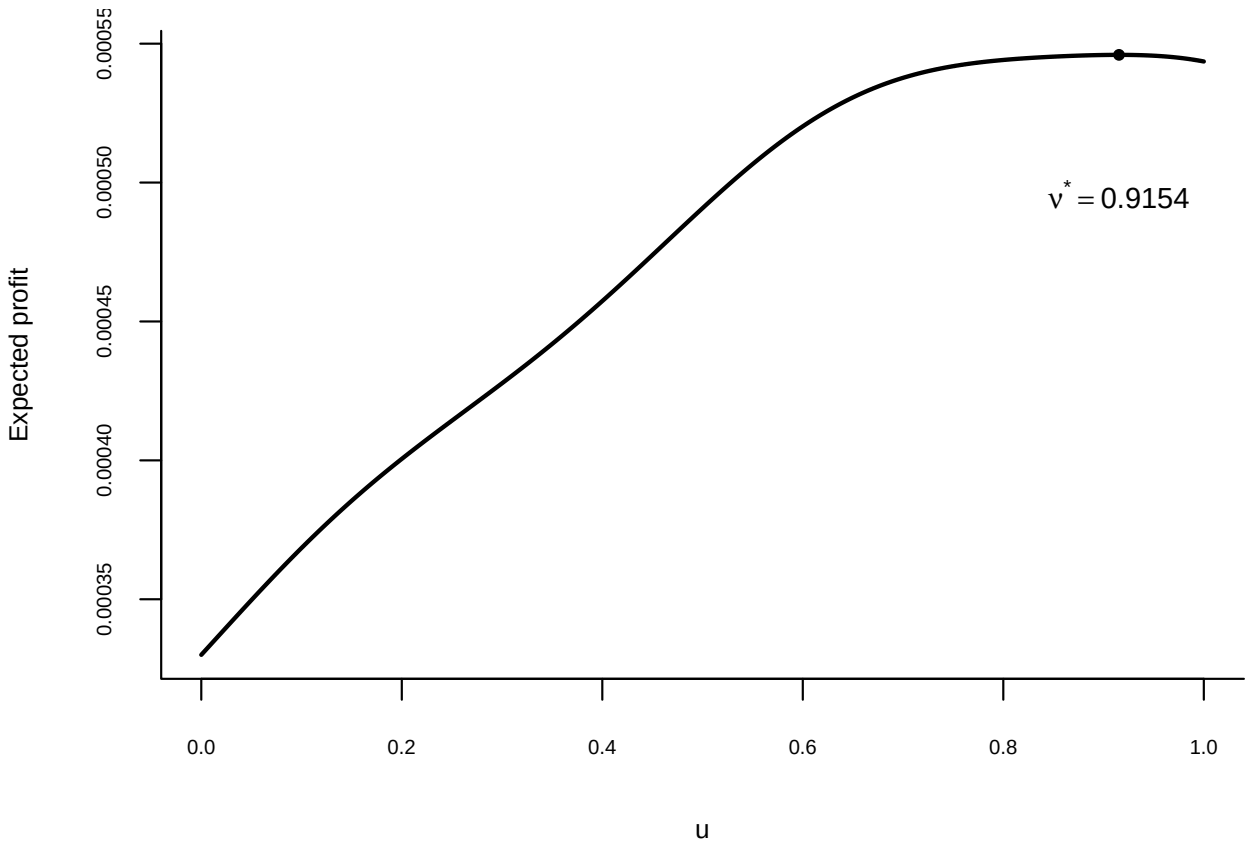


Figure 7: Expected profit function under complex MTE

4.5.2 Approximations based on different numbers of cells

We now study how the quality of approximation changes as more cells are included into the experiment. In particular, in addition to the two-cell design, we consider a three-cell design and a five-cell design, which allow us to obtain a cubic and a quintic approximation to the MTE function, respectively. Cells 1 and 2 display the same eligibility probabilities and propensity scores as above. Cell 3 has $\Pr(Z_3 = 1|\mathcal{C} = 3) = 0.5$ and $\nu(Z_3 = 1)$, chosen so that $\Pr(Z_1 = 1|\mathcal{C} = 1) \times \nu(Z_1 = 1) = \Pr(Z_3 = 1|\mathcal{C} = 3) \times \nu(Z_3 = 1)$. We set $\Pr(Z_4 = 1|\mathcal{C} = 4) = 0.25$ and $\Pr(Z_5 = 1||\mathcal{C} = 5) = 0.9$, and set $\nu(Z_5 = 1)$ so that $\Pr(Z_1 = 1|\mathcal{C} = 1) \times \nu(Z_1 = 1) = \Pr(Z_5 = 1|\mathcal{C} = 5) \times \nu(Z_5 = 1)$, obtaining $\nu(Z_5 = 1) \approx 0.288$. Finally, we pick $\nu(Z_4 = 1) = 0.17$ to increase the range of values covered by the propensity scores. Figure 8 shows the four approximated MTE functions along with the true one.

