Noncompliance and instrumental variables for 2^K factorial experiments*

Matthew Blackwell† Nicole E. Pashley‡

July 7, 2020

Abstract

Factorial experiments are widely used to assess the marginal, joint, and interactive effects of multiple concurrent factors. While a robust literature covers the design and analysis of these experiments, there is less work on how to handle treatment noncompliance in this setting. To fill this gap, we introduce a new methodology that uses the potential outcomes framework for analyzing 2^K factorial experiments with noncompliance on any number of factors. This framework builds on and extends the literature on both instrumental variables and factorial experiments in several ways. First, we define novel, complier-specific quantities of interest for this setting and show how to generalize key instrumental variables assumptions. Second, we show how partial compliance across factors gives researchers a choice over different types of compliers to target in estimation. Third, we show how to conduct inference for these new estimands from both the finite-population and superpopulation asymptotic perspectives. Finally, we illustrate these techniques by applying them to two field experiments—one on the effects of cognitive behavioral therapy on crime and the other on the effectiveness of different forms of get-out-the-vote canvassing. New easy-to-use, open-source software implements the methodology.

*Thanks to Thad Dunning, Kosuke Imai, Luke Keele, and Joel Middleton for comments. Any errors remain our own. Software to implement the methods of this paper will be available as the `factiv` package for R.

†Department of Government and Institute for Quantitative Social Science, Harvard University. web: http://www.mattblackwell.org email: mblackwell@gov.harvard.edu

‡Department of Statistics, Harvard University. Nicole Pashley was funded by the National Science Foundation Graduate Research Fellowship under Grant No. DGE1745303 while working on this paper. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. email: npashley@g.harvard.edu

1 Introduction

Researchers across the social and biomedical sciences often rely on factorial experiments to assess the effects of a number of different factors simultaneously. A 2^K factorial experiment randomly assigns units to 2^K possible treatment combinations of K binary factors. These designs have tremendous advantages. First, they allow for the estimation of both the K main effects of each factor and any interactions between the factors. Second, they allow researchers to block certain causal pathways by design and thus provide richer answers to scientific questions. Third, they are also more efficient than experiments that manipulate one factor at a time (Chapter 5, Montgomery, 2013). Such designs have a long history in statistics (Yates, 1937; Fisher, 1935) and are often of great scientific and policy relevance. However, only relatively recent literature has begun to address the design and analysis of these experiments under the so-called potential outcomes framework (Hainmueller, Hopkins, and Yamamoto, 2014; Dasgupta, Pillai, and Rubin, 2015).

A practical consideration with factorial experiments that has received relatively little attention is noncompliance with treatment assignment. This can occur when experimental units self-select into treatment in defiance of their randomized treatment assignment. When this occurs, researchers often switch focus to the intent-to-treat (ITT) effect of treatment assignment. From a scientific and policy viewpoint, however, the primary interest usually remains on the effect of the treatment actually received. In the context of single-factor experiments, researchers can address noncompliance through the use of instrumental variables (IV), which are less frequently used in factorial designs (for exceptions, see Cheng and Small, 2006; Blackwell, 2017). Indeed, the properties of IV estimators in single-factor experiments are well-studied (Angrist, Imbens, and Rubin, 1996), but the relevant estimands and estimators have yet to be developed in the factorial case.

We address this problem by introducing a framework for analyzing 2^K factorial experiments with noncompliance on any number of factors. Our contributions are several. First, we generalize the standard instrumental variables framework, including the assumptions and estimands, from the single-factor case to the factorial setting. In particular, we show how to extend key assumptions like the exclusion restriction and monotonicity and how to define novel factorial IV estimands as

ratios of intent-to-treat effects of treatment assignment on the outcome and treatment uptake. Unlike the single-factor case, there are several IV estimands in the factorial setting: main effects, two-way interactions, three-way interactions, and so on.

Second, we demonstrate how the multidimensional nature of treatment in factorial experiments complicates the interpretation of these IV estimands. A respondent might comply with their assigned value on one factor but not on another, and the number of possible compliance types grow quickly with K. To address these issues, we invoke an assumption novel to the factorial setting—the "treatment exclusion restriction"—in which the treatment receipt of a factor only depends on the treatment assignment for that factor (Blackwell, 2017). Under this and the other IV assumptions, we show that IV estimands have an interpretation as the average factorial effects of treatment received for the *marginalized compliers*—that is, those respondents who comply with treatment assignment on the active factor(s) for the main effect or interaction of interest, marginalizing over the compliance status of the other factors. One disadvantage of these effects is that the compliance group changes across the different factorial effects, and so we also introduce effects for those that would comply with assignments on all factors, whom we call *perfect compliers*.

Third, to conduct estimation and inference for these IV quantities, we explore two different frameworks: finite-population (also known as finite-sample) inference and superpopulation inference. Following Dasgupta, Pillai, and Rubin (2015) and Kang, Peck, and Keele (2018), our finite-population approach treats the potential outcomes and causal effects of interest as fixed quantities about a finite population. Variation and uncertainty in this approach come only from the random assignment of treatment. We utilize recent work on finite-population asymptotics to derive a central limit result for our intent-to-treat effects and use this to develop a procedure for generating confidence intervals (Li and Ding, 2017; Kang, Peck, and Keele, 2018). Superpopulation approaches, on the other hand, assume that the potential outcomes are random draws from an infinite superpopulation, simplifying inference considerably at the price of plausibility.

We then apply our methodology to two empirical examples. The first uses data from Blattman, Jamison, and Sheridan (2017) in which young men with a history of criminal or violent behaviors were

randomly assigned to receive some combination of cash transfers and cognitive behavioral therapy, or neither. The goal was to reduce repeats of criminal or violent behavior and the study looked at several outcomes, including economic measures, antisocial behavior, and social network measures, in the short and long term. The second example uses get-out-the-vote data from New Haven, CT (Gerber and Green, 2000) and assess the effects of three treatments: door-to-door in person canvasing, phone calls, and mailers. The outcome of interest in this second study was voter turnout.

The paper proceeds as follows. In Section 2, we introduce the setting of factorial experiments with noncompliance and outline our key assumptions, quantities of interest, and estimators. Next, in Section 3, we develop the asymptotic properties of the estimators for the instrumental variable estimands under a finite-population framework and discuss how to apply a technique from the literature on ratio estimators to construct confidence intervals. We apply these techniques to the two application in Section 4 and end with concluding thoughts in Section 5. In the Supplemental Materials, we also develop a procedure for Bayesian inference in this context and present simulation evidence for the validity of our confidence interval procedure.

2 Framework: Notation, Assumptions, and Quantities of Interest

We consider an experiment with K binary factors with levels $\{-1, +1\}$, so that $\mathcal{Z} = \{-1, +1\}^K$ is the set of all possible treatment combinations. For instance, -1 may be the control level and $+1$ the treatment level of a given factor. Thus, there are $L = 2^K$ possible treatment assignments, which we order $\{1, \ldots, L\}$ with $\boldsymbol{z}_\ell = \{z_{\ell 1}, \ldots, z_{\ell K}\}$ being the levels of each factor for treatment combination ℓ. We define the set of possible treatment uptake vectors $\boldsymbol{d}_\ell$, which have the same values and are ordered in the manner as $\boldsymbol{z}_\ell$ (i.e., $\boldsymbol{d}_\ell = \boldsymbol{z}_\ell$). Each unit may have a different potential outcome for each treatment assignment and uptake combination, $Y_i(\boldsymbol{z}, \boldsymbol{d})$. This is the value of the outcome that unit i would have if they been assigned $\boldsymbol{z}$ and taken $\boldsymbol{d}$.

Experiments with noncompliance face the problem that treatment uptake may differ from treatment assignment, and so treatment uptake will have potential outcomes as well. Let $\boldsymbol{D}_i(\boldsymbol{z}) \in \mathcal{Z}$ be

the vector of treatment uptake on each factor if unit i was assigned to treatment combination $\boldsymbol{z}$. If $\boldsymbol{D}_i(\boldsymbol{z}) = \boldsymbol{z}$ for all i and $\boldsymbol{z}$, then there is full compliance and inference can be conducted as usual. We focus on the case where $\boldsymbol{D}_i(\boldsymbol{z}) \neq \boldsymbol{z}$ for some i and $\boldsymbol{z} \in \mathcal{Z}$ and define the vector of potential outcomes indicators for each treatment uptake combination as

$$\boldsymbol{R}_i(\boldsymbol{z}) = \{\mathbb{I}(\boldsymbol{D}_i(\boldsymbol{z}) = \boldsymbol{d}_1), \ldots, \mathbb{I}(\boldsymbol{D}_i(\boldsymbol{z}) = \boldsymbol{d}_L)\}^\top.$$

Let $\boldsymbol{R}_i(\bullet)$ be the $2^K \times 2^K$ matrix with ℓth row $\boldsymbol{R}_i(\boldsymbol{z}_\ell)^\top$. For the intent-to-treat analyses, we will often work with the potential outcomes just setting the treatment assignment, $Y_i(\boldsymbol{z}) \equiv Y_i(\boldsymbol{z}, \boldsymbol{D}_i(\boldsymbol{z}))$, and we collect the L potential outcomes for unit i into the vector $\boldsymbol{Y}_i(\bullet) = \{Y_i(\boldsymbol{z}_1), \ldots, Y_i(\boldsymbol{z}_L)\}^\top$.

Let $W_{i\ell} = 1$ if $\boldsymbol{Z}_i = \boldsymbol{z}_\ell$ and 0 otherwise and $\boldsymbol{W}_i = \{W_{i1}, \ldots, W_{iL}\}$ be the vector of indicators for all treatment combinations. We assume a completely randomized design. In particular, let $\boldsymbol{W} = (\boldsymbol{W}_1, \ldots, \boldsymbol{W}_N)$ be the length LN vector of assignment indicators for all units and $\mathcal{F} = \{\boldsymbol{Y}_i(\bullet), \boldsymbol{R}_i(\bullet), i = 1, \ldots, N\}$. Consider a completely randomized design with $N_\ell = \sum_{i=1}^N W_{i\ell}$ units assigned to treatment $\boldsymbol{z}_\ell$, with $\sum_{\ell=1}^L N_\ell = N$, defined formally below.

Assumption 1 (General Completely Randomized Design)**.**

$$\mathbb{P}(\boldsymbol{W} \mid \mathcal{F}) = \begin{cases} \left(\frac{N!}{\prod_{\ell=1}^L N_\ell!}\right)^{-1} & \text{if } \sum_{i=1}^N W_{i\ell} = N_\ell \text{ for all } \ell = 1, \ldots, L \\ 0 & \text{otherwise} \end{cases}$$

Under this design, we have $\mathbb{E}\{W_{i\ell} \mid \mathcal{F}\} = (N_\ell)^{-1}$ for all $\boldsymbol{z}_\ell$, where the expectation here is over the randomization distribution. We connect the potential outcomes to the observed outcomes through a consistency assumption, $Y_i^{\text{obs}} = \sum_{\ell=1}^L W_{i\ell} Y_i(\boldsymbol{z}_\ell)$, $\boldsymbol{D}_i^{\text{obs}} = \sum_{\ell=1}^L W_{i\ell} \boldsymbol{D}_i(\boldsymbol{z}_\ell)$, and $\boldsymbol{R}_i^{\text{obs}} = \sum_{\ell=1}^L W_{i\ell} \boldsymbol{R}_i(\boldsymbol{z}_\ell)$. Note that this means that some of the $Y_i(\boldsymbol{z}, \boldsymbol{d})$ are a priori counterfactuals, because they can never be observed for a given individual.

When there is noncompliance with treatment assignment, randomization is not sufficient to identify the causal effect of treatment uptake. Several ways of addressing noncompliance have been proposed in the literature, all of which make additional assumptions beyond randomization. We follow one strain of the literature, which started with Angrist, Imbens, and Rubin (1996), and focus on two

types of assumptions: monotonicity and exclusion restrictions. We generalize these standard instrumental variables assumptions to the factorial context.

Monotonicity is a restriction on the direction of the effect of treatment assignment on treatment uptake. Let $\boldsymbol{z}^+$ be a K-vector of all $+1$ and $\boldsymbol{z}^-$ be a K-vector of all -1, with z_k^+ and z_k^- being representative kth entries. Furthermore, let $\boldsymbol{z}_{-k}$ be the vector $\boldsymbol{z}$ with the kth entry omitted and abuse notation to let $\boldsymbol{z} = (\boldsymbol{z}_{-k}, z_k)$. Let $D_{ik}(\boldsymbol{z})$ be the treatment uptake of unit i for factor k when assigned to $\boldsymbol{z}$.

Assumption 2 (Monotonicity). *$D_{ik}(\boldsymbol{z}_{-k}, z_k^+) \geq D_{ik}(\boldsymbol{z}_{-k}, z_k^-)$ for all $k \in \{1, \ldots, K\}$ and $\boldsymbol{z}_{-k}$.*

This assumption states that there are no defiers: individuals who would have treatment uptake of -1 for factor k if assigned to $+1$ of factor k and treatment uptake of $+1$ for factor k if assigned to -1 of factor k.

A standard approach in the instrumental variables literature is to assume that treatment assignment has no direct effect on the outcome, except through treatment receipt (Robins, 1989; Angrist, Imbens, and Rubin, 1996). This assumption is typically called the *exclusion restriction*, and it has a natural generalization in the factorial setting. To distinguish it from a separate exclusion restriction we define below, we call this the outcome exclusion restriction.

Assumption 3 (Outcome exclusion restriction). *For all $\boldsymbol{z}, \boldsymbol{z}' \in \mathcal{Z}$, $Y_i(\boldsymbol{z}, \boldsymbol{d}) = Y_i(\boldsymbol{z}', \boldsymbol{d})$.*

This assumption is substantive and cannot be met simply by experimental design. Finally, the factorial setting requires a novel assumption for identification of certain effects. First proposed in Blackwell (2017) for the 2×2 factorial design, the *treatment exclusion restriction* states that treatment uptake on factor k only depends on the treatment assignment for factor k, not other factors.

Assumption 4 (Treatment exclusion restriction). *For all $\boldsymbol{z} \in \mathcal{Z}$, $D_{ik}(\boldsymbol{z}) = D_{ik}(z_k)$ where z_k is the kth entry of $\boldsymbol{z}$.*

This assumption rules out interactive effects of treatment assignment on treatment uptake in the sense that it assumes no units that, say, comply on factor 1 when $z_2 = +1$ but not when $z_2 = -1$. This effectively assumes that noncompliance is factor-specific. The treatment exclusion restriction

is also a substantive assumption that restricts between-factor interactions in the effects of treatment assignment on treatment uptake. In the context of the cash and cognitive behavioral therapy experiment, for instance, this would be violated if being assigned to receive a cash transfer caused some respondents to participate in therapy when they otherwise would not. While treatment exclusion is not directly testable, some of its implications are observable. For instance, it would rule out any effect of Z_{i1} on D_{i2} or any interaction between Z_{i1} and Z_{i2} on D_{i2}. Thus, one falsification test for this assumption is to check these various effects in the assignment-uptake relationship, which we do in our empirical example below. We discuss some implications for weakening this assumption in the following section and outline further weaker assumptions of interest in the Discussion.

