PRACTICAL INSIGHTS INTO CAUSAL METHODS: NONPARAMETRIC IV BOUNDS AND ANCOVA UNDER GENERAL INTERFERENCE

by

Ralph Møller Trane

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Statistics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2023

Date of final oral examination: December 11th, 2023

The dissertation is approved by the following members of the Final Oral Committee:

Hyunseung Kang, Associate Professor, Statistics
Keith Levin, Assistant Professor, Statistics
Lu Mao, Associate Professor, Biostatistics and Medical Informatics
Guanhua Chen, Associate Professor, Biostatistics and Medical Informatics
Jiwei Zhao, Associate Professor, Biostatistics and Medical Informatics

To my wife and our son

Table of contents

Chapter 1

Introduction

Most who have any training in the realm of statistics have heard the mantra: "correlation is not causation!" Causal Inference is the study of assumptions and methods that can help us move from observed data and correlation to causal statements. It is at the core of any scientific question that aims at answering questions related to how the change in one variable might affect another.

We will not go through the vast literature on causal inference and causal methods here (see instead, for example, Hernan and Robins (2023) for a soft but thorough introduction). However, we will introduce just enough ideas to set the stage for an introduction of the two specific problems considered in Chapter 2 and Chapter 3.

Consider the question of whether a drug has an effect on a specific health outcome. To make any progress towards an answer to this question, we find a group of individuals indexed by $i = 1, ..., n$ that are at risk of having the disease. We then record two things on each individual:

1. Y_i the health outcome. This could be an indicator (did individual i get sick or not) or a measurement of kind (the blood pressure of individual i)
2. W_i the treatment indicator. This indicates whether individual i received treatment (in this example, the drug) or not.

To get to a statement about a causal effect of the treatment on the outcome, we first have to specify what is meant by a causal effect. To do so, we introduce the notion of "potential outcomes."[1] We assume each individual has two potential outcomes. $Y_i(1)$ is the outcome we would have seen if individual i would have been treated, and $Y_i(0)$ is the outcome we would have seen if individual i would not have been treated. This allows us to clearly define a causal effect of the treatment on the outcome of individual i as the difference between what would have been the outcome had the individual been treated, and what would have been the outcome had the individual not been treated: $Y_i(1) - Y_i(0)$. In reality, we only ever observe one scenario for each individual, since individual i either received the treatment or did not receive the treatment, and therefore we only observe either $Y_i(1)$ or $Y_i(0)$. This means the difference above is never observed for any individual, and therein lies the problem.

While knowing $Y_i(1) - Y_i(0)$ for every individual would be extremely useful, we have to settle for less. From here one, our main focus will be on the Average Treatment Effect, $\tau = \mathbb{E}\left[Y_i(1) - Y_i(0)\right] = \mathbb{E}\left[Y_i(1)\right] - \mathbb{E}\left[Y_i(0)\right]$. This quantity can be estimated under three core assumptions:

i. the Stable Unit Treatment Assumption (SUTVA): $Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$
ii. no unmeasured confounding (also known as exchangeability): $(Y_i(1), Y_i(0)) \perp W_i$
iii. positivity (also known as overlap): $0 < P(W_i = 1) < 1$ for all i

A simple application of conditional probabilities, law of large numbers, and Slutsky's theorem shows that under (i)-(iii), $\mathbb{E}\left[Y_i(w)\right] = \mathbb{E}\left[Y_i | W_i = w\right]$, which we can estimate using the average of the outcomes in our data where $W_i = w$. Therefore, $\hat{\tau} = \frac{\sum_{i=1} W_i Y_i}{\sum_{i=1}^{n} W_i} - \frac{\sum_{i=1} (1 - W_i) Y_i}{\sum_{i=1}^{n} 1 - W_i}$ is a consistent estimator of the ATE.

Chapter 2 and Chapter 3 consider scenarios where either (i) or (ii) above does not seem like a reasonable assumption.

[1] The potential outcomes framework was first proposed by Neyman (1923), and later extended by Rubin (1974). It is also refered to as the Rubin Causal Model or Neyman-Rubin Causal Model.

1.1 Nonparametric IV Bounds in Mendelian Randomization Studies

For now, assume that SUTVA holds. One way of ensuring assumptions (ii) and (iii) hold is by performing a Randomized Controlled Trial. That is, observe data from an experiment where the researcher randomly assigns the individuals to receive the treatment or not. This randomization ensures, by definition, that each individual has a chance of receiving treatment (i.e. positivity holds), and that the treatment assignment is completely independent of the potential outcomes (i.e. exchangeability). However, there are a wide array of questions one might ask where performing an RCT is not possible. For example, we cannot conduct an RCT exploring the effects of smoking on pregnancy, since we have strong evidence suggesting smoking leads to negative outcomes (Abraham et al., 2017). Therefore, asking pregnant people to smoke is highly unethical. Also, RCTs are often expensive and hard to perform in practice. In such cases, we seek to use observational data to gather evidence of causal effects. However, this introduces a number of problems, one of which is the possibility of unmeasured confounding.

As summarized by Davey Smith and Ebrahim (2003), observational epidemiology throughout the 1990s produced a variety of interesting findings, but most were dismissed after rigorous testing through randomized clinical trials. Davey Smith and Ebrahim (2003) suggest the most likely culprit for this discrepancy is confounding in the observational studies. The use of instrumental variables (IVs) can help alleviate this problem (see Angrist and Krueger (2001) for an overview). In essence, an instrumental variable is a variable that influences the exposure status, influences the outcome only through the exposure, and is independent of other potential confounders. If such a variable is available, the treatment effect can be estimated, even in the presence of unmeasured confounding, but good instruments are hard to come by. While IVs come with many challenges, two stand out: finding a strong instrument, and finding an instrument that is in fact independent of unmeasured confounders. As mentioned by Bound et al. (1995), using an instrument that is only weakly associated with

the treatment of interest could lead to finite-sample bias of a non-trivial magnitude. Luckily, this can to some extend be mitigated by assessing the correlation between the instrument and the exposure, often through an F-test. However, it is hard in practice to assess the independence of confounders. The use of genetic variants as instruments somewhat mitigates this issue.

In recent years, genetic variants, often in the form of single nucleotide polymorphisms (SNPs), have been used as IVs to estimate causal effects in epidemiological studies, often referred to as Mendelian randomization (MR) studies. The validity of this approach relies on a consequence of Mendel's second law, namely that the inheritance of genetic markers is in essence a random assignment. This makes genetic markers great candidates for instrumental variables when studying causal effects of certain exposures on some disease statuses, though not without limits. As mentioned by Lawlor et al. (2008), finding genetic markers that are strong instruments might be challenging, as is making sure seemingly strong associations are reliable. Many genetic associations have proven hard to replicate suggesting the associations might be spurious. None the less, in situations where a strong, reliable association between genetic marker and exposure is present, an MR analysis is a powerful tool.

Another practical challenge, as pointed out by Burgess, Scott, et al. (2015), is that of obtaining individual level data for MR analyses. However, there is a large number of databases containing summary statistics from Genome Wide Association Studies (GWAS) readily available. By gathering results from two independent GWAS results, one providing information about the genes-exposure associations and the other about the genes-outcome associations, the causal effect of the exposure on the outcome can be estimated. This approach is often refered to as a two-sample MR study. Under such a design, parametric methods are popular. For independent genetic instruments, Burgess, Scott, et al. (2015) suggest the Inverse Variance Weighted (IVW) method to combine information from multiple genetic instruments. The estimated causal effect of W on Y is found using the genetic instruments separately, and then a weighted average is created using the inverse variance of

each estimate weights. The delta method gives an approximate standard error for the IVW estimator. If the instruments are correlated, a likelihood-based method can be used, assuming the separate data sources are non-overlapping. Bowden, Davey Smith, et al. (2016) suggests the Weighted Median Estimator as an improvement over the IVW estimator that is more robust towards the inclusion of invalid instruments.

MR-Egger is another method robust to invalid instruments. Bowden, Del Greco M, et al. (2016) suggest this adaption of Egger regression to address potential violations of IV assumptions. By viewing MR analyses using multiple instruments as a meta-analysis, they show that bias induced by pleiotropy[2] can be viewed as small-sample bias, and show how Egger regression can be used to produce valid estimates of the causal effect of interest. Verbanck et al. (2018) introduced MR-PRESSO as another method for dealing with pleiotropy.

Q. Zhao et al. (2020) show that a linear model is a good approximation when pleiotropy is not an issue, but also find that in real datasets, both systematic and idiosyncratic pleiotropy is common. They use a random effects model to account for this, and propose a consistent and asymptotically normal estimator, even in the scenario where no genetic marker satisfies the IV assumptions exactly.

All of the above mentioned methods rely on relatively strong parametric assumptions which are hard to empirically verify. As an alternative, nonparametric IV bounds put aside the goal of point estimation of the causal effect and instead consider partial identification. In the following we will consider binary exposure (or treatment) W, binary outcome Y, binary instrument Z, and an arbitrary unmeasured confounder C. Based on nothing more than the standard IV assumptions, we are able to construct firm bounds on the Average Treatment Effect (ATE). Note that in this case, $\tau = \mathbb{E}\left[Y(1) - Y(0)\right]$ is by definition confined to the interval $[-1, 1]$. Typically, nonparametric IV bounds are used when the outcome, the exposure, and the instrument are all binary and simultaneously observed. In this setting, one can bound the ATE such that the width of the bounds is $1/2$, which means half the

[2]Pleiotropy is the genetic concept of a single genetic marker influencing multiple phenotypes. This opens up the possibility of the genetic marker to be associated with the outcome through other means than its influence on the exposure.

possible values can be ruled out without any furter assumptions. Robins (1989) derived a set of bounds on the ATE of AZT treatment on the time to clinical AIDS among HIV-infected subjects, and Manski (1990) presents similar bounds in a more general context. These bounds have since been referred to as the "natural" or "Robins-Manski" bounds, and rely only on one additional assumption. Assuming that the instrument Z is independent of the potential outcomes $Y(w)$ for all w, the "Robins-Manski" bounds are such that the width of the bounds is no more than $P(W = 1|Z = 0) + P(W = 0|Z = 1)$. This was pointed out by Balke and Pearl (1997), who also show that under the assumption of monotonicity, that is $P(W = 1|Z = 1, C) \geq P(W = 1|Z = 0, C)$, the width of the bounds is exactly $P(W = 1|Z = 0) + P(W = 0|Z = 1)$, which is equivalent to $1 - (P(W = 1|Z = 1) - P(W = 1|Z = 0))$. This shows that the width of the bounds is decreasing as a function of the strength of the instrument, where by "strength" we refer to how influential the instrument is on the probability of being exposed. T. S. Richardson and Robins (2014) derive the same bounds as Balke and Pearl (1997), but under a slightly stronger assumption. However, they also extend the bounds to categorical instruments with more than two levels, which is a more general version of the bounds for three-leveled instruments derived by Cheng and Small (2006). This is an important result in the context of MR studies, since genetic markers used as instruments are three-leveled.

Balke and Pearl (1997) derive bounds on the ATE by the use of linear programming to solve a system of linear constraints derived from the laws of probability. The result is not only bounds on the ATE, but also a set of constraints on other quantities of the form $P(Y = y, W = w|Z = z)$ in the shape of a set of inequalities. These can be used to falsify the IV assumptions. That is, while we cannot test if the IV assumptions hold, we can conclude that one or more of them do not if any of the aforementioned inequalities are violated.

While the bounds that can be derived when (W, Y, Z) is observed simultaneously have been extensively studied, not much has been done to extend the nonparametric IV bounds to two-sample designs. Ramsahai (2012) shows that bounds can be derived when (W, Y) and (Z, Y) are observed separately, but only for binary instruments. They follow the lead of

Balke and Pearl (1997) and use linear programming. We follow up on the work by Ramsahai (2012), extend it to allow for categorical instruments with more than two levels, and explore its utility in MR studies based on two-sample data. Despite IV bounds being an attractive alternative to the parametric models for studying exposure effects in MR studies (Didelez and Sheehan, 2007; Swanson, 2017), there is little work on actually using these bounds in the very common two-sample MR settings, i.e. two-sample designs with summary statistics, nor any practical guidance on when the bounds would be informative. The overall goal of Chapter 2 is to offer some practical guidance on using IV bounds in two-sample MR studies.

1.2 Covariate Adjustments in Randomized Controlled Trials under General Interference

When considering nonparametric IV bounds in Chapter 2, we assume that SUTVA holds. However, this is not always the case. In fact, more than a century ago, Ross (1916) wrote about the idea of "dependent happenings" to illustrate scenarios where the rate at which certain outcomes occur depends on the prevalence of the outcome, such as "infectious diseases, membership of societies and sects with propagandas, trade-unions, political parties, etc., due to propagation from within, that is, from individual to individual."(Ross, 1916) Later, Cox (1958) defined "interference between different units" as it is now most commonly referred to, i.e. two units interfere if the treatment status of one unit affects the outcome of the other unit. A common example of when this is the case is that of evaluating the effect of vaccination. Here, my outcome is dependent on whether or not the people I surround myself with are vaccinated. This constitues a violation of SUTVA in that the outcome observed when treated is not uniquely defined, but could vary depending on how many peers are treated.

Even though it was well established that SUTVA might be violated in a wide variety of practical settings, most work in causal inference over the last half of the 20th century built on SUTVA. Only recently has significant progress been made on analyzing data where

SUTVA is violated due to interference. Manski (1993) discussed the identification problem of endogenous social effects, a class of effects that had already long been discussed in sociology and social psychology as "peer influences", "neighborhood effects", "herd behavior", etc. Manski (1993) considered the identification of what is often referred to as the spillover or indirect effect, showed that this is identifiable when potential outcomes follow a linear model, and presented the negative result of non-identifiability in certain general nonparametric settings, but never touched on the notion of a direct treatment effect in the presence of interference nor the identifiability or estimation thereof. Halloran and Struchiner (1995) connected the progress made in the previous decades to the notion of "dependent happenings" in that they specifically considered estimation of treatment effects in infectious disease studies, which was one of the examples of "dependent happenings" in Ross (1916). They discussed the extension of the Rubin Causal Model (Rubin, 1974) to allow for interference (as suggested by Rubin (1990)), introduced individual level causal effects, and discussed various estimands. Sobel (2006) considered the "Moving to Opportunity" (MTO) data set, which is based on a housing mobility experiment in five cities. Residents in certain ghettos were randomly assigned to receive relocation assistance. It is claimed that the estimates presented in previous literature do not exactly correspond to traditional causal estimands because of interference. Sobel (2006) proceeded to define average causal effects in treatment groups, both direct and spillover effects, and discussed how these are linked to the estimates presented in other work dealing with this specific data set. Assuming a finite population and considering randomization inference, they showed that the standard difference-in-means estimator is unbiased under a complete randomization allocation scheme, and link various noncausal parameters to differences of causal estimands. However, all results are presented in the very specific context of the MTO data set. Rosenbaum (2007) tackled randomization inference when interference is present in a more general setting. It is shown that randomization inference is still possible without any restrictions on the interference, and that distribution-free randomization tests can be inverted to confidence statements about the magnitude of the effects. This approach, however, is limited in that it cannot distinguish between the null hypothesis of no primary effect, and the null hypothesis of no effect

at all. The former states that both treated and untreated units benefit equally from the fact that treatment has been given. This means it does not matter which units receive the treatment, as long as it is given. The latter null hypothesis states that there is no effect of the treatment at all, that is the outcome would be the same had no unit been treated. These two null hypotheses are, in practice, very different, and the fact that we cannot tell them apart illustrates the challenges of drawing any inference when interference is present.

1.2.1 Partial Interference

As mentioned above, Halloran and Struchiner (1995) introduced both individual level causal effects, and some causal estimands that might be of interest, but they stopped short of extending the ATE to the general interference setting. The complexity of allowing for general interference, they argued, means that there is little to no hope of this being of practical use. This argument, unsurprisingly, still largely rings true, and simplifying assumptions are necessary to make meaningful progress if one is to move beyond simple point estimation to inference. Hudgens and Halloran (2008) considered a two-stage randomized study as a means to estimate direct, indirect, total, and overall causal effects in the presences of interference. They assume a partial interference structure. That is, they assume that the study population consists of independent groups, where individuals within groups might interfere with each other, but individuals between groups do not. The treatment assignment is then performed in two-stages: first, each group is assigned to one of two allocation strategies, then each individual within each group is assigned treatment using the group assigned allocation strategy. They introduced estimands that have since been studied extensively, including the estimand for the direct effect that we consider estimating in Chapter 3. They considered the estimation of these under partial interference and a two-stage randomization design. But even under partial interference, variance estimation is a lost cause, as pointed out in Section 4.2.1 of Hudgens and Halloran (2008). This means that while identification of the causal estimands is possible under the two-stage design assuming partial interference, inference about these causal estimands is unobtainable.

It is important to note that "dependent happenings" is not exactly equivalent to interference as defined by Cox (1958), since the definition of interference in Cox (1958) allows for a number of different causal structures. Ogburn and VanderWeele (2014) explore the identification of causal effects under what they refer to as different "types of interference." They review the identification of causal estimands in the partial interference setting, and then discuss three types of interference. First, direct interference arise when the treatment assigned to individual i might directly influence the outcome of a peer j. Second, interference by contagion is the idea that the outcome of individual i is only influenced by a peer j's treatment assignment through its effect on j's outcome. This is the case in the infectious disease setting and is what Ross (1916) originally described as "dependent happenings." Finally, allocational interference happens when the outcome of individual i is affected by the group that that it happens to be allocated to. As an example, Ogburn and VanderWeele (2014) mention the random assignment to dorms, which happens within a block (college). While the distinction between the structure of the dependence between the observed outcomes is important, and might be useful when trying to improve on estimation techniques, it is one we will not explore further here.

Since Hudgens and Halloran (2008), different strategies to make further simplifications to the general setup have been studied. These strategies have generally seeked to impose restrictions on the structure of the interference graph (that is, limit who can interfere with whom), and restrictions on the structure of the potential outcomes (for example, assume a parametric relationship between outcome and treatment, or limit the dependencies between outcomes). Hudgens and Halloran (2008) restrict the structure of the interference graph through the partial interference assumption, and then restrict the dependency between outcomes by introducing the stratified interference assumption, which states that the dependency on peers' treatment status is only through the number of peers who are treated while it is irrelevant which peers are treated. They then proceed to present a mix of unbiased and conservative variance estimators for the causal estimands under the two-stage randomization study design.

Tchetgen and VanderWeele (2012) present further work on this same setting. They show that, even without the stratified interference assumption, conservative estimation of the variance is possible, and construct finite sample confidence intervals for binary outcome. The binary nature of the outcomes makes it possible to make progress without the stratified interference assumption. Finally, Tchetgen and VanderWeele (2012) discuss the use of observational studies in lieu of two-stage randomized studies. This idea was further expanded on by Liu, Hudgens, and Becker-Dreps (2016). They first consider Inverse Probability Weighted (IPW) estimators under general interference, but where an interference graph is observed. That is, we know exactly which units might interfere with each other. In this scenario, when the treatment assignment mechanism is known, the IPW estimator is unbiased without any further assumptions on the potential outcomes. However, for inference, they also impose the partial interference assumption, and further assume that the groups are a random sample from a super-population of groups. This induces independence between groups, and allows Liu, Hudgens, and Becker-Dreps (2016) to establish asymptotic normality of the IPW estimator as the number of groups grows, a result that is extended to a scenario where the treatment assignment mechanism is unknown, but a correctly specified propensity score model is available. It is important to note that the propensity score model in this case models the joint probability of an individuals treatment status and the treatments of all peers of the individual. Later, this was expanded to a doubly robust (DR) estimator by Liu, Hudgens, Saul, et al. (2019), still assuming partial interference with groups being a random sample from an infinite super-population of groups. They show that their DR estimator is asymptotically normal if either the propensity score model or outcome model is correctly specified.

1.2.2 General Interference

Others have tried to impose different restrictions on the interference graph and the potential outcomes. Aronow and Samii (2017) relax both the partial interference and stratified interference assumptions. They put no restrictions on the network, but require the researcher

to have knowledge of an exposure mapping, which maps the treatment vector $\boldsymbol{W}$ to an exposure value. They then consider estimating contrasts between different exposure values when randomness is entirely due to the randomization of the treatment, and conclude that such contrasts can be estimated using the Horvitz-Thompson estimator for which they derive a conservative variance estimator. Note that the stratified interference assumption is a specific kind of exposure mapping that maps the treatment vector to an exposure value given by the number of peers who are treated. While this assumption enables a wide range of results, it comes up short when considering asymptotic behavior. To get consistency and asymptotically valid confidence intervals, Aronow and Samii (2017) must also impose a restriction on the growth of the interference graph, such that the total number of peers in the graph is $o(n^2)$.

Chin (2019a) also considers the Horvitz-Thompson estimator, but for estimating the direct effect in randomized experiments. As Aronow and Samii (2017), they consider a finite population regime, but do not impose the exposure mapping assumption on the potential outcomes. They find that the Horvitz-Thompson estimator is asymptotically normal under a "local interference" assumption. This assumption restricts the growth of the network in a way that the largest number of peers $\max_i N_i$ is $o(n^{1/4})$. They also consider the difference-in-means estimator, and show that this is asymptotically normal in an "approximate local interference" setting, that is there exist neighborhoods with high interference within, but low interference between neighborhoods. This is both an assumption on how dependent potential outcomes are on each other, and that the largest number of peers is $o(n^{1/3})$.

A more general exploration of the difference-in-means estimator for estimating the direct effect is performed by Sävje et al. (2021). They also consider asymptotic behavior in a finite sample regime when treatment is randomly assigned, and show that the difference-in-means estimator is consistent under a number of different assumptions on the growth of the network. They focus on convergence rates of the difference-in-means estimator and illustrate how this varies with the depenency structure induced by the network, but do not arrive at asymptotic normality. They do, however, discuss an asymptotically conservative variance

estimator. Similarly, Ogburn, Sofrygin, et al. (2022) consider finite sample asymptotics with very limited assumptions on the potential outcomes. They consider the use of observational studies, and assume a Structural Equations Model (SEM) for the potential outcomes. They then proceed to use semiparametric theory for asymptotic results, and find that estimators based on the semiparametric influence function are asymptotically normal if $\max_i N_i = o(\sqrt{n})$. Similarly to Sävje et al. (2021), they find that the rate of convergence depends on the growth of the network.

While most of the work discussed above has mainly focused on direct effect estimation, others have explored the estimation of indirect or peer causal effects. As already mentioned, the identification of such effects was the main purpose of Manski (1993), who considered a linear model for the potential outcomes and the estimation of contagion effects. McFowland III and Shalizi (2021) further explore this under different network restrictions. They assume that the network is a realization of a random process driven by latent variables. That is, each observational unit i is associated with a latent variable U_i, and units i and j are peers with some probability. This is similar to the graphon model that we consider in Chapter 3. Of the outcomes, they assume a time series model where the outcome at time t is a linear combination of the outcome at time $t-1$, the average of peers' outcomes, and potentially the latent positions. They propose to estimate the contagion effect using the Ordinary Least Squares (OLS) estimator and show that the OLS estimator is unbiased and consistent if we can recover latent positions with some degree of success.

A different approach to the same problem is presented by Egami and Tchetgen (2021). Rather than assuming the linear model and that the graph is a realization of a random process, they utilize the existence of Negative Control Outcomes (NCO) and Negative Control Exposures (NCE), and apply a Central Limit Theorem based on the idea of ψ-dependence presented by Kojevnikov et al. (2021). In its purest form, the ψ-dependence result restricts the dependence between potential outcomes such that potential outcomes for units far from one another in the interference graph exhibit smaller degree of dependence than potential outcomes for units close to each other. While Kojevnikov et al. (2021) present their work

in a general network setting, Egami and Tchetgen (2021) apply their results to a specific causal setting. In particular, they present Assumption 3 which offers a tradeoff between graph denseness and dependency between potential outcomes. If the dependence between potential outcomes decays fast as the distance between units increases, denser graphs can be accommodated. On the other hand, if potential outcomes are strongly dependent even for units far from one another in the interference graph, a sparser graph is needed. This perfectly encapsulates the tradeoffs made in the interference litterature. Using negative controls, Egami and Tchetgen (2021) propose an estimator based on a Generalized Method of Moments approach. They consider a setup where observations are available at two time-points, baseline and follow-up, and are interested in the estimation of the effect of peers' outcomes at baseline on the outcomes observed at follow-up. They show that their proposed estimator is consistent and asymptotically normal with a convergence rate of $\sqrt{n}$.

Leung (2020) considers a somewhat similar setup to McFowland III and Shalizi (2021), but with the focus on treatment effects rather than contagion. The linear model is replaced by a more general nonparametric model, which assumes that the potential outcomes for unit i are given as a function of treatment status W_i, number of peers treated M_i, number of peers N_i, and an unobserved variable ϵ_i. They show that identification is possible under very mild assumptions, but need a weak-dependence assumption to get asymptotic results. In particular, they show that a difference-in-means style estimator is consistent and asymptotically normal if, in addition to other regularity conditions, third moments of the number of peers is bounded in probability. More over, they show that this is achievable even when the network is not completely observed.

Li and Wager (2022) combine aspects of Leung (2020) and McFowland III and Shalizi (2021). They consider interference graphs as randomly generated from a graphon model, and restrict the potential outcomes to be functions of treatment status, proportion of peers who are treated, the latent position, and an unobserved variable. Under smoothness conditions of the potential outcomes in their dependence on the proportion of peers who are treated, they show that both the Horvitz-Thompson and Hàjek estimators are consistent for the direct

effect and asymptotically normal with convergence rate $\sqrt{n}$. They present their asymptotic results for a sequence of finite populations connected by a graph that is the result of a random draw from the graphon model. This randomness from the graphon model makes it possible to handle both sparse and dense graphs, and achieve $\sqrt{n}$ convergence regardless of the growth rate of the network. It is worth noting that the graphon model itself imposes some restrictions on the growth of the network, but in general these are relatively weak. On the other hand, the smoothness assumptions on the potential outcomes are stronger than those made by Leung (2020) or Chin (2019a). Also, while the estimators can be calculated without any knowledge of the graphon, the asymptotic variance is not in general estimable, as it depends on both the graphon and the derivatives of the potential outcomes functions. Li and Wager (2022) illustrate how their results can be used in tandem with various other assumptions one might be willing to make to get conservative variance estimators.

Other aspects of causal inference in the presence of interference have also been explored. Athey et al. (2018) consider randomization inference for various network types, and in particular consider testing various non-sharp null-hypotheses. Among these, they consider the calculation of exact p-values for testing hypotheses of no spillover effects, no higher order peer effects, and competing network specifications. Choi (2017) considers estimation of the "attributable effect of treatment" assuming monotonicity. That is, the estimation of $A = \sum_{i=1}^{n} Y_i - Y_i(\boldsymbol{W})$ under the assumption that $Y_i(\boldsymbol{W}) \geq Y_i(0)$. Rather than point estimation of A, interval estimation is considered, and they show how conservative upper bounds can be derived by solving an optimization problem when it is assumed that $Y_i(0) \in \mathbb{Z}$ for all i. They also derive finite sample one-sided confidence intervals for $\sum_i Y_i(\boldsymbol{W})$ when the outcome is binary. Later, these ideas were further explored assuming binary outcomes. Choi (2023) shows that asymptotically conservative confidence intervals for otherwise unidentifiable estimands can be obtained with no further assumptions other than randomization of the treatment assignment. This work differs significantly from most of the work discussed previously in that no restrictions on the interference are necessary.

1.2.3 Covariate Adjustments

As already mentioned, variance estimation is the main challenge in dealing with interference. In much of the work done in the area, even deriving an expression for the asymptotic variance of an estimator is not straightforward. When this is done, the conclusion is often that more likely than not, the asymptotic variance of estimators under interference is larger than we would expect if no interference is present (Chin, 2019a; Li and Wager, 2022).

Adjusting for covariates can be a powerful tool to improve precision of estimators when analyzing randomized experiments, such as randomized clinical trials (RCTs). When considering the classic setting where SUTVA holds, Hernández et al. (2006), for example, showed that adjusting for covariates could lead to a 25% reduction in the required sample size. There is a vast literature dedicated to whether one should perform such adjustments, and if so, how this should be done. One popular method, the linear regression model, is often used due to its simplicity and being widely taught across disciplines. However, the validity of such adjustments are not universially accepted. Freedman (2008) critiqued the use of linear models saying that randomization alone does not offer any justification for the linearity assumption, and that an incorrectly specified linear model might lead to biased and asymptotically inefficient estimators. However, as pointed out by Lin (2013), the shortcomings brought up by Freedman (2008) are easily remedied. In fact, Tsiatis et al. (2008) showed that fitting a linear model using Ordinary Least Square (OLS) will always perform asymptotically at least as well as an unadjusted Hàjek estimator if all interaction terms between treatment indicator and covariates are included, regardless of the correctness of the model. They take this one step further, and show that the resulting estimator is asymptotically equivalent to an Augmented Inverse Probability Weighted (AIPW) estimator, which is known to be locally efficient, a fact that is obtained using semiparametric theory (see, for example, Robins and Rotnitzky (1995)). While more complex AIPW estimators might be more efficient, the simplicity of the linear regression model and the computationally effectiveness of the OLS method makes it a popular choice. To that end, the U.S. Food and Drug Administration recently published an updated set of guidelines suggesting that

covariate adjustments should always be performed when analyzing RCTs, and that a linear regression model is an acceptable choice (U.S. Food and Drug Administration. Center for Drug Evaluation and Research, 2023).

As already mentioned, most of previous research in the area of covariate adjustment in the analysis of RCTs has focused on the traditional setting where trial subjects do not interfere with each other. The work presented in Chapter 3 builds directly on the work by Li and Wager (2022), and considers the potential for variance reduction through covariance adjustments. We will explore the behavior of the general AIPW estimator when analyzing RCTs under general interference, and establish its asymptotic behavior when the interference graph is the result of a graphon model. Others have considered the AIPW estimator in the context of interference. Emmenegger et al. (2023) considers the general AIPW estimator in a general interference setting. They show that using sample splitting to evaluate the estimator, one can achieve both a central limit theorem with $\sqrt{n}$ convergence rate, and a consistent estimator for the variance. Especially the latter is noteworthy. To achieve this result, however, the growth of dependencies between units has to be limited. In contrast, the random graph asymptotic framework allows us to achieve $\sqrt{n}$ convergence rate even for dense networks, and an interpretable expression for the asymptotic variance.

One of the main results is the application to the OLS estimator. The results of Tsiatis et al. (2008) are replicated in the presence of interference. This in particular means that the OLS estimator, as long as interaction terms between covariates and the treatment assignment are included, will always perform better than an unadjusted difference-in-means estimator, regardless of the correctness of the linear model. The OLS estimator has also previously been studied in the setting of interference. Chin (2019b) consider adjusting for covariates when the potential outcomes follow a linear-in-means model, but focuses on features engineered from the graph rather than pretreatment covariates. While the linear-in-means model is of interest, we aim at investigating the behavior of the OLS estimator when the linearity assumption is not assumed to hold both to preemptively address concerns similar to those raised by Freedman (2008), and to consider the usefulness of the OLS estimator in practice

where the linearity assumption is often hard to justify.

Chapter 2

Nonparametric Bounds in Two-Sample Summary-Data Mendelian Randomization: Some Cautionary Tales for Practice

2.1 Introduction

In recent years, genetic variants, often in the form of single nucleotide polymorphisms (SNPs), have been used as instrumental variables (IV) to estimate causal effects in epidemiological studies, often referred to as Mendelian randomization (MR) studies (Davey Smith and Ebrahim, 2003; Lawlor et al., 2008; Burgess and Thompson, 2015). Typically, MR studies are based on a two-sample design where published summary statistics from two independent genome wide association studies (GWAS), with one providing information about the exposure and the other about the outcome, are used (Burgess, Butterworth, et al., 2013; Burgess, Scott, et al., 2015; Davies et al., 2018). Under a two-sample design, investigators frequently use parametric methods to study exposure effects, for instance the IVW estimator (Burgess, Butterworth, et al., 2013), MR-Egger regression (Bowden, Del Greco M, et al., 2016), the weighted median estimator (Bowden, Davey Smith, et al., 2016),

MR-PRESSO (Verbanck et al., 2018) and MR-RAPS (Q. Zhao et al., 2020); see Burgess and Thompson (2015), Burgess, Small, et al. (2017), and Slob and Burgess (2020) for recent reviews.

An alternative approach to study exposure effects using instrumental variables without parametric assumptions is through nonparametric IV bounds (Balke and Pearl, 1997; Cheng and Small, 2006; Manski, 1990; T. S. Richardson and Robins, 2014; Robins, 1989). Briefly, nonparametric IV bounds use a minimum set of assumptions to provide a range of plausible values for the exposure effect. They are typically used when the outcome, the exposure, and the instrument are all binary and are simultaneously observed; we refer to this setting as the one-sample setting to contrast it from the two-sample setting. Arguably, the most well-known IV bound is the Balke-Pearl bound (Balke and Pearl, 1997) for the average treatment effect. The Balke-Pearl bound has been extended to allow for a non-binary instrument (Cheng and Small, 2006; T. S. Richardson and Robins, 2014) and two-sample designs (Ramsahai, 2012); see Swanson et al. (2018) and references therein for a recent summary of IV bounds.

Using IV bounds can be an attractive alternative to study exposure effects in MR studies given the strong parametric assumptions accompanying most MR analyses (Didelez and Sheehan, 2007; Swanson, 2017). However, to the best of our knowledge and compared to parametric methods, there is little work on actually using these bounds in typical MR settings, i.e. two-sample designs with summary statistics, nor any practical guidance on when the bounds would be informative. For example, what kind of genetic variants provide the most informative conclusions about the exposure effect in terms of the bounds not containing the null effect? Can combining multiple variants lead to shorter and tighter bounds? How do the bounds change if many instruments are weak, which is typical in MR studies? The overall goal of this paper is to offer some practical guidance on using IV bounds in two-sample MR studies. We focus on two aspects of bounds that will better inform MR investigators about the exposure effect: (1) the length of the bounds and (2) whether the bounds cover the null effect of zero (i.e. direction/sign of the effect).

The paper is organized as follows. Section 2 reviews notation, definition, and methods for studying the exposure effect with parametric models and nonparametric bounds. Section 3 presents our results where we show the behavior of the bounds in two-sample settings when we have one or multiple instruments, and when some of the core assumptions are violated. Section 4 quantifies how bounds from two-sample data is more conservative than bounds from one-sample data. Section 5 presents the data analysis and Section 6 lays out some concrete practical recommendations for using bounds in two-sample MR studies.

2.2 Methods

2.2.1 Review: Notation, Definitions, and Assumptions

Let X and Y be binary exposure and outcome variables, respectively, Z be a categorical instrumental variable taking values in $\{0, 1, 2\}$, and U be an unmeasured confounder for the effect of X on Y. We consider trivariate instruments due to the trivariate nature of SNPs that are typically used as instruments in MR studies. Also, following prior literature on bounds (Balke and Pearl, 1997; Manski, 1990; T. S. Richardson and Robins, 2014; Swanson et al., 2018), we consider the binary exposure and binary outcome setting. In Section 2.3.4 and Appendix 2.H, we discuss how to use the bounds to detect the sign of the effect when the exposure is continuous and the outcome is binary under some an additional assumptions. In general, non-trivial bounds (e.g. bounds that do not cover the entire support of the outcome) under "standard" IV assumptions stated below are impossible when the exposure or outcome is continuous. This limitation is well-known for bound-based analyses of causal effects and investigators who wish to obtain an effect estimate with a continuous exposure or outcome may have to make untestable, parametric modeling assumptions on top of the standard IV assumptions; see Swanson(Swanson et al., 2018) and Burgess and Labrecque (Burgess and Labrecque, 2018) for further discussions.

Let $Y^{z,x}$ be the potential outcome (Rubin, 1974; Splawa-Neyman, 1923) had the subject received exposure value $X = x$ and instrument value $Z = z$. We assume the sta-

ble unit treatment value assumption (SUTVA) (Cox, 1958; Rubin, 1980), formalized as $Y = \sum_{x,z} I[Z = z, X = x]Y^{x,z}$ where $I[\cdot]$ is the indicator function.

We make the following set of assumptions about X, Y, Z, and U that are found in MR studies (Didelez and Sheehan, 2007; Wang and Tchetgen Tchetgen, 2018):

(A1) *(Relevance)*: $Z \not\perp X$

(A2) *(Independent instrument)*: $Z \perp U$

(A3) *(Exclusion restriction)*: $Y^{z,x} = Y^{z',x} = Y^x$ for all x, z, z'

(A4) *(Conditional ignorability of X, Z given U)*: $Y^{z,x} \perp Z, X|U$

Briefly, (A1) can be satisfied by finding a SNP that has been consistently associated with the exposure. (A2) and (A3) are justified by scientific theory and can be violated if the SNP (i) is in linkage disequilibrium with an unmeasured SNP that affects the exposure and the outcome or (ii) has multiple functions beyond affecting the exposure (i.e. pleiotropic), to name a few. Pleiotropy is often a great concern in MR studies; we will consider violations of (A3) in Section 2.3.3. Finally, (A4) states that if U is observed, then it is sufficient to unconfound the relationship between X and Y. For much of the paper, we will assume (A1)-(A4) hold to focus on the behavior of the bounds, even though these assumptions are important to assess in practice.

We make some additional remarks about assumptions (A1)-(A4). First, in practice, most MR studies only explicitly state assumptions (A1)-(A3) along with some parametric modeling assumptions (Burgess and Thompson, 2015); see Section 2.2.2 below. Second, one can remove (A4) and strengthen (A2) with $Z \perp U, Y^{z,x}$ without consequence on the IV bounds (T. S. Richardson and Robins, 2014). Third, under SUTVA and assumptions (A3)-(A4), we have $Y \perp Z|X, U$, which is another common way to express the exclusion restriction in MR studies (Swanson et al., 2018; Didelez and Sheehan, 2007). Fourth, for simplicity, we do not assume the existence of a potential treatment X^z.

We define instrument strength ST as

$$\text{ST} = \max_{z \neq z'} |P(X = 1|Z = z) - P(X = 1|Z = z')| \tag{2.1}$$

ST reduces to the definition of instrument strength in Balke and Pearl's bounds when the instrument is binary. ST plays a critical role in determining the length of Balke and Pearl's IV bounds(Balke and Pearl, 1997). Also, (2.1) differs from other definitions of instrument strength based on a parametric model between the exposure and the outcome, such as the concentration parameter μ^2 (Stock, Wright, et al., 2002); the concentration parameter is roughly proportional to the observed first-stage F-statistic commonly used in linear IV models to assess instrument strength. But, under some assumptions, notably that the instrument is fixed, μ^2 and ST are related by the following formula.

$$\mu^2 = \frac{\text{ST}^2}{4} \sum_{i=1}^{n} z_i^2/\sigma^2.$$

Here $z_1, ..., z_n$ are observed values of the instrument, and σ^2 is the variance of the errors in the linear, first-stage reduced model; see Appendix 2.A for more details. The important take-away from the formula is that stronger instruments as measured by ST lead to larger values of the concentration parameter μ^2.

2.2.2 Review: Parametric Models in Two-Sample Studies

To better contrast the bound-based approaches we discuss below, we briefly review parametric models used to estimate exposure effects. Formally, in two-sample MR studies, a popular parametric model for a binary exposure (Lawlor et al., 2008; Burgess, 2014; Burgess and Thompson, 2012; Millard et al., 2019) is

$$\text{logit}(P(X = 1|Z_1 = z_1, ..., Z_p = z_p, U = u)) = \gamma_0 + \sum_i \gamma_i z_i + \gamma_U u, \tag{2.2}$$

and for a binary outcome (Burgess, 2014; Burgess and Thompson, 2012) is

$$\text{logit}(P(Y = 1|X = x, U = u)) = \beta_0 + \beta_X x + \beta_U u, \tag{2.3}$$

where $\text{logit}(a) = \log(a/(1-a))$. The parameter γ_i corresponds to the effect that instrument i has on the exposure. The summary statistic reported in GWAS is the coefficient from a simple logistic regression model, i.e. the model above where $p = 1$. This summary statistic is also approximately equal to the coefficient γ_i in equation (2.2) if the instruments are independent of each other and the coefficients $\gamma_1, ..., \gamma_p$ are small, which is the case in most two-sample MR studies(Manolio, 2010). The parameter β_X corresponds to the effect that the exposure has on the outcome in the logit scale; one can compute a numerical integral to compute the effect of the exposure on the outcome in the risk difference scale; see Section 2.2.3. The parameters γ_U and β_U correspond to the magnitudes that the unmeasured confounder U has on the exposure and outcome, respectively. Typically in the analysis of two-sample MR studies, U follows a parametric distribution and each SNP is often assumed to be in Hardy-Weinberg equilibrium. In our exposition below, we will relate the analysis from nonparametric bounds to these parametric models.

2.2.3 IV Bounds Under Two-Sample Designs and Goals of Paper

The most popular design in MR studies is a two-sample design which has two separate data sources, one providing information about (X, Z) in the form of $P(X = 1|Z = z)$, $z \in \{0, 1, 2\}$, and another providing information about (Y, Z) in the form of $P(Y = 1|Z = z)$, $z \in \{0, 1, 2\}$. A two-sample design differs from a more traditional one-sample design which has a single data source providing information on all observed variables (X, Y, Z) in the form of $P(Y = y, X = x|Z = z)$ or $P(X = x, Y = y, Z = z)$. IV bounds have been well-studied in one-sample designs and there is a rich array of guidance on how to use them in practice (Balke and Pearl, 1997; T. S. Richardson and Robins, 2014; Swanson et al., 2018). However, as noted in the introduction, not much is known about the behavior of IV bounds under a two-sample design.

The goal of this paper is to offer useful practical advice on using IV bounds to study the average treatment effect (ATE), defined as

$$\begin{aligned}\text{ATE} &= E[Y^1 - Y^0] \\ &= \int P(Y=1 \mid X=1, U=u)P(U=u)du - \int P(Y=1 \mid X=0, U=u)P(U=u)du\end{aligned}$$

using $P(Y=1|Z=z)$ and $P(X=1|Z=z)$, $z \in \{0,1,2\}$ obtained from a two-sample design. Specifically, under a two-sample design and assumptions (A1)-(A4), (Ramsahai, 2012) derived the following sharp bounds for the ATE:

$$\max \left\{ \begin{array}{ll} \max_{z \neq z'} & P(Y=1|Z=z) - 2 \cdot P(Y=1|Z=z') - 2 \cdot P(X=1|Z=z') \\ \max_{z \neq z'} & P(Y=1|Z=z) + P(X=1|Z=z) \\ & \qquad -P(Y=1|Z=z') - P(X=1|Z=z') - 1 \\ \max_{z \neq z'} & 2 \cdot P(Y=1|Z=z) + 2 \cdot P(X=1|Z=z) - P(Y=1|Z=z') - 3 \\ \max_{z} & -P(Y=1|Z=z) - P(X=1|Z=z) \\ \max_{z} & P(Y=1|Z=z) + P(X=1|Z=z) - 2 \end{array} \right\}$$

$$\leq ATE \leq \tag{2.4}$$

$$\min \left\{ \begin{array}{ll} \min_{z \neq z'} & P(Y=1|Z=z) - 2 \cdot P(Y=1|Z=z') + 2 \cdot P(X=1|Z=z') + 1 \\ \min_{z \neq z'} & P(Y=1|Z=z) + 2 \cdot P(Y=1|Z=z') - 2 \cdot P(X=1|Z=z') + 1 \\ \min_{z \neq z'} & P(Y=1|Z=z) - P(X=1|Z=z) \\ & \qquad +P(X=1|Z=z') - P(Y=1|Z=z') + 1 \\ \min_{z} & P(X=1|Z=z) - P(Y=1|Z=z) + 1 \\ \min_{z} & P(Y=1|Z=z) - P(X=1|Z=z) + 1 \end{array} \right\}$$

Furthermore, the assumptions imply the following checkable constraints, which are also referred to as IV inequalities (Balke and Pearl, 1997; Diemer et al., 2020), on the observed data.

$$\min \left\{ \begin{array}{ll} \min_{z \neq z'} & P(Y=1|Z=z) - P(X=1|Z=z) \\ & \quad -P(Y=1|Z=z') - P(X=1|Z=z') + 2 \\ \min_{z \neq z'} & P(Y=1|Z=z) + P(X=1|Z=z) \\ & \quad -P(Y=1|Z=z') + P(X=1|Z=z') \\ \min_{z} & P(X=1|Z=z) \\ \min_{z} & P(Y=1|Z=z) \\ \min_{z} & 1 - P(X=1|Z=z) \\ \min_{z} & 1 - P(Y=1|Z=z) \end{array} \right\} \geq 0 \qquad (2.5)$$

In equation (2.5), we see that the constraints from the law of probability are recovered (the last four expressions in equation (2.5)) along with 12 non-trivial constraints (the first two lines in equation (2.5)). Appendix 2.B provides additional discussion on equations (2.4) and (2.5) in two-sample MR studies.

We study two properties of the above bounds that can better guide practice: (1) the length of the bounds and (2) whether the bounds cover the null effect of zero. To better understand bound-specific characteristics not due to sampling errors, we will assume we have population-level quantities of $P(Y=1|Z=z)$ and $P(X=1|Z=z)$. In practice, these are estimated summary GWAS statistics from marginal logistic models. Specifically, the marginal proportions of the outcome, exposure, and allele frequencies are used to find the intercepts inside a logistic regression model by solving $P(X=1) = \sum_{z=0}^{2} \text{expit}(\gamma_0 + \hat{\gamma}_j \cdot z) P(Z_j = z)$ and $P(Y=1) = \sum_{z=0}^{2} \text{expit}(\Gamma_0 + \hat{\Gamma}_j \cdot z) P(Z_j = z)$ for γ_0 and Γ_0, respectively; here, $\hat{\gamma}_j$ and $\hat{\Gamma}_j$ are the estimated log odds ratio of the (marginal) associations from GWAS, and expit is the inverse of the logit function. This allows us to obtain estimates of $P(Y=1|Z_j=z)$ and $P(X=1|Z_j=z)$ for every j and $z=0,1,2$.

Finally, we remark that the population-level bounds do not depend on $P(Z=z), z=0,1,2$. In particular, whether a variant is rare or common has no influence on the bounds. However, rare variants may make it difficult to estimate the conditional probabilities which make up these bounds. Since we are only examining population-level characteristics of the bounds,

we will assume $P(Z_i = 0) = P(Z_i = 2) = 0.25$ and $P(Z_i = 1) = 0.5$ when we numerically illustrate our results below.

2.3 Properties of IV Bounds

2.3.1 Length of Bounds and Coverage of Null Effect

Theorem 2.1 characterizes the length of the IV bound in equation (2.4) under a two-sample design and assumptions (A1)-(A4).

Theorem 2.1. *Under assumptions (A1)-(A4) and a two-sample design, a sharp upper bound on the length of the bound in equation* (2.4) *is* $2 - 2 \cdot ST$*, i.e. there exists a data generating process satisfying (A1)-(A4) and has length equal to* $2 - 2 \cdot ST$*.*

See Appendix 2.C for the proof, which extends Theorem 2.1 to instruments with 2, 3, or 4 categories. Compared to the Balke-Pearl IV bounds under a one-sample design where the length is $1-$ST for a binary or three-leveled IV(Balke and Pearl, 1997; T. S. Richardson and Robins, 2014), the length of two-sample IV bounds can be twice as long. Also, the length of two-sample IV bounds is only guaranteed to be less than 1 if the instrument strength ST is greater than 0.5; note that this does not imply that instruments with ST less than 0.5 has length greater than 1. In contrast, one-sample IV bounds always have length less than 1 unless ST is zero. In short, there is a cost of using a two-sample MR design instead of a one-sample MR design when performing a bound-based analysis of the ATE.

Figure 2.1a numerically illustrates the consequences of Theorem 2.1 by calculating the bounds in equation (2.4) from 10,000 randomly generated sets of values of $P(X = 1|Z = z)$ and $P(Y = 1|Z = z)$ that satisfy the IV inequalities and assumptions (A1)-(A4). We also use two real-world data examples where the causal effects are known to exist: the effect of high cholesterol on incidence of heart attacks (Cholesterol Treatment Trialists' (CTT) Collaborators, 2012), and the effect of smoking on incidence of lung cancer (Cornfield et al., 1959). We see that the length of the bounds often exceed 1 as the instrument strength

decreases. Also, the two real-world studies generally do not lead to bounds with length less than 1.

Figure 2.1b further illustrates this point by characterizing the relationship between ST and the summary statistic coefficient γ_1 from the logistic exposure model in Section 2.2.2 when $p = 1$ and U following a standard Normal. Specifically, γ_U was set to $0.1, 0.5, 1$ and 2, γ_1 varied between 0.2 and 6, and $\gamma_0 = -\gamma_1$. We see that instrument strength ST of 0.5 corresponds to a coefficient γ_1 of approximately $1.1, 1.16, 1.4$ and 1.8 if γ_U is $0.1, 0.5, 1$ and 2, respectively. Coefficients with such magnitudes are rare in GWAS where genetic variants often explain a small amount of variation in the exposure. More broadly, these values of γ_1 correspond to odds ratios between 3 and 6 and exceed some well-known magnitudes of causal effects in cancer studies, say the effect of exposure to ultraviolet radiation on the incidence of skin cancer where the odds ratios are estimated to be in the range from 1.4 to 2.22 (Schmitt et al., 2011).

Next, we examine what kind of γ_1 is needed in order for the two-sample IV bounds to exclude the null effect of zero for a specified effect size of the ATE. This question is akin to computing the power of bounds but with population-level quantities. We reuse the setup for the exposure model described above and the logistic outcome model specified in Section 2.2.2. Specifically, the coefficients for the exposure model are the same as before. For the outcome model, we vary β_X from 0.25 to 6 and set $\beta_0 = -\beta_X/2$. Then, for each parameter specification in the outcome model, we compute the corresponding ATE. Afterwards, we find the smallest γ_1 that leads to a bound that does not cover 0, but covers the ATE; see Appendix 2.D for more details. Figure 2.2 shows that even for a moderate effect size of 0.4, the corresponding γ_1 must be around 2, a tall order for most GWAS. Also, as the effect of unmeasured confounding increases via γ_U, a larger γ_1 is needed to exclude the null effect. In short, analyzing the ATE using bounds in a two-sample MR study is unlikely to be informative; the bounds will often have length greater than 1 and rarely exclude the null effect unless very strong genetic variants are used.

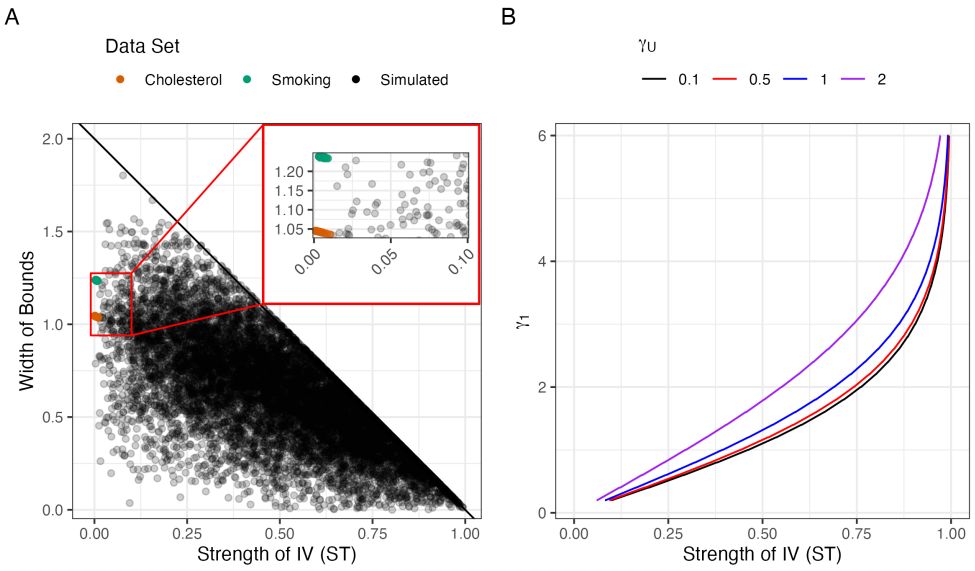


Figure 2.1: Illustration of the relationship between instrument strength, length of bounds, and coefficients from logistic regression model in two-sample MR settings. A: Relationship between instrument strength (ST) and length of the IV bounds. Black line is the upper bound on the two-sample IV bounds from Theorem 1. Black dots indicate one of the 10,000 IV bounds. Colored dots indicate bounds from real data; see Section 2.5 for details. B: Coefficients from logistic regression model and instrument strength (ST). Each color represents different magnitudes of unmeasured confounding.

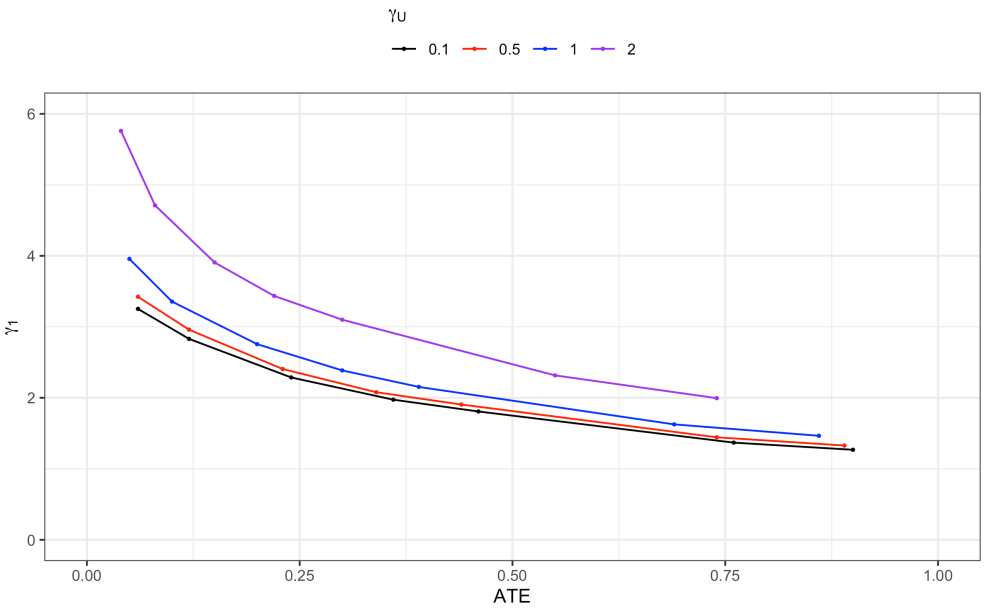


Figure 2.2: Relationship between the smallest γ_1 needed for a two-sample IV bound to exclude 0 and the average treatment effect (ATE). Each color corresponds to different levels of unmeasured confounding.

2.3.2 Would Multiple Instruments Help?

Based on the results above with a single instrument, a natural question from investigators is whether using multiple instruments can lead to more informative bounds for the ATE(Swanson, 2017). For example, suppose we aggregate two-sample IV bounds across multiple instruments by taking intersections of individual IV bounds; we refer to this as intersection bounds. This approach may be inferior to another alternative where we expand the levels of Z to accommodate multiple instruments(Swanson, 2017), but has the benefit of being applicable to most two-sample MR studies with summary statistics from GWAS. In particular, with typical summary-level data from two-sample MR studies where only marginal associations are reported, it is not possible to expand the levels of Z to accommodate multiple instruments. As we show here, the strongest instrument often determines the length of the intersection bound because the bounds from each instrument exhibit a nesting property. That is, using one bound with the strongest instrument often provides the same

amount of information about the ATE as the intersection of several bounds from multiple instruments. We will illustrate this numerically in this section; Appendix 2.E provides the exact technical conditions that lead to the nesting property.

Suppose for a moment that the true model for the data follows the models in Section 2.2.2. We consider either $p = 10$ instruments or $p = 50$ instruments with the following γ_i values:

1. *Many weak instruments*: γ_i are spread out evenly on the interval 0 to 0.2.
2. *Many strong instruments*: γ_i are spread out evenly on the interval 1 to 4.
3. *Many very weak instruments, one medium strength instrument*: γ_i, $i = 1, 2, ..., p-1$, are evenly spread out on the interval 0 to 0.01, and $\gamma_p = 0.2$.
4. *Many medium strong instruments, one strong instrument*: γ_i, $i = 1, 2, ..., p-1$, are evenly spread out on the interval 1 to 1.2, and $\gamma_p = 4$.

The first scenario mimics typical magnitudes of γ_i seen in MR studies where many genetic traits weakly contribute to the expression of complex traits (Loh et al., 2015; Shi et al., 2016; Nj et al., 2017). The third scenario represents a genetic architecture where only few genetic variants have strong effects on the exposure and the rest have weak effects (Yang et al., 2010). Scenarios 2 and 4 are similar to scenarios 1 and 3, but with larger coefficients. We don't expect to observe scenarios 2 and 4 in practice, but the magnitudes of γs in these scenarios were shown from Section 2.3.1 to produce informative bounds when $p = 1$.

For each scenario, we use monte carlo integration to obtain $P(X = 1|Z_j = z_j)$ and $P(Y = 1|Z_j = z_j)$ from the data generating model. We then use these quantities to obtain two-sample IV bounds for each of the p instruments. Figure 2.3 shows the results for $\beta_X = 0.25$ and $\beta_X = 1$, but similar trends are observed for $\beta_X = 0.5$, $\beta_X = 1.5$, or $\beta_X = 2$; see Appendix 2.F. In all scenarios, the bounds are nested within each other. Thus, if we were to aggregate bounds by taking intersections, the length of the intersection bounds will only be as strong as the bounds from the strongest instrument. Also, all bounds are non-informative, except for scenario 4 when $p = 10$ and $\beta_X = 1$.