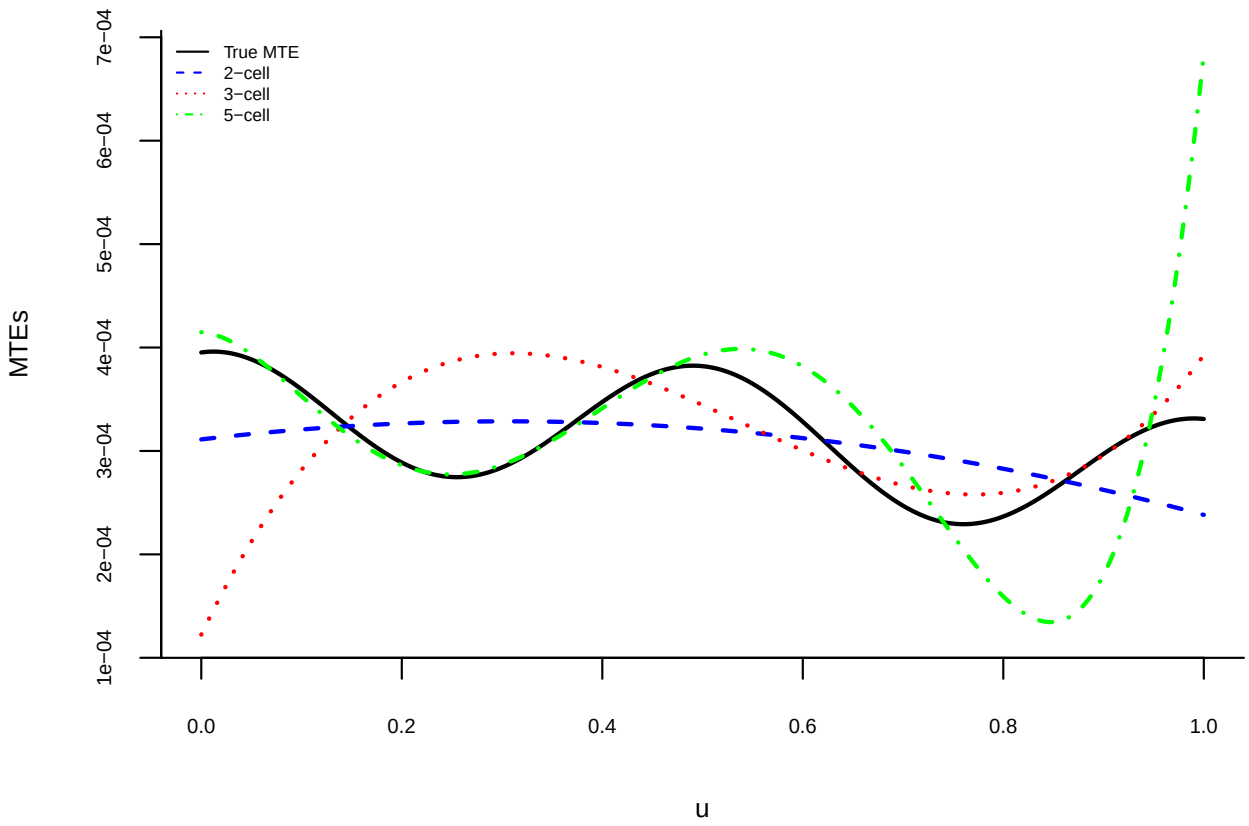


Figure 8: $\text{MTE}(u) = \urcorner\frac{1}{1+u} - \lrcorner\frac{1}{(1+u)^2} - \supset \sin^2(2\pi u)$

Seemingly as the number of cells increases, so does the quality of the approximation. Nevertheless, we note that at the extreme points of the domain these approximations might worsen because of the extrapolation imposed by the polynomial functional form. To quantify and assess its impact we compute the distance between the true and approximated MTE functions using the aforementioned sup-norm and L_2-norm. These quantities are shown in Table 5, along with the optimal decisions implied by these approximations and the losses in expected profit they imply.

