Paired 2×2 Factorial Design for Treatment Effect Identification and Estimation in the Presence of Paired Interference and Noncompliance

by
Máté Kormos

Submitted to

Central European University
Department of Economics and Business

in partial fulfillment of the requirements for the degree of
Master of Arts in Economics

Supervisor: László Mátyás

Budapest, Hungary
2018

Abstract

In my thesis, I propose an experimental design, the Paired 2×2 Factorial Design, together with two strategies for treatment effect identification and estimation from a large sample of pairs, comprised of distinguishable members (e.g. couples with a healthy and an ill member), when (i) there is interference: the potential outcomes of a pair member do not only depend on the member's own treatment participation but on his/her partner's participation too; (ii) there is (endogenous) noncompliance: the units may not comply perfectly with their treatment encouragement; (iii) the experimenter can have two different binary treatments and two different outcomes of interest for the different members in a pair; (iv) units within a pair are allowed to coordinate their treatment participation based on their encouragement. The latter, number (iv), is the main contribution of my thesis, as this has not been addressed by previous studies.

The first strategy uses only half of the sample to identify the effects of both treatments on both pair members, in certain complier subpopulations, under the usual instrumental variables assumptions. The second strategy, the main theoretical result, uses the whole sample to identify the same treatment effects, but this comes at the cost of additional strict symmetry assumptions on the members' participation willingness. To explore the impact of the violation of symmetry on consistency, some Monte Carlo results are presented.

The use of the design is illustrated, in theory, on a sample of married couples where one member suffers from depression. The treatment is an antidepressant for the depressed member and an educational program for the healthy partner.

Acknowledgements

I am grateful to my supervisor, Professor László Mátyás, for his helpful comments, and to Professor Róbert P. Lieli for the invaluable consultations. I am thankful to Professor Thomas Rooney for the proofreading. I express my thanks and appreciation to the often unjustifiably unacknowledged LaTeX and Python online community. Above all, I am indebted to my family for their support.

Contents

List of Tables

List of Figures

List of Theorems

List of Lemmas

List of Algorithms

1. Introduction

Suppose that we have a random sample of married couples where one member of each couple suffers from depression while the other member, the depressed's spouse, does not. Of interest are the effects of two binary treatments: an antidepressant for the depressed member, and an educational program about depression for the healthy spouse.

In this example, a depressed person's outcome (say, symptoms of depression) might be improved alone by his/her partner's participation in the educational program even if the depressed member does not take the antidepressant. The same is true for the healthy partner, who might feel better due to his/her depressed spouse taking the antidepressant even if he/she does not attend the educational program. If so, the Stable Unit Treatment Value Assumption (SUTVA (Rubin, 1980)), made by most works on treatment effects in the framework of the Rubin Causal Model (Rubin, 1974), is violated as interference arises within predefined pairs of units, the married couples.[1]

Hence, if a researcher wants to make causal inference about the treatment effects of the antidepressant and the educational program in a randomised controlled trial (RCT), a suitable experimental design and methods addressing interference are needed. Furthermore, it is necessary for such methods to correct for the cases when the units do not comply with their treatment encouragement instructions because their expected improvement from doing so is low or negative (in short, endogenous noncompliance, or, somewhat imprecisely, imperfect compliance). For example, if the expected improvement from participating in the educational program is low for both members, then the healthy member is less likely to participate in the educational program when he/she is encouraged to do so.

Methods capable of handling interference and imperfect compliance in RCTs have just been gaining ground recently but are still rare. Moreover, to the best of my knowledge, there exists no such method which allows for within-pair discussion and coordination of treatment participation even though this is desirable in real-life applications. To fill this gap, I propose an easy-to-implement experimental design, built on a coin-tossing protocol, the Paired 2×2 Factorial Design, and two associated large-sample identification and estimation strategies, which

[1] SUTVA states that the potential outcomes of a unit (the unit's outcome were it (not) treated) depends only on the unit's own participation in the treatment but not on others'.

(i) take into account interference present only within pairs of units, where the members of a pair are distinguishable (e.g. healthy-ill, male-female, sister-brother members in each pair in the population); and

(ii) enable the experimenter to have two, potentially member-specific binary treatments and two, potentially member-specific outcomes of interest (or one single outcome belonging to the pair); and

(iii) allow for imperfect compliance, potentially of endogeneous nature, of the units with the treatment encouragement given by the experimenter; and

(iv) allow for within-pair discussion and coordination of treatment participation.

The first strategy uses only half of the sample to identify the average treatment effects of both treatments on both pair members, in certain complier subpopulations, under the usual instrumental variables assumptions (Imbens and Angrist, 1994). The second strategy, the main theoretical result, uses the whole sample to identify the same treatment effects, but this comes at the cost of additional, strict symmetry assumptions on the members' participation willingness. Unfortunately, neither of the strategies is appropriate for identifying the joint effect of the two treatments; only a spillover-like effect can be captured by the second one.

Regarding application, the Paired 2×2 Factorial Design together with the two identification strategies constitute a ready-made tool for practical use. Implemented in Python, each step is feasible.[2] In the next the section, I give some examples for cases when the design could be used.

1.1. Application Domains, Examples

The proposed design allows for the evaluation of two, potentially different, binary treatments on two, potentially different, outcomes belonging to distinguishable pair members (for notations and definitions see Section 3.1). Thus every case in which we have a population of pairs with *distinguishable* members is a valid domain for the application. This includes cases when the treatments are the same for both pair members and/or when the two outcomes of interest are the same for both pair members. Moreover, cases where there is only a single outcome belonging to a pair, i.e. *not* two identical outcomes belonging to the two members, are included as well.

In general there exists a variable along which we can distinguish between members in a pair. To mention probably the most prevalent one, there is age: in each

[2]Find the codes at `GitHub`.

pair we are almost sure to be able to distinguish younger from older members. It is another question, however, that what sense it makes to do so, whether the interpretation is meaningful in this way. This question is to be answered by the experimenter based on the research topic.

Nevertheless, I list some possible examples in Table 1, where the distinction is natural and the design could be useful. Take, e.g. a policy evaluation issue regarding parental leave. We can distinguish between the mother and the father in each couple, and the treatment is the maternity and paternity leave, with new conditions. It is quite reasonable to assume that the parents discuss whether they want to stay at home with the baby given the chance. In this case, the proposed design could be useful to identify the effect of parental leave on the labour supply of both parents afterwards, which could affect some labour policy. Another research on parental leave could assess its effect on the social/psychological development of their child. This is an example for the case when the outcome of interest (development of children) is a single variable and we have one measurement of it per pair, not two, belonging to two parents. A typical economic example is principal-agent contracting. The treatments could be new principal and agent specific clauses in the contract, with outcomes being the satisfaction of the principal with the agent's work and the performance, or stress-level, of the agent.

Table 1: Application examples

Population		Treatment		Outcome	
Member-A	Member-B	Treatment-A	Treatment-B	Outcome-A	Outcome-B
impoverished sister	impoverished brother	special educational program	special educational program	grades	grades
husband	wife	marriage therapy for husbands	marriage therapy for wives	marriage outcome	
mother	father	maternity leave	paternity leave	labour supply	labour supply
mother	father	maternity leave	paternity leave	child development	
sister	brother	vaccination	vaccination	health status	health status
principal	agent	contract	contract	satisfaction	performance

The remainder of the thesis is organised as follows. Section 2 contains the novel components in relation to the existing literature; Section 3 describes the experimental design and the two identification strategies, while in Section 4 simulation results on consistency are reported when symmetry assumptions are violated. Finally, I conclude in Section 5.

2. Previous Work

The idea of interference is not new, with many papers written on the subject based on either randomised trials or observational data (for a review see VanderWeele et al. (2014)). Without any specific structure imposed on the nature of interference, Halloran and Struchiner (1991) conceptualise treatment effects of interest, estimands such as direct and indirect effects. Halloran and Struchiner (1995) refine these concepts aligning them with Rubin's paradigm. Rosenbaum (2007) extends Fisher's sharp null hypothesis of no effect for the interference setting.

Sobel (2006) intruduces the notion of partial interference, on which my design is built so that interference may only occur within groups, pairs, in the population but not across them. Based on Sobel (2006), Hudgens and Halloran (2008) propose a two-stage randomised design to identify and estimate total, direct and indirect treatment effects unbiasedly and consistently and derive variances of the estimators.

As for observational studies, Tchetgen and VanderWeele (2012) contributes to the literature by examining finite sample inference with inverse probability weighting estimators, which are implemented in `R` by Saul and Hudgens (2017). While they do not take into account the within-group correlation structure of treatment participation, Barkley et al. (2017) do so with a group-specific random effect logit model.

In all the randomised trial approaches above, it is assumed that the units perfectly comply with their treatment assignment instruction, which may be unrealistic. Under SUTVA, in the no interference case, the instrumental variables (IV) framework of Imbens and Angrist (1994) provides a means to correct for this and makes it possible to identify treatment effects for the subpopulation who comply with the assignment. However, Sobel (2006) proves that these IV methods do not work under interference. Another, simulation-based, evidence for the undesirable IV properties for the 2×2 factorial design is given by Merrill and McClure (2015), which however assumes that there is no interference effect of the treatments on the outcome.

A solution to this is presented by Kang and Imbens (2016) who develop a two-stage randomised design called Peer Encouragement Design. An important step to account for imperfect compliance, they put forward the idea of personalised treatment. Personalised treatment means that a unit's participation can only depend on the unit's own treatment encouragement, but not on other units' treatment encouragement. As the authors point out, this might be violated if the units are able to discuss and coordinate their participation knowing about each other's encouragement.

My method is vary similar to that of Blackwell (2017) who corrects for non-compliance in the $2{\times}2$ factorial design with IV, when there are two treatments and imperfect compliance can occur along both treatments. However, in his framework there is no paired stucture, the level of analysis is the individual/person. In other words, interference arises from the interaction of the two treatments not from another individual's treatment (he makes the assumption of no interference across people). While my paired experiment setup can be thought of as having two treatments for the same person (we consider education of the partner as the second treatment for the depressed spouse) with imperfect compliance occuring along both treatments, Blackwell does not study a paired structure. Neglecting paired structure is not necessarily a problem when we are only interested in the outcome of either the depressed or the non-depressed person (or in a single outcome per pair), yet it is restrictive when both are of interest.

More importantly, Blackwell makes the same personalised treatment assumption as Kang and Imbens (2016), corresponding to the individual-level analysis , so that participation in 'treatment-A' only depends on the encouragement of 'treatment-A' but not on that of 'treatment-B'. In some cases this may be plausible but in the paired structure it rules out within-pair discussion as Kang and Imbens (2016) writes, so it is less credible.

As opposed to this, my proposed design and strategies relax this assumption by allowing for such discussion and coordination within a pair. To the best of my knowledge, there exists no other work permitting this yet. Although the symmetry assumptions of the second strategy are strict, my work can be considered as a first step in this direction.

Compared to previous works on interference, I focus my attention on pairs, which is restrictive on one hand, but the results are more easily interpretable. This is achieved by allowing for member-specific treatments (antidepressant for the depressed member, educational program for the spouse) and also for different outcome of interests for the different members of the pairs (depression symptoms for the depressed member, some well-being/anxiety measure for the spouse). Previous studies typically define group-level averages to identify effects more easily; however this would not work with member-specific treatments and outcomes of interest.

3. Identification

In this section, after the general framework and notation is introduced (Section 3.1), I present the protocol of the Paired 2×2 Factorial Design (Section 3.3), with which are associated two identification strategies. Following the discussion of the identifying assumptions (Section 3.4), the two strategies are described.

The first strategy uses only half of the sample to infer treatment effects (Half-sample Identification Strategy, see Section 3.5). This strategy is not a theoretical advancement per se, it is merely a 'trick' to see the problem as one in the traditional Imbens and Angrist (1994) IV-setting. Consequently, the advantage of it is that the assumptions of Imbens and Angrist (1994) applied to the interference setup are sufficient for identification (see Theorem 1). The drawback is that it throws away the information present in the other half of the sample. The second strategy uses the whole sample for identification and is the main theoretical result of this thesis (Full-sample Identification Strategy, see Section 3.6 and Theorem 2). This, however, comes at the cost of additional strict assumptions.

3.1. Framework & Notation

3.1.1 Statistical Viewpoint & Sampling.

In analysing relationships between variables, multiple approaches can be taken depending on how we view our available data at hand. We can consider our available data as the whole population or a random sample drawn from the population.[3] Throughout the thesis, I take the view that there is an infinite population of pairs from which the available data are a single random sample drawn with replacement. The reason is that, even if the whole population is available, we are likely to want to sample from it due to cost considerations. Sampling randomly with replacement from the population is important because this way the sample is independently and identically distributed which renders the analysis simpler. Given this viewpoint, there are two possible scenarios.

The first scenario is the cleanest, statistically speaking. Suppose that we have the complete list of the pairs in the population of interest. *In principle*, we assign treatment to every single unit in the population in the same way as treatment assignment protocol is given in Algorithm 1 with the difference that this time we do this for the whole population, not only for the sample. Then we draw a random

[3]For further details on the implications of this on the analysis of randomised experiments see Athey and Imbens (2016)

sample of pairs with replacement (e.g. we independently generate random integers uniformly corresponding to the unique identifier of the pairs), and now we tell the members in the sampled pairs that what their in-principle treatment encoragement was. In this way, there is a theoretical chance that a pair shows up in the sample multiple times (because we sample with replacement); however the probability of this is near-zero.

In the second scenario, suppose that we do not have the list of the population, but we have a large number of available pairs. Now if we assume that this available set is itself a random sample drawn with replacement from the population of interest, we can apply the very same encouragement protocol as that in Algorithm 1 for this sample. In this way, however, the probability that a pair is present in the sample multiple times is exactly zero.

Either way, we end up with an independently and identically distributed large sample of size n from the infinite population of pairs for the analysis. Hence, every random variable introduced below is a single draw from their respective distribution and is indexed with i: $i = 1, \ldots, n$. In general, I omit the index i, except for the introduction below and when it is necessary to write it out.

3.1.2 Partial Interference

According to partial interference, the interference may occur only within specific groups, pairs in this case, but not across them. Interference stands for possible interactions between potential outcomes and treatment participation (see Section 3.1.5) within a pair.

3.1.3 Distinguishable Members

The pair members are distinguishable; thus we have member-A and member-B in each pair of the population. Equivalently, there exists at least one variable along which we can deterministically distingiush member-A from member-B in the whole population of interest. In our example, the discriminating variable is whether someone is depressed (say, member-A) or not (member-B). Other examples for divison: healthy-ill, doctor-patient, sister-brother, man-woman, younger-older pairs etc..

3.1.4 Treatment Encouragement

Let $Z_{Ai} \in \{0, 1\}$ and $Z_{Bi} \in \{0, 1\}$, $i = 1, \ldots, n$ be the binary random variables indicating the treatment encouragement (assignment) of member-A and member-B in the ith pair of the n-large sample. Then $Z_{Ai} = 1$ if and only if member-A in the ith pair is encouraged to take the member-A specific treatment, treatment-A; and $Z_{Bi} = 1$ if and only if member-B in the ith pair is encouraged to take the memer-B specific treatment, treatment-B. For example, if member-A is the depressed person, $Z_{Ai} = 1$ and $Z_{Bi} = 1$ means that in pair i we encourage the depressed person to take the antidepressant and his/her partner to enroll to the educational program.

3.1.5 Treatment Participation

Denoting the observable, actual treatment participation with D, we can distinguish between four binary random variables belonging to member-A and another four belonging to member-B. As a function of the treatment encouragements in pair i, this is compactly written as $D_{Ai}(Z_{Ai}, Z_{Bi}) \in \{0, 1\}$ and $D_{Bi}(Z_{Ai}, Z_{Bi}) \in \{0, 1\}$. That is, we have $D_{Ai}(00), D_{Ai}(10), D_{Ai}(01), D_{Ai}(11)$ and $D_{Bi}(00), D_{Bi}(10), D_{Bi}(01), D_{Bi}(11)$.[4] To ease notation, no comma is used to separate arguments when actual numbers (0,1) are used. It is important to pay attention to the order of the arguments: A comes first so that $D_{Ai}(01)$ and $D_{Bi}(01)$ describes the actual treatment participation of members when member-A is not encouraged to take his/her treatment ($Z_{Ai} = 0$) and member-B is encouraged to take his/her treatment ($Z_{Bi} = 1$).

Among the four variables, there is only one which is observable as there is only one treatment encouragement which can be given to pair i. In our example, we obviously cannot instruct the very same depressed person to take and not to take the antidepressant at the same time. Conveniently, the observable one among the four random variables can be written as

$$\begin{aligned} D_{Ai} =& Z_{Ai}Z_{Bi}D_{Ai}(11) + Z_{Ai}(1 - Z_{Bi})D_{Ai}(10) \\ &+ (1 - Z_{Ai})Z_{Bi}D_{Ai}(01) + (1 - Z_{Ai})(1 - Z_{Bi})D_{Ai}(00) \qquad (1) \\ D_{Bi} =& Z_{Ai}Z_{Bi}D_{Bi}(11) + Z_{Ai}(1 - Z_{Bi})D_{Bi}(10) \\ &+ (1 - Z_{Ai})Z_{Bi}D_{Bi}(01) + (1 - Z_{Ai})(1 - Z_{Bi})D_{Bi}(00), \qquad (2) \end{aligned}$$

where the binary treatment encouragement (Z_{Ai}, Z_{Bi}) 'activates' the appropriate random variables $(D_{Ai}(Z_{Ai}, Z_{Bi}), D_{Bi}(Z_{Ai}, Z_{Bi}))$ corresponding to the given treatment encouragement.

[4]The personalised treatment of Kang and Imbens (2016) formally states that $D_{Ai}(Z_{Ai}, Z_{Bi}) = D_{Ai}(Z_{Ai})$ and $D_{Bi}(Z_{Ai}, Z_{Bi}) = D_{Bi}(Z_{Bi})$.

Once again, suppose that member-A is the depressed person. Then by not giving the depressed person the antidepressant ($Z_{Ai} = 0$) while encouraging his/her spouse to enroll to the educational program ($Z_{Bi} = 1$), we observe $D_{Ai}(01)$ and $D_{Bi}(01)$ telling us whether the depressed member takes the pill and whether the non-depressed one enrolls to the program given the encouragement (no pill, education program).

3.1.6 Potential Outcomes

There are four potential outcomes in pair i for member-A: $Y_{Ai}(D_{Ai}, D_{Bi}) \in \mathbb{R}^1$ and four for member-B: $Y_{Bi}(D_{Ai}, D_{Bi}) \in \mathbb{R}^1$. These are four-four random variables which characterise the outcome of interest as a function of the *actual treatment participation* of both members in the pair. They can be thought of as random variables describing the outcome in four states of the world, of which we can only witness one. Similarly to the treatment encouragment, the observable one of the four is written as

$$\begin{aligned} Y_{Ai} =& D_{Ai}D_{Bi}Y_{Ai}(11) + D_{Ai}(1 - D_{Bi})Y_{Ai}(10) \\ &+ (1 - D_{Ai})D_{Bi}Y_{Ai}(01) + (1 - D_{Ai})(1 - D_{Bi})Y_{Ai}(00) \qquad (3) \\ Y_{Bi} =& D_{Ai}D_{Bi}Y_{Bi}(11) + D_{Ai}(1 - D_{Bi})Y_{Bi}(10) \\ &+ (1 - D_{Ai})D_{Bi}Y_{Bi}(01) + (1 - D_{Ai})(1 - D_{Bi})Y_{Bi}(00), \qquad (4) \end{aligned}$$

where the binary actual treatment participations (D_{Ai}, D_{Bi}) select the observable one. It might be the case that $Y_{Ai}(D_{Ai}, D_{Bi})$ and $Y_{Bi}(D_{Ai}, D_{Bi})$ measures the same thing for the members, i.e. GPA of member-A and GPA of member-B. It might also be the case that there is only a single potential outcome per pair (e.g. marriage outcome of married couples; see Section 1.1). If so, $Y_{Ai}(D_{Ai}, D_{Bi}) = Y_{Bi}(D_{Ai}, D_{Bi})$, thus A-B indexing is unnecessarry and the effects in Section 3.5 and 3.6 is to be interpreted accordingly.

In our example (member-A is the depressed), $Y_{Ai}(10)$ and $Y_{Bi}(10)$ are the outcomes of the pair members when the depressed actually takes the antidepressant and his/her spouse is not enrolled to the educational program; $Y_{Ai}(11)$ and $Y_{Bi}(11)$ are the outcomes when the depressed takes the pill and the non-depressed in enrolled to the program and so on.