2.1 Estimands

We begin by describing a set of standard linear factorial effects in the finite-population framework and then extend them to the superpopulation viewpoint below. These effects reflect differences between one half of the potential outcomes for a particular outcome versus the other. We can define these effects through the use of an L-dimensional vector $\boldsymbol{g}$ that has one half of its entries at 1 and the other half at -1 as in Dasgupta, Pillai, and Rubin (2015). There are $L-1$ such vectors and the same number of factorial effects. We can order these vectors such that the first K represent the main effects of the K factors, so that $\boldsymbol{g}_1$ corresponds to the main effect of factor 1, $\boldsymbol{g}_2$ corresponds to the main effect of factor 2, and so on. The next $\binom{K}{2}$ vectors will correspond to all two-factor interactions, and the following $\binom{K}{3}$ vectors will correspond to all three-factor interactions, and so on. This continues until $\boldsymbol{g}_{L-1}$ which corresponds to the K-way interaction between all factors. For main effects , $\boldsymbol{g}_j$ is a vector giving the level of factor j for each of the L treatment combinations. Interaction vectors are then created as element-wise products of these main effect vectors. Note that these vectors are mutually orthogonal.

With these vectors, we can define individual-level intent-to-treat factorial effects for the outcome as

$$\tau_{ij} = 2^{-(K-1)} \boldsymbol{g}_j^\top \boldsymbol{Y}_i(\bullet) = 2^{-(K-1)} \sum_{\ell=1}^{L} g_{j\ell} Y_i(\boldsymbol{z}_\ell)$$

for $i = 1, \ldots, N$ and $j = 1, \ldots, L-1$, where $g_{j\ell}$ is the ℓth entry of the $\boldsymbol{g}_j$ vector. Here, τ_{ij} is the jth factorial effect of treatment assignment on the outcome for individual i. For main effects, this is the effect of assignment to factor j, averaging over all possible assignments to other factors. For example, when $K = 2$, we have $\boldsymbol{g}_1 = (+1, -1, +1, -1)$, so that

$$\frac{1}{2}\boldsymbol{g}_1^\top \boldsymbol{Y}_i(\bullet) = \frac{1}{2}\underbrace{\{Y_i(+1,+1) - Y_i(-1,+1)\}}_{\text{effect of factor 1 when factor 2 is } +1} + \frac{1}{2}\underbrace{\{Y_i(+1,-1) - Y_i(-1,-1)\}}_{\text{effect of factor 1 when factor 2 is } -1}.$$

Writing the finite-population averages of the potential outcomes as $\overline{\boldsymbol{Y}}(\bullet) = N^{-1}\sum_{i=1}^{N}\boldsymbol{Y}_i(\bullet)$, the finite-population intent-to-treat average factorial effects are

$$\overline{\tau}_j = \frac{1}{N}\sum_{i=1}^{N}\tau_{ij} = 2^{-(K-1)}\boldsymbol{g}_j^\top \overline{\boldsymbol{Y}}(\bullet).$$

These effects marginalize over the distribution of possible assignments, weighting each possible assignment equally. While this is standard in the factorial design literature, a recent strand of work examining a specific type of factorial designs—conjoint experiments—has worked with a more general estimand that allows for researcher-specified distributions for the assignments (Hainmueller, Hopkins, and Yamamoto, 2014; de la Cuesta, Egami, and Imai, 2020). In the Supplemental Material, we discuss the straightforward extension of the present approach to those more general estimands. Finally, Egami and Imai (2019) proposed alternative quantities of interest for interactions in factorial experiments, but those average marginal interaction effects are more appropriate with factors with more than two levels.

These intent-to-treat factorial effects will not equal the true effect of treatment uptake when some units do not comply with the factors in the factorial effect. To correct this problem, the instrumental variables literature will often define the estimand of interest as the ratio of the intent-to-treat effects on the outcome and on treatment uptake (Wald, 1940). In the factorial setting, however, the definition of treatment uptake depends on the factorial effect of interest. For example, for the main effect of the first factor, we want the ITT for treatment uptake on the first factor, whereas for the interaction between the first and second factor, we want the ITT on the *interaction* between D_{i1} and D_{i2}. More generally, let $\mathcal{K}(j)$ be the set of indices of the "active" factors for factorial effect j. That is, $\mathcal{K}(j)$ are

the set of factors for which $\boldsymbol{g}_j$ is estimating the main or interaction effects. For $j = 1, \ldots, K$, this is just $\mathcal{K}(j) = \{j\}$, but for interactions, we have for example, $\mathcal{K}(K+1) = \{1, 2\}$, and so on. Define the following potential outcome of treatment uptake interaction corresponding to the jth factorial effect:

$$\widetilde{D}_{ij}(\boldsymbol{z}) = \prod_{k \in \mathcal{K}(j)} D_{ik}(\boldsymbol{z}).$$

Again, for $j \leq K$, we have $\widetilde{D}_{ij}(\boldsymbol{z}) = D_{ij}(\boldsymbol{z})$. We can collect these into a vector of potential outcomes for each treatment assignment vector $\widetilde{\boldsymbol{D}}_{ij}(\bullet) = \{\widetilde{D}_{ij}(\boldsymbol{z}_1), \ldots, \widetilde{D}_{ij}(\boldsymbol{z}_L)\}^\top$. Further, we can write these as a function of the $\boldsymbol{g}$ vectors to obtain $\widetilde{\boldsymbol{D}}_{ij}(\boldsymbol{z}) = \boldsymbol{g}_j^\top \boldsymbol{R}_i(\boldsymbol{z})$ and so $\widetilde{\boldsymbol{D}}_{ij}(\bullet) = \boldsymbol{R}_i(\bullet)\boldsymbol{g}_j$. The individual-level ITT of treatment assignment on treatment uptake for the jth factorial effect is thus

$$\delta_{ij} = 2^{-K} \boldsymbol{g}_j^\top \widetilde{\boldsymbol{D}}_{ij}(\bullet) = 2^{-K} \boldsymbol{g}_j^\top \boldsymbol{R}_i(\bullet)\boldsymbol{g}_j,$$

with $\overline{\delta}_j = N^{-1} \sum_{i=1}^N \delta_{ij}$. For example, in the two-factor case, we have

$$\begin{aligned}\delta_{i3} =& \frac{1}{4}\left\{D_{i1}(+1,+1)D_{i2}(+1,+1) - D_{i1}(-1,+1)D_{i2}(-1,+1)\right\} \\ &- \frac{1}{4}\left\{D_{i1}(+1,-1)D_{i2}(+1,-1) - D_{i1}(-1,-1)D_{i2}(-1,-1)\right\},\end{aligned}$$

so that δ_{i3} is the (scaled) interactive effect of treatment assignment on the multiplicative interaction between the two treatment uptakes. We can also write this estimand as a linear function of the potential outcomes for each assignment,

$$\delta_{ij} = \sum_{\ell=1}^{L} 2^{-K} g_{j\ell} \boldsymbol{g}_j^\top \boldsymbol{R}_i(\boldsymbol{z}_\ell),$$

where the equality comes from $\widetilde{\boldsymbol{D}}_{ij}(\boldsymbol{z}) = \boldsymbol{g}_j^\top \boldsymbol{R}_i(\boldsymbol{z})$ and $g_{j\ell}$ is the ℓth entry of $\boldsymbol{g}_j$.

For the main effect, $j = 1, \ldots, K$, and under monotonicity (but not treatment exclusion), we can show these effects represent the average proportion of compliance to factor j, marginalizing uniformly over the treatment combinations of the other factors. For higher-order effects, this quantity is a measure of the effect of the interaction in treatment assignment on the interaction of treatment uptake among the active factors. In the next section, we will see that all of these effects have a more intuitive interpretation when we invoke treatment exclusion.

We can now define the jth IV factorial effect as

$$\overline{\phi}_j = \frac{\overline{\tau}_j}{\overline{\delta}_j}.$$

We assume that $\overline{\delta}_j > 0$, which under treatment exclusion is the same as saying there are *some* compliers for the factors involved in the jth effect. Without further assumptions, $\overline{\phi}_j$ is just the ratio of two intent-to-treat factorial effects. We are able to gain an even more substantive interpretation under various exclusion restrictions on the outcome and the treatment uptake, as described in the next section.

2.2 IV estimands under exclusion restrictions

Under the IV assumptions, the various effects defined above have specific interpretations in terms of principal strata, otherwise known as compliance types. Under treatment exclusion and monotonicity, each unit can be categorized into one of 3^K types based on how treatment uptake depends on treatment assignment. Note that without the treatment exclusion restriction we would have many more compliance types, as a unit's compliance to a given factor could depend upon the 2^{K-1} possible assignments to the other factors. Thus the treatment exclusion assumption essentially makes solutions based on compliance strata more tractable. Let $\boldsymbol{T}_i \in \mathcal{T}_K = \{c, a, n\}^K$ be the K-length vector of compliance type for unit i on all K factors. Here, the compliance types of each factor are complier (c), always-taker (a), and never-taker (n), defined as follows:

$$T_{ik} = \begin{cases} c & \text{if } D_{ik}(+1) = +1, D_{ik}(-1) = -1 \\ a & \text{if } D_{ik}(+1) = +1, D_{ik}(-1) = +1 \\ n & \text{if } D_{ik}(+1) = -1, D_{ik}(-1) = -1. \end{cases}$$

Our estimands relate to these quantities in two key ways. First, under treatment exclusion and monotonicity, for any factorial effect, we have $\widetilde{\boldsymbol{D}}_{ij}(\bullet) = \boldsymbol{g}_j$ when $T_{ik} = c$ for all $k \in \mathcal{K}(j)$ and otherwise $\widetilde{\boldsymbol{D}}_{ij}(\bullet)$ is a vector that is orthogonal to $\boldsymbol{g}_j$. We define $C_{ij} = \prod_{k \in \mathcal{K}(j)} \mathbb{I}(T_{ik} = c)$ be an indicator for being a complier on all the active factors. Then for all j, we have $\delta_{ij} = C_{ij}$ and $\overline{\delta}_j = N^{-1} \sum_{i=1}^N C_{ij}$. In other words, the ITTs for treatment uptake measure compliance with the active factors for a particular factorial effect.

Second, under the treatment and outcome exclusion restrictions, the jth outcome ITT, τ_{ij}, is 0 for all units who do not comply on all the active factors in effect j, allowing us to relate these effects to the conditional effect among compliers. Let $N_j^c = \sum_{i=1}^N C_{ij}$. Noting that $\overline{\delta}_j = N_j^c/N$, we have the following:

$$\overline{\tau}_j = \frac{1}{N}\sum_{i=1}^{N} C_{ij}\tau_{ij} = \frac{\sum_{i=1}^N C_{ij}\tau_{ij}}{N_j^c} \times \overline{\delta}_j.$$

Combining these two facts, the ratio of the ITT effects under the IV assumptions (Assumptions 2, 3 and 4) is

$$\overline{\phi}_j = \frac{1}{N_j^c}\sum_{i=1}^{N} C_{ij}\tau_{ij},$$

which we refer to as the jth marginalized-complier average factorial effect (MCAFE). Because these effects condition on compliance for the active factors, we can interpret this as the average of the jth factorial effect of treatment uptake of factors in $\mathcal{K}(j)$ on the outcome among those units who comply with those active factors, marginalizing over the treatment assignments on other factors. For a main effect, for instance, we have

$$\overline{\phi}_j = \frac{1}{2^{K-1}} \sum_{\boldsymbol{z}_{-j} \in \mathcal{Z}_{-j}} \left(\frac{1}{N_j^c} \sum_{i=1}^{N} C_{ij} \left\{ Y_i(d_j = +1, \boldsymbol{z}_{-j}) - Y_i(d_j = -1, \boldsymbol{z}_{-j}) \right\} \right).$$

Here we slightly abuse notation to emphasize that it is truly treatment uptake, and not just assignment for factor j. This interpretation, while straightforward to derive, is slightly odd since it combines the effects of treatment uptake for some factors and treatment assignment for others.

How can we interpret the MCAFEs in terms of the factorial effects of treatment uptake rather than a mix of treatment uptake and assignment? We can invoke the exclusion restrictions to write the main effect MCAFEs, for instance, as

$$\begin{aligned}
\overline{\phi}_j &= \frac{1}{2^{K-1}} \sum_{\boldsymbol{z}_{-j}} \left(\frac{1}{N_j^c} \sum_{i=1}^{N} C_{ij} \left\{ Y_i(d_j = +1, \boldsymbol{D}_{i,-j}(\boldsymbol{z}_{-j})) - Y_i(d_j = -1, \boldsymbol{D}_{i,-j}(\boldsymbol{z}_{-j})) \right\} \right), \\
&= \sum_{\boldsymbol{d}_{-j}} \left(\frac{1}{N_j^c} \sum_{i=1}^{N} \omega_{ij}(\boldsymbol{d}_{-j}) C_{ij} \left\{ Y_i(d_j = +1, \boldsymbol{d}_{-j}) - Y_i(d_j = -1, \boldsymbol{d}_{-j}) \right\} \right),
\end{aligned}$$

where

$$\omega_{ij}(\boldsymbol{d}_{-j}) = \frac{1}{2^{K-1}} \sum_{\boldsymbol{z}_{-j}} \mathbb{I}\left\{ \boldsymbol{D}_{i,-j}(\boldsymbol{z}_{-j}) = \boldsymbol{d}_{-j} \right\}, \qquad \sum_{\boldsymbol{d}_{-j}} \omega_{ij}(\boldsymbol{d}_{-j}) = 1.$$

We again commit slight abuse of notation to convey the meaning in terms of treatment uptake rather than assignment. Thus, we can see that the MCAFE for the main effect of factor j is an average of complier factorial effects for treatment uptake with each individual having different weights for marginalizing over the uptake profiles. These weights depend on the unit's compliance type on the other factors. Interpretations of the higher-order MCAFEs are similar, albeit more complicated.