Table 5: Closeness to MTE function and ν^* as a function of the number of cells

Method	sup-norm	L_2-norm	ATE-norm	ν^*	Loss
Two-cell	0.00009	0.00005	-0.02445	0.8753	0.00049
Three-cell	0.00027	0.00008	0.00585	0.8824	0.00034
Five-cell	0.00035	0.00006	0.00144	1	0.00437
True	—	—	—	0.9154	—

The different criteria demonstrate that, in this example, no single approximation strictly dominates the others. Interestingly, according to the criteria that measure discrepancy between the entire MTE function and its approximation, the simplest approximation, with only two cells, performs best. In turn, the five-cell approximation is the one whose implied ATE is closest to the true ATE. Nevertheless, the three-cell approximation is the one that implies the smallest loss in expected profit.

These results further highlight that, even though they are connected, the tasks of approximating the MTE function and approximating the maximum of the expected profit function are not perfectly aligned. The extent to which these approximations are misaligned depends on the underlying MTE and cost functions.

5 Conclusion

Experiments are considered an especially attractive tool to estimate the impacts of treatments and interventions. When treatment assignment cannot be randomized, a common approach is to randomize eligibility to receive treatment instead, leading to one-sided noncompliance. Nevertheless, decision-makers who conduct experiments are often interested in obtaining information to assist them in making specific decisions, and not just measuring the effects of treatment per se. Unfortunately, typical experimental designs, such as the one where eligibility to receive treatment is randomized, do not provide enough information to assist with many decisions.

This paper proposed an approach to obtain such information. This approach combines a novel multi-cell experimental design and modern estimation techniques, where the for-

mer leads to the collection of data that contain more information about treatment effects and the latter leverages this information. Our approach leverages the method from Brinch et al. (2017), which we point out requires an arbitrary additional assumption to be applied to experiments with one-sided noncompliance.

Using data from online advertising experiments at Facebook, we addressed the performance of our proposed multi-cell experimental design vis-á-vis that of the typical experimental design. To do so, we conducted simulation exercises where we implemented the aforementioned estimators on data collected from experiments that followed each of these two designs. We found that the estimates obtained from both estimators under our design dominate those from the typical design, which, in turn, led to more accurate decision-making. This not only shows how our experimental design works in conjunction with modern estimation techniques, but also that it does lead to better estimates and decisions.

References

Baardman, L., Fata, E., Pani, A., and Perakis, G. (2019). Learning optimal online advertising portfolios with periodic budgets. *SSRN Working Paper 3346642*.

Barajas, J. and Bhamidipati, N. (2021). Incrementality testing in programmatic advertising: Enhanced precision with double-blind designs. In Leskovec, J., Grobelnik, M., Najork, M., Tang, J., and Zia, L., editors, *Proceedings of the Web Conference 2021 (WWW '21)*, pages 2818–2827, New York, USA. ACM.

Basu, A. K. and Batra, R. (1988). ADSPLIT: A multi-brand advertising budget allocation model. *Journal of Advertising*, 17(2):44–51.

Björklund, A. and Moffitt, R. (1987). The estimation of wage gains and welfare gains in self-selection models. *Review of Economics and Statistics*, 69(1):42–49.

Brinch, C. N., Mogstad, M., and Wiswall, M. (2017). Beyond LATE with a discrete instrument. *Journal of Political Economy*, 125(4):985–1039.

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics*, 9(1):247–274.

Chetty, R., Hendren, N., and Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. *American Economic Review*, 106(4):855–902.

Crépon, B., Devoto, F., Duflo, E., and Parienté, W. (2015). Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco. *American Economic Journal: Applied Economics*, 7(1):123–150.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. (2013). Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. *Quarterly Journal of Economics*, 128(2):531–580.

Daljord, Ø., Mela, Carl F. and Roos, M. J., Sprigg, J., and Yao, S. (2022). The design and targeting of compliance promotions. *Marketing Science*, Forthcoming.

Deng, A., Yuan, L.-H., and Salama-Manteau, A. (2019). Instrumental variable estimators and bounds on treatment effects. *arXiv preprint arXiv:2112.13299*.

French, E. and Song, J. (2014). The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy*, 6(2):291–337.

Geng, T., Sun, F., Wu, D., Zhou, W., Nair, H. S., and Lin, Z. (2021). Automated bidding and budget optimization for performance advertising campaigns. *SSRN Working Paper 3913039*.

Gordon, B. R., Moakler, R., and Zettelmeyer, F. (2022). Close enough? A large-scale exploration of non-experimental approaches to advertising measurement. *Marketing Science*, forthcoming.

Gordon, B. R., Zettelmeyer, F., Bhargava, N., and Chapsky, D. (2019). A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. *Marketing Science*, 38(2):193–225.