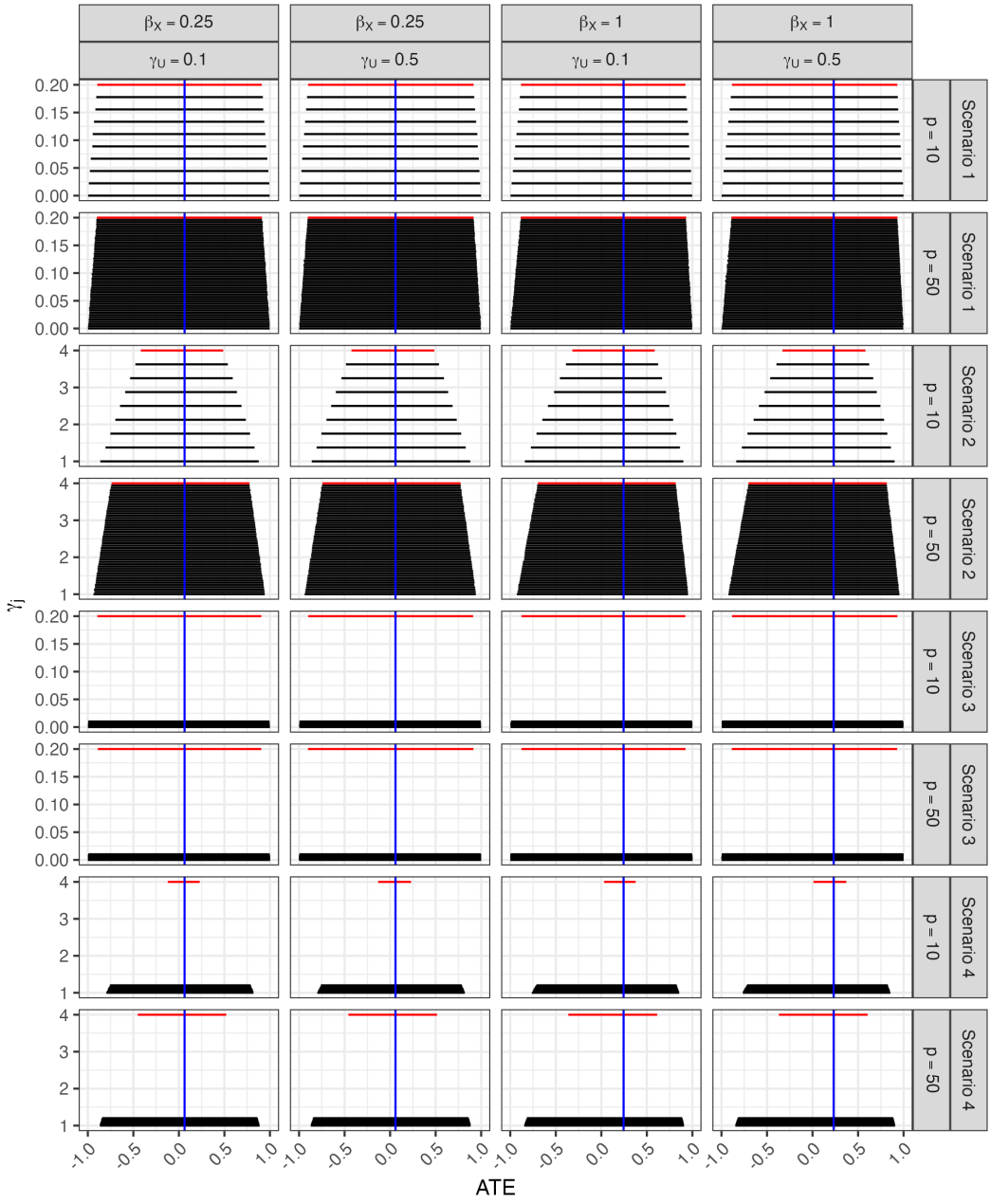


Figure 2.3: Two-sample bounds with $p = 10$ or $p = 50$ instruments. Bounds from strongest instruments are highlighted in red. Blue lines denote the true average treatment effects (ATEs). Columns represent effect size of the exposure and the unmeasured confounder on the outcome on the logit scale. Rows represents different scenarios of multiple instruments. The y-axis represents instrument strength measured by γ_j and the x-axis represents the average treatment effect.

Our results also have dire implications when some instruments turn out to be invalid. If, as suggested by Swanson (2017), we take the union of IV bounds instead of intersections so that the union bound is guaranteed to cover the true ATE so long as there is at least one valid instrument, the union bound will likely be non-informative because there was at least one IV bound in our scenario that was non-informative. Without making some assumptions about the nature of the invalid IVs when multiple IVs are used, a bound-based analysis will likely not reveal any useful information about the ATE.

Overall, combining our results in Section 2.3.1, our conclusion about using nonparametric IV bounds in two-sample MR studies is grim. A useful bound-based analysis would require very strong instruments and/or effect sizes; relatedly, the instruments must be stronger than those from one-sample studies. Also, multiple instruments are no better than having a single, strong instrument.

2.3.3 Pleiotropy

A major concern in MR studies is pleiotropy, which is a violation of the exclusion restriction (A3). In particular, a practical concern is that when (A3) is violated, the two-sample IV bounds may still produce bounds about the ATE, say in terms of length or detecting the sign of the effect, and mislead investigators about the magnitude or direction of the ATE. Or, the verifiable constraints that are part of the bounds may fail to detect the violation of (A3) and again, mislead investigators about the presence of an invalid instrument. To this end, we reuse the exposure model specified in (2.2) with $p = 1$, and use the following outcome model:

$$\text{logit}(P(Y = 1|X = x, Z = z, U = u)) = \beta_0 + \beta_X x + \beta_Z z + \beta_U u. \tag{2.6}$$

We set the coefficients $\beta_X \in \{-2, -1, 0, 1, 2\}, \beta_Z \in \{-0.5, -0.25, -0.1, 0, 0.1, 0.25, 0.5\}, \beta_0 = -\beta_X/2, \gamma_1 \in \{-0.5, -0.25, -0.1, 0, 0.1, 0.25, 0.5\}$, and $\gamma_0 = -\gamma_1$ while $\gamma_U = \beta_U = 1$. Figure 2.4 shows the results. In every single scenario, the bounds do in fact cover the ATE. That

is, weak instruments effectively dominate the behavior of the bounds, including any biases that may be incurred from a pleiotropic instrument, and produce wide bounds. Also, the verifiable constraints in equation (2.5) are never violated, suggesting that they are limited in their ability to detect violations of the assumptions when instruments are weak and in general, subject-matter knowledge may be a more powerful argument for (or against) the validity of an instrument. Appendix 2.G provides additional results concerning the effect of pleiotropy on the bounds.

2.3.4 Dichotomizing a Continuous Exposure and Effect on Two-Sample Bounds

In many MR studies, the exposure is often a continuous variable and using a bound-based analysis on such studies require dichotomizing the exposure variable. For example, in Section 2.5 where we study the effect of cholesterol on heart disease, we use a dichotomized exposure variable in order to use a bound-based analysis. Dichotomizing the exposure variable raises many important questions ranging from the interpretability of the estimator for the exposure effect to whether the estimator is actually estimating a causal quantity(Burgess and Labrecque, 2018). In this section, we show the effect of dichotomization on two-sample bounds by showing that if the exposure has a monotonic effect on the outcome, the two-sample bounds using a dichotomized exposure can be used to detect the sign of the underlying causal effect.

Let $\widetilde{X}$ be a continuous exposure, and $\widetilde{Y}^{\tilde{x}}$ be a potential outcome under the continuous exposure $\tilde{x}$. We define the dichotomization of the exposure as $X = 1[\widetilde{X} \geq c]$ for some c, and link the potential outcome under a binary exposure and the continuous exposure as $Y^1 = \widetilde{Y}^{\tilde{x}}$ for $\tilde{x} \geq c$ and $Y^0 = \widetilde{Y}^{\tilde{x}}$ for $\tilde{x} < c$. Without loss of generality, under the monotonicity assumption where the outcome increases as the exposure increases, i.e. $P(\widetilde{Y}^{\tilde{x}} \leq \widetilde{Y}^{\tilde{x}+\epsilon}) = 1$ for all $\tilde{x}$ and $\epsilon > 0$, we have for any $\tilde{x}' > \tilde{x}$, $\text{sign}\left(E[\widetilde{Y}^{\tilde{x}'} - \widetilde{Y}^{\tilde{x}}]\right) = \text{sign}\left(E[Y^1 - Y^0]\right) = 1$; see Appendix 2.H for a formal argument.

We numerically illustrate the result through a simulation study. We use a linear model for

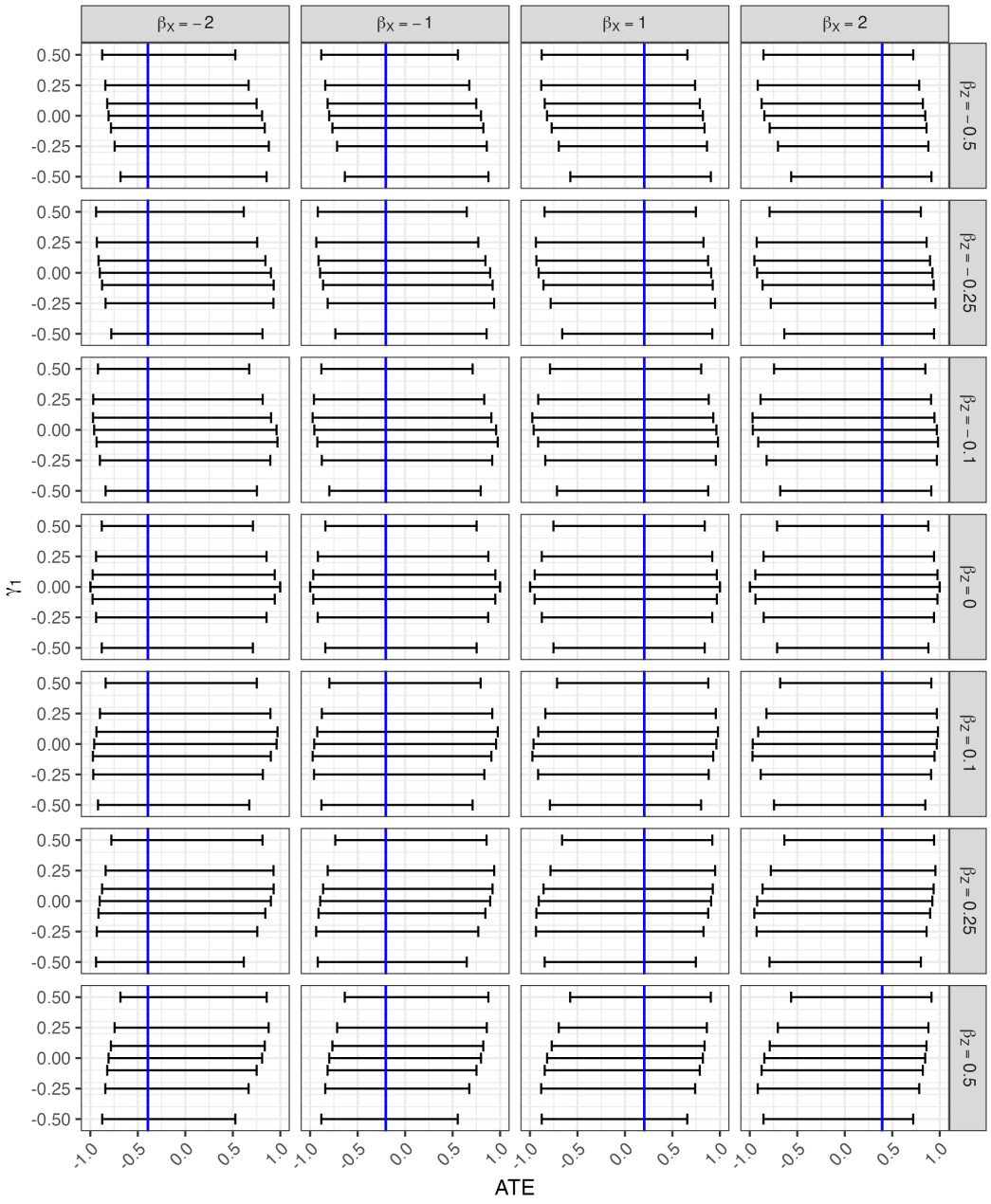


Figure 2.4: Two-sample bounds (horizontal lines) and average treatment effects (vertical blue lines) under pleiotropy. Columns represent the effect size of the exposure on the logit scale, rows represent the magnitude of violation of assumption (A3). The x-axis shows average treatment effect (ATE), and the y-axis represents instrument strength as measured by γ_i.

the continuous exposure with $\gamma_0 = 0$, and a logistic model for the outcome. We generate a data set with $10,000,000$ observations, dichotomize the exposure using the observed median of the exposure, and then calculate the nonparametric two-sample IV bounds. Figure 2.5 show parts of the results; the full set of results can be found in Appendix 2.H. We see that whenever the two-sample IV bounds based on the dichotomized exposure allow us to make conclusions about the direction of the exposure effect, the inferred direction is the same as the direction of the effect of the continuous exposure.

2.3.5 Finite-Sample Behavior of Bounds With Estimated Probabilities

The nonparametric bounds discussed above are all derived assuming population level probabilities are available. In practice, we estimate the probabilities from a sample using logistic regression; see Section 2.2.3, Section 2.5, Appendix 2.J, and Appendix 2.H.3. In particular, in Appendix 2.H.3, we show numerically that for all practical values of instrument strength in MR, incorporating additional uncertainty from estimation leads to bounds with 100% empirical confidence levels. Only when the instruments are implausibly strong do we see loss in coverage. This result is not surprising since we have shown above that two-sample IV bounds without accounting for estimation error are often wide and non-informative and adding additional uncertainty from estimation will inevitably enlarge these bounds; see Section 7 in Swanson et al. (2018) for a similar observation.

2.4 Characterizing the Loss of Information in Two-Sample MR Studies

As hinted in Theorem 2.1, the increase in the bound's length is an inevitable "cost" of using two-sample designs instead of one-sample designs in MR studies. This section examines this loss of information by creating a plausible range of the joint distribution of the outcome and

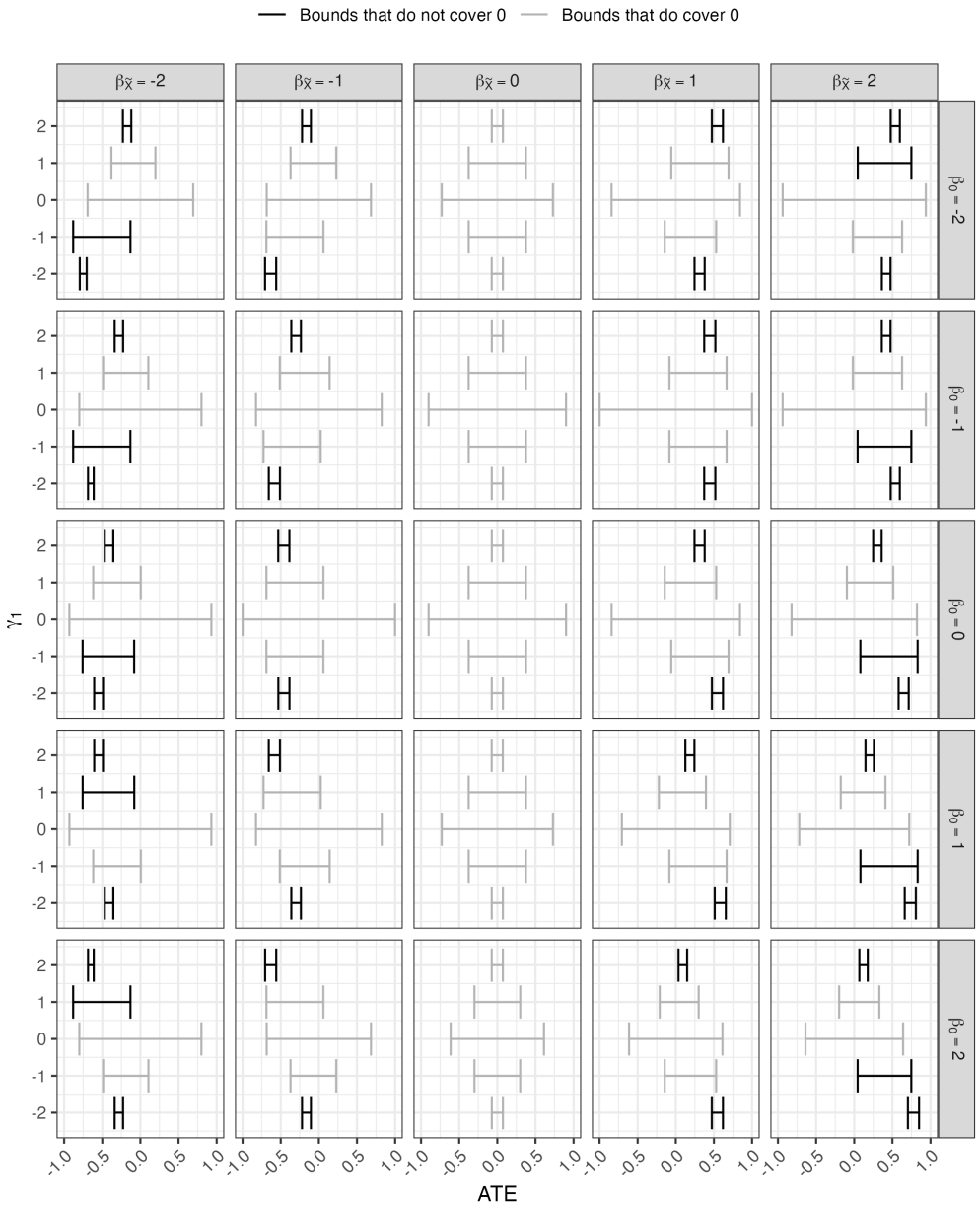


Figure 2.5: Nonparametric bounds based on a dichotomized exposure. Columns represent the effect size of the exposure on the logit scale. Rows represent different values of the intercept β_0. The y-axis shows the effect of the instrument on the continuous exposure, and the x-axis shows the average treatment effect.

the exposure given the instrument, $P(X = x, Y = y|Z = z)$, based on the observed data from two-sample MR studies; as mentioned before, this joint distribution determines the IV bounds in one-sample designs and seeing how much of it can be inferred from two-sample designs provide us with a way to characterize the loss of information from using two-sample designs.

Formally, the joint conditional distribution $P(X = x, Y = y|Z = z)$ can be written as a function of the marginal conditional distributions $P(X = x|Z = z)$ and $P(Y = y|Z = z)$ from two-sample MR studies and the conditional covariance of the exposure X and outcome Y given the instrument Z, i.e. $\text{Cov}(X, Y|Z = z)$, via the following formula

$$P(X = x, Y = y|Z = z) = P(X = x|Z = z)P(Y = y|Z = z) + (2 \cdot I[x = y] - 1)\text{Cov}(X, Y|Z = z). \tag{2.7}$$

Because $\text{Cov}(X, Y|Z = z)$ is impossible to estimate from two-sample MR studies, we take a random sample from -1 to 1, akin to placing a flat, uniform prior on -1 to 1. The sampled value of $\text{Cov}(X, Y|Z = z)$ must not only produce a proper probability distribution of $(X, Y|Z)$, but also satisfy the verifiable constraints from the IV assumptions. Specifically, $\text{Cov}(X, Y|Z = z)$ must satisfy

$$\max_z \left\{ \begin{array}{c} -P(X = 1|Z = z)P(Y = 1|Z = z) \\ -P(X = 0|Z = z)P(Y = 0|Z = z) \\ P(X = 1|Z = z)P(Y = 0|Z = z) - 1 \\ P(X = 0|Z = z)P(Y = 1|Z = z) - 1 \end{array} \right\}$$

$$\leq \text{Cov}(X,Y|Z = z) \leq$$

$$\min_z \left\{ \begin{array}{c} 1 - P(X = 1|Z = z)P(Y = 1|Z = z) \\ 1 - P(X = 0|Z = z)P(Y = 0|Z = z) \\ P(X = 1|Z = z)P(Y = 0|Z = z) \\ P(X = 0|Z = z)P(Y = 1|Z = z) \end{array} \right\}$$

Also, for any pair of $(z, z') \in \{0, 1, 2\} \times \{0, 1, 2\}$, the values of $\text{Cov}(X, Y|Z = z)$ and

$\text{Cov}(X, Y|Z = z')$ must satisfy

$$\max \left\{ \begin{array}{c} -P(X=0|Z=z)P(Y=0|Z=z) - P(X=0|Z=z')P(Y=1|Z=z') \\ P(X=1|Z=z)P(Y=0|Z=z) + P(X=1|Z=z')P(Y=1|Z=z') - 1 \\ P(X=0|Z=z')P(Y=0|Z=z') + P(X=0|Z=z)P(Y=1|Z=z) - 1 \\ -P(X=1|Z=z')P(Y=0|Z=z') - P(X=1|Z=z)P(Y=1|Z=z) \end{array} \right\}$$

$$\leq \text{Cov}(X, Y|Z = z) - \text{Cov}(X, Y|Z = z') \leq$$

$$\min \left\{ \begin{array}{c} 1 - P(X=0|Z=z)P(Y=0|Z=z) - P(X=0|Z=z')P(Y=1|Z=z') \\ P(X=1|Z=z)P(Y=0|Z=z) + P(X=1|Z=z')P(Y=1|Z=z') \\ P(X=0|Z=z')P(Y=0|Z=z') + P(X=0|Z=z)P(Y=1|Z=z) \\ 1 - P(X=1|Z=z')P(Y=0|Z=z') - P(X=1|Z=z)P(Y=1|Z=z) \end{array} \right\}$$

We sequentially sample values of $\text{Cov}(X, Y|Z = 0), \text{Cov}(X, Y|Z = 1), \text{Cov}(X, Y|Z = 2)$, such that the above inequalities are satisfied. Among such samples, we calculate the joint distribution of $P(X = x, Y = y|Z = z)$ using (2.7), leading us to a plausible set of values for the joint distribution $P(X = x, Y = y|Z = z)$.

For each plausible joint distribution $P(X = x, Y = y|Z = z)$, we use the one-sample IV bounds (Balke and Pearl, 1997; T. S. Richardson and Robins, 2014) to obtain a bound for the ATE. If a large number of one-sample IV bounds obtained from this procedure does not cover zero, then there is some evidence for a non-zero exposure effect and a one-sample MR study may yield informative bounds on the ATE. However, if a large number of the one-sample IV bounds covers zero, there is little hope of obtaining information about the ATE if we used a one-sample MR design; in other words, the one-sample IV bounds are equally likely to be conservative as the two-sample IV bounds. This approach can be extended to intersection bounds based on multiple instruments; see Appendix 2.I. For convenience, the approach is implemented as a Shiny web application for investigators to use at https://rtrane.shinyapps.io/potential-one-sample-bounds.

Table 2.1 presents nine different sets of values of the marginal distributions $P(Y|Z)$ and $P(X|Z)$ that investigators could theoretically obtain from hypothetical two-sample MR

studies. Figure 2.6 shows the one-sample IV bounds from the procedure we illustrated above.

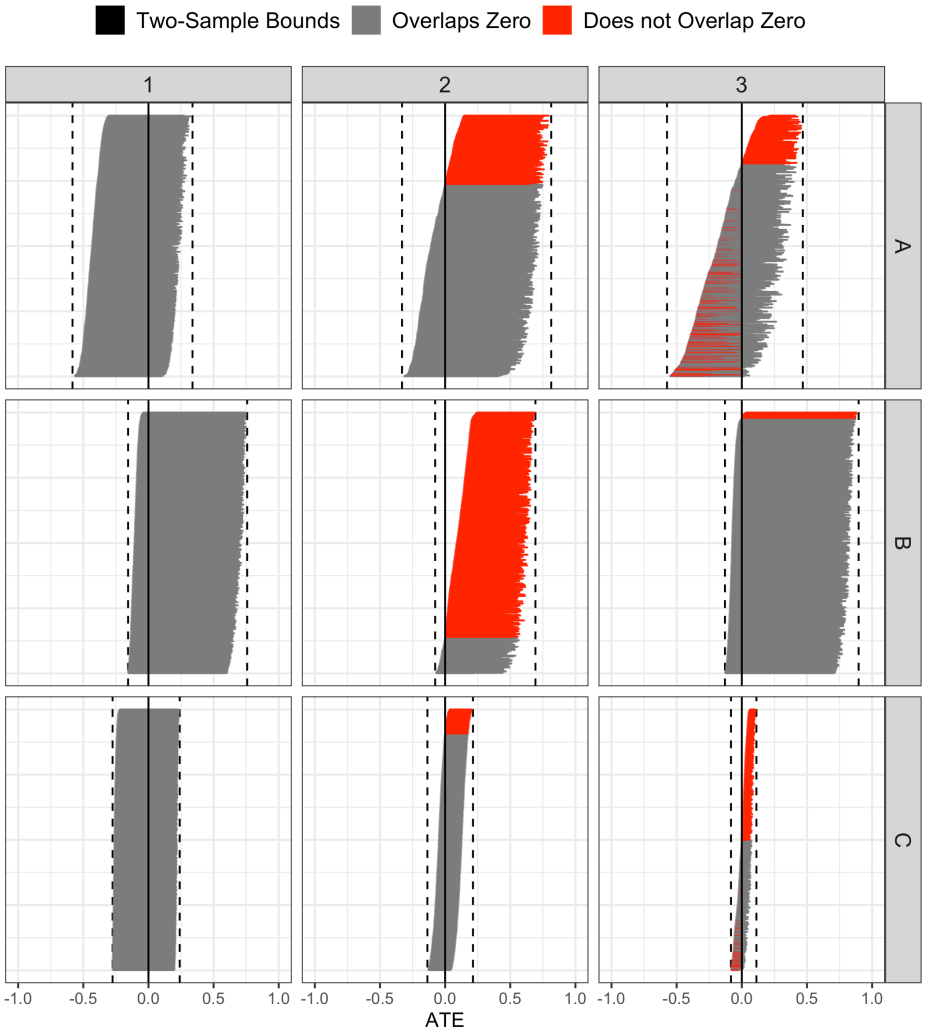


Figure 2.6: One-sample bounds (horizontal lines) and two-sample bounds (vertical dotted lines). Red color represents one-sample bounds that do not cover zero and gray color represents one-sample bounds that do cover zero.

Row A of Figure 2.6 shows three scenarios where the two-sample bounds are all centered close to zero with similar length. But, the conclusions from the one-sample bound analysis are rather different. Column 1 shows no one-sample bounds would allow us to determine the presence of a non-zero exposure effect. Column 2 indicates that about 26.3% of the

one-sample IV bounds do not contain 0 while for column 3 that number is approximately 35.9%. However, the latter includes one-sample bounds entirely above and below 0.

Row B illustrates three scenarios where the two-sample bounds are centered well above zero and have large length. We see one case where we have no hope of determining direction of the ATE from the one-sample bounds (column 1), one case where we are most likely to determine the ATE's direction (column 2), and one case where we are unlikely to determine the ATE's direction (column 3).

Row C is similar to row A in that all the two-sample bounds are centered around 0, but the lengths of the two-sample bounds are narrow. The three columns indicate similar conclusions as row A, showing that even with rather narrow two-sample bounds centered around 0, the one-sample bounds may not reveal information about presence or the direction of the exposure effect.

Overall, the proposed procedure and the examples above show that a bound-based analysis could have been useful had we used a one-sample design compared to a two-sample design. Nevertheless, we mention a word of caution when interpreting the results above, especially concerning the sampling of the covariance values. For example, a scenario like the one resulting in the bounds presented in row B, column 2 only provides information about the plausibility of different one-sample bounds; it does not provide the probability of each bound not covering zero. However, if the true ATE is in fact negative, the proposed procedure does rule out the possibility of one-sample bounds being able to ascertain this because all one-sample bounds covering a negative ATE also covers 0.

2.5 Using Bound-Based Analysis in Two Positive Control Examples

We demonstrate our findings about the behavior of two-sample IV bounds on two real MR studies. Our first study examines the effect of smoking on incidence of lung cancer and our

second study examines the effect of self-reported high cholesterol status on incidence of heart attack. The effect of smoking on lung cancer is known to be strong and positive (Smoking and Health, 1964). Also, while the exact mechanism between high cholesterol and heart disease is still being discussed (Holmes et al., 2015; T. G. Richardson et al., 2020), some meta-analyses of randomized clinical trials on the effect of cholesterol-lowering medication suggest a strong causal relationship (Cholesterol Treatment Trialists' (CTT) Collaborators, 2012; Cholesterol Treatment Trialists' (CTT) Collaborators, 2005). In both cases, we assess what conclusions can be obtained by using bound-based analyses in studies where the causal effects are strong and positive.

The study data were obtained from the UK Biobank data stored in the Integrative Epidemiology Unit (IEU) GWAS database. We use the `TwoSampleMR` R package (Hemani et al., 2018) with the recommended defaults to extract and clean the data. For more details, see Appendix 2.J.

For the effect of smoking on lung cancer, we used 84 genetic instruments, and for the effect of cholesterol on heart attack, we used 54 genetic instruments. The average instrument strengths were 0.0042 (range: 0.0032 to 0.0091) for smoking and 0.0005 (range: 0.0002 to 0.0022) for cholesterol; these values are much smaller than the ST = 0.5 needed to guarantee bounds with length less than 1. As such, the two-sample bounds in Figure 2.7 are wide; all of them have length greater than 1 and they convey no information about the causal effects of interest. Additionally, using our method from Section 2.4, the direction of the ATE would not have been detectable had we used a one-sample design; see Figure 2.8. Appendix 2.J contains additional analysis, notably demonstrating that aggregating multiple bounds through intersections are also non-informative.

Overall, while nonparametric bounds allow us to not make parametric assumptions frequent in two-sample MR analyses, they may provide little, if any, information about the exposure effects, even if the exposure effect is known to be positive and strong. Additionally, since many two-sample MR studies involve weak instruments, we believe bound-based approaches will likely have limited practical value to uncover causal effects.

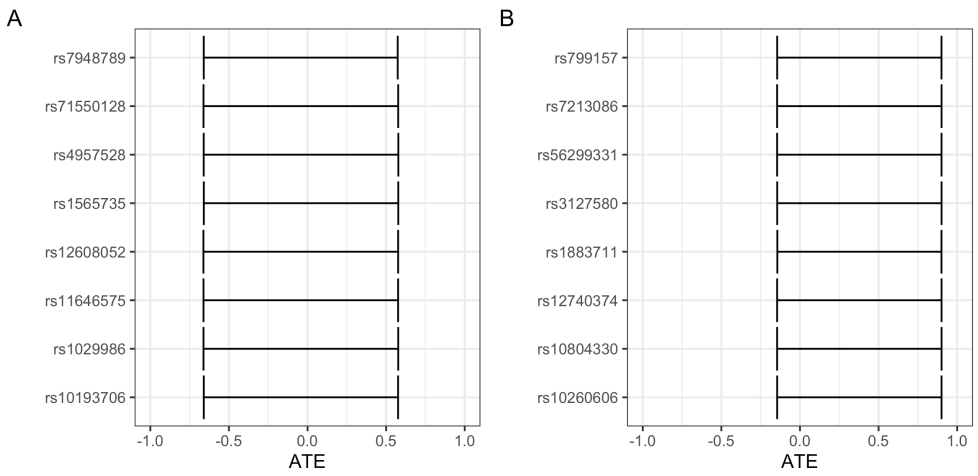


Figure 2.7: Two-sample IV bounds for the two real data examples with 8 SNPs from each data set. A: Two-sample IV bounds for the ATE of smoking on the incidence of lung cancer. B: Two-sample IV bounds for the ATE of high cholesterol on the incidence of heart attack.

2.6 Discussion

Nonparametric bounds are without a doubt an attractive concept. With a minimal set of assumptions, they let investigators obtain bounds on the average treatment effect. However, as we have seen above, in typical MR studies with two-sample summary data, a bound-based analysis may generally be uninformative for two reasons. First, while IV bounds in one-sample settings have length always less than 1, in two-sample settings, this is not always the case, and the bounds are often more conservative. Second, many genetic variants in MR studies are too weakly associated with the exposure to produce bounds with length less than 1 or bounds that exclude 0. Indeed, our two real data examples showed that despite having strong causal effects, bound-based analyses were unable to detect these effects.

We also outlined an approach to roughly quantify the information loss going from one-sample designs to two-sample designs and to assess the range of conclusions that can be drawn if we had one-sample data. We demonstrate our method to a few different settings of two-sample data and showed the range of conclusions that can be drawn about the plausible one-sample nonparametric bounds. Investigators can also use our Shiny web application to

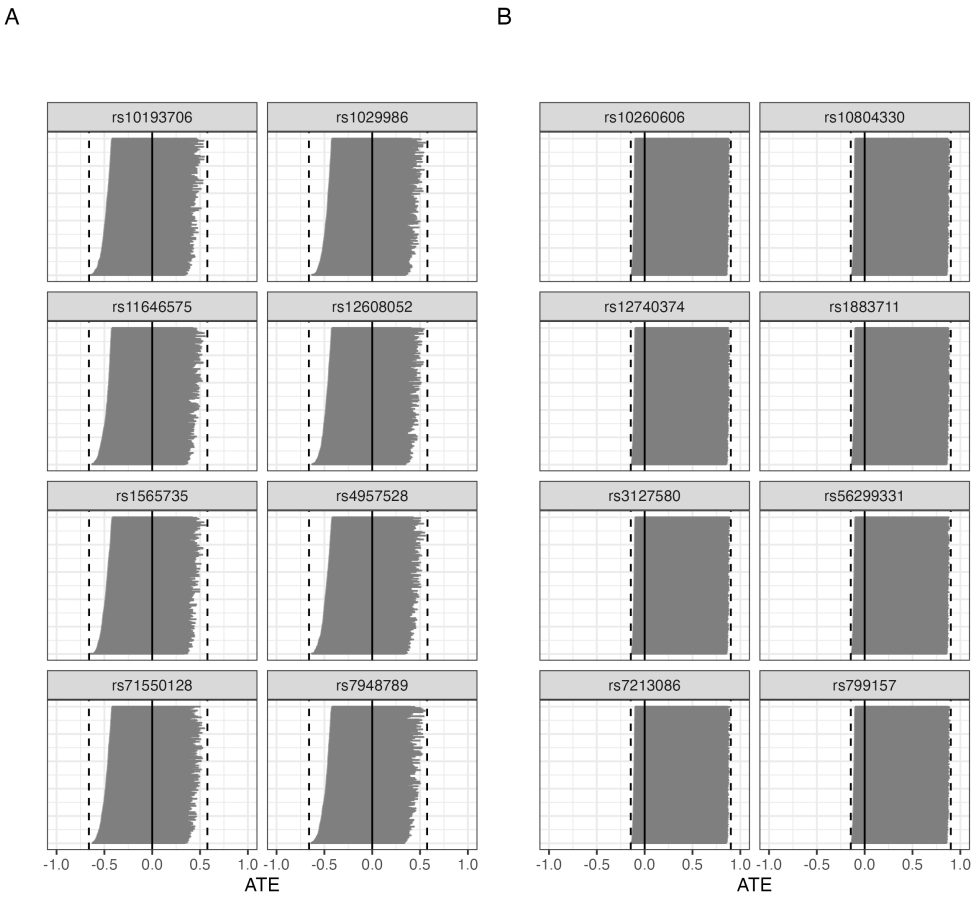


Figure 2.8: Potential one-sample IV bounds for the two real data examples using the method described in Section 2.4. A: One-sample IV bounds for the ATE of smoking on the incidence of lung cancer from 500 potential one-sample distributions for each SNP. B: One-sample IV bounds for the ATE of high cholesterol on the incidence of heart attack from 500 potential one-sample distributions for each SNP.

compare one-sample and two-sample designs for bound-based analysis.

Overall, our general recommendation for practice is that unless investigators have a very strong instrument, ideally exceeding ST > 0.5, bounds will unlikely be useful as a nonparametric analysis of the ATE, even with multiple instruments. Even if ST > 0.5, one would need strong IVs and/or strong effect sizes to make sure that the bounds do not cover 0. Finally, investigators can use our procedure above to assess whether it is worthwhile to use a one-sample MR design over a more typical (and arguably easier) two-sample MR design as the bounds under a one-sample design is generally less conservative than bounds from a two-sample design.

Nevertheless, there may be few limited, but meaningful use cases for using bounds to study the ATE in two-sample MR studies; see Diemer et al. (2020) for one example based on IV inequalities. First, when one has prior knowledge about the direction of the effect, but wish to get a better sense of its magnitude, nonparametric bounds can provide an upper limit on this magnitude. For example, when the exposure is known to cause harm or benefit, say in our smoking example, an upper bound on this effect would tell investigators about the maximum possible effect that smoking could have on increasing the incidence of lung cancer. Second, two-sample IV bounds can be used to check estimates from parametric methods to see if they lie inside of the bounds; if the estimates lie outside of the bounds, then the parametric models underlying the estimates are likely mis-specified.

We also note that as part of our study into two-sample bounds, we encountered bounds where the upper bound is smaller than the lower bound in the two-sample IV bounds above as well as existing formula for one-sample IV bounds with multi-level instruments; see Appendix 2.B for more details. In particular, we never see this behavior when the instrument is binary for both one- and two-sample data. Also, when the instrument takes on three values, we never encounter this scenario in one-sample bounds. But, when the instruments take on three values and we have two-sample data, we do see it in 0.84% of the $10,000,000$ unique probability distributions we tried. When the instrument takes on four categories, this behavior occurs in 0.04% of $10,000,000$ unique probability distributions for one-sample

data and 1.27% of $10,000,000$ unique probability distributions for two-sample data. Note that all of these bounds passed the existing falsification constraints (equation (2.5)). Our conjecture is that existing works on one-sample and two-sample bounds are correct, but the existing falsification inequalities derived from these bounds may not be tight enough under non-binary instruments to detect potential violations of the the IV assumptions. That is, both the one-sample and two-sample bounds can be computed irrespective of whether the IV assumptions are satisfied or not. But, the current falsification inequalities under non-binary IV settings may not leverage all parts of the observed data to detect violations of the IV assumptions and it may be possible to use this behavior in the resulting bounds as another falsification test in non-binary IV settings. We leave this interesting topic as future research.

Table 2.1: Values of $P(X = 1|Z = z)$ and $P(Y = 1|Z = z)$ used to illustrate our approach. For each cell (e.g. row A, column 1), we have
$\{P(X = 1|Z = 0), P(X = 1|Z = 1), P(X = 1|Z = 2)\}$ on the first row and
$\{P(Y = 1|Z = 0), P(Y = 1|Z = 1), P(Y = 1|Z = 2)\}$ on the second row.

	Column 1	Column 2	Column 3
Row A	{0.125, 0.399, 0.080}	{0.244, 0.275, 0.185}	{0.603, 0.469, 0.310}
	{0.699, 0.840, 0.742}	{0.238, 0.089, 0.146}	{0.638, 0.346, 0.719}
Row B	{0.886, 0.968, 0.874}	{0.139, 0.441, 0.334}	{0.901, 0.909, 0.935}
	{0.805, 0.822, 0.951}	{0.179, 0.359, 0.559}	{0.821, 0.810, 0.905}
Row C	{0.175, 0.079, 0.365}	{0.493, 0.911, 0.085}	{0.434, 0.045, 0.733}
	{0.599, 0.358, 0.087}	{0.360, 0.480, 0.441}	{0.747, 0.370, 0.169}

Appendices

This is the Appendix to our paper "Non-parametric Bounds in Two-Sample Summary-Data Mendelian Randomization: Some Cautionary Tales for Practice". This includes additional details on how the ST measure relates to the concentration parameter; how we obtain bounds on the Average Treatment Effect; proof of Theorem 2.1; additional details and results for the simulation behind the "power" analysis presented in Section 2.3.1; the technical conditions that lead to the nesting property of bounds discussed in Section 2.3.2; more results supporting our comments made in Section 2.3.2 in regards to the usefulness of multiple instruments; the full set of results illustrating the effect of pleiotropy on the coverage of nonparametric bounds from two-sample MR studies; full details on the effect of using a dichotomized exposure variable; simulations showing the finite-sample behavior of the nonparametric bounds from two-sample MR data; details on extending the reconstruction of one-sample distributions introduced in Section 2.4 to intersection bounds; summary statistics, and complete results for the two example analyses presented in Section 2.5. For convenience, we include a table of contents below.

2.A Relating ST to the Concentration Parameter

Stock et al. (Stock, Wright, et al., 2002) present a different measure of the strength of an instrument called the concentration parameter. We will here see how our measure of strength $\text{ST} = \max_{z \neq z'} |P(X = 1|Z = z) - P(X = 1|Z = z')|$ can be related to the concentration parameter.

The concentration parameter is defined based on a linear model of the relationship between the exposure and the instrument. Here, we will consider the model with an intercept (Stock and Yogo, 2005)

$$E[X \mid Z] = \gamma_0 + \gamma_1 Z + \epsilon,$$

where γ_0, γ_1 are real-valued parameters and Z is the vector of values of the instrument, i.e. $z_i \in \{0, 1, 2\}$. Note that we treat the above model as a working model and not a true model in roder to define the concentration parameter under a binary exposure. The concentration parameter is then defined as

$$\mu^2 = \gamma_1 Z' Z \gamma_1 / \sigma^2 = \gamma_1^2 \sum_{i=1}^{n} z_i^2 / \sigma^2$$

Notice that in this model, $P(X = 1|Z = z) = \gamma_0 + \gamma_1 z$, meaning that $\text{ST} = 2|\gamma_1|$. Hence, larger values of ST coincides with larger values of μ^2.

2.B Bounds on Average Treatment Effect

We briefly review the method presented by (Ramsahai, 2012) to bound the average treatment effect using two-sample summary data. Let $\vec{\tau}^* = \Big(P(Y = 1|X = 0, U), P(Y = 1|X = 1, U), P(X = 1|Z = 0, U), ..., P(X = 1|Z = k-1, U)\Big) \in [0, 1]^{2+k}$ and $\vec{v}^* = \Big(P(Y = 0|Z = 0, U), ..., P(Y = 1|Z = k-1, U), P(X = 0|Z = 0, U), ..., P(X = 1|Z = k-1, U), \alpha^*\Big)$ where

$$\alpha^* = P(Y = 1|X = 1, U) - P(Y = 1|X = 0, U).$$

Since $U \perp Z$, $E_U[P(X = x|Z = z, U)] = P(X = x|Z = z)$ and $E_U[P(Y = y|Z = z, U)] = P(Y = y|Z = z)$. Let $\vec{v} = E_U[\vec{v}^*] = \Big(P(Y = 0|Z = 0), ..., P(Y = 1|Z = k-1), P(X = 0|Z = 0), ..., P(X = 1|Z = k-1), \alpha\Big)$, where

$$\begin{aligned} \alpha &= E_U[P(Y = 1|X = 1, U) - P(Y = 1|X = 0, U)] \\ &= E[Y^1] - E[Y^0] = \text{ATE}. \end{aligned}$$

Note that while $\vec{\tau}^*$ and $\vec{v}^*$ are both entirely unobervable, $\vec{v}$ consists of k observable values, and one unobservable value, the ATE.

By the exclusion restriction, we have

$$P(X = x, Y = y|Z = z, U) = P(Y = 1|X = x, U)P(X = x|Z = z, U),$$

which means we can define a mapping $f : [0, 1]^{2+k} \mapsto \mathcal{V}$ such that $f(\vec{\tau}^*) = \vec{v}^*$ as

$$f(y_0, y_1, x_0, x_1, ..., x_{k-1}) = \begin{pmatrix} (1 - y_0) \cdot (1 - x_0) + (1 - y_1) \cdot x_0 \\ y_0 \cdot (1 - x_0) + y_1 \cdot x_0 \\ \vdots \\ (1 - y_0) \cdot (1 - x_{k-1}) + (1 - y_1) \cdot x_{k-1} \\ y_0 \cdot (1 - x_{k-1}) + y_1 \cdot x_{k-1} \end{pmatrix}$$

We define $\mathcal{V} = f([0, 1]^{2+k})$.

Since $\vec{v} = E_U[\vec{v}^*]$, $\vec{v}$ must be a convex combination of $\vec{v}^*$. Let $\mathcal{H}$ be the convex hull of $\mathcal{V}$. Then $\vec{v}$ will be in $\mathcal{H}$.

Now, let $\hat{\mathcal{T}}$ be the set of extreme vertices of $[0, 1]^{2+k}$, $\hat{\mathcal{V}} = f(\hat{\mathcal{T}})$, and $\hat{\mathcal{H}}$ be the convex hull of $\hat{\mathcal{V}}$. By Theorem 1 in Appendix B of (Ramsahai, 2012), $\mathcal{H} = \hat{\mathcal{H}}$. This means that $\vec{v} \in \hat{\mathcal{H}}$. Utilizing a program such as Polymake, we can describe $\mathcal{H}$ with a set of inequalities, which give us constraints that $\vec{v}$ must satisfy.

This means that we can obtain inequalities that the components of $\vec{v}$ must satisfy by describing the extreme vertices of $[0, 1]^{2+k}$, map them to $\mathcal{V}$ using the relatively simple function f, and then use polymake to find inequalities that characterize the convex hull of $f([0, 1]^{2+k})$. This gives us a set of inequalities involving the components of $\vec{v}$. Some of these will be verifiable, as they will not include the only unobservable quantity α. Others will not be verifiable, but will allow us to obtain bounds on the unobservable quantity α using the observable entries of $\vec{v}$.

Following the approach from Ramsahai (2012) as outlined above, we obtain bounds on the average treatment effect from the quantities $P(X = 1|Z = z)$ and $P(Y = 1|Z = z)$, $z = 0, 1, 2$. To do so, we first write down the most extreme values of each of $P(Y = 1|X = x, U)$ and $P(X = x|Z = z, U)$ for all $x = 0, 1$, $z = 0, 1, 2$. Since these are probabilities, the extreme values are 0 and 1.

Table 2.2: Most extreme values of $P(Y = 1|X = x, U)$ and $P(X = 1|Z = z, U)$. Here, PY1XxU = $P(Y = 1|X = x, U)$ and PX1ZzU = $P(X = 1|Z = z, U)$.

PY1X0U	PY1X1U	PY1Z0U	PX1Z1U	PX1Z2U
0	0	0	0	0
0	0	0	0	1
0	0	0	1	0
0	0	0	1	1
0	0	1	0	0
0	0	1	0	1
0	0	1	1	0
0	0	1	1	1
0	1	0	0	0

PY1X0U	PY1X1U	PY1Z0U	PX1Z1U	PX1Z2U
0	1	0	0	1
0	1	0	1	0
0	1	0	1	1
0	1	1	0	0
0	1	1	0	1
0	1	1	1	0
0	1	1	1	1
1	0	0	0	0
1	0	0	0	1
1	0	0	1	0
1	0	0	1	1
1	0	1	0	0
1	0	1	0	1
1	0	1	1	0
1	0	1	1	1
1	1	0	0	0
1	1	0	0	1
1	1	0	1	0
1	1	0	1	1
1	1	1	0	0
1	1	1	0	1
1	1	1	1	0
1	1	1	1	1

By applying the function f to each row, we get the most extreme vertices of $P(X = x|Z = z, U)$, $P(Y = y|Z = z, U)$, and α for all $x = 0, 1,\ y = 0, 1$ and $z = 0, 1, 2$.

Table 2.3: Most extreme values of $P(Y = y|Z = z)$ and $P(X = x|Z = z)$. Here, PYyZz = $P(Y = y|Z = z)$, PXxZz = $P(X = x|Z = z)$, and
$\alpha = P(Y = 1|X = 1, U) - P(Y = 1|X = 0, U)$.

PY0Z0	PY0Z1	PY0Z2	PY1Z0	PY1Z1	PY1Z2	PX0Z0	PX0Z1	PX0Z2	PX1Z0	PX1Z1	PX1Z2	α
1	1	1	0	0	0	1	1	1	0	0	0	0
0	0	0	1	1	1	1	1	1	0	0	0	-1
1	1	1	0	0	0	1	1	1	0	0	0	1
0	0	0	1	1	1	1	1	1	0	0	0	0
1	1	1	0	0	0	0	1	1	1	0	0	0
1	0	0	0	1	1	0	1	1	1	0	0	-1
0	1	1	1	0	0	0	1	1	1	0	0	1
0	0	0	1	1	1	0	1	1	1	0	0	0
1	1	1	0	0	0	1	0	1	0	1	0	0
0	1	0	1	0	1	1	0	1	0	1	0	-1
1	0	1	0	1	0	1	0	1	0	1	0	1
0	0	0	1	1	1	1	0	1	0	1	0	0
1	1	1	0	0	0	0	0	1	1	1	0	0

PY0Z0	PY0Z1	PY0Z2	PY1Z0	PY1Z1	PY1Z2	PX0Z0	PX0Z1	PX0Z2	PX1Z0	PX1Z1	PX1Z2	α
1	1	0	0	0	1	0	0	1	1	1	0	-1
0	0	1	1	1	0	0	0	1	1	1	0	1
0	0	0	1	1	1	0	0	1	1	1	0	0
1	1	1	0	0	0	1	1	0	0	0	1	0
0	0	1	1	1	0	1	1	0	0	0	1	-1
1	1	0	0	0	1	1	1	0	0	0	1	1
0	0	0	1	1	1	1	1	0	0	0	1	0
1	1	1	0	0	0	0	1	0	1	0	1	0
1	0	1	0	1	0	0	1	0	1	0	1	-1
0	1	0	1	0	1	0	1	0	1	0	1	1
0	0	0	1	1	1	0	1	0	1	0	1	0
1	1	1	0	0	0	1	0	0	0	1	1	0
0	1	1	1	0	0	1	0	0	0	1	1	-1
1	0	0	0	1	1	1	0	0	0	1	1	1
0	0	0	1	1	1	1	0	0	0	1	1	0
1	1	1	0	0	0	0	0	0	1	1	1	0
1	1	1	0	0	0	0	0	0	1	1	1	-1
0	0	0	1	1	1	0	0	0	1	1	1	1
0	0	0	1	1	1	0	0	0	1	1	1	0

Theorem 1 of Ramsahai (2012) tells us that the values of $P(X = 1|Z = z), P(Y = 1|Z = z),\ z = 0, 1, 2$ must lie in the convex hull of the vertices given by the rows in Table 2.3. This means that the vector of these values must be a convex combination of the rows in said table. Using this with the fact that they must sum to 1 is what enables us to use polymake to find inequalities that the values of $P(X = 1|Z = z)$, $P(Y = 1|Z = z)$, and α must satisfy. In this particular case, these are as presented below. This table should be read as rows of coefficients for which it holds that $\sum_{z=0}^{2} c_{X1Zz} \cdot P(X = 1|Z = z) + \sum_{z=0}^{2} c_{Y0Zz} \cdot P(Y = 0|Z = z) + c_{Y1Z0} \cdot P(Y = 1|Z = 0) + c_{\alpha}\alpha \geq 0$.

Table 2.4: Results from polymake. Columns with all zeroes have been removed.

c_{Y0Z0}	c_{Y0Z1}	c_{Y0Z2}	c_{Y1Z0}	c_{X1Z0}	c_{X1Z1}	c_{X1Z2}	c_{α}
2	0	-1	0	2	0	0	-1
1	0	-1	1	0	0	0	0
1	-1	0	1	0	0	0	0
1	-1	0	0	1	1	0	0
1	0	-1	0	1	0	1	0
2	0	-1	1	1	0	-1	-1
2	-1	0	1	1	-1	0	-1
2	0	-2	1	0	0	2	1
2	-1	0	1	-1	1	0	1
4	0	-2	3	0	0	-2	-1
2	-2	0	1	0	2	0	1
4	-1	0	2	-2	0	0	1
4	0	-1	2	-2	0	0	1

c_{Y0Z0}	c_{Y0Z1}	c_{Y0Z2}	c_{Y1Z0}	c_{X1Z0}	c_{X1Z1}	c_{X1Z2}	c_{α}
2	0	-1	1	-1	0	1	1
1	0	-1	1	0	0	1	1
3	-1	0	2	-1	-1	0	0
2	-1	0	0	2	0	0	-1
4	-2	0	3	0	-2	0	-1
3	0	-1	2	-1	0	-1	0
1	-1	0	1	0	1	0	1
1	-1	1	1	0	1	-1	1
1	0	0	1	0	-1	0	0
1	0	0	1	0	0	-1	0
1	0	1	1	0	0	-1	1
2	-1	2	2	0	0	-2	1
1	1	0	1	0	-1	0	1
0	1	0	1	1	-1	0	1
0	0	1	1	1	0	-1	1
2	2	-1	2	0	-2	0	1
2	1	-1	2	0	-1	-1	0
2	-1	1	2	0	-1	-1	0
0	0	0	1	1	0	0	1
1	1	-1	1	0	-1	1	1
0	0	0	0	1	0	0	0
2	0	0	1	-1	0	0	1
0	0	1	1	-1	0	1	-1
0	0	0	0	0	1	0	0
1	-1	1	1	0	-1	1	-1
-1	2	0	0	0	2	0	-1
2	0	-1	2	0	0	-1	-1
1	0	1	3	-2	0	0	-1
1	1	0	2	-1	-1	0	0
0	1	-1	0	0	1	1	0
0	1	0	1	-1	1	0	-1
0	0	1	0	0	0	0	0
-1	0	1	1	2	0	0	1
3	-2	1	3	0	-2	0	-1
0	0	0	0	0	0	1	0
0	-1	1	0	0	1	1	0
0	1	0	0	0	0	0	0
1	1	0	3	-2	0	0	-1
1	0	0	1	-1	0	0	0
0	2	-1	0	0	2	0	-1
1	0	2	2	0	0	-2	1
0	0	0	1	0	0	0	0
1	-2	1	1	0	2	0	1
2	-1	0	2	0	-1	0	-1
1	1	-1	1	0	1	-1	-1
-1	0	1	0	1	0	1	0

c_{Y0Z0}	c_{Y0Z1}	c_{Y0Z2}	c_{Y1Z0}	c_{X1Z0}	c_{X1Z1}	c_{X1Z2}	c_α
1	0	0	0	1	0	0	-1
-1	0	2	0	0	0	2	-1
1	2	0	2	0	-2	0	1
1	1	-2	1	0	0	2	1
-1	1	0	0	1	1	0	0
0	1	0	0	0	1	0	-1
0	0	1	0	0	0	1	-1
1	0	0	2	-1	0	0	-1
-1	1	0	1	2	0	0	1
3	1	-2	3	0	0	-2	-1
0	-1	2	0	0	0	2	-1
1	0	1	2	-1	0	-1	0
1	0	0	0	0	0	0	0

The matrix presented in the table above simplifies to the following set of bounds on the average treatment effect. These are obtained by considering the rows above where $c_\alpha \neq 0$.

$$\max \left\{ \begin{array}{ll} \max_{i \neq j} & P(Y=1|Z=i) - 2 \cdot P(Y=1|Z=j) - 2 \cdot P(X=1|Z=j) \\ \max_{i \neq j} & P(Y=1|Z=i) + P(X=1|Z=i) - P(Y=1|Z=j) - P(X=1|Z=j) - 1 \\ \max_{i \neq j} & 2 \cdot P(Y=1|Z=i) + 2 \cdot P(X=1|Z=i) - P(Y=1|Z=j) - 3 \\ \max_{i} & -P(Y=1|Z=i) - P(X=1|Z=i) \\ \max_{i} & P(Y=1|Z=i) + P(X=1|Z=i) - 2 \end{array} \right\}$$

$$\leq \alpha \leq \tag{2.8}$$

$$\min \left\{ \begin{array}{ll} \min_{i \neq j} & P(Y=1|Z=i) - 2 \cdot P(Y=1|Z=j) + 2 \cdot P(X=1|Z=j) + 1 \\ \min_{i \neq j} & P(Y=1|Z=i) + 2 \cdot P(Y=1|Z=j) - 2 \cdot P(X=1|Z=j) + 1 \\ \min_{i \neq j} & P(Y=1|Z=i) - P(X=1|Z=i) + P(X=1|Z=j) - P(Y=1|Z=j) + 1 \\ \min_{i} & P(X=1|Z=i) - P(Y=1|Z=i) + 1 \\ \min_{i} & P(Y=1|Z=i) - P(X=1|Z=i) + 1 \end{array} \right\} \tag{2.9}$$

Furthermore, we obtain the following checkable constraints from the rows where $\alpha = 0$:

$$\min \left\{ \begin{array}{ll} \min_{i \neq j} & P(Y=1|Z=i) - P(X=1|Z=i) - P(Y=1|Z=j) - P(X=1|Z=j) + 2 \\ \min_{i \neq j} & P(Y=1|Z=i) + P(X=1|Z=i) - P(Y=1|Z=j) + P(X=1|Z=j) \\ \min_{i} & P(X=1|Z=i) \\ \min_{i} & P(Y=1|Z=i) \\ \min_{i} & 1 - P(X=1|Z=i) \\ \min_{i} & 1 - P(Y=1|Z=i) \end{array} \right\} \geq 0 \tag{2.10}$$

We notice that the constraints from the law of probability are recovered (the last four expressions above) along with 12 non-trivial constraints.

These bounds involve 24 different expressions on both the lower and upper end, making an algebraic exploration of the bounds very challenging. However, by imposing two additional monotonicity assumptions, the bounds reduce to just three expressions on the lower end and three expressions on the upper end. This is done by removing rows in the matrix of extreme vertices where the monotonicity assumptions are violated before using Polymake to get the inequalities. The additinal assumptions are

(A5) *(Monotonicity between Z and X)* $P(X = 1|Z = z, U) \leq P(X = 1|Z = z + 1, U)$ for $z = 0, 1, 2$

(A6) *(Monotonicity between Z and Y)* $P(Y = 1|Z = z, U) \leq P(Y = 1|Z = z + 1, U)$ for $z = 0, 1, 2$

and the resulting bounds are

$$\max\left\{\begin{array}{c} P(Y=1|Z=k) - 2\cdot P(Y=1|Z=0) - P(X=1|Z=0) \\ P(Y=1|Z=k) - P(Y=1|Z=0) + P(X=1|Z=2) - P(X=1|Z=0) - 1 \\ 2\cdot P(Y=1|Z=k) - P(Y=1|Z=0) + P(X=1|Z=k) - 2 \end{array}\right\}$$
$$\leq ATE \leq$$
$$\min\left\{\begin{array}{c} P(X=1|Z=0) - P(Y=1|Z=0) + 1 \\ P(Y=1|Z=k) - P(Y=1|Z=0) + P(X=1|Z=0) - P(X=1|Z=k) + 1 \\ P(Y=1|Z=k) - P(X=1|Z=k) + 1 \end{array}\right\}$$

It should be noted that imposing these additional assumptions do not tighten the bounds, it only simplifies the expressions.

We encountered one surprise when studying the behavior of the bounds in (2.8). Of 10,123 randomly generated sets of values for $P(X = 1|Z = z), P(Y = 1|Z = z),\ z = 0, 1, 2,\ 123$ resulted in bounds where the upper limit is smaller than the lower limit without violating any of the verifiable constraints presented in (2.10). Table 2.5 gives the values of the marginal conditional distributions with the strength of the IV, the corresponding bounds, and the length. It is notable that the IVs are rather strong in all cases where we see the bounds flip, but the bounds themselves and the length vary quite a bit.