Of course, treatment exclusion is a strong assumption that may not hold in practice, so it is helpful to understand how we can interpret these IV estimands under weaker assumptions. Let $T_{ik}(\boldsymbol{z}_{-k}) \in \{c, a, n\}$ be the compliance status for unit i on factor k when the other factors are set to $\boldsymbol{z}_{-k}$ and $C_{ik}(\boldsymbol{z}_{-k})$ be an indicator for if i is a complier for k in that case. Assuming monotonicity (Assumption 2), we can show that the individual ITTs for treatment uptake on the main effects can be written as

$$\delta_{ij} = \frac{1}{2^{K-1}} \sum_{\boldsymbol{z}_{-j} \in \mathcal{Z}_{-j}} C_{ij}(\boldsymbol{z}_{-j})$$

which is the marginalized compliance rate for unit i on factor j, where the marginalization is over the assignments to other factors. If we further assume the outcome exclusion restriction (Assumption 3), then for the main effects $j \in \{1, \ldots, K\}$, the MCAFE has a similar interpretation as under treatment exclusion as a weighted average of complier average effects of treatment uptake on factor j, marginalizing over the treatment assignments on the other factors,

$$\overline{\phi}_j = \frac{1}{2^{K-1}} \sum_{\boldsymbol{z}_{-j} \in \mathcal{Z}_{-j}} \left(\frac{1}{N_j^c(\boldsymbol{z}_{-j})} \sum_{i=1}^{N} C_{ij}(\boldsymbol{z}_{-j}) \left\{ Y_i(d_j = +1, \boldsymbol{z}_{-j}) - Y_i(d_j = -1, \boldsymbol{z}_{-j}) \right\} \right),$$

where $N_j^c(\boldsymbol{z}_{-j}) = \sum_{i=1}^{N} C_{ij}(\boldsymbol{z}_{-j})$.

We can even obtain an interpretation of main effect MCAFEs in terms of treatment uptake on all factors under a weaker version of treatment exclusion. In the context of interference in randomized experiments, Imai, Jiang, and Malai (2020) proposed the following assumption that limited how one factor may influence uptake on other factors.

Assumption 5 (Weak treatment exclusion)**.** *For any $\boldsymbol{z}_{-j} \in \mathcal{Z}_{-j}$, if $D_{ij}(+1, \boldsymbol{z}_{-j}) = D_{ij}(-1, \boldsymbol{z}_{-j})$, then $\boldsymbol{D}_i(+1, \boldsymbol{z}_{-j}) = \boldsymbol{D}_i(-1, \boldsymbol{z}_{-j})$.*

In words, this assumption says that if unit i is a noncomplier on factor j when the other factors are assigned to $\boldsymbol{z}_{-j}$, then the assignment of factor j should not affect the treatment uptake of the other factors for the assignment $\boldsymbol{z}_{-j}$. Under Assumptions 2, 3, and 5, we can show that

$$\overline{\phi}_j = \sum_{\boldsymbol{z}_{-j} \in \mathcal{Z}_{-j}} \omega(\boldsymbol{z}_{-j}) \left(\frac{1}{N_j^c(\boldsymbol{z}_{-j})} \sum_{i=1}^{N} C_{ij}(\boldsymbol{z}_{-j}) \left\{ Y_i(d_j = +1, \boldsymbol{D}_{i,-j}(+1, \boldsymbol{z}_{-j})) - Y_i(d_j = -1, \boldsymbol{D}_{i,-j}(-1, \boldsymbol{z}_{-j})) \right\} \right),$$

where

$$\omega(\boldsymbol{z}_{-j}) = \frac{N_j^c(\boldsymbol{z}_{-j})}{\sum_{i=1}^{N} \sum_{\boldsymbol{z}_{-j} \in \mathcal{Z}_{-j}} C_{ij}(\boldsymbol{z}_{-j})}, \qquad \sum_{\boldsymbol{z}_{-j} \in \mathcal{Z}_{-j}} \omega(\boldsymbol{z}_{-j}) = 1.$$

Thus, $\overline{\phi}_j$ is a weighted average of different complier-specific effects induced by changing the assignment of factor j from level -1 to $+1$. Each effect in the weighted average is conditional on compliers for a specific assignment profile for the other factors and the weights depend on the compliance rate of factor j for that profile. Unlike with treatment exclusion, assignment on factor j may affect uptake on other factors, so it is possible that $\boldsymbol{D}_{i,-j}(+1, \boldsymbol{z}_{-j}) \neq \boldsymbol{D}_{i,-j}(-1, \boldsymbol{z}_{-j})$ and thus these complier-specific effects are not simply the conditional effects of uptake on factor j as above. In this case, the MCAFEs combines two types of effects of treatment uptake: the "direct" effect of uptake on factor j induced by assignment on j and the "indirect" effect of uptake on other factors also induced by assignment on j.

We note two limitations with this weaker version of treatment exclusion. First, the interpretation of interactions is much more complicated because joint compliance across two factors is not identified under this assumption. It appears that if we wish to cleanly identify the interactive effects of treatment uptake on the outcome, we have to restrict interactive effects of treatment assignment on uptake. Second, and related, there is no way to identify perfect compliers, discussed in the following section, under weak treatment exclusion, meaning we may be limited to providing bounds for the effects discussed below.

2.3 Perfect complier effects

One issue with the marginalized-complier effects is that the conditioning set changes depending on the factorial effect under study. This makes, for instance, the main effect of factor 1 and the interaction effect of factor 1 and 2 difficult to compare. The first MCAFE will only average over compliers for

factor 1, while the latter will focus on compliers for factor 1 and factor 2. One way to resolve this issue to estimate effects for those units that would comply with all factors—whom we call *perfect compliers*. The definition of these effects requires more care than the previous estimands.

We can identify the perfect compliers by applying the K-way interaction to any vector of potential outcomes for specific treatment uptake vectors. Specifically, let $P_i = \prod_{k=1}^{K} \mathbb{I}\left(T_{ik} = c\right)$ be an indicator for being a perfect complier. From the above discussion, we have

$$\delta_{i,L-1} = 2^{-K} \boldsymbol{g}_{L-1}^{T} \boldsymbol{R}_i(\bullet)^{\top} \boldsymbol{g}_{L-1} = C_{i,L-1} = P_i.$$

We can use a similar approach to identify the potential outcomes for the perfect compliers. Let $\boldsymbol{H}_i(\boldsymbol{z}) = Y_i(\boldsymbol{z})\boldsymbol{R}_i(\boldsymbol{z})$ so that

$$\boldsymbol{H}_i(\boldsymbol{z}) = \{Y_i(\boldsymbol{z})\mathbb{I}(\boldsymbol{D}_i(\boldsymbol{z}) = \boldsymbol{d}_1), \ldots, Y_i(\boldsymbol{z})\mathbb{I}(\boldsymbol{D}_i(\boldsymbol{z}) = \boldsymbol{d}_L)\}^{\top},$$

and let $\boldsymbol{H}_i(\bullet)$ be the $L \times L$ matrix with ℓth row $\boldsymbol{H}_i(\boldsymbol{z}_\ell)^{\top}$. Then under Assumptions 1– 4 and a similar argument to above, we can show that

$$\boldsymbol{g}_{L-1} \circ \boldsymbol{H}_i(\bullet)^{\top} \boldsymbol{g}_{L-1} = Y_i(\bullet) P_i.$$

Thus, we can write the jth ITT for unit i, if unit i is a perfect complier, as

$$\tau_{ij,p} = 2^{-(K-1)} \left(\boldsymbol{g}_j \circ \boldsymbol{g}_{L-1}\right)^{\top} \boldsymbol{H}_i(\bullet)^{\top} \boldsymbol{g}_{L-1} = P_i \tau_{ij}$$

As with τ_{ij} and δ_{ij}, we can write this quantity as a linear function of the potential outcomes for each assignment,

$$\tau_{ij,p} = \sum_{\ell=1}^{L} g_{L-1,\ell} \left(\boldsymbol{g}_j \circ \boldsymbol{g}_{L-1}\right)^{\top} \boldsymbol{H}_i(\boldsymbol{z}_\ell),$$

again where $g_{L-1,\ell}$ is the ℓth entry in the $\boldsymbol{g}_{L-1}$ vector. Let $\overline{\boldsymbol{H}}(\bullet) = N^{-1} \sum_{i=1}^{N} \boldsymbol{H}_i(\bullet)$ the be population average of $\boldsymbol{H}_i(\bullet)$. Then we can define the population effects as

$$\begin{aligned}
\overline{\tau}_{j,p} &= 2^{-(K-1)} \left(\boldsymbol{g}_j \circ \boldsymbol{g}_{L-1}\right)^{\top} \overline{\boldsymbol{H}}(\bullet)^{\top} \boldsymbol{g}_{L-1} \\
&= \frac{1}{N} \sum_{i=1}^{N} P_i \tau_{ij} \\
&= \left(\frac{1}{N_p} \sum_{i=1}^{N} P_i \tau_{ij} \right) \frac{N_p}{N},
\end{aligned}$$

where N_p is the number of perfect compliers in the finite population. Noting from our earlier discussion that $\overline{\delta}_{L-1} = N_p/N$, we can define

$$\overline{\gamma}_j = \frac{\overline{\tau}_{j,p}}{\overline{\delta}_{L-1}} = \frac{1}{N_p} \sum_{i=1}^{N} P_i \tau_{ij}$$

The $\overline{\gamma}_j$ represent the jth average factorial effect among the perfect compliers, which we refer to as the jth perfect complier average factorial effect (PCAFE).

2.4 Superpopulation estimands

We now take an alternative point of view—that the sample of units is actually a draw from an infinite superpopulation. Now, the potential outcomes are themselves random variables and not fixed quantities as in the finite-population point of view. Under treatment exclusion in particular, we define the probability of a particular compliance type, $\boldsymbol{t} \in \mathcal{T}_K$ as $\rho_{\boldsymbol{t}} = \mathbb{P}(\boldsymbol{T}_i = \boldsymbol{t})$. We can relate the finite-population quantities $\overline{\delta}_k$ to these values by considering the limit of a series of growing finite populations with units sampled from a larger fixed population. For example, for any main effect, we have

$$\lim_{N\to\infty} \overline{\delta}_k = \sum_{\boldsymbol{t}:t_k=c} \rho_{\boldsymbol{t}} = \mathbb{P}(C_{ik} = 1).$$

Let $\mathbb{E}[\cdot]$ be the expectation operator that averages over both randomization and sampling from the superpopulation. Then, we can define the superpopulation version of the marginalized-complier average factorial effect as

$$\overline{\phi}_j^{\text{sp}} = \mathbb{E}[\tau_{ij} \mid C_{ij} = 1] = \lim_{N\to\infty} \overline{\phi}_j.$$

We can also define a similar superpopulation version of the perfect complier average factorial effect as

$$\overline{\gamma}_j^{\text{sp}} = \mathbb{E}[\tau_{ij} \mid P_i = 1] = \lim_{N\to\infty} \overline{\gamma}_j.$$

Finally, we can define $\overline{\tau}_j^{\text{sp}}$, $\overline{\tau}_{j,p}^{\text{sp}}$, and $\overline{\delta}_j^{\text{sp}}$ in a similar manner.

2.5 Estimators

We can define the following natural in-sample estimators for the population (of units in the study) or superpopulation potential outcomes:

$$\overline{Y}^{\text{obs}}(\boldsymbol{z}_\ell) = \frac{1}{N_\ell}\sum_{i=1}^{N} W_{i\ell} Y_i^{\text{obs}}, \qquad \overline{\boldsymbol{R}}^{\text{obs}}(\boldsymbol{z}_\ell) = \frac{1}{N_\ell}\sum_{i=1}^{N} W_{i\ell} \boldsymbol{R}_i^{\text{obs}},$$

$$\overline{\boldsymbol{H}}^{\text{obs}}(\boldsymbol{z}_\ell) = \frac{1}{N_\ell}\sum_{i=1}^{N} W_{i\ell} \boldsymbol{H}_i(\boldsymbol{z}_\ell).$$

These lead to the natural estimators for the various ITT effects:

$$\widehat{\tau}_j = \sum_{\ell=1}^{L} 2^{-(K-1)} g_{j\ell} \overline{Y}^{\text{obs}}(\boldsymbol{z}_\ell), \qquad \widehat{\delta}_j = \sum_{\ell=1}^{L} 2^{-K} g_{j\ell} \boldsymbol{g}_j^\top \overline{\boldsymbol{R}}^{\text{obs}}(\boldsymbol{z}_\ell),$$

$$\widehat{\tau}_{j,p} = \sum_{\ell=1}^{L} 2^{-(K-1)} g_{L-1,\ell} \left(\boldsymbol{g}_j \circ \boldsymbol{g}_{L-1}\right)^\top \overline{\boldsymbol{H}}^{\text{obs}}(\boldsymbol{z}_\ell).$$

Under a completely randomized design, we have $\mathbb{E}\left[\overline{Y}^{\text{obs}}(\boldsymbol{z}) \mid \mathcal{F}\right] = \overline{Y}(\boldsymbol{z})$, which implies that $\widehat{\tau}_j$ is unbiased for $\overline{\tau}_j$ when averaging over the randomization distribution. The same result holds for $\widehat{\delta}_j$ and $\widehat{\tau}_{j,p}$ for $\overline{\delta}_j$ and $\overline{\tau}_{j,p}$, respectively. Importantly, these results do not depend on any of the instrumental variable assumptions and hold by experimental design. Finally, we can define estimators for the MCAFE and the PCAFE as:

$$\widehat{\phi}_j = \widehat{\tau}_j / \widehat{\delta}_j \qquad \widehat{\gamma}_j = \widehat{\tau}_{j,p} / \widehat{\delta}_{L-1}.$$

Each of these estimators has a similar form to the classic Wald estimator: ratios of ITT effects on the outcome to ITT effects on (some function of) treatment uptake.

3 Inference

Inference for instrumental variables estimators has generally followed two broad approaches. First, and more traditionally, one can assume that the data are a random sample from an infinite superpopulation and derive the asymptotic distribution of the various estimators from the central limit

theorem and the delta method. This approach has the advantage that the subsamples corresponding to each treatment assignment vector, $\boldsymbol{z}_\ell$, can be thought of as independent random samples from different population distributions, which greatly simplifies derivation of the large-sample distribution. This approach considers variation in the estimates both from the randomization of $\boldsymbol{Z}_i$ and the random sampling from the superpopulation. The second approach to inference is to take the finite-population quantities $\overline{\phi}_j$ and $\overline{\gamma}_j$ as the quantities of interest and consider the behavior of the estimators over the distribution of the treatment assignments induced by randomization (Fisher, 1935; Imbens and Rosenbaum, 2005). This approach has the advantage that it hews closely to the design of the original experiment and is well-defined even when it is difficult to imagine a hypothetical superpopulation. Below, we present results for the finite-population setting and then show how they change when targeting inference to a superpopulation.