Green, D. P., Gerber, A. S., and Nickerson, D. W. (2003). Getting out the vote in the local elections: Results from six door-to-door canvassing experiments. *Journal of Politics*, 65(4):653–663.

Gui, G., Nair, H. S., and Niu, F. (2021). Auction throttling and causal inference of online advertising effects. *arXiv preprint arXiv:2112.15155*.

Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–738.

Holthausen Jr., D. M. and Assmus, G. (1982). Advertising budget allocation under uncertainty. *Management Science*, 28(5):487–499.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Johnson, G. A. (2022). Inferno: A guide to field experiments in online display advertising.

Johnson, G. A., Lewis, R. A., and Nubbemeyer, E. I. (2017a). The online ad effectiveness funnel & carryover: Lessons from 432 field experiments. *SSRN Working Paper 2701578*.

Johnson, G. A., Lewis, R. A., and Reiley, D. H. (2016). Location, location, location: Repetition and proximity increase advertising effectiveness. *SSRN Working Paper 2268215*.

Johnson, G. A., Lewis, R. A., and Reiley, D. H. (2017b). When less is more: Data and power in advertising experiments. *Marketing Science*, 36(1):43–53.

Kowalski, A. E. (2021). Reconciling seemingly contradictory results from the Oregon Health Insurance Experiment and the Massachussets Health Reform. *Review of Economics and Statistics*, forthcoming.

Kowalski, A. E. (2023). Behavior within a clinical trial and implications for mammography guidelines. *Review of Economic Studies*, 90(1):432–462.

Lewis, R. A. and Reiley, D. H. (2014). Online ads and offline sales: Measuring the effect of retail advertising via a controlled experiment on Yahoo! *Quantitative Marketing and Economics*, 12(3):43–53.

Lewis, R. A. and Wong, J. (2018). Incrementality bidding & attribution.

Manski, C. F. (2004). Monotone treatment response. *Econometrica*, 65(6):1311–1334.

Mealli, F., Imbens, G. W., Ferro, S., and Biggeri, A. (2004). Analyzing a randomized trial on breast self-examination with noncompliance and missing outcomes. *Biostatistics*, 59(2):207–222.

Miguel, E. and Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.

Moffitt, R. (2008). Estimating marginal treatment effects in heterogeneous populations. *Annales d'Économie et Statistique*, 91/92:239–261.

Mogstad, M., Santos, A., and Torgovitsky, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, 86(5):1589–1619.

Olsen, R. J. (1980). A least squares correction for selectivity bias. *Econometrica*, 48(7):1815–1820.

Pani, A. A., Raghavan, S., and Sahin, M. (2017). Large-scake advertising portfolio optimization in online marketing.

Sahni, N. S., Narayanan, S., and Kalyanam, K. (2019). An experimental investigation of the effects of retargeted advertising: The role of frequency and timing. *Journal of Marketing Research*, 56(3):401–418.

Schochet, P. Z., Burghardt, J., and Sheena, M. (2008). Does Job Corps work? Impact findings from the National Job Corps Study. *American Economic Review*, 98(5):855–902.

Sethi, S. P. (1977). Optimal advertising for the Nerlove-Arrow model under a budget constraint. *Journal of the Operational Research Society*, 3(28):683–693.

Simon, H. (1982). ADPULS: An advertising model with wearout and pulsation. *Journal of Marketing Research*, 19(3):352–363.

Sommer, A. and Zeger, S. L. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine*, 10(1):45–52.

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, 70(1):331–341.

Waisman, C., , Nair, H. S., and Carrion, C. (2022a). Online causal inference for advertising in real-time bidding auctions. *arXiv preprint arXiv:1908.08600*.

Waisman, C., Sahni, N. S., Nair, H. S., and Lin, X. (2022b). Parallel experimentation on advertising platforms. *arXiv preprint arXiv:1903.11198*.

Zhao, K., Hua, J., Yan, L., Zhang, Q., Xu, H., and Yang, C. (2019). A unified framework for marketing budget allocation. In Teredesai, A. and Kumar, V., editors, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining(KDD '19)*, pages 1820–1830, New York, USA. ACM.

A Incorporating observable characteristics

We now demonstrate how to incorporate observable characteristics, captured in a vector X, into the model above. Three changes have to be made. First, equation (3) is replaced with:

$$D = \mathbb{1}\{\nu(Z, X) \geq U\}. \tag{A.1}$$

Importantly, notice that the error in the selection, U, remains additively separable.