3.2. Variables in the Sample

To understand how the proposed estimator works, a few more notations have to be brought in. Let y_{Ai} (y_{Bi}) denote the sample analogue of the observable outcome for member-A (member-B); let $\boldsymbol{d}_{Ai} \equiv [1, d_{Ai}, d_{Bi}, d_{Ai}d_{Bi}]' \in \{0,1\}^{4\times1}$ and $\boldsymbol{d}_{Bi} \equiv [1, d_{Bi}, d_{Ai}, d_{Ai}d_{Bi}]' \in \{0,1\}^{4\times1}$, where d_{Ai} (d_{Bi}) is the sample-realisation of treatment participation of member-A (member-B) in pair i; and last let $\boldsymbol{z}_{Ai} \equiv [1, z_{Ai}, z_{Bi}, z_{Ai}z_{Bi}]' \in \{0,1\}^{4\times1}$ and $\boldsymbol{z}_{Bi} \equiv [1, z_{Bi}, z_{Ai}, z_{Ai}z_{Bi}]' \in \{0,1\}^{4\times1}$, where z_{Ai} (z_{Bi}) is the sample-realisation of the treatment encouragement of member-A (member-B) in pair i. Broadcasting them to $\boldsymbol{y}_A \equiv [y_{A1}, \ldots, y_{An}]' \in \mathbb{R}^{n\times1}$ and $\boldsymbol{y}_B \equiv [y_{B1}, \ldots, y_{Bn}]' \in \mathbb{R}^{n\times1}$, $\boldsymbol{D}'_A \equiv [\boldsymbol{d}_{A1}, \ldots, \boldsymbol{d}_{An}] \in \{0,1\}^{4\times n}$ and $\boldsymbol{D}'_B \equiv [\boldsymbol{d}_{B1}, \ldots, \boldsymbol{d}_{Bn}] \in \{0,1\}^{4\times n}$, $\boldsymbol{Z}'_A \equiv [\boldsymbol{z}_{A1}, \ldots, \boldsymbol{z}_{An}] \in \{0,1\}^{4\times n}$ and $\boldsymbol{Z}'_B \equiv [\boldsymbol{z}_{B1}, \ldots, \boldsymbol{z}_{Bn}] \in \{0,1\}^{4\times n}$ facilitates more compact notation, with '$'$' indicating the transpose.

3.3. Paired 2×2 Factorial Design

After having obtained a large random sample of pairs, the next step in the experiment is to assign treatments to the units (e.g. which depressed person is encouraged to take the antidepressant and which healthy person is encouraged to enroll to the educational program). The exact, algorithmic procedure of doing so is the experimental design/protocol, which is presented in this section.

In our case, a good starting point is the 2×2 factorial design, which is suitable for exploring the interaction effect between two binary treatments on a signle unit (Cheng, 2013). The depression example can be thought of as having two treatment for the same unit/person (we consider education of the partner as the second treatment for the depressed spouse) which gives rise to the factorial design.

The treatment encouragement protocol is specified, somewhat formally, in Algorithm 1, and is illustrated with the depression example in Algorithm 2. The encouragement mechanism is fairly simple and the advantage of it is that upon treatment assignment, the experimenter does not have to take into account the paired structure: it is only later, in the treatment effect estimation phase, when it is necessary to keep track of the paired structure. In contrast to the personalised treatment assumption of Kang and Imbens (2016), the experimenter does not have to care about whether the members in a pair know about each other's encouragement or not, as discussion and coordination between them is allowed for in the estimation phase.

Algorithm 1 Paired 2×2 Factorial Design

draw an n-large i.i.d. sample from the population of (member-A, member-B) pairs
for k=1:2n **do** ▷ for each *unit*, k, in the sample
  draw x from Bernoulli(P) i.i.d. with $P = 0.5$
  if $x \geq 0.5$ **then**
    if person is member-A **then**
      encouragement(person) $\leftarrow$ 'take treatment-A!'
    else
      encouragement(person) $\leftarrow$ 'take treatment-B!'
    $z_k \leftarrow 1$ ▷ dummy indicating treatment
  else
    encouragement(person) $\leftarrow$ excluded from treatment
    $z_k \leftarrow 0$
Output: n pairs with member-specific treatment assignment: $\{(z_{Ai}, z_{Bi})\}_{i=1}^{n}$

Algorithm 2 Paired 2×2 Factorial Design example

draw an i.i.d. sample from the population of (depressed, not depressed) pairs
for each *person* in the sample **do**
  flip a fair coin: $\mathbb{P}\,(\text{head}) \equiv P = 0.5$
  if head **then**
    if person is depressed **then**
      encouragement(person) $\leftarrow$ 'take the pill!'
    else
      encouragement(person) $\leftarrow$ 'enroll to educational program!'
  else
    encouragement(person) $\leftarrow$ cannot access to pill/education
Output: n pairs with member-specific treatment assignment

The encouragement protocol leads to four goups of pairs along treatment assignment: based on the observed values of (Z_A, Z_B) we have $\mathcal{G}_{Z_A,Z_B} : \mathcal{G}_{00}, \mathcal{G}_{10}, \mathcal{G}_{01}, \mathcal{G}_{11}$. In our example (member-A is depressed), these are the pairs with treatment encouragement: (no pill, no education), (pill, education), (no pill, education), (pill, education), respectively. It is easy to see that for $P = 0.5$ the groups consist of approximately the same number of pairs, i.e. the proportion of pairs in each group is roughly 25%, even though we neglected the paired structure during encouragement.[5] These groups play a role in the different identification strategies in Section 3.5 and 3.6.

[5]See the Monte Carlo simulation of the assignment in `assignment_mechanism.py`, which verifies the 25% proportion.

3.4. Identifying Assumptions

Once the treatments are assigned to the units based on the protocol in Algorithm 1, the experimenter observes whether they comply and participate in the treatment, and records their outcome of interest (e.g. symptoms of depression). Before analysing these data to identify treatment effects with the strategies in Section 3.5 or 3.6, it is crucial that the experimenter is aware of the underlying assumptions. In this section, the Identifying Assumptions, i.e. the sufficient conditions to identify treatment effects, are outlined. The two (half- and full-sample) identification strategies depend on different set of assumptions: the half-sample identification relies solely on assumptions $A1 - A6$, while full-sample identification requires all assumptions $A1 - A8$. First, I describe the assumptions formally, and then in a more intuitive way.

Some assumptions address the treatment encouragement and thus are completely under the control of the experimenter; they are met by following Algorithm 1 (these assumptions are not in italic). Other assumptions concern the participation behaviour of the subjects and hence are not under the control of the experimenter (indicated in *italic*). In between these is *One-sided noncompliance*, which depends on both the experimenter and the nature of the treatments. An in-depth view on assumptions concerning the distribution of treatment participation ($A4 - A8$) is provided in Appendix A.

Identifying Assumptions Usual IV assumptions, extended for interference:

A1 Exclusion:
$$Y_A(D_A, D_B, Z_A, Z_B) = Y_A(D_A, D_B)$$
$$Y_B(D_A, D_B, Z_A, Z_B) = Y_B(D_A, D_B)$$

A2 Random assignment:
$$[Y_A(D_A, D_B), Y_B(D_A, D_B), D_A(Z_A, Z_B), D_B(Z_A, Z_B)] \perp\!\!\!\perp [Z_A, Z_B]$$

A3 i.i.d. assigment: $Z_A \perp\!\!\!\perp Z_B$ and $\mathbb{P}(Z_A = 1) = \mathbb{P}(Z_B = 1) \equiv P \in (0, 1)$

A4 *One-sided noncompliance:*
$$D_A(Z_A = 0, Z_B) = D_B(Z_A, Z_B = 0) = 0 \forall Z_A, Z_B$$

A5 *Monotonicity:*
$$D_A(Z_A = 1, Z_B = 0) \leq D_A(Z_A = 1, Z_B = 1)$$
$$D_B(Z_A = 0, Z_B = 1) \leq D_B(Z_A = 1, Z_B = 1)$$

A6 *Invertibility:*
$$\mathbb{P}(D_A(Z_A = 1, Z_B = 1) = 1, D_A(Z_A = 1, Z_B = 0) = 1) \neq 0$$
$$\mathbb{P}(D_B(Z_A = 1, Z_B = 1) = 1, D_B(Z_A = 0, Z_B = 1) = 1) \neq 0$$

Symmetry assumptions:

***A*7** *Joint willingness:*

$$\mathbb{P}\left(D_A(Z_A=1,Z_B=1)=1\right)=\mathbb{P}\left(D_B(Z_A=1,Z_B=1)=1\right)$$

***A*8** *Spouse sensitivity:*

$$\begin{aligned}&\mathbb{P}\left(D_A(Z_A=1,Z_B=1)=1,D_A(Z_A=1,Z_B=0)=0\right)\\=&\mathbb{P}\left(D_B(Z_A=1,Z_B=1)=1,D_B(Z_A=0,Z_B=1)=0\right).\end{aligned}$$

3.4.1 Exclusion

The outcome variable is not influenced directly by the assignment, only by the actual treatment participation. That is, the fact itself that the unit is instructed to take the treatment has no effect whatsoever on the outcome of interest – it is only the fact whether the pair members participate in the treatments which may influence the outcome.

3.4.2 Random Assignment

The assignment must be independent of the outcomes and the actual treatment participation. This rules out that the experimenter encourges those units to take the treatment who is more likely to (i) benefit from it (e.g. will have higher/lower outcome) or (ii) follow their treatment instruction (e.g. will have higher participation). Hence, the experimenter must randomise the treatment assignment.

3.4.3 i.i.d. Assignment

Every single *unit* in the sample is independently encouraged, with the same probability, to take his/her appropriate member-specific treatment. As a result, the experimenter does not have to take into account who is the pair of whom during the assignment procedure; the only thing has to be tracked is whether the unit is member-A or member-B in the pair in which he/she is.

3.4.4 One-sided Noncompliance

Whenever a unit is not encouraged to take the treatment, he/she is enforcably excluded from participation, i.e. has no access to the treatment.

3.4.5 Monotonicity

If a member is willing to take the treatment when his/her partner is excluded from doing so, he/she must also take the treatment whenever they are both encouraged to take the treatment.

3.4.6 Invertibility

There must be such member-A and member-B units in the population who participate when he/she is encouraged but his/her pair is excluded from the treatment. In fact, these units form the complier subpopulations for whom the average effects are identified (see later). Intuitively, if there are no units who respect their own encouragment, ignoring their partner's, we cannot identify and estimate the effects.

3.4.7 Joint Willingness

For member-A the willingness to participate when both members in the pair are encouraged is the same as the willingness of member-B.

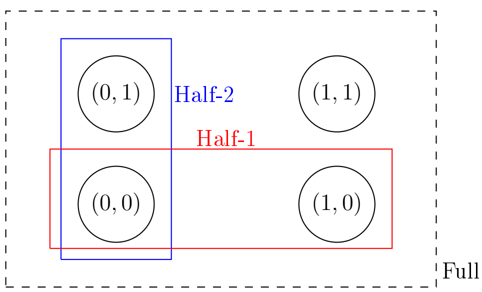
3.4.8 Spouse Sensitivity

Member-A and member-B are equally sensitive to the exclusion of their spouse from the treatment assignment in terms of their participation willingness.

3.5. Half-sample Identification Strategy

Having discussed the assumptions, we can turn to the Half-sample Identification Strategy, which is rewarding as fewer assumptions are sufficient to identify treatment effects, but is costly as we throw away half of the sample. Figure 1 illustrates how the half-sample and the full-sample strategies relate to one another.

Figure 1: Half- and Full-sample strategies, with (Z_A, Z_B) values in the nodes



As discussed earlier, the sample can be split up into groups $\mathcal{G}_{00}, \mathcal{G}_{10}, \mathcal{G}_{01}, \mathcal{G}_{11}$ based on the sample values of (Z_A, Z_B) . In the Full-sample Identification Strategy we use the observations in all four groups. In the Half-sample Identification Strategy we use observations in either groups $\mathcal{G}_{00}$ and $\mathcal{G}_{10}$ (Half-1), or $\mathcal{G}_{00}$ and $\mathcal{G}_{01}$ (Half-2). In our example, these are the pairs: Half-1: (no pill, no education) and (pill, no education), or Half-2: (no pill, no education) and (no pill, education). That is, when one member is excluded from the treatment, so there are exactly two half-sample strategies.

What the half-sample strategy exploits is $A4$ One-sided noncompliance. Because under Identifying Assumptions $A1$ Exclusion and $A4$ One-sided noncompliance, the observable outcome variable can be written as

Half-1

$$\begin{aligned} Y_A^{\text{Half-1}} &= D_A Y_A(10) + (1 - D_A) Y_A(00) \\ &= Y_A(00) + (Y_A(10) - Y_A(00)) D_A \\ Y_B^{\text{Half-1}} &= D_A Y_B(10) + (1 - D_A) Y_B(00) \\ &= Y_B(00) + (Y_B(10) - Y_B(00)) D_A \end{aligned}$$

Half-2

$$\begin{aligned} Y_A^{\text{Half-2}} &= D_B Y_A(01) + (1 - D_B) Y_A(00) \\ &= Y_A(00) + (Y_A(01) - Y_A(00)) D_B \\ Y_B^{\text{Half-2}} &= D_B Y_B(01) + (1 - D_B) Y_B(00) \\ &= Y_B(00) + (Y_B(01) - Y_B(00)) D_B. \end{aligned}$$

From this follows Theorem 1.

Theorem 1 (Half-sample Identification)**.** *Applying the IV method of Imbens and Angrist (1994) in each of the cases (Half-1 and Half-2) separately (that is, instrumenting D_A with Z_A and D_B with Z_B), the Paired 2×2 Factorial Design and Identifying Assumptions $A4 - A6$ are sufficient to identify:*

[1] *Average baseline outcomes:* $\mathbb{E}[Y_A(00)]$ *and* $\mathbb{E}[Y_B(00)]$

[2] *Average effect of own treatment for compliers (ATEO):*

$\mathbb{E}[Y_A(10) - Y_A(00) \mid D_A(10) = 1, D_A(11) = 1]$

$\mathbb{E}[Y_B(01) - Y_B(00) \mid D_B(01) = 1, D_B(11) = 1]$

[3] *Average effect of partner's treatment for those with complier partner (ATEP):*

$\mathbb{E}[Y_A(01) - Y_A(00) \mid D_B(01) = 1, D_B(11) = 1]$

$\mathbb{E}[Y_B(10) - Y_B(00) \mid D_A(10) = 1, D_A(11) = 1]$.

Proof: see Appendix D.

According to Theorem 1 we can identify the following averages for the member-A:

[1] $\mathbb{E}[Y_A(00)]$ which is the average baseline level for member-A in the whole population, the expected outcome when none of the members is treated. In our example (member-A is the depressed), this is the average outcome of the depressed member when he/she does not take the antidepressant, and nor does his/her spouse participate in the educational program.

[2] $\mathbb{E}[Y_A(10) - Y_A(00) \mid D_A(11) = 1, D_A(10) = 1]$ which is the average treatment effect of his/her own (member-specific) treatment on member-A, in the complier subpopulation of member-A's, i.e. those who respect their own treatment encouragment regardless of their partner's access to treatment. In our example, this is the average treatment effect of the antidepressant (given to the depressed member) on the depressed member when the non-depressed has no access to the educational program for the (sub)population of those depressed people who are willing to take the antidepressant regardless of their spouse's access to the educational program.

[3] $\mathbb{E}[Y_A(01) - Y_A(00) \mid D_B(11) = 1, D_B(01) = 1]$ which is the average treatment effect of his/her partner's (member-specific) treatment on member-A, in the subpopulation of member-A's who have complier partners, i.e. those whose partner respect their treatment encouragement, regardless of member-A's access to treatment. In our example, this is the average treatment effect of the educational program (given to the non-depressed member) on the depressed member when the depressed has no access to the antidepressant for the (sub)population of those depressed people who have partners that are willing to take the educational program regardless of the depressed's access to the antidepressant.

Similarly, for member-B:

[1] $\mathbb{E}[Y_B(00)]$ which is the average baseline level for member-B in the whole population, the expected outcome when none of the members is treated. In our example, thsi is the average outcome of the non-depressed member when neither him/her nor his/her spouse receives their corresponding treatment.

[2] $\mathbb{E}[Y_B(01) - Y_B(00) \mid D_B(11) = 1, D_B(01) = 1]$ which is the average treatment effect of his/her own (member-specific) treatment on member-B, in the complier subpopulation of member-B's, i.e. those who respect their own treatment

encouragment regardless of their partner's access to treatment. In our example, this is the average treatment effect of the educational program (given to the non-depressed member) on the non-depressed member when the depressed has no access to the antidepressant for the (sub)population of those non-depressed people who are willing to take the educational program regardless of their spouse's access to the antidepressant.

[3] $\mathbb{E}\left[Y_B(10) - Y_B(00) \mid D_A(11) = 1, D_A(10) = 1\right]$ which is the average treatment effect of his/her partner's (member-specific) treatment on member-B, in the subpopulation of member-B's who have complier partners, i.e. those whose partner respect their treatment encouragement, regardless of member-B's access to treatment. In our example, this is the average treatment effect of the antidepressant (given to the depressed member) on the non-depressed member when the non-depressed has no access to the educational program for the (sub)population of those non-depressed people who have partners that are willing to take the antidepressant regardless of the non-depressed's access to the educational program.

There are only two half-sample strategies in Figure 1. Why only two if there are more possibilities? In the same spirit, it is tempting to pick any other two groups and compare the outcomes similarly. The problem is that comparison will not be simpler then in the full-sample case and/or there is not enough variation in the instrument(s) (treatment encoragement(s)) or the correlation between the encouragement and the participation is insufficient (singular matrix). I do not provide details on this in this thesis – it can be shown by going through the initial steps of the proof of Theorem 2 in the Appendix.

3.6. Full-sample Identification Strategy

The Full-sample Identification Strategy, the main theoretical advancement of my work, is also illustrated in Figure 1. The name is self-explanatory: the strategy has the advantage over the half-sample one that we use information present in the whole sample. But the experimenter has to be aware of the high price paid for the extra information: Identifying Assumptions $A7 - A8$. These symmetry assumptions are the most restrictive of all, and are unlikely to meet exactly in practice. However, were it not for $A7 - A8$, the IV-based identification, which I use, would become intractable.[6]

[6]Even with $A7 - A8$ holding, the proof is quite involved (see Appendix E).

If the experimenter is willing to make Identifying Assumptions $A1 - A8$ then he/she can refer to Theorem 2 for the identifiable treatment effects. The identified effects $[1]-[3]$ are identical to those in Theorem 1 and are described in Section 3.5. What, in addition, is identified is effect [4]. This is a spillover-like effect which is quite hard to interpret (in words).