We first attributed this to the transition from one-sample to two-sample bounds, but later realized similar scenarios arise when dealing with one-sample bounds from four category IVs. We present some of these in Table 2.6, where we include the one-sample distributions with the strengths of the IVs, and the length of the resulting bounds. Again, it is interesting to see the large span of lengths and strengths present.

To further explore this, we randomly generated $10,000,000$ distributions of both $P(X, Y|Z)$ and $P(X|Z), P(Y|Z)$ for an instrument Z with $k = 2, 3, 4$ categories. In Table 2.7, we tabulate the one-sample distributions based on whether or not they violate the verifiable constraints, and whether the bounds are flipped. Table 2.8 is the equivalent table for two-sample distributions. We note that we do not see any flipped bounds for one-sample bounds

when the instrument is binary or with three categories that meet the verifiable constraints, but we do see this when the instrument has four categories, albeit very rarely. For two-sample bounds, we see flipped bounds that are not caught by the verifiable constraints even when the instrument has three categories, and in more than one percent of the distributions when the instrument has four categories.

We have been unable to unearth a reason for why we see this phenomenon. One possible explanation is that the distributions that result in flipped bounds violate some uncheckable assumption.

Table 2.5: Marginal conditional probabilities resulting in bounds where the upper bound is smaller than the lower bound.

P(X=1|Z=0)	P(X=1|Z=1)	P(X=1|Z=2)	P(Y=1|Z=0)	P(Y=1|Z=1)	P(Y=1|Z=2)	Strength	Lower Bound	Upper Bound	Length
0.2309955	0.3669268	0.9387298	0.8850137	0.3013143	0.9801302	0.7077343	0.5364056	-0.0067221	-0.5431277
0.9404491	0.4742722	0.1448868	0.0262469	0.5741507	0.1155472	0.7955623	0.0532826	-0.4025552	-0.4558377
0.8243777	0.0826950	0.6396267	0.0984834	0.0536095	0.6267494	0.7416826	0.3541403	-0.0785379	-0.4326782
0.6253430	0.7940521	0.0769966	0.7125237	0.1332569	0.0937761	0.7170556	0.3709784	-0.0341142	-0.4050925
0.4687418	0.9885571	0.0147455	0.4269904	0.0952051	0.1145516	0.9738116	0.1683963	-0.2136943	-0.3820906
0.2384690	0.9589127	0.4551064	0.9411639	0.8220534	0.2995920	0.7204437	0.2623402	-0.1057977	-0.3681380
0.1201855	0.5087544	0.6903413	0.1553146	0.7813318	0.0153936	0.5701558	0.2303316	-0.1312272	-0.3615588
0.0558596	0.8249922	0.5150187	0.1693588	0.0317164	0.6019942	0.7691326	0.1515574	-0.1885458	-0.3401031
0.0601930	0.7105220	0.7764157	0.0349669	0.6138605	0.1288649	0.7162227	0.4235408	0.0910378	-0.3325030
0.9689451	0.3369273	0.0921191	0.9728974	0.3379845	0.6435396	0.8768260	0.5457005	0.2351435	-0.3105570
0.0272617	0.9602504	0.7090107	0.9941238	0.7603751	0.5393045	0.9329888	-0.0980534	-0.3944198	-0.2963664
0.8593575	0.5455747	0.0954651	0.7493743	0.2343858	0.8692962	0.7638924	-0.0169223	-0.3132765	-0.2963542
0.0051370	0.7930864	0.6854693	0.0171757	0.5039197	0.0258429	0.7879494	0.4592943	0.1768274	-0.2824669
0.8095621	0.0899196	0.7315497	0.1398438	0.0112235	0.5721541	0.7196425	0.3698677	0.0884094	-0.2814583
0.0312864	0.5136612	0.7187288	0.1782691	0.7144743	0.0839332	0.6874423	0.2953632	0.0159345	-0.2794287
0.2841081	0.4642261	0.9303618	0.9272837	0.3015191	0.8563395	0.6462537	0.2718836	0.0151680	-0.2567156
0.7020589	0.0426525	0.7537495	0.8146495	0.9551254	0.3030152	0.7110970	-0.2695984	-0.5219304	-0.2523321
0.7299439	0.7079992	0.0126445	0.4179246	0.9411138	0.9059591	0.7172993	-0.1196986	-0.3687044	-0.2490059
0.8553215	0.1611814	0.3987327	0.0868026	0.0650961	0.5766878	0.6941401	0.1241329	-0.1137256	-0.2378585
0.7503627	0.8262444	0.0255938	0.9023691	0.4826617	0.9697816	0.8006505	-0.1771982	-0.4057139	-0.2285157
0.7516532	0.1293625	0.6636683	0.2319998	0.0773707	0.8011377	0.6222907	0.3876713	0.1595554	-0.2281159
0.1892072	0.6542341	0.6029697	0.9717090	0.8941221	0.2186525	0.4650268	-0.1219402	-0.3463509	-0.2244107
0.9351863	0.1648035	0.3655840	0.1803887	0.1576169	0.6793117	0.7703828	0.0344709	-0.1889068	-0.2233777
0.8913881	0.2924893	0.1391987	0.0678851	0.5562612	0.1311623	0.7521894	0.0155394	-0.2032671	-0.2188065
0.2004629	0.8817321	0.4467427	0.2410824	0.0446975	0.7057212	0.6812692	-0.1773694	-0.3797903	-0.2024209
0.2713706	0.9177118	0.2155938	0.0584116	0.0235335	0.5341155	0.7021180	-0.1254488	-0.3224721	-0.1970232
0.1716186	0.9793879	0.4387238	0.0758875	0.0913810	0.4572813	0.8077692	-0.0377310	-0.2332949	-0.1955639
0.0346134	0.8601421	0.5243412	0.7170224	0.9940138	0.4402146	0.8255286	0.2680971	0.0753966	-0.1927005
0.0517557	0.9490455	0.4763609	0.2257054	0.0428283	0.4666474	0.8972898	-0.0882749	-0.2790819	-0.1908070
0.2097271	0.7849572	0.5591844	0.9851851	0.7694310	0.2353843	0.5752301	-0.1266079	-0.3155315	-0.1889237
0.8533233	0.5437889	0.3202183	0.0278734	0.0138157	0.8263378	0.5331050	-0.2888714	-0.4772378	-0.1883664
0.0781475	0.4316186	0.9562902	0.6056942	0.2534086	0.8616394	0.8781427	0.3824505	0.1983152	-0.1841354
0.7343532	0.7111032	0.0863323	0.4004145	0.9342732	0.9323079	0.6480209	-0.1096618	-0.2915366	-0.1818748
0.4855778	0.2600183	0.9736867	0.3390356	0.9283873	0.7874292	0.7136685	0.1831962	0.0022975	-0.1808987

(continued)

P(X=1|Z=0)	P(X=1|Z=1)	P(X=1|Z=2)	P(Y=1|Z=0)	P(Y=1|Z=1)	P(Y=1|Z=2)	Strength	Lower Bound	Upper Bound	Length
0.6368154	0.0572293	0.8159708	0.5109590	0.0158577	0.1663634	0.7587416	0.3647850	0.1898262	-0.1749588
0.8824330	0.1367268	0.3081087	0.0653359	0.1951474	0.6000460	0.7457061	-0.0637026	-0.2342401	-0.1705375
0.8090247	0.3226145	0.5675011	0.9402684	0.9741885	0.3180210	0.4864103	0.1805653	0.0148730	-0.1656923
0.4510693	0.0872080	0.9033969	0.5323388	0.1710303	0.0969452	0.8161888	0.0158620	-0.1452420	-0.1611040
0.1518352	0.6975145	0.6509167	0.0629987	0.8097783	0.1657477	0.5456793	0.3801104	0.2198838	-0.1602266
0.0653620	0.3813488	0.9612892	0.9275631	0.4953530	0.7515764	0.8959272	-0.0696219	-0.2290492	-0.1594273
0.2032074	0.7755576	0.4991361	0.7865987	0.9554554	0.2348516	0.5723502	0.2271745	0.0680689	-0.1591056
0.0233274	0.6660489	0.8176706	0.8429973	0.2798561	0.7213751	0.7943432	-0.2017648	-0.3594838	-0.1577189
0.9294752	0.2110150	0.4387583	0.1560685	0.0882931	0.6040925	0.7184602	0.0054762	-0.1509059	-0.1563822
0.1670113	0.6894123	0.4795673	0.0041910	0.8002859	0.0345400	0.5224010	0.4578813	0.3096595	-0.1482218
0.3785346	0.9143229	0.1322393	0.3764540	0.9927913	0.6755701	0.7820836	0.4377743	0.2897923	-0.1479819
0.1776605	0.3763786	0.8762187	0.2525663	0.7852824	0.1601145	0.6985582	-0.0751713	-0.2174909	-0.1423196
0.7676593	0.0086728	0.5238627	0.3109642	0.8841540	0.9821670	0.7589865	-0.2989048	-0.4399984	-0.1410937
0.8834087	0.2154675	0.5237259	0.9402145	0.9094435	0.4479360	0.6679412	0.1993104	0.0599839	-0.1393265
0.2128945	0.6634662	0.7020688	0.9859116	0.2297734	0.8227277	0.4891743	-0.1801804	-0.3162608	-0.1360804
0.8197957	0.4539939	0.2933378	0.1292782	0.6944266	0.0241216	0.5264579	0.0595077	-0.0754615	-0.1349692
0.8932091	0.2573860	0.3789772	0.8683447	0.8850420	0.3218777	0.6358231	0.2012298	0.0665657	-0.1346641
0.3852521	0.7681010	0.1679198	0.6200211	0.0286245	0.1269667	0.6001813	0.0302481	-0.0989742	-0.1292223
0.4450183	0.3448027	0.9580487	0.0334938	0.6223715	0.0373602	0.6132460	-0.3346527	-0.4637484	-0.1290957
0.9626206	0.3323393	0.3615993	0.8971357	0.8947940	0.3577061	0.6302814	0.3618066	0.2327966	-0.1290100
0.9579589	0.2856719	0.2557011	0.0294142	0.0312341	0.4495460	0.7022578	-0.1842660	-0.3066353	-0.1223693
0.2722892	0.1030317	0.9532750	0.3335194	0.0179986	0.1046059	0.8502432	0.0914587	-0.0308574	-0.1223161
0.2075435	0.6267518	0.9907035	0.0610969	0.8711902	0.5325762	0.7831600	0.3339092	0.2125552	-0.1213540
0.1309917	0.9511009	0.6110001	0.0092469	0.1382892	0.3862037	0.8201092	0.1057264	-0.0118269	-0.1175533
0.9469203	0.4771290	0.2975224	0.8483259	0.2756656	0.8366797	0.6493979	0.3148269	0.1973510	-0.1174758
0.9141838	0.3947449	0.2582693	0.1776121	0.6284717	0.0485084	0.6559145	0.0149163	-0.1016151	-0.1165314
0.2539480	0.3283935	0.9257231	0.5855638	0.1211694	0.0074839	0.6717752	-0.3135619	-0.4220422	-0.1084803
0.7554315	0.0394385	0.8166883	0.9193390	0.1504442	0.4920783	0.7772497	0.5395735	0.4314412	-0.1081323
0.5322302	0.8442719	0.1311744	0.7227207	0.1174348	0.2652317	0.7130975	-0.0700917	-0.1763950	-0.1063033
0.1022484	0.7850567	0.3114329	0.9983873	0.9750404	0.6040354	0.6828082	-0.0838413	-0.1882423	-0.1044009
0.8859779	0.1854690	0.2675919	0.9352886	0.8113619	0.3954484	0.7005089	0.2470847	0.1436625	-0.1034222
0.8858413	0.0577413	0.7457014	0.9231434	0.9814877	0.6837953	0.8281000	-0.0658260	-0.1636975	-0.0978715
0.5688937	0.0533840	0.9092544	0.4161218	0.0847550	0.1385937	0.8558704	0.1398438	0.0425567	-0.0972870
0.0111502	0.5785773	0.7360408	0.9491940	0.9715842	0.4417906	0.7248905	-0.3414676	-0.4342969	-0.0928294
0.8016434	0.0919814	0.6269118	0.0598012	0.0080604	0.4024806	0.7096620	0.2023970	0.1138349	-0.0885621

(continued)

P(X=1|Z=0)	P(X=1|Z=1)	P(X=1|Z=2)	P(Y=1|Z=0)	P(Y=1|Z=1)	P(Y=1|Z=2)	Strength	Lower Bound	Upper Bound	Length
0.5613155	0.3343263	0.9641096	0.1739435	0.9413168	0.6466249	0.6297833	0.0475254	-0.0400375	-0.0875629
0.9421035	0.7800406	0.0170238	0.6536674	0.8584000	0.0860958	0.9250797	0.6521608	0.5647278	-0.0874330
0.4856718	0.1412137	0.8327200	0.2353279	0.7698770	0.8171080	0.6915064	0.0643282	-0.0219988	-0.0863269
0.7587967	0.2217142	0.4642144	0.1261614	0.0095185	0.6397095	0.5370825	0.1772441	0.0950201	-0.0822241
0.8476325	0.0321449	0.5761561	0.7137147	0.9222930	0.4156565	0.8154876	-0.2929622	-0.3646398	-0.0716776
0.8443266	0.0231323	0.6135112	0.5114541	0.9662261	0.9901356	0.8211943	-0.3041605	-0.3747334	-0.0705729
0.7090756	0.0306938	0.8591612	0.8275547	0.1987801	0.4221209	0.8284674	0.3686070	0.2983647	-0.0702424
0.5210445	0.6877412	0.1936365	0.2077578	0.8583608	0.8895555	0.4941047	-0.1155538	-0.1840802	-0.0685264
0.7325333	0.0360979	0.7452189	0.9243027	0.1841382	0.4150783	0.7091209	0.4838304	0.4154162	-0.0684143
0.3112649	0.5408216	0.7700621	0.0719339	0.8911155	0.9844600	0.4587973	0.4371103	0.3713461	-0.0657642
0.6839198	0.0601158	0.7429099	0.3546209	0.0832522	0.8458772	0.6827941	0.5591411	0.4955250	-0.0636161
0.4925476	0.1475428	0.6432137	0.1357593	0.7295215	0.9418075	0.4956709	0.0342830	-0.0281982	-0.0624812
0.0567614	0.4716677	0.8412115	0.9781020	0.6182925	0.8866750	0.7844501	-0.1625195	-0.2243887	-0.0618691
0.1902110	0.3836209	0.9071890	0.8456573	0.3088491	0.0296753	0.7169780	-0.5392827	-0.6006846	-0.0614020
0.3772296	0.8822068	0.2883994	0.2173902	0.9350335	0.7191264	0.5938073	0.4170904	0.3559363	-0.0611541
0.5973862	0.8450983	0.2624347	0.1392309	0.6156584	0.9712264	0.5826636	-0.2177176	-0.2783525	-0.0606348
0.6339672	0.0297922	0.8123455	0.7376053	0.9506195	0.2630108	0.7825533	-0.5198657	-0.5786439	-0.0587783
0.0823461	0.5840173	0.6679903	0.9677474	0.8284869	0.2712011	0.5856442	-0.4461926	-0.4996015	-0.0534089
0.6535119	0.8883952	0.1073055	0.2820041	0.7154519	0.8117950	0.7810897	-0.0743099	-0.1269749	-0.0526651
0.7404535	0.1312750	0.4474163	0.1314948	0.9068344	0.9347602	0.6091785	-0.3671417	-0.4196239	-0.0524822
0.0820021	0.8994346	0.3178099	0.4734612	0.1446546	0.8253918	0.8174325	-0.2855348	-0.3349518	-0.0494170
0.0143154	0.1408971	0.9883829	0.5259441	0.4011591	0.9257180	0.9740675	0.4270428	0.3779018	-0.0491410
0.5142074	0.8446779	0.0753746	0.5067568	0.0715657	0.1808748	0.7693032	-0.0057421	-0.0529810	-0.0472389
0.1391137	0.4452852	0.7319911	0.0201224	0.4730480	0.0227584	0.5928773	0.1545757	0.1084867	-0.0460890
0.7671998	0.0911903	0.9424491	0.7190755	0.0257481	0.5228183	0.8512587	0.4851985	0.4416630	-0.0435356
0.2249334	0.9771968	0.6502243	0.9434316	0.7995282	0.4743734	0.7522634	0.0790767	0.0373769	-0.0416998
0.9124694	0.5503730	0.0400667	0.7951134	0.6099932	0.9632078	0.8724027	-0.1948275	-0.2362891	-0.0414616
0.1645046	0.8060324	0.5635964	0.9246119	0.7605022	0.3061245	0.6415279	-0.1730552	-0.2140902	-0.0410350
0.7079565	0.5723802	0.2806847	0.8839699	0.2430289	0.9515723	0.4272719	-0.0591760	-0.0987463	-0.0395703
0.2097282	0.9124687	0.2747676	0.2570863	0.1285457	0.7024909	0.7027405	-0.2311382	-0.2703369	-0.0391987
0.9736240	0.0208031	0.3737885	0.9045140	0.4334044	0.2716260	0.9528209	0.4846500	0.4464234	-0.0382266
0.1845828	0.1851770	0.8937890	0.8433725	0.4857333	0.9516657	0.7092062	0.2051761	0.1681541	-0.0370221
0.1904095	0.9898458	0.0778574	0.3241436	0.0396418	0.5826816	0.9119883	-0.4464247	-0.4830894	-0.0366648
0.3058563	0.8758829	0.3221585	0.8338573	0.0715108	0.2981029	0.5700266	-0.4066656	-0.4426015	-0.0359359
0.5517228	0.8850872	0.1379439	0.7797196	0.3208303	0.1888349	0.7471432	0.1261619	0.0917667	-0.0343952

(continued)

P(X=1|Z=0)	P(X=1|Z=1)	P(X=1|Z=2)	P(Y=1|Z=0)	P(Y=1|Z=1)	P(Y=1|Z=2)	Strength	Lower Bound	Upper Bound	Length
0.0614376	0.2965834	0.9979328	0.0027831	0.1401460	0.0597136	0.9364952	0.0117046	-0.0165844	-0.0282890
0.8779495	0.4096741	0.2304406	0.7998226	0.4274697	0.9938156	0.6475089	-0.0719255	-0.0992804	-0.0273549
0.6979215	0.7737010	0.0234315	0.9852010	0.4651610	0.8182570	0.7502694	-0.0989160	-0.1244899	-0.0255739
0.6623782	0.7107869	0.1608789	0.9024376	0.2805005	0.8890312	0.5499081	-0.1508689	-0.1758042	-0.0249354
0.4107040	0.6300393	0.0755462	0.7135503	0.0247311	0.2318819	0.5544931	0.0986941	0.0758333	-0.0228608
0.2389620	0.9996788	0.3607017	0.1224239	0.2775328	0.6499732	0.7607167	-0.0727986	-0.0942652	-0.0214665
0.2466505	0.3150522	0.9973913	0.7941729	0.4943148	0.9589104	0.7507408	0.4182885	0.3992699	-0.0190186
0.1047963	0.5872602	0.6265764	0.1702907	0.0689137	0.7661262	0.5217801	0.2159521	0.1971807	-0.0187714
0.6454304	0.5477765	0.0021959	0.8270074	0.1628806	0.2007895	0.6432345	0.4210367	0.4032008	-0.0178359
0.0147348	0.9403617	0.7719393	0.1339251	0.5201033	0.7372833	0.9256270	0.4399636	0.4221999	-0.0177637
0.6149141	0.1287129	0.8052456	0.3774013	0.9281094	0.7809966	0.6765327	-0.2049168	-0.2213916	-0.0164747
0.6318831	0.8417779	0.1046526	0.1803197	0.6822984	0.0227946	0.7371254	0.4274041	0.4145748	-0.0128292
0.4658334	0.1177519	0.8202813	0.3008471	0.8740505	0.7295855	0.7025294	-0.2011135	-0.2117500	-0.0106365
0.4692894	0.9793264	0.2505315	0.6858286	0.3586177	0.0507586	0.7287948	0.0832484	0.0727541	-0.0104943
0.9053262	0.4920161	0.2908324	0.8237065	0.8801458	0.1128271	0.6144939	0.3452384	0.3365678	-0.0086706
0.8400507	0.6066834	0.0207922	0.8392446	0.3014262	0.1199182	0.8192585	0.5578239	0.5502410	-0.0075829
0.2986999	0.3574011	0.7508847	0.7003727	0.1246649	0.9739429	0.4521849	0.3249903	0.3213192	-0.0036711
0.0463115	0.4417234	0.7452841	0.1110238	0.4748895	0.0612693	0.6989726	0.1602189	0.1570808	-0.0031381
0.8543023	0.0104242	0.1896705	0.9925313	0.2311163	0.0674310	0.8438782	0.6262363	0.6260467	-0.0001896

Table 2.6: Lower and Upper limits of bounds where the upper limit is less than the lower limit for trivariate distributions with four category instruments.

Lower	Upper	Strength	Length
0.1796920	0.0395535	0.0853119	-0.1401385
-0.0038326	-0.1264492	0.1539099	-0.1226166
-0.0169573	-0.1304422	0.2235469	-0.1134849
-0.0620851	-0.1743916	0.0805434	-0.1123066
0.0996764	-0.0065497	0.2112420	-0.1062260
-0.0348047	-0.1393748	0.1884223	-0.1045701
-0.0097177	-0.1102060	0.0874967	-0.1004882
-0.0470850	-0.1435686	0.1458296	-0.0964835
-0.1052398	-0.1993785	0.2667633	-0.0941387
0.1097975	0.0268471	0.1774704	-0.0829504
0.1884781	0.1110487	0.3297432	-0.0774293
0.0174359	-0.0580424	0.2058740	-0.0754784
-0.0530855	-0.1187770	0.2521754	-0.0656915
0.0534080	-0.0107149	0.1509847	-0.0641230
-0.0660707	-0.1258819	0.2831483	-0.0598112
0.3495840	0.2945716	0.3633999	-0.0550124
0.1665198	0.1136389	0.2131245	-0.0528809
-0.0356540	-0.0879713	0.2476628	-0.0523173
0.1089847	0.0575836	0.1941017	-0.0514012
0.0086756	-0.0338341	0.2340061	-0.0425097
0.1335166	0.0930974	0.4555966	-0.0404192
0.1163970	0.0761754	0.1573917	-0.0402216
-0.1249197	-0.1611461	0.1712798	-0.0362264
-0.1252239	-0.1581375	0.1035529	-0.0329136
-0.2954311	-0.3273509	0.3077593	-0.0319199
0.0274287	-0.0007244	0.0813449	-0.0281530
-0.1317444	-0.1586467	0.3469784	-0.0269023
0.1050533	0.0818064	0.2388595	-0.0232469
-0.1980031	-0.2156885	0.2205149	-0.0176854
0.0408272	0.0265662	0.1314643	-0.0142609
0.1255375	0.1131666	0.0426523	-0.0123709
-0.1421790	-0.1523644	0.1409053	-0.0101854
-0.0997312	-0.1083943	0.3816466	-0.0086630
-0.0304169	-0.0353880	0.1323408	-0.0049711
0.0094786	0.0046709	0.2838685	-0.0048077
-0.0217285	-0.0245811	0.3531008	-0.0028526
-0.0563955	-0.0583218	0.4092683	-0.0019263

Table 2.7: For each category, $10,000,000$ distributions of $P(X, Y|Z)$ were randomly generated.

k	Verifiable Constraints Violated	upper < lower	n	Proportion
2	FALSE	FALSE	9611081	0.9611
	TRUE	FALSE	291802	0.0292
	TRUE	TRUE	97117	0.0097
3	FALSE	FALSE	8822427	0.8822
	TRUE	FALSE	682620	0.0683
	TRUE	TRUE	494953	0.0495
4	FALSE	FALSE	7802425	0.7802
	FALSE	TRUE	3797	0.0004
	TRUE	FALSE	1013534	0.1014
	TRUE	TRUE	1180244	0.1180

Table 2.8: For each category, $10,000,000$ distributions of $P(X|Z)$ and $P(Y|Z)$ were randomly generated.

k	Verifiable Constraints Violated	upper < lower	n	Proportion
2	FALSE	FALSE	8333580	0.8334
	TRUE	FALSE	925808	0.0926
	TRUE	TRUE	740612	0.0741
3	FALSE	FALSE	6000482	0.6000
	FALSE	TRUE	83565	0.0084
	TRUE	FALSE	1513834	0.1514
	TRUE	TRUE	2402119	0.2402
4	FALSE	FALSE	3951817	0.3952
	FALSE	TRUE	127177	0.0127
	TRUE	FALSE	1567644	0.1568
	TRUE	TRUE	4353362	0.4353

2.C Proof of Theorem

We present the proof of Theorem 2.1.

First of all, recall that the nonparametric bounds for instruments with $k = 2, 3, 4$ can be written as

$$\max \left\{ \begin{array}{ll} \max_{z \neq z'} & P(Y=1|Z=z) - 2 \cdot P(Y=1|Z=z') - 2 \cdot P(X=1|Z=z') \\ \max_{z \neq z'} & P(Y=1|Z=z) + P(X=1|Z=z) - P(Y=1|Z=z') - P(X=1|Z=z') - 1 \\ \max_{z \neq z'} & 2 \cdot P(Y=1|Z=z) + 2 \cdot P(X=1|Z=z) - P(Y=1|Z=z') - 3 \\ \max_{z} & -P(Y=1|Z=z) - P(X=1|Z=z) \\ \max_{z} & P(Y=1|Z=z) + P(X=1|Z=z) - 2 \end{array} \right\}$$

$$\leq ATE \leq \tag{2.11}$$

$$\min \left\{ \begin{array}{ll} \min_{z \neq z'} & P(Y=1|Z=z) - 2 \cdot P(Y=1|Z=z') + 2 \cdot P(X=1|Z=z') + 1 \\ \min_{z \neq z'} & P(Y=1|Z=z) + 2 \cdot P(Y=1|Z=z') - 2 \cdot P(X=1|Z=z') + 1 \\ \min_{z \neq z'} & P(Y=1|Z=z) - P(X=1|Z=z) + P(X=1|Z=z') - P(Y=1|Z=z') + 1 \\ \min_{z} & P(X=1|Z=z) - P(Y=1|Z=z) + 1 \\ \min_{z} & P(Y=1|Z=z) - P(X=1|Z=z) + 1 \end{array} \right\}$$

By definition, $\text{ST} = \max_{z,z'} |P(X=1|Z=z) - P(X=1|Z=z')|$. I.e. $\text{ST} = P(X = 1|Z = z_{max}) - P(X = 1|Z = z_{min})$ where $z_{max} = \arg\max_z P(X = 1|Z = z)$ and $z_{min} = \arg\min_z P(X = 1|Z = z)$.

Next, for the second expression in the lower bound,

$$\begin{aligned} &\max_{z \neq z'} P(Y=1|Z=z) - P(Y=1|Z=z') + P(X=1|Z=z) - P(X=1|Z=z') - 1 \\ &\quad \geq P(Y=1|Z=z_{max}) - P(Y=1|Z=z_{min}) + P(X=1|Z=z_{max}) - P(X=1|Z=z_{min}) - 1 \\ &\quad = P(Y=1|Z=z_{max}) - P(Y=1|Z=z_{min}) + \text{ST} - 1 \end{aligned}$$

and therefore the lower bound is bounded below by $\text{LB} \geq P(Y = 1|Z = z_{max}) - P(Y = 1|Z = z_{min}) + \text{ST} - 1$.

Similarly, for the third expression in the upper bound,

$$\begin{aligned} &\min_{z \neq z'} P(Y=1|Z=z) - P(Y=1|Z=z') - P(X=1|Z=z) + P(X=1|Z=z') + 1 \\ &\quad \leq P(Y=1|Z=z_{max}) - P(Y=1|Z=z_{min}) - P(X=1|Z=z_{max}) + P(X=1|Z=z_{min}) + 1 \\ &\quad = P(Y=1|Z=z_{max}) - P(Y=1|Z=z_{min}) - \text{ST} + 1 \end{aligned}$$

and so the upper bound is bounded above by $\text{UB} \leq P(Y = 1|Z = z_{max}) - P(Y = 1|Z = z_{min}) - \text{ST} + 1$.

This leads us to conclude that the length can be bounded from above:

$$\begin{aligned}\text{length} = \text{UB} - \text{LB} &\leq P(Y=1|Z=z_{max}) - P(Y=1|Z=z_{min}) - \text{ST} + 1 \\ &\quad - (P(Y=1|Z=z_{max}) - P(Y=1|Z=z_{min}) + \text{ST} - 1) \\ &= 2 - 2 \cdot \text{ST}.\end{aligned}$$

□

2.D Simulation Setup and Results

Here we provide details on the simulation used to obtain the results presented in Section 2.3.1.

Since GWAS results are most often reported as summary statistics and coefficients from a logistic model, we use monte carlo integration to show the relationship between ST and coefficients in a logistic model. We use the model introduced in Section 2.2.2 with $p = 1$. Throughout, we set $\gamma_0 = -\gamma_1$ and $\beta_0 = -\beta_X/2$. This is done to maximize the differences between probabilities $P(X = 1|Z = z)$, $z = 0, 1, 2$, and $P(Y = 1|Z = z)$, $z = 0, 1, 2$. For simplicity, we also keep $\beta_U = \gamma_U$.

For each combination of values of the coefficients $\gamma_1, \gamma_U, \beta_X$ listed below, $10,000,000$ realizations of the unmeasured confounder U are drawn from a standard normal distribution. For each realization, a value of Z is drawn such that $P(Z = 0) = P(Z = 2) = 0.25$, and $P(Z = 1) = 0.5$. Next, values of X and Y are generated using these values such that $\text{logit}(P(X = 1|Z = z, U = u)) = \gamma_0 + \gamma_1 z + \gamma_U u$ and $\text{logit}(P(Y = 1|X = x, U = u)) = \beta_0 + \beta_X x + \beta_U u$. This results in $10,000,000$ realizations of (X, Y, Z, U). From these, we find the marginal probabilities $P(X = 1|Z = z)$ and $P(Y = 1|Z = z)$, $z = 0, 1, 2$, the values of $\text{ST} = \max_{z \neq z'} |P(X = 1|Z = z) - P(X = 1|Z = z')|$ and the $\text{ATE} = P(Y = 1|X = 1) - P(Y = 1|X = 0)$.

Table 2.9: The monte carlo integration was performed for all combinations of values of the coefficients γ_1, γ_U, and β_X presented below.

β_1	γ_1	γ_U
0.25, 0.5, 1, 1.5, 2, 4, 6	0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 2.2, 2.4, 2.6, 2.8, 3, 3.2, 3.4, 3.6, 3.8, 4, 4.2, 4.4, 4.6, 4.8, 5, 5.2, 5.4, 5.6, 5.8, 6	0.1, 0.5, 1, 2

Each set of marginal probabilities leads us to a set of non-parametric bounds from two-sample data. These are shown on Figure 2.9 together with the ATE, while Figure 2.1b shows the values of γ_1 plotted against ST.

To find the smallest value of γ_1 that results in bounds excluding 0, we fit a loess curve to the lower bounds in Figure 2.9, and find the value where this curve crosses 0. This results in the values depicted on Figure 2.2.

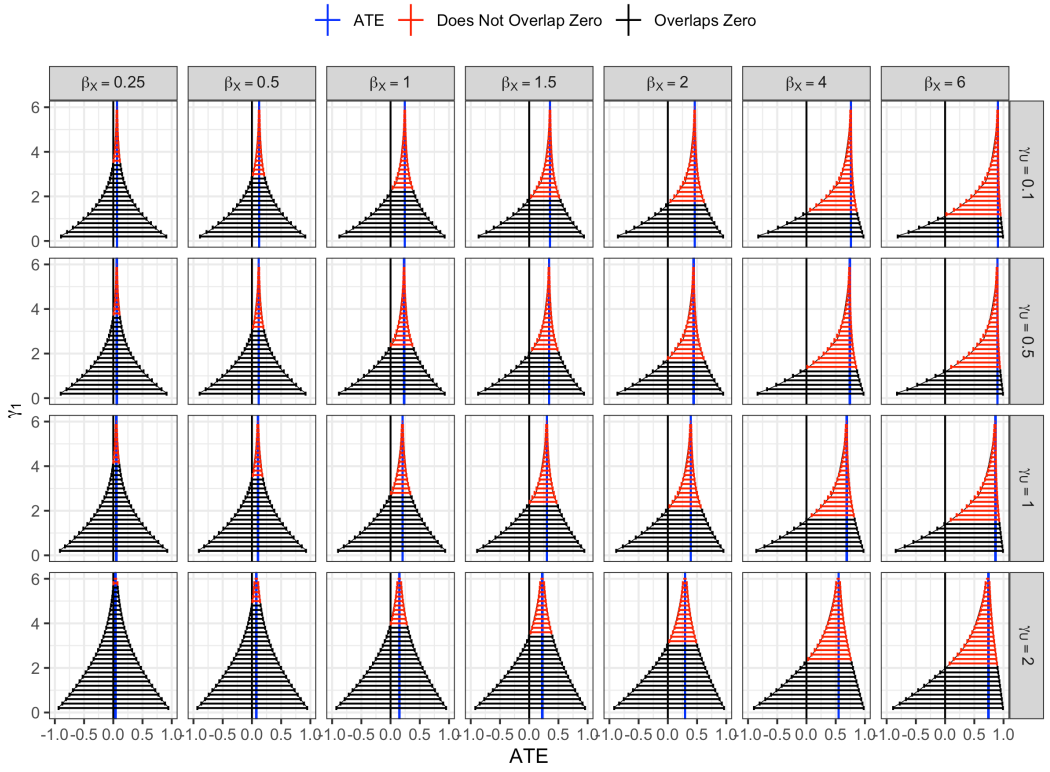


Figure 2.9: Bounds based on simulations as described. Upper and lower bounds are connected by a curve (dotted lines) based on a loess extrapolation. This curve is used to find the smallest coefficients needed to detect direction as plotted on Figure 2.2.

2.E Nestedness of Nonparametric Bounds under Monotonicity

Here we will show that the nonparametric bounds are nested as instrument strength increases when the logistic regression model holds for both exposure and outcome, and the instruments are independent and identically distributed. In particular, we assume that

$$\text{logit}(P(X = 1|Z_1 = z_1, ..., Z_p = z_p, U = u)) = \gamma_0 + \sum_{j=1}^{p} \gamma_j z_j + \gamma_U u,$$

and

$$\text{logit}(P(Y = 1|X = x, U = u)) = \beta_0 + \beta_X x + \beta_U u.$$

Finally, we assume that $P(Z_i = z) = p_z$ for $z = 0, 1, 2$ for any i.

Under the exposure and outcome models specified, assumptions (A5) and (A6) hold. We note that the bounds under (A5) and (A6) can be written as

$$\max \left\{ \begin{array}{c} P(Y=1|Z=k) - P(Y=1|Z=0) + P(X=1|Z=k) - P(X=1|Z=0) - 1 \\ 2 \cdot P(Y=1|Z=k) - P(Y=1|Z=0) + P(X=1|Z=k) - 2 \\ P(Y=1|Z=k) - 2 \cdot P(Y=1|Z=0) - P(X=1|Z=0) \end{array} \right\} \begin{array}{c}(L1)\\(L2)\\(L3)\end{array}$$

$$\leq ATE \leq$$

$$\min \left\{ \begin{array}{c} 1 - P(Y=1|Z=0) + P(X=1|Z=0) \\ P(Y=1|Z=k) - P(Y=1|Z=0) - P(X=1|Z=k) + P(X=1|Z=0) + 1 \\ P(Y=1|Z=k) - P(X=1|Z=k) + 1 \end{array} \right\} \begin{array}{c}(U1)\\(U2)\\(U3)\end{array}$$

when $k = 2, 3, 4$. We can rewrite $P(Y = 1|Z = z)$ as a function of $P(X = 1|Z = z)$ as follows:

$$\begin{aligned} P(Y = 1|Z = z) &= \sum_x P(Y = 1|X = x, Z = z)P(X = x|Z = z) \\ &= P(Y = 1|X = 0, Z = z)\{1 - P(X = 1|Z = z)\} + P(Y = 1|X = 1, Z = z)P(X = 1|Z = k) \\ &= P(X = 1|Z = z) \cdot (P(Y = 1|X = 1, Z = z) - P(Y = 1|X = 0, Z = z)) + P(Y = 1|X = 0, Z = z) \\ &= P(X = 1|Z = z) \cdot (E_U[E[Y|X = 1, U = u]] - E_U[E[Y|X = 0, U = u]]) + E_U[E[Y|X = 0, U = u]] \\ &= P(X = 1|Z = z) \cdot \text{ATE} + E_U[E[Y|X = 0, U = u]] \end{aligned}$$

Assumptions (A5) and (A6) imply that ATE ≥ 0. Now, we can rewrite the bounds as a function of $P(X = 1|Z = k), P(X = 1|Z = 0)$ and ATE:

$$\max \left\{ \begin{array}{c} (1+ATE)(P(X=1|Z=k)-P(X=1|Z=0))-1 \\ (2\cdot \text{ATE}+1)P(X=1|Z=k)+E_U\left[E[Y|X=0,U=u]\right]-\text{ATE}\cdot P(X=1|Z=0)-2 \\ P(X=1|Z=k)\cdot \text{ATE}-P(X=1|Z=0)\cdot(2\cdot\text{ATE}+1)-E_U\left[E[Y|X=0,U=u]\right] \end{array} \right\} \begin{array}{c}(L1)\\(L2)\\(L3)\end{array}$$

$$\leq ATE \leq$$

$$\min \left\{ \begin{array}{c} 1+P(X=1|Z=0)\cdot(1-\text{ATE})-E_U\left[E[Y|X=0,U=u]\right] \\ P(X=1|Z=0)(1-\text{ATE})-P(X=1|Z=k)\cdot(1-\text{ATE})+1 \\ -P(X=1|Z=k)\cdot(1-\text{ATE})+E_U\left[E[Y|X=0,U=u]\right]+1 \end{array} \right\} \begin{array}{c}(U1)\\(U2)\\(U3)\end{array}$$

This means that the lower bound is non-decreasing as a function of $P(X=1|Z=k)$ and non-increasing as a function of $P(X=1|Z=0)$. Since $1-\text{ATE}\geq 0$, the upper bound is non-increasing as a function of $P(X=1|Z=k)$ and non-decreasing as a function of $P(X=1|Z=0)$.

In a logistic model where all instruments $Z_1,...,Z_p$ are independent and have the same distribution, i.e. $P(Z_i=z)=p_z$ for $i=1,...,p$ and $z=0,1,2$, the rank of the coefficients $\gamma_1,...,\gamma_p$ gives the rank of $P(X=1|Z_j=2)$ and $-P(X=1|Z_j=0)$, i.e. the instrument with the largest coefficient has the largest value of $P(X=1|Z_j=2)$ and the smallest value of $P(X=1|Z_j=0)$. To see this, assume without loss of generality $\gamma_1\geq\gamma_j$ for any j. Then

$$\begin{aligned} &P(X=1|Z_1=2,Z_2=z_2,...,Z_p=z_p,U=u) \\ &\quad=\text{expit}\left(\gamma_0+2\cdot\gamma_1+\sum_{j=2}^{p}\gamma_j z_j+\gamma_U u\right) \\ &\quad\geq\text{expit}\left(\gamma_0+\gamma_1 z_i+2\cdot\gamma_i+\sum_{j\notin\{1,i\}}^{p}\gamma_j z_j+\gamma_U u\right) \\ &\quad=P(X=1|Z_1=z_i,Z_i=2,Z_2=z_2,...,Z_p=z_p,U=u). \end{aligned}$$

By the law of total expectation,

$$\begin{aligned} &P(X=1|Z_1=2,Z_2=z_2,...,Z_p=z_p) \\ &\quad=E[X|Z_1=2,Z_2=z_2,...,Z_p=z_p] \\ &\quad=E_U[E[X|Z_1=2,Z_2=z_2,...,Z_p=z_p,U=u]] \\ &\quad=E_U[P(X=1|Z_1=2,Z_2=z_2,...,Z_p=z_p,U=u)] \\ &\quad\geq E_U[P(X=1|Z_1=z_i,Z_i=2,Z_2=z_2,...,Z_p=z_p,U=u)] \\ &\quad=P(X=1|Z_1=z_i,Z_i=2,Z_2=z_2,...,Z_p=z_p). \end{aligned}$$

Using Bayes' Theorem, the laws of probability, and the independence of the instruments,

$$
\begin{aligned}
P(X=1|Z_1=2) &= \frac{P(X=1,Z_1=2)}{P(Z_1=2)} \\
&= \frac{\sum_{z_2,\ldots,z_p} P(X=1,Z_1=2,Z_2=z_2,\ldots,Z_p=z_p)}{P(Z_1=2)} \\
&= \sum_{z_2,\ldots,z_p} P(X=1|Z_1=2,Z_2=z_2,\ldots,Z_p=z_p)P(Z_2=z_2)\cdots P(Z_p=z_p) \\
&= \sum_{z_2,\ldots,z_p} P(X=1|Z_1=2,Z_2=z_2,\ldots,Z_p=z_p)p_{z_2}\cdots p_{z_p} \\
&\geq \sum_{z_2,\ldots,z_p} P(X=1|Z_1=z_j,Z_j=2,Z_2=z_2,\ldots,Z_p=z_p)p_{z_2}\cdots p_{z_p} \\
&= P(X=1|Z_j=2)
\end{aligned}
$$

where the inequality follows from the inequality illustrated above. Similarly, it can be seen that $P(X=1|Z_1=0) \leq P(X=1|Z_j=0)$ for all j. Therefore, Z_1 gives the largest lower bound and smallest upper bound of any instrument. Under (A5) and (A6), instrument strength simplifies to

$$\text{ST} = \max_{z \neq z'} |P(X=1|Z=z) - P(X=1|Z=z')| = P(X=1|Z=k) - P(X=1|Z=0).$$

which means Z_1 is also the strongest instrument. So, under the conditions specified, the strongest instrument will give bounds that are nested in the bounds based on weaker instruments.

2.F Would Multiple Instruments Help? Additional Figures

The following figures provide the full results described in Section 2.3.2. Figure 2.10 illustrate how similarly sized γ_i coefficients lead to weaker instruments when p is increased. Figures 2.11, 2.12, 2.13, and 2.14 show the full results for all four scenarios.

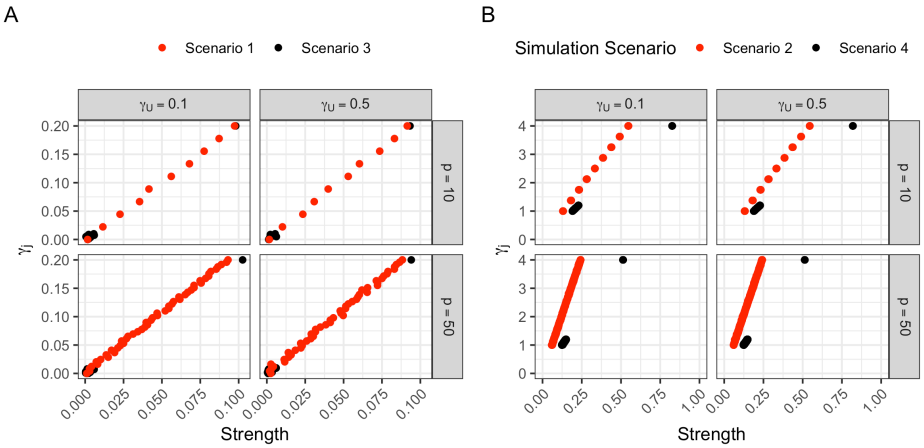


Figure 2.10: Figure showing the dilution effect described in Section 2.3.2 in each of the four scenarios. When p is larger, similar sized coefficients lead to lower strength. The effect is smaller when we are in a scenario where one coefficient is relatively much larger than the rest, rather than when the coefficients are evenly spread out. A: Scenarios 1 and 3. B: Scenarios 2 and 4.

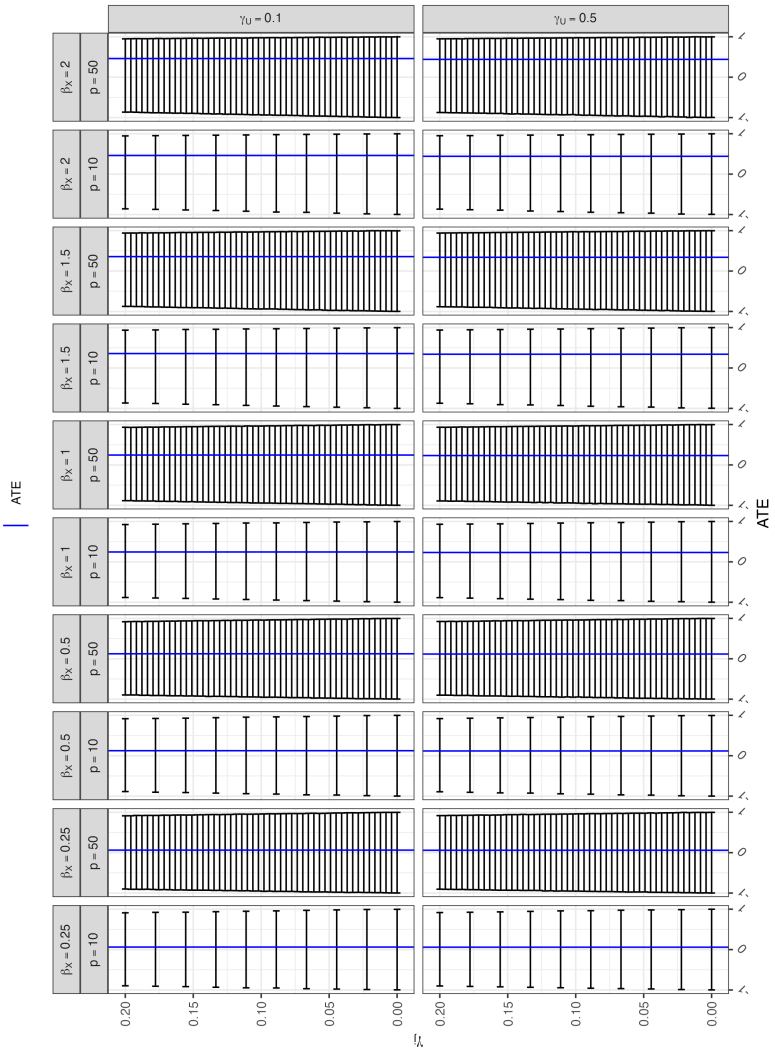


Figure 2.11: Bounds based on monte carlo integration with 1,000,000 resamples in scenario 1.

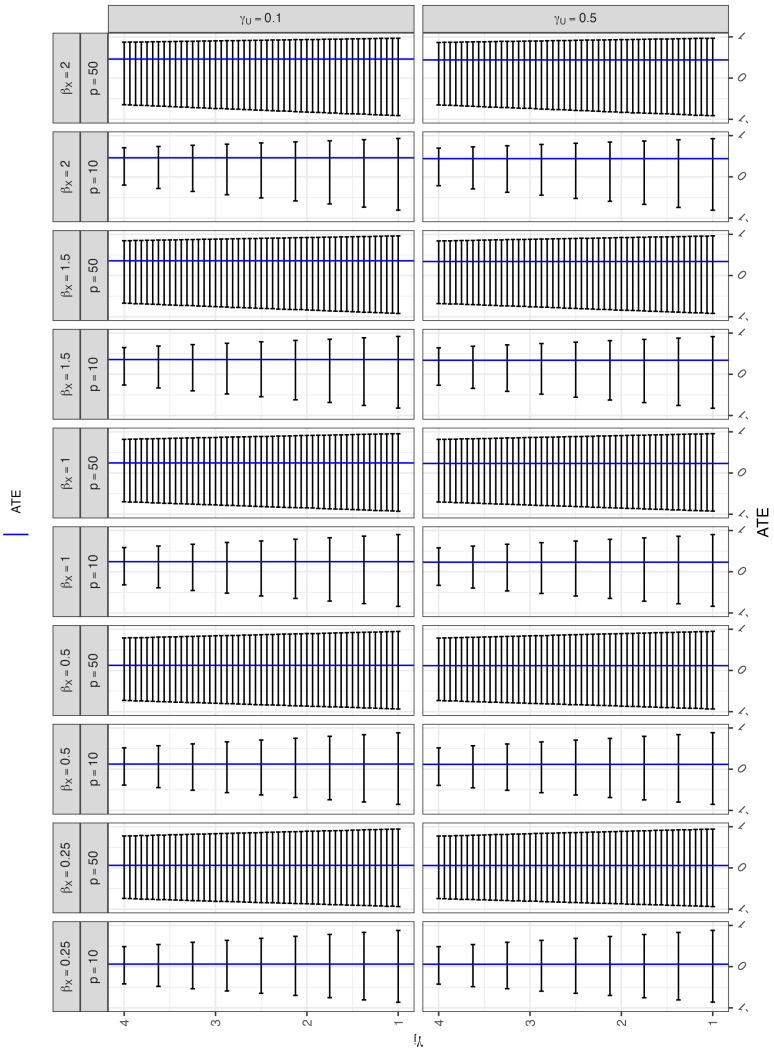


Figure 2.12: Bounds based on monte carlo integration with 1,000,000 resamples in scenario 2.

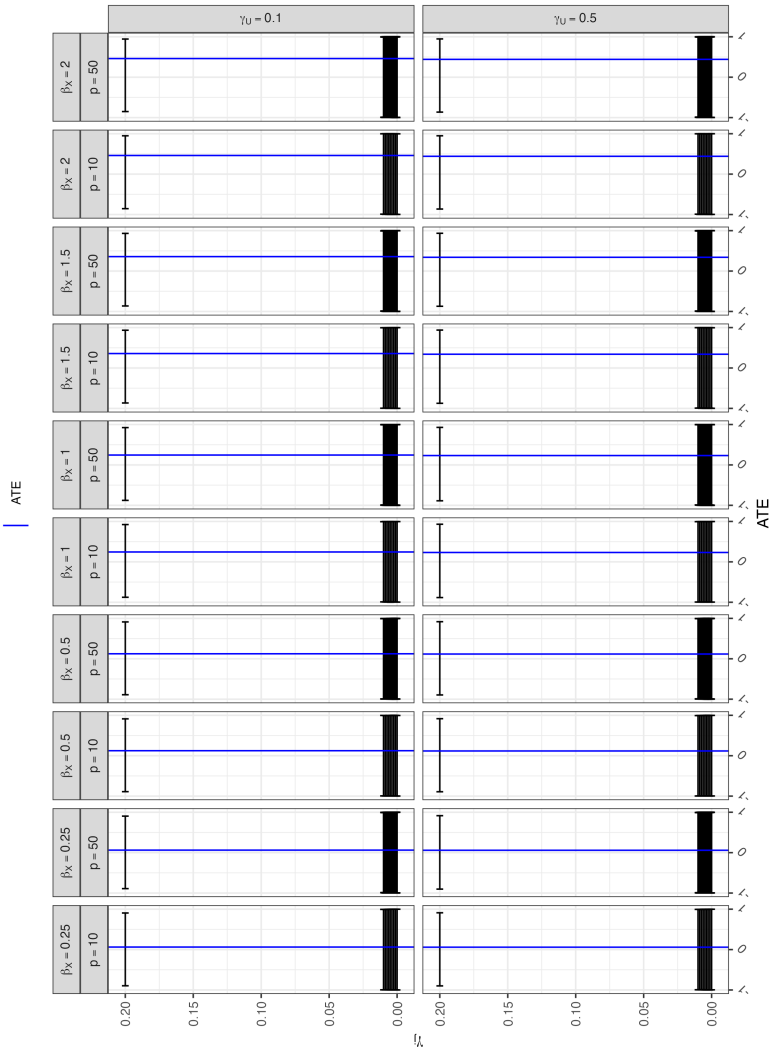


Figure 2.13: Bounds based on monte carlo integration with 1,000,000 resamples in scenario 3.

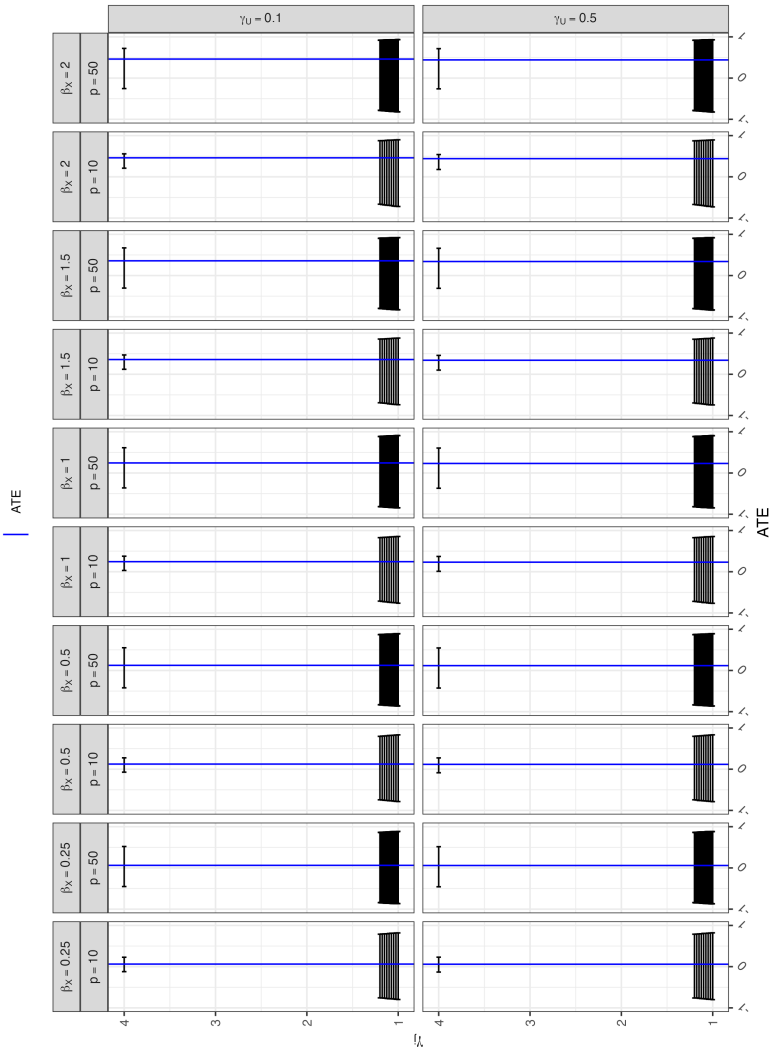


Figure 2.14: Bounds based on monte carlo integration with 1,000,000 resamples in scenario 4.

2.G Coverage Under Pleiotropic Genetic Markers

A common concern in MR studies is the possibility of using a genetic marker that is pleiotropic, i.e. that affects the likelihood of both exposure and outcome, and thereby violate the exclusion restriction assumption (A3). To assess the potential damage such a violation might bring, we simulate $10,000,000$ observation from the set of exposure and outcome models described in Section 2.3.3 with varying values of the coefficients γ_1, β_Z, and β_X. The results are presented on Figures 2.15 and 2.16. These figures show the nonparametric bounds (horizontal lines) and actual Average Treatment Effects (solid dots). Blue lines indicate that the bounds do in fact cover the true ATE, while red lines indicate that the bounds do not. Finally, the lines are dotted if any of the verifiable constraints are violated in the given scenario.

These two figures reveal a few noteworthy observations. For coefficients in line with what we see in MR studies (Figure 2.15), there are two important points to make. First, we see that the ATE is always covered by the bounds. In other words, pleiotropy on the scale we would expect to see in MR analyses do not cause bounds to not cover the ATE. Second, the verifiable constraints are not powerful enough to detect violation of the exclusion restriction in most MR studies.

For Figure 2.16, we pushed the size of the coefficients to a scale that would provide more information about the power of the verifiable constraints, and the power of the nonparametric bounds. We see that either the effect of the instrument on the outcome or the ATE has to be very large for the bounds to not cover the ATE. Likewise, the effect of the instrument on the outcome has to be large for the verfiable constraints to be violated.

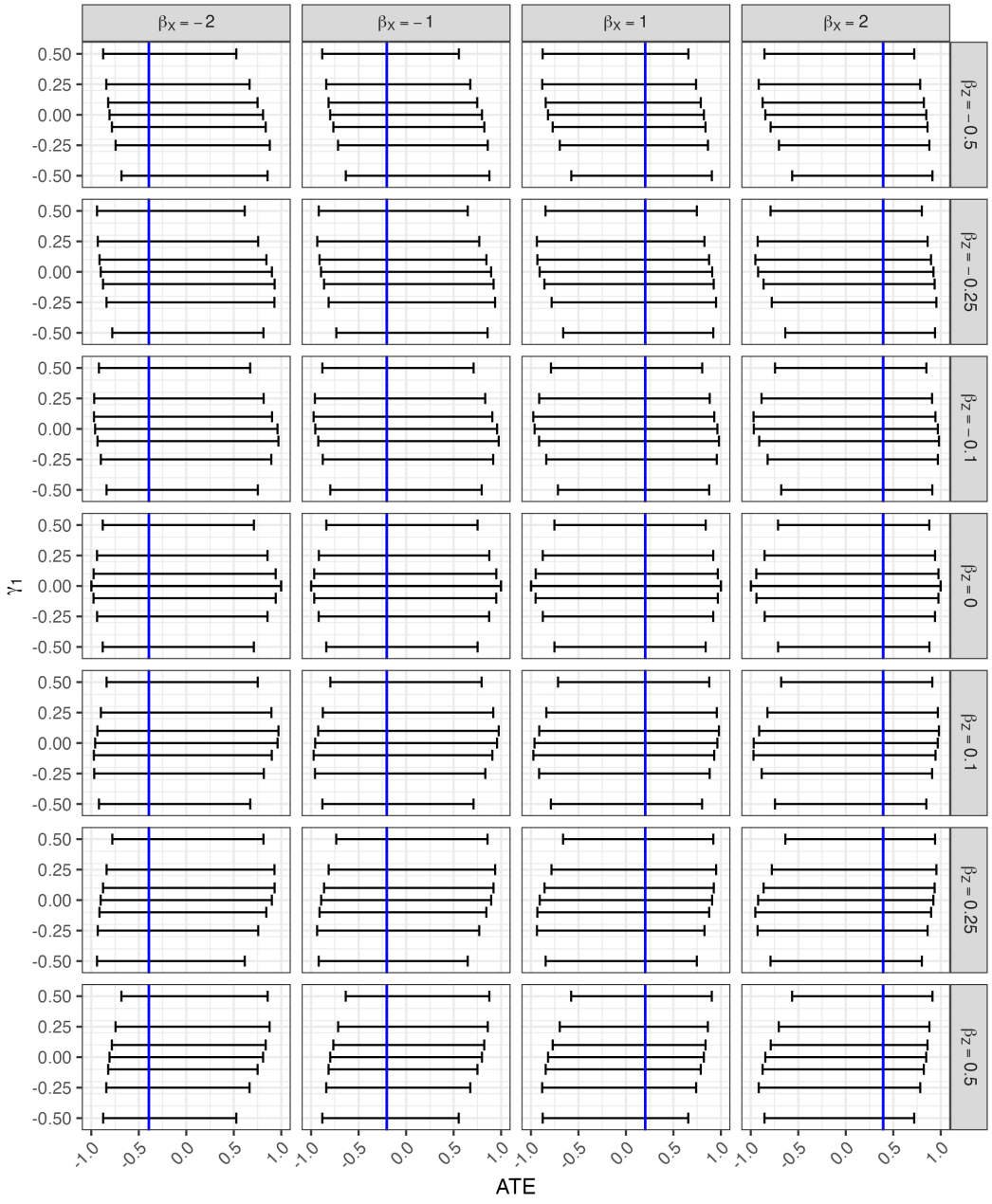


Figure 2.15: Two-sample bounds (horizontal lines) and average treatment effects (vertical blue lines) under pleiotropy. Columns represent effect size of the exposure on the logit scale, rows represent the magnitude of the violation of assumption (A3). X-axis show average treatment effect, and Y-axis represent instrument strength.