Second, Assumption 1 is replaced with:

Assumption A.1.
(i) $U \perp\!\!\!\perp Z|X$, *where* $\perp\!\!\!\perp$ *denotes conditional statistical independence.*
(ii) $\mathbb{E}[Y_d|Z, X, U] = \mathbb{E}[Y_d|X, U]$ *and* $\mathbb{E}\left[Y_d^2\right] < \infty$ *for* $d \in \{0, 1\}$.
(iii) U *is continuously distributed conditional on* X.

Finally, we replace Assumption 2 with:

Assumption A.2.
(i) $\Pr(Z_c = z|X, \mathcal{C} = c) \in (0, 1)$ *for all* X, c *and all* z.
(ii) $\nu(Z_c = 1|X) \equiv \Pr(D = 1|X, Z_c = 1) \in (0, 1)$ *for all* c *and* X.
(iii) $\nu(Z_c = 1|X) \neq \nu(Z_{c'} = 1|X)$ *for all* $c \neq c'$ *and* X.

In summary, Assumptions A.1 and A.2 simply add conditioning on X to Assumptions 1 and 2. The equivalence between this model and that of Imbens and Angrist (1994) remains. Furthermore, the MTR and MTE functions become $m_d(u, x) \equiv \mathbb{E}[Y_d|U = u, X = x]$, where $d \in \{0, 1\}$, and $\text{MTE}(u, x) \equiv \mathbb{E}[Y_1 - Y_0|U = u, X = x] = m_1(u, x) - m_0(u, x)$, respectively.

B Rewriting firm's decision problem

We now demonstrate how the firm's decision problem given in equation (2) can be rewritten as in equation (7). We begin by deriving the terms in equation (6). First, we have

that:

$$\begin{aligned}
\mathbb{E}\left[Y_1|U \leq \nu\right] &= \int_0^{\nu} \int_{y_1 \in \mathcal{Y}_1} y_1 \frac{f(y_1, u)}{\Pr(U \leq \nu)} du dy_1 \\
&= \int_0^{\nu} \left(\int_{y_1 \in \mathcal{Y}_1} y_1 f(y_1|u) dy_1 \right) \frac{f(u)}{\nu} du \\
&= \int_0^{\nu} \mathbb{E}\left[Y_1|U = u\right] \times \frac{1}{\nu} du \\
&\equiv \int_0^{\nu} m_1(u) \frac{1}{\nu} du, \qquad \text{(B.1)}
\end{aligned}$$

where in the third equality we used that $U \sim U(0,1)$ and defined $m_1(u) \equiv \mathbb{E}\left[Y_1|U = u\right]$. Second,

$$\begin{aligned}
\mathbb{E}\left[Y_0|U > \nu\right] &= \int_{\nu}^1 \int_{y_0 \in \mathcal{Y}_0} y_0 \frac{f(y_0, u)}{\Pr(U > \nu)} du dy_0 \\
&= \int_{\nu}^1 \left(\int_{y_0 \in \mathcal{Y}_0} y_0 f(y_0|u) dy_0 \right) \frac{f(u)}{1-\nu} du \\
&= \int_{\nu}^1 \mathbb{E}\left[Y_0|U = u\right] \times \frac{1}{1-\nu} du \\
&\equiv \int_{\nu}^1 m_0(u) \frac{1}{1-\nu} du, \qquad \text{(B.2)}
\end{aligned}$$

where in the third equality we also used that $U \sim U(0,1)$ and defined $m_0(u) \equiv \mathbb{E}\left[Y_0|U = u\right]$.

Plugging equations (B.1) and (B.2) back into expression (2) yields:

$$\begin{aligned}
&\max_{\nu \in [0,1]} \left(\delta \times \left\{ \int_0^{\nu} m_1(u) du + \int_{\nu}^1 m_0(u) du \right\} - \kappa(\nu) \right) \\
&\max_{\nu \in [0,1]} \left(\delta \times \int_0^1 m_0(u) du + \delta \times \left\{ \int_0^{\nu} \left[m_1(u) - m_0(u) \right] du \right\} - \kappa(\nu) \right) \\
&\max_{\nu \in [0,1]} \left(\delta \times \mathbb{E}\left[Y_0\right] + \delta \times \int_0^{\nu} \text{MTE}(u) du - \kappa(\nu) \right) \\
&\max_{\nu \in [0,1]} \left(\delta \times \int_0^{\nu} \text{MTE}(u) du - \kappa(\nu) \right),
\end{aligned}$$

which establishes expression (7).