Theorem 2 (Full-sample Identification)**.** *Let*

$$\hat{\boldsymbol{\theta}}_A \equiv \left(n^{-1}\boldsymbol{Z}_A'\boldsymbol{D}_A\right)^{-1} n^{-1}\boldsymbol{Z}_A'\boldsymbol{y}_A = \left(n^{-1}\sum_{i=1}^{n}\boldsymbol{z}_{Ai}\boldsymbol{d}_{Ai}'\right)^{-1} n^{-1}\sum_{i=1}^{n}\boldsymbol{z}_{Ai}y_{Ai}$$

$$\hat{\boldsymbol{\theta}}_B \equiv \left(n^{-1}\boldsymbol{Z}_B'\boldsymbol{D}_B\right)^{-1} n^{-1}\boldsymbol{Z}_B'\boldsymbol{y}_B = \left(n^{-1}\sum_{i=1}^{n}\boldsymbol{z}_{Bi}\boldsymbol{d}_{Bi}'\right)^{-1} n^{-1}\sum_{i=1}^{n}\boldsymbol{z}_{Bi}y_{Bi}.$$

Then the Paired 2×2 Factorial Design and Identifiying Assumptions $A4-A8$ are sufficient to identify:

[1] *Average baseline outcome:* $\mathbb{E}\left[Y_A(00)\right]$

[2] *Average effect of own treatment for compliers (ATEO):*

$$\mathbb{E}\left[Y_A(10)-Y_A(00)\mid D_A(11)=1,D_A(10)=1\right]$$

[3] *Average effect of partner's treatment for those with complier partner (ATEP):*

$$\mathbb{E}\left[Y_A(01)-Y_A(00)\mid D_B(11)=1,D_B(01)=1\right]$$

[4] *Spillover-like effect:*

$$\begin{aligned}
&\mathbb{E}\left[Y_A(11)-Y_A(10)-\left[Y_A(10)-Y_A(00)\right]\mid D_A(11)=1,D_B(11)=1\right]\\
&+\frac{r}{\bar{q}}\left\{\mathbb{E}\left[Y_A(10)-Y_A(00)\mid D_A(11)=1,D_A(10)=0\right]\right.\\
&\left.-\mathbb{E}\left[Y_A(10)-Y_A(00)\mid D_A(11)=1,D_A(10)=1\right]\right\}\\
&+\frac{r}{\bar{q}}\left\{\mathbb{E}\left[Y_A(01)-Y_A(00)\mid D_B(11)=1,D_B(01)=0\right]\right.\\
&\left.-\mathbb{E}\left[Y_A(01)-Y_A(00)\mid D_B(11)=1,D_B(01)=1\right]\right\}
\end{aligned}$$

as the probability limit plim $\hat{\boldsymbol{\theta}}_A$*, and:*

[1] *Average baseline outcome:* $\mathbb{E}\left[Y_B(00)\right]$

[2] *Average effect of own effect for compliers (ATEO):*

$$\mathbb{E}\left[Y_B(01)-Y_B(00)\mid D_B(11)=1,D_B(01)=1\right]$$

[3] *Average effect of partner's treatment for those with complier partner (ATEP):*

$$\mathbb{E}\left[Y_B(10)-Y_B(00)\mid D_A(11)=1,D_A(10)=1\right]$$

[4] *Spillover-like effect:*

$$\mathbb{E}\left[Y_B(11) - Y_B(01) - [Y_B(01) - Y_B(00)] \mid D_B(11) = 1, D_A(11) = 1\right]$$
$$+ \frac{r}{\bar{q}} \{\mathbb{E}\left[Y_B(01) - Y_B(00) \mid D_B(11) = 1, D_B(01) = 0\right]$$
$$- \mathbb{E}\left[Y_B(01) - Y_B(00) \mid D_B(11) = 1, D_B(01) = 1\right]\}$$
$$+ \frac{r}{\bar{q}} \{\mathbb{E}\left[Y_B(10) - Y_B(00) \mid D_A(11) = 1, D_A(10) = 0\right]$$
$$- \mathbb{E}\left[Y_B(10) - Y_B(00) \mid D_A(11) = 1, D_A(10) = 1\right]\}$$

as the probability limit plim $\hat{\boldsymbol{\theta}}_B$*, where* $r \equiv \mathbb{P}\left(D_A(11) = 1, D_A(10) = 0\right) = \mathbb{P}\left(D_B(11) = 1, D_B(01) = 0\right)$ *and* $\bar{q} \equiv \mathbb{P}\left(D_A(11) = 1, D_B(11) = 1\right)$.

Proof: see Appendix E.

It may as well concern us what happens if we refer to the full-sample strategy while assumptions $A7 - A8$ do not hold. To examine this case in one particular setup, I conduct a simulation study, which is reported in the next section.

4. Violation of Symmetry Assumptions: a Monte Carlo Study of Consistency

To explore the impact of the violation of the symmetry assumptions $A7 - A8$ of the full-sample strategy, I carry out a Monte Carlo study to examine how off the estimates are compared to the case when the assumptions are met. For the simulation the distributions of the relvant random variables have to be specified. The distribution of the treatment encouragements (Z_A, Z_B) is known, however those of the treatment participations and the potential outcomes are not entirely. Hence, an arbitrary decision has to be made on these distributions and, as a consequence, the result of the simulation is *conditional* on their specification.

First, I construct two types of probability distributions for the treatment participation.[7] One of the so-constructed distributions fulfills Identifying Assumptions $A7 - A8$:

$$\mathbb{P}\left(D_A(11) = 1\right) - \mathbb{P}\left(D_B(11) = 1\right) = 0 \tag{5}$$

$$\mathbb{P}\left(D_A(11) = 1, D_A(10) = 0\right) - \mathbb{P}\left(D_B(11) = 1, D_B(01) = 0\right) = 0, \tag{6}$$

[7] Each type is a multivariate Bernoulli distribution (Dai et al., 2012) of the vector $(D_A(10), D_A(11), D_B(01), D_B(11)) \in \{0,1\}^4$. Note that $D_A(00) = D_A(01) = D_B(00) = D_B(10) = 0$ because of $A4$ One-sided noncompliance, hence we do not have to include them in the distributions; further note that the distribution of $(D_A(10), D_A(11), D_B(01), D_B(11))$ is zero for some combinations because of $A5$ Monotonicity.

while the other is crafted on purpose to violate $A7 - A8$:

$$\mathbb{P}\left(D_A(11)=1\right)-\mathbb{P}\left(D_B(11)=1\right)=d_1 \tag{7}$$

$$\mathbb{P}\left(D_A(11)=1, D_A(10)=0\right)-\mathbb{P}\left(D_B(11)=1, D_B(01)=0\right)=d_2 \tag{8}$$

$$d_1 \neq 0 \tag{9}$$

$$d_2 \neq 0. \tag{10}$$

Having fixed the deviations at eirher zero or not, the distribution of treatment participation is not yet fully specified as there remains a lot of free parameters to assign value to. (For further details on their choice and on how the Identifying Assumptions $A4 - A8$ are present in the distribution, refer to Appendix A and the Python codes.[8])

Second, the distribution of the potential outcomes *given* the treatment participations remains to specified. This is a conditional distribution, the advantage of which is that it captures endogenous self-selection into the treatment which necessitates the IV approach in the first place. In other words, the distribution represents the situation when noncompliance occurs because the unit knows that participation will not, on average, mean a better outcome for him/her than opting out. For sake of simplicity, I work with a multivariate normal distribution, where the mean vector depends on the treatment participation[9], but the covariance matrix does not. That is, the units are sensitive to the expected outcomes but not to their variation. Then how the mean vector is specified is, again, arbitrary. In this study, I choose a few behavioural patterns based on which I build the conditional mean vector. These are the following: (i) higher outcome is better – units aim to achieve high expected change in outcomes; and (ii) laziness – if no strictly positive expected change in outcome, then no participation; and (iii) lazy altruism toward spouse – participation even if the unit is worse-off as long as the sum of the expected changes for both members is strictly positive. The covariance matrix is specified so that any two potential outcomes are positively correlated, which is often the case in (social science) applications. (For further details refer to Appendix B.)

Finally, having specified the distributions, we can see how sensitive the consistency is to the violation of symmetry assumptions $A7 - A8$. To get a clear picture, I

[8] For now, let us just remark that the parameters are given so that the distribution is approximately uniform over the combinations of the values of $(D_A(10), D_A(11), D_B(01), D_B(11))$, except those probabilites directly affected by d_1, d_2.

[9] i.e. on the the value of $(D_A(10),\ D_A(11),\ D_B(01),\ D_B(11))$

trace out the bias of the four-four treatment effects estimators (four for member-A and B) *as compared to the case when* $A7 - A8$ *hold*[10], and their variances for different combinations of d_1, d_2. (For exact details of the implementation, see Algorithm 3 in Appendix C and the Python codes.) Using 1000 Monte Carlo repetitions, I conducted the experiment with sample sizes $n = 250, 300, 400$, the results of which are in Figures 2, 3, 4 respectively.[11]

The baseline estimator when symmetry is violated ($\hat{\theta}_1$) does not seem to exhibit any regular pattern as a function of the deviations (d_1, d_2); however the bias as compared to the symmetrical case shrinks, as does the variance, as the sample size grows. Thus the baseline estimator seems to show mean square convergence (which implies consistency) irrespective of whether symmetry holds.

The same irregularity goes for the bias of the own effect estimator for member-A ($\hat{\theta}_2^A$) and the partner effect estimator for member-B ($\hat{\theta}_3^B$). Their variances, on the other hand, exhibit the same behaviour as a function of the deviations, and are decaying to zero in the neighbourhood of $d_1 = 0, d_2 = 0$ as n grows. This suggests that the consistency of these estimators is somewhat robust to minor violations of symmetry assumptions.

Similarly, the bias of the own effect estimator for member-B ($\hat{\theta}_2^B$) and the bias of the partner effect estimator for member-A ($\hat{\theta}_3^A$) display the same pattern. So does their variances, which is more sensitive to the deviation in d_2. In the light of this, the consistency of these estimators is less robust to minor violations of symmetry assumptions.

The bias of the estimator of the spillover-like effect ($\hat{\theta}_4$) reveals a regular pattern, just like its variance, for both members. Even though the variance is by far the largest for the spillover-like effect, it slowly decreses in the sample size. Judging by the figures, the consistency of these estimators is probably more robust to minor violations of symmetry assumptions than $\hat{\theta}_2^B$ and $\hat{\theta}_3^A$, but less than $\hat{\theta}_2^A$ and $\hat{\theta}_3^B$.

[10] That is, (expected value when symmetry does not hold)-(expected value when symmetry holds).

[11] White blank areas in the figures are points in the (d_1, d_2) space for which the probability mass function of the treatment participation is invalid (negative) given the chosen values of other parameters of the distribution.

Figure 2: Monte Carlo results when symmetry does not hold, $n = 250$

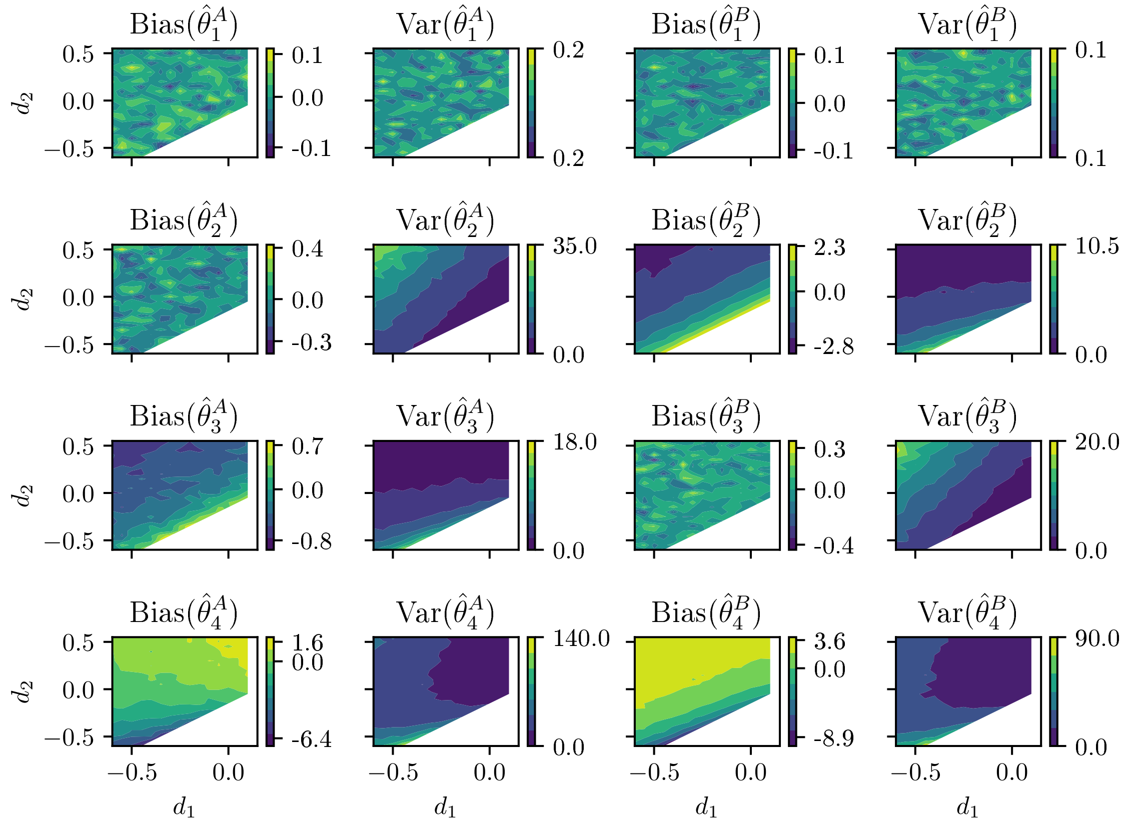


Figure 3: Monte Carlo results when symmetry does not hold, $n = 300$

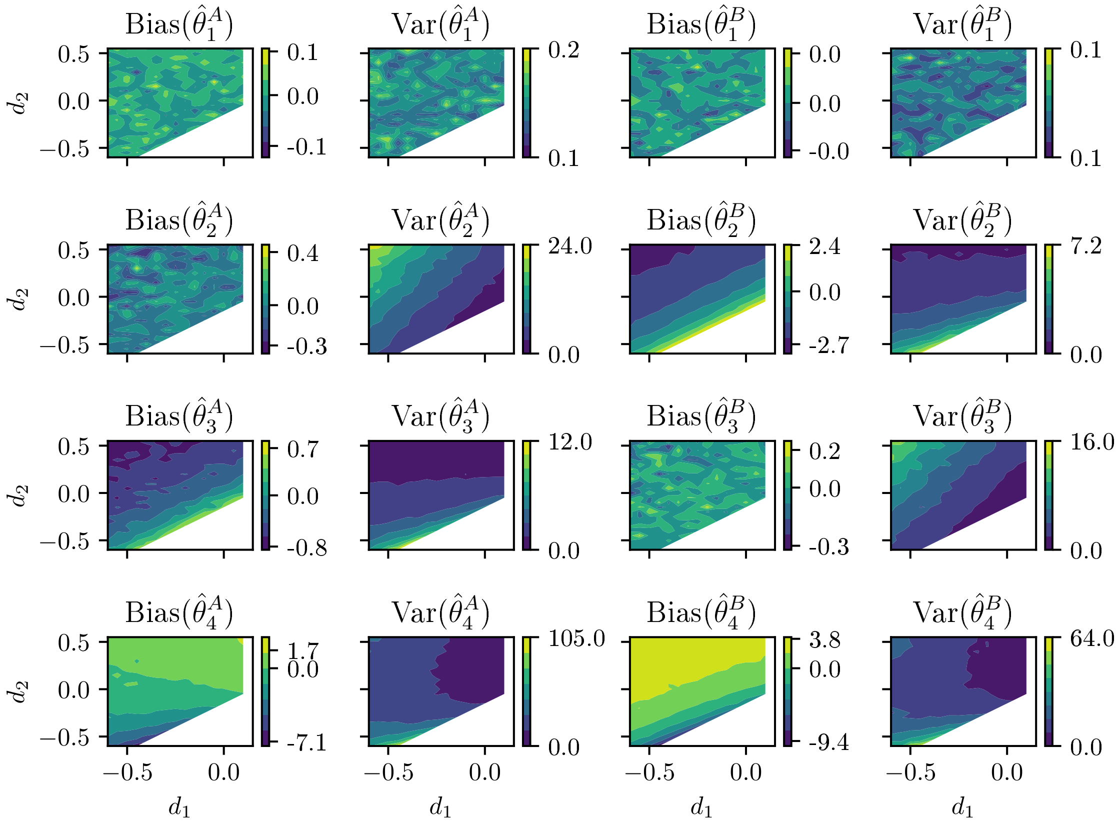
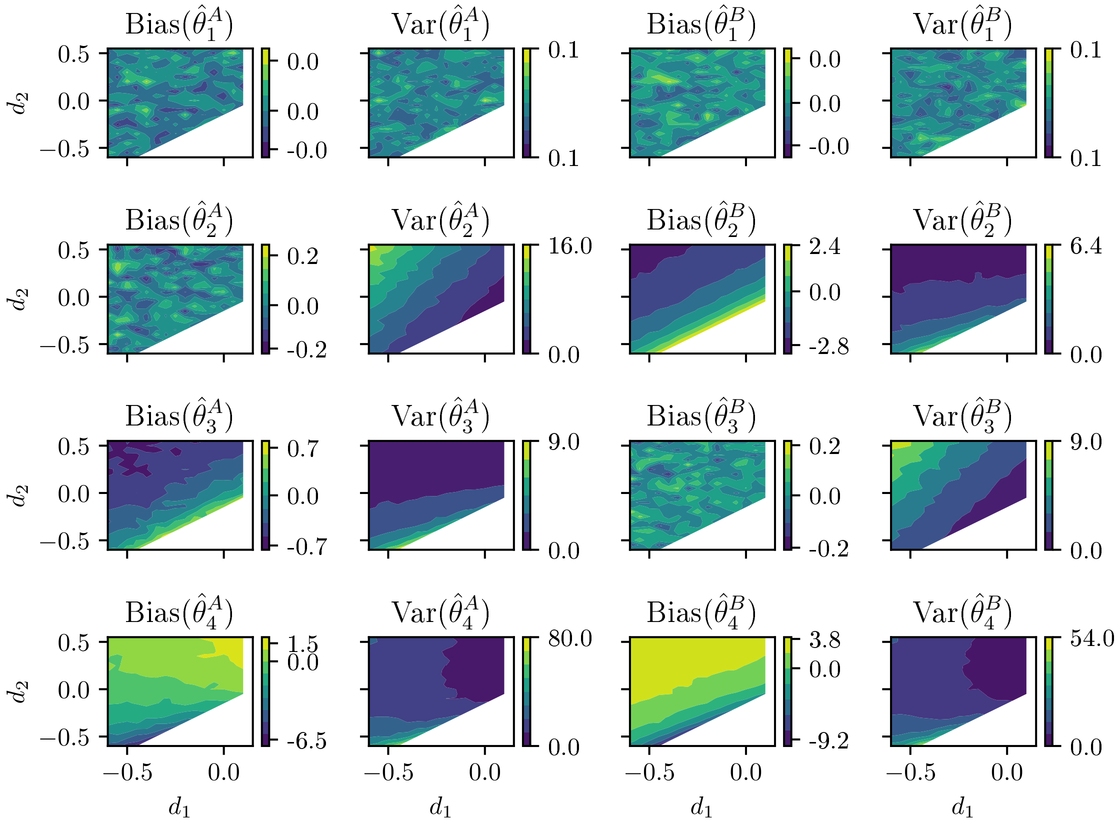


Figure 4: Monte Carlo results when symmetry does not hold, $n = 400$



5. Conclusion, Limitations and Further Research

In my thesis, I proposed an experimental design and two treatment effect identification strategies (based on instrumental variables) when there is interference within predefined pairs, made up by distinguishable members (e.g. healthy-ill members), and there is imperfect compliance, probably of endogenous nature, with treatment encouragement. As opposed to previous works addressing these two issues, my strategies allow for within-pair discussion and coordination of treatment participation. Both strategies are suitable for identifying the average treatment effects of two, potentially member-specific, binary treatments on two, potentially member-specific, outcomes belonging to the different pair members (or on one single outcome belonging to a pair) in complier subpopulations. Hence, we can identify an own effect and a partner effect for both members separately. Unfortunately, the joint effect of the two treatments is not identified.

One of the strategies throws away half of the sample but requires less assumptions. The other one, the main theoretical contribution of this thesis, uses the whole sample but comes at the high cost of strict symmetry assumptions on the members' participation willingness. Even though a Monte Carlo study suggests that the consistency of the proposed estimators in this second strategy is somewhat robust to minor violation of symmetry, this is hardly generalisable as the simulation results are conditional on specific, arbitrarily chosen distributions. Thus, extra care should be taken if the design is applied to real-life problems – the safe choice is obviously the half-sample strategy. Furthermore, I would recommend the use of bootstrap resampling to explore the properties of the estimators of the two strategies in the data at hand.

In addition to the symmetry assumptions, there are some drawbacks of my method. First, generalisation to more than two units in a group is hardly feasible due to the exponentially growing number of combinations. Second, working with distinguishable pair members has its advantages on one hand (member-specific treatments and outcomes), but is restrictive on the other because pairs with indistinguishable members require a different identification strategy. To the best of my kowledge, there is no such strategy which permits coordination of treatment participation, so this remains a subject of further research.

Last, but not least, future research topics could include inference, that is, to establish the asymptotic distribution of the estimators. Besides, one could consider blocking on pretreatment variables, i.e. to implement the Paired 2×2 Factorial Design within a block design, to reduce asymptotic variance.

References

Athey, S. and Imbens, G. (2016). The Econometrics of Randomized Experiments. *ArXiv e-prints*.

Barkley, B. G., Hudgens, M. G., Clemens, J. D., Ali, M., and Emch, M. E. (2017). Causal Inference from Observational Studies with Clustered Interference. *ArXiv e-prints*.

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

Cheng, C.-S. (2013). *Theory of Factorial Design: Single- and Multi-Stratum Experiments*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor and Francis, Hoboken, NJ.

Dai, B., Ding, S., and Wahba, G. (2012). Multivariate Bernoulli distribution. *ArXiv e-prints*.

Halloran, M. E. and Struchiner, C. J. (1991). Study Designs for Dependent Happenings. *Epidemiology*, 2(5):331–338.

Halloran, M. E. and Struchiner, C. J. (1995). Causal Inference in Infectious Diseases. *Epidemiology*, 6(2):142–151.