Once an asymptotic distribution has been established, there are several ways to construct confidence intervals for the types of ratio estimators we defined above. The standard way to construct confidence intervals for, say, $\widehat{\phi}_j$ would be to use the delta method on the ratio of $\widehat{\tau}_j$ and $\widehat{\delta}_j$ to obtain an estimator of its asymptotic variance, $\widehat{V}_j$. Then, a 95% confidence interval could be obtained from $\widehat{\phi}_j \pm 1.96 \times \widehat{V}_j$. Unfortunately, this approach, which is based on a Taylor expansion, can be a poor approximation when the denominator is close to 0 (in our case, when there are relatively few compliers).An alternative approach, first proposed by Fieller (1954), uses a carefully chosen test statistic and inverts it to construct the confidence intervals. The key to this approach is that the variance of the test statistic under the null can be written as a quadratic function of null hypothesis of the true effect, allowing the confidence intervals to achieve nominal coverage even when the denominator is close to zero. The trade-off is that these confidence intervals can have infinite length in some samples. See Supplementary Material C for simulations exploring the performance of the different confidence interval methods and for MCAFE vs PCAFE estimators.

3.1 Expectation and variances in the finite population

Although we cannot directly calculate the expectations and variances of our ratio estimators in the finite population, we can derive these properties for their numerators and denominators. Let $\boldsymbol{U}_i(\boldsymbol{z}) = \{\boldsymbol{H}_i(\boldsymbol{z}), \boldsymbol{R}_i(\boldsymbol{z})\}^\top$ be the vector of all $2L$ potential outcomes for unit i under treatment assignment $\boldsymbol{z}$ and let $\overline{\boldsymbol{U}}(\boldsymbol{z})$ be the vector of $2L$ finite-population means. Similarly, let $\widehat{\boldsymbol{U}}(\boldsymbol{z})$ be the vector of estimated means based on treatment assignment. The quantities of interest defined in previous sections are linear combinations of these potential outcomes.

Combining all of the above estimands, we are interested in $r = 3L - 3$ of these effects; $L - 1$ intent-to-treat factorial effects on the outcome, $\overline{\tau}_j$, $L-1$ effects among the perfect compliers, $\overline{\tau}_{j,p}$, and $L - 1$ intent-to-treat effects on the treatment uptake indicators, $\overline{\delta}_j$. As in Li and Ding (2017), we can write our vector of estimands using coefficient matrices $\boldsymbol{Q}_\ell \in \mathbb{R}^{r\times 2L}$ so that we have

$$\boldsymbol{\theta}_i = \sum_{\ell=1}^{L} \boldsymbol{Q}_\ell \boldsymbol{U}_i(\boldsymbol{z}_\ell) \qquad \boldsymbol{\theta}_i = \{\tau_{i1}, \ldots, \tau_{i,L-1}, \tau_{i1,p}, \ldots, \tau_{i,L-1,p}, \delta_{i1}, \ldots, \delta_{i,L-1}\}^\top.$$

Averaging over units, we can write the vector of estimands as

$$\boldsymbol{\theta} = \sum_{\ell=1}^{L} \boldsymbol{Q}_\ell \overline{\boldsymbol{U}}(\boldsymbol{z}_\ell), \qquad \boldsymbol{\theta} = \{\overline{\tau}_1, \ldots, \overline{\tau}_{L-1}, \overline{\tau}_{1,p}, \ldots, \overline{\tau}_{L-1,p}, \overline{\delta}_1, \ldots, \overline{\delta}_{L-1}\}^\top.$$

Furthermore, we can write the vector of estimators for these quantities defined above as $\widehat{\boldsymbol{\theta}} = \sum_{\ell=1}^{L} \boldsymbol{Q}_\ell \widehat{\boldsymbol{U}}(\boldsymbol{z}_\ell)$, where the first entry of $\widehat{\boldsymbol{\theta}}$ is $\widehat{\tau}_1$ and the other values are defined similarly. For our particular quantities of interest, we have

$$\boldsymbol{Q}_\ell = \begin{pmatrix} 2^{-(K-1)} g_{1\ell} \mathbf{1}_L^\top & \mathbf{0}_L^\top \\ \vdots & \vdots \\ 2^{-(K-1)} g_{L-1,\ell} \mathbf{1}_L^\top & \mathbf{0}_L^\top \\ 2^{-(K-1)} g_{L-1,\ell} (\boldsymbol{g}_1 \circ \boldsymbol{g}_{L-1})^\top & \mathbf{0}_L^\top \\ \vdots & \vdots \\ 2^{-(K-1)} g_{L-1,\ell} (\boldsymbol{g}_{L-1} \circ \boldsymbol{g}_{L-1})^\top & \mathbf{0}_L^\top \\ \mathbf{0}_L^\top & 2^{-K} g_{1\ell} \boldsymbol{g}_1^\top \\ \vdots & \vdots \\ \mathbf{0}_L^\top & 2^{-K} g_{L-1,\ell} \boldsymbol{g}_{L-1}^\top \end{pmatrix},$$

where the exact formulations of each block come from the above definitions of the estimands.

To assess the asymptotic distribution of the these estimators, we now define several variance and covariance terms. In particular, let

$$\boldsymbol{S}_\ell^2 = \frac{1}{N-1}\sum_{i=1}^{N}[\boldsymbol{U}_i(\boldsymbol{z}_\ell) - \overline{\boldsymbol{U}}(\boldsymbol{z}_\ell)][\boldsymbol{U}_i(\boldsymbol{z}_\ell) - \overline{\boldsymbol{U}}(\boldsymbol{z}_\ell)]^\top$$

and

$$\boldsymbol{S}_{\boldsymbol{\theta}}^2 = \frac{1}{N-1}\sum_{i=1}^{N}[\boldsymbol{\theta}_i - \boldsymbol{\theta}][\boldsymbol{\theta}_i - \boldsymbol{\theta}]^\top.$$

The first of these, $\boldsymbol{S}_\ell^2$ is the variance of the potential outcomes under treatment assignment $\boldsymbol{z}_\ell$, and the second, $\boldsymbol{S}_{\boldsymbol{\theta}}^2$, is the covariance matrix of the individual-level treatment effects. Note that while $\boldsymbol{S}_\ell^2$ can be identified under the present experimental design, $\boldsymbol{S}_{\boldsymbol{\theta}}^2$ cannot be identified because it would require observing individual-level treatment effects. In particular, we can use the sample variance within each treatment arm to estimate $\boldsymbol{S}_\ell^2$,

$$\boldsymbol{s}_\ell^2 = \frac{1}{N_\ell - 1}\sum_{i:W_{i\ell}=1}\left\{\boldsymbol{U}_i - \widehat{\boldsymbol{U}}(\boldsymbol{z}_\ell)\right\}\left\{\boldsymbol{U}_i - \widehat{\boldsymbol{U}}(\boldsymbol{z}_\ell)\right\}^\top.$$

Under Assumption 1 and over the randomization distribution, $\widehat{\boldsymbol{\theta}}$ has mean $\boldsymbol{\theta}$ and covariance

$$\text{cov}(\widehat{\boldsymbol{\theta}}) = \sum_{\ell=1}^{L}\frac{1}{N_\ell}\boldsymbol{Q}_\ell\boldsymbol{S}_\ell^2\boldsymbol{Q}_\ell^\top - \frac{1}{N}\boldsymbol{S}_{\boldsymbol{\theta}}^2,$$

by Theorem 3 of Li and Ding (2017). This result is a finite-population result and requires no assumptions on the data generating process of the outcomes.

A conservative estimator for the covariance of $\widehat{\boldsymbol{\theta}}$ can be $\widehat{\boldsymbol{V}} = \sum_{\ell=1}^{L} N_\ell^{-1}\boldsymbol{Q}_\ell\boldsymbol{s}_\ell^2\boldsymbol{Q}_\ell^\top$. Given the above result, this will overestimate the covariance of $\widehat{\boldsymbol{\theta}}$ by $N^{-1}\boldsymbol{S}_{\boldsymbol{\theta}}^2$. This latter quantity is generally unestimable because estimating it would require observing the joint distribution of different potential outcomes, $\{\boldsymbol{U}_i(\boldsymbol{z}_1), \ldots, \boldsymbol{U}_i(\boldsymbol{z}_L)\}$. Under the additional stringent assumption that all of the individual-level effects are additive, $\boldsymbol{S}_{\boldsymbol{\theta}}^2$ will be equal to 0 because the effects do not vary across units. In the IV context, however, additive treatment effects are awkward because they would rule out heterogeneous treatment effects that the compliance framework is designed to address.

3.2 Asymptotic distribution under a finite-population approach

In this subsection, we take a finite-population approach to asymptotics that treats $\Pi_N = \{\boldsymbol{U}_1(\boldsymbol{z}_1), \ldots, \boldsymbol{U}_N(\boldsymbol{z}_L)\}$ as a set of fixed population quantities and all randomness comes from the distribution of $\boldsymbol{Z}_i$. To perform asymptotics in this setting, we embed Π_N into a hypothetical sequence of finite populations that grow in size and investigate the properties of our estimators along that sequence (see Lehmann and D'Abrera, 1975; Lehmann, 1999; Li and Ding, 2017, for more on this approach). We assume that we are in a setting where as N increases, N_ℓ also increases without bound for all ℓ. In particular, we assume that N_ℓ/N has a positive limiting value for all ℓ throughout.

We can start by getting a consistency result.

Theorem 1 (Consistency)**.** *Under Assumption 1 and the assumption that* $(1 - N_\ell/N)\boldsymbol{S}_\ell^2/N_\ell \to 0$ *as* $N \to \infty$, $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \xrightarrow{p} 0$ *as* $N \to \infty$.

Proof. From finite population results in, for instance, Rosén (1964) and Scott and Wu (1981), the assumption that $(1 - N_\ell/N)\boldsymbol{S}_\ell^2/N_\ell \to 0$ gives us that $\widehat{\boldsymbol{U}}(\boldsymbol{z}) - \overline{\boldsymbol{U}}(\boldsymbol{z}) \xrightarrow{p} 0$ as $N \to \infty$ for all $\boldsymbol{z}$. Therefore

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \sum_{\ell=1}^{L} \boldsymbol{Q}_\ell \widehat{\boldsymbol{U}}(\boldsymbol{z}_\ell) - \sum_{\ell=1}^{L} \boldsymbol{Q}_\ell \overline{\boldsymbol{U}}(\boldsymbol{z}_\ell) = \sum_{\ell=1}^{L} \boldsymbol{Q}_\ell \left(\widehat{\boldsymbol{U}}(\boldsymbol{z}_\ell) - \overline{\boldsymbol{U}}(\boldsymbol{z}_\ell) \right) \xrightarrow{p} 0.$$

□

We now move on to distributional results. In order to conduct inference on $\widehat{\boldsymbol{\theta}}$, we need to know not only its moments, but also its distribution. While it is possible to computationally approximate the randomization distribution of $\widehat{\boldsymbol{\theta}}$ under a null hypothesis about $\boldsymbol{\theta}$, this approach can be quite complicated and even infeasible when entertaining non-sharp null hypotheses (Kang, Peck, and Keele, 2018). Instead, we rely on finite-population asymptotics to derive an approximation of the distribution $\widehat{\boldsymbol{\theta}}$ as in Li and Ding (2017) and Kang, Peck, and Keele (2018). In this framework, we can derive asymptotic normality of our estimators under a limitation on how much a unit can dominate the population variance. In particular, define the maximum squared distance of the qth coordinate of $\boldsymbol{Q}_\ell \boldsymbol{U}_i(\boldsymbol{z}_\ell)$ from its

population mean,

$$m_\ell(q) = \max_{1\le i\le N}\left[\boldsymbol{Q}_\ell \boldsymbol{U}_i(\boldsymbol{z}_\ell) - \boldsymbol{Q}_\ell \overline{\boldsymbol{U}}(\boldsymbol{z}_\ell)\right]_q^2 \quad 1 \le q \le r,$$

the finite-population variance of the qth coordinate of $\boldsymbol{Q}_\ell \boldsymbol{U}_i(\boldsymbol{z}_\ell)$,

$$v_\ell(q) = \frac{1}{N-1}\sum_{i=1}^{N}\left[\boldsymbol{Q}_\ell \boldsymbol{U}_i(\boldsymbol{z}_\ell) - \boldsymbol{Q}_\ell \overline{\boldsymbol{U}}(\boldsymbol{z})_\ell\right]_q^2 \quad 1 \le q \le r,$$

and the finite-population variance of the qth coordinate of $\boldsymbol{\theta}$,

$$v_{\boldsymbol{\theta}}(q) = \frac{1}{N-1}\sum_{i=1}^{N}[\boldsymbol{\theta}_i - \boldsymbol{\theta}]_q^2 \quad 1 \le q \le r.$$

Li and Ding (2017) derive the following assumptions that are sufficient for asymptotic normality.

Assumption 6. *As* $N \to \infty$,

$$max_\ell max_{1\le q\le r}\frac{1}{N_\ell^2}\frac{m_\ell(q)}{\sum_{\ell'} N_{\ell'}^{-1} v_{\ell'}(q) - N^{-1} v_{\boldsymbol{\theta}}(q)} \to 0$$

Roughly speaking, this assumption limits how a particular unit can dominate the variance of $\boldsymbol{Q}_\ell \boldsymbol{U}_i(\boldsymbol{z}_\ell)$, uniformly across all assignment vectors and components of $\boldsymbol{\theta}$. While this assumption is general and difficult to interpret, Li and Ding (2017) demonstrate several more interpretable conditions that imply this assumption. Finally, we impose a regularity condition on the correlation matrix of $\widehat{\boldsymbol{\theta}}$.

Assumption 7. *The correlation matrix of* $\widehat{\boldsymbol{\theta}}$ *has limiting value* $\boldsymbol{\Sigma}$.

Lemma 1. *Under Assumption 1, 6 and 7, by Theorem 4 of Li and Ding (2017), we have*

$$\left(\frac{\widehat{\tau}_1 - \overline{\tau}_1}{\sqrt{var(\widehat{\tau}_1)}}, \ldots, \frac{\widehat{\tau}_{L-1} - \overline{\tau}_{L-1}}{\sqrt{var(\widehat{\tau}_{L-1})}}, \frac{\widehat{\tau}_{1,c} - \overline{\tau}_{1,c}}{\sqrt{var(\widehat{\tau}_{1,c})}}, \ldots, \frac{\widehat{\tau}_{L-1,c} - \overline{\tau}_{L-1,c}}{\sqrt{var(\widehat{\tau}_{L-1,c})}}, \frac{\widehat{\delta}_1 - \overline{\delta}_1}{\sqrt{var(\widehat{\delta}_1)}}, \ldots, \frac{\widehat{\delta}_{L-1} - \overline{\delta}_{L-1}}{\sqrt{var(\widehat{\delta}_{L-1})}}\right) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}).$$

These results do not rely on any of the instrumental variable assumptions (monotonicity and the exclusion restrictions), and so we can conduct inference on these quantities as ITT effects even if the IV assumptions are suspect. These quantities will gain the additional interpretations in terms of complier effects, as discussed earlier, if the IV assumptions hold.

To get an asymptotic, finite-population distributional result for our IV estimators, which are all ratio estimators, we can use a finite-population delta method (Pashley, 2019).