C From MTE function to treatment effect parameters

Our main objects of interest are the MTR and MTE functions. This is in part motivated by the the firm's decision problem we presented in Sections 2.1 and 2.3. An additional motivation is that, as shown, for example, in Heckman and Vytlacil (2005), we can use the MTE function to recover other typical treatment effect parameters of interest. This is because we can write most target parameters in the following form:

$$\theta = \mathbb{E}_Z \left[\int_0^1 m_1(u)\omega_1(u,Z)d\mu(u) \right] + \mathbb{E}_Z \left[\int_0^1 m_0(u)\omega_0(u,Z)d\mu(u) \right], \tag{C.1}$$

where $\omega_1(\cdot)$ and $\omega_0(\cdot)$ are identified weighing functions and $\mu(\cdot)$ is an appropriate integrating measure. The subscript Z in the expectations above indicates that they are taken with respect to Z. Hence, if we know the MTR functions we can compute many target parameters we might be interested in.

Examples of such target parameters are the ATE:

$$\begin{aligned} \text{ATE} &\equiv \mathbb{E}\left[Y_1 - Y_0\right] \\ &= \int_0^1 m_1(u)du - \int_0^1 m_0(u)du; \end{aligned} \tag{C.2}$$

the ATT:

$$\begin{aligned} \text{ATT} &\equiv \mathbb{E}\left[Y_1 - Y_0 | D = 1\right] \\ &= \mathbb{E}_Z \left[\int_0^1 \frac{\mathbb{1}\{u \leq \nu(Z)\}}{\Pr(D=1)} m_1(u)du \right] - \mathbb{E}_Z \left[\int_0^1 \frac{\mathbb{1}\{u \leq \nu(Z)\}}{\Pr(D=1)} m_0(u)du \right] \\ &= \frac{1}{\Pr(D=1)} \left\{ \mathbb{E}_Z \left[\int_0^{\nu(Z)} m_1(u)du \right] - \mathbb{E}_Z \left[\int_0^{\nu(Z)} m_0(u)du \right] \right\}; \end{aligned} \tag{C.3}$$

the ATU:

$$\begin{aligned} \text{ATU} &\equiv \mathbb{E}\left[Y_1 - Y_0 | D = 0\right] \\ &= \mathbb{E}_Z \left[\int_0^1 \frac{\mathbb{1}\{u > \nu(Z)\}}{\Pr(D=0)} m_1(u)du \right] - \mathbb{E}_Z \left[\int_0^1 \frac{\mathbb{1}\{u > \nu(Z)\}}{\Pr(D=0)} m_0(u)du \right] \\ &= \frac{1}{\Pr(D=0)} \left\{ \mathbb{E}_Z \left[\int_{\nu(Z)}^1 m_1(u)du \right] - \mathbb{E}_Z \left[\int_{\nu(Z)}^1 m_0(u)du \right] \right\}; \end{aligned} \tag{C.4}$$

and the LATE:

$$\begin{aligned}\text{LATE}\left(z, z'\right) &\equiv \mathbb{E}\left[Y_1 - Y_0 \middle| \nu(z) < U \leq \nu(z')\right] \\ &= \int_0^1 \frac{\mathbb{1}\{\nu(z) < u \leq \nu(z')\}}{\nu(z') - \nu(z)} m_1(u)du - \int_0^1 \frac{\mathbb{1}\{\nu(z) < u \leq \nu(z')\}}{\nu(z') - \nu(z)} m_0(u)du \\ &= \frac{1}{\nu(z') - \nu(z)} \left\{ \int_{\nu(z)}^{\nu(z')} m_1(u)du - \int_{\nu(z)}^{\nu(z')} m_0(u)du \right\}.\end{aligned} \tag{C.5}$$

Finally, notice that if we set $\mu(\cdot)$ to be the Dirac measure and $\omega_1(\cdot)$ and $\omega(\cdot)$ to be -1, 0 or 1 as appropriate, we obtain the MTR and MTE functions back.

D MTE function under different assumptions

We now consider two commonly made assumptions on the DGP to assess whether they impose enough structure on the resulting MTE function to imply specific guidance on how to choose the number of cells or values for the propensity score during the experiment. Under each assumption, we present an example such that no specific guidance is obtained.

D.1 Monotonic MTE function

One possible assumption the researcher might be willing to make is that the marginal treatment effect function is monotonic. This is necessarily satisfied when the MTE function is linear, an assumption that is commonly made, as is by construction the approximation to the MTE function that can be obtained from a single-cell design. We now provide an example where monotonicity of the MTE function does not necessarily aid in choosing C or $\nu(Z_c = 1)$.