Hudgens, M. G. and Halloran, M. E. (2008). Toward Causal Inference with Interference. *Journal of the American Statistical Association*, 103(482):832–842.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

Kang, H. and Imbens, G. (2016). Peer Encouragement Designs in Causal Inference with Partial Interference and Identification of Local Average Network Effects. *ArXiv e-prints*.

Merrill, P. D. and McClure, L. A. (2015). Dichotomizing Partial Compliance and Increased Participant Burden in Factorial Designs: The Performance of Four Noncompliance Methods. 16:523.

Rosenbaum, P. R. (2007). Interference Between Units in Randomized Experiments. *Journal of the American Statistical Association*, 102(477):191–200.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5):688.

Rubin, D. B. (1980). Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Journal of the American Statistical Association*, 75(371):591–593.

Saul, B. and Hudgens, M. (2017). A Recipe for Inferference: Start with Causal Inference. Add Interference. Mix Well with R. *Journal of Statistical Software, Articles*, 82(2):1–21.

Sobel, M. E. (2006). What Do Randomized Studies of Housing Mobility Demonstrate? Causal Inference in the Face of Interference. *Journal of the American Statistical Association*, 101(476):1398–1407.

Tchetgen, E. J. T. and VanderWeele, T. J. (2012). On Causal Inference in the Presence of Interference. *Statistical Methods in Medical Research*, 21(1):55–75. PMID: 21068053.

VanderWeele, T. J., Tchetgen, E. J. T., and Halloran, M. E. (2014). Interference and Sensitivity Analysis. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, 29(4):687.

Appendices

A. Distribution of Treatment Participation

The purpose of examining the distribution of the treatment participation is threefold: to see what the Identifying Assumptions $A5-A8$ require from the distribution to exhibit (Appendix A.1); to derive an intermediate result from these which is later used in the proof of Theorem 2 (Appendix A.2); and to implement the Monte Carlo simulation in Section 4 (Appendix A.3). As previously, I use the abbreviations $D_A(10)$ for $D_A(Z_A = 1, Z_B = 0)$, $D_A(01)$ for $D_A(Z_A = 0, Z_B = 1)$, $D_B(10)$ for $D_B(Z_A = 1, Z_B = 0)$, $D_B(01)$ for $D_B(Z_A = 0, Z_B = 1)$ and so on. During the analysis, *A4 One sided-noncompliance is assumed to hold*, leading to the degenerate random variables $D_A(00) = D_A(01) = D_B(00) = D_B(10) = 0$. Hence the distribution boils down to that of $\boldsymbol{\mathcal{D}} \equiv [D_A(10), D_A(11), D_B(01), D_B(11)]'$. This is a 4-variate Bernoulli distribution supported on $\{0,1\}^4$ and can be characterised by 16 non-negative parameters (Dai et al., 2012), each of which indicates the joint prabability of a certain combination of zeros and ones:

$$
\begin{aligned}
p_{0000} &\equiv \mathbb{P}\left(D_A(10) = 0, D_A(11) = 0, D_B(01) = 0, D_B(11) = 0\right)\\
p_{1000} &\equiv \mathbb{P}\left(D_A(10) = 1, D_A(11) = 0, D_B(01) = 0, D_B(11) = 0\right)\\
p_{0100} &\equiv \mathbb{P}\left(D_A(10) = 0, D_A(11) = 1, D_B(01) = 0, D_B(11) = 0\right)\\
&\vdots\\
p_{0001} &\equiv \mathbb{P}\left(D_A(10) = 0, D_A(11) = 0, D_B(01) = 0, D_B(11) = 1\right)\\
p_{1111} &\equiv \mathbb{P}\left(D_A(10) = 1, D_A(11) = 1, D_B(01) = 1, D_B(11) = 1\right).
\end{aligned}
$$

Let p's be denoted by $\mathcal{P}$, then the probability mass function of $\boldsymbol{\mathcal{D}}$ is

$$
\begin{aligned}
f_{\mathcal{P}}(\boldsymbol{\mathcal{D}}) =& p_{1111}^{D_A(10)D_A(11)D_B(01)D_B(11)} p_{0111}^{[1-D_A(01)][D_A(11)D_B(01)D_B(11)]} \times \dots\\
&\times p_{0000}^{[1-D_A(10)][1-D_A(11)][1-D_B(01)][1-D_B(11)]}.
\end{aligned}
$$

A.1. Implications of Identifying Assumptions

A5 Monotonicity. Monotonicity requires $D_A(10) \leq D_A(11)$ and $D_B(01) \leq D_B(11)$. Hence whenever $D_A(10) > D_A(11)$ or $D_B(01) > D_B(11)$ we need $f_{\mathcal{P}}(\boldsymbol{\mathcal{D}}) = 0$ which is met by $p_{10..} = p_{..10} = 0$, or for more explicit form see Lemma 1.

Lemma 1 (Conditions for $A5$ Monotonicity)**.** *For Identifying Assumption $A5$ Monotonicity to hold the condition*

$$p_{1000} = p_{0010} = p_{0110} = p_{1010} = p_{1001} = p_{1011} = p_{1110} = 0.$$

is necessary and sufficient.

A6 Invertibility. Having clarified the implications of $A5$, let us turn our attention to $A6$ Invertibility which has $\mathbb{P}\left(D_A(11) = 1, D_A(10) = 1\right) \neq 0$ and $\mathbb{P}\left(D_B(11) = 1, D_B(01) = 1\right) \neq 0$. By the law of total probability and $A5$ Monotonicity we can express these in terms of $\mathcal{P}$:

$$\begin{aligned}
\mathbb{P}\left(D_A(11) = 1, D_A(10) = 1\right) &= \mathbb{P}\left(D_A(10) = 1, D_A(11) = 1, D_B(01) = 1, D_B(11) = 1\right) \\
&\quad + \mathbb{P}\left(D_A(10) = 1, D_A(11) = 1, D_B(01) = 0, D_B(11) = 1\right) \\
&\quad + \mathbb{P}\left(D_A(10) = 1, D_A(11) = 1, D_B(01) = 0, D_B(11) = 0\right) \\
&= p_{1111} + p_{1101} + p_{1100} \\
\mathbb{P}\left(D_B(11) = 1, D_B(01) = 1\right) &= \mathbb{P}\left(D_A(10) = 1, D_A(10) = 1, D_B(01) = 1, D_B(11) = 1\right) \\
&\quad + \mathbb{P}\left(D_A(10) = 0, D_A(11) = 1, D_B(01) = 1, D_B(11) = 1\right) \\
&\quad + \mathbb{P}\left(D_A(10) = 0, D_A(11) = 0, D_B(01) = 1, D_B(11) = 1\right) \\
&= p_{1111} + p_{0111} + p_{0011}.
\end{aligned}$$

Hence

$$\begin{aligned}
\mathbb{P}\left(D_A(11) = 1, D_A(10) = 1\right) \neq 0 &\iff p_{1111} + p_{1101} + p_{1100} \neq 0 \\
&\iff \min(\{p_{1111}, p_{1101}, p_{1100}\}) > 0 \\
\mathbb{P}\left(D_B(11) = 1, D_B(01) = 1\right) \neq 0 &\iff p_{1111} + p_{0111} + p_{0011} \neq 0 \\
&\iff \min(\{p_{1111}, p_{0111}, p_{0011}\}) > 0,
\end{aligned}$$

from which follows Lemma 2.

Lemma 2 (Conditions for *A6* Invertibility)**.** *For Identifying Assumptions $A6$ Invertibility to hold, assuming that $A5$ Monotonicity holds, the conditions*

$$\begin{aligned}
&\min(\{p_{1111}, p_{1101}, p_{1100}\}) > 0 \\
&\min(\{p_{1111}, p_{0111}, p_{0011}\}) > 0
\end{aligned}$$

are necessary and sufficient. This implies that the condition $p_{1111} > 0$ *is sufficient for* $A6$ *to hold.*

A7 Joint willingness. Joint willingness has $\mathbb{P}(D_A(11) = 1) = \mathbb{P}(D_B(11) = 1)$. We continue in the same fashion, expressing the assumption with $\mathcal{P}$. By the law of total probability and by $A5$ Monotonicity holding:

$$\begin{aligned}
\mathbb{P}(D_A(11) = 1) &= \mathbb{P}(D_A(10) = 1, D_A(11) = 1, D_B(01) = 1, D_B(11) = 1) \\
&+ \mathbb{P}(D_A(10) = 0, D_A(11) = 1, D_B(01) = 1, D_B(11) = 1) \\
&+ \mathbb{P}(D_A(10) = 1, D_A(11) = 1, D_B(01) = 0, D_B(11) = 1) \\
&+ \mathbb{P}(D_A(10) = 0, D_A(11) = 1, D_B(01) = 0, D_B(11) = 1) \\
&+ \mathbb{P}(D_A(10) = 1, D_A(11) = 1, D_B(01) = 0, D_B(11) = 0) \\
&+ \mathbb{P}(D_A(10) = 0, D_A(11) = 1, D_B(01) = 0, D_B(11) = 0) \\
&= p_{1111} + p_{0111} + p_{1101} + p_{0101} + p_{1100} + p_{0100} \\
\mathbb{P}(D_B(11) = 1) &= \mathbb{P}(D_A(10) = 1, D_A(11) = 1, D_B(01) = 1, D_B(11) = 1) \\
&+ \mathbb{P}(D_A(10) = 0, D_A(11) = 1, D_B(01) = 1, D_B(11) = 1) \\
&+ \mathbb{P}(D_A(10) = 1, D_A(11) = 1, D_B(01) = 0, D_B(11) = 1) \\
&+ \mathbb{P}(D_A(10) = 0, D_A(11) = 1, D_B(01) = 0, D_B(11) = 1) \\
&+ \mathbb{P}(D_A(10) = 0, D_A(11) = 0, D_B(01) = 1, D_B(11) = 1) \\
&+ \mathbb{P}(D_A(10) = 0, D_A(11) = 0, D_B(01) = 0, D_B(11) = 1) \\
&= p_{1111} + p_{0111} + p_{1101} + p_{0101} + p_{0011} + p_{0001},
\end{aligned}$$

which implies

$$\mathbb{P}(D_A(11) = 1) = \mathbb{P}(D_B(11) = 1) \iff$$
$$p_{1100} + p_{0100} = p_{0011} + p_{0001},$$

thus we get Lemma 3.

Lemma 3 (Conditions for $A7$ Joint willingness)**.** *For Identifying Assumptions* $A7$ *Joint willingness to hold, assuming that* $A5$ *Monotonicity holds, the condition*

$$p_{1100} + p_{0100} - (p_{0011} + p_{0001}) = 0$$

is necessary and sufficient.

A8 Spouse sensitivity. Spouse sensitivity has: $\mathbb{P}\left(D_A(10)=0, D_A(11)=1\right) = \mathbb{P}\left(D_B(01)=0, D_B(11)=1\right)$. Repeating our routine, by the law of total probability and $A5$:

$$\begin{aligned}
\mathbb{P}\left(D_A(10)=0, D_A(11)=1\right) &= \mathbb{P}\left(D_A(10)=0, D_A(11)=1, D_B(01)=1, D_B(11)=1\right) \\
&+ \mathbb{P}\left(D_A(10)=0, D_A(11)=1, D_B(01)=0, D_B(11)=1\right) \\
&+ \mathbb{P}\left(D_A(10)=0, D_A(11)=1, D_B(01)=0, D_B(11)=0\right) \\
&= p_{0111} + p_{0101} + p_{0100} \\
\mathbb{P}\left(D_j'(01)=0, D_B(11)=1\right) &= \mathbb{P}\left(D_A(10)=1, D_A(11)=1, D_B(01)=0, D_B(11)=1\right) \\
&+ \mathbb{P}\left(D_A(10)=0, D_A(11)=1, D_B(01)=0, D_B(11)=1\right) \\
&+ \mathbb{P}\left(D_A(10)=0, D_A(11)=0, D_B(01)=0, D_B(11)=1\right) \\
&= p_{1101} + p_{0101} + p_{0001},
\end{aligned}$$

which implies

$$\begin{aligned}
\mathbb{P}\left(D_A(10)=0, D_A(11)=1\right) = \mathbb{P}\left(D_j'(01)=0, D_B(11)=1\right) \iff \\
p_{0111} + p_{0100} = p_{1101} + p_{0001},
\end{aligned}$$

thus we get Lemma 4.

Lemma 4 (Conditions for $A8$ Spouse sensitivity)**.** *For Identifying Assumptions $A8$ Spouse sensitvity to hold, assuming that $A5$ Monotonicity holds, the condition*

$$p_{0111} + p_{0100} - (p_{1101} + p_{0001}) = 0$$

is necessary and sufficient.

A.2. Ramification of $A7 - A8$: a Useful Intermediate Result

The symmetry assumptions $A7 - A8$ have an important implication, which is used in the proof of Theorem 2 in Appendix E. The implication itself is a symmetry condition and is stated in Lemma 5.

Lemma 5 (Consequence of $A5$, $A7$, $A8$)**.** *If symmetry assumptions $A7 - A8$ and $A5$ Monotonicity holds, then it must be that*

$$\mathbb{P}\left(D_A(10)=1\right) = \mathbb{P}\left(D_B(01)=1\right).$$

Proof of Lemma 5. The proof is straightforward, we proceed in the the familiar way, expressing the probabilities with $\mathcal{P}$. By the law of total probability and $A5$ Monotonicity holding:

$$\begin{aligned}
\mathbb{P}\left(D_A(10)=1\right) &= \mathbb{P}\left(D_A(10)=1, D_A(11)=1, D_B(01)=1, D_B(11)=1\right) \\
&+ \mathbb{P}\left(D_A(10)=1, D_A(11)=1, D_B(01)=0, D_B(11)=1\right) \\
&+ \mathbb{P}\left(D_A(10)=1, D_A(11)=1, D_B(01)=0, D_B(11)=0\right) \\
&= p_{1111} + p_{1101} + p_{1100} \\
\mathbb{P}\left(D_B(01)=1\right) &= \mathbb{P}\left(D_A(10)=1, D_A(11)=1, D_B(01)=1, D_B(11)=1\right) \\
&+ \mathbb{P}\left(D_A(10)=0, D_A(11)=1, D_B(01)=1, D_B(11)=1\right) \\
&+ \mathbb{P}\left(D_A(10)=0, D_A(11)=0, D_B(01)=1, D_B(11)=1\right) \\
&= p_{1111} + p_{0111} + p_{0011},
\end{aligned}$$

so

$$\begin{aligned}
\mathbb{P}\left(D_A(10)=1\right) = \mathbb{P}\left(D_B(01)=1\right) &\iff \\
p_{1101} + p_{1100} = p_{0111} + p_{0011}.
\end{aligned}$$

We know from Lemmas 3 and 4 that whenever $A5$ holds, the necessary and sufficient conditions are

$$p_{1100} + p_{0100} - (p_{0011} + p_{0001}) = 0 \tag{11}$$

$$p_{0111} + p_{0100} - (p_{1101} + p_{0001}) = 0 \tag{12}$$

for $A7 - A8$ to hold. In other words, if $A5$ and $A7 - A8$ holds then these must be true. But if both hold then the necessary and sufficient condition for Lemma 5 is true, thus Lemma 5 must hold. This is seen by subsracting (12) from (11).

A.3. Constructing the Distribution of Treatment Participation for the Monte Carlo Study

Let us quickly review what is known so far about the distribution. The 4-variate Bernoulli distribution is fully specified by 16 parameters, $\mathcal{P}$, that are constrained by Identifying Assumptions $A5 - A8$. $A5$ Monotonicity requires 7 parameters to be zero (Lemma 1), and for $A6$ Invertibility $p_{1111} > 0$ is sufficient, and $A7 - A8$ impose two resrictions on the parameters. To assess the impacts of the violation of the symmetry assumptions, we want to construct a distribution which respects the requirements of $A5 - A6$ but allows for deliberate discrepancies in $A7 - A8$. In other words, we want to specify d_1 and d_2 in

$$\begin{aligned}
\mathbb{P}\left(D_A(11)=1\right)-\mathbb{P}\left(D_B(11)=1\right)&=d_1\\
\mathbb{P}\left(D_A(11)=1, D_A(10)=0\right)-\mathbb{P}\left(D_B(11)=1, D_B(01)=0\right)&=d_2.
\end{aligned}$$

A specification $d_1 \neq 0$ means that $A7$ is violated, and a specification $d_2 \neq 0$ means that $A8$ is violated. Based on the previous results of expressing these probabilities in terms of $\mathcal{P}$, we can write

$$\begin{aligned}
\mathbb{P}\left(D_A(11)=1\right)-\mathbb{P}\left(D_B(11)=1\right)&=p_{0001}-p_{0100}\\
&\quad -p_{1100}+p_{0011}\\
\mathbb{P}\left(D_A(11)=1, D_A(10)=0\right)-\mathbb{P}\left(D_B(11)=1, D_B(01)=0\right)&=p_{0001}-p_{0100}\\
&\quad +p_{1101}-p_{0111}
\end{aligned}$$

and then want to solve for p's given the d's:

$$\begin{bmatrix} -1 & 1 & 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} p_{0001} \\ p_{0100} \\ p_{1100} \\ p_{0011} \\ p_{1101} \\ p_{0111} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix},$$

from which

$$\begin{bmatrix} -1 & 1 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} p_{0001} \\ p_{0100} \\ p_{1100} \\ p_{0011} \\ p_{1101} \\ p_{0111} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 - d_1 \end{bmatrix}.$$

The system is underdetermined and there is no contradiction, thus we can have

$$p_{0001} = -d_1 + p_{0100} + p_{1100} - p_{0011} \tag{13}$$

$$p_{0111} = d_2 - d_1 + p_{1100} - p_{0011} + p_{1101}. \tag{14}$$

So given d_1, d_2, we can freely choose p_{0100}, p_{1100}, p_{0011} and p_{1101} subject to non-negativity contraints. Tthe relationships between the parameters, and what can be chosen freely and what is affected are summarised in Table 2.

Table 2: Distribution of treatment participation

Parameter	How is the parameter specified? 0 by $A5$?	Free to choose?	Determined by what?
p_{0000}		✓s.t. ≥ 0	
p_{1000}	Yes, 0		
p_{0100}		✓ s.t. ≥ 0	
p_{0010}	Yes, 0		
p_{0001}			✓ and d_1, d_2
p_{1100}[A]		✓✓ s.t. ≥ 0	
p_{0110}	Yes, 0		
p_{0011}[B]		✓✓ s.t. ≥ 0	
p_{1010}	Yes, 0		
p_{0101}		✓s.t. ≥ 0	
p_{1001}	Yes, 0		
p_{0111}[B]			✓ and d_1, d_2
p_{1011}[A]	Yes, 0		
p_{1101}[A]		✓ s.t. ≥ 0	
p_{1110}[A]	Yes, 0		
p_{1111}[A, B]		✓s.t. > 0	

[A] determines complier subpopulation for the treatment effects: $D_A(10) = 1$ and $D_A(11) = 1$
[B] determines complier subpopulation for the treatment effects: $D_B(01) = 1$ and $D_B(11) = 1$
[A, B] determines both above

This is complete in theory, but it provides only guidance in practice, as one has to make sure that the probability mass function adds up to one. To ensure that this is fulfilled, the discrepancies d_1, d_2 hold, and one can specify the free-to-choose probabilities in an intuitive way without having to worry about normalisation, I apply a multi-stage procedure.[12] In the first stage, the user specifies the discrepancies, d_1, d_2 and *weights* for the free-to-choose parameters. The non-negative weights are intuitive and what matters is their ratio. Then a quasi-normalisation is done with

[12] See the Python code for formal details.

the sum of weights on the free-to-choose parameters, taking into account their future role in determining p_{0001} and p_{0111}. Next, the two parameters, p_{0001} and p_{0111}, are computed as specified by Equations (13), (14) using the quasi-normalised weights from the previous step as the right-hand-side probabilities and using d_1, d_2. The final step is a normalisation, after which we obtain a valid p.m.f., where the ratios and the discrepancies specified in the first step approximately hold.

Following the steps specified above, we end up with a distribution which fulfills $A5$ and $A6$, and either fulfills both $A7$ and $A8$, or fails to fulfill one or both of them. In Appendix C, where the details of the Monte Carlo simulation are outlined in Algorithm 3, the former is denoted by $f_{\boldsymbol{\mathcal{D}},S}$, while the latter by $f_{\boldsymbol{\mathcal{D}},\neg S}$, where $\boldsymbol{\mathcal{D}} = [D_A(10), D_A(11), D_B(01), D_B(11)]'$.