Lemma 2. *Under Assumption 1, 6 and 7, assumptions for Theorem 1, and also assuming that $\overline{\delta}_j$ has a non-zero limiting value, we have the following asymptotic normality result for our MCAFE estimators:*

$$\frac{\widehat{\phi}_j - \overline{\phi}_j}{\sqrt{\frac{1}{\overline{\delta}_j^2} var(\widehat{\tau}_j) + \overline{\phi}_j^2 \frac{1}{\overline{\delta}_j^2} var(\widehat{\delta}_j) - 2\overline{\phi}_j \frac{1}{\overline{\delta}_j^2} cov(\widehat{\tau}_j, \widehat{\delta}_j)}} \xrightarrow{d} N(0,1).$$

It is straightforward to extend this result to the PCAFEs. Although the delta method is typically associated with a superpopulation perspective, this is a finite-population asymptotic result only requiring standard assumptions on the asymptotic variance and that $\overline{\delta}_j$ has a non-zero limiting value, which under monotonicity and outcome exclusion is the same as assuming that the proportion of compliers for that particular effect has a non-zero limiting value.

We can construct confidence intervals directly from this distribution by estimating the variance as $\frac{1}{\widehat{\delta}_j^2}\widehat{\text{var}}(\widehat{\tau}_j) + \widehat{\phi}_j^2 \frac{1}{\widehat{\delta}_j^2}\widehat{\text{var}}(\widehat{\delta}_j) - 2\widehat{\phi}_j \frac{1}{\widehat{\delta}_j^2}\widehat{\text{cov}}(\widehat{\tau}_j, \widehat{\delta}_j)$. However, we employ a useful trick in the next section to create intervals with potential benefits in terms of coverage and behavior with small compliance probabilities.

Before moving on to this confidence interval method, we give a final consistency result for our ratio estimators:

Lemma 3. *Assume either of the following two sets of conditions:*

(a) the assumptions of Theorem 1 and the additional assumptions that all components of $\boldsymbol{\theta}$ have finite limiting values, and in particular non-zero limiting values for the $\overline{\delta}_j$; OR

(b) the assumptions of Lemma 2 and the additional assumption that $\boldsymbol{S}_t^2$ and $\boldsymbol{S}_{\boldsymbol{\theta}}^2$ have finite limiting values.

Then $\widehat{\phi}_j - \overline{\phi}_j \xrightarrow{p} 0$ and $\widehat{\gamma}_j - \overline{\gamma}_j \xrightarrow{p} 0$ as $N \to \infty$, for all $j \in \{1, \ldots, L-1\}$.

Lemma 3 requires additional regularity conditions on the sequence of finite populations beyond those required in Theorem 1 to avoid situations where the ratio of the population ITTs diverges as $N \to \infty$.

3.3 Constructing confidence intervals for IV effects: Fieller's method

The results of the previous section can be used directly to generate confidence intervals. Here we present a method to create intervals originally from Fieller (1954) and used in Kang, Peck, and Keele (2018) and Li and Ding (2017) in the context of instrumental variables, which performs better with low rates of compliance. We can begin from the result of Lemma 2 to derive this method but it is traditional instead to consider the hypothesis test of a particular value, $H_0 : \overline{\phi}_j = \phi_{j0}$, which can be rewritten as $H_0 : \overline{\tau}_j - \phi_{j0}\overline{\delta}_j = 0$. Following Fieller (1954) and Kang, Peck, and Keele (2018), we use the following test statistic to assess this hypothesis:

$$T(\phi_{j0}) = \widehat{\tau}_j - \phi_{j0}\widehat{\delta}_j,$$

for which it is straightforward to use the asymptotic results previously used to derive the (asymptotic) variance as

$$\sigma^2(\phi_{j0}) = \text{var}(\widehat{\tau}_j) + \phi_{j0}^2\text{var}(\widehat{\delta}_j) - 2\phi_{j0}\text{cov}(\widehat{\tau}_j, \widehat{\delta}_j).$$

We can then obtain $\widehat{\text{var}}(\widehat{\tau}_j)$, $\widehat{\text{var}}(\widehat{\delta}_j)$, and $\widehat{\text{cov}}(\widehat{\tau}_j, \widehat{\delta}_j)$ from $\widehat{\boldsymbol{V}}$ for all j and create the following estimator for the variance of the test statistic:

$$\widehat{\sigma}^2(\phi_{j0}) = \widehat{\text{var}}(\widehat{\tau}_j) + \phi_{j0}^2\widehat{\text{var}}(\widehat{\delta}_j) - 2\phi_{j0}\widehat{\text{cov}}(\widehat{\tau}_j, \widehat{\delta}_j).$$

Under the above results about the approximate normality of these quantities, the typical way to assess this hypothesis is to reject the null if

$$\left|\frac{T(\phi_{j0})}{\widehat{\sigma}(\phi_{j0})}\right| \geq z_{1-\alpha/2}$$

for some pre-specified choice of α. We could then construct a $1-\alpha$ confidence interval for this quantity by inverting the test:

$$\left\{\phi_{j0} : \left|\frac{T(\phi_{j0})}{\widehat{\sigma}(\phi_{j0})}\right| \leq z_{1-\alpha/2}\right\} = \left\{\phi_{j0} : T(\phi_{j0})^2 \leq \widehat{\sigma}^2(\phi_{j0})z_{1-\alpha/2}^2\right\}.$$

Noting that $T(\phi_{j0})^2 = (\widehat{\tau}_j^2 - 2\phi_{j0}\widehat{\tau}_j\widehat{\delta}_j + \phi_{j0}^2\widehat{\delta}_j^2)$, then this implies we can generate the $1-\alpha$ confidence interval by finding: $\{\phi_{j0} : a\phi_{j0}^2 + b\phi_{j0} + c < 0\}$, where

$$
\begin{aligned}
a &= \widehat{\delta}_j^2 - z_{1-\alpha/2}^2 \widehat{\text{var}}(\widehat{\delta}_j) \\
b &= -2\left(\widehat{\tau}_j\widehat{\delta}_j - z_{1-\alpha/2}^2 \widehat{\text{cov}}(\widehat{\tau}_j, \widehat{\delta}_j)\right) \\
c &= \widehat{\tau}_j^2 - z_{1-\alpha/2}^2 \widehat{\text{var}}(\widehat{\tau}_j).
\end{aligned}
$$

As in the case of Fieller (1954), Li and Ding (2017), and Kang, Peck, and Keele (2018), the type of interval generated by this quadratic inequality can take several forms: closed interval, disjoint union of tail intervals, or an infinite-length interval that cover the real line. A similar derivation holds for hypotheses about the perfect complier effects, $\overline{\gamma}_j$, replacing $\widehat{\tau}_j$ with $\widehat{\tau}_{j,p}$.

3.4 Inference under a superpopulation model

If we assume that the data are random samples from an infinite superpopulation, some aspects of inference become simpler. In particular, we can view the observations of Y_i for $\boldsymbol{Z}_i = \boldsymbol{z}$ to be a random sample from the superpopulation distribution of $Y_i(\boldsymbol{z})$, independent from the samples of the other treatment assignments. This means that $\widehat{\boldsymbol{V}}$ is a consistent estimator for the asymptotic covariance of $\widehat{\boldsymbol{\theta}}$. In addition, under mild regularity conditions $\sqrt{N}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ converges in distribution to $N(\boldsymbol{0}, \boldsymbol{V})$. Thus, one can derive confidence intervals for the superpopulation parameters using $\widehat{\boldsymbol{V}}$ and applying either the above delta method or test-inversion methods.

In the Supplemental Material B we describe a Bayesian approach to inference in this setting as that is a popular way to study both factorial experiments (Dasgupta, Pillai, and Rubin, 2015) and instrumental variables (Imbens and Rubin, 1997).

4 Empirical Applications

We now apply our framework to two empirical settings. The first is a field experiment with two factors studying strategies for reducing criminal behavior and has relatively high compliance, both marginally and jointly. The second is a voter mobilization study with three factors and relatively low compliance rates across two of the factors. Both of these studies demonstrate how the choice of

which sample to target—marginalized compliers or perfect compliers–can strongly affect both point estimates and the uncertainty of our inferences.

4.1 The effect of cash transfers and cognitive behavioral therapy on crime and violence

Blattman, Jamison, and Sheridan (2017) investigate a field experiment to assess various strategies for reducing violence and crime among poor young men who had previously engaged in such behavior. In particular, the experiment attempted to compare the use of cash transfers to the use of cognitive behavioral therapy, the latter of which attempted to boost a host of noncognitive skills that would reduce antisocial behaviors. The study randomly assigned these two factors, in independent lotteries, to a sample of 999 high-risk men in Liberia aged 18 to 35. The independent lotteries means that the number of units actually assigned to each of the four treatment combinations is random. However, if we condition on the number of units assigned to each treatment combination, we are exactly back in our (unbalanced) factorial setting with randomization as in Assumption 1. See Pashley, Basse, and Miratrix (2020) for discussion of the validity of this style of conditional as-if analysis. There were additional complications to the experimental design, including some blocking, but we ignore these issues for clarity of the illustration. The cash treatment was equivalent to around three months' wages and the group cognitive behavioral therapy (CBT) was an eight-week program that emphasized self-control, forward-looking behavior, and nonviolence. They then investigate the effects of these treatments on a range of outcomes both in the short-term (2–5 weeks after treatment) and the long-term (12–13 months after treatment), including economic outcomes, antisocial behaviors, and a variety of intermediary outcomes. The economic outcomes include income, consumption, and savings, whereas the antisocial behaviors include measures of drug dealing, fights, weapon-carrying, and thefts.

Blattman, Jamison, and Sheridan (2017) investigated the intent-to-treat effects of cash and CBT, but there was noncompliance for both the cash and CBT treatments. Perhaps unsurprisingly, the cash treatment had a relatively high marginal compliance rate of 0.98. Following Blattman, Jamison, and Sheridan (2017), we define compliance with the CBT treatment as attending 80% of the group therapy

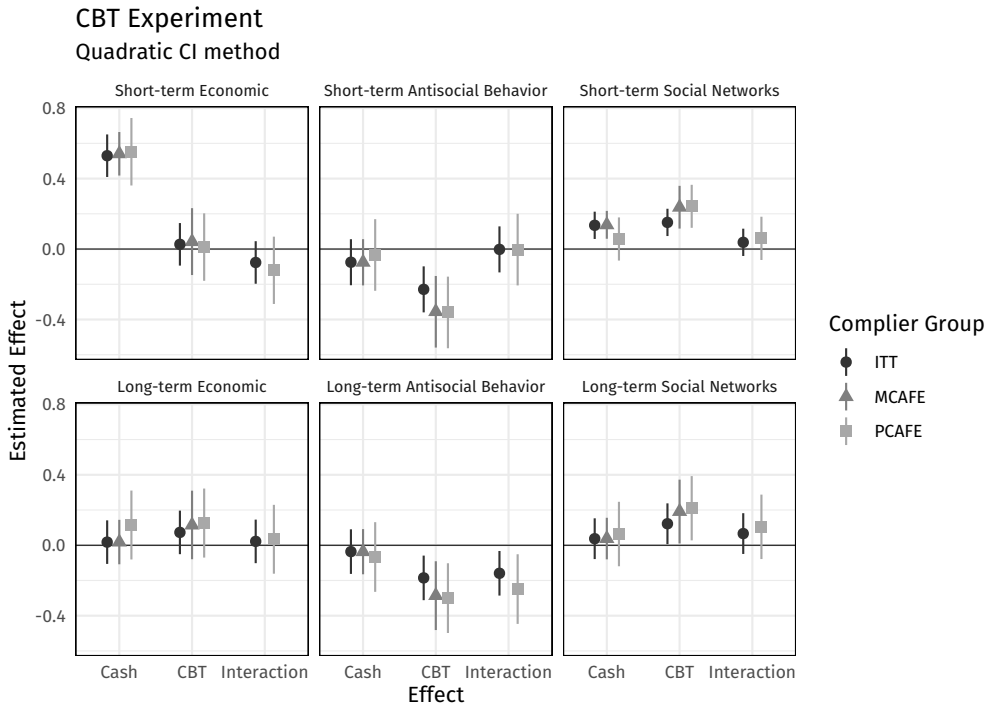

Figure 1: Estimated marginal and perfect complier factorial effects cash transfers and cognitive behavioral therapy. Lines show confidence intervals obtained by the method of Section 3.3.

sessions, and with that definition the marginal compliance rate for the CBT treatment is 0.63. Given the high rate of compliance for the cash treatment, the perfect compliance rate is very similar to the marginal compliance rate for CBT. This definition of compliance for CBT has a disadvantage in that it might make the outcome exclusion restriction less plausible since it would require participants who were assigned to control to have the same outcomes as those who are assigned to CBT and attended 75% of CBT sessions. In Supplemental Material D, we estimate the MCAFEs and PCAFEs under an alternative definition of treatment uptake as attending any CBT sessions, which largely sidesteps these concerns.

Figure 1 displays the ITTs, MCAFEs, and PCAFEs of this experiment on the summary indices of economic measure, antisocial behaviors, and one of the intermediary measures, the quality of social networks. Substantively, these show similar results to the original findings of Blattman, Jamison, and Sheridan (2017), with the cash treatment increasing the short-term economic outcomes for participants, while CBT lowers antisocial behavior and increases the quality of social networks. Given how

CBT noncompliance drives most of the noncompliance in this experiment, the effects of CBT are very similar across the MCAFEs and PCAFEs, whereas both of these tend to differ from the ITTs. Interestingly, the PCAFEs for the cash treatment do sometimes differ from the MCAFEs. The estimated perfect complier effect of cash on long-term economic outcomes is larger than the estimated marginal complier effect, and for short-term quality of social networks, the MCAFE for cash is statistically significant whereas the PCAFE is not.

One way that the factorial setting differs from the univariate IV setting is that it is less clear how ITT factorial effects, $\overline{\tau}_j$, relate to the perfect complier effects, $\overline{\gamma}_j$. With univariate IV, the standard ratio of ITT to compliance rate means that the ITT is will be mechanically lower than the complier average treatment effect, since the compliance rate is less than 1. This basic relationship also holds for the MCAFEs, $\overline{\phi}_j$, and the typical ITT factorial effects, $\overline{\tau}_j$. This does not necessarily hold for PCAFEs since they are the ratio $\overline{\tau}_{j,p}/\overline{\delta}_{L-1}$, and so they will be larger in magnitude than $\overline{\tau}_{j,p}$ but not necessarily larger in magnitude than the usual ITT factorial effect, $\overline{\tau}_j$. For instance, on long-term antisocial behavior, the ITT interaction between cash and CBT is -0.32, whereas the estimated perfect complier interaction is about 25% smaller in magnitude at -0.24. This highlights how important it can be to move past ITTs and investigate noncompliance in the factorial setting.