To this end, Assume that $\text{MTE}(u) = \frac{2.7}{2+2^{10u}}$. This function is not only monotonic, but it also is strictly convex and nonnegative. Hence, it is a fairly "well-behaved" function.

In particular, the monotonicity might suggest that a linear approximation might be satisfactory. However, D.1 shows that not to be the case because a linear approximation might cover a wide range of negative values, which is a marked difference from the true MTE function. Furthermore, a linear approximation is highly susceptible to values of the propensity score. As Figure D.1 shows, depending on these values the resulting linear curve can have very distinct slopes.

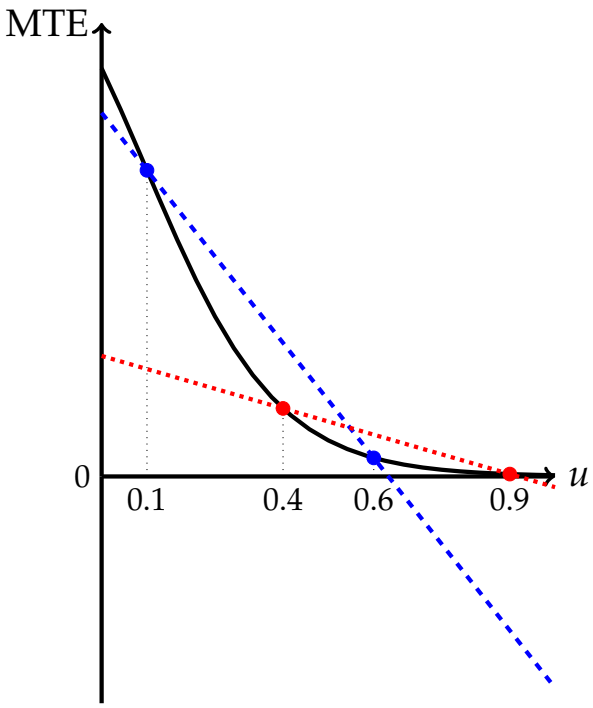


Figure D.1: Different linear approximations for an MTE curve

Given this MTE function, it might seem like having one additional cell might suffice to obtain a satisfactory approximation. Nevertheless, this need not be the case. Figure D.2 depicts two quadratic approximations to the MTE function. As we can see, neither approximation is particularly good, and one fails to capture the monotonicity of the MTE function. This is a result of the range of values taken by the propensity score, which, in both cases, is limited.

On the other hand, as perhaps expected, when the propensity score covers a wider range of values, the quality of the approximation can be high, as we show in Figure D.3. Note, however, that this approximation, unlike the true MTE function, displays negative values.

A priori, it is unclear, however, the curvature of the function, and therefore what the exact range of propensity score values should be.

D.2 Monotone treatment response

A different and arguably stronger assumption is that of monotone treatment response (Manski, 2004). It implies that the treatment effects themselves always have the same sign, which implies that so does the MTE function. However, the previous example suggests that this will not suffice to make this MTE function sufficiently "well-behaved" for us to

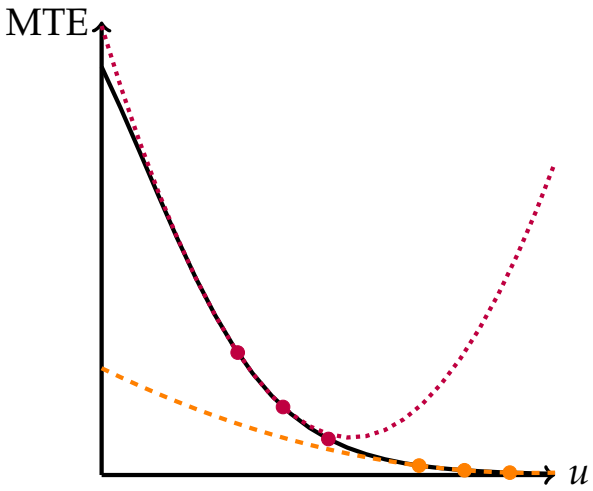


Figure D.2: Different quadratic approximations for an MTE curve

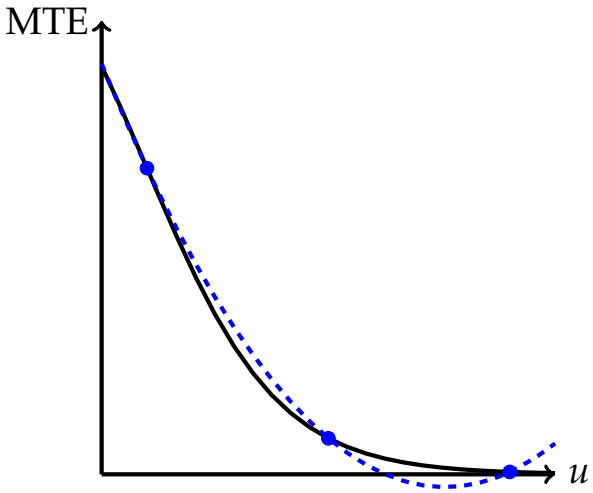


Figure D.3: Different quadratic approximations for an MTE curve

obtain precise guidance for the specific design of the experiment. Indeed, the example below confirms this to be the case.