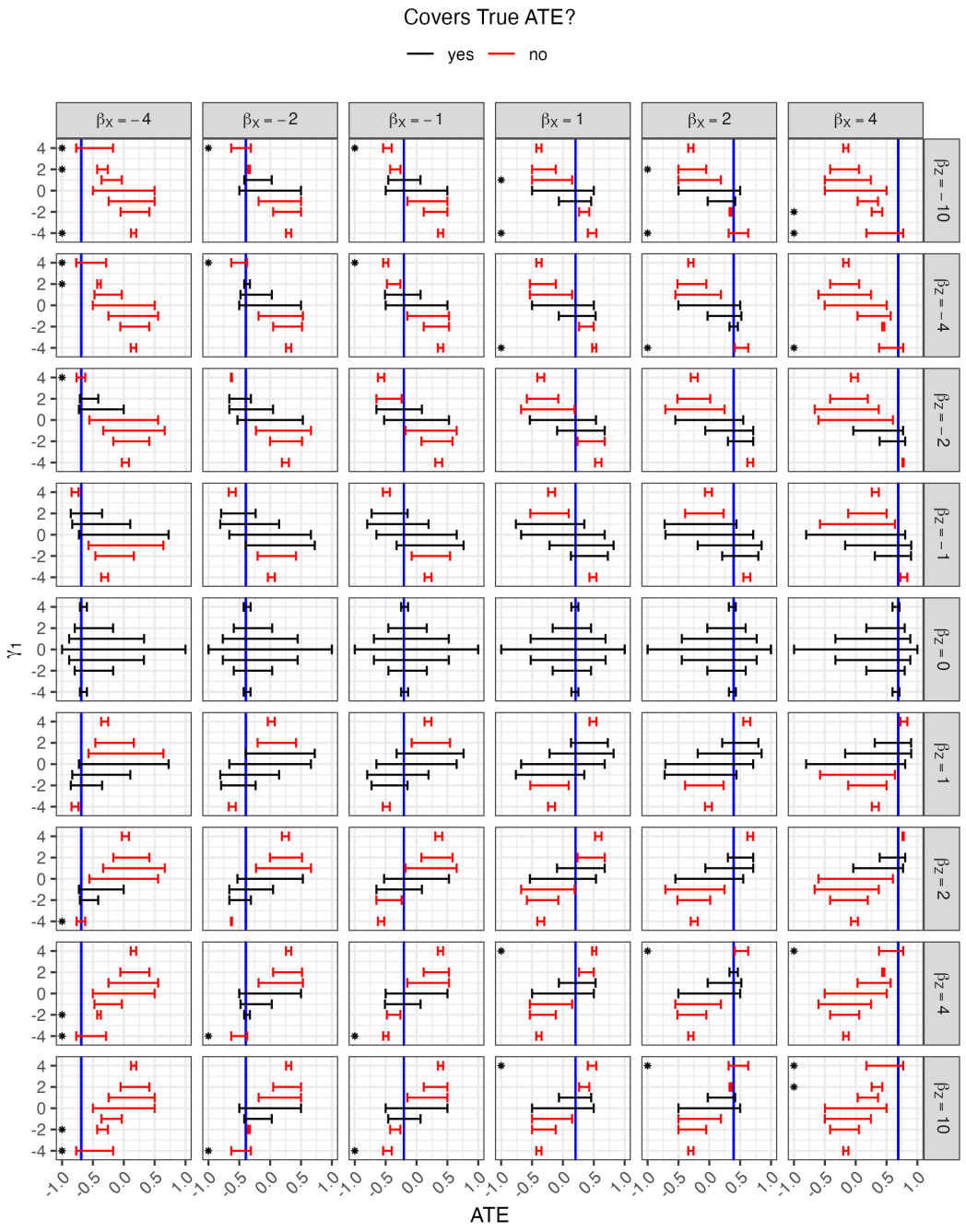


Figure 2.16: Two-sample bounds (horizontal lines) and average treatment effects (vertical blue lines) under pleiotropy. Columns represent effect size of the exposure on the logit scale, rows represent the magnitude of the violation of assumption (A3). X-axis show average treatment effect, and Y-axis represent instrument strength. Cases where any of the verifiable constraints were violated are denoted by an asterix to the left.

2.H Dichotomized Exposure Variable

In many MR studies, the exposure and/or the outcome is dichotomization of a continuous variable. For example, in Section 2.5 we explore the effect of having high cholesterol on the incidence of heart attack. The exposure variable in this example is a dichotomization of the continuous measure of cholesterol.

In this section, we will explore the effect of using a dichotomized version of a continuous variable in nonparametric bounds with a focus on detection of the direction of the causal effect. The main takeaway is that if the analysis using the dichotomized version is able to conclude the direction of the causal effect, the detected direction is the same as the direction of the causal effect of the underlying continuous variable. There are other concerns one should be aware of when using a dichotomization of a continuous variable as a binary exposure (Burgess and Labrecque, 2018).

2.H.1 Theoretical Setup

Let X be a continuous exposure, and Y^x the binary potential outcome. For simplicity, we will work under the monotonicity assumption that $P(Y^x \leq Y^{x+\epsilon}) = 1$ for all x and all $\epsilon > 0$. This assumption simplifies the math, and it is often implicitly made in mendelian randomization studies that rely on regression frameworks. Also note that this assumption gives us the direction of the causal effect of X on Y. In this case, it is positive. Next, let $\widetilde{X}$ be a dichotomization of the exposure, that is $\widetilde{X} = 1[X \geq c]$ for some c.

Now, we define the potential outcomes under the dichotomized exposure as being one of the potential outcomes under the continuous exposure. I.e. $\widetilde{Y}^1 = Y^x$ for some $x \geq c$, and $\widetilde{Y}^0 = Y^x$ for some $x < c$.

If we consider the average treatment effect under the dichotomized exposure, we see that

$$\begin{aligned} E[\widetilde{Y}^1 - \widetilde{Y}^0] &= \sum_{k \in \{-1,0,1\}} k \cdot P(\widetilde{Y}^1 - \widetilde{Y}^0 = k) \\ &= P(\widetilde{Y}^1 - \widetilde{Y}^0 = 1) - P(\widetilde{Y}^1 - \widetilde{Y}^0 = -1) \end{aligned}$$

Note that if $\widetilde{Y}^1 - \widetilde{Y}^0 = -1$, then $\widetilde{Y}^1 < \widetilde{Y}^0$. Since $\widetilde{Y}^1 = Y^{x'}$ for some $x' \geq c$, and $\widetilde{Y}^0 = Y^{x''}$ for some $x'' < c$,

$$P(\widetilde{Y}^1 < \widetilde{Y}^0) = P(Y^{x'} < Y^{x''}) = 0$$

since $P(Y^x \leq Y^{x+\epsilon}) = 1$ for all x, and all $\epsilon > 0$. Therefore, $P(\widetilde{Y}^1 - \widetilde{Y}^0 = -1) = 0$, which leads us to conclude that

$$E[\widetilde{Y}^1 - \widetilde{Y}^0] = P(\widetilde{Y}^1 - \widetilde{Y}^0 = 1) \geq 0$$

and therefore the sign of the causal effect is preserved.

Similar calculations show that if we instead assume $P(Y^x \geq Y^{x+\epsilon}) = 1$, then

$$E[\widetilde{Y}^1 - \widetilde{Y}^0] = -P(\widetilde{Y}^1 - \widetilde{Y}^0 = -1) \leq 0$$

2.H.2 Simulation Results

To investigate what behavior we can expect in our logistic regression framework, we carried out a simulation study. We use a very similar framework as what is introduced in Section 2.2.2. In particular, we let the exposure follow

$$\widetilde{X} = \gamma_0 + \gamma_1 Z + U + \epsilon_{\widetilde{X}},$$

where Z is the instrument taking values in $\{0, 1, 2\}$, U is a binary unmeasured confounder, and $\epsilon_{\widetilde{X}}$ is the error term following the standard normal distribution. The outcome follows the logistic model

$$\text{logit}(P(Y = 1|\widetilde{X} = x, U = u)) = \beta_0 + \beta_{\widetilde{X}} x + u + \epsilon_Y.$$

where ϵ_Y is the error term following the standard normal distribution.

Based on these two models, we simulate $10,000,000$ observations for each combination of parameters given in Table 2.10 based on which we find $P(Y = 1|Z = z)$, and $P(X = 1|Z = z)$, where $X = 1[\widetilde{X} < M_{\widetilde{X}}]$, where $M_{\widetilde{X}}$ is the median of $\widetilde{X}$.

The resulting bounds can be found in Figures 2.17-2.21. We notice that when we are able to detect a direction of the ATE, it is consistently the correct direction.

Table 2.10: Parameters used in simulation to explore the behavior of the bounds when using an exposure that is a dichotomization of an underlying continuous variable.

γ_0	γ_1	β_0	$\beta_{\widetilde{X}}$
-2, -1, 0, 1, 2	-2, -1, 0, 1, 2	-2, -1, 0, 1, 2	-2, -1, 0, 1, 2

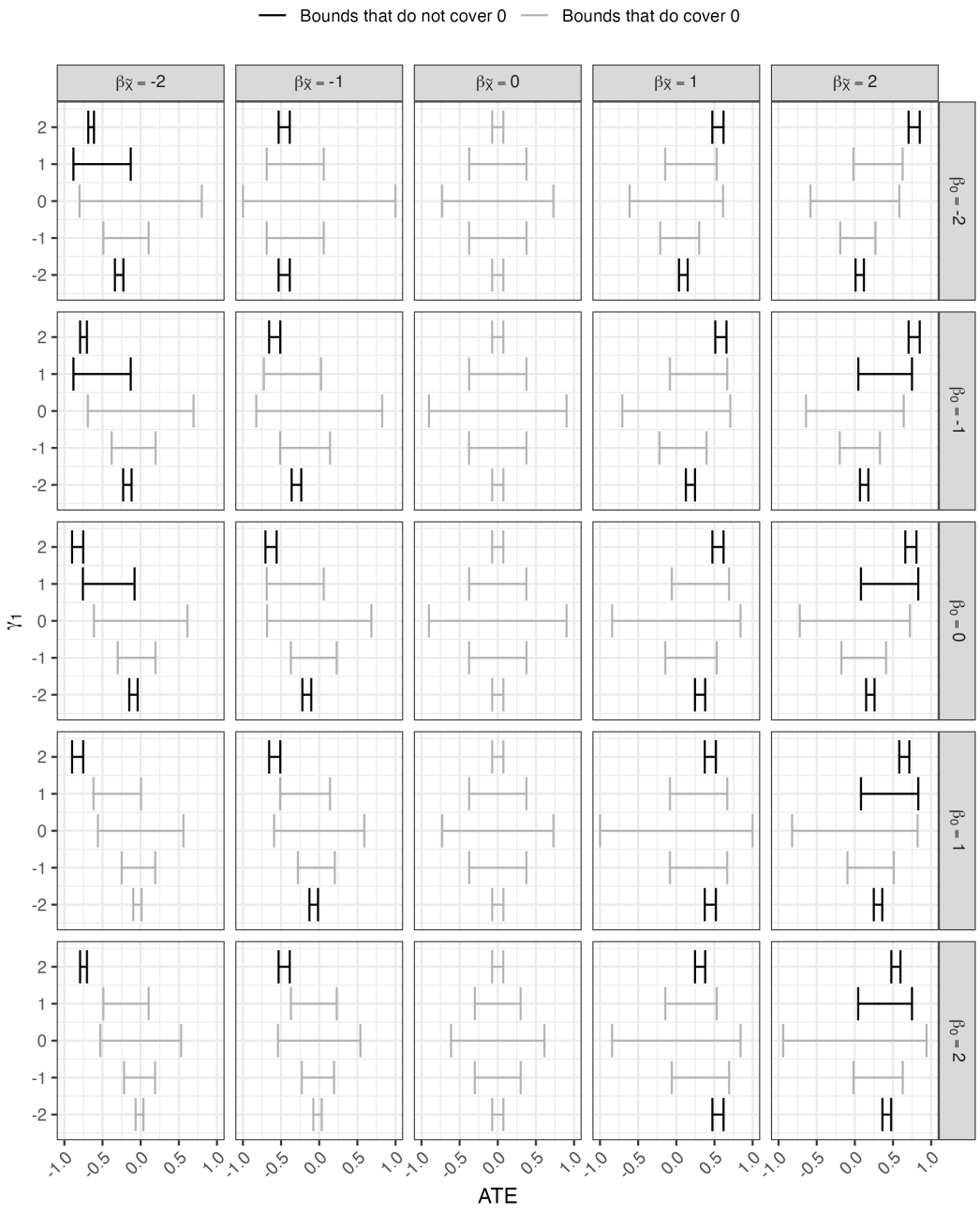


Figure 2.17: Nonparametric bounds based on a dichotomized exposure; here $\gamma_0 = -2$. Columns represent the effect size of the exposure on the logit scale. Rows represent different values of the intercept β_0. The y-axis shows the effect of the instrument on the continuous exposure, and the x-axis shows the average treatment effect.

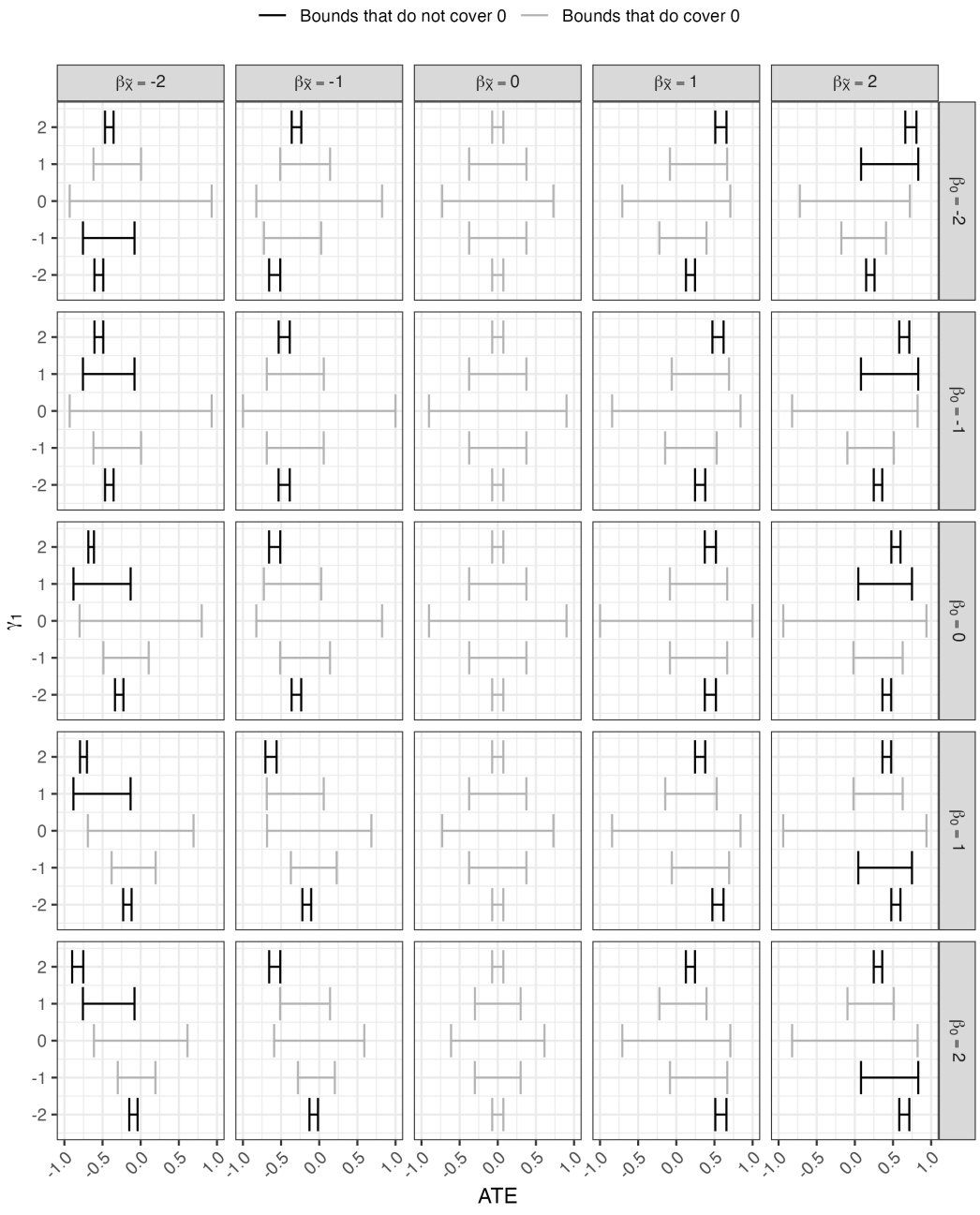


Figure 2.18: Nonparametric bounds based on a dichotomized exposure; here $\gamma_0 = -1$. Columns represent the effect size of the exposure on the logit scale. Rows represent different values of the intercept β_0. The y-axis shows the effect of the instrument on the continuous exposure, and the x-axis shows the average treatment effect.

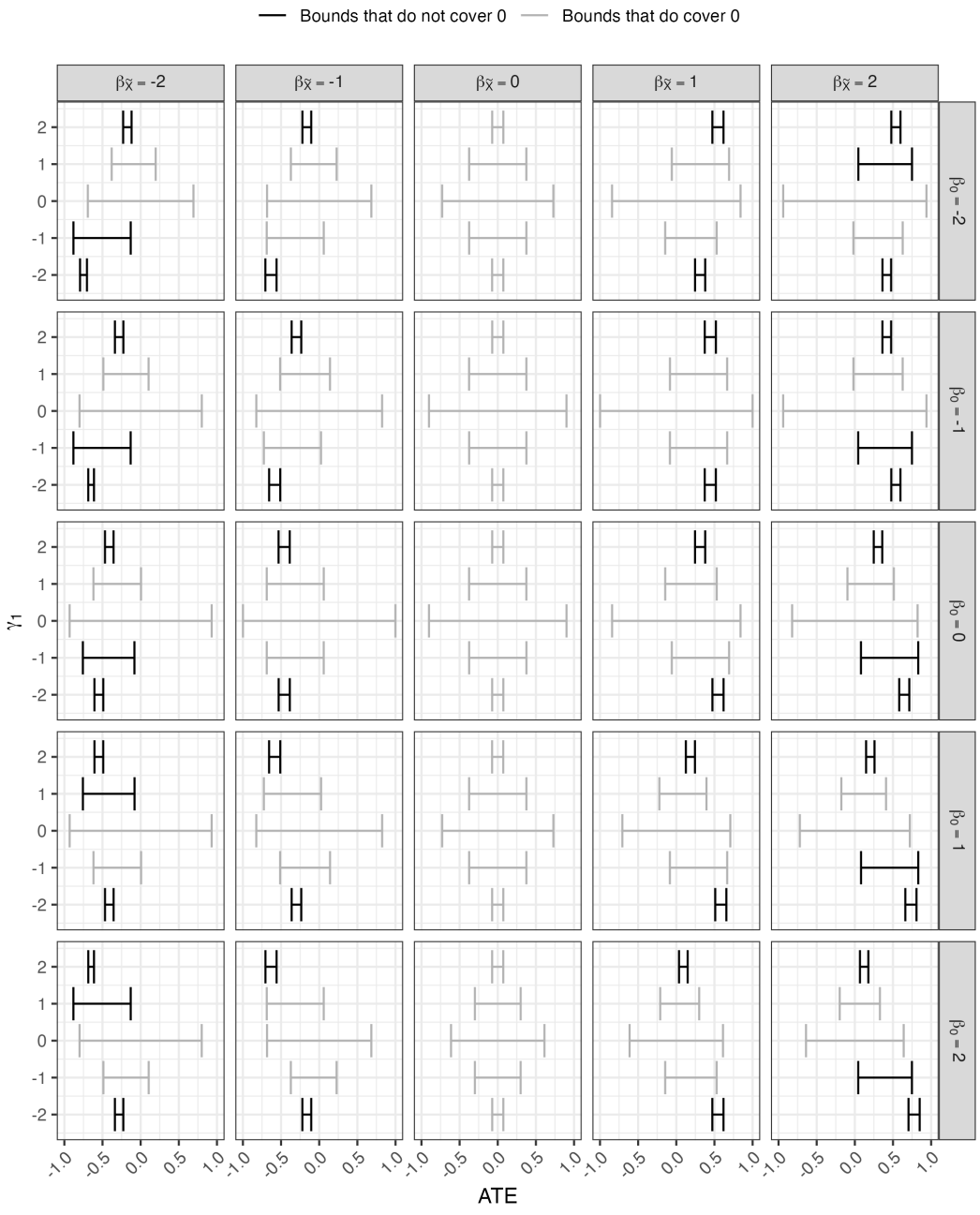


Figure 2.19: Nonparametric bounds based on a dichotomized exposure; here $\gamma_0 = 0$. Columns represent the effect size of the exposure on the logit scale. Rows represent different values of the intercept β_0. The y-axis shows the effect of the instrument on the continuous exposure, and the x-axis shows the average treatment effect.

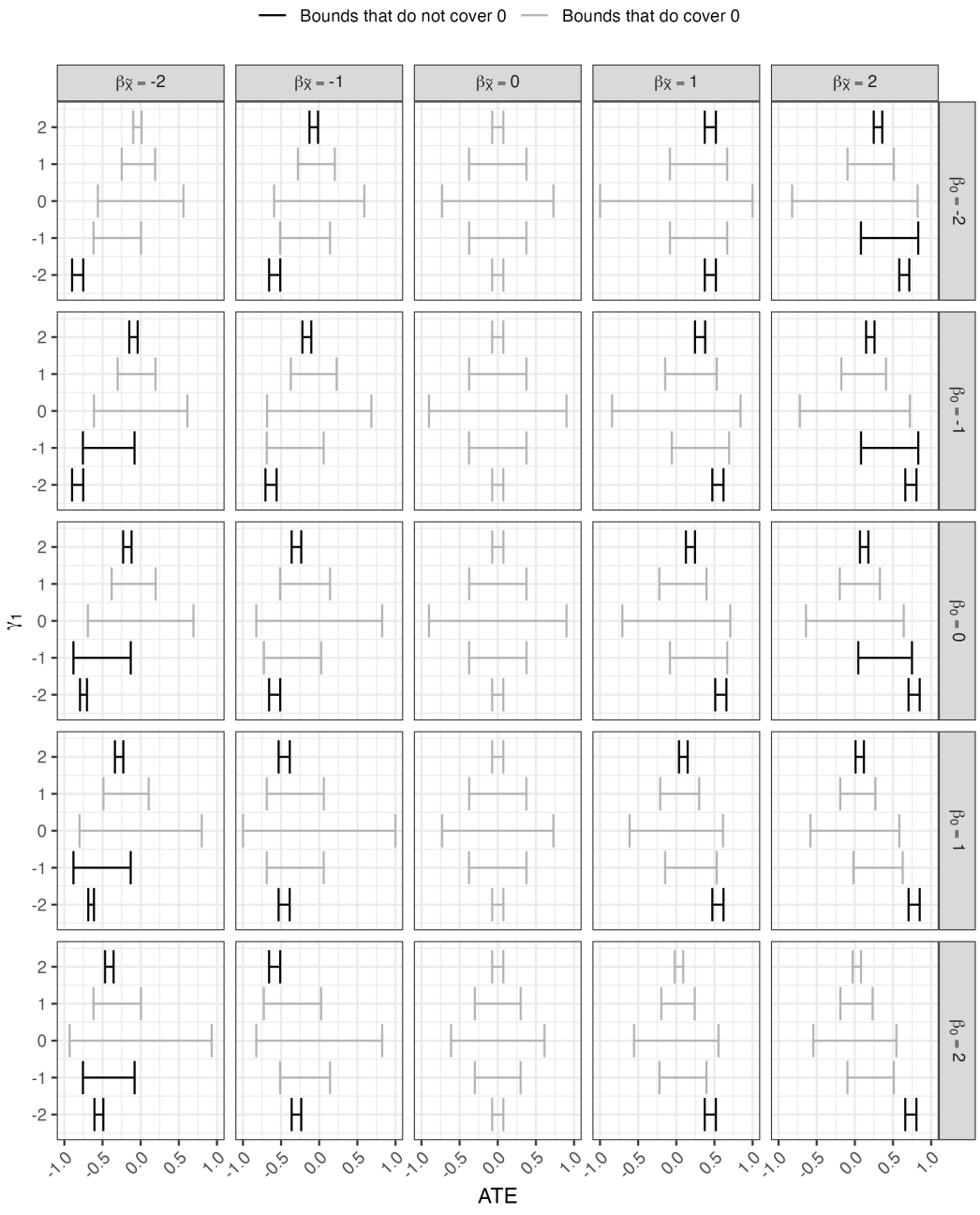


Figure 2.20: Nonparametric bounds based on a dichotomized exposure; here $\gamma_0 = 1$. Columns represent the effect size of the exposure on the logit scale. Rows represent different values of the intercept β_0. The y-axis shows the effect of the instrument on the continuous exposure, and the x-axis shows the average treatment effect.

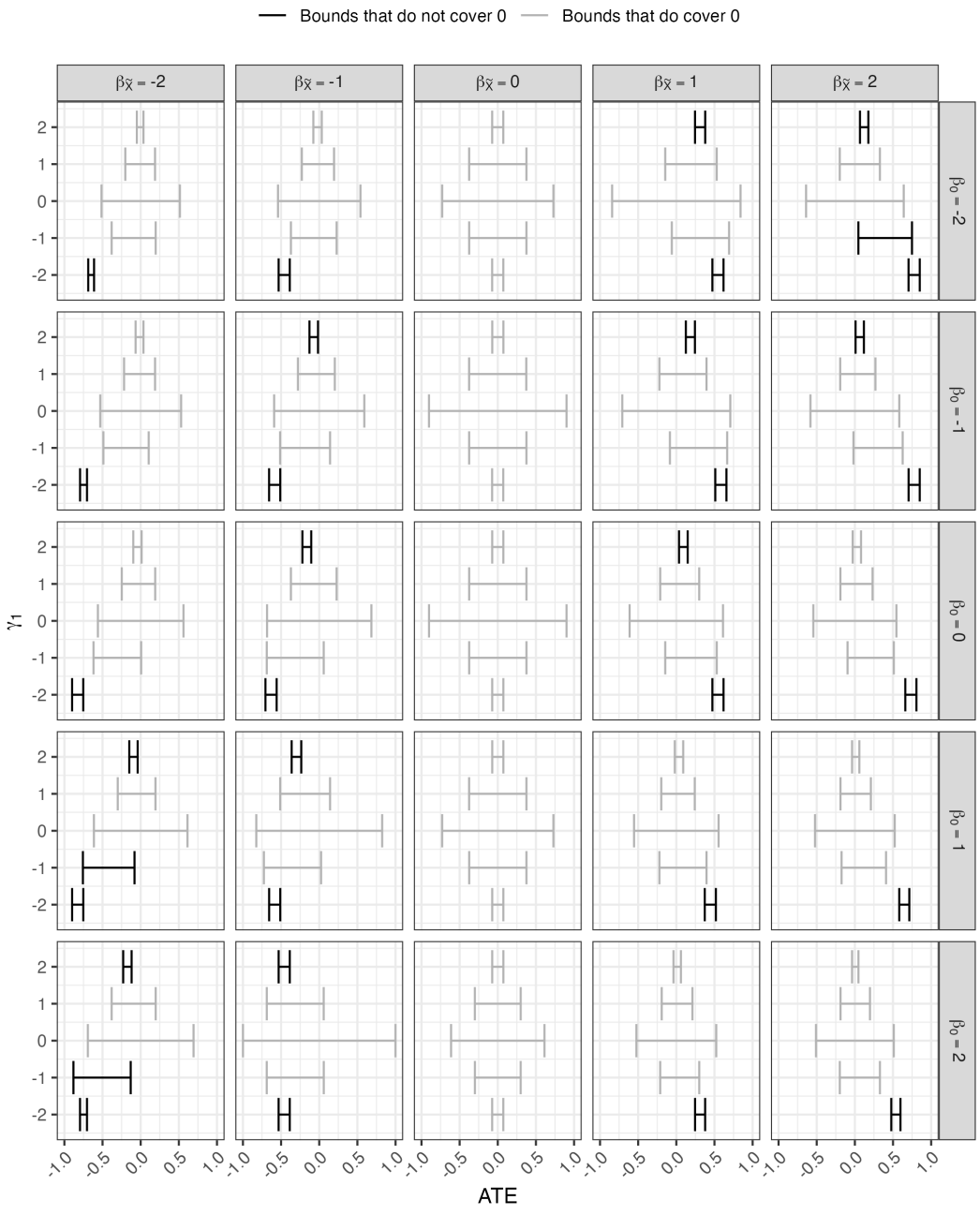


Figure 2.21: Nonparametric bounds based on a dichotomized exposure; here $\gamma_0 = 2$. Columns represent the effect size of the exposure on the logit scale. Rows represent different values of the intercept β_0. The y-axis shows the effect of the instrument on the continuous exposure, and the x-axis shows the average treatment effect.

2.H.3 Finite-Sample Behavior of Bounds With Estimated Probabilities

We use the same exposure and outcome models as presented in Section 2.2.3. In particular, we let the exposure X be binary with

$$\text{logit}(P(X = 1|Z = z, U = u)) = \gamma_0 + \gamma_1 \cdot z + u$$

and the outcome Y be binary with

$$\text{logit}(P(Y = 1|X = x, U = u)) = \beta_0 + \beta_X \cdot x + u.$$

We vary $\gamma_1 \in \{-4, -2, -1 - 0.5, -0.25, -0.1, 0, 0.1, 0.25, 0.5, 1, 2, 4\}$ and $\beta_X \in \{-4, -2, -1, 1, 2, 4\}$, while we set $\gamma_0 = -\gamma_1$, $\beta_0 = -\beta_X/2$ and let U follow a standard Normal. We then generate 5000 data sets with 2000 observations in each. The resulting coverages are presented in Tables 2.11 and 2.12. We see that coverage is essentially 100% unless the coefficients are very extreme. In particular, for any values that we might realistically encounter in a two-sample MR study, we would expect 100% coverage.

Table 2.11: Coverage of bounds with sample size 2000 based on 5000 simulated data sets.

γ_1	β_X	ATE	ST	Coverage
	-2	-0.393	0.205	1
	-1	-0.204	0.204	1
	1	0.204	0.204	1
	2	0.394	0.204	1
	-2	-0.393	0.103	1
	-1	-0.204	0.103	1
	1	0.204	0.103	1
	2	0.393	0.104	1
	-2	-0.393	0.041	1
	-1	-0.204	0.042	1
	1	0.204	0.041	1
	2	0.393	0.042	1
	-2	-0.393	0.001	1
	-1	-0.204	0.001	1
	1	0.204	0.001	1
	2	0.393	0.001	1
	-2	-0.393	0.042	1
	-1	-0.204	0.041	1
	1	0.204	0.041	1
	2	0.393	0.041	1
	-2	-0.393	0.103	1
	-1	-0.204	0.103	1
	1	0.204	0.103	1
	2	0.393	0.103	1
	-2	-0.393	0.204	1
	-1	-0.204	0.204	1
	1	0.204	0.204	1
	2	0.393	0.204	1

Table 2.12: Coverage of bounds with sample size 2000 based on 5000 simulated data sets.

γ_1	β_X	ATE	ST	Coverage
	-4	-0.689	0.944	0.8086
	-2	-0.393	0.944	0.8862
	-1	-0.204	0.944	0.8952
	1	0.204	0.944	0.9056
	2	0.394	0.944	0.8682
	4	0.689	0.944	0.8236
	-4	-0.689	0.689	1.0000
	-2	-0.393	0.689	1.0000
	-1	-0.204	0.689	1.0000
	1	0.204	0.689	1.0000
	2	0.394	0.689	1.0000
	4	0.689	0.689	1.0000
	-4	-0.689	0.394	1.0000
	-2	-0.393	0.394	1.0000
	-1	-0.204	0.393	1.0000
	1	0.204	0.394	1.0000
	2	0.393	0.393	1.0000
	4	0.689	0.394	1.0000
	-4	-0.689	0.393	1.0000
	-2	-0.393	0.393	1.0000
	-1	-0.204	0.393	1.0000
	1	0.204	0.393	1.0000
	2	0.393	0.394	1.0000
	4	0.689	0.394	1.0000
	-4	-0.689	0.690	1.0000
	-2	-0.393	0.689	1.0000
	-1	-0.204	0.689	1.0000
	1	0.204	0.689	1.0000
	2	0.393	0.690	1.0000
	4	0.689	0.689	1.0000
	-4	-0.689	0.944	0.8086
	-2	-0.393	0.944	0.8798
	-1	-0.204	0.944	0.8992
	1	0.204	0.944	0.8952
	2	0.394	0.944	0.8758
	4	0.689	0.944	0.8188

2.I Sampling of Intersection Bounds From Two Instruments

To extend our method for sampling plausible joint distributions of $P(X = x, Y = y|Z = z)$ to the scenario where we have multiple instruments available, we simply repeat the one instrument sampling for each instrument. This is equivalent to assuming that the covariances of X and Y given Z_1 are independent of the covariances of X and Y given Z_2. Once we have obtained bounds for each instrument, we take the intersection to get the intersection bounds.

Specifically, say we get bounds $(LB_{1i}, UB_{1i}), i = 1, 2, ..., m$ by sampling m trivariate distributions based on the information we have on (X, Z_1) and (Y, Z_1), and bounds $(LB_{2i}, UB_{2i}), i = 1, 2, ..., m$ by sampling m trivariate distributions based on the information we have on (X, Z_2) and (Y, Z_2). We then create the intersection bounds as $\left(\max_{z\in1,2} LB_{zi}, \min_{z\in1,2} UB_{zi}\right), i = 1, 2, ..., m$. This, under the assumption that $\text{Cov}(X, Y|Z_1 = z)$ and $\text{Cov}(X, Y|Z_2 = z)$ are independent of each other, gives us a sample from the posterior distribution of intersection bounds. We can use this to assess the potential usefulness of aggregating information from two sets of trivariate data, (X, Y, Z_1) and (X, Y, Z_2), using intersection bounds.

2.J Additional Summary Statistics and Figures for Analyses

We present expanded results to complement the analyses in Section 2.5.

We use the `TwoSampleMR` R package (Hemani et al., 2018) to extract and preprocess the data for our analyses. For preprocessing, we followed the defaults of the R pacakge where linkage disequilibrium based clumping ($r^2 \geq 0.001$ within a $10,000$ kb window using $p < 5 \times 10^{-8}$ as the level of significance) were performed such that only independent instruments with significant associations were used in the analysis. Afterwards, we obtain the estimated coefficients corresponding to the effects of the SNPs on the exposure and the outcome from a logistic model. Since estimates of the intercept are not included in these reported results, but the marginal proportions of the outcome, exposure, and allele frequencies are known, we find the intercepts by solving $P(X = 1) = \sum_{z=0}^{2} \text{logit}(\gamma_0 + \hat{\gamma}_j \cdot z) \cdot P(Z_j = z)$ and $P(Y = 1) = \sum_{z=0}^{2} \text{logit}(\Gamma_0 + \hat{\Gamma}_j \cdot z) \cdot P(Z_j = z)$ for Γ_0 and Γ_0, respectively. This allows us to obtain estimates of $P(Y = 1|Z_j = z)$ and $P(X = 1|Z_j = z)$ for every j and $z = 0, 1, 2$.

Data on smoking was obtained from the data entry ID ukb-d-20116_0, data on lung cancer was from data entry ID ukb-d-40001_C349, data on cholesterol was from data entry ID ukb-a-108, and data on heart attack was from data entry ID ukb-a-434.

2.J.1 Effect of Smoking on Lung Cancer

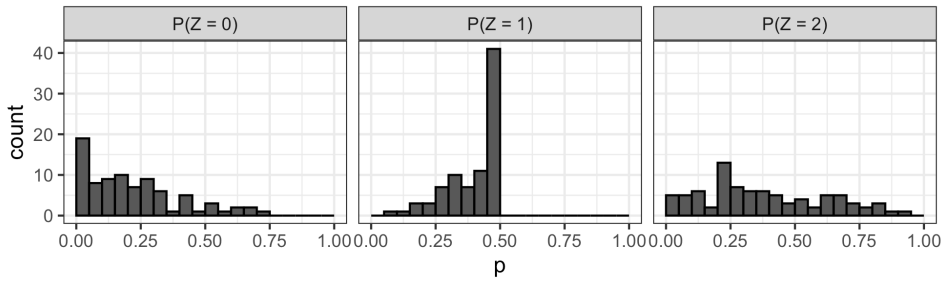


Figure 2.22: Histograms of the marginal distribution of instruments, $P(Z = z), z = 0, 1, 2$, estimated after preprocessing.

Table 2.14: Coefficients from GWAS results of logistic regression of the SNPs on smoking status and lung cancer status. Intercepts (β_0 and γ_0) are inferred, while slopes (β_X and γ_1) are as reported.

SNP	β_X	β_0	γ_1	γ_0
rs10173733	-0.0065148	0.1773766	0.0033363	-1.987122
rs10193706	-0.0117667	0.1807753	-0.0015310	-1.981684
rs10233018	-0.0076551	0.1771914	0.0050495	-1.988150
rs10274594	0.0078326	0.1617046	-0.0015364	-1.981589
rs1029986	-0.0070208	0.1754303	0.0035498	-1.986088
rs10774625	0.0074868	0.1621777	-0.0084158	-1.974806
rs10813628	-0.0068761	0.1762662	0.0051706	-1.988156
rs10897561	-0.0066917	0.1782117	0.0066835	-1.991747
rs10905461	0.0072731	0.1658787	-0.0058844	-1.980131
rs10914684	0.0077356	0.1591408	-0.0026047	-1.979616
rs10956808	0.0076247	0.1607905	-0.0063546	-1.975802
rs11103667	-0.0086047	0.1835048	0.0063118	-1.993343
rs11127913	0.0081801	0.1596256	-0.0033969	-1.978997
rs11429972	0.0083148	0.1640148	-0.0096129	-1.976695
rs11611651	-0.0119868	0.1914724	0.0013059	-1.985521
rs11631530	-0.0099863	0.1872160	-0.0047887	-1.974691
rs11646575	-0.0082446	0.1788545	0.0012319	-1.984521
rs11693702	-0.0080254	0.1781679	0.0046224	-1.988077
rs117435980	-0.0092037	0.1849986	-0.0054804	-1.973970
rs12042107	0.0071759	0.1631404	-0.0020557	-1.981288
rs12244388	-0.0104344	0.1834505	0.0019355	-1.985707
rs12450028	-0.0070626	0.1788556	-0.0024536	-1.979923
rs12479064	-0.0080362	0.1823251	-0.0088600	-1.969116
rs12487411	0.0075048	0.1616745	-0.0077980	-1.974913
rs12608052	0.0067542	0.1631129	-0.0048100	-1.978521
rs12725407	0.0081386	0.1564297	-0.0067998	-1.972138

Table 2.14: Coefficients from GWAS results of logistic regression of the SNPs on smoking status and lung cancer status. Intercepts (β_0 and γ_0) are inferred, while slopes (β_X and γ_1) are as reported. *(continued)*

SNP	β_X	β_0	γ_1	γ_0
rs12886628	-0.0071010	0.1743626	-0.0018595	-1.981891
rs12910916	-0.0090138	0.1838027	0.0026458	-1.987308
rs13100688	0.0072663	0.1604864	-0.0055464	-1.976186
rs1492546	-0.0068801	0.1757890	0.0040638	-1.986797
rs1499982	-0.0114648	0.1730098	0.0024892	-1.983878
rs1549213	0.0085270	0.1634849	0.0056335	-1.987184
rs1561195	-0.0078947	0.1771393	0.0072232	-1.990046
rs1565735	0.0115901	0.1510915	-0.0072487	-1.971566
rs16951001	-0.0066035	0.1772784	0.0070226	-1.991313
rs17003752	0.0098606	0.1526117	-0.0055424	-1.973591
rs17151637	0.0075112	0.1588020	-0.0027771	-1.979146
rs1899896	-0.0079928	0.1808293	0.0047935	-1.989876
rs2240294	0.0069566	0.1618616	-0.0078381	-1.974429
rs2416770	-0.0064888	0.1756858	-0.0035668	-1.979794
rs264974	0.0093111	0.1600323	-0.0047198	-1.978291
rs2675609	0.0081586	0.1635228	-0.0069708	-1.977953
rs2797116	0.0079136	0.1580011	-0.0039635	-1.977330
rs2867749	0.0069446	0.1601396	-0.0032894	-1.978658
rs299688	-0.0072721	0.1737306	-0.0019058	-1.982055
rs326341	0.0065809	0.1627032	0.0031753	-1.986468
rs35891966	0.0147752	0.1421811	-0.0122161	-1.960473
rs379525	-0.0064906	0.1763327	-0.0018594	-1.981209
rs42417	-0.0070331	0.1739582	0.0003829	-1.983375
rs4566215	0.0066219	0.1634100	-0.0035546	-1.979817
rs4910656	0.0068438	0.1605890	-0.0006962	-1.982221
rs4957528	-0.0084750	0.1731252	0.0036288	-1.984649
rs523528	0.0080708	0.1629116	0.0029251	-1.985564
rs528301	-0.0086008	0.1773068	0.0124616	-1.994333
rs55921136	0.0085950	0.1559000	-0.0069653	-1.972040
rs568599	-0.0067027	0.1757286	0.0043346	-1.987105
rs5850689	0.0119733	0.1608296	-0.0038879	-1.980291
rs60745548	0.0071946	0.1656670	0.0062353	-1.986552
rs6141314	-0.0080616	0.1818108	0.0010534	-1.984733
rs6265	0.0101598	0.1531146	-0.0043806	-1.976031
rs6433897	-0.0072353	0.1734104	-0.0011588	-1.982527
rs6676022	0.0115926	0.1492373	-0.0153059	-1.956268
rs6690680	0.0088409	0.1547067	-0.0050219	-1.974679
rs6828849	0.0067122	0.1617773	0.0008050	-1.984076
rs71550128	-0.0073950	0.1762278	0.0034139	-1.986200

Table 2.14: Coefficients from GWAS results of logistic regression of the SNPs on smoking status and lung cancer status. Intercepts (β_0 and γ_0) are inferred, while slopes (β_X and γ_1) are as reported. *(continued)*

SNP	β_X	β_0	γ_1	γ_0
rs72505558	0.0067437	0.1614885	-0.0009876	-1.981950
rs72678864	0.0097538	0.1534836	-0.0034394	-1.977455
rs7333559	0.0080523	0.1662222	-0.0183846	-1.975467
rs7451586	-0.0066732	0.1775422	0.0027432	-1.986404
rs748828	0.0086213	0.1572389	-0.0047229	-1.976368
rs7528604	0.0068658	0.1618157	-0.0001820	-1.982931
rs7567570	-0.0091324	0.1727617	-0.0002451	-1.983053
rs763053	0.0080618	0.1570972	-0.0069210	-1.972409
rs76608582	0.0182891	0.1347646	-0.0048192	-1.973958
rs772921	0.0072725	0.1600453	-0.0054837	-1.975937
rs77878475	0.0125950	0.1465726	0.0010985	-1.985146
rs7870475	-0.0071900	0.1771594	0.0082598	-1.991835
rs7948789	-0.0161713	0.1894568	0.0009336	-1.984284
rs883403	0.0094240	0.1536556	-0.0014726	-1.980646
rs9375371	-0.0073963	0.1804155	-0.0069852	-1.972929
rs9381917	0.0112569	0.1493838	-0.0155636	-1.955201
rs9423279	0.0076695	0.1643324	0.0046716	-1.986350
rs9487626	0.0131029	0.1648247	-0.0136868	-1.978168
rs9835772	-0.0078024	0.1814198	-0.0031275	-1.978401

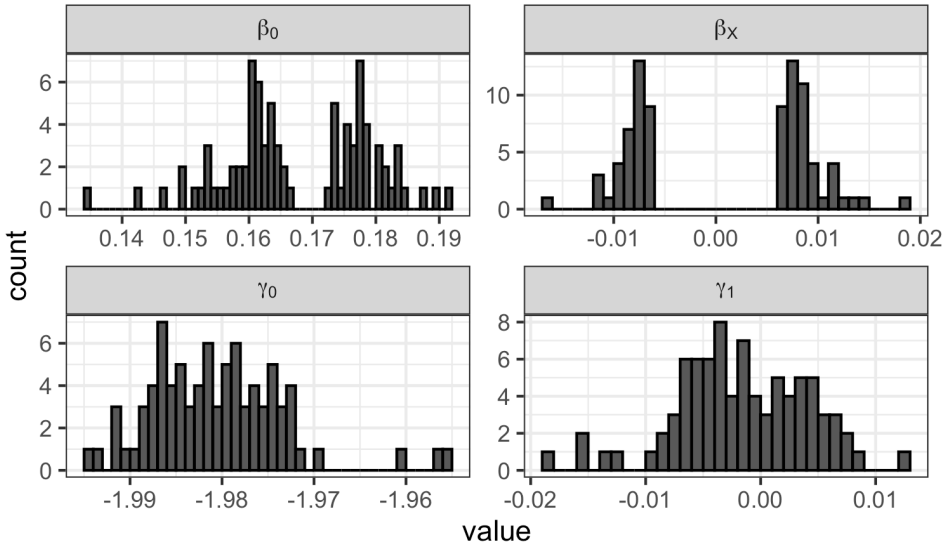


Figure 2.23: Histograms of the coefficients from GWAS results of logistic regression of the SNPs on smoking status and lung cancer status. Intercepts (β_0 and γ_0) are inferred, while slopes (β_X and γ_1) are as reported.

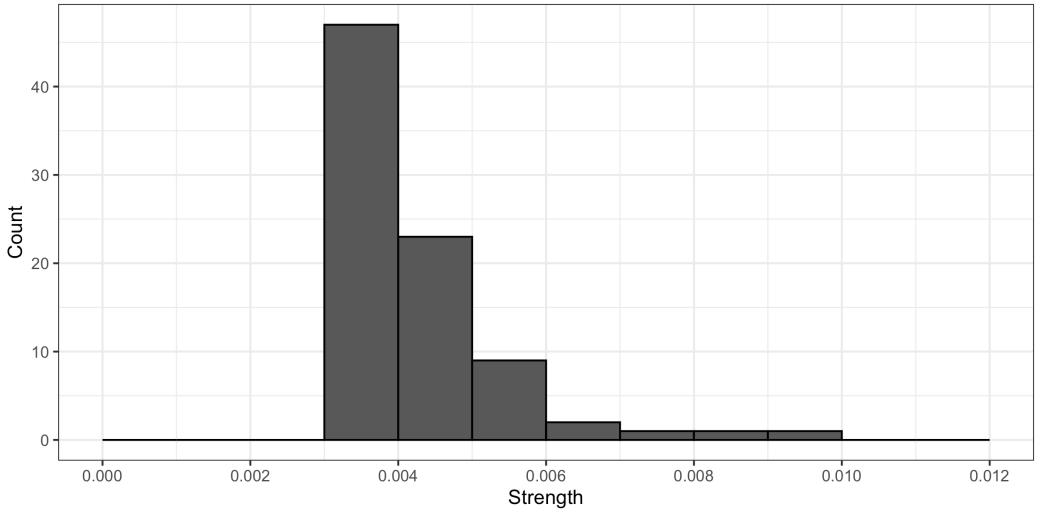


Figure 2.24: Histogram of strengths of IVs on the exposure. Here, SNPs are IVs, and smoking status (ever/never) is exposure. We see that all IVs are very weak, with the largest value just below 0.01.

Table 2.13: Table of the marginal distribution of instruments, $P(Z = z), z = 0, 1, 2$, estimated after preprocessing for analysis.

SNP	P(Z = 2)	P(Z = 1)	P(Z = 0)	SNP	P(Z = 2)	P(Z = 1)	P(Z = 0)
rs10173733	0.3562119	0.4812460	0.1625421	rs2797116	0.5370791	0.3915554	0.0713655
rs10193706	0.2254196	0.4987283	0.2758521	rs2867749	0.4639468	0.4343792	0.1016740
rs10233018	0.2458307	0.4999649	0.2542044	rs299688	0.0806544	0.4066855	0.5126601
rs10274594	0.2540510	0.4999674	0.2459816	rs326341	0.2745833	0.4988473	0.2265693
rs1029986	0.1723980	0.4856208	0.3419813	rs35891966	0.8609698	0.1338295	0.0052006
rs10774625	0.2457332	0.4999633	0.2543035	rs379525	0.2690001	0.4993042	0.2316957
rs10813628	0.2349574	0.4995333	0.2655093	rs42417	0.0959979	0.4276747	0.4763274
rs10897561	0.4140371	0.4588401	0.1271228	rs4566215	0.2184561	0.4978736	0.2836703
rs10905461	0.0654474	0.3807590	0.5537936	rs4910656	0.4334112	0.4498570	0.1167317
rs10914684	0.4570550	0.4380069	0.1049382	rs4957528	0.0432505	0.3294341	0.6273153
rs10956808	0.3337643	0.4879181	0.1783175	rs523528	0.1717181	0.4853414	0.3429405
rs11103667	0.6528207	0.3103050	0.0368743	rs528301	0.2006916	0.4945891	0.3047192
rs11127913	0.3717426	0.4759287	0.1523286	rs55921136	0.6351822	0.3236020	0.0412158
rs11429972	0.1128192	0.4461330	0.4410478	rs568599	0.2090011	0.4963306	0.2946684
rs11611651	0.8323808	0.1599365	0.0076827	rs5850689	0.1341980	0.4642649	0.4015371
rs11631530	0.7779345	0.2081429	0.0139226	rs60745548	0.0747101	0.3972427	0.5280472
rs11646575	0.3149600	0.4925059	0.1925340	rs6141314	0.5735637	0.3675524	0.0588839
rs11693702	0.2849095	0.4977193	0.2173712	rs6265	0.6582586	0.3061456	0.0355959
rs117435980	0.6998026	0.2734789	0.0267185	rs6433897	0.0693372	0.3879647	0.5426982
rs12042107	0.2025948	0.4950210	0.3023842	rs6676022	0.7713790	0.2138057	0.0148153
rs12244388	0.4404143	0.4464457	0.1131399	rs6690680	0.7094689	0.2656618	0.0248694
rs12450028	0.4293549	0.4517938	0.1188513	rs6828849	0.3395694	0.4863129	0.1741177
rs12479064	0.6268375	0.3297864	0.0433761	rs71550128	0.2008017	0.4946147	0.3045837
rs12487411	0.2788384	0.4984262	0.2227354	rs72505558	0.3617072	0.4794276	0.1588652
rs12608052	0.2306302	0.4992191	0.2701507	rs72678864	0.6825787	0.2872090	0.0302123
rs12725407	0.6546886	0.3088794	0.0364320	rs7333559	0.0439935	0.3315056	0.6245008
rs12886628	0.1124522	0.4457734	0.4417744	rs7451586	0.3541182	0.4819202	0.1639616
rs12910916	0.6206505	0.3343265	0.0450230	rs748828	0.5139770	0.4058898	0.0801332
rs13100688	0.3932914	0.4676762	0.1390324	rs7528604	0.3213716	0.4910497	0.1875787
rs1492546	0.2022894	0.4949531	0.3027575	rs7567570	0.0299625	0.2862686	0.6837689
rs1499982	0.0221071	0.2531548	0.7247382	rs763053	0.6013164	0.3482591	0.0504245
rs1549213	0.1285982	0.4600154	0.4113864	rs76608582	0.9070039	0.0907272	0.0022689
rs1561195	0.2279701	0.4989841	0.2730458	rs772921	0.4315416	0.4507533	0.1177051
rs1565735	0.6376078	0.3217914	0.0406009	rs77878475	0.8356836	0.1569474	0.0073690
rs16951001	0.3380123	0.4867519	0.1752358	rs7870475	0.2763346	0.4986816	0.2249839
rs17003752	0.7420669	0.2387323	0.0192008	rs7948789	0.3767706	0.4740916	0.1491378
rs17151637	0.5166809	0.4042486	0.0790705	rs883403	0.7156415	0.2606289	0.0237296
rs1899896	0.4934387	0.4180265	0.0885349	rs9375371	0.5345687	0.3931467	0.0722846
rs2240294	0.3093641	0.4936820	0.1969539	rs9381917	0.8063218	0.1832649	0.0104133
rs2416770	0.2199058	0.4980707	0.2820235	rs9423279	0.1179428	0.4509704	0.4310869
rs264974	0.2640248	0.4996173	0.2363579	rs9487626	0.0332246	0.2981030	0.6686724
rs2675609	0.1387352	0.4674731	0.3937917	rs9835772	0.5737177	0.3674477	0.0588346

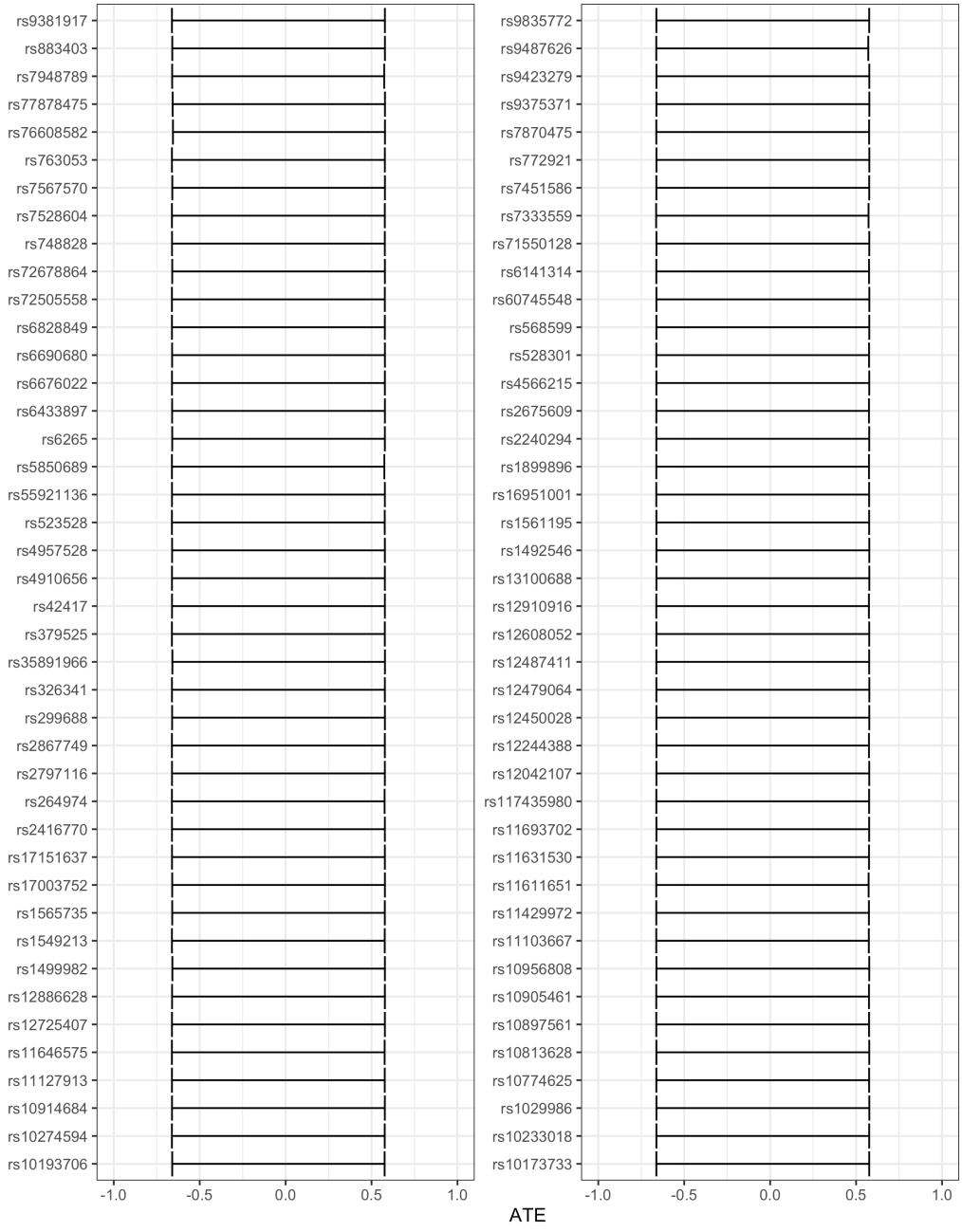


Figure 2.25: Nonparametric two-sample IV bounds on the average treatment effect of smoking on the incidence of lung cancer.