B. Conditional Distribution of Potential Outcomes

For the Monte Carlo simulation it is necessary to specify the distribution of the random variables involved. The joint 4-variate distribution characterising the treatment participation is given in Appendix A, with emphasis on Monte Carlo in Appendix A.3. In this section, I specify the distribution of the potential outcomes conditional on the treatment participation. As noted in Section 4, the virtue of the conditional distribution is that it represents endogenous self-selection into the treatment. The conditional distribution is denoted by $f_{\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{D}}}(. \mid \boldsymbol{\Theta}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{\mathcal{D}}))$ 8-variate p.d.f., where $\boldsymbol{\mathcal{Y}} = [Y_A(00), Y_A(10), Y_A(01), Y_A(11), Y_B(00), Y_B(10), Y_B(01), Y_B(11)]'$ and $\boldsymbol{\mathcal{D}} = [D_A(10), D_A(11), D_B(01), D_B(11)]'$. To simplify analysis, I choose to work with a 8-variate normal distribution, the covariance matrix of which does not depend on the treatment participation, unlike its mean vector, so $\boldsymbol{\Theta}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{\mathcal{D}}) = \{\boldsymbol{\mu}(\boldsymbol{\mathcal{D}}), \boldsymbol{\Sigma}\}$.

As discussed in Section 4, the dependence of the expected potential outcomes on the treatment participation reflects certain behavioural patterns of the subjects. In Section 4, three principles are listed (higher outcome is better; laziness, lazy altruism toward spouse), which summarises the following exact rules on participation:

- Let improvement mean a zero or positive expected change in the potential outcome stemming from participation.
- Let decline mean a strictly negative expected change in the potential outcome stemming from participation.
- Let welfare-change mean the sum of the expected changes in the potential outcomes of both members stemming from participation.

- If no member has strict improvement, no member participates.
- If both members have strict improvement, both participate.
- Member-A (Member-B) does not participate if there is no strictly positive welfare-change and there is own decline.
- Member-A (Member-B) participates if there is no strictly positive welfare-change but there is own strictly postive improvement.
- Member-A participates if $|\text{decline}_A| <$ improvement$_B$.
- Member-B participates if $|\text{decline}_B| <$ improvement$_A$.

Then $\boldsymbol{\mu}(\boldsymbol{\mathcal{D}})$ is chosen to comply with these rules and is given in Table 3 for those values of $\boldsymbol{\mathcal{D}}$ which have non-zero probability by $A5$ Monotonicity.

What remains to be specified is the covariance matrix, $\boldsymbol{\Sigma}$. Before the simulation, I initialise a random positive semidefinite matrix in $\mathbb{R}^{8\times 8}$, where the covariance between any two potential outcomes is positive, which is likely the case in a lot of applications in social sciences. Then I use this matrix throughout every step of the simulation afterwards.

Table 3: Mean vector of potential outcomes conditional on the treatment particiption

$\boldsymbol{\mathcal{D}}' \backslash \boldsymbol{\mu}'$ of	$Y_A(00)$	$Y_A(10)$	$Y_A(01)$	$Y_A(11)$	$Y_B(00)$	$Y_B(10)$	$Y_B(01)$	$Y_B(11)$
$[0,0,0,0]$	0	0	0	0	1	1	1	1
$[0,1,0,0]$	0	0	0	1	1	1	1	0
$[0,0,0,1]$	0	0	0	-1	1	1	1	2
$[1,1,0,0]$	0	1	0	1	1	1	0	0
$[0,0,1,1]$	0	-1	0	-1	1	1	2	2
$[0,1,0,1]$	0	0	0	4	1	1	1	5
$[0,1,1,1]$	0	0	3	6	1	1	0	7
$[1,1,0,1]$	0	-1	0	6	1	3	1	7
$[1,1,1,1]$	0	5	6	8	1	4	8	10

C. Implementation of the Monte Carlo simulation

This section describes the technical details of the Monte Carlo simulation, described in Section 4, to explore the effects of the violation of symmetry assumptons $A7-A8$. The pseuso-code of implementation is in Algorithm 3. Pay attention to the order of arguments in the treatment paticipation and potential outcomes in the pseudo code.

Due to lack of space, no keyword arguments are possible. Hence, whenever letters A and B are used the first arguments always belong to member-A and the second to member-B. When the generic j, j' indexing is used the first argument always belongs to member-j and the second to member-j'.

Algorithm 3 Monte Carlo simulation

let $\circ$ denote the Hadamard, i.e. element-wise product
$n \leftarrow$ large sample size
$R \leftarrow$ Monte Carlo repetitions
construct $f_{\boldsymbol{\mathcal{D}},S}$ 4-variate p.m.f. such that it fulfils symm. assumptions $A7 - A8$
construct $f_{\boldsymbol{\mathcal{D}},\neg S}$ 4-variate p.m.f. such that it does not fulfil $A7 - A8$, where
 $\boldsymbol{\mathcal{D}} = [D_A(10), D_A(11), D_B(01), D_B(11)]'$
construct $f_{\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{D}}}(. \mid \boldsymbol{\Theta}_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{\mathcal{D}}))$ 8-variate p.d.f., where
 $\boldsymbol{\mathcal{Y}} = [Y_A(00), Y_A(10), Y_A(01), Y_A(11), Y_B(00), Y_B(10), Y_B(01), Y_B(11)]'$
 ▷ Draw a sample from DGP at each Monte Carlo repetition
for $r = 1 : R$ **do**
  ▷ Treatment assignment
  $\boldsymbol{z}_A, \boldsymbol{z}_B \leftarrow n \times 1$ vectors generated by Algorithm 1
  ▷ One-sided noncompliance
  $\boldsymbol{d}_A(00), \boldsymbol{d}_A(01), \boldsymbol{d}_B(00), \boldsymbol{d}_B(10) \leftarrow \mathbf{0}_{n\times 1}$ zero vectors
  ▷ DGP when symmetry holds (S) and does not hold ($\neg S$)
  for $s = S, \neg S$ **do**
   ▷ Treatment participation: potential
   $\boldsymbol{d}_{A,s}(10), \boldsymbol{d}_{A,s}(11), \boldsymbol{d}_{B,s}(01), \boldsymbol{d}_{B,s}(11) \leftarrow$ each row is i.i.d. draw from $f_{\boldsymbol{\mathcal{D}},s}$
   ▷ Treatment participation: observed
   $\boldsymbol{d}_{A,s} \leftarrow \boldsymbol{z}_A \circ \boldsymbol{z}_B \circ \boldsymbol{d}_{A,s}(11) + \boldsymbol{z}_A \circ (\mathbf{1} - \boldsymbol{z}_B) \circ \boldsymbol{d}_{A,s}(10)$
    $+ (\mathbf{1} - \boldsymbol{z}_A) \circ \boldsymbol{z}_B \circ \boldsymbol{d}_A(01) + (\mathbf{1} - \boldsymbol{z}_A) \circ (\mathbf{1} - \boldsymbol{z}_B) \circ \boldsymbol{d}_A(00)$
   $\boldsymbol{d}_{B,s} \leftarrow \boldsymbol{z}_A \circ \boldsymbol{z}_B \circ \boldsymbol{d}_{B,s}(11) + \boldsymbol{z}_A \circ (\mathbf{1} - \boldsymbol{z}_B) \circ \boldsymbol{d}_B(10)$
    $+ (\mathbf{1} - \boldsymbol{z}_A) \circ \boldsymbol{z}_B \circ \boldsymbol{d}_{B,S}(01) + (\mathbf{1} - \boldsymbol{z}_A) \circ (\mathbf{1} - \boldsymbol{z}_B) \circ \boldsymbol{d}_B(00)$
   ▷ Potential outcomes
   $\boldsymbol{y}_{A,s}(00), \boldsymbol{y}_{A,s}(10), \boldsymbol{y}_{A,s}(01), \boldsymbol{y}_{A,s}(11), \boldsymbol{y}_{B,s}(00), \boldsymbol{y}_{B,s}(10), \boldsymbol{y}_{B,s}(01), \boldsymbol{y}_{B,s}(11) \leftarrow$ each row, i, is i.i.d. draw from
   $f_{\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{D}}}(. \mid \boldsymbol{\Theta}_{\boldsymbol{\mathcal{Y}}}([d_{A,s,i}(10), d_{A,s,i}(11), d_{B,s,i}(01), d_{B,s,i}(11)]))$
  ▷ Estimators
  for $j = A, B$ **do** ▷ $j' = \{A, B\} \setminus \{j\}$
   ▷ Observed outcome
   $\boldsymbol{y}_{j,S} \leftarrow \boldsymbol{d}_{j,S} \circ \boldsymbol{d}_{j',S} \circ \boldsymbol{y}_{j,S}(11) + \boldsymbol{d}_{j,S} \circ (\mathbf{1} - \boldsymbol{d}_{j',S}) \circ \boldsymbol{y}_{j,S}$
    $+ (\mathbf{1} - \boldsymbol{d}_{j,S}) \circ \boldsymbol{d}_{j',S} \circ \boldsymbol{y}_{j,S}(01) + (\mathbf{1} - \boldsymbol{d}_{j,S}) \circ (\mathbf{1} - \boldsymbol{d}_{j',S}) \circ \boldsymbol{y}_{j,S}(00)$
   $\boldsymbol{y}_{j,\neg S} \leftarrow \boldsymbol{d}_{j,\neg S} \circ \boldsymbol{d}_{j',\neg S} \circ \boldsymbol{y}_{j,\neg S}(11) + \boldsymbol{d}_{j,\neg S} \circ (\mathbf{1} - \boldsymbol{d}_{j',\neg S}) \circ \boldsymbol{y}_{j,\neg S}$
    $+ (\mathbf{1} - \boldsymbol{d}_{j,\neg S}) \circ \boldsymbol{d}_{j',\neg S} \circ \boldsymbol{y}_{j,\neg S}(01) + (\mathbf{1} - \boldsymbol{d}_{j,\neg S}) \circ (\mathbf{1} - \boldsymbol{d}_{j',\neg S}) \circ \boldsymbol{y}_{j,\neg S}(00)$
   $\boldsymbol{D}_{j,S} \leftarrow [\mathbf{1}_{n\times 1}, \boldsymbol{d}_{j,S}, \boldsymbol{d}_{j',S}, \boldsymbol{d}_{j,S} \circ \boldsymbol{d}_{j',S}]$
   $\boldsymbol{D}_{j,\neg S} \leftarrow [\mathbf{1}_{n\times 1}, \boldsymbol{d}_{j,\neg S}, \boldsymbol{d}_{j',\neg S}, \boldsymbol{d}_{j,\neg S} \circ \boldsymbol{d}_{j',\neg S}]$
   $\boldsymbol{Z}_j \leftarrow [\mathbf{1}_{n\times 1}, \boldsymbol{z}_j, \boldsymbol{z}_{j'}, \boldsymbol{z}_j \circ \boldsymbol{z}_{j'}]$
   $\hat{\boldsymbol{\theta}}^n_{j,S}(r) \leftarrow (\boldsymbol{Z}_j' \boldsymbol{D}_{j,S})^{-1} \boldsymbol{Z}_j' \boldsymbol{y}_{j,S}$ ▷ consistent estimator under symm. ass.
   $\hat{\boldsymbol{\theta}}^n_{j,\neg S}(r) \leftarrow (\boldsymbol{Z}_j' \boldsymbol{D}_{j,\neg S})^{-1} \boldsymbol{Z}_j' \boldsymbol{y}_{j,\neg S}$ ▷ estimator when symm. ass. violated
 ▷ Check consistency: consistent for member-j if
$\mathbb{E}\left[\hat{\boldsymbol{\theta}}^n_{j,\neg S}\right] - \mathbb{E}\left[\hat{\boldsymbol{\theta}}^n_{j,S}\right] \approx R^{-1} \sum_{r=1}^{R} \hat{\boldsymbol{\theta}}^n_{j,\neg S}(r) - R^{-1} \sum_{r=1}^{R} \hat{\boldsymbol{\theta}}^n_{j,S}(r) \overset{n\to\infty}{\longrightarrow} \mathbf{0}_{4\times 1}$ and
$\mathbb{V}\left[\hat{\boldsymbol{\theta}}^n_{j,\neg S}\right] \approx R^{-1} \sum_{r=1}^{R} \hat{\boldsymbol{\theta}}^n_{j,\neg S}(r) \hat{\boldsymbol{\theta}}^n_{j,\neg S}(r)' - R^{-2} \left(\sum_{r=1}^{R} \hat{\boldsymbol{\theta}}^n_{j,\neg S}(r) \right) \left(\sum_{r=1}^{R} \hat{\boldsymbol{\theta}}^n_{j,\neg S}(r) \right)' \overset{n\to\infty}{\longrightarrow} \mathbf{0}_{4\times 4}$

D. Proof of Theorem 1

The proof of Theorem 1 directly follows from the results of Imbens and Angrist (1994). When using $\mathcal{G}_{10}$ and $\mathcal{G}_{00}$, we set $D(1) \equiv D_A(Z_A = 1, Z_B = 0)$, $D(0) \equiv D_A(Z_A = 0, Z_B = 0)$, and $Y(1) \equiv Y_j(D_A = 1, D_B = 0)$, $Y(0) \equiv Y_j(D_A = 0, D_B = 0)$ for $j \in \{A, B\}$. Directly applying the results of Imbens and Angrist (1994) gives the own effect for member-A (in the subpopulation of complier member-A's) and the partner effect for member-B (in the subpopulation with complier member-A partner):

$$\mathbb{E}\left[Y_A(D_A = 1, D_B = 0) - Y_A(D_A = 0, D_B = 0) \mid D_A(Z_A = 1, Z_B = 0) = 1\right]$$
$$\mathbb{E}\left[Y_B(D_A = 1, D_B = 0) - Y_B(D_A = 0, D_B = 0) \mid D_A(Z_A = 1, Z_B = 0) = 1\right].$$

Now under $A5$ Monotonicity the condition $D_A(Z_A = 1, Z_B = 0) = 1$ is equivalent to the condition $D_A(Z_A = 1, Z_B = 0) = 1$ and $D_A(Z_A = 1, Z_B = 1) = 1$.

When using $\mathcal{G}_{01}$ and $\mathcal{G}_{00}$, we set $D(1) \equiv D_B(Z_A = 0, Z_B = 1)$, $D(0) \equiv D_B(Z_A = 0, Z_B = 0)$, and $Y(1) \equiv Y_j(D_A = 0, D_B = 1)$, $Y(0) \equiv Y_j(D_A = 0, D_B = 0)$ for $j \in \{A, B\}$. Then we can immediately apply the proof of Imbens and Angrist (1994) again to obtain the own effect for member-B (in the subpopulation of complier member-B's) and the partner effect for member-A (in the subpopulation with complier member-B partner):

$$\mathbb{E}\left[Y_A(D_A = 0, D_B = 1) - Y_A(D_A = 0, D_B = 0) \mid D_B(Z_A = 0, Z_B = 1) = 1\right]$$
$$\mathbb{E}\left[Y_B(D_A = 0, D_B = 1) - Y_B(D_A = 0, D_B = 0) \mid D_B(Z_A = 0, Z_B = 1) = 1\right].$$

Now under $A5$ Monotonicity the condition $D_B(Z_A = 0, Z_B = 1) = 1$ is equivalent to the condition $D_B(Z_A = 0, Z_B = 1) = 1$ and $D_B(Z_A = 1, Z_B = 1) = 1$.

The results of Imbens and Angrist (1994) apply as $A5$ Monotonicity together with $A6$ Invertibility implies that $\mathbb{P}\left(D_A(10) = 1\right) \neq 0$ and $\mathbb{P}\left(D_B(01) = 1\right) \neq 0$. Thus the proof of Theorem 1 is complete ■

E. Proof of Theorem 2

E.1. Generic Notation

To make the proof of Theorem 2 generic for both member-A and member-B, I adopt a general j, j' indexing so that $j \in \{A, B\}$ and $j' \in \{A, B\} \setminus \{j\}$. That is, when j is fixed at A, j' stands for member-B, and when j represents member-B, j' denotes member-A.

Treatment encouragement. Z_{ji} denotes the member-j specific treatment encouragement and is 1 if and only if member-j in pair i is encouraged to take the member-j specific treatment. Similarly, $Z_{j'i}$ denotes the member-j' specific treatment encoragement in pair i and is 1 if and only if member-j' in pair i is encouraged to take the member-j' specific treatment. Suppose that $j \equiv A$, and member-A is the depressed person. Then $Z_{ji} = 1$ means that the depressed person in pair i is encouraged to take the treatment.

Treatment participation. Let $D_{ji}(Z_{ji}, Z_{ij'})$ denote the treatment participation of member-j when the treatment encouragement of member-j in pair i is Z_{ji}, while that of member-j' is $Z_{j'i}$. It is again important to pay attention to the order of the arguments: $D_{ji}(01)$ is the treatment participation of member-j, and $D_{j'i}(01)$ is the treatment participation of member-j' in pair i when member-j is not encouraged to take his/her treatment, while member-j' is encouraged (i.e. the first argument always stands for member-j). Then the observable treatment participation is written as

$$\begin{aligned} D_{ji} =& Z_{ji}Z_{j'i}D_{ji}(11) + Z_{ji}(1-Z_{j'i})D_{ji}(10) \\ &+ (1-Z_{ji})Z_{j'i}D_{ji}(01) + (1-Z_{ji})(1-Z_{j'i})D_{ji}(00) \qquad (15) \\ D_{j'i} =& Z_{ji}Z_{j'i}D_{j'i}(11) + Z_{ji}(1-Z_{j'i})D_{j'i}(10) \\ &+ (1-Z_{ji})Z_{j'i}D_{j'i}(01) + (1-Z_{ji})(1-Z_{j'i})D_{j'i}(00). \qquad (16) \end{aligned}$$

Suppose that $j \equiv A$, and member-A is the depressed person. Then by not giving the depressed person the antidepressant ($Z_{ij} = 0$) while encouraging his/her spouse to enroll to the educational program ($Z_{ij'} = 1$), we observe $D_{ij}(01)$ and $D_{ij'}(01)$ telling us whether the depressed member takes the pill and whether the non-depressed one enrolls to the program given the encouragement (no pill, education program).

Potential outcomes. There are four potential outcomes in pair i for member-j: $Y_{ji}(D_{ji}, D_{j'i}) \in \mathbb{R}^1$ and four for member-j': $Y_{j'i}(D_{ji}, D_{j'i}) \in \mathbb{R}^1$. Similarly to the treatment encouragment, the observable one of the four is written as

$$\begin{aligned} Y_{ji} =& D_{ji}D_{j'i}Y_{ji}(11) + D_{ji}(1-D_{j'i})Y_{ji}(10) \\ &+ (1-D_{ji})D_{j'i}Y_{ji}(01) + (1-D_{ji})(1-D_{j'i})Y_{ji}(00) \qquad (17) \\ Y_{j'i} =& D_{ji}D_{j'i}Y_{j'i}(11) + D_{ji}(1-D_{j'i})Y_{j'i}(10) \\ &+ (1-D_{ji})D_{j'i}Y_{j'i}(01) + (1-D_{ji})(1-D_{j'i})Y_{j'i}(00), \qquad (18) \end{aligned}$$

where the binary actual treatment participation $(D_{ji}, D_{j'i})$ selects the observable one. In our example ($j \equiv A =$ depressed member), $Y_{ji}(10)$ and $Y_{j'i}(10)$ are the outcomes of the pair members when the depressed actually takes the antidepressant and his/her spouse is not enrolled to the educational program; $Y_{ji}(11)$ and $Y_{j'i}(11)$ are the outcomes when the depressed takes the pill and the non-depressed in enrolled to the program and so on.

Variables in the sample. Let y_{ji} denote the sample analogue of the observable outcome for member-j; let $\boldsymbol{d}_{ji} \equiv [1, d_{ji}, d_{j'i}, d_{ji}d_{j'i}]' \in \{0,1\}^{4\times 1}$, where d_{ji} is the sample-realisation of treatment participation of member-j in pair i; and last let $\boldsymbol{z}_{ji} \equiv [1, z_{ji}, z_{j'i}, z_{ji}z_{j'i}]' \in \{0,1\}^{4\times 1}$, where z_{ji} is the sample-realisation of the treatment encouragement of member-j in pair i. Broadcasting them to $\boldsymbol{y}_j \equiv [y_{j1}, \ldots, y_{jn}]' \in \mathbb{R}^{n\times 1}$, $\boldsymbol{D}_j' \equiv [\boldsymbol{d}_{j1}, \ldots, \boldsymbol{d}_{jn}] \in \{0,1\}^{4\times n}$, $\boldsymbol{Z}_j' \equiv [\boldsymbol{z}_{j1}, \ldots, \boldsymbol{z}_{jn}] \in \{0,1\}^{4\times n}$ facilitates more compact notation.