Assume that $Y_0|U = u \sim N\left(0.9u - 3.8u^2 + 3.3u^3, 1\right)$ and $Y_1|Y_0 \sim TN\left(0, 1, Y_0, +\infty\right)$, so that Y_1 follows a standard normal distribution truncated from below at Y_0. The resulting MTE function is shown in Figure D.4. Unsurprisingly, monotone treatment response ensures that the function always has the same sign, but is not even enough to impose, for instance, monotonicty. Consequently, on its own this assumption is not helpful in informing specifically how the experiment should be designed and implemented.

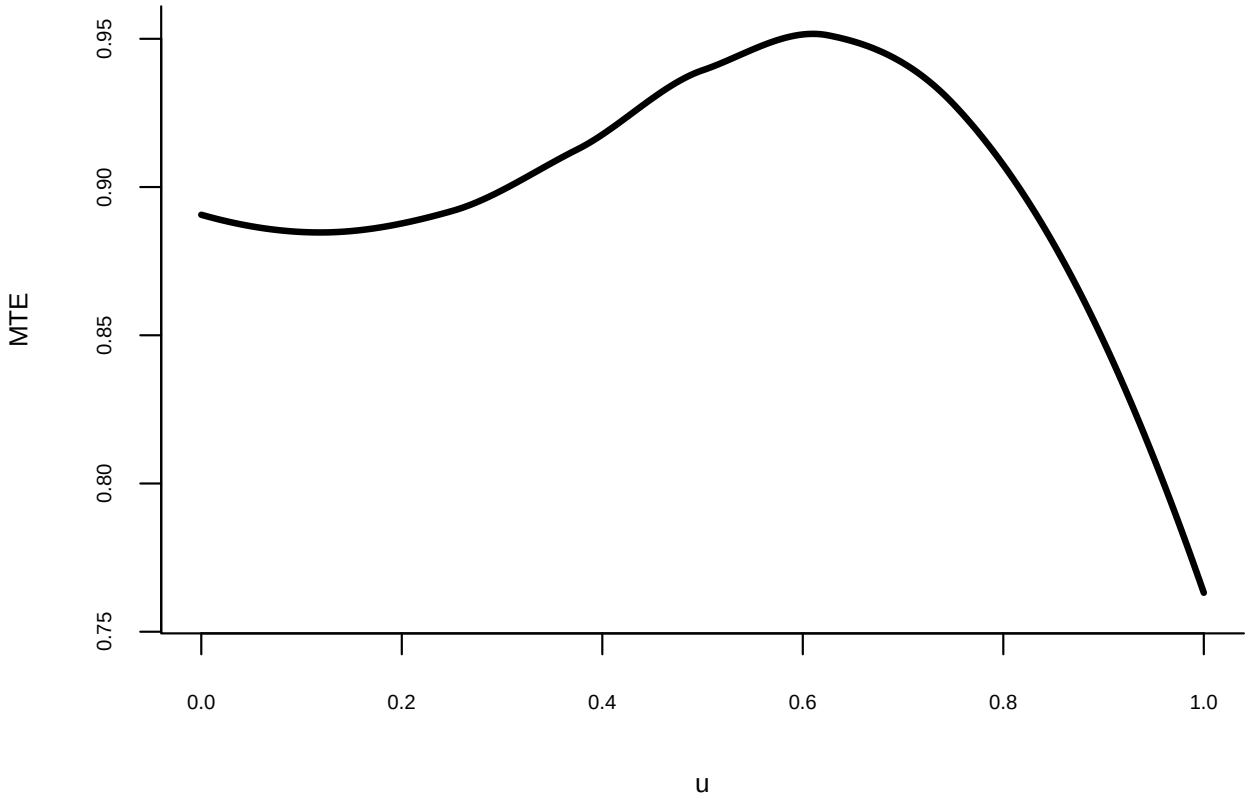


Figure D.4: MTE function under monotone treatment response