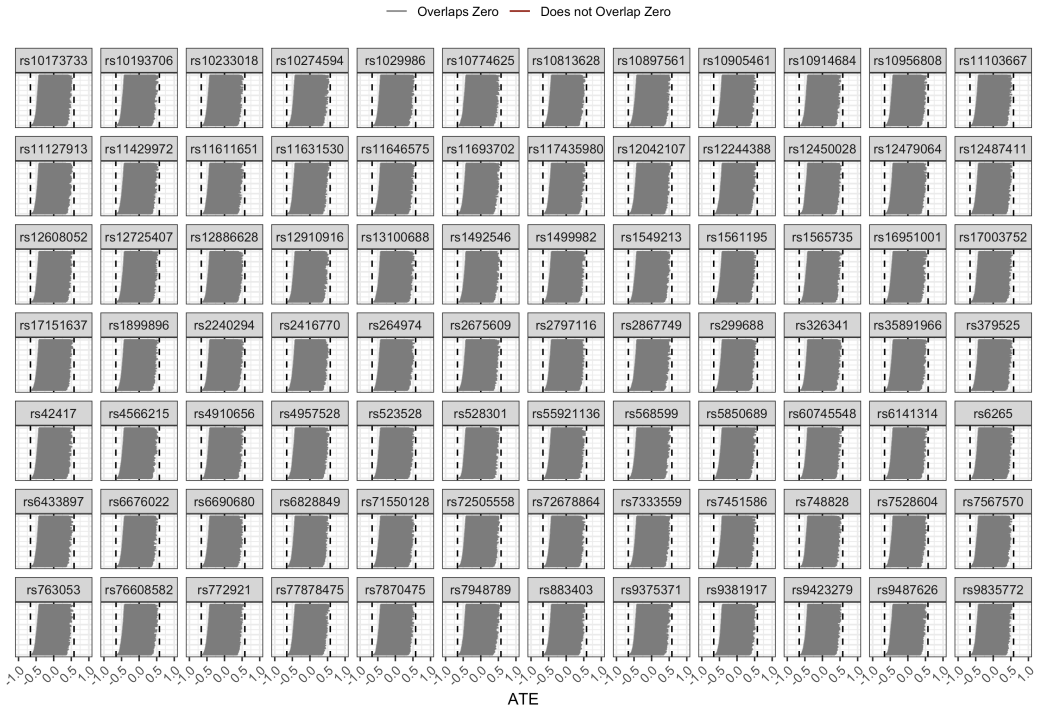


Figure 2.26: 500 sets of bounds of the average treatment effect of smoking on lung cancer for each of the 84 SNPs. Each bound is based on a set of values for the trivariate distribution randomly sampled. Bounds are color coded to show if they overlap 0 (grey) or do not (red). All bounds overlap 0.

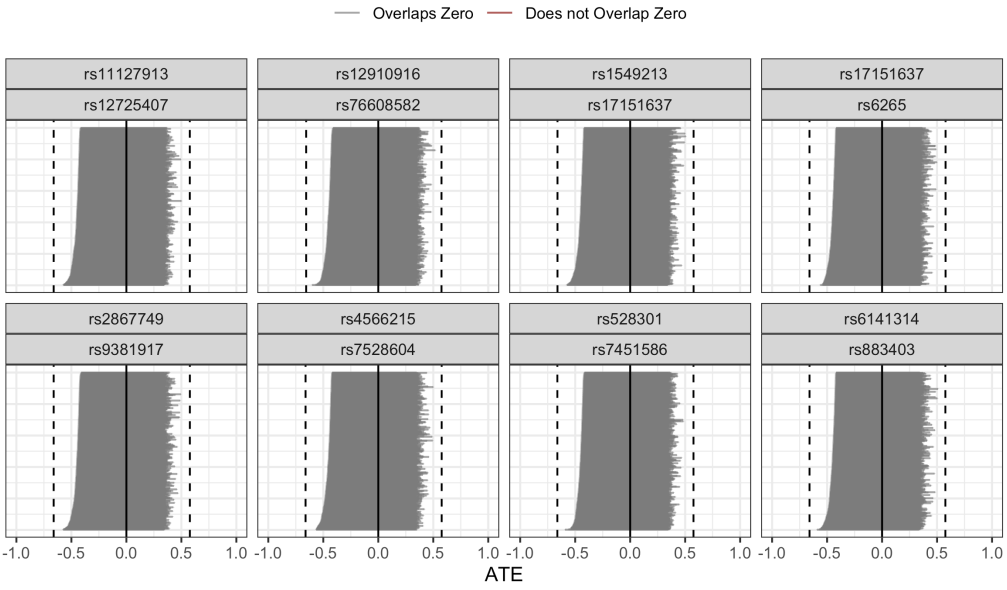


Figure 2.27: Intersection bounds of the average treatment effect of smoking on lung cancer based on randomly sampled trivariate distributions from pairs of SNPs. These 8 pairs were randomly chosen from all possible pairs.

2.J.2 Effect of High Cholesterol on Heart Attack

Table 2.16: Coefficients from GWAS results of logistic regression of the SNPs on high cholesterol and heart attack status. Intercepts (β_0 and γ_0) are inferred, while slopes (β_X and γ_1) are as reported.

SNP	β_X	β_0	γ_1	γ_0
rs10096633	-0.0089830	-3.727152	-0.0012995	-1.966860
rs10260606	0.0076950	-3.755485	0.0007029	-1.970288
rs10410835	0.0071078	-3.749661	0.0007948	-1.969894
rs10504255	-0.0056764	-3.739063	-0.0000742	-1.969088
rs10804330	-0.0050169	-3.737181	-0.0012539	-1.967709
rs112019714	0.0251675	-3.791824	0.0025525	-1.974100
rs11580878	-0.0051399	-3.737725	-0.0006621	-1.968472
rs11591147	-0.0476105	-3.649365	-0.0054389	-1.958449
rs117733303	0.0311528	-3.804047	0.0116909	-1.992088

Table 2.16: Coefficients from GWAS results of logistic regression of the SNPs on high cholesterol and heart attack status. Intercepts (β_0 and γ_0) are inferred, while slopes (β_X and γ_1) are as reported. *(continued)*

SNP	β_X	β_0	γ_1	γ_0
rs12471811	0.0084776	-3.758037	0.0000048	-1.969147
rs1260326	-0.0102312	-3.734879	-0.0003941	-1.968828
rs12740374	-0.0183231	-3.714419	-0.0025251	-1.965207
rs12916	0.0104793	-3.755479	0.0006700	-1.969941
rs1367117	0.0155585	-3.763513	0.0011495	-1.970658
rs1601935	-0.0061378	-3.738671	-0.0007014	-1.968655
rs1883025	-0.0069826	-3.732469	-0.0013153	-1.967173
rs1883711	0.0241076	-3.789616	0.0026734	-1.974319
rs2125345	-0.0056374	-3.734933	-0.0009408	-1.967809
rs2237107	-0.0070166	-3.731732	-0.0007194	-1.967993
rs2244608	0.0070205	-3.752512	0.0010406	-1.970563
rs2618567	-0.0047485	-3.739660	-0.0007455	-1.968630
rs2738447	0.0081671	-3.749563	0.0016947	-1.970520
rs28601761	-0.0140739	-3.726664	-0.0011169	-1.967847
rs28807203	-0.0106943	-3.722554	-0.0002164	-1.968726
rs3127580	0.0076693	-3.755804	0.0022978	-1.973006
rs34042070	0.0094413	-3.758272	0.0002698	-1.969577
rs34707604	0.0058521	-3.751591	0.0002016	-1.969438
rs3918226	0.0081783	-3.757916	0.0028105	-1.974301
rs4299376	-0.0111342	-3.735719	-0.0012431	-1.968335
rs4470903	0.0067035	-3.753387	0.0014579	-1.971420
rs456598	0.0065720	-3.754166	0.0005768	-1.970127
rs4704727	0.0074887	-3.747988	0.0007432	-1.969643
rs472495	0.0064154	-3.747379	0.0004743	-1.969469
rs56299331	0.0057258	-3.752033	0.0001068	-1.969308
rs57180587	0.0081592	-3.756830	0.0013685	-1.971475
rs58542926	-0.0146353	-3.715853	-0.0013536	-1.966636
rs58691354	0.0074756	-3.755521	0.0000196	-1.969171
rs59950280	0.0058286	-3.750690	0.0004805	-1.969780
rs6090040	-0.0055812	-3.737545	-0.0007168	-1.968450
rs622871	0.0065093	-3.746991	0.0013161	-1.969966
rs635634	0.0098788	-3.758987	0.0014151	-1.971442
rs6458349	0.0056558	-3.746031	0.0007529	-1.969556
rs6511720	-0.0261322	-3.696906	-0.0030216	-1.963813
rs7012637	0.0047984	-3.747932	0.0002456	-1.969396
rs7213086	0.0047773	-3.747169	0.0007846	-1.969840
rs73534263	0.0071810	-3.755717	0.0000767	-1.969275
rs7412	-0.0374088	-3.674234	-0.0038000	-1.962153
rs74617384	0.0190473	-3.777927	0.0069894	-1.981990

Table 2.16: Coefficients from GWAS results of logistic regression of the SNPs on high cholesterol and heart attack status. Intercepts (β_0 and γ_0) are inferred, while slopes (β_X and γ_1) are as reported. *(continued)*

SNP	β_X	β_0	γ_1	γ_0
rs7534572	0.0081187	-3.748658	0.0005830	-1.969551
rs7707394	0.0061511	-3.750841	0.0000817	-1.969243
rs77542162	0.0253674	-3.792474	0.0020548	-1.973154
rs799157	-0.0108031	-3.741956	-0.0003979	-1.969103
rs9376091	-0.0053004	-3.735070	-0.0005561	-1.968317
rs964184	-0.0215630	-3.737246	-0.0013629	-1.968778

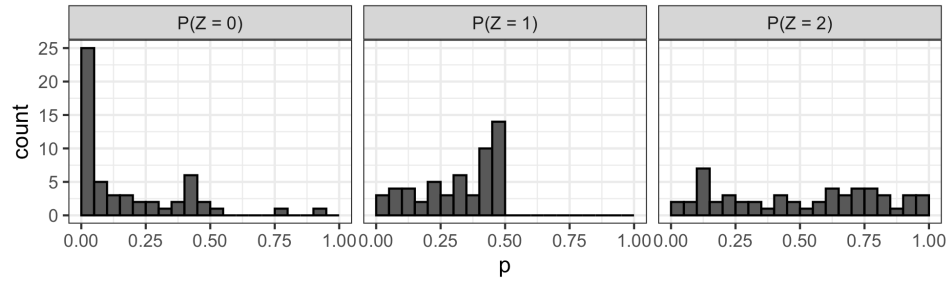


Figure 2.28: Histograms of the marginal distribution of instruments, $P(Z = z), z = 0, 1, 2$, estimated after preprocessing.

Table 2.15: Table of the marginal distribution of instruments, $P(Z = z), z = 0, 1, 2$, estimated after preprocessing for analysis.

SNP	P(Z = 2)	P(Z = 1)	P(Z = 0)	SNP	P(Z = 2)	P(Z = 1)	P(Z = 0)
rs10096633	0.7682873	0.2164654	0.0152473	rs3918226	0.8434773	0.1498658	0.0066569
rs10260606	0.6689457	0.2978906	0.0331637	rs4299376	0.1044835	0.4375111	0.4580055
rs10410835	0.2261041	0.4987999	0.2750961	rs4470903	0.6122421	0.3404338	0.0473241
rs10504255	0.1141345	0.4474070	0.4384585	rs456598	0.7353800	0.2443260	0.0202940
rs10804330	0.3246447	0.4902626	0.1850927	rs4704727	0.1153479	0.4485623	0.4360899
rs112019714	0.9445278	0.0546808	0.0007914	rs472495	0.1219232	0.4545036	0.4235732
rs11580878	0.2532012	0.4999796	0.2468192	rs56299331	0.6368870	0.3223300	0.0407830
rs11591147	0.9653935	0.0343018	0.0003047	rs57180587	0.7289642	0.2496596	0.0213762
rs117733303	0.9629825	0.0366685	0.0003491	rs58542926	0.8541959	0.1400626	0.0057415
rs12471811	0.7974669	0.1910863	0.0114469	rs58691354	0.7129641	0.2628159	0.0242201
rs1260326	0.1542518	0.4769944	0.3687538	rs59950280	0.4469685	0.4431771	0.1098545
rs12740374	0.6060342	0.3448956	0.0490702	rs6090040	0.2300488	0.4991705	0.2707808
rs12916	0.3593703	0.4802094	0.1604203	rs622871	0.0988228	0.4310763	0.4701008
rs1367117	0.4370916	0.4480749	0.1148336	rs635634	0.6627002	0.3027276	0.0345722
rs1601935	0.1186871	0.4516457	0.4296671	rs6458349	0.0768498	0.4007364	0.5224138
rs1883025	0.5579089	0.3780482	0.0640429	rs6511720	0.7764852	0.2093975	0.0141172
rs1883711	0.9385769	0.0604497	0.0009733	rs7012637	0.2755284	0.4987592	0.2257124
rs2125345	0.4990744	0.4147551	0.0861704	rs7213086	0.2001050	0.4944520	0.3054430
rs2237107	0.6333104	0.3249953	0.0416944	rs73534263	0.7971401	0.1913739	0.0114861
rs2244608	0.4686429	0.4318641	0.0994929	rs7412	0.8445834	0.1488576	0.0065590
rs2618567	0.1161249	0.4492923	0.4345829	rs74617384	0.8447171	0.1487357	0.0065473
rs2738447	0.1661712	0.4829396	0.3508892	rs7534572	0.1255675	0.4575751	0.4168575
rs28601761	0.3342690	0.4877820	0.1779490	rs7707394	0.4169078	0.4575523	0.1255398
rs28807203	0.9046336	0.0929773	0.0023890	rs77542162	0.9546715	0.0448029	0.0005257
rs3127580	0.7081492	0.2667336	0.0251172	rs799157	0.0018869	0.0831041	0.9150089
rs34042070	0.6625016	0.3028808	0.0346176	rs9376091	0.5451282	0.3863995	0.0684722
rs34707604	0.5518930	0.3820040	0.0661030	rs964184	0.0174433	0.2292594	0.7532973

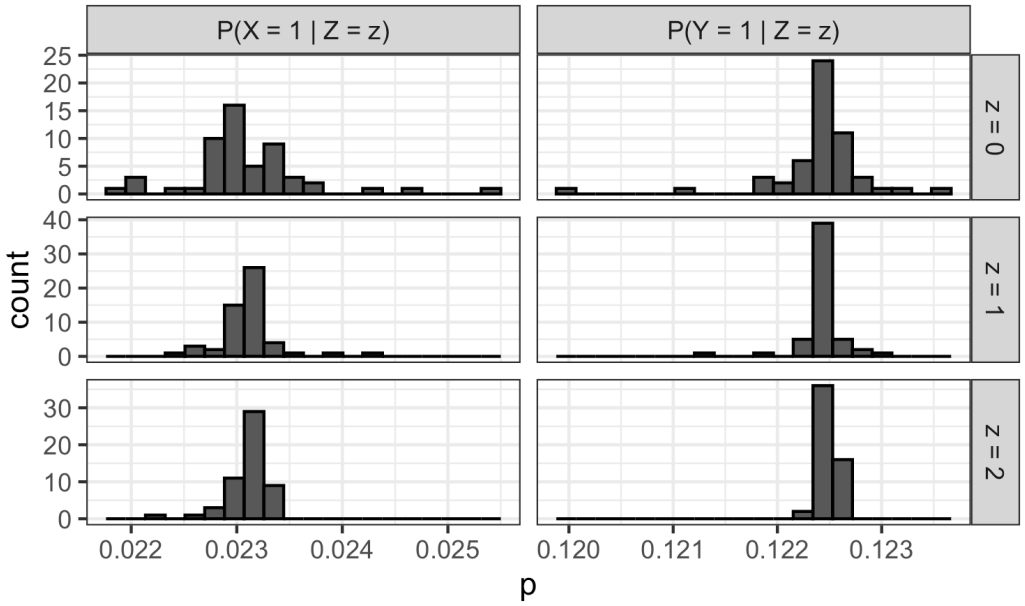


Figure 2.29: Histograms of the marginal conditional probabilities $P(X = 1|Z = z), z = 0, 1, 2$ and $P(Y = 1|Z = z), z = 0, 1, 2$.

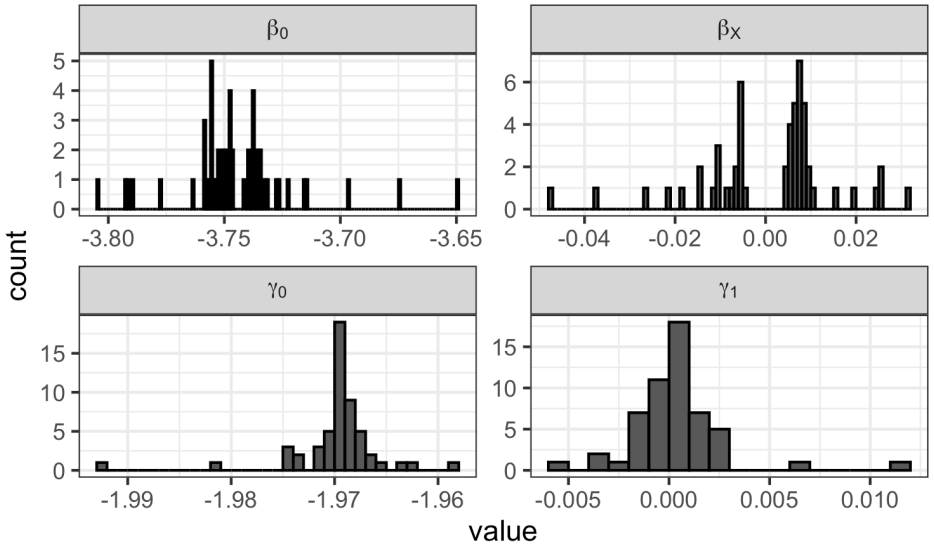


Figure 2.30: Histograms of the coefficients from GWAS results of logistic regression of the SNPs on high cholesterol and heart attack, respectively. Intercepts (β_0 and γ_0) are inferred, while slopes (β_X and γ_1) are as reported.

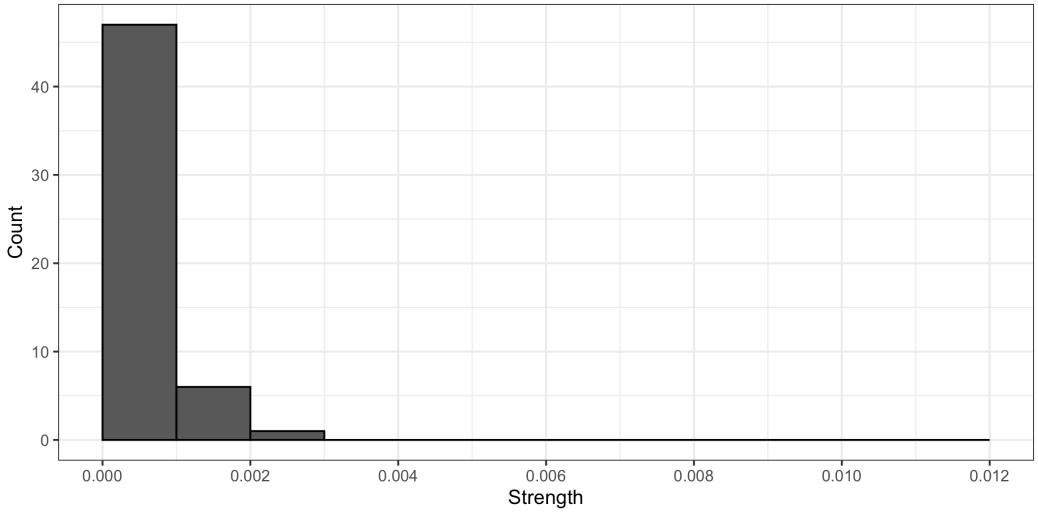


Figure 2.31: Histogram of strengths of IVs on the exposure. Here, SNPs are IVs, and high cholesterol is the exposure. We see that all IVs are very weak, with the largest value below 0.003.

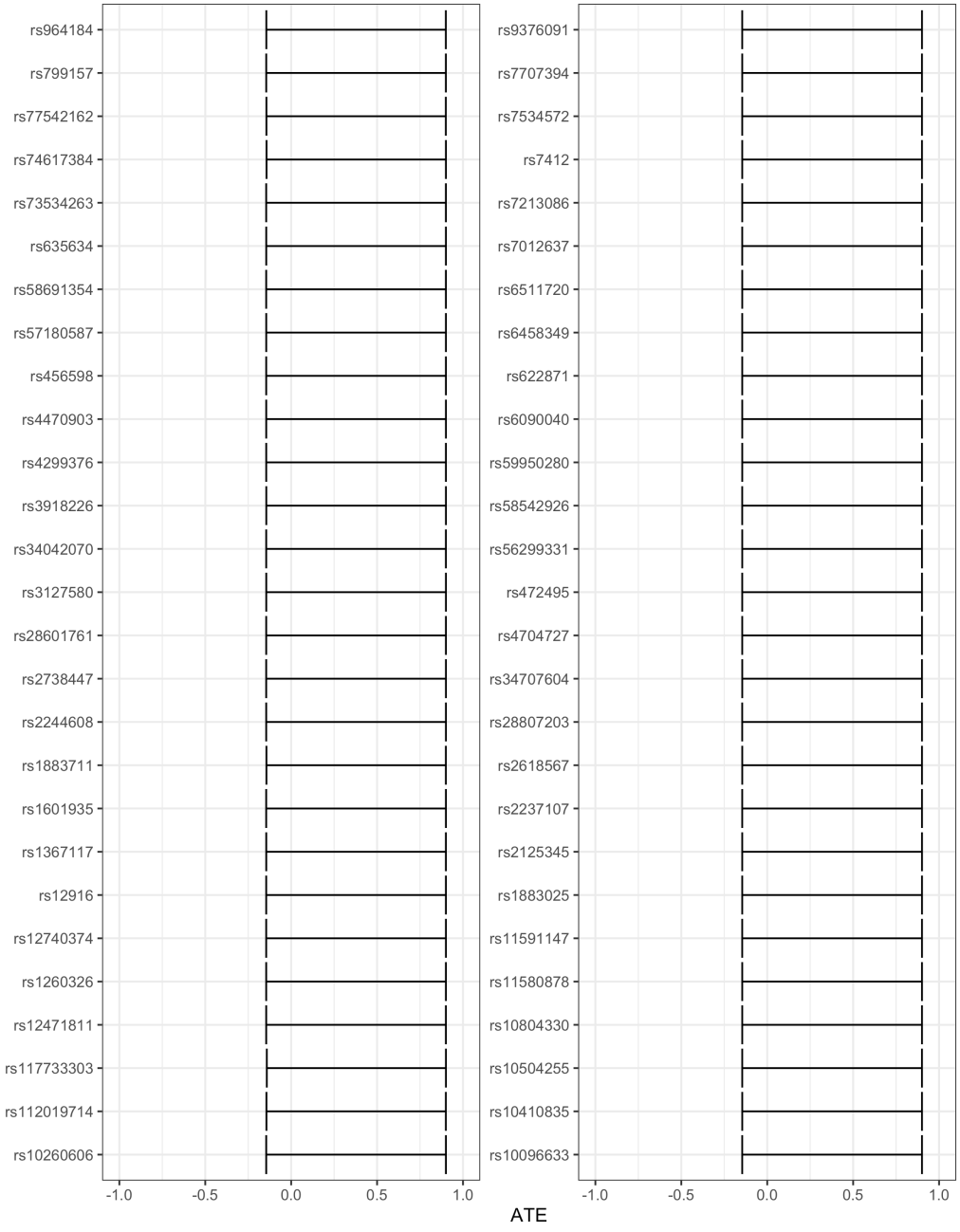


Figure 2.32: Nonparametric two-sample IV bounds on the average treatment effect of high cholesterol on the incidence of heart attack.

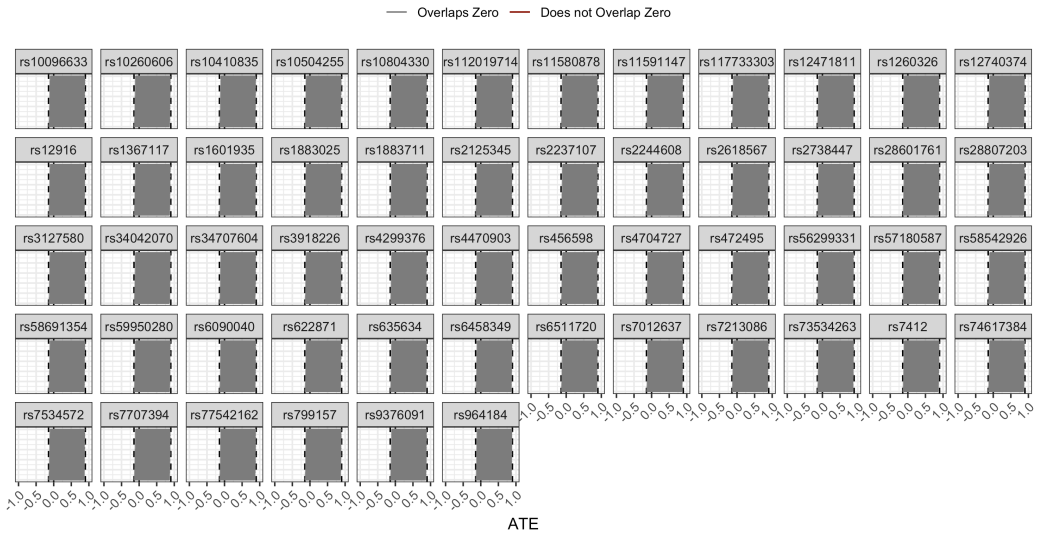


Figure 2.33: 500 sets of bounds of the average treatment effect of high cholesterol on heart attack for each of the 54 SNPs. Each bound is based on a set of values for the trivariate distribution randomly sampled. Bounds are color coded to show if they overlap 0 (grey) or do not (red). All bounds overlap 0.

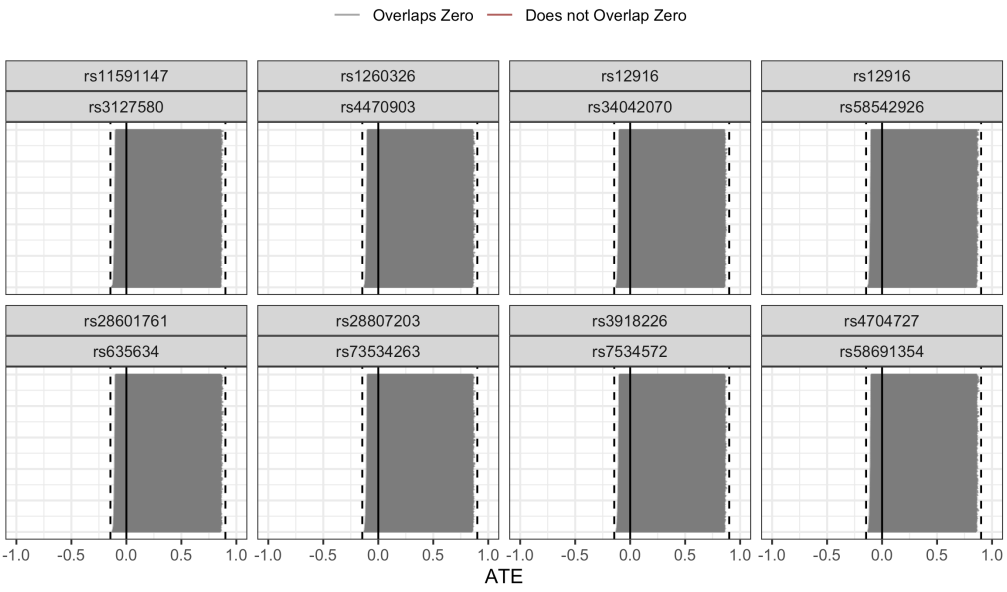


Figure 2.34: Intersection bounds of the average treatment effect of high cholesterol on heart attack based on randomly sampled trivariate distributions from pairs of SNPs. These 8 pairs were randomly chosen from all possible pairs.

Chapter 3

Covariate Adjustments in Randomized Controlled Trials under Network Interference

3.1 Introduction

Covariate adjustment methods are popular in randomized experiments, such as randomized clinical trials (RCTs), when trying to improve precision of estimators of the Average Treatment Effect (ATE). There is a vast literature dedicated to whether one should perform such adjustments, and if so, how this should be done. Tsiatis et al. (2008) provide a flexible and principled approach that address most arguments against covariate adjustment. However, most of previous research has focused on the traditional setting where trial subjects do not interfere with each other. While this assumption has long been known to be unrealistic in many settings (Cox, 1958), there is still much to learn about the behavior of even simple estimators of the ATE when this assumption is relaxed.

In the case where interference between experimental units is negligible, semiparametric theory helps derive a locally efficient estimator (Robins and Rotnitzky, 1995). This leads to the well-known Augmented Inverse Probability Weighted (AIPW) estimator, which besides being locally efficient also possess the doubly-robust characteristic (Kurz, 2022) meaning even

under misspecification of either the propensity score or the outcome model, the estimator is still well-behaved. This is a very flexible estimator that can be used with a plethora of models for estimating the potential outcomes, both modern machine learning and classical statistical methods.

In practice, the classical Analysis of Covariance (ANCOVA) is a popular choice to estimate the ATE while adjusting for baseline covariates. While it is simple in nature, Tsiatis et al. (2008) showed how the resulting estimator for the ATE is asymptotically equivalent to an AIPW estimator, and therefore suggest the ANCOVA estimator is a suitable choice for covariate adjustment in randomized experiments. Among the benefits of ANCOVA is its computational efficiency as it can be estimated using Ordinary Least Squares (OLS) estimation. Also, since ANCOVA methods are taught to a wide audience from a wide array of disciplines, it is familiar to most researchers, and as such often used.

While both the AIPW estimator and the OLS estimator have been studied extensively in the classical setting where interference is negligible, their properties under general interference are still relatively unknown. We aim to characterize the asymptotic behavior of both of these estimators under a somewhat general interference scheme, when the interference network can be explained as a graphon model and the spillover effect is determined solely through the proportion of neighbors who receive treatment. These two assumptions are not novel: Li and Wager (2022) applied the random graph asymptotic paradigm to the estimation of the ATE under general interference, and Hudgens and Halloran (2008) considered what they refer to as stratified interference. Later, Sävje et al. (2021) generalized this concept to the idea of exposure mappings as a way to allow for a more general interference network but still have an interference mechanism that allows for asymptotic analyses of estimators.

Our main contributions to this area are two-fold: first, we show that the AIPW estimator is asymptotically normal and provide an exact expression for the asymptotic variance. Though this does not lead to a variance estimator, it does provide insights into the asymptotic behavior of the AIPW estimator. Second, we extend the result of Tsiatis et al. (2008) to show that the ANCOVA estimator is still asymptotically equivalent to an AIPW estimator

under the general interference setup considered here, which implies it is also asymptotically normal. The practical consequences are substantial: any analysis performed under the assumption of no interference using an AIPW estimator, including the ANCOVA estimator, lead to unbiased and consistent results, even if the no interference assumption is invalid. Also, going forward, the well-known and widely used ANCOVA estimator can still play a role when analyzing data from randomized experiments on networks. However, there is a caveat: the OLS variance estimator is not reliable, and as such any inference should be considered with skepticism.

The rest of this chapter is organized as follows. Section 3.2 sets up the randomized experiment scenario we consider, introduces the estimand of interest, and necessary assumptions related to the potential outcomes and the random graph model. Section 3.3 introduces the Augmented Inverse Probability Weighting estimators, conditions needed to establish consistency and asymptotic normality, and shows that the ANCOVA estimator is asymptotically equivalent to an AIPW style estimator and therefore asymptotically normal. Section 3.4 goes beyond the theoretical results, and tries to illustrate some of the challenges that are still unsolved, such as standard error estimation, testing of a sharp null hypothesis, how to improve precision of the AIPW estimator, and the potential costs and benefits of the use of sample splitting as suggested by Emmenegger et al. (2023). Finally, Section 3.5 presents results of analyzing the Cai et al. (2015b) study using the AIPW estimator.

3.2 Setup, Estimands, and Graphon Asymptotics

We consider collecting data on $n \in \mathbb{N}$ units indexed by $i = 1, ..., n$. Each unit is randomized according to a binary treatment $W_i \in \{0, 1\}$ such that $W_i \sim \text{Bernoulli}(\pi)$ for a $\pi \in (0, 1)$ chosen by the researcher. For each unit, we observe pretreatment covariates $\boldsymbol{X}_i \in \mathbb{R}^p$ along with an outcome $Y_i \in \mathbb{R}$. For the covariates, we assume $\boldsymbol{X}_i$ is i.i.d. and each of the p covariates have finite variance. We assume the existence of potential outcomes $Y_i(\boldsymbol{w})$ for $\boldsymbol{w} \in \{0, 1\}^n$, in line with the Neyman-Rubin causal model (Neyman, 1923; Rubin, 1974), such that the

observed outcome is exactly the potential outcome of the observed treatment allocation vector $\boldsymbol{W} = \begin{bmatrix} W_1 & ... & W_n \end{bmatrix}'$, i.e. $Y_i = Y_i(\boldsymbol{W})$. We also introduce the notation $Y_i(w_i, \boldsymbol{w}_{-i})$ for the potential outcome of unit i when unit i receives treatment w_i and the remaining units receive treatments according to the vector $\boldsymbol{w}_{-i} = (w_1, ..., w_{i-1}, w_{i+1}, ..., w_n) \in \{0,1\}^{n-1}$.

Following Cox (1958), we say two units i and j interfere with each other if the potential outcome of unit i is affected by the treatment of unit j. That is, $Y_i(\boldsymbol{w}) \neq Y_i(\boldsymbol{w}')$ for $\boldsymbol{w}, \boldsymbol{w}' \in \{0,1\}^n$ with $w_k = w'_k$ for all $k \neq j$ and $w_j \neq w'_j$. Also, following Li and Wager (2022), we use an interference graph $\mathcal{G}$ with n nodes and adjacency matrix $E \in \{0,1\}^{n \times n}$ to denote the interference between all n units in the study. Specifically, each node in $\mathcal{G}$ represents an observational unit and for a pair of units i and j, $E_{ij} = 1$ if units i and j interfere, and $E_{ij} = 0$ otherwise. Two units i and j are neighbors if $E_{ij} = 1$. We adopt the convention $E_{ii} = 0$.

We focus our efforts on the estimation of the direct effect,

$$\bar{\tau}_{\text{DIR}} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[Y_i(1, \boldsymbol{W}_{-i}) - Y_i(0, \boldsymbol{W}_{-i}) \middle| Y_i(\cdot) \right]. \tag{3.1}$$

In words, $\bar{\tau}_{\text{DIR}}$ is the average, over the n experimental units, of the expected difference between being treated and not being treated. The inner expectation is taken over the random variable $\boldsymbol{W}_{-i}$. This estimand has become the de facto extension of the Average Treatment Effect when interference is present in recent years (Hudgens and Halloran, 2008; Tchetgen and VanderWeele, 2012; Sävje et al., 2021; Hu et al., 2022). This is often referred to as the Global Average Treatment Effect (GATE) or the Expected Average Treatment Effect (EATE). When no interference is present, i.e. when $Y_i(w_i, \boldsymbol{w}_{-i}) = Y_i(w_i)$ for all $\boldsymbol{w}_{-i} \in \{0,1\}^{n-1}$, $\mathbb{E}\left[Y_i(1, \boldsymbol{W}_{-i}) - Y_i(0, \boldsymbol{W}_{-i}) | Y_i(\cdot) \right] = Y_i(1) - Y_i(0)$, and $\bar{\tau}_{\text{DIR}} = \frac{1}{n} \sum_{i=1}^{n} Y_i(1) - Y_i(0)$ is the classic Average Treatment Effect. If the potential outcomes $Y_i(\cdot)$ are sampled according to certain sampling models, one can consider the limiting population estimand $\tau_{\text{DIR}} = \lim_{n \to \infty} \mathbb{E}\left[\bar{\tau}_{\text{DIR}} \right]$.

Following the work of Li and Wager (2022), we introduce a few assumptions on the graph

and the potential outcomes.

Assumption 3.1 (Undirected Relationships)**.** The interference graph $\mathcal{G}$ is undirected, i.e. $E_{ij} = E_{ji}$ for all $i \neq j$.

Assumption 3.2 (Random Graph)**.** $\mathcal{G}$ follows a *random graphon model*: each unit is associated with a latent variable $U_i \overset{\text{iid}}{\sim} \text{Uniform}[0,1]$, and the probability that units i and j are connected in the graph $\mathcal{G}$ is $G_n(U_i, U_j)$ for a symmetric measureable function $G_n : [0,1]^2 \mapsto [0,1]$. I.e. $E_{ij} \sim \text{Bernoulli}(G_n(U_i, U_j))$ independently for all $i < j$. The function G_n is referred to as the *graphon*.

Random graphon models are inspired by work on partially exchangeable arrays (Aldous, 1981; Lovász and Szegedy, 2006). Note that the graphon in the assumption above could depend on n. The next assumption describes this dependency, which allows us to consider asymptotic behaviors.

Assumption 3.3 (Graphon Sequence)**.** The graphon G_n from Assumption 3.2 is assumed to be of the form

$$G_n(U_i, U_j) = \min\left\{1, \rho_n G(U_i, U_j)\right\}$$

for some symmetric function $G : [0,1]^2 \mapsto [0, \infty)$, and a sequence $0 < \rho_n \leq 1$. The sequence ρ_n is either such that $\rho_n \to 0$ with $n\rho_n \to \infty$ as $n \to \infty$, in which case G_n is a *sparse* graphon, or $\rho_n = 1$, in which case G_n is a *dense* graphon.

Assuming that the interference graph is a realization of a stochastic process is different than the assumption that the interference graph is non-random often made in the early work on estimation of causal effects under interference (Hudgens and Halloran, 2008; Liu, Hudgens, and Becker-Dreps, 2016). Asymptotic behavior in a fixed graph setting can be derived. Sävje et al. (2021) shows that the Horvitz-Thompson and Hájek estimators are consistent for $\bar{\tau}_{\text{DIR}}$, but with a rate of convergence that depends on the denseness of the graph, which is assumed to be relatively sparse, and without any asymptotic normality results. As is evident by Assumption 3.3, the random graph asymptotic framework considered by Li and

Wager (2022) and in this work, allows for both sparse and dense graphs. Similar to Li and Wager (2022), we view this random interference graph as a useful, technical device to gain insights about the statistical properties of estimators under general interference, most notably $\sqrt{n}$ consistency and asymptotic normality.

The following additional assumptions on the graphon are requried for the central limit theorem derived by Li and Wager (2022), and the central limit theorem we will derive later.

Assumption 3.4 (Additional Graphon Assumptions)**.**

i. There exists some $c_l > 0$ such that $c_l \leq g_1(u) = \int_0^1 \min\{1, G(u,t)\}dt$ for any $u \in [0,1]$.
ii. The graphon has finite second moment, i.e. $\mathbb{E}\left[G(U_1, U_2)^k\right] \leq c_u^k, k = 1, 2$.
iii. The sparsity controlling sequence ρ_n satisfies $\liminf \log \rho_n / \log n > -1$.

In addition to the stochastic behavior of the graph, we also introduce the following assumptions on the behavior of the observed and potential outcomes.

Assumption 3.5 (Causal Identifying Assumptions)**.** We assume the following about the observed outcomes $\boldsymbol{Y}$, potential outcomes $\boldsymbol{Y}(\boldsymbol{w})$ and treatment assignments $\boldsymbol{W}$:

a. Consistency: $Y_i = Y_i(\boldsymbol{W})$
b. Complete Ignorability: $\boldsymbol{W} \perp \boldsymbol{Y}(\boldsymbol{w})$
c. Positivity: $0 < \pi < 1$

Assumption 3.6 (Anonymous Interference)**.** The potential outcomes of unit i only depends on the treatments of its neighbors through the quantity $\frac{\sum_{j\neq i} E_{ij} W_j}{\sum_{j\neq i} E_{ij}}$, that is, only the proportion of peers who are treated is relevant to the potential outcomes of unit i. More formally,

$$Y_i(w_i, \boldsymbol{w}_{-i}) = Y_i(w_i, \boldsymbol{w}'_{-i})$$

for all $\boldsymbol{w}_{-i}$ and $\boldsymbol{w}'_{-i}$ with

$$\frac{\sum_{j\neq i} E_{ij} w_j}{\sum_{j\neq i} E_{ij}} = \frac{\sum_{j\neq i} E_{ij} w'_j}{\sum_{j\neq i} E_{ij}}.$$

In the case $\sum_{j\neq i} E_{ij} = 0$, we define $\frac{\sum_{j\neq i} E_{ij} W_j}{\sum_{j\neq i} E_{ij}} = 0$.

Assumption 3.7 (Class of Potential Outcomes Functions)**.** The potential outcomes follow a potential outcome function $f : \{0,1\} \times [0,1] \times \mathbb{R}^p \times [0,1] \mapsto \mathbb{R} \times \mathbb{R}$ such that

$$Y_i(w_i, \boldsymbol{w}_{-i}) = f\left(w_i, \frac{\sum_{j \neq i} E_{ij} w_j}{\sum_{j \neq i} E_{ij}}, \boldsymbol{X}_i, U_i; \epsilon_i\right),$$

where ϵ_i is iid that allows for random error. Furthermore,

$$|f(w, p, \boldsymbol{x}, u; \epsilon)| \leq B,$$
$$\left|\frac{d^m}{dp^m} f(w, p, \boldsymbol{x}, u; \epsilon)\right| \leq B \text{ for } m = 1, 2, 3.$$

uniformly for $(w, p, \boldsymbol{x}, u; \epsilon) \in \{0,1\} \times [0,1] \times \mathbb{R}^p \times [0,1] \times \mathbb{R}$.

Assumption 3.5 provides extensions of often-used assumptions in traditional causal inference literature on the relationship between the observed outcomes, the potential outcomes, and the treatment assignment mechanism. Assumption 3.6 characterizes the structure of the interference between units. Specifically, it only allows for interference through the number of neighbors who are treated with no effect of *which* neighbors are treated. Under partial interference, Hudgens and Halloran (2008) called this "stratified interference". This can also be viewed as a special case of an exposure mapping as defined by Aronow and Samii (2017). Assumption 3.7 states that the potential outcomes function is three times differentiable, and the first three derivatives along with the function itself are bounded. This is, in essence, a smoothness assumption which assures that the potential outcomes behave nicely with small variations of the proportion of peers who have received treatment.

We want to emphasize a small deviation from Li and Wager (2022). In Assumption 3.7, we do not allow the functional form of the potential outcomes to be different between individuals. This simplification is made solely to ease notation, intuition and remove a layer of complexity – all results still hold if one sticks to the Li and Wager (2022) assumption that potential outcomes functions are drawn from a class of functions. Nonetheless, we find it conceptually easier to consider a setting where the potential outcomes of all experimental units follow the same function of individual level characteristics.

3.3 Estimating the direct effect

The classic Horvitz-Thompson estimator is often used for estimation of the average treatment effect under no interference

$$\hat{\tau}^{\mathrm{HT}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{W_i Y_i}{\pi} - \frac{(1-W_i)Y_i}{1-\pi}\right). \tag{3.2}$$

Under no interference, this simple estimator is asymptotically Normal, but is less efficient than the Hájek estimator, also known as the difference in means estimator (Rosenbaum, 1987),

$$\begin{aligned}\hat{\tau}^{\mathrm{HA}} &= \sum_{i=1}^{n}\left(\frac{W_i Y_i}{\sum_{i=1}^{n} W_i} - \frac{(1-W_i)Y_i}{\sum_{i=1}^{n}(1-W_i)}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{W_i Y_i}{\hat{\pi}} - \frac{(1-W_i)Y_i}{1-\hat{\pi}}\right)\end{aligned} \tag{3.3}$$

where $\hat{\pi} = \frac{1}{n}\sum_{i=1}^{n} W_i$ would be a natural estimator of π if this were unknown.

It is well known that $\hat{\tau}^{\mathrm{HT}}$ is unbiased for our estimand in (3.1) if we condition on the graph $\mathcal{G}$ and the potential outcomes, i.e. consider $\mathcal{G}$ and $\boldsymbol{Y}(\cdot)$ fixed (see, for example, Section 2 of Li and Wager (2022)). Under Assumptions 3.1, 3.5, 3.6, and 3.7, both $\hat{\tau}^{\mathrm{HT}}$ and $\hat{\tau}^{\mathrm{HA}}$ are consistent, and if we further include Assumptions 3.2, 3.3, and 3.4, $\hat{\tau}^{\mathrm{HT}}$ and $\hat{\tau}^{\mathrm{HA}}$ are asymptotically normal *without conditioning* on the graph and potential outcomes (Li and Wager, 2022). However, when interference is present, it is no longer true that $\hat{\tau}^{\mathrm{HA}}$ is universally at least as efficient as $\hat{\tau}^{\mathrm{HT}}$.

As mentioned by Li and Wager (2022), interference effects are expected to inflate the asymptotic variances of both the Horvitz-Thompson and the Hájek estimators in most settings. This makes it particularly interesting to consider covariate adjustments in the estimation of the direct effect as a way to reduce the asymptotic variance, and ensure higher efficiency. We consider two classes of augmented inverse probability weighting (AIPW) estimators.

The first class arise from augmenting the Horvitz-Thompson estimator,

$$\begin{aligned}\hat{\tau}^{\mathrm{HT}}_{\mathrm{AIPW}} = \\ \frac{1}{n}\sum_{i=1}^{n}\frac{W_i\left(Y_i - \hat{Y}_n^{(1,\boldsymbol{w}_{-i})}\right)}{\pi} + \hat{Y}_n^{(1,\boldsymbol{w}_{-i})} \\ - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-W_i)\left(Y_i - \hat{Y}_n^{(0,\boldsymbol{w}_{-i})}\right)}{1-\pi} + \hat{Y}_n^{(0,\boldsymbol{w}_{-i})},\end{aligned} \tag{3.4}$$

and the second class from augmenting the Hájek estimator,

$$\begin{aligned}\hat{\tau}^{\mathrm{HA}}_{\mathrm{AIPW}} = \\ \frac{1}{n}\sum_{i=1}^{n}\frac{W_i\left(Y_i - \hat{Y}_{i,n}^{(1,\boldsymbol{w}_{-i})}\right)}{\hat{\pi}} + \hat{Y}_{i,n}^{(1,\boldsymbol{w}_{-i})} \\ - \frac{1}{n}\sum_{i=1}^{n}\frac{(1-W_i)\left(Y_i - \hat{Y}_{i,n}^{(0,\boldsymbol{w}_{-i})}\right)}{1-\hat{\pi}} + \hat{Y}_{i,n}^{(0,\boldsymbol{w}_{-i})},\end{aligned} \tag{3.5}$$

In general, $\hat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})}$ are simply random variables that behave nicely, but typically these will be estimates of the potential outcomes $Y_i(w, \boldsymbol{w}_{-i})$.

The estimators in Equation 3.4 and Equation 3.5 are the same as have been used in the classical setting (Robins, Rotnitzky, and L. P. Zhao, 1995; Rotnitzky et al., 2012). It has long been known that in the classical setting of no interference, $\hat{\tau}^{\mathrm{HA}}_{\mathrm{AIPW}}$ is locally semiparametric efficient if $\hat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})}$ is estimated using a flexible model (Robins and Rotnitzky, 1995; Rotnitzky et al., 2012), which makes it a natural choice to consider when trying to reduce variance.

Both $\hat{\tau}^{\mathrm{HT}}_{\mathrm{AIPW}}$ and $\hat{\tau}^{\mathrm{HA}}_{\mathrm{AIPW}}$ are consistent estimators of $\bar{\tau}_{\mathrm{DIR}}$, conditionally on the graph and the potential outcomes, if the random variables $\hat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})}$ are asymptotically well-behaved.

Proposition 3.1 (Consistency)**.** *Suppose Assumptions 3.1, 3.6 and 3.7 hold. If the random*

variables $\widehat{Y}_{i,n}^{(w,\boldsymbol{W}_{-i})}$ *are such that*

$$\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\widehat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})} \to_p 0$$
$$\frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\pi}{1-\pi}\widehat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})} \to_p 0$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{Y}_{i,n}^{(w,\boldsymbol{W}_{-i})} = O_p(1)$$

for $w \in \{0,1\}$*, then the estimators* $\hat{\tau}_{AIPW}^{HT}$ *and* $\hat{\tau}_{AIPW}^{HA}$ *are consistent for* $\bar{\tau}_{DIR}$*, i.e.* $\hat{\tau}_{AIPW}^{HT} - \bar{\tau}_{DIR} \to_p 0$ *and* $\hat{\tau}_{AIPW}^{HA} - \bar{\tau}_{DIR} \to_p 0$*, conditional on the graph and the potential outcomes.*

See proof in Appendix 3.C.

In practice, $\widehat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})}$ is often a model fitted to the data. But notice that no assumptions are made regarding the correctness of the models. Similar to the no interference case, the AIPW estimator is robust towards model misspecification, i.e. the random variables $\widehat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})}$ could be completely unrelated to the potential outcomes $Y_i(\boldsymbol{w})$, and we would still end up with a consistent estimator, as long as the assumptions of Proposition 3.1 hold. This can be assured by choosing specific models for $\widehat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})}$. For example, if one chooses to use an estimate of $\mathbb{E}\left[Y(1,\boldsymbol{W}_{-i})|\boldsymbol{X}_i\right]$ for $\widehat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})}$ that is estimated using a well-behaved estimator, the estimators $\hat{\tau}_{\text{AIPW}}^{\text{HT}}$ and $\hat{\tau}_{\text{AIPW}}^{\text{HA}}$ are consistent; see Lemma 3.1 and the subsequent discussion for more.

With consistency established, we will now consider a set of sufficient conditions that guarantee asymptotic normality of both $\hat{\tau}_{\text{AIPW}}^{\text{HT}}$ and $\hat{\tau}_{\text{AIPW}}^{\text{HA}}$. Theorem 3.1 also provides an explicit expression for the asymptotic variance. Later, this will help us compare the asymptotic efficiency of $\hat{\tau}_{\text{AIPW}}^{\text{HA}}$ compared to its unadjusted sibling $\hat{\tau}^{\text{HA}}$ under a set of specific functions for f, G_n, and specific predictive model for $\widehat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})}$.

Theorem 3.1 (Central Limit Theorem)**.** *Suppose Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7 hold, and further assume the existence of random variables* $\widetilde{Y}_i^{(w)}, w \in \{0,1\}$*, such*

that $(W_i, \boldsymbol{X}_i, U_i, \widetilde{Y}_i^{(0)}, \widetilde{Y}_i^{(1)}; \epsilon_i)$ *are iid. Also, assume*

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(1)}\right) \\ &\qquad = o_p(1/\sqrt{n}) \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\pi}{1-\pi}\left(\hat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(0)}\right) \\ &\qquad = o_p(1/\sqrt{n}). \end{aligned} \tag{3.7}$$

Then

$$\begin{aligned} &\sqrt{n}\left(\hat{\tau}_{AIPW}^{HT}-\bar{\tau}_{DIR}\right) \to_d N(0,\pi(1-\pi)\mathbb{E}\left[\left(R_i-A_i+Q_i\right)^2\right]), \\ &\sqrt{n}\left(\hat{\tau}_{AIPW}^{HA}-\bar{\tau}_{DIR}\right) \to_d N\left(0,\pi(1-\pi)\left(Var[R_i-A_i+Q_i]+\left(\mathbb{E}\left[Q_i\right]\right)^2\right)\right), \end{aligned} \tag{3.8}$$

where

$$\begin{aligned} R_i &= \frac{f(1,\boldsymbol{X}_i,U_i,\pi;\epsilon_i)}{\pi}+\frac{f(0,\boldsymbol{X}_i,U_i,\pi;\epsilon_i)}{1-\pi}, \\ A_i &= \frac{\widetilde{Y}_i^{(1)}}{\pi}+\frac{\widetilde{Y}_i^{(0)}}{1-\pi} \\ Q_i &= \mathbb{E}\left[\frac{G(U_i,U_j)(f'(1,\pi,\boldsymbol{X}_j,U_j;\epsilon_j)-f'(0,\pi,\boldsymbol{X}_j,U_j;\epsilon_j))}{\mathbb{E}\left[G(U_i,U_j)|U_j\right]}\middle| U_i\right]. \end{aligned}$$

If $\sqrt{n}\rho_n \to \infty$*, then with* $\sigma_0^2 = Var[f(1,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)-f(0,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)]$*,*

$$\begin{aligned} &\sqrt{n}\left(\hat{\tau}_{AIPW}^{HT}-\tau_{DIR}\right) \to_d N(0,\sigma_0^2+\pi(1-\pi)\mathbb{E}\left[\left(R_i-A_i+Q_i\right)^2\right]), \\ &\sqrt{n}\left(\hat{\tau}_{AIPW}^{HA}-\tau_{DIR}\right) \to_d N\left(0,\sigma_0^2+\pi(1-\pi)\left(Var[R_i-A_i+Q_i]+\left(\mathbb{E}\left[Q_i\right]\right)^2\right)\right), \end{aligned} \tag{3.9}$$

See proof in Appendix 3.C.

Lunceford and Davidian (2004) consider the asymptotic behavior of the estimator $\hat{\tau}_{\text{AIPW}}^{\text{HA}}$

under no interference when $\widetilde{Y}_i^{(w)} = \mathbb{E}\left[Y_i(w)|\boldsymbol{X}_i\right]$, and show that the asymptotic variance is

$$\begin{aligned}\mathbb{E}\left[\frac{\left(Y(1)-\mathbb{E}\left[Y(1)\right]\right)^2}{\pi}+\frac{\left(Y(0)-\mathbb{E}\left[Y(0)\right]\right)^2}{1-\pi}\right] \\ -\left(\mathbb{E}\left[\sqrt{\tfrac{1-\pi}{\pi}}\left(\widetilde{Y}_i^{(1)}-\mathbb{E}\left[Y(1)\right]\right)+\sqrt{\tfrac{\pi}{1-\pi}}\left(\widetilde{Y}_i^{(0)}-\mathbb{E}\left[Y(0)\right]\right)\right]\right)^2.\end{aligned} \tag{3.10}$$

When no units interfere with one another, $f(w,\pi,\boldsymbol{x},u;\epsilon) = f(w,\pi',\boldsymbol{x},u;\epsilon)$ for all $\pi,\pi' \in (0,1)$, and therefore $Q_i = 0$, in which case it can be shown that the asymptotic variance in Equation 3.9 is equivalent to the expression in Equation 3.10.

The practical implications of this result are mixed. First, this means any analysis performed using an augmented IPW estimator under the assumption of no interference will still yield a consistent estimate of the direct effect, and the estimator is asymptotically normal. Second, and on a more negative note, any consistent estimator of the variance of $\hat{\tau}^{\text{HT}}_{\text{AIPW}}$ or $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$ would be hard to come by. Even a conservative estimator of the variance in Equation 3.9 would be useful, but the search for such is complicated by the fact that Q_i is entirely unknown, and its correlation with the observed outcomes could be in either direction. Further restrictions of the graphon G and/or the potential outcomes function f would help in this quest, but we leave this as a topic for future research. One can also use this result to perform an ad hoc assessment similar to that of Section 3 of Li and Wager (2022).

Third, the asymptotic variances of Theorem 3.1 does provide some guidance for how to create $\hat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})}$. To reduce variance, choosing $\hat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})}$ such that A_i is similar to $R_i + Q_i$ is the goal, but doing so is not an easy feat. For example, consider the hypothetical situation where the functional form of f is known to be linear in its arguments. In this case, $f'(1,\pi,\boldsymbol{X}_j,U_j) - f'(0,\pi,\boldsymbol{X}_j,U_j) = 0$ and therefore $Q_i = 0$. Intuitively, we would like to use an estimate of $f(w,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)$ for $\hat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})}$, but since U_i is unobservable, its effect on the potential outcome cannot be corrected for in our estimate. If the latent position U_i is very influential on the potential outcomes, the difference $\hat{Y}_{i,n}^{(w)} - f(w,\pi,\boldsymbol{X}_i,U_i;\epsilon)$ could be very large. We consider this case in greater detail in Section 3.3.1.1.

Despite the challenges one might have in estimating $f(w,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)$ even in the most

optimistic hypothetical, it is nonetheless the most straight forward strategy in improving precision of the estimator $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$. In the more general case where $Q_i \neq 0$, the easiest way of minimizing the difference $R_i + Q_i - A_i$ would be to use $f(w, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)$ as a target for $\widehat{Y}^{(w,\boldsymbol{w}_{-i})}_{i,n}$. First, this can be consistently estimated. Second, since Q_i is a function of the latent position U_i which depends on the edge probabilities $G(U_i, U_j)$ and the difference $f'(1, \pi, \boldsymbol{X}_i, U_i; \epsilon_i) - f'(0, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)$, targeting Q_i would be a much more daunting task. While this might be possible in certain instances, for example with large amounts of data to inform the estimation of the difference in derivatives, or the ability to observe multiple networks that one has reason to believe are drawn from similar graphon models, targeting $f(w, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)$ is a more feasible approach that can be accomplished using traditional statistical methods or modern machine learning tools.

From a practical perspective, targeting $f(w, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)$ would often be approached by estimating some conditional expectation of Y_i given treatment status and covariates. Here, the challenge lies in the fact that the observed data $\{Y_i, \boldsymbol{X}_i, W_i\}$ are not independent, since $Y_i \not\perp W_j$ if $E_{ij} = 1$. However, for n large enough the data will behave as if they are approximately i.i.d., and therefore it is our belief that many methods will perform quite well, even without adjusting for the dependency between observational units. It is important to keep in mind that the rate of convergence is driven by the number of neighbors. In a large, dense network, each observational unit will have a larger number of neighbors, and since the treatment is given randomly, the observed proportion of peers who are treated M_i/N_i will be closer to the limit π, and so variation due to interference will be more neglible than in a smaller, sparse network. This point, that data from more connected networks would behave more like independent data, does seem counterintuitive. It is a result of the anonymous interference assumption (Assumption 3.6) and the assumption that the potential outcomes are smooth (Assumption 3.7). This combination implies that when $M_i/N_i \approx \pi$, the potential outcomes $Y_i(w, \boldsymbol{W}_{-i}) = f(w, M_i/N_i, \boldsymbol{X}_i, U_i; \epsilon_i) \approx f(w, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)$, which means the observed data are approximately independent and identically distributed. Also, while the latent positions U_i are unobservable, they can be estimated with some nice asymptotic guarantees, at least for the special cases where the graphon is a random dot product graph

(Sussman et al., 2014). This might be useful in improving precision. Though we do not have any theoretical results to back this up, we do explore this idea numerically in Section 3.4.3.