Finally, we might worry about treatment exclusion in this case because perhaps being assigned cash makes compliance with the CBT condition more likely. We can investigate this by checking if there is an interaction between cash and CBT assignment on uptake for CBT. If there was a significant interaction, this might cast doubt on treatment exclusion. In this case, however, such an interaction is negligible: the proportion CBT uptake is 0.647 when cash is assigned and 0.628 when no cash was assigned ($p = 0.686$). While not a confirmation of treatment exclusion, this might give us some confidence that it is a reasonable assumption in this setting.

4.2 The effect of political canvassing on voter turnout

A large literature in political science uses field experiments to examine the effectiveness of various strategies for encouraging voter turnout in elections. These strategies include phone calls, door-

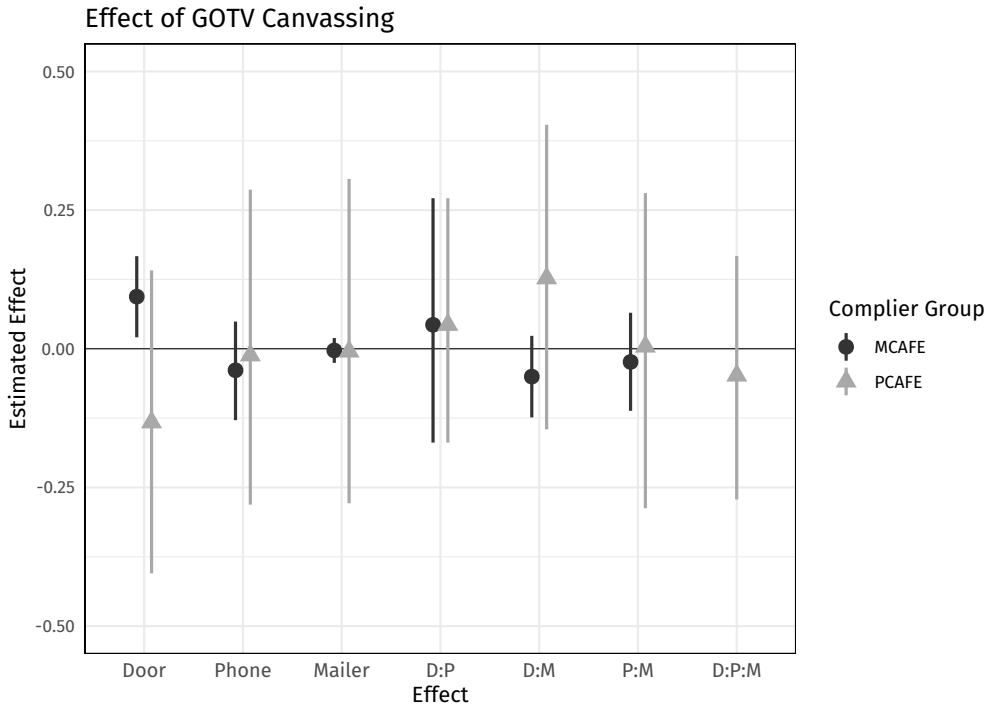

Figure 2: Estimated marginal and perfect complier factorial effects of get-out-the-vote methods on voter turnout.

to-door canvassing, mailers, and more. A ubiquitous problem with these field experiments is non-compliance because relatively few people are willing and able to speak with political canvassers on the phone or at the doorstep. We apply the above framework to a particular get-out-the-vote field experiment fielded in New Haven ahead of the 1998 general election in New Haven, CT (Gerber and Green, 2000). In the original experiment, $N = 23,450$ households were randomly assigned three factors: a door-to-door canvassing visit, a phone call, or a mailer sent to their home. Note that door-to-door canvassing was randomized independently of the other two factors, so again here we are performing a conditional analysis when analyzing as a factorial design, conditioning on the number of people actually assigned to each treatment combination. All of the factors involved messages that encouraged voter turnout. Randomization was done at the household level and the outcome is whether anyone in the household voted in the 1998 general election. Previous studies have analyzed various aspects of this experiment, both substantively and methodologically (Gerber and Green, 2000; Imai, 2005; Hansen and Bowers, 2009; Blackwell, 2017).

Noncompliance in this voter mobilization setting usually occurs when a resident fails to answer the door for an in-person canvassing attempt or fails to answer the phone for a phone canvassing attempt. We could also imagine noncompliance on the mailers factor, but this is difficult to measure—we would have to know if a person both received the mailer and read it closely enough to get the message. Thus, for the purposes of this application, we assume perfect compliance on the mailers factor. One advantage of our approach is that all estimands, estimators, and confidence intervals are well-defined even when some of the factors have perfect compliance. It also emphasizes the benefits of our MCAFE quantities which can be calculated on any given factor without knowing compliance information for other factors. We estimate that the marginal compliance rates for in-person and phone canvassing is 0.296 and 0.282, respectively. The perfect compliance rate, on the other hand, is just 0.104.

Figure 2 shows the estimated MCAFEs and PCAFEs for this voter mobilization study with 95% confidence intervals using the Fieller method. The main substantive takeaway from the results is that only in-person canvassing appears to have a positive and statistically significant effect on turnout, at least for marginal compliers. Other MCAFEs, while sometimes having large point estimates, all have confidence intervals that include 0. The effects for perfect compliers also all have confidence intervals that include zero, and all of these intervals are much wider than for marginal compliers. This demonstrates the loss of precision when attempting to make inferences about a smaller group, even if the resulting coefficients are more directly comparable. Even with that increase in uncertainty, there are striking differences between the point estimates of the PCAFEs and MCAFEs, which could also reflect how the perfect compliers in this setting might be behavioral outliers. Given that the in-person canvassing was done during the day, these are people who are home and willing to talk to about political campaigns in person or over the phone. We may expect these individuals to have different responses to canvassing attempts than the population at large.

5 Conclusion

In this paper we have presented a new framework for 2^K factorial experiments with noncompliance on any number of factors. Under standard instrumental variable assumptions and a treatment exclusion restriction unique to this setting, we showed how there are several ways to define compliance and we exploited this to define two broad classes of factorial effects: those for marginal compliers and those for perfect compliers. Furthermore, we detailed several ways to estimate and make inferences about these quantities of interest.

There are several avenues for extending this framework. The first would be to consider how to proceed with the identification and estimation of bounds for either the overall average factorial effect or various complier factorial effects when the assumptions maintained in this paper do not hold. In particular, the treatment exclusion restriction assumption can be restrictive in that it rules out many types of interactions for compliance. This is especially limiting because interactions are often the target of inference in factorial experiments. Another way to extend this setting would be to allow for more than two levels for each factor given these types of designs are quite common in the social and biomedical sciences. Finally, there are many situations where the compliance status is unknown or only known for a subset of individuals, as in the mailers in the GOTV New Haven experiment. In these settings, it would be useful to use partial identification and bounds to understand what can be learned about the effect of treatment uptake.

References

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–455.

Blackwell, Matthew. 2017. "Instrumental Variable Methods for Conditional Effects and Causal Interaction in Voter Mobilization Experiments." *Journal of the American Statistical Association* 112 (518): 590-599.

Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Vio-

lence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review* 107 (April): 1165-1206.

Cheng, Jing, and Dylan S Small. 2006. "Bounds on causal effects in three-arm trials with non-compliance." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 68 (November): 815–836.

Dasgupta, Tirthankar, Natesh S. Pillai, and Donald B. Rubin. 2015. "Causal inference from 2^K factorial designs by using potential outcomes." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 77 (September): 727–753.

de la Cuesta, Brandon, Naoki Egami, and Kosuke Imai. 2020. "Improving the External Validity of Conjoint Analysis: The Essential Role of Profile Distribution." *Political Analysis*. Forthcoming. **URL:** *https://imai.fas.harvard.edu/research/files/conjoint.pdf*

Egami, Naoki, and Kosuke Imai. 2019. "Causal Interaction in Factorial Experiments: Application to Conjoint Analysis." *Journal of the American Statistical Association* 114 (526): 529-540.

Fieller, E. C. 1954. "Some Problems in Interval Estimation." *Journal of the Royal Statistical Society. Series B (Methodological)* 16 (2): 175–185.

Fisher, R.A. 1935. *The design of experiments.* Edinburgh: Oliver and Boyd.

Gerber, Alan S., and Donald P. Green. 2000. "The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment." *American Political Science Review* 94 (03): 653–663.

Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22 (January): 1–30.

Hansen, Ben B, and Jake Bowers. 2009. "Attributing effects to a cluster-randomized get-out-the-vote campaign." *Journal of the American Statistical Association* 104 (487): 873–885.

Imai, Kosuke. 2005. "Do get-out-the-vote calls reduce turnout? The importance of statistical methods for field experiments." *American Political Science Review* 99 (02): 283–300.

Imai, Kosuke, Zhichao Jiang, and Anup Malai. 2020. "Causal Inference with Interference and Noncompliance in Two-Stage Randomized Experiments." *Journal of the American Statistical Association*. Forthcoming.

Imbens, Guido W., and Donald B. Rubin. 1997. "Bayesian inference for causal effects in randomized experiments with noncompliance." *The Annals of Statistics* 25 (February): 305–327.

Imbens, Guido W., and Paul R. Rosenbaum. 2005. "Robust, Accurate Confidence Intervals with a Weak Instrument: Quarter of Birth and Education." *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 168 (1): 109–126.

Kang, Hyunseung, Laura Peck, and Luke Keele. 2018. "Inference for instrumental variables: a randomization inference approach." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181 (4): 1231–1254.

Lehmann, Erich Leo. 1999. *Elements of large sample theory*. New York: Springer.

Lehmann, Erich Leo., and Howard JM D'Abrera. 1975. *Nonparametrics: Statistical methods based on ranks*. San Francisco: Holden-Day, Inc.

Li, Xinran, and Peng Ding. 2017. "General Forms of Finite Population Central Limit Theorems with Applications to Causal Inference." *Journal of the American Statistical Association* 112 (520): 1759-1769.

Montgomery, Douglas C. 2013. *Design And Analysis of Experiments*. 8th ed. John Wiley & Sons.

Pashley, Nicole E. 2019. "Note on the Delta Method for Finite Population Inference with Applications to Causal Inference.".
URL: *https://arxiv.org/abs/1910.09062*

Pashley, Nicole E, Guillaume W. Basse, and Luke W. Miratrix. 2020. "Conditional As-If Analyses in Randomized Experiments.". Working paper.

Robins, James M. 1989. "The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies." In *Health Service Research Methodology: A Focus on AIDS*, ed. Mulley A. Sechrest L., Freeman H. Washington, DC: U.S. Public Health Service, National Center for Health Services Research.

Rosén, Bengt. 1964. "Limit theorems for sampling from finite populations." *Arkiv för Matematik* 5 (5): 383–424.

Scott, Alastair, and Chien-Fu Wu. 1981. "On the asymptotic distribution of ratio and regression estimators." *Journal of the American Statistical Association* 76 (373): 98–102.

Wald, Abraham. 1940. "The Fitting of Straight Lines if Both Variables are Subject to Error." *The Annals of Mathematical Statistics* 11 (09): 284–300.

Yates, F. 1937. "Design and analysis of factorial experiments." *Technical Communication 35*.

Supplemental Materials (to appear online)

A Main Effect Estimands and Estimators with Alternative Weighting

When defining the factorial main effects we may wish to marginalize over a distribution other than the uniform distribution. For instance, in conjoint experiments one may wish to use the profile distribution as in de la Cuesta, Egami, and Imai (2020). For the main effect for factor j, this would be a reweighting of the conditional effects of factor j, conditioning on the assignment of the other factors. That is, we can define the conditional effect of factor j, conditional on the assignment of all other factors, $\boldsymbol{z}_{[-j]}$, as

$$\tau_{ij}(\boldsymbol{z}_{[-j]}) = Y_i(+1, \boldsymbol{z}_{[-j]}) - Y_i(-1, \boldsymbol{z}_{[-j]}).$$

Then our factorial effect is

$$\tau_{ij} = 2^{-(K-1)} \boldsymbol{g}_j^\top \boldsymbol{Y}_i(\bullet) = 2^{-(K-1)} \sum_{\boldsymbol{z}_{[-j]}} \tau_{ij}(\boldsymbol{z}_{[-j]}).$$

Instead, we may be interested in

$$\widetilde{\tau}_{ij} = \sum_{\boldsymbol{z}_{[-j]}} P^*(\boldsymbol{z}_{[-j]}) \tau_{ij}(\boldsymbol{z}_{[-j]}) = 2^{-(K-1)} \widetilde{\boldsymbol{g}}_j^\top \boldsymbol{Y}_i(\bullet),$$

where $P^*(\boldsymbol{z}_{[-j]}) > 0$, $\sum_{\boldsymbol{z}_{[-j]}} P^*(\boldsymbol{z}_{[-j]}) = 1$, and $\tilde{\boldsymbol{g}}_j$ is an appropriately reweighted version of $\boldsymbol{g}_j$. Note that under treatment exclusion and monotonicity

$$\tau_{ij}(\boldsymbol{z}_{[-j]}) = C_{ij} \tau_{ij}(\boldsymbol{z}_{[-j]})$$

and so

$$\widetilde{\tau}_{ij} = C_{ij} \widetilde{\tau}_{ij}.$$

Thus, our estimands under the IV and exclusion restriction assumptions follow naturally. For example, we can define the MCAFE for factor j as

$$\widetilde{\phi}_j = \frac{\widetilde{\overline{\tau}}_j}{\overline{\delta}_j} = \frac{1}{N_j^c} \sum_{i=1}^{N} C_{ij} \widetilde{\tau}_{ij}$$

and the PCAFE for factor j as

$$\widetilde{\overline{\gamma}}_j = \frac{\widetilde{\overline{\tau}}_{j,c}}{\overline{\delta}_{L-1}} = \frac{1}{N_p} \sum_{i=1}^{N} P_i \widetilde{\tau}_{ij}.$$

Once we have defined these reweighted estimands, estimation follows naturally from the methods in the main text. Under treatment exclusion, our estimator $\widehat{\delta}_j$ can remain unchanged, though a reweighted estimator can also be used as the conditional ITT effects are all the same under treatment exclusion. Our estimators for the numerators of our estimands become

$$\widehat{\widetilde{\tau}}_j = \sum_{\ell=1}^{L} 2^{-(K-1)} \widetilde{g}_{j\ell} \overline{Y}^{\text{obs}}(\boldsymbol{z}_\ell)$$

and

$$\widehat{\widetilde{\tau}}_{j,c} = \sum_{\ell=1}^{L} 2^{-(K-1)} \widetilde{g}_{L-1,\ell} \left(\boldsymbol{g}_j \circ \boldsymbol{g}_{L-1}\right)^\top \overline{\boldsymbol{H}}^{\text{obs}}(\boldsymbol{z}_\ell).$$

Inference will follow Section 3 by simply redefining our $\boldsymbol{Q}_\ell$ matrix.