E.2. Statement of Theorem 2

Theorem 2 states that the certain treatment effects are identified as plim $\hat{\boldsymbol{\theta}}_j$ if Identifying Assumptions are met, where the IV-based estimator is

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_j &\equiv \left(n^{-1}\boldsymbol{Z}_j'\boldsymbol{D}_j\right)^{-1} n^{-1}\boldsymbol{Z}_j'\boldsymbol{y}_j \\
&= \left(n^{-1}\sum_{i=1}^{n}\boldsymbol{z}_{ji}\boldsymbol{d}_{ji}'\right)^{-1} n^{-1}\sum_{i=1}^{n}\boldsymbol{z}_{ji}y_{ji} \\
\boldsymbol{d}_{ji} &\equiv [1, d_{ji}, d_{j'i}, d_{ji}d_{j'i}]' \in \{0,1\}^{4\times 1} \\
\boldsymbol{z}_{ji} &\equiv [1, z_{ji}, z_{j'i}, z_{ji}z_{j'i}]' \in \{0,1\}^{4\times 1}
\end{aligned}$$

for $j \in \{A, B\}$ and $j' \in \{A, B\} \setminus \{j\}$. To prove this, we examine plim $\hat{\boldsymbol{\theta}}_j$, i.e. the probability limit of $\hat{\boldsymbol{\theta}}_j$, or equivalently $\boldsymbol{\theta}_j : \lim_{n\to\infty} \mathbb{P}\left(||\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j||_2^2 > \varepsilon\right) = 0$ for any $\varepsilon > 0$. By the continuous mapping property of the probability limit and by the Weak

Law of Large Numbers for i.i.d. data (as the data across pairs are i.i.d.), we have

$$\begin{aligned}
\operatorname{plim} \hat{\boldsymbol{\theta}}_j &= \operatorname{plim} \left(n^{-1} \sum_{i=1}^{n} \boldsymbol{z}_{ji} \boldsymbol{d}'_{ji} \right)^{-1} n^{-1} \sum_{i=1}^{n} \boldsymbol{z}_{ji} y_{ji} \\
&= \left(\operatorname{plim} n^{-1} \sum_{i=1}^{n} \boldsymbol{z}_{ji} \boldsymbol{d}'_{ji} \right)^{-1} \operatorname{plim} n^{-1} \sum_{i=1}^{n} \boldsymbol{z}_{ji} y_{ji} \\
&= \mathbb{E}\left[\boldsymbol{z}_{ji} \boldsymbol{d}'_{ji} \right]^{-1} \mathbb{E}\left[\boldsymbol{z}_{ji} y_{ji} \right] \\
&= \mathbb{E}\left[\begin{bmatrix} 1 \\ Z_j \\ Z_{j'} \\ Z_j Z_{j'} \end{bmatrix} \begin{bmatrix} 1 & D_j & D_{j'} & D_j D_{j'} \end{bmatrix} \right]^{-1} \mathbb{E}\left[\begin{bmatrix} 1 \\ Z_j \\ Z_{j'} \\ Z_j Z_{j'} \end{bmatrix} Y_j \right] \\
&\equiv \boldsymbol{M}^{-1} \boldsymbol{v}.
\end{aligned} \tag{19}$$

Thus we need to compute the appropriate expectations and calculate the product. Though this might seem easy, were it not for the simplifying symmetry assumptions, the procedure would be unwieldy even when symbolic math packages are used, mainly due to the 4-by-4 matrix. The proof is organised into sections, so that in Appendix E.3 $\boldsymbol{M}$, and in Appendix E.4 $\boldsymbol{v}$ are computed, and then the product $\boldsymbol{M}^{-1}\boldsymbol{v}$ is evaluated and the results are summarised in Appendix E.6.

E.3. Computing $\boldsymbol{M}$

In this section, I compute $\boldsymbol{M}$, which is defined in (19). For a start, because of $A4$ One-sided noncompliance ($D_j(0, Z_{j'}) = D_{j'}(Z_j, 0) = 0 \forall Z_j, Z_{j'}$) we can rewrite the observable treatment participations in (20) and (22) as

$$\begin{aligned}
D_j =& Z_j Z_{j'} D_j(11) + Z_j(1 - Z_{j'}) D_j(10) \\
&+ (1 - Z_j) Z_{j'} D_j(01) + (1 - Z_j)(1 - Z_{j'}) D_j(00) \\
=& Z_j Z_{j'} [D_j(11) - D_j(10)] + Z_j D_j(10) \\
D_{j'} =& Z_j Z_{j'} D_{j'}(11) + Z_j(1 - Z_{j'}) D_{j'}(10) \\
&+ (1 - Z_j) Z_{j'} D_{j'}(01) + (1 - Z_j)(1 - Z_{j'}) D_{j'}(00) \\
=& Z_j Z_{j'} [D_{j'}(11) - D_{j'}(01)] + Z_{j'} D_{j'}(01).
\end{aligned}$$

Now we can focus on $\boldsymbol{M}$:

$$
\begin{aligned}
\boldsymbol{M} &= \mathbb{E}\left[\begin{bmatrix} 1 \\ Z_j \\ Z_{j'} \\ Z_j Z_{j'} \end{bmatrix} \begin{bmatrix} 1 & D_j & D_{j'} & D_j D_{j'} \end{bmatrix}\right] \\
&= \mathbb{E}\left[\begin{bmatrix} 1 & D_j & D_{j'} & D_j D_{j'} \\ Z_j & Z_j D_j & Z_j D_{j'} & Z_j D_j D_{j'} \\ Z_{j'} & Z_{j'} D_j & Z_{j'} D_{j'} & Z_{j'} D_j D_{j'} \\ Z_j Z_{j'} & Z_j Z_{j'} D_j & Z_j Z_{j'} D_{j'} & Z_j Z_{j'} D_j D_{j'} \end{bmatrix}\right] \\
&= \begin{bmatrix} 1 & \mathbb{E}[D_j] & \mathbb{E}[D_{j'}] & \mathbb{E}[D_j D_{j'}] \\ \mathbb{E}[Z_j] & \mathbb{E}[Z_j D_j] & \mathbb{E}[Z_j D_{j'}] & \mathbb{E}[Z_j D_j D_{j'}] \\ \mathbb{E}[Z_{j'}] & \mathbb{E}[Z_{j'} D_j] & \mathbb{E}[Z_{j'} D_{j'}] & \mathbb{E}[Z_{j'} D_j D_{j'}] \\ \mathbb{E}[Z_j Z_{j'}] & \mathbb{E}[Z_j Z_{j'} D_j] & \mathbb{E}[Z_j Z_{j'} D_{j'}] & \mathbb{E}[Z_j Z_{j'} D_j D_{j'}] \end{bmatrix}.
\end{aligned}
$$

In the following, these expectations are computed row-by-row, one-by-one. At each element I use: $A2$ Random assignment to factor the expectation of the product of the assigment(s) and participation(s) into the product of the expected assignment(s) and expected participation(s); $A3$ i.i.d. assignement to have $\mathbb{E}[Z_j Z_{j'} X] = \mathbb{E}[Z_j]\,\mathbb{E}[Z_{j'}]\,\mathbb{E}[X] = PP\,\mathbb{E}[X] = P^2\,\mathbb{E}[X]$ for any random variable X which is not Z_j or $Z_{j'}$. Also note that for any binary random variable W we have $\mathbb{E}[W] = \mathbb{P}(W=1)$ and that $W^2 = W$. So, let us proceed with the expectations.

Row of $\boldsymbol{M}$: 1st

$$
\begin{aligned}
\mathbb{E}[D_j] =& \mathbb{E}[Z_j Z_{j'}[D_j(11) - D_j(10)] + Z_j D_j(10)] \\
=& P^2\,\mathbb{E}[D_j(11) - D_j(10)] + P\,\mathbb{E}[D_j(10)] \\
=& P^2\,\mathbb{E}[D_j(11) - D_j(10)] + P\mathbb{P}(D_j(10) = 1) \quad A5 \implies \\
=& P^2[0\mathbb{P}(D_j(11) = 0, D_j(10) = 0) + 1\mathbb{P}(D_j(11) = 1, D_j(10) = 0) \\
& + 0\mathbb{P}(D_j(11) = 1, D_j(10) = 1)] + P\mathbb{P}(D_j(10) = 1) \\
=& P^2\mathbb{P}(D_j(11) = 1, D_j(10) = 0) + P\mathbb{P}(D_j(10) = 1). \qquad (20)
\end{aligned}
$$

Similarly for $D_{j'}$

$$
\begin{aligned}
\mathbb{E}[D_{j'}] &= \mathbb{E}[Z_j Z_{j'}[D_j(11) - D_j(01)] + Z_{j'} D_j(01)] \qquad (21) \\
&= P^2\mathbb{P}(D_{j'}(11) = 1, D_{j'}(01) = 0) + P\mathbb{P}(D_{j'}(01) = 1). \qquad (22)
\end{aligned}
$$

Then

$$
\begin{aligned}
D_j D_{j'} =&\{Z_j Z_{j'}[D_j(11)-D_j(10)]+Z_j D_j(10)\} \\
&\times \{Z_j Z_{j'}[D_{j'}(11)-D_{j'}(01)]+Z_{j'} D_{j'}(01)\} \\
=&Z_j^2 Z_{j'}^2[D_j(11)-D_j(10)][D_{j'}(11)-D_{j'}(01)]+ \\
&+Z_j Z_{j'}^2[D_j(11)-D_j(10)]D_{j'}(01) \\
&+Z_j^2 Z_{j'}[D_{j'}(11)-D_{j'}(01)]D_j(10) \\
&+Z_j Z_{j'} D_j(10) D_{j'}(01)
\end{aligned} \tag{23}
$$

$Z \in \{0,1\} \implies$

$$
\begin{aligned}
\mathbb{E}\left[D_j D_{j'}\right] = P^2\{&\,\mathbb{E}\left[(D_j(11)-D_j(10))(D_{j'}(11)-D_{j'}(01))\right] \\
&+\mathbb{E}\left[(D_j(11)-D_j(10))D_{j'}(01)\right] \\
&+\mathbb{E}\left[(D_{j'}(11)-D_{j'}(01))D_j(10)\right] \\
&+\mathbb{E}\left[D_j(10)D_{j'}(01)\right]\} \\
= P^2\{&\,\mathbb{E}\left[(D_j(11)-D_j(10))(D_{j'}(11)-D_{j'}(01)\right] \\
&+\mathbb{E}\left[D_j(11)D_{j'}(01)+D_{j'}(11)D_j(10)-D_{j'}(01)D_j(10)\right]\} \\
= P^2&\,\mathbb{E}\left[D_j(11)D_{j'}(11)\right] \\
= P^2&\mathbb{P}\left(D_j(11)=1, D_{j'}(11)=1\right).
\end{aligned}
$$

Row of M: 2nd

$$
\begin{aligned}
\mathbb{E}\left[Z_j D_j\right] &= \mathbb{E}\left[Z_j\{Z_j Z_{j'}[D_j(11)-D_j(10)]+Z_j D_j(10)\}\right] \\
&= \mathbb{E}\left[Z_j^2 Z_{j'}[D_j(11)-D_j(10)]+Z_j^2 D_j(10)\right] \quad Z_j \in \{0,1\} \implies \\
&= \mathbb{E}\left[Z_j Z_{j'}[D_j(11)-D_j(10)]+Z_j D_j(10)\right] \\
&= \mathbb{E}\left[D_j\right] \\
&= P^2\mathbb{P}\left(D_j(11)=1, D_j(10)=0\right)+P\mathbb{P}\left(D_j(10)=1\right).
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[Z_j D_{j'}\right] &= \mathbb{E}\left[Z_j\{Z_j Z_{j'}[D_{j'}(11)-D_{j'}(01)]+Z_{j'} D_{j'}(01)\}\right] \quad Z_j \in \{0,1\} \implies \\
&= \mathbb{E}\left[Z_j Z_{j'}[D_{j'}(11)-D_{j'}(01)]+Z_j Z_{j'} D_{j'}(01)\right] \\
&= \mathbb{E}\left[Z_j Z_{j'} D_{j'}(11)\right] \\
&= P^2\mathbb{P}\left(D_{j'}(11)=1\right).
\end{aligned}
$$

(23) and $Z_j \in \{0,1\} \implies$

$$\begin{aligned}\mathbb{E}\left[Z_j D_j D_{j'}\right] &= \mathbb{E}\left[D_j D_{j'}\right] \\ &= P^2 \mathbb{P}\left(D_j(11) = 1, D_{j'}(11) = 1\right).\end{aligned}$$

Row of M: 3rd

$$\begin{aligned}\mathbb{E}\left[Z_{j'} D_j\right] &= \mathbb{E}\left[Z_{j'}\{Z_j Z_{j'}[D_j(11) - D_j(10)] + Z_j D_j(10)\}\right] \\ &= \mathbb{E}\left[Z_{j'} Z_j [D_j(11) - D_j(10)] + Z_{j'} Z_j D_j(10)\right] \\ &= \mathbb{E}\left[Z_{j'} Z_j D_j(11)\right] \\ &= P^2 \mathbb{P}\left(D_j(11) = 1\right).\end{aligned}$$

(22) and $Z_{j'} \in \{0,1\} \implies$

$$\begin{aligned}\mathbb{E}\left[Z_{j'} D_{j'}\right] &= \mathbb{E}\left[D_{j'}\right] \\ &= P^2 \mathbb{P}\left(D_{j'}(11) = 1, D_{j'}(01) = 0\right) + P\mathbb{P}\left(D_{j'}(01) = 1\right).\end{aligned}$$

(23) and $Z_{j'} \in \{0,1\} \implies$

$$\begin{aligned}\mathbb{E}\left[Z_{j'} D_j D_{j'}\right] &= \mathbb{E}\left[D_j D_{j'}\right] \\ &= P^2 \mathbb{P}\left(D_j(11) = 1, D_{j'}(11) = 1\right).\end{aligned}$$

Row of M: 4th

(20) and $Z_j \in \{0,1\} \implies$

$$\begin{aligned}\mathbb{E}\left[Z_j Z_{j'} D_j\right] &= \mathbb{E}\left[Z_{j'} D_j\right] \\ &= P^2 \mathbb{P}\left(D_j(11) = 1\right).\end{aligned}$$

(22) and $Z_{j'} \in \{0,1\} \implies$

$$\begin{aligned}\mathbb{E}\left[Z_j Z_{j'} D_{j'}\right] &= \mathbb{E}\left[Z_j D_{j'}\right] \\ &= P^2 \mathbb{P}\left(D_{j'}(11) = 1\right).\end{aligned}$$

(23) and $Z_j, Z_{j'} \in \{0,1\} \implies$

$$\begin{aligned}\mathbb{E}\left[Z_j Z_{j'} D_j D_{j'}\right] &= \mathbb{E}\left[D_j D_{j'}\right] \\ &= P^2 \mathbb{P}\left(D_j(11)=1, D_{j'}(11)=1\right).\end{aligned}$$

Next, I invoke the symmetry assumptions and our intermediate result from Lemma 5, and introduce new notation for compactness:

$$\begin{aligned}
&q \equiv \mathbb{P}\left(D_j(11)=1\right) = \mathbb{P}\left(D_{j'}(11)=1\right) \quad \text{by } A7 \\
&r \equiv \mathbb{P}\left(D_j(11)=1, D_j(10)=0\right) = \mathbb{P}\left(D_{j'}(11)=1, D_{j'}(01)=0\right) \quad \text{by } A8 \\
&p \equiv \mathbb{P}\left(D_j(10)=1\right) = \mathbb{P}\left(D_{j'}(01)=1\right) \quad \text{by Lemma 5} \\
&\bar{q} \equiv \mathbb{P}\left(D_j(11)=1, D_{j'}(11)=1\right) \quad \text{new notation,}
\end{aligned}$$

so that we can compactly expresss the expectations of $\boldsymbol{M}$ row-by-row:

Row of $\boldsymbol{M}$: 1st

$$\begin{aligned}\mathbb{E}\left[D_j\right] &= P^2 r + Pp \\ \mathbb{E}\left[D_{j'}\right] &= P^2 r + Pp \\ \mathbb{E}\left[D_j D_{j'}\right] &= P^2 \bar{q}\end{aligned}$$

Row of $\boldsymbol{M}$: 2nd

$$\begin{aligned}\mathbb{E}\left[Z_j D_j\right] &= P^2 r + Pp \\ \mathbb{E}\left[Z_j D_{j'}\right] &= P^2 q \\ \mathbb{E}\left[Z_j D_j D_{j'}\right] &= P^2 \bar{q}\end{aligned}$$

Row of $\boldsymbol{M}$: 3rd

$$\begin{aligned}\mathbb{E}\left[Z_{j'} D_j\right] &= P^2 q \\ \mathbb{E}\left[Z_{j'} D_{j'}\right] &= P^2 r + Pp \\ \mathbb{E}\left[Z_{j'} D_j D_{j'}\right] &= P^2 \bar{q}\end{aligned}$$

Row of $\boldsymbol{M}$: 4th

$$\begin{aligned}
\mathbb{E}\left[Z_j Z_{j'} D_j\right] &= P^2 q \\
\mathbb{E}\left[Z_j Z_{j'} D_{j'}\right] &= P^2 q \\
\mathbb{E}\left[Z_j Z_{j'} D_j D_{j'}\right] &= P^2 \bar{q}.
\end{aligned}$$

For further simplification let

$$\begin{aligned}
a &\equiv P^2 r + P p \\
b &\equiv P^2 \bar{q} \\
c &\equiv P^2 q
\end{aligned}$$

so we can write

$$\boldsymbol{M} = \begin{bmatrix} 1 & a & a & b \\ P & a & c & b \\ P & c & a & b \\ P^2 & c & c & b \end{bmatrix}.$$

The letters a, b, c will not be used later on, their only purpose is to see how $\boldsymbol{M}$ can be written in the simplest form and thus to check the conditions for $\boldsymbol{M}^{-1}$ to exist.[13] So there exists $\boldsymbol{M}^{-1}$ if and only if

$$\begin{aligned}
a &\neq c \\
Pp &\neq P^2(q-r) \\
p &\neq P(q-r),
\end{aligned}$$

where by $A5$ Monotonicity

$$\begin{aligned}
q - r &= \mathbb{P}\left(D_j(11) = 1\right) - \mathbb{P}\left(D_j(11) = 1, D_j(10) = 0\right) \\
&= \mathbb{P}\left(D_j(11) = 1, D_j(10) = 1\right) \\
p &= \mathbb{P}\left(D_j(10) = 1\right) \\
&= \mathbb{P}\left(D_j(10) = 1, D_j(11) = 1\right),
\end{aligned}$$

[13] Another purpose is that the reader can follow more easily `symbolic_inverversion.py`.

so

$$
\begin{aligned}
p &\neq P(q-r) \\
\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right) &\neq P\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right) \\
\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right) &\neq 0.
\end{aligned}
$$

Note that this is where $A6$ Invertibility stems from.

E.4. Computing $\boldsymbol{v}$

Having computed $\boldsymbol{M}$ and checked the necessary and sufficient condition for $\boldsymbol{M}^{-1}$ to exist, now we can analyse $\boldsymbol{v}$ defined in (19):

$$
\begin{aligned}
\boldsymbol{v} &= \mathbb{E}\left[\begin{bmatrix} 1 \\ Z_j \\ Z_{j'} \\ Z_j Z_{j'} \end{bmatrix} Y_j\right] \\
&= \begin{bmatrix} \mathbb{E}\left[Y_j\right] \\ \mathbb{E}\left[Z_j Y_j\right] \\ \mathbb{E}\left[Z_{j'} Y_j\right] \\ \mathbb{E}\left[Z_j Z_{j'} Y_j\right] \end{bmatrix}
\end{aligned}
$$

This analysis is more tedious, but the procdure is the same: computing expectations one-by-one. Again, I use $A2$ Random assignment and $A3$ I.i.d. assignment throughout the computation. Before this, let us rewrite the potential outcomes in (17) as

$$
\begin{aligned}
Y_j =& D_j D_{j'} Y_j(11) + D_j(1-D_{j'})Y_j(10) \\
&+ (1-D_j)D_{j'}Y_j(01) + (1-D_j)(1-D_{j'})Y_j(00) \\
=& D_j D_{j'}[Y_j(11) - Y_j(01) - (Y_j(10) - Y_j(00))] \\
&+ D_j[Y_j(10) - Y_j(00)] + D_{j'}[Y_j(01) - Y_j(00)] + Y_j(00) \qquad (24)
\end{aligned}
$$

and introduce $\Delta_j(D_j, D_{j'}|\tilde{D_j}, \tilde{D_{j'}}) \equiv Y_j(D_j, D_{j'}) - Y_j(\tilde{D_j}, \tilde{D_{j'}})$. Then we can proceed with the expectations one by one.