The two additional assumptions introduced in Theorem 3.1 (Equation 3.6 and Equation 3.7) are high level assumptions that provide little insight to which models can be used. The following lemma presents more manageable conditions that are sufficient to guarantee Equation 3.6 and Equation 3.7.

Lemma 3.1. *Assume* $\hat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})} = \hat{h}_n^{(w)}(\boldsymbol{X}_i, M_i/N_i, U_i)$*, where* $\hat{h}_n^{(w)}$ *are random functions such that there exists a P-Donsker class* $\mathcal{F}$ *with* $P(\hat{h}_n^{(w)} \in \mathcal{F}) \to 1$*. We assume there exists* $h^{(w)} \in L_2(P), w \in \{0,1\}$ *with* $\int \left(\hat{h}_n^{(w)}(x,\pi,u) - h^{(w)}(x,\pi,u)\right)^2 dP_{(\boldsymbol{X}_i,U_i)}(x,u) \to_p 0$*. Then Equation 3.6 and Equation 3.7 hold with* $\widetilde{Y}_i^{(w)} = h^{(w)}(\boldsymbol{X}_i, \pi, U_i)$ *if one of the following is true:*

a. $\hat{h}_n^{(w)}(\boldsymbol{X}_i, M_i/N_i, U_i) = \hat{h}_n^{(w)}(\boldsymbol{X}_i, U_i)$
b. $\hat{h}_n^{(w)}(\boldsymbol{X}_i, M_i/N_i, U_i) = \hat{f}_n(\boldsymbol{X}_i, U_i) + \hat{\beta} M_i/N_i$ *with* $\hat{f}_n \to_p f$ *and* $\hat{\beta}_n \to_p \beta \in \mathbb{R}$

See proof in Appendix 3.C.

Lemma 3.1 first introduces a set of general assumptions on the predictive models $\hat{h}_n^{(w)}$. These guarantee that the predictive models are asymptotically well-behaved and converge to functions $h^{(w)}$. Second, Lemma 3.1 provides two choices that the researcher is free to make that guarantees that $\hat{\tau}_{\text{AIPW}}^{\text{HT}}$ and $\hat{\tau}_{\text{AIPW}}^{\text{HA}}$ are both asymptotically normal. The first choice is to only use the i.i.d. covariates $\boldsymbol{X}_i$ along with latent positions (in practice, estimated latent positions), but exclude the observed proportion of peers who are treated in our predictive models. This immediately implies that if data were analyzed as if there were no interference, i.e. we only adjust for i.i.d. covariates, the resulting estimator is asymptotically normal. The second choice is to include the observed proportion of peers who are treated as a predictor, but include it in a partial linear model. This might not be the most flexible of models, but it is nonetheless a proof of concept. Recall, these are sufficient, but not necessary, conditions for asymptotic normality. There are possibly many more models in which we can include

the observed proportion of peers who are treated as a covariate, and achieve an estimator that is asymptotically normal.

3.3.1 ANCOVA Estimator as a Hájek Style AIPW Estimator

Under the classical SUTVA assumption, estimating the average treatment effect using an Analysis of Covariance (ANCOVA) model is asymptotically equivalent to an AIPW estimator (Tsiatis et al., 2008). The simplicity of the ANCOVA model makes it a widely taught and used statistical method, and so being able to give similar theoretical guarantees for the ANCOVA model as those given for a more complex AIPW estimator is of practical significance. Here, we expand the work of Tsiatis et al. (2008) to show that similar results hold in the general interference setup considered throughout this paper.

Consider the ANCOVA model postulating a linear relationship between outcome, covariates, and treatment,

$$\mathbb{E}\left[Y_i | \boldsymbol{X}_i, W_i\right] = \gamma_0 + \gamma_{\boldsymbol{X}}^T \boldsymbol{X}_i + \beta_W W_i. \tag{3.11}$$

An estimator for $\bar{\tau}_{\text{DIR}}$ is given by the estimate for β_W using ordinary least squares. We will refer to this estimator as $\hat{\beta}^{\text{ANCOVA}}$. As mentioned by Tsiatis et al. (2008), this estimator can be written as

$$\begin{aligned} &\hat{\beta}^{\text{ANCOVA}} \\ &= \left\{1 - \frac{n^2}{n_0 n_1}(n^{-1}d_1)^T \hat{\Sigma}_{\boldsymbol{XX}}^{-1}(n^{-1}d_1)\right\}^{-1} \\ &\qquad \left\{\hat{\tau}^{\text{HA}} - \frac{n}{n_0 n_1}\sum_{i=1}^{n}(W_i - \hat{\pi})\hat{\Sigma}_{\boldsymbol{XY}}^T \hat{\Sigma}_{\boldsymbol{XX}}^{-1}\boldsymbol{X}_i\right\}, \end{aligned} \tag{3.12}$$

where $n_1 = \sum_{i=1}^n W_i$, $n_0 = n - n_1$, $d_1 = \sum_{i=1}^n (W_i - \hat{\pi})\boldsymbol{X}_i$, $\hat{\Sigma}_{\boldsymbol{XX}} = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T$, and $\hat{\Sigma}_{\boldsymbol{XY}} = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(Y_i - \bar{Y})$.[1] In the absence of interference, this estimator is asymptotically equivalent to an estimator of the same form as $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$ – Tsiatis et al.

[1] We use $\bar{\boldsymbol{X}}$ for the p dimensional vector with i'th entry $\frac{1}{n}\sum_{j=1}^n X_{ij}$, that is the average value of the j'th covariate, and $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ for the average observed outcome.

(2008) arrive at this conclusion by showing that the first factor in Equation 3.12 goes to 1 in probability, and the second factor is exactly of the desired form with $\hat{Y}_{i,n}^{(w)} = \hat{\Sigma}_{\boldsymbol{X}Y}^{T}\hat{\Sigma}_{\boldsymbol{X}\boldsymbol{X}}^{-1}\boldsymbol{X}_i$.

Here, we will show that the ANCOVA estimator is consistent and asymptotically Normal, even in the presence of interference.

Proposition 3.2. *Under Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7, $\hat{\beta}^{ANCOVA}$ is consistent for $\bar{\tau}_{DIR}$, and asymptotically normal in that*

$$\sqrt{n}\left(\hat{\beta}^{ANCOVA} - \bar{\tau}_{DIR}\right) \to_d N\left(0, \pi(1-\pi)\left(Var\left[\hat{\tau}^{HA}\right] + Var[A_i] - 2Cov[A_i, R_i]\right)\right),$$

where

$$\begin{aligned}
Var\left[\hat{\tau}^{HA}\right] &= Var[R_i + Q_i] + (\mathbb{E}[Q_i])^2, \\
R_i &= \frac{f(1, X_i, U_i, \pi; \epsilon_i)}{\pi} + \frac{f(0, X_i, U_i, \pi; \epsilon_i)}{1-\pi}, \\
A_i &= \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\Sigma_{\boldsymbol{X}\boldsymbol{f}}^{T}\Sigma_{\boldsymbol{X}\boldsymbol{X}}^{-1}\boldsymbol{X}_i, \\
Q_i &= \mathbb{E}\left[\frac{G(U_i, U_j)(f'(1, \pi, \boldsymbol{X}_j, U_j) - f'(0, \pi, \boldsymbol{X}_j, U_j))}{\mathbb{E}\left[G(U_i, U_j)|U_j\right]}\middle| U_i\right]
\end{aligned}$$

with

$$\begin{aligned}
\Sigma_{\boldsymbol{X}f} &= \mathbb{E}[\boldsymbol{X}_i f(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)] - \mathbb{E}[\boldsymbol{X}_i]\,\mathbb{E}[f(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)] \\
\Sigma_{\boldsymbol{X}\boldsymbol{X}} &= \mathbb{E}[(\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])(\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])^T]
\end{aligned}$$

See proof in Appendix 3.C.

This result carries some implications for practitioners. For one, it illustrates that if a researcher analyze data under the assumption of no interference using an ANCOVA model to adjust for covariates, the estimator is still consistent and asymptotically normal. Again, notice the robustness to model misspecification. Unfortunately, any inference made about the direct effect comes with much uncertainty as variance estimation is not feasible in the general setting, and conclusions should therefore be taken with some skepticism.

One immediate consequence of the result above is that the estimator $\hat{\beta}^{\text{ANCOVA}}$ is asymptotically more efficient than $\hat{\tau}^{\text{HA}}$ if and only if $\text{Var}[A_i] - 2\text{Cov}[R_i, A_i] < 0$. Due to the exact form of the ANCOVA estimator $\hat{\beta}^{\text{ANCOVA}}$, we can further simplify this expression.

The corollary below does that for a univariate model, but the result can be extended to multivariate models.

Corollary 3.1. *Consider the ANCOVA model in Equation 3.11 where* $X_i \in \mathbb{R}$. *Let* $\rho_{W_i} = Cov[f(W_i, \pi, X_i, U_i), X_i]$, $\rho_0 = Cov[f(0, \pi, X_i, U_i), X_i]$, *and* $\rho_0 = Cov[f(0, \pi, X_i, U_i), X_i]$. *We can characterize the relationship between* $Var\left[\hat{\beta}^{ANCOVA}\right]$ *and* $Var\left[\hat{\tau}^{HA}\right]$ *as follows:*

i. $Var\left[\hat{\beta}^{ANCOVA}\right] < Var\left[\hat{\tau}^{HA}\right]$ *if* $\rho_{W_i} \neq 0$ *and* $\frac{(1-\pi)\rho_1 + \pi\rho_0}{\rho_{W_i}} > \frac{1}{2}$.

ii. $Var\left[\hat{\beta}^{ANCOVA}\right] > Var\left[\hat{\tau}^{HA}\right]$ *if* $\rho_{W_i} \neq 0$ *and* $\frac{(1-\pi)\rho_1 + \pi\rho_0}{\rho_{W_i}} < \frac{1}{2}$.

iii. $Var\left[\hat{\beta}^{ANCOVA}\right] = Var\left[\hat{\tau}^{HA}\right]$ *if either* X_i *and* f *are uncorrelated, i.e.* $\rho_{W_i} = 0$, *or* $\frac{(1-\pi)\rho_1 + \pi\rho_0}{\rho_{W_i}} = \frac{1}{2}$.

See proof in Appendix 3.C.

Proposition 3.2 reveals that if the covariate X_i has a similar effect on potential outcomes when treated as it does on potential outcomes when not treated, then the $\hat{\beta}^{\text{ANCOVA}}$ is always more efficient than $\hat{\tau}^{\text{HA}}$, and the adjusting for covariates using the ANCOVA model is recommendable. As we will see in the next section, this is not always the case, and hence there are situations where $\hat{\beta}^{\text{ANCOVA}}$ is *less* efficient than the simpler $\hat{\tau}^{\text{HA}}$.

One way to mitigate this is by including an interaction term between the treatment and covariate in the ANCOVA model. As mentioned by Tsiatis et al. (2008), this model can be expressed in terms of centered versions of the outcome, treatment, and control as

$$\begin{aligned} &\mathbb{E}\left[Y_i - \mathbb{E}\left[Y_i\right] | W_i, \boldsymbol{X}_i\right] \\ &\quad = \beta_X^T \left(\boldsymbol{X}_i - \mathbb{E}\left[\boldsymbol{X}_i\right]\right) + \beta_W \left(W_i - \mathbb{E}\left[W_i\right]\right) \\ &\qquad + \beta_{XW}^T \left(\boldsymbol{X}_i - \mathbb{E}\left[\boldsymbol{X}_i\right]\right)\left(W_i - \mathbb{E}\left[W_i\right]\right) \end{aligned} \tag{3.13}$$

This model is fitted by ordinary least squares regression of $Y_i - \bar{Y}$ on $\boldsymbol{X}_i - \bar{\boldsymbol{X}}$,

$\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)(W_i - \hat{\pi})$, and $W_i - \hat{\pi}$ since $\hat{\pi} = \bar{W}$. This estimator can be written as

$$\begin{aligned}
&\hat{\beta}^{\text{ANCOVA2}} \\
&= \left\{1 - \frac{n^2}{n_0 n_1}(n^{-1}d_2)^T D^{-1}(n^{-1}d_2)\right\}^{-1} \\
&\quad \left\{\hat{\tau}^{\text{HA}} - \frac{n}{n_0 n_1} d_2^T D^{-1} \begin{pmatrix} \widehat{\Sigma}_{\boldsymbol{X}Y} \\ \widehat{\Sigma}_{\boldsymbol{X}YW} \end{pmatrix}\right\},
\end{aligned} \tag{3.14}$$

where $d_2 = \left(\left(\sum_{i=1}^n (W_i - \hat{\pi})\boldsymbol{X}_i\right)^T, \sum_{i=1}^n (W_i - \hat{\pi})^2(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_i^T)\right)^T$, $\widehat{\Sigma}_{\boldsymbol{X}YW} = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(Y_i - \bar{Y})(W_i - \hat{\pi})$, and

$$D = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T & \frac{1}{n}\sum_{i=1}^n (W_i - \hat{\pi})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T \\ \frac{1}{n}\sum_{i=1}^n (W_i - \hat{\pi})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T & \frac{1}{n}\sum_{i=1}^n (W_i - \hat{\pi})^2(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T \end{pmatrix}.$$

Proposition 3.3 shows that this estimator is also consistent for $\bar{\tau}_{\text{DIR}}$ and asymptotically normal.

Proposition 3.3. *Under Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7, $\hat{\beta}^{ANCOVA2}$ is consistent for $\bar{\tau}_{DIR}$, and asymptotically normal in that*

$$\sqrt{n}\left(\hat{\beta}^{ANCOVA2} - \bar{\tau}_{DIR}\right) \to_d N\left(0, \pi(1-\pi)\left(Var\left[\hat{\tau}^{HA}\right] + Var[A_i] - 2Cov[A_i, R_i]\right)\right),$$

where

$$\begin{aligned}
Var\left[\hat{\tau}^{HA}\right] &= Var[R_i + Q_i] + (\mathbb{E}[Q_i])^2, \\
R_i &= \frac{f(1, X_i, U_i, \pi; \epsilon_i)}{\pi} + \frac{f(0, X_i, U_i, \pi; \epsilon_i)}{1-\pi}, \\
A_i &= \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\left[\pi\Sigma^{(0)}_{\boldsymbol{X}f} + (1-\pi)\Sigma^{(1)}_{\boldsymbol{X}f}\right]^T \Sigma^{-1}_{\boldsymbol{X}\boldsymbol{X}}\boldsymbol{X}_i, \\
Q_i &= \mathbb{E}\left[\frac{G(U_i, U_j)(f'(1, \pi, \boldsymbol{X}_j, U_j) - f'(0, \pi, \boldsymbol{X}_j, U_j))}{\mathbb{E}\left[G(U_i, U_j)|U_j\right]}\middle| U_i\right]
\end{aligned}$$

See proof in Appendix 3.C.

Similar to Corollary 3.1, we here characterize the relationship between the variance of the Hájek estimator and the variance of $\hat{\beta}^{\text{ANCOVA2}}$.

Corollary 3.2. *If the ANCOVA model in Equation 3.13 includes a single covariate, then* $Var\left[\hat{\beta}^{ANCOVA2}\right] \leq Var\left[\hat{\tau}^{HA}\right]$ *with equality if and only if* $\pi Cov[f(0,\pi,X_i,U_i;\epsilon_i),X_i] + (1-\pi)Cov[f(1,\pi,X_i,U_i;\epsilon_i),X_i] = 0$.

See proof in Appendix 3.C.

Corollary 3.2 shows us that if we fit an ANCOVA model and include the interaction between the covariate and the treatment indicator, the resulting estimator will always be at least as efficient as the unadjusted estimator $\hat{\tau}^{\text{HA}}$. This behavior is similar to the behavior of the ANCOVA estimators when no interference is present (Tsiatis et al., 2008). This means we can always improve precision in randomized clinical trials by adjusting for baseline covariates.

In the following two subsections, we consider two specific potential outcomes functions, and show how we can use the results above to gain insights into the efficiency of the $\hat{\beta}^{\text{ANCOVA}}$ and $\hat{\beta}^{\text{ANCOVA2}}$ estimators.

3.3.1.1 Example: Linear Potential Outcome Function

Consider using $\hat{\beta}^{\text{ANCOVA}}$ to estimate the ATE in a scenario where the potential outcomes for an individual with treatment assignment w, proportion p of neighbors are treated, pre-treatment covariate x, and latent position u, is

$$f(w,p,x,u;\epsilon) = \beta_W \cdot w + \beta_X \cdot x + p + u + \epsilon.$$

In this scenario, $f'(w,p,x,u;\epsilon) = 1$ for all $(w,p,x,u;\epsilon)$, and therefore $Q_i = 0$. This means that the graphon does not influence the asymptotic variance. Under these assumptions, the asymptotic variances of $\hat{\tau}^{\text{HA}}$ and $\hat{\beta}^{\text{ANCOVA}}$ are

$$\begin{aligned}
\text{Var}\left[\hat{\tau}^{\text{HA}}\right] &= \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\left(\beta_X^2 \text{Var}\left[X_i\right] + \text{Var}\left[U_i\right] + \text{Var}\left[\epsilon_i\right]\right) \\
\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] &= \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\left(\text{Var}\left[U_i\right] + \text{Var}\left[\epsilon_i\right]\right).
\end{aligned}$$

It is clear that $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] \leq \text{Var}\left[\hat{\tau}^{\text{HA}}\right]$ with equality when $\beta_X^2 = 0$. We also immediately see that the relative efficiency of the ANCOVA estimator relative to the Hájek estimator is quadratic in the coefficient β_X. See Appendix 3.D.1 for a detailed derivation.

One final note on this specific setup: a consistent estimator for the variance of $\hat{\beta}^{\text{ANCOVA}}$ can be obtained if one is able to consistently estimate the variance of ϵ_i. While we do not present a formal proof, under the assumptions presented here, this is a much less daunting task than a general estimator for the asymptotic variance in Proposition 3.2.

3.3.1.2 Example: Quadratic Outcome Function

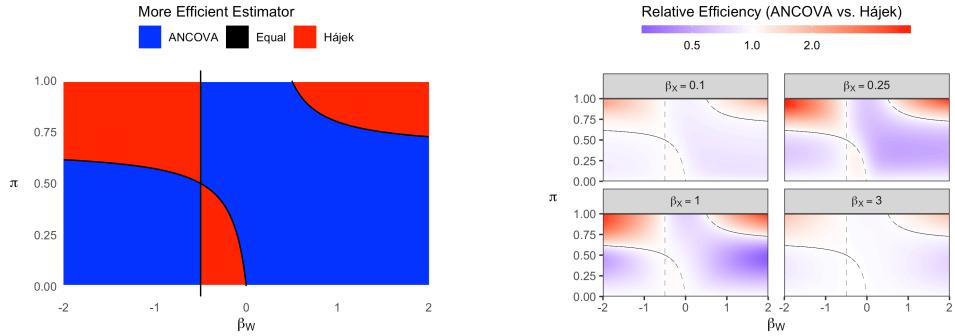
Next, a quadratic potential outcomes function,

$$f(w, p, x, u; \epsilon) = \frac{1}{2}\left(\beta_W w + \beta_X x + p \cdot u\right)^2 + \epsilon/5.$$

Here, the Q_i term is not constantly equal to 0, and the graphon therefore influences the asymptotic behavior of the estimators $\hat{\beta}^{\text{ANCOVA}}$ and $\hat{\beta}^{\text{ANCOVA2}}$. We will consider the graphon model G_1; see Appendix 3.A for details. The expressions for the variances are much more complex here than what we derived under the linear potential outcomes function above; the detailed derivations are included in the Appendix (specifically Appendix 3.D.2).

Under this potential outcomes function, it can be verified that $\text{Cov}\left[f(w, \pi, X_i, U_i; \epsilon_i), X_i\right] = \beta_W \beta_X \cdot w + \beta_X \frac{\pi}{2}$. We can characterize the difference $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] - \text{Var}\left[\hat{\tau}^{\text{HA}}\right]$ using Corollary 3.1. Interestingly, under this particular potential outcomes function, the difference $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] - \text{Var}\left[\hat{\tau}^{\text{HA}}\right]$ does not depend on β_X with the exception of $\beta_X = 0$, which implies the variances are the same. Figure 3.1(a) shows the sign of the difference $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] - \text{Var}\left[\hat{\tau}^{\text{HA}}\right]$ for (β_W, π) in $[-2, 2] \times (0.01, 0.99)$ for any $\beta_X \neq 0$; that is, we see when $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] < \text{Var}\left[\hat{\tau}^{\text{HA}}\right]$, $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] > \text{Var}\left[\hat{\tau}^{\text{HA}}\right]$, and $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] = \text{Var}\left[\hat{\tau}^{\text{HA}}\right]$. We see that for certain values of β_W, for example $\beta_W = 0.25$, the ANCOVA estimator is more efficient for any value of π. However, there are values of β_W where the

ANCOVA estimator is less efficient than the Hájek estimator for a number of values of π. If, for example, the researcher decided to run an experiment assigning treatments with probability $\pi = 2/3$, and it happened to be that $\beta_W = -1$, then the Hájek estimator would be the more efficient estimator. It is also worth noting that for the commonly used $\pi = 1/2$, $\hat{\beta}^{\text{ANCOVA}}$ is at least as efficient as the unadjusted estimator $\hat{\tau}^{\text{HA}}$.



(a) Dichotomized picture of the difference between $\text{Var}\left[\hat{\tau}^{\text{HA}}\right]$ and $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right]$. This picture is identical for all $\beta_X \neq 0$.

(b) The relative efficiency $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] / \text{Var}\left[\hat{\tau}^{\text{HA}}\right]$. Note how each of the four panels mirror the trend shown in (a).

Figure 3.1: Heatmaps indicating when $\text{Var}\left[\hat{\tau}^{\text{HA}}\right] < \text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right]$, $\text{Var}\left[\hat{\tau}^{\text{HA}}\right] = \text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right]$, $\text{Var}\left[\hat{\tau}^{\text{HA}}\right] > \text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right]$ as a function of π and β_W.

While it is somewhat simple to characterize *when* the ANCOVA estimator is more efficient than the Hájek estimator, it is more complicated to describe the relative efficiency. This not only depends on π and β_W, but also β_X.

Figure 3.1(b) shows heatmaps of the relative efficiency of the ANCOVA estimator relative to the Hájek estimator for four different values of β_X across values of β_W and π. While each of the four panels reiterate the main points of Figure 3.1, there are many areas where the estimators are not exactly equally efficient, but the relative efficiency is very close to 1. Also, in the scenarios where one is more efficient than the other, the magnitude of the relative efficiency varies dramatically depending on the value of β_X.

Figure 3.2 recreates Figure 3.1(b) for the relative efficiency of $\hat{\beta}^{\text{ANCOVA2}}$ relative to $\hat{\tau}^{\text{HA}}$. The most striking difference is that the relative efficiency is never above 1, as we expected per Corollary 3.2. Again, the relative efficiency is highly dependend on the coefficients β_X

and β_W, and the chosen propensity π.

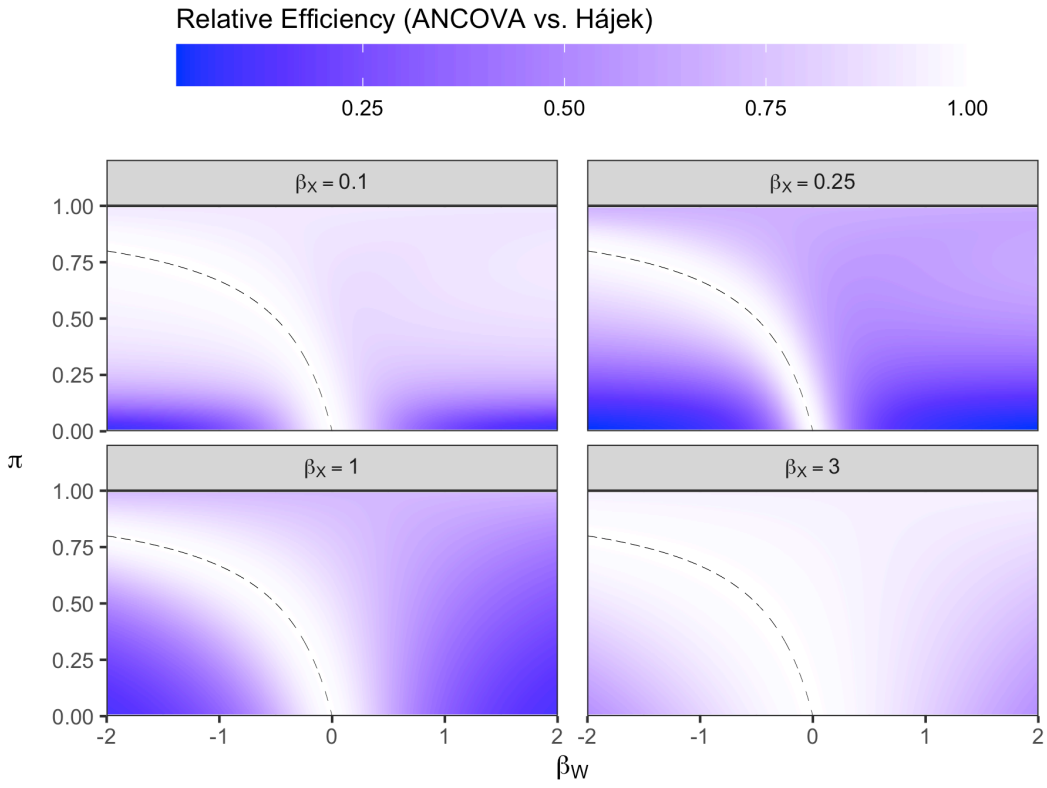


Figure 3.2: Relative efficiency (as $\text{Var}\left[\hat{\beta}^{\text{ANCOVA2}}\right]/\text{Var}\left[\hat{\tau}^{\text{HA}}\right]$) for combinations of β_W, π, and β_X. Dashed lines show the values where $\text{Var}\left[\hat{\beta}^{\text{ANCOVA2}}\right] = \text{Var}\left[\hat{\tau}^{\text{HA}}\right]$, and indicate where the relative efficiency is exactly equal to 1.

This example shows that if one wants to use the popular and well known ANCOVA model to adjust for covariates when estimating $\bar{\tau}_{\text{DIR}}$, an interaction term between covariates and treatment assignment is crucial to guarantee variance reduction.

3.4 Numerical Results

In this section, we present important observations regarding the asymptotic behavior of the AIPW estimators based on simulation results. This section is split into four subsections: first, we consider the standard error of the ANCOVA estimator. Second, we look at testing a sharp null hypothesis of no treatment effect. Third, we will consider the use of more flexible predictive models for the AIPW estimator. Finally, we consider the estimation of the AIPW estimator using the sample splitting algorithm proposed by Emmenegger et al. (2023).

All the simulations performed in this section are done in a super-population setting. That is, the potential outcomes are **not** kept fixed between simulations, but new values of covariates and error terms drawn for each simulation. This means the target estimand in the simulations is τ_{DIR}.

3.4.1 Is the OLS variance estimator of any use?

In Section 3.3.1, we saw that the OLS estimator $\widehat{\beta}^{\text{ANCOVA2}}$ is a reasonable choice for estimating $\bar{\tau}_{\text{DIR}}$ and τ_{DIR}. We also derived the asymptotic variance, but did not provide an estimator. To the best of our knowledge, a consistent variance estimator Though not theoretically proven to be consistent, it seems natural to ask if the OLS variance estimator can be useful in practice. Here, we will explore the behavior of the OLS variance estimator.

We perform a simulation study to gain some insights to the behavior of the OLS variance estimator. For various potential outcomes functions and various graphon models, we simulate 3000 data sets of sample size 1000. For each data set, we estimate τ_{DIR} using ANCOVA for various working models, and calculate the estimate standard error using the square root of classic OLS variance estiamtor. We compare the 3000 standard error estimates to the standard deviation of the 3000 ANCOVA estimates to get a sense of the bias and variability of the OLS variance estimator.

For simplicity, we present a few of the scenarios considered. We restrict ourselves to one

graphon model (G_1) and use $\rho_n = n^{-1/3} = 0.1$, but note that the results were similar across various graphon models and ρ_n sequences. We present results for potential outcomes functions f_1, f_2, and f_4 (see Appendix 3.A.2 for details) for $\beta_X \in \{0.5, 5\}$, and for propensities $\pi \in \{0.3, 0.5, 0.9\}$. We consider $\widehat{\beta}^{\text{ANCOVA}}$ and $\widehat{\beta}^{\text{ANCOVA2}}$, and variants thereof which include the observed proportion of peers treated as a covariate. Note that we do not have theoretical results for the estimators including network summaries, but this is common practice, and as such we found it to be of interest to include these here. For each of the four estimators, we get the corresponding SE estimates based on the OLS variance estimator.

Figure 3.3 shows the results. First of all, we see that the accuracy of the OLS variance estimator is highly dependend on the true potential outcomes function, and how the covariate X_i influences the potential outcomes (here, through the coefficient β_X). This is to be expected. It is also worth noting that the OLS variance estimator is not biased in a consistent way. In some scenarios, it is too liberal, in others too conservative. This makes it hard to rely on this estimator in practice, since the direction of the bias will be unknown.

There are many variance estimators that improve on the OLS variance estimator. Two such examples are the so-called "sandwich" estimator and the bootstrap variance estimator. In our experience, the sandwich estimator provides results very similar to the OLS variance estimator. Estimating the variance using a bootstrap approach is tempting in this situation, since the outcomes behave as if i.i.d. asymptotically, which is an assumption required for the bootstrap. Figure 3.4 replicates Figure 3.3, but using the bootstrap to estimate the variance rather than the OLS variance estimator. The overall picture is similar in that the bootstrap variance estimator is sometimes liberal, sometimes conservative. Also, when compared to the OLS variance estimator, it sometimes performs better (as seems to be the case in the third column) and sometimes worse (see the fourth column).

A more interesting observation from Figures 3.3 and 3.4 is how decisions made by the researcher affects the performance of the OLS and bootstrap variance estimators. For the OLS variance estimator, choosing an extreme value for the propensity π of 0.9 can help you get a less biased estimator (see the last two columns of Figure 3.3), but it might also lead

to a more biased estimator (see the first column of Figure 3.3). Also, choosing to include the observed proportion of peers treated as an extra covariate in the working model might increase bias (column 5 of Figure 3.3) or have little to no effect on the bias of the OLS variance estimator. Most interestingly, the OLS variance estimator often performs better when the interaction term is not included, but choosing to not include the interaction term is ill-advised, as discussed in the previous section. The bootstrap variance estimator seems to be less sensitive to the working model. It performs similarly regardless of whether interactions and/or observed proportion of peers are included or not.

The observations made above do not bode well for the use of the classic OLS variance estimator or the naive bootstrap estimator in general. In particular, it is problematic that the direction of the bias is not consistent. We have not been able to uncover any pattern in the direction of the bias, but suspect heterogeneity of the treatment effect is of importance. As mentioned above, we also considered other variance estimators related to the OLS variance estimator, such as the sandwich estimator using heteroscedastic and autocorrelation consistent and clustered covariance estimators, but these did not seem to improve on the overall performance. We do think this is an area that is worth exploring in the future, as a conservative variance estimator would be of great benefit for practical applications. The bootstrap's robustness towards the chosen working model seems promising, and there might be an adjustment to be made to get more consistent behavior.

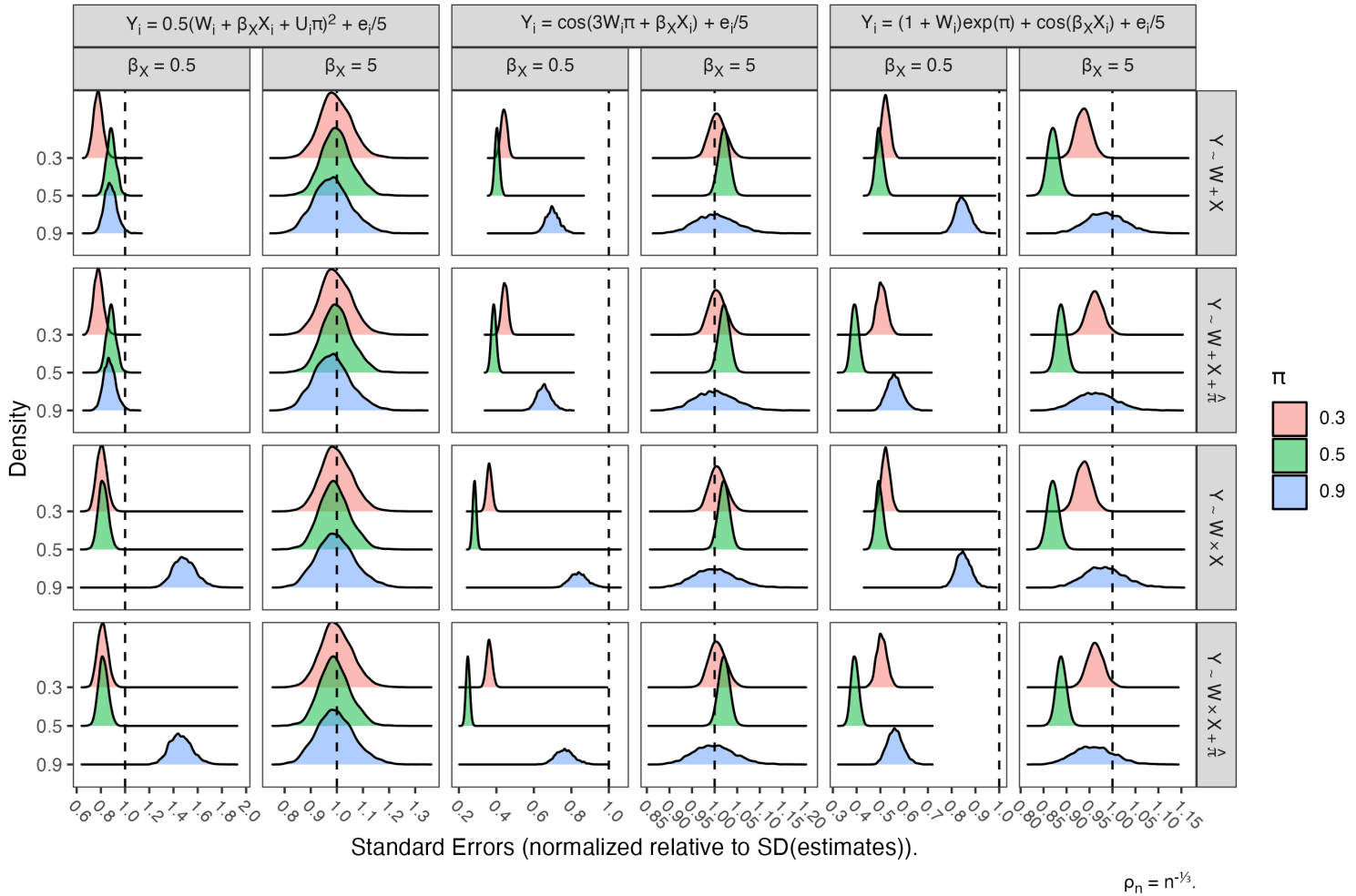


Figure 3.3: Density curves for ratios of standard error estimates using OLS variance estimator over standard deviation of the 3000 estimates of $\bar{\tau}_{\text{DIR}}$.

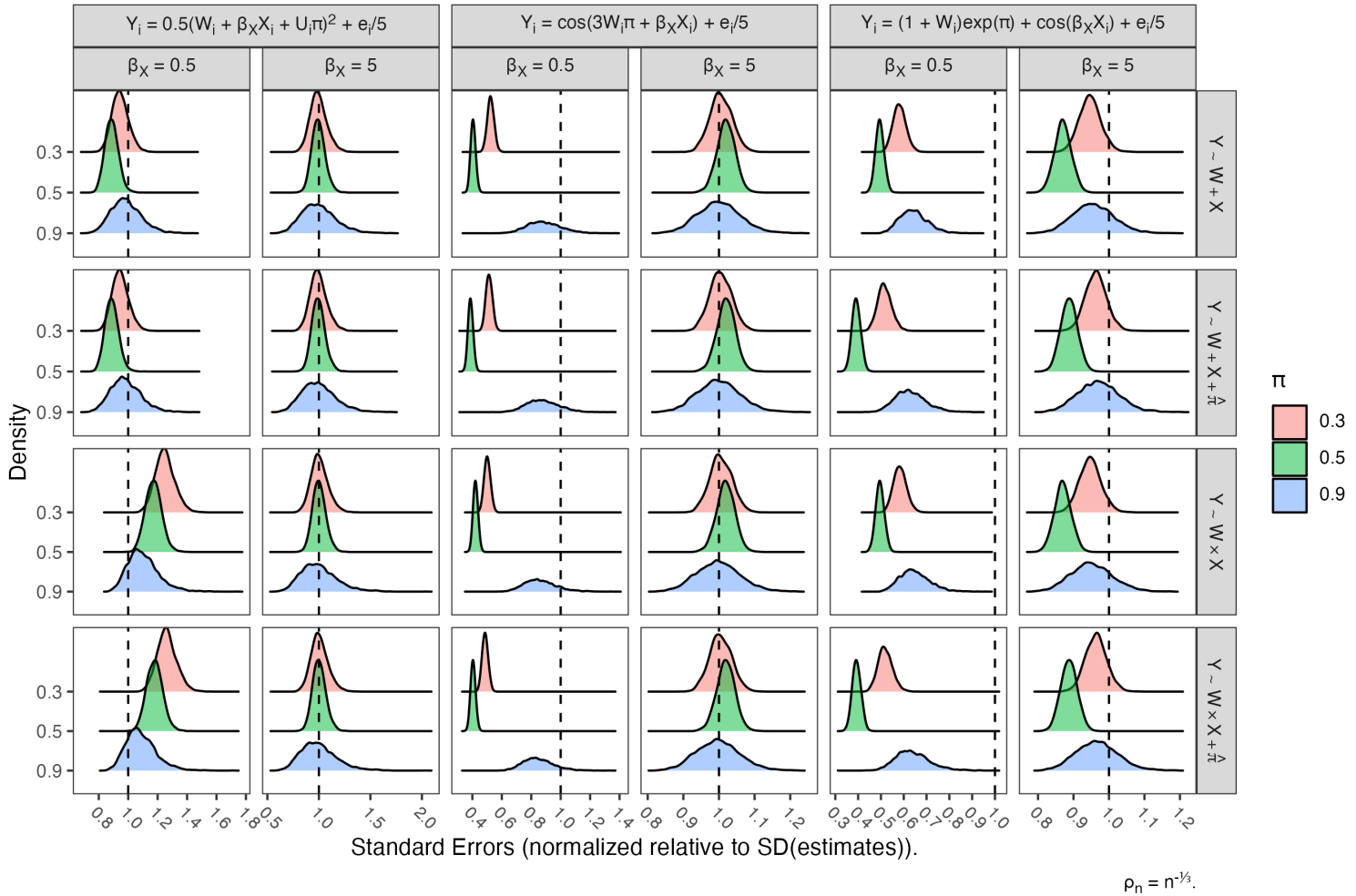


Figure 3.4: Density curves for ratios of standard error estimates using a naive bootstrap variance estimator over standard deviation of the 3000 estimates of $\bar{\tau}_{\text{DIR}}$.

3.4.2 Testing of No Direct Effect

While the variance of $\hat{\beta}^{\text{ANCOVA2}}$ cannot be estimated in general, we can consistently estimate the asymptotic variance under a sharp null hypothesis of no direct treatment effect.

For notational simplicity, consider testing the null hypothesis that $f(1, \pi, X_i, U_i; \epsilon_i) = f(0, \pi, X_i, U_i; \epsilon_i)$ when a single covariate is available. Under this null hypothesis, the derivatives $f'(1, \pi, X_i, U_i; \epsilon_i)$ and $f'(0, \pi, X_i, U_i; \epsilon_i)$, both taken with respect to π, are equal, and therefore $Q_i = 0$. Also, since $f(1, \pi, X_i, U_i; \epsilon_i) = f(0, \pi, X_i, U_i; \epsilon_i)$ for all i, $\sigma_0^2 = \text{Var}\left[f(1, \pi, X_i, U_i; \epsilon_i) - f(0, \pi, X_i, U_i; \epsilon_i)\right] = 0$. So, in this case,

$$\text{Var}\left[\hat{\beta}^{\text{ANCOVA2}}\right] = \pi(1-\pi)\text{Var}\left[R_i - B_i\right].$$

Since $f(1, \pi, X_i, U_i; \epsilon_i) = f(0, \pi, X_i, U_i; \epsilon_i)$,

$$R_i = \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right) f(1, \pi, X_i, U_i; \epsilon_i)$$

and

$$\Sigma_{Xf}^{(1)} = \Sigma_{Xf}^{(0)} = \text{Cov}\left[f(1, \pi, X_i, U_i; \epsilon_i), X_i\right]$$

So,

$$R_i - B_i = \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\left\{f(1, \pi, X_i, U_i; \epsilon_i) - \text{Cov}\left[f(1, \pi, X_i, U_i; \epsilon_i), X_i\right]\text{Var}\left[X_i\right]^{-1} X_i\right\},$$

and therefore

$$\begin{aligned}\text{Var}\left[R_i - B_i\right] = &\left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)^2 \\ &\left\{\text{Var}\left[f(1, \pi, X_i, U_i; \epsilon_i)\right] - \text{Cov}\left[f(1, \pi, X_i, U_i; \epsilon_i), X_i\right]^2 \text{Var}\left[X\right]^{-1}\right\}.\end{aligned}$$

This means we can estimate the variance of $\hat{\beta}^{\text{ANCOVA2}}$ as

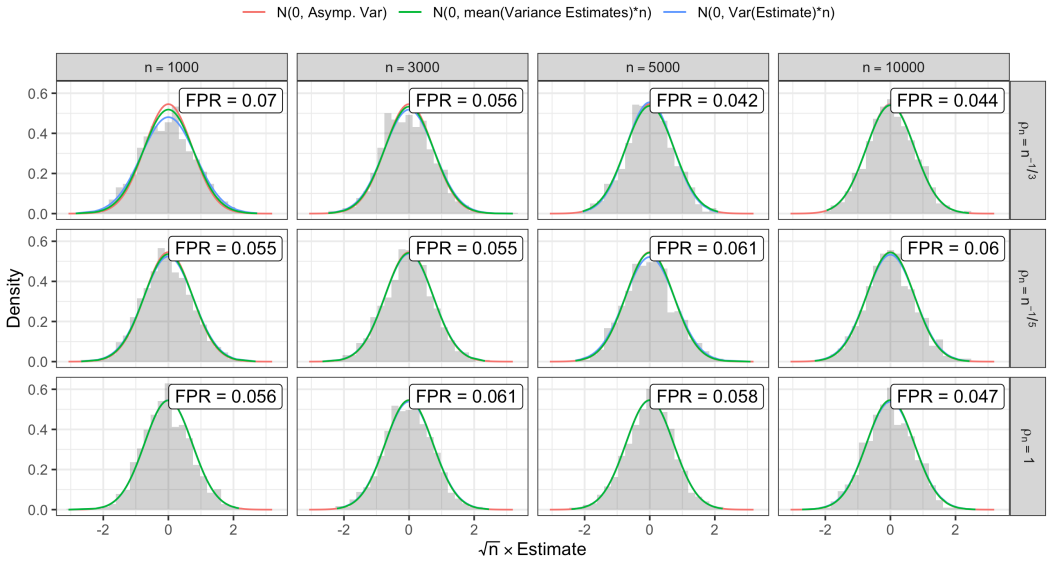
$$\widehat{\text{Var}}\left[\hat{\beta}^{\text{ANCOVA2}}\right] = \pi(1-\pi)\left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)^2 \left\{\widehat{\text{Var}}\left[Y_i\right] - \widehat{\text{Cov}}\left(Y_i, X_i\right)^2 / \widehat{\text{Var}}\left[X_i\right]\right\}, \quad (3.15)$$

since the variance of the observed outcomes converges to the variance of the potential outcomes function f, and the covariance of the observed outcomes and the covariates converges to the covariance of the potential outcomes function and the covariates (see proof of Proposition 3.3 in Appendix 3.C). While this estimator is consistent, it seems that it is biased in finite samples. We illustrate this through a few simulations.

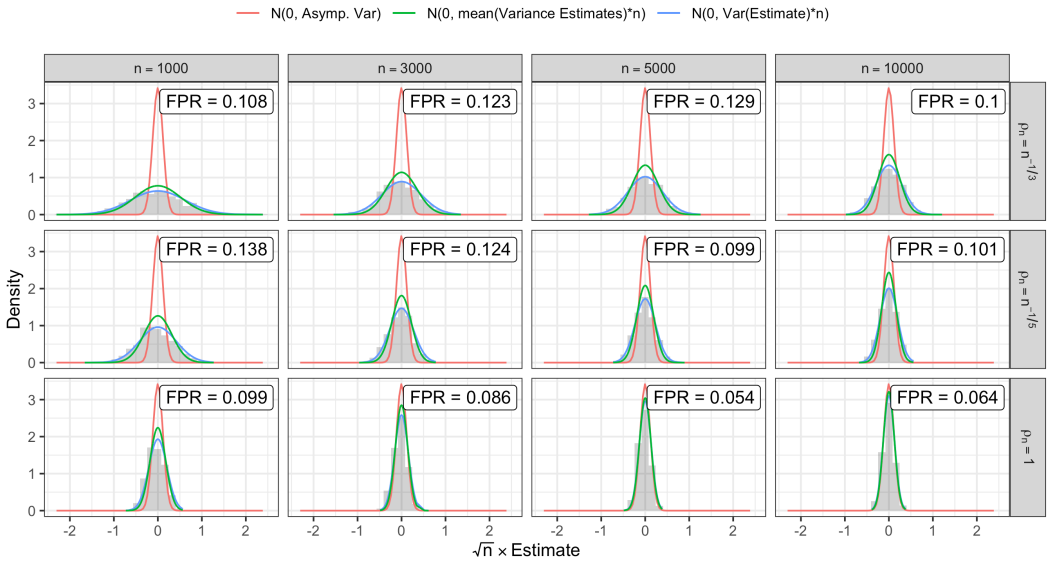
We consider two potential outcomes functions: $f(w, \pi, x, u; e) = (1 + u)^2 \cdot (1 + 0.5 \cdot x) \cdot \exp(\pi)/5 + e/5$ and $f(w, \pi, x, u; e) = \exp(\pi) + \cos(0.5 \cdot x) + e/5$. In either case, $\bar{\tau}_{\text{DIR}} = 0$ and therefore $\tau_{\text{DIR}} = 0$. For each combination of potential outcomes function, $n \in \{1000, 3000, 5000, 10000\}$, and $\rho_n = \{n^{-1/3}, n^{-1/5}, 1\}$, we simulate 1000 data sets, get estimates of τ_{DIR} using $\hat{\beta}^{\text{ANCOVA2}}$, and estimate the variance using Equation 3.15.

Figure 3.5 presents the results histograms of the estimates of τ_{DIR} using $\hat{\beta}^{\text{ANCOVA2}}$ with density curves for three normal distributions with different variances: the asymptotic variance, evaluated using a monte carlo simulation scheme; the average of the 1000 variance estimates found using Equation 3.15; and the variance of the 1000 estimates obtained using $\hat{\beta}^{\text{ANCOVA2}}$. Each panel also shows the false positive rate of testing the null hypothesis with $\alpha = 0.05$. Figure 3.5(a) shows that for the potential outcomes function $f(w, \pi, x, u; e) = (1 + u)^2 \cdot (1 + 0.5 \cdot x) \cdot \exp(\pi)/5 + e/5$, the variance estimator seems to be fairly unbiased, and close to the asymptotic variance. This is not exactly the case for the potential outcomes function $f(w, \pi, x, u; e) = \exp(\pi) + \cos(0.5 \cdot x) + e/5$. Figure 3.5(b) indicates that the variance estimator generally is too liberal resulting in a larger than expected false positive rate. We notice that for the dense case, i.e. when $\rho_n = 1$, the increasing sample size does seem to lead to better performance of the variance estimator, but for the sparse case where $\rho_n \to 0$, the convergence rate seems to influence the performance.

Figure 3.5(b) also provides a suggestion to why the variance estimator based on Equation 3.15 does not perform well. In the sparse settings, we still see a clear separation between the asymptotic variance and the variance of the estimates, even for $n = 10000$, suggesting that larger sample size is necessary for the estimator to behave as its asymptotic limit. This is further reinforced by Figure 3.6, which shows the distributions of the



(a) Potential outcomes function $f(w, \pi, x, u; e) = (1 + u)^2 \cdot (1 + 0.5 \cdot x) \cdot \exp(\pi)/5 + e/5$



(b) Potential outcomes function $f(w, \pi, x, u; e) = \exp(\pi) + \cos(0.5 \cdot x) + e/5$

Figure 3.5: Histograms of $\hat{\beta}^{\text{ANCOVA2}}$ from various combinations of sample size, ρ_n, and potential outcomes function. Each histogram is based on 1000 simulated data sets. Overlaid are three normal curves with varying variances: the asymptotic variance, the average of the 1000 variance estimates, and the variance of the 1000 estimates. FPR = False Positive Rate of testing the null hypothesis with $\alpha = 0.05$. FPR = False Positive Rate of testing the null hypothesis with $\alpha = 0.05$.

standard errors across the 1000 simulated data sets for the potential outcomes function $f(w, \pi, x, u; e) = \exp(\pi) + \cos(0.5 \cdot x) + e/5$ along with the asymptotic standard error and the standard error of the 1000 estimates from $\hat{\beta}^{\text{ANCOVA2}}$. We see that while the standard error of the 1000 estimates does seem to converge towards the asymptotic standard error in all scenarios, this convergence happens at a slower rate when $\rho_n \to 0$. This is not surprising. The asymptotic normality of $\hat{\beta}^{\text{ANCOVA2}}$ is derived using a Taylor expansion, where the contribution from higher order terms converge to 0 at a rate determined by $n\rho_n$. More interestingly, we notice that the estimated standard errors are almost entirely underestimating the standard deviation, suggesting a finite sample correction might be possible. We leave this as an area for further exploration in future research.

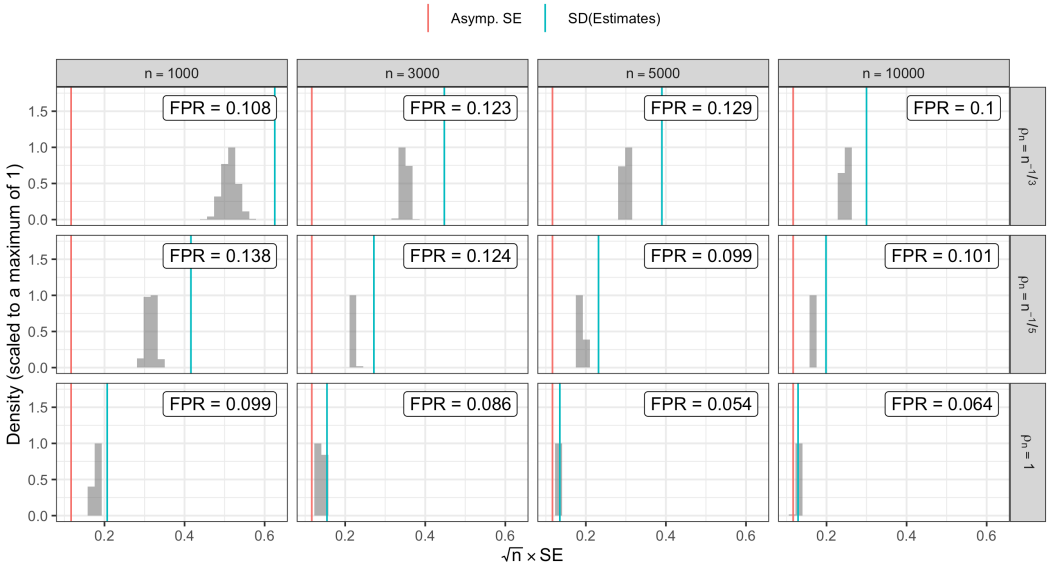


Figure 3.6: Distributions of Standard Error estimates across 1000 simulated data sets for various values of sample size n and ρ_n with potential outcomes function $f(w, \pi, x, u; e) = \exp(\pi) + \cos(0.5 \cdot x) + e/5$.

3.4.3 Extra flexibility: improvement, or unnecessary complexity?

So far, we have considered asymptotic properties of the ANCOVA estimator. We saw how it is possible to derive explicit expressions for this estimator for specific graphon and

potential outcomes functions. This is possible because of the nice asymptotic behavior of the ANCOVA estimator where limiting functions of $\hat{h}_n^{(w)}$ are somewhat simple to derive. Here, we will consider more complex functions with the goal of bigger variance reduction than what the ANCOVA estimator can provide. Due to the nature of these estimators, the assumptions of Theorem 3.1 cannot be verified. In particular, explicit expressions of $h^{(w)}$ are not obtained. We will validate the normality of the estimators through simulations, but do not have the opportunity to validate the asymptotic variance expression, and will instead compare estimated values of asymptotic variances based on simulation results.

We want to illustrate different approaches to reduce variance of the estimator $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$ through choices one can make in building the models $\hat{h}_n^{(w)}$. To do so, we seek to compare the distributions of $\hat{\tau}^{\text{HA}}$ and $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$ for various combinations of potential outcomes functions f, graphons, sequences for ρ_n, and propensities π. We generate 3000 data sets of sample size 1000 for each setting and obtain 3000 estimates using the estimators of interest. The main concern is the estimated variances as it is well understood that the estimators are consistent and asymptotically normal.

For $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$, we use a non-parametric regression model for $\hat{h}_n^{(0)}$ and $\hat{h}_n^{(1)}$. For each generated data set, these are fitted using the two subsets of the data where $W_i = 0$ and $W_i = 1$. Specifically, $\hat{h}_n^{(w)}$ is fitted using the `npreg` function from the R-package `np` with the observed outcome Y as the dependent variable, and the covariate X_i as the independent variable.

The main points of this section are three fold.

1. A more flexible model can lead to significant efficiency gains.

Using a more flexible model for $\hat{h}_n^{(w)}$ can lead to significant efficiency gains. Consider comparing the unadjusted estimator $\hat{\tau}^{\text{HA}}$, the ANCOVA based estimators $\hat{\beta}^{\text{ANCOVA}}$ and $\hat{\beta}^{\text{ANCOVA2}}$, and the AIPW-style estimator $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$ using the non-parametric regression models for $\hat{h}_n^{(w)}$. Figure 3.7 shows ridgeline plots of the four estimators for the linear and quadratic models considered in the previous section, and the graphon G_1 also considered in the previous section. Here, $\beta_W = \beta_X = 1$, $\pi = 0.9$, and $\rho_n = n^{-1/3}$. Vertical lines indicate τ_{DIR}.

From Figure 3.7, the observations made in Section 3.3.1.1 and Section 3.3.1.2 once again surface. For the quadratic potential outcomes function (left panel), we see that the variance of $\hat{\beta}^{\text{ANCOVA}}$ is slightly larger than the variance of Hájek. The inclusion of an interaction term, which results in the estimator $\hat{\beta}^{\text{ANCOVA2}}$, results in smaller variance than the Hájek estimator. For the linear potential outcomes function (right panel), we see that the Hájek estimator has larger variance than $\hat{\beta}^{\text{ANCOVA}}$ and $\hat{\beta}^{\text{ANCOVA2}}$, as is expected. Figure 3.7 also shows the ridgeline plot of $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$ with the non-parametric models used for $\hat{h}^{(w)}_n$. We see that this added flexibility results in smaller variance with little to no additional bias when the potential outcomes function is of the quadratic form, while $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$ performs very similar to $\hat{\beta}^{\text{ANCOVA}}$ and $\hat{\beta}^{\text{ANCOVA2}}$ when the potential outcomes function is linear, i.e. when the model fitted using ANCOVA is the true model.

Figure 3.7 is summarized in Table 3.1 as estimates of the bias and the variances of the estimators. Here, the effect of the more flexible model is very noticeable. When the potential outcomes function is of the quadratic form, the relative efficiency gain of using the more flexible model for $\hat{h}^{(w)}_n$ over the $\hat{\beta}^{\text{ANCOVA2}}$ is similar to the efficiency gain when adjusting for covariates in the first place.

Table 3.1: Bias (Asymptotic Variance Estimate $\times 10^3$) of $\hat{\tau}^{\text{HA}}$, $\hat{\beta}^{\text{ANCOVA}}$, $\hat{\beta}^{\text{ANCOVA}}_2$, and AIPW Estimator using non-parametric regression of Y on X for $\hat{h}^{(w)}_n$. Bias estimated as average deviation from $\bar{\tau}_{\text{DIR}}$ across 3000 simulations of sample size 1000. Variance is the empirical variance of the estimates across 3000 simulations of sample size 1000. $\pi = 0.9$ and $\rho_n = n^{-1/3}$.