B Bayesian Inference for Factorial Experiments with Noncompliance

One convenient and popular way to conduct inference for a randomized experiment under noncompliance is to use the Bayesian framework. This approach has been applied to the study of single-factor randomized experiments with noncompliance (Imbens and Rubin, 1997) and to 2^K factorial experiments with perfect compliance (Dasgupta, Pillai, and Rubin, 2015), but not to the combination of these two settings (to our knowledge). While Bayesian methods can used to conduct inference on finite-population quantities of interest (see, e.g., Dasgupta, Pillai, and Rubin, 2015, Section 6), this requires parametric assumptions about the joint distribution of the potential outcomes for any particular unit. Of course, the data are completely uninformative about, for instance, the correlation between potential outcomes for a unit and so posterior inference would be driven by prior assumptions on these parameters. Thus, we follow Imbens and Rubin (1997) and focus on conducting inference for superpopulation parameters, noting that this should generally be conservative relative to the finite-population setting.

The goal of this analysis will be to infer the posterior distribution of the missing potential outcomes and use these posteriors to make inferences about the causal quantities of interest. To do so, we first note that in this setting each unit has the following random variables: $\boldsymbol{Z}_i, \boldsymbol{D}_i(\bullet), \boldsymbol{Y}_i(\bullet)$, where $\boldsymbol{D}_i(\bullet)$ is the $L \times K$ matrix with rows $\boldsymbol{D}_i(\boldsymbol{z}_\ell)^\top$. In addition to these, there is a unit's compliance type, $\boldsymbol{T}_i$, but this is a deterministic function of $\boldsymbol{D}_i(\bullet)$. Of course, for each unit, we only observe one row of $\boldsymbol{D}_i(\bullet)$ and one value in $\boldsymbol{Y}_i(\bullet)$. We collect each of these into matrices $\boldsymbol{Z}$, $\boldsymbol{D}(\bullet)$, $\boldsymbol{Y}(\bullet)$ of dimensions $N \times K$, $N \times LK$ and $N \times L$ with rows $\boldsymbol{Z}_i^\top$, $\text{vec}(\boldsymbol{D}_i(\bullet))^\top$, and $\boldsymbol{Y}_i(\bullet)^\top$, respectively. Let $\boldsymbol{Y}^{\text{obs}}$ and $\boldsymbol{D}^{\text{obs}}$ be the observed entries of the $\boldsymbol{Y}(\bullet)$ and $\boldsymbol{D}(\bullet)$, with $\boldsymbol{Y}^{\text{mis}}$ and $\boldsymbol{D}^{\text{mis}}$ being the missing values and where missingness is defined here by treatment assignment.

Using the random assignment of $\boldsymbol{Z}$, we can characterize the joint distribution of these as

$$f(\boldsymbol{Z}, \boldsymbol{D}(\bullet), \boldsymbol{Y}(\bullet)) = f(\boldsymbol{D}(\bullet), \boldsymbol{Y}(\bullet) \mid \boldsymbol{Z}) f(\boldsymbol{Z}) = f(\boldsymbol{D}(\bullet), \boldsymbol{Y}(\bullet)) f(\boldsymbol{Z}).$$

It will be more convenient to work with the distribution of the compliance types rather than the potential outcomes for treatment. Thus, we use $f(\boldsymbol{T}, \boldsymbol{Y}(\bullet))$ and assume this distribution is independent and identically distributed with the following factorization:

$$f(\boldsymbol{T}, \boldsymbol{Y}(\bullet) \mid \boldsymbol{\rho}, \boldsymbol{\psi}) = \prod_{i=1}^{N} f(\boldsymbol{T}_i \mid \boldsymbol{\rho}) f(\boldsymbol{Y}_i(\bullet) \mid \boldsymbol{T}_i, \boldsymbol{\psi}).$$

The $\boldsymbol{\rho}$ and $\boldsymbol{\psi}$ are parameter vectors with priors $p(\boldsymbol{\rho}, \boldsymbol{\psi})$. Here, $\boldsymbol{\psi}$ with entries $\psi_{\boldsymbol{z}|\boldsymbol{t}}$ are the parameters of the distribution of $Y_i(\boldsymbol{z})$ conditional on $\boldsymbol{T}_i = \boldsymbol{t}$. With this, we can write

$$f(\boldsymbol{T}, \boldsymbol{Y}(\bullet)) = \int \prod_{i=1}^{N} f(\boldsymbol{T}_i \mid \boldsymbol{\rho}) f(\boldsymbol{Y}_i(\bullet) \mid \boldsymbol{T}_i, \boldsymbol{\psi}) p(\boldsymbol{\rho}, \boldsymbol{\psi}) d\boldsymbol{\rho} d\boldsymbol{\psi},$$

and the posterior distribution of these parameters can be written as

$$p(\boldsymbol{\rho}, \boldsymbol{\psi} \mid \boldsymbol{Z}^{\text{obs}}, \boldsymbol{D}^{\text{obs}}, \boldsymbol{Y}^{\text{obs}}) \propto p(\boldsymbol{\rho}, \boldsymbol{\psi}) \int \int \left[\prod_{i=1}^{N} f(\boldsymbol{T}_i \mid \boldsymbol{\rho}) f(\boldsymbol{Y}_i(\bullet) \mid \boldsymbol{T}_i, \psi) d\boldsymbol{Y}_i^{\text{mis}} d\boldsymbol{D}_i^{\text{mis}} \right].$$

Let $\Omega(\boldsymbol{z}, \boldsymbol{d})$ indicate the set of units that have $\boldsymbol{Z}_i = \boldsymbol{z}$ and $\boldsymbol{D}_i = \boldsymbol{d}$. It is easy to show that for each of these subsets, the above integration over the missing potential outcomes will be a mixture over different potential outcome distributions, with the compliance probabilities compatible with that $\boldsymbol{z}$

and $\boldsymbol{d}$ as weights. Let $g(\boldsymbol{z}, \boldsymbol{d}) \in \mathcal{T}_K$ be all of the compliance types such that $\boldsymbol{D}_i(\boldsymbol{z}) = \boldsymbol{d}$. Then, for any particular combination $\boldsymbol{z}$ and $\boldsymbol{d}$, we have

$$\int \int f(\boldsymbol{T}_i \mid \boldsymbol{\rho}) f(\boldsymbol{Y}_i(\bullet) \mid \boldsymbol{T}_i, \psi) d\boldsymbol{Y}_i^{\text{mis}} d\boldsymbol{D}_i^{\text{mis}} = \sum_{\boldsymbol{t} \in g(\boldsymbol{z}, \boldsymbol{d})} \rho_{\boldsymbol{t}} f(Y_i^{\text{obs}} \mid \psi_{\boldsymbol{z}|\boldsymbol{t}}).$$

Combining all the different treatment assignment-uptake combinations, we can write the posterior of the superpopulation parameters as

$$p(\boldsymbol{\rho}, \boldsymbol{\psi} \mid \boldsymbol{Z}^{\text{obs}}, \boldsymbol{D}^{\text{obs}}, \boldsymbol{Y}^{\text{obs}}) \propto p(\boldsymbol{\rho}, \boldsymbol{\psi}) \prod_{\boldsymbol{z}, \boldsymbol{d}} \prod_{i \in \Omega(\boldsymbol{z}, \boldsymbol{d})} \sum_{\boldsymbol{t} \in g(\boldsymbol{z}, \boldsymbol{d})} \rho_{\boldsymbol{t}} f(Y_i^{\text{obs}} \mid \psi_{\boldsymbol{z}|\boldsymbol{t}}).$$

To make inferences on these parameters, we take a data augmentation approach and instead investigate the joint posterior of $(\boldsymbol{\rho}, \boldsymbol{\psi})$ and $\boldsymbol{T}_i$. Computationally, we do this through a Gibbs sampler, iteratively drawing from two distributions:

1. $p(\boldsymbol{T}_i \mid \boldsymbol{Z}^{\text{obs}}, \boldsymbol{D}^{\text{obs}}, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{\rho}, \boldsymbol{\psi})$ for all i;
2. $p(\boldsymbol{\rho}, \boldsymbol{\psi} \mid \boldsymbol{Z}^{\text{obs}}, \boldsymbol{D}^{\text{obs}}, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{T}_i)$.

The compliance statuses for a unit with $\boldsymbol{Z}_i = \boldsymbol{z}$ and $\boldsymbol{D}_i = \boldsymbol{d}$ can be drawn from a multinomial distribution with probabilities:

$$\mathbb{P}(\boldsymbol{T}_i = \boldsymbol{t} \mid \boldsymbol{Z}^{\text{obs}}, \boldsymbol{D}^{\text{obs}}, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{\rho}, \boldsymbol{\psi})) = \frac{\rho_{\boldsymbol{t}} f(Y_i^{\text{obs}} \mid \psi_{\boldsymbol{z}|\boldsymbol{t}})}{\sum_{\boldsymbol{t} \in g(\boldsymbol{z}, \boldsymbol{d})} \rho_{\boldsymbol{t}} f(Y_i^{\text{obs}} \mid \psi_{\boldsymbol{z}|\boldsymbol{t}})}.$$

With a Dirichlet conjugate prior for the compliance probabilities and these draws of compliance types, one can draw the $\boldsymbol{\rho}$ from a standard multinomial distribution. Finally, the distribution parameters, $\boldsymbol{\psi}$, can be obtained from standard posterior update steps within compliance types if typical conjugate priors are used. For example, if the outcome is binary, one could use a binomial model for the outcome, in which case the $\psi_{\boldsymbol{z}|\boldsymbol{t}}$ would be response probabilities and a Beta prior distribution would lead to a simple update step.

One advantage of the Bayesian approach is that it is relatively straightforward to weaken the instrumental variable assumptions and still conduct valid inference on various quantities of interest. For example, Imbens and Rubin (1997) applied these techniques to the single-factor case and analyzed

the posterior distribution of the complier average treatment effect in situations where the exclusion restriction holds and does not hold. This type of analysis could be extended to this setting for both the outcome and treatment exclusion restrictions.

C Simulation Evidence

We conducted simulations to address the performance of our MCAFE and PCAFE estimators and confidence building methods under different scenarios. We focus on a simple 2×2 factorial experiment. In the simulations we varied the probability of being a perfect complier (0.05, 0.1, 0.2, 0.5, 0.75) and the probability of being treated on the first factor (0.25, 0.5, 0.75). For the second factor we kept the treatment probability at 0.5. We also varied the the correlation of potential outcomes (0, 0.5, 1), but this was found to not substantially change the comparison of the methods so we only show results for correlation of 0.5, for simplicity. We ran two versions of the simulation, one for the finite population and one for the superpopulation settings. For each setting, we had a sample size of 1000 units.

To generate potential outcomes, first compliance type was drawn with probability p of being a perfect complier and probability $(1-p)/8$ of all other compliance types. Compliance type then determines the treatment uptake potential outcomes for each unit. For the finite population simulation, the probabilities for each compliance type were drawn exactly for the sample. For the superpopulation simulation, compliance type for each sample was drawn from a multinomial distribution, with appropriate probabilities, for each type.

Once compliance type for each unit in the sample is set, the potential outcomes were drawn according to a distribution specific to that compliance type. In general, the potential outcomes were drawn independently for each unit from a multivariate normal distribution with means specific to the unit's compliance type as given in Table SM.1, marginal variance of 1, and correlation of 0.5 between potential outcomes. The means were chosen such that the compliers had the following properties:

$$E[Y_i(-1,-1) \mid \boldsymbol{T}_i = cc] = 0, \qquad \overline{\gamma}_1^{\text{sp}} = 1, \qquad \overline{\gamma}_2^{\text{sp}} = 2, \qquad \overline{\gamma}_3^{\text{sp}} = -1.$$

We then assumed being an always-taker on one factor increased the potential outcome when not

assigned to treatment for that factor but decreased the treatment effect of the other factor (if the unit was a complier on that factor). For never-takers on one factor, we assumed that the treatment effect for the other factor (if the unit was a complier on that factor) was increased. For the finite population simulation these potential outcomes were drawn once and drawn such that the empirical distribution matched this "population" distribution. For the superpopulation simulation the potential outcomes were redrawn on each run without this empirical criteria.

Compliance type	$z = (-1,-1)$	$z = (-1,+1)$	$z = (+1,-1)$	$z = (+1,+1)$
cc	0	3	2	3
ca	1	1	1.5	1.5
cn	0	0	2	2
ac	1	2.5	1	2.5
aa	1	1	1	1
an	1	1	1	1
nc	0	3	0	3
na	1	1	1	1
nn	0	0	0	0

Table SM.1: Superpopulation expectations of potential outcomes by treatment assignment and compliance type.

Each scenario was run 2000 times and the performance of PCAFE and MCAFE estimators, as well as the two versions of obtaining confidence intervals, were compared. Figures SM.3 and SM.4 show the coverage of the different intervals under the finite-population and superpopulation models, respectively. We see that the standard method for creating confidence intervals tends to be more conservative than the Fieller method, as expected, especially for the PCAFE, where we are estimating effects for a smaller proportion of the data. It is interesting to note that this is true in the finite population and superpopulation. Also as expected, we see more conservative intervals for the finite-population than the superpopulation. We see the conservativeness reduced as the proportion of compliers increases but it is not not linear in compliance probability. Rather, the conservativeness seems to reach a minimum around 0.5 in both the finite-population and superpopulation results.

Figures SM.3 and SM.4 show the median interval length for the different methods under the finite-population and superpopulation models, respectively. For the MCAFEs, both methods of in-

terval creation perform very similarly, with the points overlapping across the board. For the PCAFEs we see that both methods have very large median interval lengths when there is a small compliance probability, Fieller's method being even more extreme.

Based on the simulation results, we see a benefit of using Fieller's method of interval construction for the PCAFEs when there is a low compliance probability, in terms of reducing overcoverage. However, this reduced overcoverage may came at the, somewhat paradoxical, cost of having even more large, non-informative intervals. In terms of MCAFEs, both methods perform similarly though Fieller's method may reduce overcoverage a little.