Row of v: 1st, $\mathbb{E}[Y_j]$. (24) $\implies$

$$\begin{aligned}
\mathbb{E}[Y_j] =& \mathbb{E}[D_j D_{j'}\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j\Delta_j(10|00) + D_{j'}\Delta_j(01|00) + Y_j(00)] \\
=& P^2\,\mathbb{E}[D_j(11)D_{j'}(11)\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j(11)\Delta_j(10|00) + D_{j'}(11)\Delta_j(01|00) + Y_j(00)] \\
&+ P(1-P)\,\mathbb{E}[D_j(10)D_{j'}(10)\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j(10)\Delta_j(10|00) + D_{j'}(10)\Delta_j(01|00) + Y_j(00)] \\
&+ (1-P)P\,\mathbb{E}[D_j(01)D_{j'}(01)\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j(01)\Delta_j(10|00) + D_{j'}(01)\Delta_j(01|00) + Y_j(00)] \\
&+ (1-P)^2\,\mathbb{E}[D_j(00)D_{j'}(00)\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j(00)\Delta_j(10|00) + D_{j'}(00)\Delta_j(01|00) + Y_j(00)] \qquad (25)
\end{aligned}$$

$$(26)$$

Now by $A4$ One-sided noncompliance ($D_j(0, Z_{j'}) = D_{j'}(Z_j, 0) = 0 \forall Z_j, Z_{j'}$)

$$\begin{aligned}
\mathbb{E}[Y_j] =& P^2 \underbrace{\mathbb{E}[D_j(11)D_{j'}(11)\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j(11)\Delta_j(10|00) + D_{j'}(11)\Delta_j(01|00) + Y_j(00)]}_{\equiv\alpha} \\
&+ P(1-P)\underbrace{\mathbb{E}[D_j(10)\Delta_j(10|00) + Y_j(00)]}_{\equiv\beta} \\
&+ (1-P)P\underbrace{\mathbb{E}[D_{j'}(01)\Delta_j(01|00) + Y_j(00)]}_{\equiv\gamma} \\
&+ (1-P)^2\underbrace{\mathbb{E}[Y_j(00)]}_{\equiv\psi},
\end{aligned}$$

so

$$\begin{aligned}
\mathbb{E}\left[Y_j\right] =& P^2\alpha + P(1-P)\beta + (1-P)P\gamma + (1-P)^2\psi \\
=& P^2\alpha + P\beta - P^2\beta + P\gamma - P^2\gamma + (1-P)\psi - (P-P^2)\psi \\
=& P^2(\alpha - (\beta+\gamma)) + P(\beta+\gamma) + (1-P)\psi - (P-P^2)\psi \\
=& P^2(\alpha - (\beta+\gamma-\psi)) + P(\beta+\gamma-2\psi) + \psi.
\end{aligned}$$

What we want to take advantage of now is $A5$ Monotonicity, which is best done by expressing as many terms as we can with $D_j(11) - D_j(10)$ and $D_{j'}(11) - D_{j'}(01)$. This is why the greek letters are introduced in the first place. Let us see how it works:

$$\begin{aligned}
\beta+\gamma-\psi &= \mathbb{E}\left[D_j(10)\Delta_j(10|00) + Y_j(00)\right] + \mathbb{E}\left[D_{j'}(01)\Delta_j(01|00) + Y_j(00)\right] - \mathbb{E}\left[Y_j(00)\right] \\
&= \mathbb{E}\left[D_j(10)\Delta_j(10|00) + D_{j'}(01)\Delta_j(01|00) + Y_j(00)\right] \qquad (27) \\
\alpha - (\beta+\gamma-\psi) &= \mathbb{E}\left[D_j(11)D_{j'}(11)\{\Delta_j(11|01) - \Delta_j(10|00)\} + (D_j(11) - D_j(10))\Delta_j(10|00) + (D_{j'}(11) - D_{j'}(01))\Delta_j(01|00)\right],
\end{aligned}$$

which simplifies by $A5$ Monotonicity as follows

$$\begin{aligned}
\mathbb{E}\left[D_j(11)D_{j'}(11)\{\Delta_j(11|01) - \Delta_j(10|00)\}\right] &= \mathbb{E}\left[\Delta_j(11|01) - \Delta_j(10|00) \mid D_j(11)=1, D_{j'}(11)=1\right]\mathbb{P}\left(D_j(11)=1, D_{j'}(11)=1\right) \\
\mathbb{E}\left[(D_j(11) - D_j(10))\Delta_j(10|00)\right] &= \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right]\mathbb{P}\left(D_j(11)=1, D_j(10)=0\right) \\
\mathbb{E}\left[(D_{j'}(11) - D_{j'}(01))\Delta_j(01|00)\right] &= \mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11)=1, D_{j'}(10)=0\right]\mathbb{P}\left(D_{j'}(11)=1, D_{j'}(01)=0\right),
\end{aligned}$$

thus

$$\begin{aligned}\alpha-(\beta+\gamma-\psi)=\,&\mathbb{E}\left[\Delta_j(11|01)-\Delta_j(10|00)\mid D_j(11)=1,D_{j'}(11)=1\right]\mathbb{P}\left(D_j(11)=1,D_{j'}(11)=1\right)\\&+\mathbb{E}\left[\Delta_j(10|00)\mid D_j(11)=1,D_j(10)=0\right]\mathbb{P}\left(D_j(11)=1,D_j(10)=0\right)\\&+\mathbb{E}\left[\Delta_j(01|00)\mid D_{j'}(11)=1,D_{j'}(10)=0\right]\mathbb{P}\left(D_{j'}(11)=1,D_{j'}(01)=0\right).\end{aligned}$$

Next, by (27) and the definition of $\psi=\mathbb{E}\left[Y_j(00)\right]$,

$$\begin{aligned}\beta+\gamma-2\psi&=\mathbb{E}\left[D_j(10)\Delta_j(10|00)+D_{j'}(01)\Delta_j(01|00)\right]\\&=\mathbb{E}\left[\Delta_j(10|00)\mid D_j(10)=1\right]\mathbb{P}\left(D_j(10)=1\right)+\mathbb{E}\left[\Delta_j(01|00)\mid D_{j'}(01)=1\right]\mathbb{P}\left(D_{j'}(01)=1\right).\end{aligned}$$

Putting together the greek-letter expressions, we finally obtain the first row of $\boldsymbol{v}$:

$$\begin{aligned}\mathbb{E}\left[Y_j\right]&=P^2(\alpha-(\beta+\gamma-\psi))+P(\beta+\gamma-2\psi)+\psi\\&=P^2\{\mathbb{E}\left[\Delta_j(11|01)-\Delta_j(10|00)\mid D_j(11)=1,D_{j'}(11)=1\right]\mathbb{P}\left(D_j(11)=1,D_{j'}(11)=1\right)\\&\qquad+\mathbb{E}\left[\Delta_j(10|00)\mid D_j(11)=1,D_j(10)=0\right]\mathbb{P}\left(D_j(11)=1,D_j(10)=0\right)\\&\qquad+\mathbb{E}\left[\Delta_j(01|00)\mid D_{j'}(11)=1,D_{j'}(10)=0\right]\mathbb{P}\left(D_{j'}(11)=1,D_{j'}(01)=0\right)\}\\&\quad+P\{\mathbb{E}\left[\Delta_j(10|00)\mid D_j(10)=1\right]\mathbb{P}\left(D_j(10)=1\right)+\mathbb{E}\left[\Delta_j(01|00)\mid D_{j'}(01)=1\right]\mathbb{P}\left(D_{j'}(01)=1\right)\}\\&\quad+\mathbb{E}\left[Y_j(00)\right].\end{aligned}$$

Row of $\boldsymbol{v}$: 2nd, $\mathbb{E}\left[Z_j Y_j\right]$. (24) and$A4$ One-sided noncompliance $\implies$

$$
\begin{aligned}
\mathbb{E}\left[Z_j Y_j\right] =& \mathbb{E}\left[Z_j\{D_j D_{j'}\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j\Delta_j(10|00) + D_{j'}\Delta_j(01|00) + Y_j(00)\}\right] \\
=& P^2\,\mathbb{E}\left[D_j(11)D_{j'}(11)\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j(11)\Delta_j(10|00) + D_{j'}(11)\Delta_j(01|00) + Y_j(00)\right] \\
&+ P(1-P)\,\mathbb{E}\left[D_j(10)\Delta_j(10|00) + Y_j(00)\right] \\
=& P^2\,\mathbb{E}\left[D_j(11)D_{j'}(11)\{\Delta_j(11|01) - \Delta_j(10|00)\} + (D_j(11) - D_j(10))\Delta_j(10|00) + D_{j'}(11)\Delta_j(01|00)\right] \\
&+ P\,\mathbb{E}\left[D_j(10)\Delta_j(10|00) + Y_j(00)\right].
\end{aligned} \tag{28}
$$

Note how the last two terms in (25) are not present in (28) because of $A4$ One-sided noncompliance. Next, as we did for $\mathbb{E}\left[Y_j\right]$, using $A5$ Monotonicity:

$$
\begin{aligned}
\mathbb{E}\left[Z_j Y_j\right] = P^2\{&\mathbb{E}\left[\Delta_j(11|01) - \Delta_j(10|00) \mid D_j(11) = 1, D_{j'}(11) = 1\right]\mathbb{P}\left(D_j(11) = 1, D_{j'}(11) = 1\right) \\
&+ \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 0\right]\mathbb{P}\left(D_j(11) = 1, D_j(10) = 0\right) \\
&+ \mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1\right]\mathbb{P}\left(D_{j'}(11) = 1\right)\} \\
&+ P\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10) = 1\right]\mathbb{P}\left(D_j(10) = 1\right) \\
&+ P\,\mathbb{E}\left[Y_j(00)\right].
\end{aligned}
$$

Row of v: 3rd, $\mathbb{E}\left[Z_{j'}Y_j\right]$. (24) and $A4$ One-sided noncompliance $\implies$

$$\begin{aligned}
\mathbb{E}\left[Z_{j'}Y_j\right] =& \mathbb{E}\left[Z_{j'}\{D_jD_{j'}\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j\Delta_j(10|00) + D_{j'}\Delta_j(01|00) + Y_j(00)\}\right] \\
=& P^2\,\mathbb{E}\left[D_j(11)D_{j'}(11)\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j(11)\Delta_j(10|00) + D_{j'}(11)\Delta_j(01|00) + Y_j(00)\right] \\
& + (1-P)P\,\mathbb{E}\left[D_{j'}(01)\Delta_j(01|00)\right] \qquad (29) \\
=& P^2\,\mathbb{E}\left[D_j(11)D_{j'}(11)\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j(11)\Delta_j(10|00) + (D_{j'}(11) - D_{j'}(01))\Delta_j(01|00)\right] \\
& + P\,\mathbb{E}\left[D_{j'}(01)\Delta_j(01|00) + Y_j(00)\right]
\end{aligned}$$

Note again how the two terms in (25) are not present in (29) because of $A4$ One-sided noncompliance. Next, as we did for $\mathbb{E}\left[Y_j\right]$, using $A5$ Monotonicity:

$$\begin{aligned}
\mathbb{E}\left[Z_{j'}Y_j\right] = P^2\{& \mathbb{E}\left[\Delta_j(11|01) - \Delta_j(10|00) \mid D_j(11) = 1, D_{j'}(11) = 1\right]\mathbb{P}\left(D_j(11) = 1, D_{j'}(11) = 1\right) \\
& + \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1\right]\mathbb{P}\left(D_j(11) = 1\right) \\
& + \mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1, D_{j'}(01) = 0\right]\mathbb{P}\left(D_{j'}(11) = 1, D_{j'}(01) = 0\right)\} \\
& + P\,\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(01) = 1\right]\mathbb{P}\left(D_{j'}(01) = 1\right) \\
& + P\,\mathbb{E}\left[Y_j(00)\right].
\end{aligned}$$

Row of v: 4th, $\mathbb{E}\left[Z_jZ_{j'}Y_j\right]$. (24) and $A4$ One-sided noncompliance $\implies$

$$\begin{aligned}
\mathbb{E}\left[Z_jZ_{j'}Y_j\right] =& \mathbb{E}\left[Z_jZ_{j'}\{D_jD_{j'}\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j\Delta_j(10|00) + D_{j'}\Delta_j(01|00) + Y_j(00)\}\right] \\
=& P^2\,\mathbb{E}\left[D_j(11)D_{j'}(11)\{\Delta_j(11|01) - \Delta_j(10|00)\} + D_j(11)\Delta_j(10|00) + D_{j'}(11)\Delta_j(01|00) + Y_j(00)\right]
\end{aligned}$$

Doing what has been done with the three previous rows:

$$\begin{aligned}\mathbb{E}\left[Z_j Z_{j'} Y_j\right] = P^2\{&\mathbb{E}\left[\Delta_j(11|01) - \Delta_j(10|00) \mid D_j(11) = 1, D_{j'}(11) = 1\right]\mathbb{P}\left(D_j(11) = 1, D_{j'}(11) = 1\right)\\ &+ \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1\right]\mathbb{P}\left(D_j(11) = 1\right)\\ &+ \mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1, D_{j'}(01) = 0\right]\mathbb{P}\left(D_{j'}(11) = 1\right)\}\\ +&P^2\,\mathbb{E}\left[Y_j(00)\right].\end{aligned}$$

Before $\boldsymbol{M}^{-1}\boldsymbol{v}$ is computed in Appendix E.5, I introduce new notation for further simplification:

$$\begin{aligned}e_1 &\equiv \mathbb{E}\left[\Delta_j(11|01) - \Delta_j(10|00) \mid D_j(11) = 1, D_{j'}(11) = 1\right]\\ e_2 &\equiv \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 0\right]\\ e_3 &\equiv \mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1, D_{j'}(01) = 0\right]\\ e_4 &\equiv \mathbb{E}\left[\Delta_j(10|00) \mid D_j(10) = 1\right]\\ e_5 &\equiv \mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(01) = 1\right]\\ e_6 &\equiv \mathbb{E}\left[Y_j(00)\right]\\ e_7 &\equiv \mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1\right]\\ e_8 &\equiv \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1\right].\end{aligned}$$

Recalling the notation in Appendix E.3,

$$\begin{aligned}
&q \equiv \mathbb{P}\left(D_j(11)=1\right)=\mathbb{P}\left(D_{j'}(11)=1\right) \\
&r \equiv \mathbb{P}\left(D_j(11)=1, D_j(10)=0\right)=\mathbb{P}\left(D_{j'}(11)=1, D_{j'}(01)=0\right) \\
&p \equiv \mathbb{P}\left(D_j(10)=1\right)=\mathbb{P}\left(D_{j'}(01)=1\right) \\
&\bar{q} \equiv \mathbb{P}\left(D_j(11)=1, D_{j'}(11)=1\right),
\end{aligned}$$

we can now express $\boldsymbol{v}$ as

$$\boldsymbol{v}=\begin{bmatrix} \mathbb{E}\left[Y_j\right] \\ \mathbb{E}\left[Z_j Y_j\right] \\ \mathbb{E}\left[Z_{j'} Y_j\right] \\ \mathbb{E}\left[Z_j Z_{j'} Y_j\right] \end{bmatrix}=\begin{bmatrix} P^2\{e_1\bar{q}+e_2 r+e_3 r\}+Pe_4 p+Pe_5 p+e_6 \\ P^2\{e_1\bar{q}+e_2 r+e_7 q\}+Pe_4 p+Pe_6 \\ P^2\{e_1\bar{q}+e_8 q+e_3 r\}+Pe_5 p+Pe_6 \\ P^2\{e_1\bar{q}+e_8 q+e_7 q\}+P^2 e_6 \end{bmatrix}. \tag{30}$$

E.5. Putting the Pieces Together: Computing $\boldsymbol{M}^{-1}\boldsymbol{v}$

Having computed $\boldsymbol{M}$ and $\boldsymbol{v}$ in Appendix E.3 and E.4, the final step in the proof of Theorem 2 is evaluating $\boldsymbol{M}^{-1}\boldsymbol{v} \in \mathbb{R}^{4\times 1}$. In this 4-long column vector, each row is a linear combination of the elements of $\boldsymbol{v}$, where the weigths are the elements in the corresponding rows of $\boldsymbol{M}^{-1}$. *A*6 Invertibility ensures that $\boldsymbol{M}^{-1}$ exists (see the end of E.3). Suppose that

$$\boldsymbol{M}^{-1} = \begin{bmatrix} m_1^{(1)} & m_2^{(1)} & m_3^{(1)} & m_4^{(1)} \\ m_1^{(2)} & m_2^{(2)} & m_3^{(2)} & m_4^{(2)} \\ m_1^{(3)} & m_2^{(3)} & m_3^{(3)} & m_4^{(3)} \\ m_1^{(4)} & m_2^{(4)} & m_3^{(4)} & m_4^{(4)} \end{bmatrix},$$

then the rth row for $r = 1, 2, 3, 4$ in $\boldsymbol{M}^{-1}\boldsymbol{v}$, $(\boldsymbol{M}^{-1}\boldsymbol{v})_r$ is given by

$$(\boldsymbol{M}^{-1}\boldsymbol{v})_r = m_1^{(r)}v_1 + m_2^{(r)}v_2 + m_3^{(r)}v_3 + m_4^{(r)}v_4.$$

If we substitute in the results from (30) we obtain

$$\begin{aligned}(\boldsymbol{M}^{-1}\boldsymbol{v})_r =& m_1^{(r)}[P^2\{e_1\bar{q} + e_2r + e_3r\} + Pe_4p + Pe_5p + e_6] \\ &+ m_2^{(r)}[P^2\{e_1\bar{q} + e_2r + e_7q\} + Pe_4p + Pe_6] \\ &+ m_3^{(r)}[P^2\{e_1\bar{q} + e_8q + e_3r\} + Pe_5p + Pe_6] \\ &+ m_4^{(r)}[P^2\{e_1\bar{q} + e_8q + e_7q\} + P^2e_6],\end{aligned}$$

and after collecting terms

$$
\begin{aligned}
(\boldsymbol{M}^{-1}\boldsymbol{v})_r =& e_1[P^2\bar{q}\{m_1^{(r)} + m_2^{(r)} + m_3^{(r)} + m_4^{(r)}\}] \\
&+ e_2[P^2 r(m_1^{(r)} + m_2^{(r)})] \\
&+ e_3[P^2 r(m_1^{(r)} + m_3^{(r)})] \\
&+ e_4[Pp(m_1^{(r)} + m_2^{(r)})] \\
&+ e_5[Pp(m_1^{(r)} + m_3^{(r)})] \\
&+ e_6[m_1^{(r)} + Pm_2^{(r)} + Pm_3^{(r)} + P^2 m_4^{(r)}] \\
&+ e_7[P^2 q(m_2^{(r)} + m_4^{(r)})] \\
&+ e_8[P^2 q(m_3^{(r)} + m_4^{(r)})] \\
\equiv& \sum_{k=1}^{8} e_k w_k^{(r)}.
\end{aligned}
$$

At this point, to compute w's, it is convenient to rely on a symbolic math package, `SymPy`. The code in `symbolic_inversion.py` computes $\{w_k^{(r)}\}_{k=1}^8$ for $r = 1, 2, 3, 4$, and the values are shown in Table 4.

Table 4: Output from `symbolic_inversion.py`

$\{w_k^{(r)}\}_{k=1}^8\backslash$Row	1	2	3	4
$w_1^{(r)}$	0	0	0	1
$w_2^{(r)}$	0	$\frac{Pr}{-Pq+Pr+p}$	0	$\frac{-Pqr}{\bar{q}(-Pq+Pr+p)}$
$w_3^{(r)}$	0	0	$\frac{Pr}{-Pq+Pr+p}$	$\frac{-Pqr}{\bar{q}(-Pq+Pr+p)}$
$w_4^{(r)}$	0	$\frac{p}{-Pq+Pr+p}$	0	$\frac{-pq}{\bar{q}(-Pq+Pr+p)}$
$w_5^{(r)}$	0	0	$\frac{p}{-Pq+Pr+p}$	$\frac{-pq}{\bar{q}(-Pq+Pr+p)}$
$w_6^{(r)}$	1	0	0	0
$w_7^{(r)}$	0	0	$\frac{-Pq}{-Pq+Pr+p}$	$\frac{q(Pr+p)}{\bar{q}(-Pq+Pr+p)}$
$w_8^{(r)}$	0	$\frac{-Pq}{-Pq+Pr+p}$	0	$\frac{q(Pr+p)}{\bar{q}(-Pq+Pr+p)}$

In the following, I plug in these values to $\sum_{k=1}^{8} e_k w_k^{(r)}$ to evaluate $\boldsymbol{M}^{-1}\boldsymbol{v}$ row by row (weights taking on the value 0 are not written out). However, before that let us examine $-Pq + Pr + p$ as it is present in all denominators.