	Bias (Variance $\times 10^3$)
$Y_i = 0.5 * (\beta_W * W_i + \beta_X * X_i + U_i * \hat{\pi}_i)^2 + \epsilon_i/5$	
Hájek	0.00035(11.395)
ANCOVA ($Y \sim W + X$)	−0.00486(14.937)
ANCOVA ($Y \sim W * X$)	0.00713(7.984)
NP Regression ($Y \sim X$)	0.01596(2.448)
$Y_i = W_i + \beta_X * X_i + \hat{\pi}_i + U_i + \epsilon_i$	
Hájek	0.00262(13.161)
ANCOVA ($Y \sim W + X$)	−0.00130(1.925)
ANCOVA ($Y \sim W * X$)	−0.00120(1.935)
NP Regression ($Y \sim X$)	−0.00105(2.084)

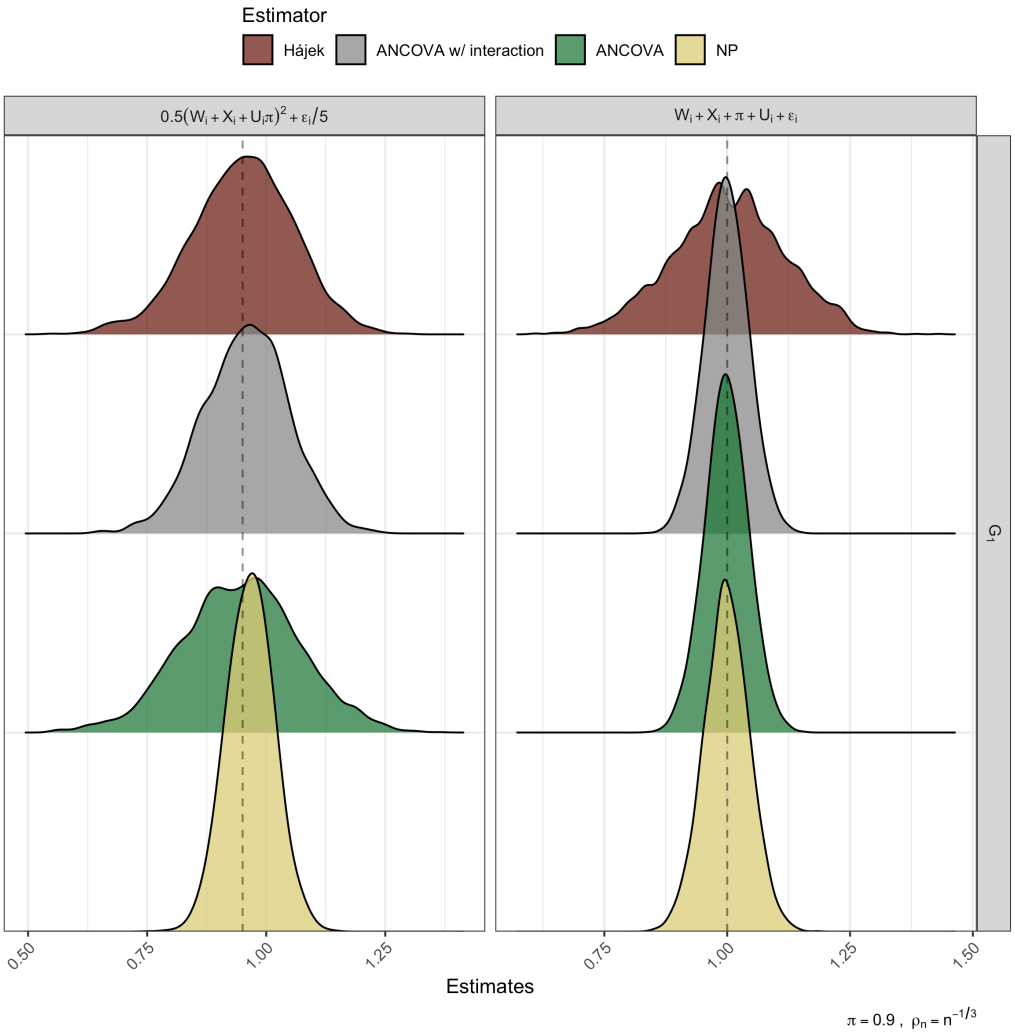


Figure 3.7: "Empirical" distributions of the Hájek estimator, ANCOVA estimator, and an AIPW style estimator using a non-parametric regression for the functions $\widehat{h}_n^{(w)}$ for graphon model G_1. Columns indicate potential outcomes functions.

2. Eigenvectors can help reduce variation due to the effect of the latent positions.

Including eigenvectors in the predictive models $\hat{h}_n^{(w)}$ might help reduce variance if the latent positions U_i influence the potential outcomes through f. The motivation behind this move is simple: by assumption, the potential outcomes functions f are functions of the treatment status W_i, the proportion of peers treated, the covariate X_i, and the latent position U_i. The latent positions are unfortunately unobservable, but we might be able to capture some of the variation due to the latent positions using the eigenvectors.

Figure 3.8 depicts the distributions of 3000 estimates using the Hájek estimator, the ANCOVA estimator $\hat{\beta}^{\text{ANCOVA2}}$, an ANCOVA estimator where 30 eigenvectors are included as predictors in the model, $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$ with $\hat{h}_n^{(w)}$ as above, and $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$ where three eigenvectors were included alongside X_i as predictors in the nonparametric regression models $\hat{h}_n^{(w)}$. We consider a scenario where $\pi = 0.9$, $\rho_n = n^{-1/7}$, and two graphon models, G_1 and G_4 (see Appendix 3.A.1 for details on the graphon models).

The panels on the left of Figure 3.8 show the results when the potential outcomes function is the quadratic model considered earlier with $\beta_W = 1$ and $\beta_X = 0.5$. As expected, the ANCOVA estimator is less efficient than the Hájek estimator, and including eigenvectors does not change that. The use of the more flexible nonparametric regression for the predictive models improves efficiency, and the further inclusion of the eigenvectors lead to a similar reduction (see Table 3.2(a) for numeric values).

The panels in the middle of Figure 3.8 show a scenario where the potential outcomes function is

$$f(w, p, x, u; \epsilon) = \tfrac{1}{5}(1+u)^2 \cdot (1 + w + 0.5 \cdot x) \cdot \exp(p) + \epsilon/5.$$

For this potential outcomes function, including eigenvectors lead to improvements for both the ANCOVA and nonparametric regression estimators. Interestingly, the nonparametric regression based estimator does not perform noticeably better than the ANCOVA estimator when only the covariate is used in the regression, but when the eigenvectors are included, the efficiency gains for the nonparametric regression based estimator are much better than those

obtained by the ANCOVA estimator. This is not surprising. In this particular potential outcomes function, the effect of the covariate is linear, and the ANCOVA estimator therefore is sufficient when trying to correct for variation in the covariate. The latent variable, however, influences the potential outcomes in a nonlinear fashion, and the nonparametric regression is therefore more equipped to capture any variation that can be attributed to the latent variable.

The panels on the right of Figure 3.8 show the results for a third potential outcomes function, namely

$$f(w, p, x, u; \epsilon) = \cos(3 \cdot w \cdot p + 0.5 \cdot x) + \epsilon/5.$$

Here, the potential outcomes do not depend on the latent positions directly. From the figure and Table 3.2 we see that the inclusion of the eigenvectors do not seem to negatively affect the efficiency of the estimators. This is comforting as in practice we would not know whether the latent positions would have any direct effect on the potential outcomes.

Table 3.2: Bias (Asymptotic Variance Estimate $\times 10^3$) of ANCOVA estimator, AIPW Estimator using non-parametric regression of Y on X for $\hat{h}_n^{(w)}$, and AIPW Estimator using non-parametric regression of Y on X and three eigenvectors for $\hat{h}_n^{(w)}$. Bias estimated as average deviation from $\bar{\tau}_{\text{DIR}}$ across 3000 simulations of sample size 1000. Variance calculated as the empirical variance of the estimates across 3000 simulations of sample size 1000. $\pi = 0.9$ and $\rho_n = n^{-1/7}$.

	Bias (Variance $\times 10^3$)	
	G_1	G_4
$Y_i = 0.5 * (\beta_W * W_i + \beta_X * X_i + U_i * \hat{\pi}_i)^2 + \epsilon_i/5$		
Hájek	−0.00061(2.001)	−0.00165(2.100)
ANCOVA ($Y \sim X \times W$)	0.00112(1.209)	0.00112(1.250)
ANCOVA ($Y \sim X \times W + \sum_{k=1}^{30} \widehat{\text{PC}}_k$)	0.00175(1.388)	−0.00236(1.327)
NP Regression($Y \sim X$)	0.00332(0.872)	0.00300(0.950)
NP Regression($Y \sim X + \widehat{\text{PC}}_1 + \widehat{\text{PC}}_2 + \widehat{\text{PC}}_3$)	0.00814(0.512)	0.00762(0.577)
$Y_i = \cos(3 * \beta_W * W_i * \hat{\pi}_i + \beta_X * X_i) + \epsilon_i/5$		

Hájek	0.00263(0.475)	0.00335(0.481)
ANCOVA ($Y \sim X \times W$)	0.00071(0.482)	0.00135(0.484)
ANCOVA ($Y \sim X \times W + \sum_{k=1}^{30} \widehat{\mathrm{PC}}_k$)	0.00087(0.497)	0.00074(0.491)
NP Regression($Y \sim X$)	−0.00096(0.304)	−0.00015(0.330)
NP Regression($Y \sim X + \widehat{\mathrm{PC}}_1 + \widehat{\mathrm{PC}}_2 + \widehat{\mathrm{PC}}_3$)	−0.00126(0.310)	−0.00047(0.335)
$Y_i = (1+U_i)^2 * (1 + W_i + \beta_X * X_i) * \exp(\hat{\pi}_i)/5 + \epsilon_i/5$		
Hájek	−0.00046(7.070)	−0.00364(7.849)
ANCOVA ($Y \sim X \times W$)	−0.00126(3.355)	−0.00236(3.884)
ANCOVA ($Y \sim X \times W + \sum_{k=1}^{30} \widehat{\mathrm{PC}}_k$)	−0.00015(2.359)	−0.00694(2.274)
NP Regression($Y \sim X$)	−0.00163(3.500)	−0.00277(4.108)
NP Regression($Y \sim X + \widehat{\mathrm{PC}}_1 + \widehat{\mathrm{PC}}_2 + \widehat{\mathrm{PC}}_3$)	−0.00189(0.880)	−0.00122(1.318)

3. The observed proportion of peers treated may help finite sample efficiency.

Including the observed proportion of peers treated M_i/N_i as a predictor in the predictive models $\hat{h}_n^{(w)}$ might lead to finite sample efficiency gains, though we do not expect to see any reduction in the asymptotic variance. From Theorem 3.1, we note that the asymptotic variance does not depend on M_i/N_i, and asymptotically $M_i/N_i = \pi$ with probability 1.

Figure 3.9 shows a scenario where including the proportion of peers treated as a covariate leads to some variance reduction. We consider the linear outcomes function along side the potential outcomes function

$$f(w, p, x, u; \epsilon) = (1 + w)\, e^p + \cos(x) + \epsilon/5 \tag{3.16}$$

and graphon models G_5 and G_6 (see Appendix 3.A.1). We consider four estimators: $\hat{\beta}^{\mathrm{ANCOVA2}}$ as already discussed in detail, an ANCOVA estimator where the observed proportion of peers who are treated is included as an additional covariate along side the interaction term, $\hat{\tau}^{\mathrm{HA}}_{\mathrm{AIPW}}$ with $\hat{h}_n^{(w)}$ being a non-parametric regression of the observed outcomes Y_i on X_i, and finally $\hat{\tau}^{\mathrm{HA}}_{\mathrm{AIPW}}$ where the observed proportion of peers treated

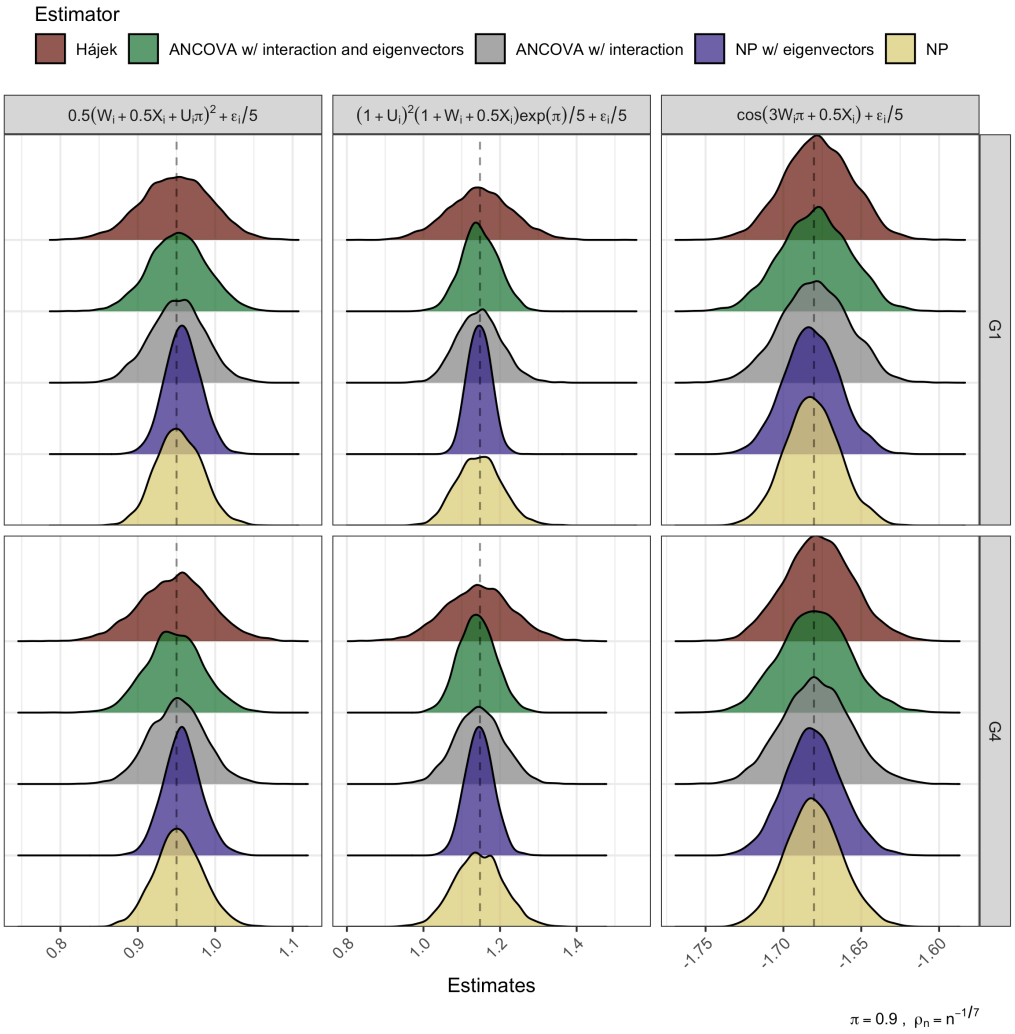


Figure 3.8: "Empirical" distributions of the ANCOVA estimator, and two AIPW style estimator using non-parametric regression for the functions $\widehat{h}_n^{(w)}$ for graphon model G_1 and G_4. ANCOVA w/ interaction and eigenvectors includes 30 eigen vectors. AIPW using non-parametric regresion with eigenvectors include three eigenvectors of the observed adjacency matrix. Columns indicate potential outcomes functions.

is included in addition to X_i as a predictor in the non-parametric regression model. Figure 3.9 illustrates how the effect of including $\hat{\pi}_i$ as a covariate highly depends on the potential outcomes function. For the linear potential outcomes function, the benefit is small. However, for the model in Equation 3.16, there is significant variance reduction in going from the ANCOVA estimator to the AIPW style estimator, and an even greater reduction when including the proportion of peers treated as a predictor. There is also some degree of variance reduction when the observed proportion of peers who are treated is included in the ANCOVA model, though it is not as significant. Table 3.3 give bias and variance of the scenarios shown in Figure 3.9.

Table 3.3: Bias (Asymptotic Variance Estimate $\times 10^3$) of ANCOVA estimator, AIPW Estimator using non-parametric regression of Y on X for $\hat{h}_n^{(w)}$, and AIPW Estimator using non-parametric regression of Y on X and $\hat{\pi}_i$ for $\hat{h}_n^{(w)}$. Bias estimated as average deviation from $\bar{\tau}_{\text{DIR}}$ across 3000 simulations of sample size 1000. Variance calculated as the empirical variance of the estimates across 3000 simulations of sample size 1000. $\pi = 0.3$ and $\rho_n = n^{-1/3}$.

	Bias (Variance $\times 10^3$)	
	G_5	G_6
$Y_i = (1 + \beta_W * W_i) * \exp(\hat{\pi}_i) + \cos(\beta_X * X_i) + \epsilon_i/5$		
ANCOVA ($Y \sim X \times W$)	0.00632(1.932)	0.01109(2.276)
ANCOVA ($Y \sim X \times W + \hat{\pi}$)	0.00835(1.446)	0.01270(1.532)
NP Regression ($Y \sim X$)	0.00608(1.005)	0.01166(1.348)
NP Regression ($Y \sim X + \hat{\pi}$)	0.00990(0.526)	0.01450(0.586)
$Y_i = W_i + \beta_X * X_i + \hat{\pi}_i + U_i + \epsilon_i$		
ANCOVA ($Y \sim X \times W$)	−0.00037(0.901)	−0.00060(1.002)
ANCOVA ($Y \sim X \times W + \hat{\pi}$)	0.00081(0.813)	0.00039(0.849)
NP Regression ($Y \sim X$)	−0.00045(0.928)	−0.00060(1.022)
NP Regression ($Y \sim X + \hat{\pi}$)	−0.00158(0.845)	0.00205(0.858)

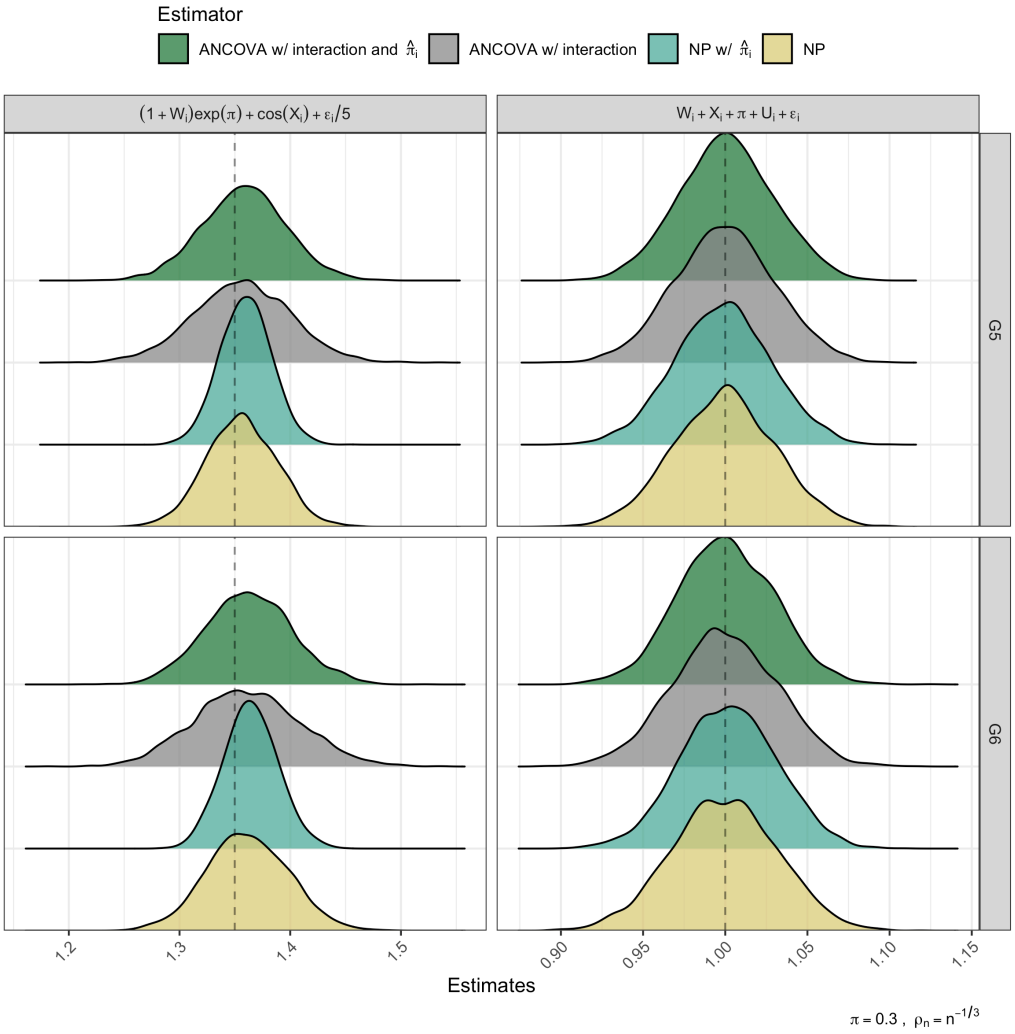


Figure 3.9: "Empirical" distributions of the ANCOVA estimator, and two AIPW style estimator using non-parametric regression for the functions $\widehat{h}_n^{(w)}$ for graphon model G_1. One AIPW only include the covariate X_i, while the other also include the observed proportion of peers who are treated as a predictor. Columns indicate potential outcomes functions.

3.4.4 Does Sample Splitting Help?

Emmenegger et al. (2023) suggests the use of sample splitting when estimating the AIPW estimator. Sample splitting typically relies on the independence of experimental units to create two separate data sets, one used to fit your predictive model, and another used to estimate the ATE. When interference is present, this is not generally possible, as the outcomes can be arbitrarily linked. However, Emmenegger et al. (2023) suggest we avoid this problem by fitting the nuisance functions needed for prediction on the complement of each split; see Algorithm 1 of Emmenegger et al. (2023) for details.

We wanted to assess the benefits of this flavor of sample splitting in a simulation setup similar to that in Section 3.4.3. However, due to the sample splitting method, the estimation procedure suggested by Emmenegger et al. (2023) was often not feasible. For more dense graphons with a more extreme propensity, the chances of having the complement of a given split be null is rather large. Therefore, we narrowed the scope of the comparison of the AIPW estimator as we have presented it here, and the estimation procedure suggested by Emmenegger et al. (2023). We consider two graphons (G_3 and G_6; see Appendix 3.A.1 for details), three potential outcomes functions (models 1, 2, and 3; see Appendix 3.A.2 for details), four values of β_X (0.5, 1, 2, and 5), three propensities (0.3, 0.5, 0.9), and $\rho_n = n^{-1/3}$. However, for G_3 with $\pi = 0.9$, we often have a null set for the complement, and as such cannot consistently estimate the direct effect using the sample splitting algorithm.

As we see from Table 3.4, for the graphon G_3, using sample splitting leads to very similar biases and asymptotic variances as if sample splitting was not used when estimating the direct effect.

Table 3.5 shows similar results for the graphon G_6. However, for a more extreme propensity of 0.9, sample splitting seem to introduce some finite sample bias and result in larger variance. Figure 3.10 illustrates this across different potential outcomes functions, and for different values of β_X.

The sample splitting algorithm presented by Emmenegger et al. (2023) does not seem to

improve on the AIPW estimator, introduces finite sample bias and variance inflation in certain settings, and at times might not be a feasible approach at all. However, the sample splitting algorithm is not without merit, the biggest being the ability to estimate the variance of the estimator, something that is not possible in general.

Table 3.4: Bias and variance comparison for AIPW estimator. We compare using sample splitting and not. Overall, bias and variance are similar for the G_3 graphon when the propensity is either 0.3 or 0.5 across various values of β_X.

			Bias (Variance)	
β_X	π	n	AIPW	AIPW w/ SS algorithm
$Y_i = 0.5 * (\beta_W * W_i + \beta_X * X_i + U_i * \hat{\pi}_i)^2 + \epsilon_i/5$				
0.5	0.3	3000	-0.0016 (0.39)	-0.0095 (0.59)
0.5	0.5	3000	0.0002 (0.48)	-0.0085 (0.49)
1.0	0.3	3000	-0.0045 (1.23)	-0.0116 (2.04)
1.0	0.5	3000	0.0004 (1.29)	-0.0118 (1.30)
2.0	0.3	3000	-0.0156 (4.92)	-0.0126 (8.04)
2.0	0.5	3000	-0.0018 (4.79)	-0.0161 (4.84)
5.0	0.3	3000	-0.0975 (40.29)	-0.0118 (58.77)
5.0	0.5	3000	0.0003 (38.69)	-0.0160 (36.60)
$Y_i = \cos(3 * \beta_W * W_i * \hat{\pi}_i + \beta_X * X_i) + \epsilon_i/5$				
0.5	0.3	3000	-0.0123 (1.21)	-0.0100 (1.27)
0.5	0.5	3000	-0.0021 (2.07)	0.0039 (2.06)
1.0	0.3	3000	-0.0086 (0.87)	-0.0086 (1.06)
1.0	0.5	3000	-0.0010 (1.52)	0.0000 (1.52)
2.0	0.3	3000	-0.0025 (0.50)	-0.0020 (0.76)
2.0	0.5	3000	-0.0000 (1.04)	-0.0016 (1.05)
5.0	0.3	3000	0.0000 (0.50)	0.0002 (0.88)
5.0	0.5	3000	-0.0000 (1.00)	0.0006 (1.05)
$Y_i = -\exp(U_i) * (1 + W_i) * \cos(3 * W_i * \hat{\pi}_i) + \beta_X * \cos(X_i) + \epsilon_i/5$				
0.5	0.3	3000	0.0479 (15.79)	0.0494 (15.66)
0.5	0.5	3000	0.0045 (27.18)	-0.0032 (26.96)
1.0	0.3	3000	0.0474 (16.01)	0.0492 (15.90)
1.0	0.5	3000	-0.0007 (27.96)	-0.0125 (27.69)
2.0	0.3	3000	0.0484 (15.26)	0.0518 (15.05)
2.0	0.5	3000	0.0046 (28.59)	-0.0137 (28.06)
5.0	0.3	3000	0.0514 (15.43)	0.0551 (15.06)
5.0	0.5	3000	0.0011 (27.16)	-0.0298 (26.37)

Table 3.5: Bias and variance comparison for AIPW estimator. We compare using sample splitting and not. Overall, bias and variance are similar for the G_6 graphon for various values of π and β_X, but with some cases of sample splitting hurting overall performance.

			Bias (Variance)	
β_X	π	n	AIPW	AIPW w/ SS algorithm
$Y_i = 0.5 * (\beta_W * W_i + \beta_X * X_i + U_i * \hat{\pi}_i)^2 + \epsilon_i/5$				
0.5	0.3	3000	-0.0012 (0.45)	-0.0101 (0.64)
0.5	0.5	3000	0.0002 (0.56)	-0.0097 (0.57)
0.5	0.9	3000	0.0038 (0.96)	-0.0119 (1.97)
1.0	0.3	3000	-0.0044 (1.23)	-0.0137 (1.95)
1.0	0.5	3000	-0.0005 (1.38)	-0.0154 (1.36)
1.0	0.9	3000	0.0161 (2.53)	-0.0148 (7.81)
2.0	0.3	3000	-0.0158 (4.89)	-0.0171 (7.48)
2.0	0.5	3000	-0.0001 (5.01)	-0.0190 (4.85)
2.0	0.9	3000	0.0692 (10.67)	-0.0201 (31.07)
5.0	0.3	3000	-0.0964 (42.68)	-0.0149 (43.79)
5.0	0.5	3000	0.0057 (38.01)	-0.0175 (29.37)
5.0	0.9	3000	0.4244 (147.18)	-0.0392 (199.70)
$Y_i = \cos(3 * \beta_W * W_i * \hat{\pi}_i + \beta_X * X_i) + \epsilon_i/5$				
0.5	0.3	3000	-0.0392 (1.55)	-0.0371 (1.61)
0.5	0.5	3000	-0.0053 (2.58)	0.0014 (2.54)
0.5	0.9	3000	0.0225 (0.40)	0.0405 (1.22)
1.0	0.3	3000	-0.0260 (1.06)	-0.0272 (1.25)
1.0	0.5	3000	-0.0026 (1.66)	-0.0025 (1.65)
1.0	0.9	3000	0.0119 (0.97)	0.0200 (6.05)
2.0	0.3	3000	-0.0072 (0.66)	-0.0067 (0.90)
2.0	0.5	3000	-0.0004 (1.04)	-0.0016 (1.04)
2.0	0.9	3000	0.0108 (1.90)	0.0008 (13.94)
5.0	0.3	3000	-0.0004 (0.67)	-0.0005 (0.83)
5.0	0.5	3000	-0.0006 (1.09)	0.0002 (1.10)
5.0	0.9	3000	0.0012 (2.22)	0.0007 (12.76)
$Y_i = -\exp(U_i) * (1 + W_i) * \cos(3 * W_i * \hat{\pi}_i) + \beta_X * \cos(X_i) + \epsilon_i/5$				
0.5	0.3	3000	0.1333 (17.64)	0.1339 (17.48)
0.5	0.5	3000	0.0150 (32.64)	0.0067 (32.26)
0.5	0.9	3000	-0.0917 (4.84)	-0.1134 (5.04)
1.0	0.3	3000	0.1299 (17.70)	0.1307 (17.48)
1.0	0.5	3000	0.0109 (30.76)	-0.0015 (30.27)
1.0	0.9	3000	-0.0922 (5.03)	-0.1266 (5.37)
2.0	0.3	3000	0.1357 (18.62)	0.1367 (18.29)
2.0	0.5	3000	0.0099 (32.16)	-0.0082 (31.36)
2.0	0.9	3000	-0.0994 (4.81)	-0.1432 (5.22)
5.0	0.3	3000	0.1359 (17.29)	0.1354 (16.74)
5.0	0.5	3000	0.0134 (30.92)	-0.0143 (29.79)
5.0	0.9	3000	-0.1176 (5.29)	-0.1769 (5.65)

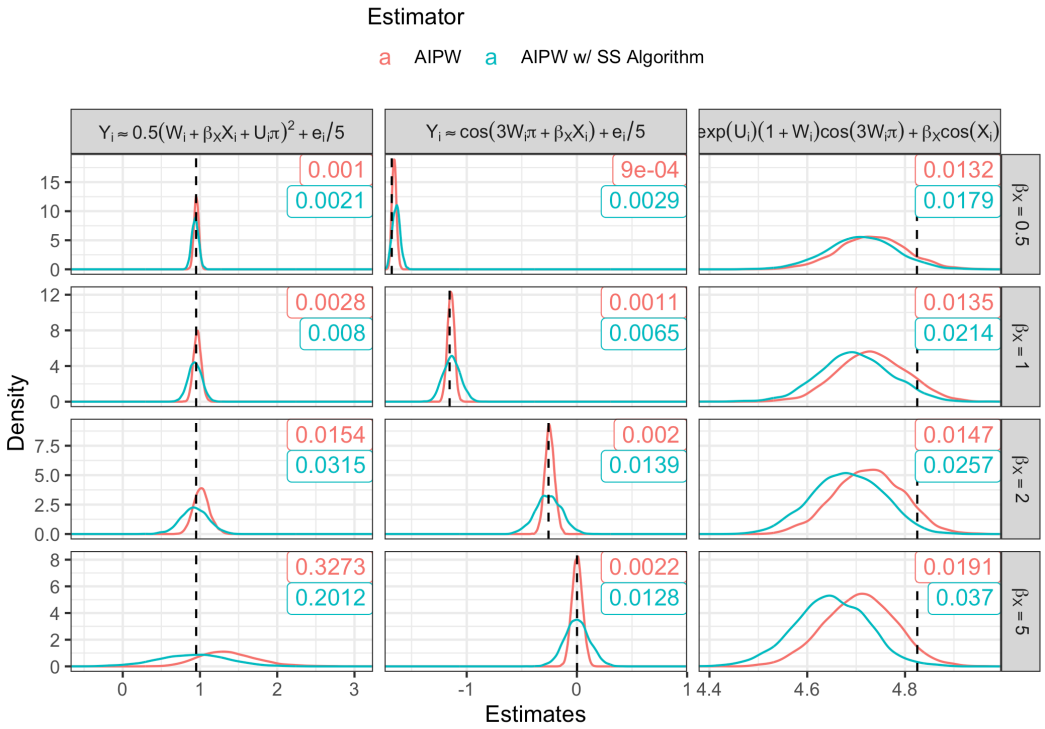


Figure 3.10: Density curves for estimates based on the AIPW estimator evaluating the use of sample-splitting. Here, we consider the graphon G_6 with propensity $\pi = 0.9$ across different potential outcomes functions and for various values of β_X. $\rho_n = n^{-1/3}$.

3.5 Data Application

We consider the field experiment presented by Cai et al. (2015b). This experiment was designed as a randomized experiment based on the introduction of a new weather insurance policy specifically aimed at rice farmers in China. During the roll out of this new policy, two rounds of information sessions were held across 185 villages in rural China. The households were randomized to attend either a simple 20 minute long session, or an intensive 45 minute long session, and to either attend the first or second round of sessions. This randomization scheme allows Cai et al. (2015b) to estimate both direct effects and indirect treatment effects due to knowledge sharing across social bonds. For our purposes, we simplify the treatment to a binary "intensive session" status indicator, and discard the timing of the session. The data is freely available; see Cai et al. (2015a).

To prepare the data for analysis, we construct the adjacency matrix from the network data given following Chin (2019b). We did see duplicated network edges in the data, which results in a non-binary adjacency matrix. We fixed the adjacency matrix to be binary, which results in slightly different estimates using the Hájek estimator than what is presented by Chin (2019b).

It is important to note that these data do not satisfy the assumptions underlying our theoretical results. In particular, since the outcome considered here is binary, i.e. whether or not participants signed up for the insurance policy, the smoothness assumption in Assumption 3.7 is violated. However, due to a lack of available data from randomized experiments on networks, our options are limited, and therefore we proceed with the analysis.

The estimated direct effects are presented in Table 3.6. We present estimates using a number of estimators. First, the Hájek and Horvitz-Thompson estimators. Second, two ANCOVA models were fit to the data. For both, we included only the covariates that seemed most correlated with the outcome; these were educational level, risk aversion (on a scale from 0 to 1), an indicator for whether any disaster had happened in the previous year, loss in yield due to distasters last year (as a percentage), and area of rice production. The first

ANCOVA model only included the covariates and interactions with the treatment indicator, while the second also added the observed proportion of peers treated as an extra covariate.

Finally, estimates using the AIPW estimator with a non-parametric regression model are presented. The same set of covariates were included as for the ANCOVA models. Again, we include the results of this estimator both with and without the observed proportion of peers treated as a covariate. We include the result of the sample-splitting approach suggested by Emmenegger et al. (2023).

We also present a variety of standard error estimates. For the Hájek estimator, we present the naive variance estimator, that is the sum of the variance of treated outcomes and the variance of control outcomes. For all but the sample splitting approach, we include naive bootstrap standard errors, i.e. a bootstrap is performed as if the data were independent. For the sample splitting approach, the standard error estimator presented by Emmenegger et al. (2023) is used. It is important to note that the standard error estimate based on Emmenegger et al. (2023) is the only one of these that have some theoretical guarantees; as noted in Section 3.4.1, the OLS and naive bootstrap standard error estimates are most likely biased, and we have no way of knowing in which direction.

We note that all point estimates are of similar magnitudes. This is unsurprising; all of these methods offer unbiased point estimates.

For the standard error estimates, we note that the naive estimate of the standard error for the Hájek estimator is the smallest, the OLS based estimates the largest, and the sample splitting based estimator somewhere in between. The bootstrap standard error estimates for the Hájek and ANCOVA estimators are very similar to the naive and OLS standard error estimates, respectively. For the AIPW estimator, the naive bootstrap standard error estimate is similar to the Emmenegger standard error, but is slightly smaller.

Table 3.6: Estimates of the Direct Effect using the various estimators considered previously. Here, "w/ ppt" indicates the use of proportion of peers treated as a covariate. "using SS" indicates the AIPW estimate was obtained using the sample splitting procedure proposed by Emmenegger et al., 2023.

Estimator	Direct Effect Estimate	SE Estimator	SE
Horvitz-Thompson	0.07262	Naive Bootstrap	0.01890
Hájek	0.07335	Naive	0.01416
Hájek	0.07335	Naive Bootstrap	0.01440
ANCOVA	0.06237	Naive Bootstrap	0.04256
ANCOVA	0.06237	OLS	0.04316
ANCOVA w/ ppt	0.05655	Naive Bootstrap	0.04247
ANCOVA w/ ppt	0.05655	OLS	0.04321
AIPW	0.07154	Naive Bootstrap	0.02077
AIPW w/ ppt	0.07117	Naive Bootstrap	0.02149
AIPW using SS	0.06903	Emmennegger	0.02454
AIPW w/ ppt using SS	0.06748	Emmennegger	0.02452

3.6 Conclusions

The asymptotic behavior of estimators of the Average Treatment Effect when general interference is allowed is still largely unknown. Often, interference will inflate the variance of estimators, and therefore variance reduction techniques are of interest. We show how the random graph framework proposed by Li and Wager (2022) can be used to get asymptotic results for the Augmented Inverse Probability Estimator, and use this to evaluate the efficiency of one of the most popular covariate adjustment techniques, the ANCOVA, when analyzing randomized experiments. We illustrate the use of the studied estimator by analyzing the Cai et al. (2015b) study.

Our main conclusions are two fold. First, ANCOVA still provides a consistent and asymptotically normal estimator for the ATE, even when misspecified. Furthermore, as long as interactions between the treatment indicator and covariates are included in the working model, the ANCOVA estimator is at least as efficient as the Hájek estimator. This implies that using the ANCOVA method to adjust for covariates in randomized experiments is sensible, even when interference is present. From a practical perspective, this is encouraging.

Second, we see that estimating the standard error of the AIPW estimator in general, and the ANCOVA estimator in particular, is still an unsolved problem. We show through simulations that both the OLS and the bootstrap standard error estimators perform poorly, but more importantly, inconsistently. This means that inference based on the AIPW or the ANCOVA estimator is, as of now, not possible in practice.

While the direct practical implications of our work are mixed, we are able to exactly characterize the asymptotic behavior of widely used estimators in a general interference setting. This can help up better understand the challenges still to be solved when the goal is to obtain statistical inference from randomized experiments where experimental units interfere with each other.

Appendices

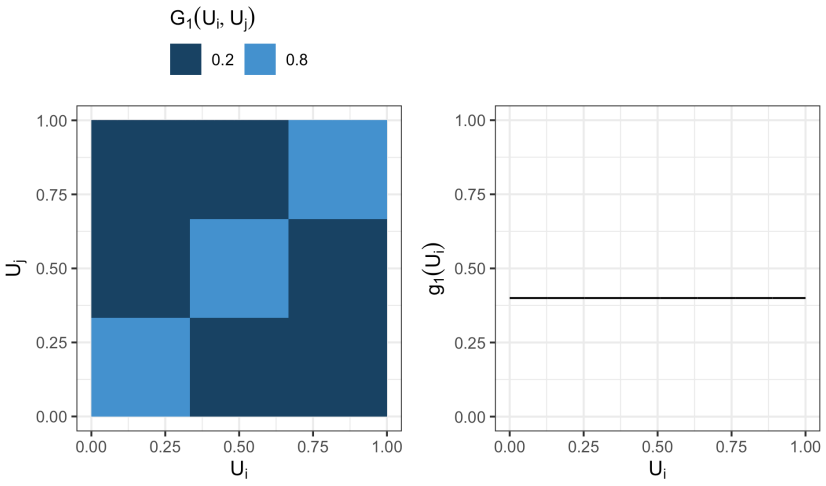
3.A Simulation Details

3.A.1 Graphons

The six graphons considered are included below. While we have not considered results for all of the graphons, they are all included here for future reference. The expressions for $G(U_i, U_j)$ and $g(U_i) = \mathbb{E}\left[G(U_i, U_j)|U_i\right]$ are given, and both are presented visually. The former is presented as a heat map, the latter as a plot of $g(u)$ as a function of u. The relationship between the two: $g(u)$ is the average of all values along the vertical line $U_i = u$ on the heatmap.

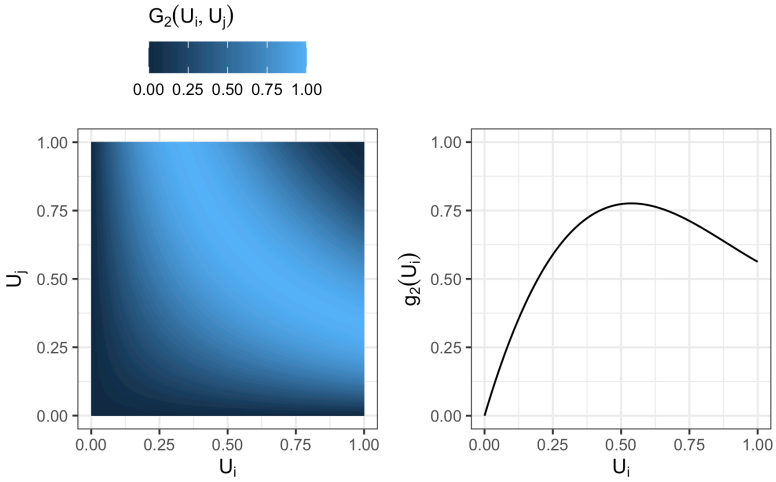
Graphon 1:

$$\begin{aligned} G_1(U_i, U_j) &= \frac{1}{5} + \\ &\quad \frac{3}{5}\left(\mathbb{1}\left[U_i < \tfrac{1}{3}\right] \cdot \mathbb{1}\left[U_j < \tfrac{1}{3}\right] + \right. \\ &\quad\quad \mathbb{1}\left[\tfrac{1}{3} \leq U_i \leq \tfrac{2}{3}\right] \cdot \mathbb{1}\left[\tfrac{1}{3} \leq U_j \leq \tfrac{2}{3}\right] + \\ &\quad\quad \left. \mathbb{1}\left[U_i > \tfrac{2}{3}\right] \cdot \mathbb{1}\left[U_j > \tfrac{2}{3}\right]\right) \\ g_1(U_i) &= \mathbb{E}\left[G(U_i, U_j)|U_i\right] \\ &= \frac{2}{5} \end{aligned}$$



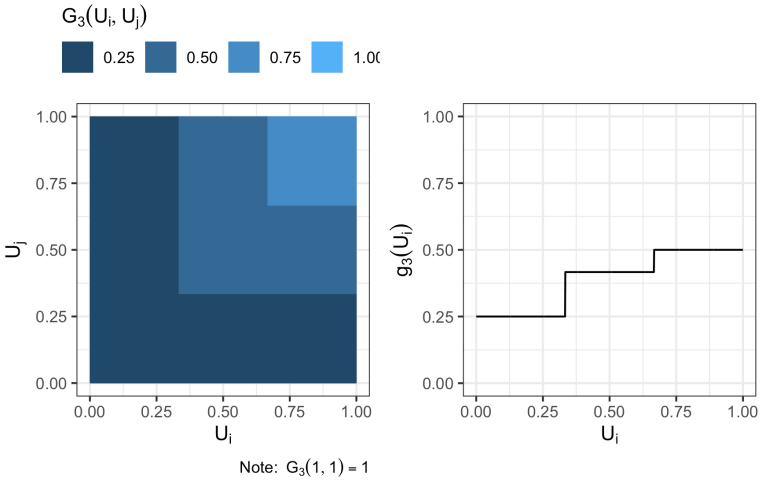
Graphon 2:

$$G_2(U_i, U_j) = \frac{27}{4}\left(U_i \cdot U_j - 2 \cdot U_i^2 \cdot U_j^2 + U_i^3 \cdot U_j^3\right)$$
$$g_2(U_i) = \frac{27}{4}\left(\frac{U_i}{2} - 2 \cdot \frac{U_i^2}{3} + \frac{U_i^3}{4}\right)$$



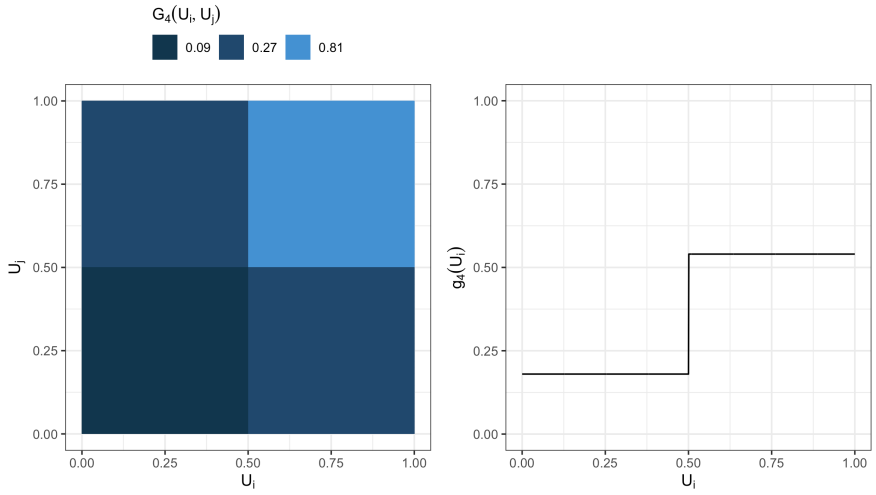
Graphon 3:

$$G_3(U_i, U_j) = \frac{1}{4} + \frac{\lfloor 3 \cdot \min(U_i, U_j) \rfloor}{4}$$
$$g_3(U_i) = \frac{1}{4} + \frac{1}{6}\mathbb{1}\left[U_i \geq \tfrac{1}{3}\right] \cdot \mathbb{1}\left[U_i < \tfrac{2}{3}\right] + \frac{1}{4}\mathbb{1}\left[U_i \geq \tfrac{2}{3}\right]$$



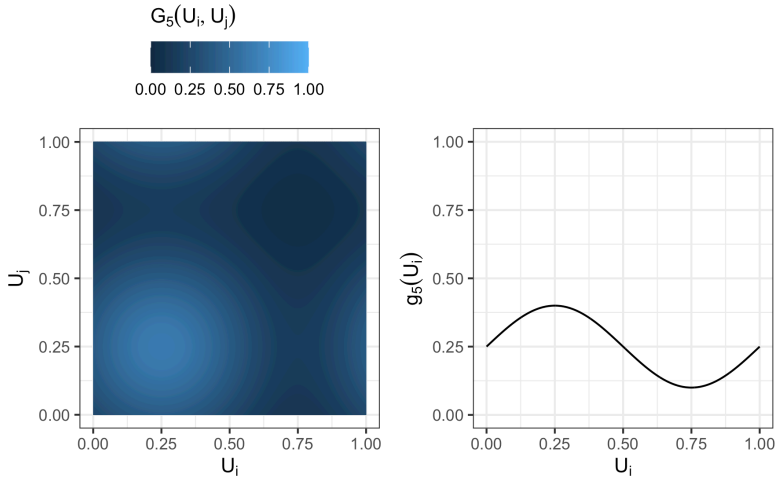
Graphon 4:

$$G_4(U_i, U_j) = \left(\frac{3}{10} + \frac{3}{5}\mathbb{1}\left[U_i > \tfrac{1}{2}\right]\right) \cdot \left(\frac{3}{10} + \frac{3}{5}\mathbb{1}\left[U_j > \tfrac{1}{2}\right]\right)$$
$$g_4(U_i) = \left(\frac{3}{10} + \frac{3}{5}\mathbb{1}\left[U_i > \tfrac{1}{2}\right]\right) \cdot \frac{3}{5}$$



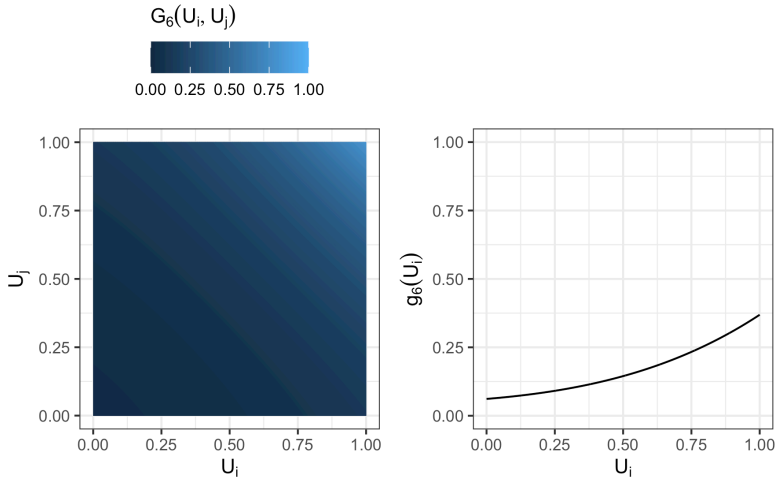
Graphon 5:

$$G_5(U_i, U_j) = \left(\frac{3}{10} \cdot \sin\left(2 \cdot \pi \cdot U_i\right) + \frac{1}{2}\right) \cdot \left(\frac{3}{10} \cdot \sin\left(2 \cdot \pi \cdot U_j\right) + \frac{1}{2}\right)$$
$$g_5(U_i) = \left(\frac{3}{20} \cdot \sin\left(2 \cdot \pi \cdot U_i\right) + \frac{1}{4}\right)$$



Graphon 6:

$$G_6(U_i, U_j) = \left(\frac{1}{20} \cdot (U_i + 1)^4 + \frac{1}{10}\right)\left(\frac{1}{20} \cdot (U_j + 1)^4 + \frac{1}{10}\right)$$
$$g_6(U_i) = \left(\frac{1}{20} \cdot (U_i + 1)^4 + \frac{1}{10}\right)\frac{41}{100}$$



3.A.2 Potential Outcomes Functions

These are the potential outcomes functions we considered. Again, we did not consider result for all functions here, but these are all included for future reference.

Potential Outcomes Function 1:

$$f(w,p,x,u;\epsilon) = \frac{1}{2}\left(w + \beta_X \cdot x + u \cdot p\right)^2 + \frac{\epsilon_i}{5}$$
$$f'(w,p,x,u;\epsilon) = u\left(w + \beta_X \cdot x + u \cdot p\right)$$

Potential Outcomes Function 2:

$$f(w,p,x,u;\epsilon) = \cos\left(3 \cdot w \cdot p + \beta_X \cdot x\right) + \frac{\epsilon_i}{5}$$
$$f'(w,p,x,u;\epsilon) = -3 \cdot w \cdot \sin\left(3 \cdot w \cdot p + \beta_X \cdot x\right)$$

Potential Outcomes Function 3:

$$f(w,p,x,u;\epsilon) = -\exp(u) \cdot (1+w) \cdot \cos(3 \cdot w \cdot p) + \beta_X * \cos(x) + \frac{\epsilon_i}{5}$$
$$f'(w,p,x,u;\epsilon) = 3 \cdot w \cdot \exp(u) \cdot (1+w) \cdot \cos(3 \cdot w \cdot p)$$

Potential Outcomes Function 4:

$$f(w,p,x,u;\epsilon) = (1+w) \cdot \exp(p) + cos(\beta_X \cdot x) + \frac{\epsilon_i}{5}$$
$$f'(w,p,x,u;\epsilon) = (1+w) \cdot \exp(p)$$

Potential Outcomes Function 5:

$$f(w,p,x,u;\epsilon) = \frac{1}{5}(1+u)^2 \cdot (1 + w + \beta_X \cdot x) \cdot \exp(p) + \frac{\epsilon_i}{5}$$
$$f'(w,p,x,u;\epsilon) = \frac{1}{5}(1+u)^2 \cdot (1 + w + \beta_X \cdot x) \cdot \exp(p)$$

Potential Outcomes Function 6:

$$f(w,p,x,u;\epsilon) = w + \beta_X \cdot x + p + u + \epsilon_i$$
$$f'(w,p,x,u;\epsilon) = 1$$

Potential Outcomes Function 7:

$$f(w,p,x,u;\epsilon) = w + \beta_X \cdot x + p^2 + u + \epsilon_i$$
$$f'(w,p,x,u;\epsilon) = 2 \cdot p$$

3.B Additional Results Needed

Proposition 3.4. *If Equation 3.6 and Equation 3.7 hold, then*

$$\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\hat{\pi}}\right)\left(\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(1)}\right)=o_p(1/\sqrt{n}) \tag{3.17}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\hat{\pi}}{1-\hat{\pi}}\left(\hat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(0)}\right)=o_p(1/\sqrt{n}). \tag{3.18}$$

Proof. We will show that Equation 3.17 holds. Let $\epsilon > 0$ and $\eta > 0$ be given. We want to show that there exists an integer $N \in \mathbb{N}$ such that for all $n > N$,

$$P\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{W_i}{\hat{\pi}}\right)\left(\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(1)}\right)\right|<\epsilon\right)\geq 1-\eta.$$

Let $g_n : (0,1) \mapsto \mathbb{R}$ be the function

$$g_n(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{W_i}{x}\right)\left(\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(1)}\right).$$

Since g_n is continuous in a neighborhood around π, there exists some $\delta > 0$ such that $|g_n(x) - g_n(\pi)| < \epsilon/2$ for all x with $|x - \pi| < \delta$.

Since $\hat{\pi} \to_p \pi$, there is some $N_1 \in \mathbb{N}$ such that for all $n \geq N_1$, $P(|\hat{\pi} - \pi| < \delta) \geq 1 - \eta/2$.

Since, by assumption, $g_n(\pi) = o_p(1)$, there also exists some $N_2 \in \mathbb{N}$ such that for all $n \geq N_2$, $P(|g_n(\pi)| < \epsilon/2) \geq 1 - \eta/2$.

Now, let $N = \max\{N_1, N_2\}$. Then for any $n > N$,

$$\begin{aligned}
P(|g_n(\hat{\pi})|<\epsilon) &\geq P(|g_n(\hat{\pi})-g_n(\pi)|+|g_n(\pi)|<\epsilon)\\
&\geq P(\{g_n(\hat{\pi})-g_n(\pi)|<\epsilon/2\}\cap\{|g_n(\pi)|<\epsilon/2\})\\
&=1-P(\{g_n(\hat{\pi})-g_n(\pi)|\geq\epsilon/2\}\cup\{|g_n(\pi)|\geq\epsilon/2\})\\
&\geq 1-[P(g_n(\hat{\pi})-g_n(\pi)|\geq\epsilon/2)+P(|g_n(\pi)|\geq\epsilon/2)]\\
&\geq 1-(\eta/2+\eta/2)\\
&=1-\eta.
\end{aligned}$$

Therefore, $g_n(x) = o_p(1)$, and so Equation 3.17 holds.

□

Proposition 3.5. *Under the assumptions and conditions stated in Proposition 3.4,*

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(W_i-\hat{\pi})\left\{\frac{\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}}{\hat{\pi}}+\frac{\hat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})}}{1-\hat{\pi}}\right\}\\
&\quad=\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)\left\{\frac{\widetilde{Y}_i^{(1)}-\mathbb{E}\left[\widetilde{Y}_i^{(1)}\right]}{\pi}+\frac{\widetilde{Y}_i^{(0)}-\mathbb{E}\left[\widetilde{Y}_i^{(0)}\right]}{1-\pi}\right\}\\
&\qquad+o_p(1/\sqrt{n}).
\end{aligned}$$

Proof. We will show that

$$\frac{1}{n}\sum_{i=1}^{n}(W_i-\hat{\pi})\frac{\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}}{\hat{\pi}}=\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)\frac{\widetilde{Y}_i^{(1)}-\mathbb{E}\left[\widetilde{Y}_i^{(1)}\right]}{\pi}+o_p(1/\sqrt{n})$$

and

$$\frac{1}{n}\sum_{i=1}^{n}(W_i-\hat{\pi})\frac{\hat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})}}{1-\hat{\pi}}=\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)\frac{\widetilde{Y}_i^{(0)}-\mathbb{E}\left[\widetilde{Y}_i^{(0)}\right]}{1-\pi}+o_p(1/\sqrt{n}).$$

First, note that

$$\frac{1}{n}\sum_{i=1}^{n}(W_i-\hat{\pi})\frac{\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}}{\hat{\pi}}=\frac{1}{n}\sum_{i=1}^{n}(W_i-\hat{\pi})\frac{\widetilde{Y}_i^{(1)}}{\hat{\pi}}+\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\hat{\pi}}\right)\left(\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(1)}\right).$$

The second term in the above decomposition is $o_p(1/\sqrt{n})$ by Proposition 3.4.

For the first term,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(W_i-\hat{\pi})\frac{\widetilde{Y}_i^{(1)}}{\hat{\pi}}\\
&\quad=\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)\frac{\widetilde{Y}_i^{(1)}}{\hat{\pi}}-\frac{\hat{\pi}-\pi}{\hat{\pi}}\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^{(1)}.
\end{aligned}$$

Since $\{\widetilde{Y}_i^{(1)}\}_{i=1}^n$ is a sequence of i.i.d. random variables, $\frac{1}{n}\sum_{i=1}^n\widetilde{Y}_i^{(1)}=\mathbb{E}\left[\widetilde{Y}_i^{(1)}\right]+O_p(1/\sqrt{n})$. Since $\frac{\hat{\pi}-\pi}{\hat{\pi}}=o_p(1)$, we have

$$\begin{aligned}
&\frac{\hat{\pi}-\pi}{\hat{\pi}}\frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i^{(1)}=\\
&\quad=\frac{\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)}{\hat{\pi}}\mathbb{E}\left[\widetilde{Y}_i^{(1)}\right]+\frac{\hat{\pi}-\pi}{\hat{\pi}}O_p(1/\sqrt{n})\\
&\quad=\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)\frac{\mathbb{E}\left[\widetilde{Y}_i^{(1)}\right]}{\hat{\pi}}+o_p(1/\sqrt{n}),
\end{aligned}$$

and therefore

$$\frac{1}{n}\sum_{i=1}^{n}(W_i-\hat{\pi})\frac{\widetilde{Y}_i^{(1)}}{\hat{\pi}}=\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)\frac{\widetilde{Y}_i^{(1)}-\mathbb{E}\left[\widetilde{Y}_i^{(1)}\right]}{\hat{\pi}}+o_p(1/\sqrt{n}).$$

Continuing with the sum on the right,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)\frac{\widetilde{Y}_i^{(1)}-\mathbb{E}\left[\widetilde{Y}_i^{(1)}\right]}{\hat{\pi}} \\
&\quad=\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)\frac{\widetilde{Y}_i^{(1)}-\mathbb{E}\left[\widetilde{Y}_i^{(1)}\right]}{\pi} \\
&\qquad+\left(\frac{1}{\hat{\pi}}-\frac{1}{\pi}\right)\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)\left(\widetilde{Y}_i^{(1)}-\mathbb{E}\left[\widetilde{Y}_i^{(1)}\right]\right).
\end{aligned}$$

Since $\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)\left(\widetilde{Y}_i^{(1)}-\mathbb{E}\left[\widetilde{Y}_i^{(1)}\right]\right)$ is an average of mean-zero i.i.d. random variables, it is $O_p(1/\sqrt{n})$. $\frac{1}{\hat{\pi}}-\frac{1}{\pi}=o_p(1)$, and therefore

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(W_i-\hat{\pi})\frac{\widetilde{Y}_i^{(1)}}{\hat{\pi}} \\
&\quad=\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)\frac{\widetilde{Y}_i^{(1)}-\mathbb{E}\left[\widetilde{Y}_i^{(1)}\right]}{\pi}+o_p(1/\sqrt{n})
\end{aligned}$$

A parallel argument shows us that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(W_i-\hat{\pi})\frac{\widetilde{Y}_i^{(0)}}{1-\hat{\pi}} \\
&\quad=\frac{1}{n}\sum_{i=1}^{n}(W_i-\pi)\frac{\widetilde{Y}_i^{(0)}-\mathbb{E}\left[\widetilde{Y}_i^{(0)}\right]}{1-\pi}+o_p(1/\sqrt{n})
\end{aligned}$$

□

Proposition 3.6. *Under Assumptions 3.1, 3.2, 3.3, 3.4 and 3.7,*

$$\widehat{\Sigma}_{\boldsymbol{X}Y}=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})(Y_i-\bar{Y})=O_p(1/\sqrt{n})$$

and

$$\widehat{\Sigma}_{\boldsymbol{X}YW}=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})(W_i-\hat{\pi})(Y_i-\bar{Y})=O_p(1/\sqrt{n})$$

Proof. We will show that $\hat{\Sigma}_{\boldsymbol{X}Y}=O_p(1/\sqrt{n})$, and leave the proof of the second result to the reader. We want to show that $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})(Y_i-\bar{Y})=O_p(1)$. First, note that $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})(Y_i-\bar{Y})=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})Y_i-\frac{\bar{Y}}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})$. For the second term, $\bar{Y}=O_p(1)$ since $\bar{Y}\to_p\mathbb{E}\left[f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)\right]$ (see proof of Proposition 3.2 for details), and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})=O_p(1)$ since $\boldsymbol{X}_i$ are i.i.d., so $\frac{\bar{Y}}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})=O_p(1)$.