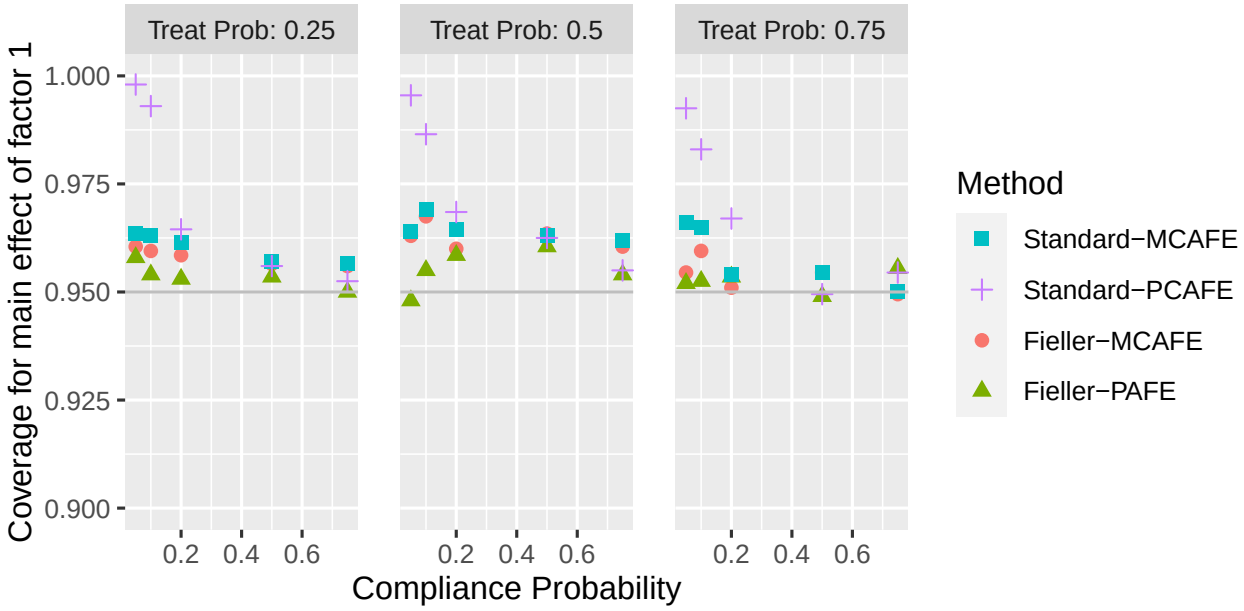

Figure SM.3: Finite population results for coverage of the first factor using the standard and Fieller methods for building confidence intervals.

D Additional Results for Empirical Applications

In the main text, we reported the MCAFEs and PCAFEs for the CBT experiment using the definition of compliance used by Blattman, Jamison, and Sheridan (2017), which was that the respondent attended more than 19 classes of the CBT treatment, or 80% of classes. One downside to this definition of compliance is that it makes the outcome exclusion restriction perhaps difficult to maintain. This

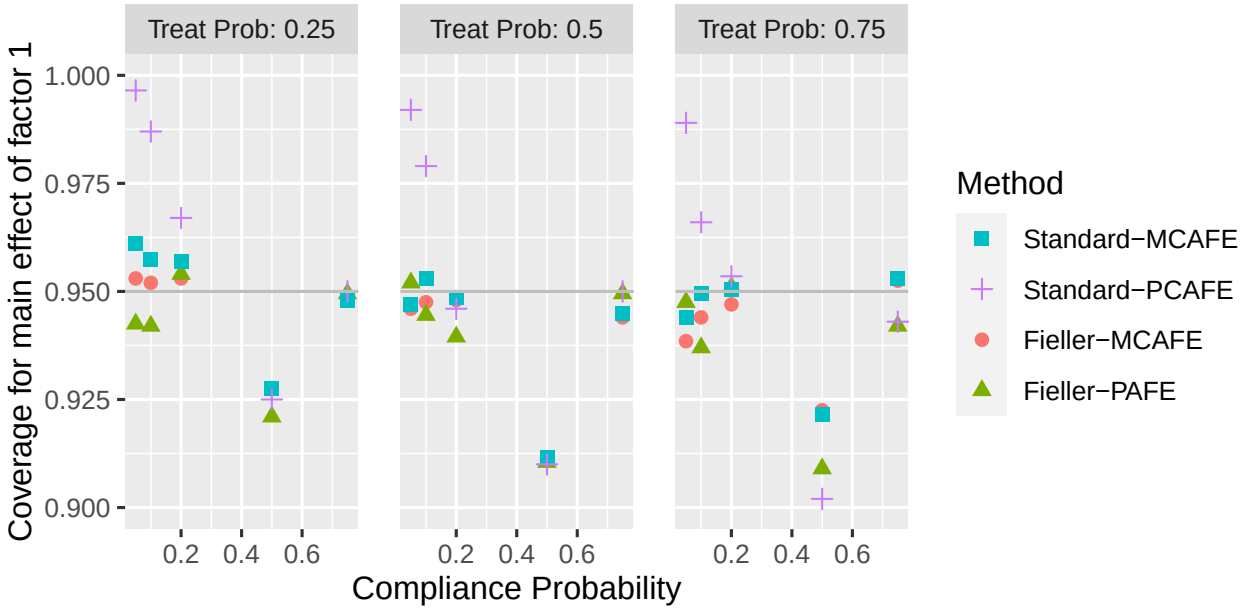

Figure SM.4: Superpopulation results for coverage of the first factor using the standard and Fieller methods for building confidence intervals.

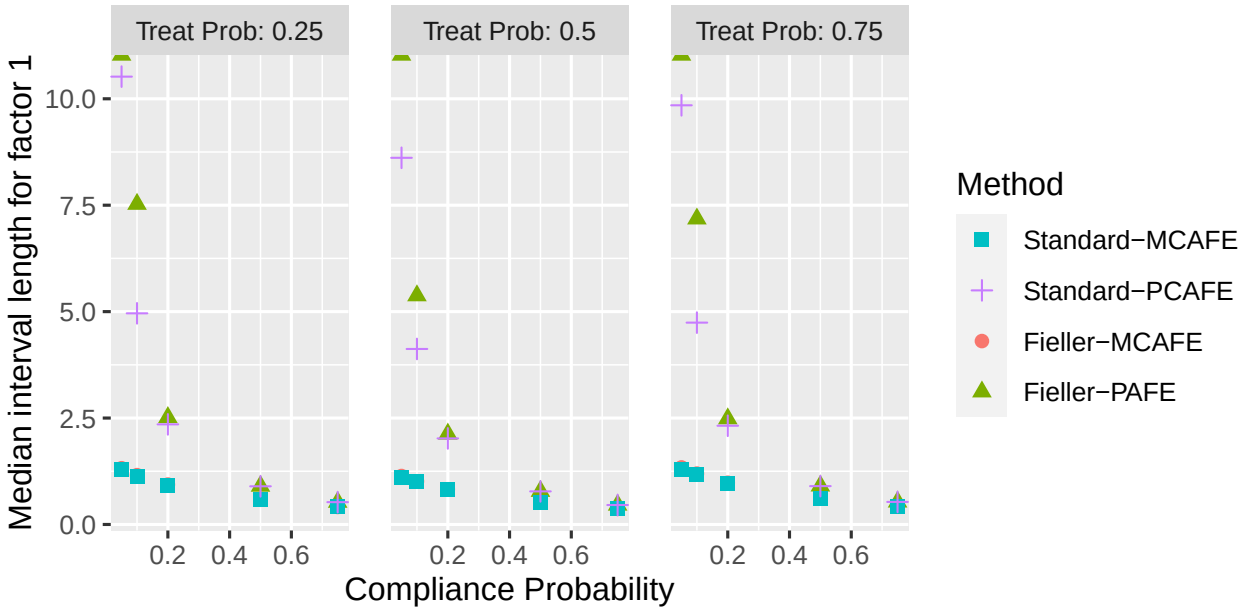

Figure SM.5: Finite population results for interval length of the first factor using the standard and Fieller methods for building confidence intervals.

is because the outcome exclusion restriction would imply that, for example, being assigned to CBT and attending 18 classes has the same effect as not being assigned to CBT at all, for all individuals. To

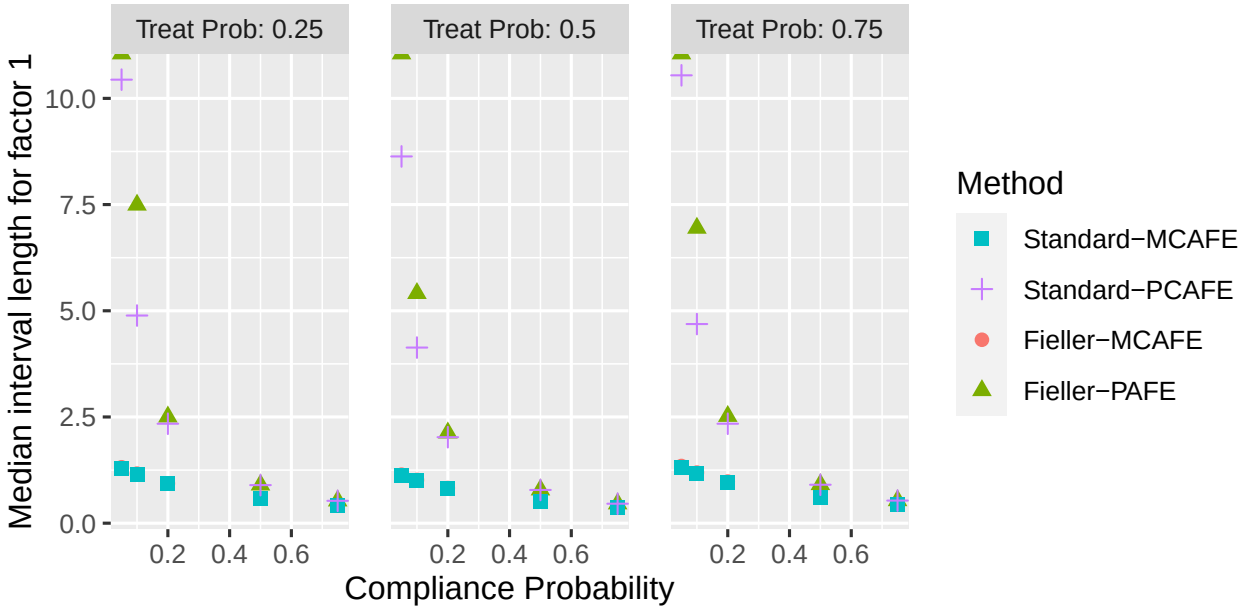

Figure SM.6: Superpopulation results for interval length of the first factor using the standard and Fieller methods for building confidence intervals.

guard against this potential issue, we now present results under a different measure of compliance: whether or not the participant attended any CBT sessions. As one might expect, this vastly increases the marginal compliance rate with CBT to 0.959 and improves the perfect compliance rate to 0.954. Given these improvements to the compliance rates, we should expect that there would be very little variation between the ITTs, the MCAFEs, and the PCAFEs. Figure SM.7 show the results, which are very consistent with this expectation. There are minor variations between the MCAFEs and the PCAFEs, but they are considerably more muted compared to the 80% definition of compliance.

Of course, this alternative definition of treatment uptake has its own problems. In particular, attending only a few sessions may provide extremely little benefit and so somewhat nullifies the goals of an instrumental variable analysis in this setting. A more accurate representation for this setting may be that we have a 2×2 factorial experiment for assignment, but a continuous variable for treatment uptake on one of the factors. It would be interesting to extend the theoretical developments of this paper to this setting of fractional compliance in future work.

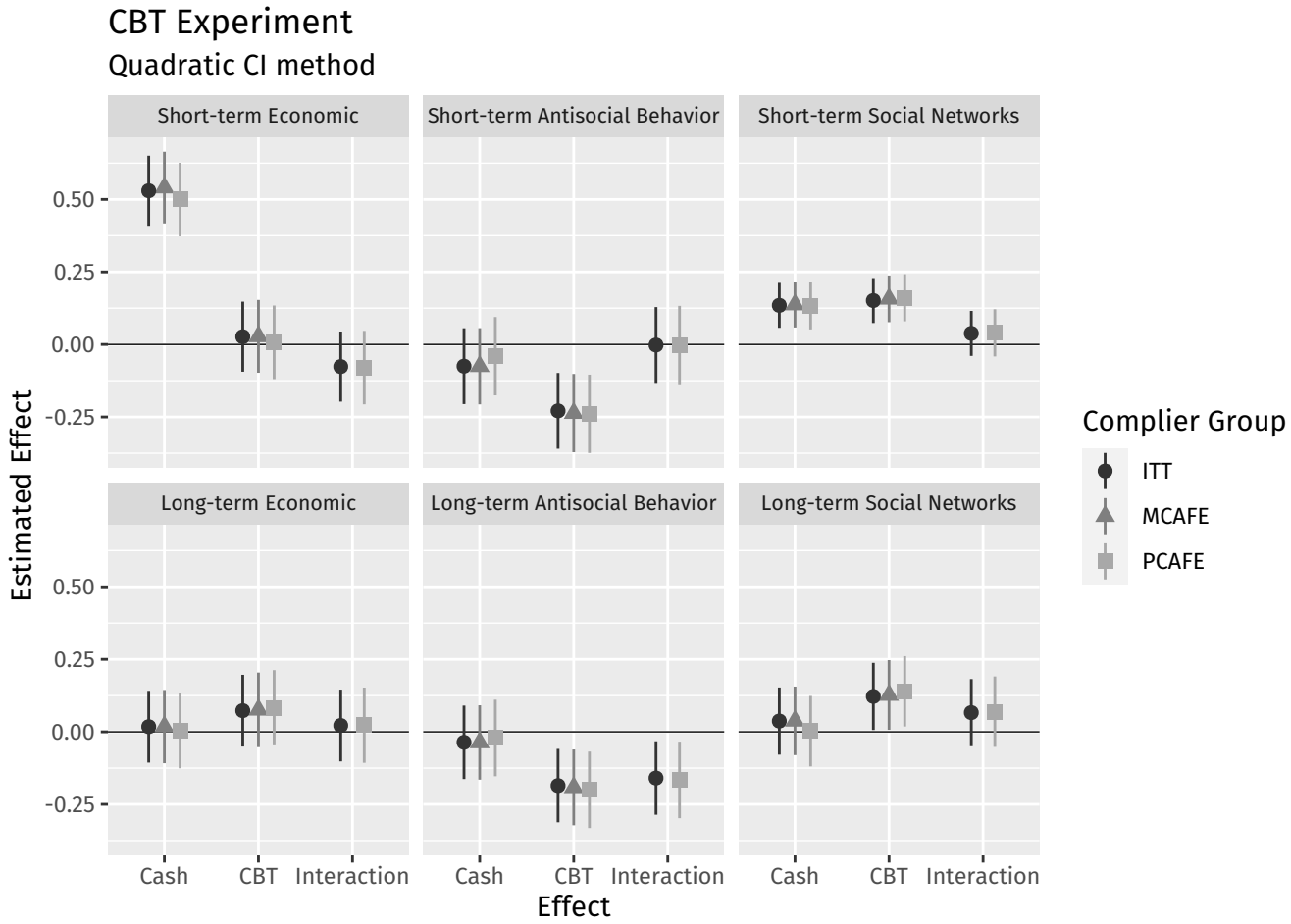

Figure SM.7: Results of the Blattman, Jamison, and Sheridan (2017) replication using "attended any CBT sessions" as the treatment uptake definition.