Denominator: $-\boldsymbol{Pq}+\boldsymbol{Pr}+\boldsymbol{p}$.

By the definition of r, q, p, $A5$ Monotonicity ($D_j(10) \leq D_j(11)$), and the law of total probability (LTP):

$$
\begin{aligned}
r &= \mathbb{P}\left(D_j(11)=1, D_j(10)=0\right) && (31)\\
q &= \mathbb{P}\left(D_j(11)=1\right) \quad LTP \implies \\
&= \mathbb{P}\left(D_j(11)=1, D_j(10)=1\right)+\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right) && (32)\\
p &= \mathbb{P}\left(D_j(10)=1\right) \quad LTP \implies \\
&= \mathbb{P}\left(D_j(10)=1, D_j(11)=0\right)+\mathbb{P}\left(D_j(10)=1, D_j(11)=1\right) \quad A5 \implies \\
&= \mathbb{P}\left(D_j(10)=1, D_j(11)=1\right) \implies \\
-Pq+Pr+p &= P(r-q)+p \\
&= P(\mathbb{P}\left(D_j(11)=1, D_j(10)=0\right)-[\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right)+\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right)])+p \\
&= -P\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right)+p \\
&= -P\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right)+\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right) \\
&= (1-P)\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right), && (33)
\end{aligned}
$$

which is non-zero because $P \in (0,1)$ and $\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right) \neq 0$ by $A6$ Invertibility. Now we can focus on the linear combinations.

Row of $\boldsymbol{M}^{-1}\boldsymbol{v}$: 1st, $(\boldsymbol{M}^{-1}\boldsymbol{v})_1$.

$$
(\boldsymbol{M}^{-1}\boldsymbol{v})_1 = w_6^{(1)} e_6 = 1e_6 = \mathbb{E}\left[Y_j(00)\right].
$$

Hence the first claim of Theorem 2 is proved.

Row of $M^{-1}v$: 2nd, $(M^{-1}v)_2$.

$$
\begin{aligned}
(\boldsymbol{M}^{-1}\boldsymbol{v})_2 =& w_2^{(2)}e_2 + w_4^{(2)}e_4 + w_8^{(2)}e_8 \\
=& \frac{1}{-Pq+Pr+p}\left\{Pr\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] + p\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1\right] - Pq\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1\right]\right\} \\
=& \frac{1}{-Pq+Pr+p}\left\{P[r\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] - q\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1\right]] + p\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1\right]\right\}
\end{aligned} \tag{34}
$$

At this point, to figure out what $q\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1\right]$ is, we need to think backwards using the law of total probability and the definition of conditional probability.

$$
\begin{aligned}
\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1\right] =& \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right]\mathbb{P}\left(D_j(10)=0 \mid D_j(11)=1\right) \\
&+ \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right]\mathbb{P}\left(D_j(10)=1 \mid D_j(11)=1\right) \\
=& \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right]\frac{\mathbb{P}\left(D_j(10)=0, D_j(11)=1\right)}{\mathbb{P}\left(D_j(11)=1\right)} \\
&+ \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right]\frac{\mathbb{P}\left(D_j(10)=1, D_j(11)=1\right)}{\mathbb{P}\left(D_j(11)=1\right)} \\
=& \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right]\frac{\mathbb{P}\left(D_j(10)=0, D_j(11)=1\right)}{q} \\
&+ \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right]\frac{\mathbb{P}\left(D_j(10)=1, D_j(11)=1\right)}{q},
\end{aligned}
$$

which in turn implies

$$q\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1\right] = \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right]\mathbb{P}\left(D_j(10)=0, D_j(11)=1\right) + \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right]\mathbb{P}\left(D_j(10)=1, D_j(11)=1\right). \tag{35}$$

Applying the same logic to $p\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1\right]$ and keeping in mind that by $A5$ Monotonicity $p = \mathbb{P}\left(D_j(11)=1, D_j(10)=1\right)$ leads to

$$p\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1\right] = \mathbb{P}\left(D_j(11)=1, D_j(10)=1\right)\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1, D_j(11)=1\right]. \tag{36}$$

Plugging in the above expressions, the definition of r, and the denominator, into (34):

$$\begin{aligned}
(\boldsymbol{M}^{-1}\boldsymbol{v})_2 &= \frac{1}{-Pq+Pr+p}\left\{P[r\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] - q\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1\right]] + p\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1\right]\right\} \\
&= \frac{1}{-Pq+Pr+p}\{P\{\mathbb{P}\left(D_j(11)=1, D_j(10)=0\right)\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] \\
&\qquad - [\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right]\mathbb{P}\left(D_j(10)=0, D_j(11)=1\right) \\
&\qquad\quad + \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right]\mathbb{P}\left(D_j(10)=1, D_j(11)=1\right)]\} \\
&\qquad + \mathbb{P}\left(D_j(11)=1, D_j(10)=1\right)\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1, D_j(11)=1\right]\} \\
&= \frac{1}{-Pq+Pr+p}(1-P)\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right)\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1, D_j(11)=1\right]
\end{aligned}$$

Finally, substituting in $-Pq + Pr + p = (1-P)\mathbb{P}\left(D_j(11) = 1, D_j(10) = 1\right)$ (see (33)) gives us

$$
\begin{aligned}
(\boldsymbol{M}^{-1}\boldsymbol{v})_2 =& \frac{1}{-Pq + Pr + p}(1-P)\mathbb{P}\left(D_j(11) = 1, D_j(10) = 1\right)\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10) = 1, D_j(11) = 1\right] \\
=& \frac{1}{(1-P)\mathbb{P}\left(D_j(11) = 1, D_j(10) = 1\right)}(1-P)\mathbb{P}\left(D_j(11) = 1, D_j(10) = 1\right)\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10) = 1, D_j(11) = 1\right] \\
=& \,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10) = 1, D_j(11) = 1\right].
\end{aligned}
$$

Hence the second claim of Theorem 2 is proved.

Row of $\boldsymbol{M}^{-1}\boldsymbol{v}$: 3rd, $(\boldsymbol{M}^{-1}\boldsymbol{v})_3$.

Exploiting symmetry, we can proceed in this case exactly as with the second row.

$$
\begin{aligned}
(\boldsymbol{M}^{-1}\boldsymbol{v})_3 =& w_3^{(3)}e_3 + w_5^{(3)}e_5 + w_7^{(3)}e_7 \\
=& \frac{1}{-Pq + Pr + p}\left\{Pr\,\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1, D_{j'}(01) = 0\right] + p\,\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(01) = 1\right] - Pq\,\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1\right]\right\} \\
=& \frac{1}{-Pq + Pr + p}\left\{P[r\,\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1, D_{j'}(01) = 0\right] - q\,\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1\right]] + p\,\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(01) = 1\right]\right\} \\
=& \frac{1}{-Pq + Pr + p}(1-P)\mathbb{P}\left(D_{j'}(11) = 1, D_{j'}(01) = 1\right)\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(01) = 1, D_{j'}(11) = 1\right] \\
=& \frac{1}{(1-P)\mathbb{P}\left(D_{j'}(11) = 1, D_{j'}(01) = 1\right)}(1-P)\mathbb{P}\left(D_{j'}(11) = 1, D_{j'}(01) = 1\right)\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(01) = 1, D_{j'}(11) = 1\right] \\
=& \,\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(01) = 1, D_{j'}(11) = 1\right].
\end{aligned}
$$

Hence the third claim of Theorem 2 is proved.

Row of $\boldsymbol{M}^{-1}\boldsymbol{v}$: 4th, $(\boldsymbol{M}^{-1}\boldsymbol{v})_4$.

For the last element we have the most non-zero weigths in Table 4, amounting to

$$\begin{aligned}(\boldsymbol{M}^{-1}\boldsymbol{v})_4 =& w_1^{(4)}e_1 + w_2^{(4)}e_2 + w_3^{(4)}e_3 + w_4^{(4)}e_4 + w_5^{(4)}e_5 + w_7^{(4)}e_7 + w_8^{(4)}e_8 \\ =& w_1^{(4)}e_1 + (w_2^{(4)}e_2 + w_4^{(4)}e_4 + w_8^{(4)}e_8) + (w_3^{(4)}e_3 + w_5^{(4)}e_5 + w_7^{(4)}e_7).\end{aligned}$$

Grouping the sum as indicated with the brackets renders the computation easier. In the following, I evaluate the sum group by group.

- $w_1^{(4)}e_1$

$$w_1^{(4)}e_1 = 1e_1 = \mathbb{E}\left[\Delta_j(11|01) - \Delta_j(10|00) \mid D_j(11) = 1, D_{j'}(11) = 1\right] \tag{37}$$

- $w_2^{(4)}e_2 + w_4^{(4)}e_4 + w_8^{(4)}e_8$

$$\begin{aligned}w_2^{(4)}e_2 + w_4^{(4)}e_4 + w_8^{(4)}e_8 =& \frac{1}{\bar{q}(-Pq + Pr + p)}\{-Pqre_2 - pqe_4 + q(Pr + p)e_8\} \\ =& \frac{1}{\bar{q}(-Pq + Pr + p)}\{Pqr(e_8 - e_2) + pq(e_8 - e_4)\}.\end{aligned}$$

Next, I compute $Pqr(e_8 - e_2)$ and $pq(e_8 - e_4)$.

•• $Pqr(e_8 - e_2)$ From the definition of e_8 and from(35), we know that

$$\begin{aligned}
qre_8 =& qr\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1\right] \\
=& \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 0\right] \mathbb{P}\left(D_j(10) = 0, D_j(11) = 1\right) r \\
&+ \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 1\right] \mathbb{P}\left(D_j(10) = 1, D_j(11) = 1\right) r,
\end{aligned}$$

where by definition $\mathbb{P}\left(D_j(10) = 0, D_j(11) = 1\right) = r$, and by the law of total probability and $A5$ Monotonicity

$$r = \mathbb{P}\left(D_j(11) = 1, D_j(10) = 0\right) = \mathbb{P}\left(D_j(11) = 1\right) - \mathbb{P}\left(D_j(11) = 1, D_j(10) = 1\right) = q - \mathbb{P}\left(D_j(11) = 1, D_j(10) = 1\right) \iff \tag{38}$$

$$\mathbb{P}\left(D_j(11) = 1, D_j(10) = 1\right) = q - r, \tag{39}$$

so

$$\begin{aligned}
qr\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1\right] =& \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 0\right] \mathbb{P}\left(D_j(10) = 0, D_j(11) = 1\right) r \\
&+ \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 1\right] \mathbb{P}\left(D_j(10) = 1, D_j(11) = 1\right) r \\
=& \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 0\right] r^2 \\
&+ \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 1\right] (q - r) r.
\end{aligned}$$

By the definition of e_2

$$qre_2 = \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 0\right] qr,$$

so

$$
\begin{aligned}
Pqr(e_8-e_2) = P\{&\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] r^2 \\
&+ \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right](q-r)r \\
&- \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] qr\} \\
= P\{&\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right](q-r)r - \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right](q-r)r\} \\
= P&(q-r)r\{\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right] - \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right]\}.
\end{aligned}
$$

•• $pq(e_8-e_4)$ From the definition of e_8 and from(35), we know that

$$
\begin{aligned}
pqe_8 =& pq\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1\right] \\
=& \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] \mathbb{P}\left(D_j(10)=0, D_j(11)=1\right) p \\
&+ \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right] \mathbb{P}\left(D_j(10)=1, D_j(11)=1\right) p,
\end{aligned}
$$

and from (39) $\mathbb{P}\left(D_j(10)=1, D_j(11)=1\right) = q-r$, and by definition $\mathbb{P}\left(D_j(10)=0, D_j(11)=1\right) = r$ so

$$
pqe_8 = \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] rp + \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right](q-r)p.
$$

From (36) we kow that

$$
\begin{aligned}
pe_4 =& p\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1\right] \\
=& \mathbb{P}\left(D_j(11)=1, D_j(10)=1\right)\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1\right] \\
=& (q-r)\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1\right] \implies \\
pqe_4 =& q(q-r)\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1\right].
\end{aligned}
$$

It follows then that

$$
\begin{aligned}
pq(e_8-e_4) =& \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] rp + \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right](q-r)p \\
&- q(q-r)\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(10)=1\right] \\
=& \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right]((q-r)p-(q-r)q) + \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] rp \\
=& \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right](q-r)(p-q) + \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] rp \\
=& \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right](q-r)(-r) + \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] rp,
\end{aligned}
$$

where the last step is possible due to $A5$ Monotonicity and the law of total probability:

$$\begin{aligned}
p-q =& \mathbb{P}\left(D_j(10)=1\right)-\mathbb{P}\left(D_j(11)=1\right)\\
=& \left(\mathbb{P}\left(D_j(10)\right)=1, D_j(11)=1\right)+\mathbb{P}\left(D_j(10)=1, D_j(11)=0\right))\\
&-\left(\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right)+\mathbb{P}\left(D_j(11)=1, D_j(10)=0\right)\right)\\
=& -\mathbb{P}\left(D_j(11)=1, D_j(10)=0\right)\\
=& -r
\end{aligned} \tag{40}$$

by the definition of r.

Having computed $Pqr(e_8-e_2)$ and $pq(e_8-e_4)$ we can obtain

$$\begin{aligned}
Pqr(e_8-e_2)+pq(e_8-e_4) =& P(q-r)r\{\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right]-\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right]\}\\
&+\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right](q-r)(-r)+\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right]rp\\
=&(q-r)r(P-1)\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right]\\
&+(-P(q-r)r+rp)\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right]\\
=& r\{(1-P)(r-q)\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right]\\
&+\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right](-P(q-r)+p)\}
\end{aligned}$$

Note that $(1-P)(r-q) = (1-p)(-1)\mathbb{P}\left(D_j(11)=1, D_j(10)=1\right) = -Pq+Pr+p$ by (40) and (33). Thus

$$\begin{aligned}
w_2^{(4)}e_2 + w_4^{(4)}e_4 + w_8^{(4)}e_8 =& \frac{1}{\bar{q}(-Pq+Pr+p)}\{Pqr(e_8-e_2)+pq(e_8-e_4)\} \\
=& \frac{1}{\bar{q}(-Pq+Pr+p)}\{r\{(1-P)(r-q)\,\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right] \\
&+ \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right](-P(q-r)+p)\} \\
=& \frac{r}{\bar{q}}\{\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=0\right] - \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11)=1, D_j(10)=1\right]\}.
\end{aligned}$$

- $w_3^{(4)}e_3 + w_5^{(4)}e_5 + w_7^{(4)}e_7$

$$\begin{aligned}
w_3^{(4)}e_3 + w_5^{(4)}e_5 + w_7^{(4)}e_7 =& \frac{1}{\bar{q}(-Pq+Pr+p)}\{-Pqre_3 - pqe_5 + q(Pr+p)e_7\} \\
=& \frac{1}{\bar{q}(-Pq+Pr+p)}\{Pqr(e_7-e_3)+pq(e_7-e_5)\}.
\end{aligned}$$

Due to the symmetry assumptions and the definitions of e's, this sum behaves exactly as $Pqr(e_8-e_2)+pq(e_8-e_4)$, only that this time we have $D_{j'}(11)$ instead of $D_j(11)$, $D_{j'}(01)$ instead of $D_j(10)$ and $\Delta_j(01|00)$ instead of $\Delta_j(10|00)$. So the proof is exactly the same, and we have

$$w_3^{(4)}e_3 + w_5^{(4)}e_5 + w_7^{(4)}e_7 = \frac{r}{\bar{q}}\{\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11)=1, D_{j'}(01)=0\right] - \mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11)=1, D_{j'}(01)=1\right]\}.$$

Putting all this together leaves us with

$$
\begin{aligned}
(\boldsymbol{M}^{-1}\boldsymbol{v})_4 =& w_1^{(4)}e_1 + (w_2^{(4)}e_2 + w_4^{(4)}e_4 + w_8^{(4)}e_8) + (w_3^{(4)}e_3 + w_5^{(4)}e_5 + w_7^{(4)}e_7) \\
=& \mathbb{E}\left[\Delta_j(11|01) - \Delta_j(10|00) \mid D_j(11) = 1, D_{j'}(11) = 1\right] \\
&+ \frac{r}{\bar{q}}\{\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 0\right] - \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 1\right]\} \\
&+ \frac{r}{\bar{q}}\{\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1, D_{j'}(01) = 0\right] - \mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1, D_{j'}(01) = 1\right]\}.
\end{aligned}
$$

Hence the fourth statement of Theorem 2 is proved.

E.6. Overview of the Proof

Let

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_j &\equiv \left(n^{-1}\boldsymbol{Z}_j'\boldsymbol{D}_j\right)^{-1} n^{-1}\boldsymbol{Z}_j'\boldsymbol{y}_j \\
&= \left(n^{-1}\sum_{i=1}^{n}\boldsymbol{z}_{ji}\boldsymbol{d}_{ji}'\right)^{-1} n^{-1}\sum_{i=1}^{n}\boldsymbol{z}_{ji}y_{ji} \\
\boldsymbol{d}_{ji} &\equiv [1, d_{ji}, d_{j'i}, d_{ji}d_{j'i}]' \in \{0,1\}^{4\times 1} \\
\boldsymbol{z}_{ji} &\equiv [1, z_{ji}, z_{j'i}, z_{ji}z_{j'i}]' \in \{0,1\}^{4\times 1}
\end{aligned}
$$

for $j \in \{A, B\}$ and $j' \in \{A, B\} \setminus \{j\}$.

Examining $\operatorname{plim}\hat{\boldsymbol{\theta}}_j$, i.e. the probability limit of $\hat{\boldsymbol{\theta}}_j$, or equivalently $\boldsymbol{\theta}_j : \lim_{n\to\infty}\mathbb{P}\left(||\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j||_2^2 > \varepsilon\right) = 0$ for any $\varepsilon > 0$ leads to

$$
\begin{aligned}
\operatorname{plim}\hat{\boldsymbol{\theta}}_j &= \operatorname{plim}\left(n^{-1}\sum_{i=1}^{n}\boldsymbol{z}_{ji}\boldsymbol{d}'_{ji}\right)^{-1} n^{-1}\sum_{i=1}^{n}\boldsymbol{z}_{ji}y_{ji} \\
&= \left(\operatorname{plim} n^{-1}\sum_{i=1}^{n}\boldsymbol{z}_{ji}\boldsymbol{d}'_{ji}\right)^{-1} \operatorname{plim} n^{-1}\sum_{i=1}^{n}\boldsymbol{z}_{ji}y_{ji} \\
&= \mathbb{E}\left[\boldsymbol{z}_{ji}\boldsymbol{d}'_{ji}\right]^{-1}\mathbb{E}\left[\boldsymbol{z}_{ji}y_{ji}\right] \\
&= \mathbb{E}\left[\begin{bmatrix}1\\ Z_j\\ Z_{j'}\\ Z_jZ_{j'}\end{bmatrix}\begin{bmatrix}1 & D_j & D_{j'} & D_jD_{j'}\end{bmatrix}\right]^{-1}\mathbb{E}\left[\begin{bmatrix}1\\ Z_j\\ Z_{j'}\\ Z_jZ_{j'}\end{bmatrix}Y_j\right] \\
&\equiv \boldsymbol{M}^{-1}\boldsymbol{v}
\end{aligned}
$$

by the continuous mapping property of the probability limit and by the Weak Law of Large Numbers for i.i.d. data (as the data across pairs are i.i.d.).

I proved that Identifying Assumptions $A1 - A8$ are sufficient to establish

$$
\begin{aligned}
(\boldsymbol{M}^{-1}\boldsymbol{v})_1 =& \mathbb{E}\left[Y_j(00)\right] \\
(\boldsymbol{M}^{-1}\boldsymbol{v})_2 =& \mathbb{E}\left[\Delta_j(10|00) \mid D_j(10) = 1, D_j(11) = 1\right] \\
(\boldsymbol{M}^{-1}\boldsymbol{v})_3 =& \mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(01) = 1, D_{j'}(11) = 1\right] \\
(\boldsymbol{M}^{-1}\boldsymbol{v})_4 =& \mathbb{E}\left[\Delta_j(11|01) - \Delta_j(10|00) \mid D_j(11) = 1, D_{j'}(11) = 1\right] \\
& + \frac{r}{\bar{q}}\{\mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 0\right] - \mathbb{E}\left[\Delta_j(10|00) \mid D_j(11) = 1, D_j(10) = 1\right]\} \\
& + \frac{r}{\bar{q}}\{\mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1, D_{j'}(01) = 0\right] - \mathbb{E}\left[\Delta_j(01|00) \mid D_{j'}(11) = 1, D_{j'}(01) = 1\right]\}.
\end{aligned}
$$

Thus the proof of Theorem 2 is complete ■