For the first term, we use the Taylor expansion of Y_i that is possible by Assumption 3.7. This implies

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})Y_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right).$$

The first sum can be split into two. $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{X}_i f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) = O_p(1)$ since this is a sum of iid terms. $\frac{\bar{X}}{\sqrt{n}}\sum_{i=1}^{n} f_i(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) = O_p(1)$ since $\bar{X} = O_p(1)$ and the terms in the sum are iid, so $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_i(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) = O_p(1)$. Therefore, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) = O_p(1)$.

To see that the second sum is also $O_p(1)$, we will consider each element of the vector separately. Let $X_{i,k}$ be the value of the kth covariate for observational unit i and $\bar{X}_k$ the average of the kth covariate. We first see that all elements of the vector have mean zero:

$$\begin{aligned}
&\mathbb{E}\left[(\boldsymbol{X}_i-\bar{\boldsymbol{X}})f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right)\right]\\
&\quad=\mathbb{E}\left[\mathbb{E}\left[(\boldsymbol{X}_i-\bar{\boldsymbol{X}})f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right)|\mathcal{G}\right]\right]\\
&\quad=\mathbb{E}\left[(\boldsymbol{X}_i-\bar{\boldsymbol{X}})f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\right]\mathbb{E}\left[\mathbb{E}\left[\left(\frac{M_i}{N_i}-\pi\right)|\mathcal{G}\right]\right]\\
&\quad=0,
\end{aligned}$$

where the last equality follows from $\mathbb{E}\left[\mathbb{E}\left[\left(\frac{M_i}{N_i}-\pi\right)|\mathcal{G}\right]\right]=0$. So, if we can show the second moment is bounded, then we have the desired result.

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_{i,k}-\bar{X}_k)f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right)\right)^2\right]\\
&\quad=\mathbb{E}\left[(X_{i,k}-\bar{X}_k)^2 f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)^2\left(\frac{M_i}{N_i}-\pi\right)^2\right]\\
&\qquad+(n-1)\mathbb{E}\left[(X_{i,k}-\bar{X}_k)f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right)(X_{j,k}-\bar{X}_k)f'(W_j,\pi_j^*,\boldsymbol{X}_j,U_j;\epsilon_j)\left(\frac{M_j}{N_j}-\pi\right)\right]
\end{aligned}$$

By Assumption 3.7, $f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)^2 \le B^2$. Since $X_{i,k}$ has finite second moment by assumption, $\mathbb{E}\left[(X_{i,k}-\bar{X}_k)^2\right]<\infty$. By Lemma 15 of Li and Wager (2022), $\mathbb{E}\left[(M_i/N_i-\pi)\right] \le \frac{C}{n\rho_n}$. So,

$$\begin{aligned}
&\mathbb{E}\left[(X_{i,k}-\bar{X}_k)^2 f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)^2\left(\frac{M_i}{N_i}-\pi\right)^2\right]\\
&\quad\le B^2\mathbb{E}\left[(X_{i,k}-\bar{X}_k)^2\right]\frac{C}{n\rho_n}.
\end{aligned}$$

Next, note that

$$\mathbb{E}\Bigg[(X_{i,k}-\bar{X}_k)f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right) \\ (X_{j,k}-\bar{X}_k)f'(W_j,\pi_j^*,\boldsymbol{X}_j,U_j;\epsilon_j)\left(\frac{M_j}{N_j}-\pi\right)|E_{ij}=0\Bigg]=0,$$

since $\frac{M_i}{N_i}-\pi$ is independent of the rest of the factors when $E_{ij}=0$. So, since $P(E_{ij}=1)=\mathbb{E}\left[\mathbb{E}\left[E_{ij}|U_i,U_j\right]\right]=\mathbb{E}\left[G_n(U_i,U_j)\right]\leq\rho_n c_u$ by Assumption 3.4 and $|f'|\leq B$, we can use Hölder's inequality to get

$$\begin{aligned}
&\mathbb{E}\left[(X_{i,k}-\bar{X}_k)f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right)(X_{j,k}-\bar{X}_k)f'(W_j,\pi_j^*,\boldsymbol{X}_j,U_j;\epsilon_j)\left(\frac{M_j}{N_j}-\pi\right)\right] \\
&\quad\leq\mathbb{E}\Bigg[(X_{i,k}-\bar{X}_k)f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right) \\
&\qquad(X_{j,k}-\bar{X}_k)f'(W_j,\pi_j^*,\boldsymbol{X}_j,U_j;\epsilon_j)\left(\frac{M_j}{N_j}-\pi\right)|E_{ij}=1\Bigg]\rho_n c_u \\
&\quad\leq\mathbb{E}\left[(X_{i,k}-\bar{X}_k)^2f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)^2\left(\frac{M_i}{N_i}-\pi\right)^2|E_{ij}=1\right]\rho_n c_u \\
&\quad\leq B^2\rho_n c_u\mathbb{E}\left[(X_{i,k}-\bar{X}_k)^2\right]\mathbb{E}\left[\left(\frac{M_i}{N_i}-\pi\right)^2|E_{ij}=1\right].
\end{aligned}$$

By definition, $\frac{M_i}{N_i}-\pi=\frac{\sum_{l\neq i}E_{il}(W_l-\pi)}{N_i}$, and therefore

$$\mathbb{E}\left[\left(\frac{M_i}{N_i}-\pi\right)^2|E_{ij}=1\right]=\mathbb{E}\left[\frac{1}{N_i^2}\sum_{l_1\neq i}\sum_{l_2\neq i}\mathbb{E}\left[E_{l_1j}E_{l_2j}(W_{l_1}-\pi)(W_{l_2}-\pi)|E_{ij}=1,N_i\right]|E_{ij}=1\right]$$

For any $l_1\neq l_2$, $\mathbb{E}\left[E_{l_1j}E_{l_2j}(W_{l_1}-\pi)(W_{l_2}-\pi)|E_{ij}=1,N_i\right]=0$ since $W_{l_1}\perp$

$W_{l_2}, E_{l_1 j}, E_{l_2 j}, E_{ij}$. So,

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{M_i}{N_i}-\pi\right)^2 |E_{ij}=1\right] \\
&= \mathbb{E}\left[\frac{1}{N_i^2}\sum_{l\neq i}\mathbb{E}\left[E_{li}(W_l-\pi)^2|E_{ij}=1,N_i\right]|E_{ij}=1\right] \\
&= \pi(1-\pi)\mathbb{E}\left[\frac{1}{N_i^2}\sum_{l\neq i}\mathbb{E}\left[E_{li}|E_{ij}=1,N_i\right]|E_{ij}=1\right] \\
&= \pi(1-\pi)\mathbb{E}\left[\frac{1}{N_i}|E_{ij}=1\right] \\
&= \pi(1-\pi)\mathbb{E}\left[\frac{1}{1+\sum_{l\neq i,j}E_{il}}|E_{ij}=1\right] \\
&= \pi(1-\pi)\mathbb{E}\left[\mathbb{E}\left[\frac{1}{1+\sum_{l\neq i,j}E_{il}}|U_i,E_{ij}=1\right]|E_{ij}=1\right] \\
&= \pi(1-\pi)\mathbb{E}\left[\mathbb{E}\left[\frac{1}{1+\sum_{l\neq i,j}E_{il}}|U_i\right]|E_{ij}=1\right].
\end{aligned}$$

Given U_i, $\sum_{l\neq i,j}E_{il} \sim$ Binomial$(n-2, g_n(U_i))$, so $\mathbb{E}\left[\frac{1}{1+\sum_{l\neq i,j}E_{il}}|U_i\right] \leq \frac{1}{(n-2)g_n(U_i)} \leq \frac{1}{(n-2)\rho_n c_l}$ where the first inequality follows from properties of the binomial distribution (see, for example, Cribari-Neto et al. (2000)) and the second follows from (i) of Assumption 3.4. So,

$$\mathbb{E}\left[\left(\frac{M_i}{N_i}-\pi\right)^2 |E_{ij}=1\right] \leq \frac{\pi(1-\pi)}{(n-2)\rho_n c_l}.$$

This leads us to conclude that

$$\begin{aligned}
&(n-1)\mathbb{E}\left[(X_{i,k}-\bar{X}_k)f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right)(X_{j,k}-\bar{X}_k)f'(W_j,\pi_j^*,\boldsymbol{X}_j,U_j;\epsilon_j)\left(\frac{M_j}{N_j}-\pi\right)\right] \\
&\quad\leq (n-1)B^2\rho_n c_u\mathbb{E}\left[(X_{i,k}-\bar{X}_k)^2\right]\mathbb{E}\left[\left(\frac{M_i}{N_i}-\pi\right)^2|E_{ij}=1\right] \\
&\quad\leq (n-1)B^2\rho_n c_u\mathbb{E}\left[(X_{i,k}-\bar{X}_k)^2\right]\frac{\pi(1-\pi)}{(n-2)\rho_n c_l} \\
&\quad= O(1).
\end{aligned}$$

Therefore,

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n(X_{i,k}-\bar{X}_k)f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right)\right)^2\right] = O(1),$$

and so

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_{i,k}-\bar{X}_k)f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right)=O_p(1).$$

This concludes the proof that $\hat{\Sigma}_{\boldsymbol{X}Y}=O_p(1/\sqrt{n})$.

□

3.C Proofs

Proposition 3.1 (Consistency)**.** *Suppose Assumptions 3.1, 3.6 and 3.7 hold. If the random variables* $\hat{Y}_{i,n}^{(w,\boldsymbol{W}_{-i})}$ *are such that*

$$\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})} \to_p 0$$
$$\frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\pi}{1-\pi}\hat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})} \to_p 0$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\hat{Y}_{i,n}^{(w,\boldsymbol{W}_{-i})} = O_p(1)$$

for $w \in \{0,1\}$*, then the estimators* $\hat{\tau}^{HT}_{AIPW}$ *and* $\hat{\tau}^{HA}_{AIPW}$ *are consistent for* $\bar{\tau}_{DIR}$*, i.e.* $\hat{\tau}^{HT}_{AIPW} - \bar{\tau}_{DIR} \to_p 0$ *and* $\hat{\tau}^{HA}_{AIPW} - \bar{\tau}_{DIR} \to_p 0$*, conditional on the graph and the potential outcomes.*

Proof of 3.1. Consider $\hat{\tau}^{\text{HT}}_{\text{AIPW}}$. Note,

$$\hat{\tau}^{\text{HT}}_{\text{AIPW}} = \hat{\tau}^{\text{HT}} + \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})} - \frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\pi}{1-\pi}\hat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})}. \tag{3.19}$$

By assumption, the two sums are $o_p(1)$. By Li and Wager (2022), $\hat{\tau}^{\text{HT}}$ is consistent for $\bar{\tau}_{\text{DIR}}$, and therefore is $\hat{\tau}^{\text{HT}}_{\text{AIPW}}$ consistent.

Now we turn our attention to $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$. We can express this as $\hat{\tau}^{\text{HA}} + o_p(1)$. Let $\hat{\pi} = \frac{1}{n}\sum_{i=1}^{n} W_i$. Then

$$\hat{\tau}^{\text{HA}}_{\text{AIPW}} = \hat{\tau}^{\text{HA}} + \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\hat{\pi}}\right)\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})} - \frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\hat{\pi}}{1-\hat{\pi}}\hat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})}. \tag{3.20}$$

We will prove $\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\hat{\pi}}\right)\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})} = o_p(1)$ and leave the second sum to the reader. We start with a small rewrite:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\hat{\pi}}\right)\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})} \\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{\pi}{\hat{\pi}}\right)\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})} + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\pi}{\hat{\pi}}-\frac{W_i}{\hat{\pi}}\right)\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})} \\
&\quad= \left(1-\frac{\pi}{\hat{\pi}}\right)\frac{1}{n}\sum_{i=1}^{n}\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})} + \frac{\pi}{\hat{\pi}}\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}.
\end{aligned}$$

By assumption, $\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})} = o_p(1)$. Since $\pi/\hat{\pi} \to 1$ as $n \to \infty$, the second term is $o_p(1)$. For the first sum, $\left(1-\frac{\pi}{\hat{\pi}}\right) = o_p(1)$. By assumption, $\frac{1}{n}\sum_{i=1}^{n}\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})} = O_p(1)$, and so the first sum is $o_p(1)$. This completes the proof.

□

Theorem 3.1 (Central Limit Theorem)**.** *Suppose Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7 hold, and further assume the existence of random variables* $\widetilde{Y}_i^{(w)}, w \in \{0,1\}$*, such that* $(W_i, \boldsymbol{X}_i, U_i, \widetilde{Y}_i^{(0)}, \widetilde{Y}_i^{(1)}; \epsilon_i)$ *are iid. Also, assume*

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(1)}\right) \\ &\qquad = o_p(1/\sqrt{n}) \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\pi}{1-\pi}\left(\hat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(0)}\right) \\ &\qquad = o_p(1/\sqrt{n}). \end{aligned} \tag{3.7}$$

Then

$$\begin{aligned} &\sqrt{n}\left(\hat{\tau}_{AIPW}^{HT}-\bar{\tau}_{DIR}\right) \to_d N(0,\pi(1-\pi)\mathbb{E}\left[\left(R_i-A_i+Q_i\right)^2\right]), \\ &\sqrt{n}\left(\hat{\tau}_{AIPW}^{HA}-\bar{\tau}_{DIR}\right) \to_d N\left(0,\pi(1-\pi)\left(Var[R_i-A_i+Q_i]+\left(\mathbb{E}\left[Q_i\right]\right)^2\right)\right), \end{aligned} \tag{3.8}$$

where

$$\begin{aligned} R_i &= \frac{f(1,\boldsymbol{X}_i,U_i,\pi;\epsilon_i)}{\pi}+\frac{f(0,\boldsymbol{X}_i,U_i,\pi;\epsilon_i)}{1-\pi}, \\ A_i &= \frac{\widetilde{Y}_i^{(1)}}{\pi}+\frac{\widetilde{Y}_i^{(0)}}{1-\pi} \\ Q_i &= \mathbb{E}\left[\frac{G(U_i,U_j)(f'(1,\pi,\boldsymbol{X}_j,U_j;\epsilon_j)-f'(0,\pi,\boldsymbol{X}_j,U_j;\epsilon_j))}{\mathbb{E}\left[G(U_i,U_j)|U_j\right]}\middle| U_i\right]. \end{aligned}$$

If $\sqrt{n}\rho_n \to \infty$*, then with* $\sigma_0^2 = Var[f(1,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)-f(0,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)]$*,*

$$\begin{aligned} &\sqrt{n}\left(\hat{\tau}_{AIPW}^{HT}-\tau_{DIR}\right) \to_d N(0,\sigma_0^2+\pi(1-\pi)\mathbb{E}\left[\left(R_i-A_i+Q_i\right)^2\right]), \\ &\sqrt{n}\left(\hat{\tau}_{AIPW}^{HA}-\tau_{DIR}\right) \to_d N\left(0,\sigma_0^2+\pi(1-\pi)\left(Var[R_i-A_i+Q_i]+\left(\mathbb{E}\left[Q_i\right]\right)^2\right)\right), \end{aligned} \tag{3.9}$$

Proof of Theorem 3.1. First, recall the rewrite of the estimator in Equation 3.19

$$\hat{\tau}_{\mathrm{AIPW}}^{\mathrm{HT}} = \hat{\tau}_{\mathrm{DIR}}^{\mathrm{HT}} + \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\hat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})} - \frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\pi}{1-\pi}\hat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})}. \tag{3.21}$$

Li and Wager (2022) showed that the Horvitz-Thompson estimator can be written as

$$\hat{\tau}_{\mathrm{DIR}}^{\mathrm{HT}} = \bar{\tau}_{\mathrm{DIR}} + \frac{1}{n}\sum_{i=1}^{n}(R_i+Q_{n,i})(W_i-\pi) + o_p(1/\sqrt{n})$$

where $Q_{n,i} = \mathbb{E}\left[G_n(U_i,U_j)\frac{(f'(1,\pi,\boldsymbol{X}_j,U_j)-f'(0,\pi,\boldsymbol{X}_j,U_j))}{\mathbb{E}\left[G_n(U_i,U_j)|U_j\right]}|U_i\right]$ which converges to Q_i.

Note we can write

$$\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\widehat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}=\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\widehat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(1)}\right)+\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\widetilde{Y}_i^{(1)}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\pi}{1-\pi}\widehat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})}=\frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\pi}{1-\pi}\left(\widehat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(0)}\right)+\frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\pi}{1-\pi}\widetilde{Y}_i^{(0)}.$$

We plug these three expressions into Equation 3.21 to get

$$\begin{aligned}
\widehat{\tau}_{\text{AIPW}}^{\text{HA}}-\bar{\tau}_{\text{DIR}} &= \frac{1}{n}\sum_{i=1}^{n}(R_i+Q_{n,i})(W_i-\pi)+o_p(1/\sqrt{n})\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\widehat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(1)}\right)+\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\widetilde{Y}_i^{(1)}\\
&\quad-\frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\pi}{1-\pi}\left(\widehat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(0)}\right)-\frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\pi}{1-\pi}\widetilde{Y}_i^{(0)}\\
&= \frac{1}{n}\sum_{i=1}^{n}(R_i+Q_{n,i})(W_i-\pi)\\
&\quad-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\widetilde{Y}_i^{(1)}}{\pi}+\frac{\widetilde{Y}_i^{(0)}}{1-\pi}\right)(W_i-\pi)\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\widehat{Y}_{i,n}^{(1,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(1)}\right)\\
&\quad-\frac{1}{n}\sum_{i=1}^{n}\frac{W_i-\pi}{1-\pi}\left(\widehat{Y}_{i,n}^{(0,\boldsymbol{W}_{-i})}-\widetilde{Y}_i^{(0)}\right)\\
&\quad+o_p(1/\sqrt{n})\\
&= \frac{1}{n}\sum_{i=1}^{n}(R_i-A_i+Q_{n,i})(W_i-\pi)+o_p(1/\sqrt{n}),
\end{aligned}$$

where the last equality follows from Equation 3.6 and Equation 3.7. $\{(R_i-A_i+Q_{n,i})(W_i-\pi)\}_{i=1}^{n}$ is a sequence of i.i.d. random variables, and so the central limit theorem tells us that

$$\sqrt{n}(\widehat{\tau}_{\text{AIPW}}^{\text{HT}}-\bar{\tau}_{\text{DIR}})\to_d N(0,\text{Var}\left[(R_i-A_i+Q_{n,i})(W_i-\pi)\right]).$$

Since $\mathbb{E}\left[(R_i-A_i+Q_{n,i})(W_i-\pi)\right]=0$,

$$\begin{aligned}
&\text{Var}\left[(R_i-A_i+Q_{n,i})(W_i-\pi)\right]\\
&\quad=\mathbb{E}\left[(R_i-A_i+Q_{n,i})^2(W_i-\pi)^2\right]\\
&\quad=\pi(1-\pi)\mathbb{E}\left[(R_i-A_i+Q_{n,i})^2\right],
\end{aligned}$$

which by the dominated convergence theorem converges to $\pi(1-\pi)\mathbb{E}\left[(R_i-A_i+Q_i)^2\right]$.

For $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$, Li and Wager (2022) showed that we can rewrite this as

$$\hat{\tau}^{\text{HA}}_{\text{AIPW}} = \bar{\tau}_{\text{DIR}} + \frac{1}{n}\sum_{i=1}^{n}(R_i - \mathbb{E}\left[R_i\right] + Q_{n,i})(W_i - \pi) + o_p(1/\sqrt{n}).$$

$\hat{\tau}^{\text{HA}}$ can be rewritten as in Equation 3.20. Using these two rewrites and Proposition 3.5, we see that

$$\begin{aligned}
\hat{\tau}^{\text{HA}}_{\text{AIPW}} - \bar{\tau}_{\text{DIR}} &= \frac{1}{n}\sum_{i=1}^{n}(R_i - \mathbb{E}\left[R_i\right] + Q_{n,i})(W_i - \pi) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{W_i}{\hat{\pi}}\right)\hat{Y}^{(1,\boldsymbol{W}_{-i})}_{i,n} - \frac{1}{n}\sum_{i=1}^{n}\frac{W_i - \hat{\pi}}{1-\hat{\pi}}\hat{Y}^{(0,\boldsymbol{W}_{-i})}_{i,n} \\
&\quad + o_p(1/\sqrt{n}) \\
&= \frac{1}{n}\sum_{i=1}^{n}(R_i - \mathbb{E}\left[R_i\right] + Q_{n,i})(W_i - \pi) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}(W_i - \hat{\pi})\left\{\frac{\hat{Y}^{(1,\boldsymbol{W}_{-i})}_{i,n}}{\hat{\pi}} + \frac{\hat{Y}^{(0,\boldsymbol{W}_{-i})}_{i,n}}{1-\hat{\pi}}\right\} \\
&\quad + o_p(1/\sqrt{n}) \\
&= \frac{1}{n}\sum_{i=1}^{n}(R_i - A_i - \mathbb{E}\left[R_i - A_i\right] + Q_{n,i})(W_i - \pi) \\
&\quad + o_p(1/\sqrt{n}).
\end{aligned}$$

Since $\{(R_i - A_i - \mathbb{E}\left[R_i - A_i\right] + Q_{n,i})(W_i - \pi)\}_{i=1}^{n}$ is a sequence of i.i.d. random variables, the Central Limit Theorem tells us

$$\sqrt{n}(\hat{\tau}^{\text{HA}}_{\text{AIPW}} - \bar{\tau}_{\text{DIR}}) \to_d N(0, \lim_{n\to\infty}\sigma_n^2),$$

where

$$\begin{aligned}
\sigma_n^2 &= \text{Var}((R_i - A_i - \mathbb{E}\left[R_i - A_i\right] + Q_{n,i})(W_i - \pi)) \\
&= \mathbb{E}\left[(R_i - A_i - \mathbb{E}\left[R_i - A_i\right] + Q_{n,i})^2\right]\mathbb{E}\left[(W_i - \pi)^2\right] \\
&= \pi(1-\pi)\left(\text{Var}(R_i - A_i - \mathbb{E}\left[R_i - A_i\right] + Q_{n,i}) + (\mathbb{E}\left[R_i - A_i - \mathbb{E}\left[R_i - A_i\right] + Q_{n,i}\right])^2\right) \\
&= \pi(1-\pi)\left(\text{Var}(R_i - A_i + Q_{n,i}) + (\mathbb{E}\left[Q_{n,i}\right])^2\right),
\end{aligned}$$

Taking the limit of this expression gives the desired result.

For the limiting population estimands, note that $\bar{\tau}_{\text{DIR}} = \frac{1}{n}\sum_{i=1}^{n} f(1,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) - f(0,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) + o_p(1/\sqrt{n})$ and $\tau_{\text{DIR}} = \mathbb{E}\left[f(1,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) - f(0,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)\right]$ if

$\sqrt{n}\rho_n \to \infty$, and so

$$\begin{aligned}
&\hat{\tau}^{\mathrm{HT}}_{\mathrm{AIPW}} - \tau_{\mathrm{DIR}} \\
&\quad = \hat{\tau}^{\mathrm{HT}}_{\mathrm{AIPW}} - \bar{\tau}_{\mathrm{DIR}} + \bar{\tau}_{\mathrm{DIR}} - \tau_{\mathrm{DIR}} \\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\left[(R_i - A_i + Q_{n,i})(W_i - \pi) + f(1,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) - f(0,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) - \tau_{\mathrm{DIR}}\right] \\
&\qquad + o_p(1/\sqrt{n}).
\end{aligned}$$

This is a sum of iid random variables with mean 0, and therefore the central limit theorem gives us that

$$\sqrt{n}\left(\hat{\tau}^{\mathrm{HT}}_{\mathrm{AIPW}} - \tau_{\mathrm{DIR}}\right) \to_d N(0, \sigma^2_\infty),$$

where

$$\sigma^2_\infty = \lim_{n\to\infty} \mathbb{E}\left[\left((R_i - A_i + Q_{n,i})(W_i - \pi) + f(1,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) - f(0,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) - \tau_{\mathrm{DIR}}\right)^2\right]$$

Since W_i is assigned independently,

$$\mathbb{E}\left[(R_i - A_i + Q_{n,i})(W_i - \pi)(f(1,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) - f(0,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) - \tau_{\mathrm{DIR}})\right] = 0,$$

and so

$$\begin{aligned}
&\mathbb{E}\left[\left(R_i - A_i + Q_{n,i})(W_i - \pi) + f(1,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) - f(0,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) - \tau_{\mathrm{DIR}}\right)^2\right] \\
&= \mathbb{E}\left[(R_i - A_i + Q_{n,i})^2(W_i - \pi)^2\right] + \mathbb{E}\left[\left(f(1,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) - f(0,\pi,\boldsymbol{X}_i,U_i;\epsilon_i) - \tau_{\mathrm{DIR}}\right)^2\right] \\
&= \pi(1-\pi)\mathbb{E}\left[(R_i - A_i + Q_{n,i})^2\right] + \sigma^2_0.
\end{aligned}$$

Taking the limit gives the desired result. An analogue argument gives the result for $\hat{\tau}^{\mathrm{HA}}_{\mathrm{AIPW}}$.

□

Lemma 3.1. *Assume $\hat{Y}^{(w,\boldsymbol{w}_{-i})}_{i,n} = \hat{h}^{(w)}_n(\boldsymbol{X}_i, M_i/N_i, U_i)$, where $\hat{h}^{(w)}_n$ are random functions such that there exists a P-Donsker class $\mathcal{F}$ with $P(\hat{h}^{(w)}_n \in \mathcal{F}) \to 1$. We assume there exists $h^{(w)} \in L_2(P), w \in \{0,1\}$ with $\int \left(\hat{h}^{(w)}_n(x,\pi,u) - h^{(w)}(x,\pi,u)\right)^2 dP_{(\boldsymbol{X}_i,U_i)}(x,u) \to_p 0$. Then Equation 3.6 and Equation 3.7 hold with $\widetilde{Y}^{(w)}_i = h^{(w)}(\boldsymbol{X}_i,\pi,U_i)$ if one of the following is true:*

a. $\hat{h}^{(w)}_n(\boldsymbol{X}_i, M_i/N_i, U_i) = \hat{h}^{(w)}_n(\boldsymbol{X}_i, U_i)$
b. $\hat{h}^{(w)}_n(\boldsymbol{X}_i, M_i/N_i, U_i) = \hat{f}_n(\boldsymbol{X}_i, U_i) + \hat{\beta} M_i/N_i$ *with* $\hat{f}_n \to_p f$ *and* $\hat{\beta}_n \to_p \beta \in \mathbb{R}$

Proof of 3.1. Here, we show that Equation 3.6 follows as postulated; Equation 3.7 also follows by parallel arguments.

First, note we can rewrite Equation 3.6 as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\hat{h}_n^{(1)}\left(\boldsymbol{X}_i, M_i/N_i, U_i\right) - h^{(1)}(\boldsymbol{X}_i,\pi,U_i)\right) \\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\hat{h}_n^{(1)}\left(\boldsymbol{X}_i, M_i/N_i, U_i\right) - \hat{h}_n^{(1)}(\boldsymbol{X}_i,\pi,U_i)\right) \\
&\qquad + \frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\hat{h}_n^{(1)}\left(\boldsymbol{X}_i, \pi, U_i\right) - h^{(1)}(\boldsymbol{X}_i,\pi,U_i)\right).
\end{aligned}$$

(a) and (b) both imply that the first of these two averages is $o_p(1/\sqrt{n})$; we will provide details below. First, we will show that the second average is $o_p(1/\sqrt{n})$ by results from empirical process theory. This follows directly from applying Theorem 6.15 of Bolthausen et al. (2002). We include a detailed "translation" of our specific situation to the notation used by Bolthausen et al. (2002).

Let $\hat{f}_n$ be the random function $(1-\frac{w}{\pi})\,\hat{h}_n^{(1)}(x,\pi,u)$ (note: the randomness comes from the estimation of $\hat{h}_n^{(1)}$, which implicitly is a function of the observed data $(\boldsymbol{Y},\boldsymbol{W},\boldsymbol{E},\boldsymbol{X})$), and f_0 be the function $(1-\frac{w}{\pi})\,h^{(1)}(x,\pi,u)$. We see that

$$\begin{aligned}
P\hat{f}_n &= \int\left(1-\frac{w}{\pi}\right)\hat{h}_n^{(1)}(x,\pi,u)dP_{(W_i,\boldsymbol{X}_i,U_i)}(w,x,u) \\
&= \int\left(1-\frac{w}{\pi}\right)dP_{W_i}(w)\int\hat{h}_n^{(1)}(x,\pi,u)dP_{(\boldsymbol{X}_i,U_i)}(x,u) \\
&= 0,
\end{aligned}$$

since $P_{(W_i,\boldsymbol{X}_i,U_i)}$ is the joint measure of $(W_i,\boldsymbol{X}_i,U_i)$, and $W_i \perp (\boldsymbol{X}_i,U_i)$. We want to show

$$\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\hat{h}_n^{(1)}\left(\boldsymbol{X}_i,\pi,U_i\right) - h^{(1)}(\boldsymbol{X}_i,\pi,U_i)\right) = \mathbb{G}_n(\hat{f}_n - f_0) \to 0$$

in probability. This holds by Theorem 6.15 of Bolthausen et al. (2002) if $P(\hat{f}_n - f_0)^2 \to 0$ in probability. Since $P_{(W_i,\boldsymbol{X}_i,U_i)}(w,x,u) = P_{W_i}(w)P_{(\boldsymbol{X}_i,U_i)}(x,u)$,

$$\begin{aligned}
P(\hat{f}_n - f_0)^2 &= \int\left(\left(1-\frac{w}{\pi}\right)\hat{h}_n^{(1)}(x,\pi,u) - \left(1-\frac{w}{\pi}\right)h^{(1)}(x,\pi,u)\right)^2 P_{(W_i,\boldsymbol{X}_i,U_i)}(w,x,u) \\
&= \int\int\left(1-\frac{w}{\pi}\right)^2\left(\hat{h}_n^{(1)}(x,\pi,u) - h^{(1)}(x,\pi,u)\right)^2 P_{W_i}(w)P_{(\boldsymbol{X}_i,U_i)}(x,u) \\
&= \left\{\left(1-\frac{1}{\pi}\right)^2\pi + (1-\pi)\right\}\int\left(\hat{h}_n^{(1)}(x,\pi,u) - h^{(1)}(x,\pi,u)\right)^2 P_{(\boldsymbol{X}_i,U_i)}(x,u)
\end{aligned}$$

which by assumption converges to 0 in probability. Therefore, by Theorem 6.15 of Bolthausen et al. (2002), we can conclude that $\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\hat{h}_n^{(1)}\left(\boldsymbol{X}_i,\pi,U_i\right) - h^{(1)}(\boldsymbol{X}_i,\pi,U_i)\right) = o_p(1/\sqrt{n})$.

We now show that (a) implies

$$\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\hat{h}_n^{(1)}\left(\boldsymbol{X}_i, M_i/N_i, U_i\right)-\hat{h}_n^{(1)}(\boldsymbol{X}_i,\pi,U_i)\right)=o_p(1/\sqrt{n})$$

Since by (a), $\hat{h}_n^{(w)}(\boldsymbol{X}_i, M_i/N_i, U_i) = \hat{h}_n^{(w)}(\boldsymbol{X}_i, U_i)$, we see that $\hat{h}_n^{(w)}(\boldsymbol{X}_i, M_i/N_i, U_i) - \hat{h}_n^{(w)}(\boldsymbol{X}_i, \pi, U_i) = 0$, and so

$$\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\hat{h}_n^{(1)}\left(\boldsymbol{X}_i, M_i/N_i, U_i\right)-\hat{h}_n^{(1)}(\boldsymbol{X}_i,\pi,U_i)\right)=0.$$

Finally, assume (b). Since $\hat{h}_n^{(w)}(\boldsymbol{X}_i, M_i/N_i, U_i) = \hat{f}_n(\boldsymbol{X}_i,\pi,U_i) + \hat{\beta}\left(\frac{M_i}{N_i}-\pi\right)$, we see that the difference is $\hat{h}_n^{(w)}(\boldsymbol{X}_i, M_i/N_i, U_i) - \hat{h}_n^{(w)}(\boldsymbol{X}_i,\pi,U_i) = \hat{\beta}\left(\frac{M_i}{N_i}-\pi\right)$. Therefore,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\hat{h}_n^{(1)}\left(\boldsymbol{X}_i, M_i/N_i, U_i\right)-\hat{h}_n^{(1)}(\boldsymbol{X}_i,\pi,U_i)\right)\\ &\quad=\frac{\hat{\beta}_n}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\frac{M_i}{N_i}-\pi\right)\end{aligned}$$

By assumption, $\hat{\beta}_n \to_p \beta$. Therefore, if we can show $\frac{1}{n}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\frac{M_i}{N_i}-\pi\right) = o_p(1/\sqrt{n})$, the result follows from Slutsky's Theorem.

We need to show $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\frac{M_i}{N_i}-\pi\right)\to_p 0$, i.e.

$$\lim_{n\to\infty} P\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\frac{M_i}{N_i}-\pi\right)\right|\geq\epsilon\right)=0 \quad \text{for any } \epsilon>0.$$

Since each term in the sum has mean zero by the independence of W_i and M_i/N_i, Chebyshev's inequality tells us that it is sufficient to show that the second moment converges to 0 as $n\to\infty$. Therefore, consider the second moment,

$$\begin{aligned}&\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{W_i}{\pi}\right)\left(\frac{M_i}{N_i}-\pi\right)\right)^2\right]\\ &\quad=\mathbb{E}\left[\left(1-\frac{W_i}{\pi}\right)^2\left(\frac{M_i}{N_i}-\pi\right)^2\right]\\ &\qquad+(n-1)\mathbb{E}\left[\left(1-\frac{W_i}{\pi}\right)\left(\frac{M_i}{N_i}-\pi\right)\left(1-\frac{W_j}{\pi}\right)\left(\frac{M_j}{N_j}-\pi\right)\right],\end{aligned}$$

where $i\neq j$.

Since $\frac{M_i}{N_i} = \frac{\sum_{j \neq i} E_{ij} W_j}{\sum_{j \neq i} E_{ij}}$, the first term is

$$\mathbb{E}\left[\left(1 - \frac{W_i}{\pi}\right)^2 \left(\frac{M_i}{N_i} - \pi\right)^2\right] = \mathbb{E}\left[\left(1 - \frac{W_i}{\pi}\right)^2\right] \mathbb{E}\left[\left(\frac{M_i}{N_i} - \pi\right)^2\right] = \left(\frac{1}{\pi} - 1\right) \mathbb{E}\left[\left(\frac{M_i}{N_i} - \pi\right)^2\right].$$

Since $\mathbb{E}\left[\left(\frac{M_i}{N_i} - \pi\right)^2\right] = O_p(1/(\rho_n n)^2)$ (see Lemma 15 in Appendix B of Li and Wager (2022)), this term converges to 0 as $n \to \infty$.

Now, consider the second term. To ease notation, let $\widetilde{W}_i = 1 - W_i/\pi$. Then,

$$\begin{aligned}
&(n-1)\mathbb{E}\left[\left(1 - \frac{W_i}{\pi}\right)\left(\frac{M_i}{N_i} - \pi\right)\left(1 - \frac{W_j}{\pi}\right)\left(\frac{M_j}{N_j} - \pi\right)\right] \\
&= (n-1)\mathbb{E}\left[\widetilde{W}_i \widetilde{W}_j \frac{\sum_{k \neq i} E_{ik}(W_k - \pi)}{N_i} \frac{\sum_{l \neq j} E_{jl}(W_l - \pi)}{N_j}\right] \\
&= (n-1)\mathbb{E}\left[\frac{\sum_{k \neq i} \sum_{l \neq j} \widetilde{W}_i E_{ik}(W_k - \pi) \widetilde{W}_j E_{jl}(W_l - \pi)}{N_i N_j}\right] \\
&= (n-1)\mathbb{E}\left[\mathbb{E}\left[\frac{\sum_{k \neq i} \sum_{l \neq j} \widetilde{W}_i E_{ik}(W_k - \pi) \widetilde{W}_j E_{jl}(W_l - \pi)}{N_i N_j} \middle| N_i, N_j\right]\right] \\
&= (n-1)\mathbb{E}\left[\frac{\sum_{k \neq i} \sum_{l \neq j} \mathbb{E}\left[\widetilde{W}_i E_{ik}(W_k - \pi) \widetilde{W}_j E_{jl}(W_l - \pi) \middle| N_i, N_j\right]}{N_i N_j}\right].
\end{aligned}$$

Notice that $\widetilde{W}_i \perp E_{ik}(W_k - \pi)\widetilde{W}_j E_{jl}(W_l - \pi) | N_i, N_j$ for all $l \neq i$ (since $i \neq j$). So,

$$\mathbb{E}\left[\widetilde{W}_i E_{ik}(W_k - \pi)\widetilde{W}_j E_{jl}(W_l - \pi) \middle| N_i, N_j\right] = \mathbb{E}\left[\widetilde{W}_i\right] \mathbb{E}\left[E_{ik}(W_k - \pi)\widetilde{W}_j E_{jl}(W_l - \pi) \middle| N_i, N_j\right] = 0.$$

The same holds for all $k \neq j$. Therefore,

$$\begin{aligned}
&(n-1)\mathbb{E}\left[\frac{\sum_{k \neq i} \sum_{l \neq j} \mathbb{E}\left[\widetilde{W}_i E_{ik}(W_k - \pi)\widetilde{W}_j E_{jl}(W_l - \pi) \middle| N_i, N_j\right]}{N_i N_j}\right] \\
&= (n-1)\mathbb{E}\left[\frac{\mathbb{E}\left[\widetilde{W}_i E_{ij}(W_j - \pi)\widetilde{W}_j E_{ji}(W_i - \pi) \middle| N_i, N_j\right]}{N_i N_j}\right] \\
&= (n-1)\mathbb{E}\left[\frac{\mathbb{E}\left[E_{ij}(W_j - \pi)^2 (W_i - \pi)^2 \middle| N_i, N_j\right]}{N_i N_j \pi^2}\right] \\
&= (n-1)\mathbb{E}\left[\frac{E_{ij}(W_j - \pi)^2 (W_i - \pi)^2}{\pi^2 N_i N_j}\right].
\end{aligned}$$

where the second to last equality follows since $\widetilde{W}_i = \frac{-1}{\pi}(W_i - \pi)$. By the law of total

expectation,

$$\begin{aligned}\mathbb{E}\left[\frac{E_{ij}(W_j-\pi)^2(W_i-\pi)^2}{\pi^2 N_i N_j}\right] &= \sum_{x=0,1}\mathbb{E}\left[\frac{E_{ij}(W_j-\pi)^2(W_i-\pi)^2}{\pi^2 N_i N_j}\middle| E_{ij}=x\right]P(E_{ij}=x)\\ &= \frac{\mathbb{E}\left[(W_j-\pi)^2\right]^2}{\pi^2}\,\mathbb{E}\left[\frac{1}{N_i N_j}\middle| E_{ij}=1\right]\mathbb{E}\left[E_{ij}\right]\\ &= (1-\pi)^2\,\mathbb{E}\left[\frac{1}{N_i N_j}\middle| E_{ij}=1\right]\mathbb{E}\left[E_{ij}\right]\end{aligned}$$

Now, $\mathbb{E}\left[E_{ij}\right] = \mathbb{E}\left[\mathbb{E}\left[E_{ij}|U_i,U_j\right]\right] = \mathbb{E}\left[G_n(U_i,U_j)\right] = \mathbb{E}\left[\min\{\rho_n G(U_i,U_j),1\}\right] \leq \rho_n\mathbb{E}\left[G(U_i,U_j)\right] \leq \rho_n c_u$, where the last inequality follows by (ii) of Assumption 3.4.

Next, we consider $\mathbb{E}\left[\frac{1}{N_iN_j}\middle| E_{ij}=1\right]$. By Hölder's inequality, $\mathbb{E}\left[\frac{1}{N_iN_j}\middle| E_{ij}=1\right] \leq \sqrt{\mathbb{E}\left[\frac{1}{N_i^2}\middle| E_{ij}=1\right]^2\mathbb{E}\left[\frac{1}{N_j^2}\middle| E_{ij}=1\right]^2} = \mathbb{E}\left[\frac{1}{N_i^2}\middle| E_{ij}=1\right]$, since $E_{ij}=E_{ji}$. So,

$$\begin{aligned}\mathbb{E}\left[\frac{1}{N_iN_j}\middle| E_{ij}=1\right] &= \mathbb{E}\left[\frac{1}{N_i^2}\middle| E_{ij}=1\right]\\ &= \mathbb{E}\left[\frac{1}{(1+\sum_{k\neq i,j}E_{ik})^2}\middle| E_{ij}=1\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{1}{(1+\sum_{k\neq i,j}E_{ik})^2}\middle| U_i\right]\right].\end{aligned}$$

Note that $\sum_{k\neq i,j}E_{ik}|U_i \sim \text{Binomial}(n-2,\mathbb{E}\left[G_n(U_i,U_j)|U_i\right])$. Therefore, by properties of the binomial distribution (see, for example, Cribari-Neto et al. (2000)) and the fact that $\mathbb{E}\left[G_n(U_i,U_j)|U_i\right] \geq \rho_n c_l$ (by (i) of Assumption 3.4), $\mathbb{E}\left[\mathbb{E}\left[\frac{1}{(1+\sum_{k\neq i,j}E_{ik})^2}\middle| U_i\right]\right] \leq \frac{M}{(n-2)^2\mathbb{E}\left[G_n(U_i,U_j)|U_i\right]^2} = \frac{M}{(n-2)^2\rho_n^2 c_l^2}$ for an appropriate constant $M>0$.

Finally, this let us conclude that for an appropriate constant $M>0$

$$\begin{aligned}&(n-1)\mathbb{E}\left[\left(1-\frac{W_i}{\pi}\right)\left(\frac{M_i}{N_i}-\pi\right)\left(1-\frac{W_j}{\pi}\right)\left(\frac{M_j}{N_j}-\pi\right)\right]\\ &\leq (n-1)(1-\pi)^2\mathbb{E}\left[E_{ij}\right]\mathbb{E}\left[\frac{1}{N_i^2}\middle| E_{ij}\right]\\ &\leq (n-1)(1-\pi)^2\rho_n c_u\frac{M}{(n-2)^2\rho_n^2c_l^2},\end{aligned}$$

which converges to 0 as $n\to\infty$. This implies $\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(1-\frac{W_i}{\pi}\right)\left(\frac{M_i}{N_i}-\pi\right)\right)^2\right]\to 0$, and therefore $\frac{1}{n}\sum_{i=1}^n\left(1-\frac{W_i}{\pi}\right)\hat{\beta}_n\left(\frac{M_i}{N_i}-\pi\right) = o_p(1/\sqrt{n})$.

□

Proposition 3.2. *Under Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7,* $\hat{\beta}^{ANCOVA}$ *is*

consistent for $\bar{\tau}_{DIR}$, and asymptotically normal in that

$$\sqrt{n}\left(\hat{\beta}^{ANCOVA} - \bar{\tau}_{DIR}\right) \to_d N\left(0, \pi(1-\pi)\left(Var\left[\hat{\tau}^{HA}\right] + Var[A_i] - 2Cov[A_i, R_i]\right)\right),$$

where

$$\begin{aligned}
Var\left[\hat{\tau}^{HA}\right] &= Var[R_i + Q_i] + (\mathbb{E}[Q_i])^2, \\
R_i &= \frac{f(1, X_i, U_i, \pi; \epsilon_i)}{\pi} + \frac{f(0, X_i, U_i, \pi; \epsilon_i)}{1-\pi}, \\
A_i &= \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right) \Sigma_{\boldsymbol{Xf}}^T \Sigma_{\boldsymbol{XX}}^{-1} \boldsymbol{X}_i, \\
Q_i &= \mathbb{E}\left[\frac{G(U_i, U_j)(f'(1, \pi, \boldsymbol{X}_j, U_j) - f'(0, \pi, \boldsymbol{X}_j, U_j))}{\mathbb{E}\left[G(U_i, U_j)|U_j\right]}\middle| U_i\right]
\end{aligned}$$

with

$$\begin{aligned}
\Sigma_{\boldsymbol{X}f} &= \mathbb{E}[\boldsymbol{X}_i f(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)] - \mathbb{E}[\boldsymbol{X}_i]\,\mathbb{E}[f(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)] \\
\Sigma_{\boldsymbol{XX}} &= \mathbb{E}[(\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])(\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])^T]
\end{aligned}$$

Proof of 3.2. We show that $\hat{\beta}^{\text{ANCOVA}}$ is asymptotically equivalent to an estimator of the form $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$ with

$$\hat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})} = \hat{h}_n^{(w)}(\boldsymbol{X}_i) = \hat{\Sigma}_{\boldsymbol{XY}}^T \hat{\Sigma}_{\boldsymbol{XX}}^{-1} \boldsymbol{X}_i$$

which converge to

$$\widetilde{Y}_i^{(w)} = h^{(w)}(\boldsymbol{X}_i) = \Sigma_{\boldsymbol{Xf}}^T \Sigma_{\boldsymbol{XX}}^{-1} \boldsymbol{X}_i.$$

We then rewoke the result in Lemma 3.1(a) to arrive at the desired conclusion.

As shown by Tsiatis et al. (2008), the estimator $\hat{\beta}^{\text{ANCOVA}}$ can be written as

$$\begin{aligned}
&\hat{\beta}^{\text{ANCOVA}} \\
&= \left\{1 - \frac{n^2}{n_0 n_1}(n^{-1}d_1)^T \hat{\Sigma}_{\boldsymbol{XX}}^{-1}(n^{-1}d_1)\right\}^{-1} \\
&\qquad \left\{\hat{\tau}^{\text{HA}} - \frac{n}{n_0 n_1}\sum_{i=1}^n (W_i - \hat{\pi})\hat{\Sigma}_{\boldsymbol{XY}}^T \hat{\Sigma}_{\boldsymbol{XX}}^{-1} \boldsymbol{X}_i\right\}.
\end{aligned}$$

Note that the second factor of $\hat{\beta}^{\text{ANCOVA}}$ is exactly of the wanted form, so we simply need

$$\frac{n^2}{n_0 n_1}(n^{-1}d_1)^T \hat{\Sigma}_{\boldsymbol{XX}}^{-1}(n^{-1}d_1) \to_p 0.$$

First, since the covariates are i.i.d. with finite variance by assumption, $\hat{\Sigma}_{\boldsymbol{XX}} \to_p \Sigma_{\boldsymbol{XX}}$. Second, since $\{(W_i, \boldsymbol{X}_i)\}_{i=1}^n$ is a sequence of i.i.d. random variables,

$$\frac{1}{n}\sum_{i=1}^n (W_i - \hat{\pi})\boldsymbol{X}_i \to_p \mathbb{E}[(W_i - \hat{\pi})\boldsymbol{X}_i] = \mathbb{E}[W_i - \hat{\pi}]\,\mathbb{E}[\boldsymbol{X}_i] = 0.$$

Finally, $\frac{n_1}{n} \to_p \pi$ and $\frac{n_0}{n} \to_p 1-\pi$. Therefore,

$$\frac{n^2}{n_0 n_1}(n^{-1}d_1)^T \hat{\Sigma}_{\boldsymbol{XX}}^{-1}(n^{-1}d_1) \to_p \frac{1}{\pi(1-\pi)} \left(\mathbb{E}\left[(W_i - \hat{\pi})\boldsymbol{X}_i\right]\right)^2 \Sigma_{\boldsymbol{XX}}^{-1} = 0.$$

Next, we aim to establish convergence in probability of $\hat{h}_n^{(w)}(\boldsymbol{X}_i) = \hat{\Sigma}_{\boldsymbol{XY}}\hat{\Sigma}_{\boldsymbol{XX}}^{-1}$ to $h^{(w)}(\boldsymbol{X}_i) = \Sigma_{\boldsymbol{Xf}}\Sigma_{\boldsymbol{XX}}^{-1}\boldsymbol{X}_i$. This entails making sure that $\hat{\Sigma}_{\boldsymbol{XX}} \to_p \Sigma_{\boldsymbol{XX}}$ and $\hat{\Sigma}_{\boldsymbol{XY}} \to_p \Sigma_{\boldsymbol{Xf}}$. The former is well-known, since $\{\boldsymbol{X}_i\}$ are i.i.d. For the latter, by definition

$$\begin{aligned}
\hat{\Sigma}_{\boldsymbol{XY}} &= \frac{1}{n}\sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(Y_i - \bar{Y}) \\
&= \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i Y_i - \left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i\right)\left(\frac{1}{n}\sum_{i=1}^n Y_i\right)
\end{aligned}$$

Consider the average of the outcomes. By Assumption 3.7, we can write the observed outcome Y_i using the Taylor expansion around π as

$$\begin{aligned}
Y_i &= f(W_i, M_i/N_i, \boldsymbol{X}_i, U_i; \epsilon_i) \\
&= f(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i) \\
&\quad + f'(W_i, \pi_i^*, \boldsymbol{X}_i, U_i; \epsilon_i)\left(\frac{M_i}{N_i} - \pi\right)
\end{aligned}$$

for some π_i^* between π and M_i/N_i.

First, by Hölder's inequality and Lemma 15 of Li and Wager (2022)

$$\begin{aligned}
\mathbb{E}\left[f'(W_i, \pi_i^*, \boldsymbol{X}_i, U_i; \epsilon_i)\left(\frac{M_i}{N_i} - \pi\right)\right] &\le \mathbb{E}\left[f'(W_i, \pi_i^*, \boldsymbol{X}_i, U_i; \epsilon_i)^2\right]\mathbb{E}\left[\left(\frac{M_i}{N_i} - \pi\right)^2\right] \\
&\le B^2 \frac{C_1}{n\rho_n c_l}.
\end{aligned}$$

Next, consider the second moment:

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n f'(W_i, \pi_i^*, \boldsymbol{X}_i, U_i; \epsilon_i)\left(\frac{M_i}{N_i} - \pi\right)\right)^2\right] \\
&\quad = \frac{1}{n}\mathbb{E}\left[f'(W_i, \pi_i^*, \boldsymbol{X}_i, U_i; \epsilon_i)^2\left(\frac{M_i}{N_i} - \pi\right)^2\right] \\
&\qquad + \frac{n-1}{n}\mathbb{E}\left[f'(W_i, \pi_i^*, \boldsymbol{X}_i, U_i; \epsilon_i)\left(\frac{M_i}{N_i} - \pi\right) f'(W_j, \pi_j^*, \boldsymbol{X}_j, U_j; \epsilon_i)\left(\frac{M_j}{N_j} - \pi\right)\right] \qquad (i \ne j).
\end{aligned}$$

Since $|f'| \le B$, $\mathbb{E}\left[f'(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)^2\left(\frac{M_i}{N_i} - \pi\right)^2\right] \le B^2\mathbb{E}\left[\left(\frac{M_i}{N_i} - \pi\right)^2\right] \le \frac{B^2 C_1}{n\rho_n c_l}$. By

Hölder's Inequality,

$$\begin{aligned}
&\mathbb{E}\left[f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right)f'(W_j,\pi_j^*,\boldsymbol{X}_j,U_j)\left(\frac{M_j}{N_j}-\pi\right)\right]\\
&\quad\le \mathbb{E}\left[f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)^2\left(\frac{M_i}{N_i}-\pi\right)^2\right]\\
&\quad\le \frac{B^2C_1}{n\rho_n c_l}.
\end{aligned}$$

So,

$$\frac{1}{n}\sum_{i=1}^n f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right)=O_p(1/\sqrt{n\rho_n})=o_p(1)$$

since $n\rho_n\to\infty$ as $n\to\infty$. I.e.

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n Y_i &= \frac{1}{n}\sum_{i=1}^n f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)+\frac{1}{n}\sum_{i=1}^n f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right)\\
&= \frac{1}{n}\sum_{i=1}^n f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)+o_p(1).
\end{aligned}$$

and therefore $\frac{1}{n}\sum_{i=1}^n Y_i\to_p \frac{1}{n}\sum_{i=1}^n f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)$, which is an average of iid. random variables, and hence $\frac{1}{n}\sum_{i=1}^n Y_i\to_p \mathbb{E}\left[f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)\right]$.

The last piece of the puzzle is to show that $\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_iY_i$ converges in probability to $\mathbb{E}\left[\boldsymbol{X}_if(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)\right]$. We will use the same approach as above, and consider each covariate separately, since convergence in probability of the vector $\frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_iY_i$ to a limit is equivalent to element-wise convergence. In the following, let $X_{i,k}$ be the value of the k'th covariate for the i'th individual. We rewrite the sample average as

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n X_{i,k}Y_i &= \frac{1}{n}\sum_{i=1}^n X_{i,k}f(W_i,M_i/N_i,\boldsymbol{X}_i,U_i;\epsilon_i)\\
&= \frac{1}{n}\sum_{i=1}^n X_{i,k}f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)\\
&\quad+\frac{1}{n}\sum_{i=1}^n X_{i,k}f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right).
\end{aligned}$$

Once again by Hölder's inequality,

$$\begin{aligned}
&\mathbb{E}\left[X_{i,k}f'(W_i,\pi_i^*,\boldsymbol{X}_i,U_i;\epsilon_i)\left(\frac{M_i}{N_i}-\pi\right)\right]\\
&\quad\le B^2\mathbb{E}\left[X_{i,k}^2\right]\mathbb{E}\left[\left(\frac{M_i}{N_i}-\pi\right)^2\right]\\
&\quad\le B^2\nu_k\frac{C}{n\rho_n c_l},
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_{i,k} f'(W_i, \pi_i^*, \boldsymbol{X}_i, U_i; \epsilon_i)\left(\frac{M_i}{N_i} - \pi\right)\right)^2\right] \\
&\quad = \frac{1}{n}\mathbb{E}\left[X_{i,k}^2 f'(W_i, \pi_i^*, \boldsymbol{X}_i, U_i; \epsilon_i)^2\left(\frac{M_i}{N_i} - \pi\right)^2\right] \\
&\qquad + \frac{n-1}{n}\mathbb{E}\left[X_{i,k} f'(W_i, \pi_i^*, \boldsymbol{X}_i, U_i; \epsilon_i)\left(\frac{M_i}{N_i} - \pi\right) X_{j,k} f'(W_j, \pi_j^*, \boldsymbol{X}_j, U_j)\left(\frac{M_j}{N_j} - \pi\right)\right] \\
&\quad \leq \frac{B^2\nu_k C_1}{n\rho_n c_l} + \frac{n-1}{n}\frac{B^2\nu_k C_1}{n\rho_n c_l} \\
&= O(1/(n\rho_n))
\end{aligned}$$

where $\nu_k > 0$ such that $\mathbb{E}\left[X_{i,k}^2\right] \leq \nu_k$ (which is guaranteed to exist by the assumption of finite variance of the covariates). So,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} X_{i,k} f'(W_i, \pi_i^*, \boldsymbol{X}_i, U_i; \epsilon_i)\left(\frac{M_i}{N_i} - \pi\right) \\
&\qquad = \mathbb{E}\left[X_{i,k} f'(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)\left(\frac{M_i}{N_i} - \pi\right)\right] + O_p(1/(n\rho_n)) \\
&\qquad = o_p(1).
\end{aligned}$$

This completes the proof that the k'th element of $\hat{\Sigma}_{\boldsymbol{XY}}$ converges in probability to $\mathbb{E}\left[X_{i,k} f(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)\right]$, and therefore

$$\hat{\Sigma}_{\boldsymbol{XY}} \to_p \mathbb{E}\left[\boldsymbol{X}_i f(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)\right] - \mathbb{E}\left[\boldsymbol{X}_i\right]\mathbb{E}\left[f(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)\right].$$

Since $\hat{Y}_{i,n}^{(w,\boldsymbol{w}_{-i})} = \hat{h}_n^{(w)}(\boldsymbol{X}_i)$ and $\widetilde{Y}_i^{(w)} = h^{(w)}(\boldsymbol{X}_i)$ are functions of only $\boldsymbol{X}_i$, (a) of Lemma 3.1 holds. Note that $\hat{h}_n^{(w)}$ and $h^{(w)}$ belong to the parametric class of functions $\mathcal{D} = \{f_\theta(\boldsymbol{x}) = \theta^T\boldsymbol{x} : \theta \in \mathbb{R}^p\}$, which is Lipschitz in θ since

$$|f_{\theta_1}(\boldsymbol{x}) - f_{\theta_2}(\boldsymbol{x})| \leq ||\theta_1^T - \theta_2^T|| \cdot ||\boldsymbol{X}_i||.$$

$\mathbb{E}\left[||\boldsymbol{X}_i||^2\right] < \infty$ since $\mathrm{Var}\left[\boldsymbol{X}_i\right] < \infty$, and therefore $\mathcal{D}$ is P-Donsker (see, for example, Example 19.7 of van der Vaart (1998)).

Finally, we see

$$\begin{aligned}
\int\left(\hat{h}_n^{(w)}(x) - h^{(w)}(x)\right)^2 dP_{\boldsymbol{X}_i}(x) &= \int\left(\left[\hat{\Sigma}_{\boldsymbol{XY}}^T\hat{\Sigma}_{\boldsymbol{XX}}^{-1} - \Sigma_{\boldsymbol{Xf}}^T\Sigma_{\boldsymbol{XX}}^{-1}\right]\boldsymbol{X}_i\right)^2 dP_{\boldsymbol{X}_i}(x) \\
&\leq ||\hat{\Sigma}_{\boldsymbol{XY}}^T\hat{\Sigma}_{\boldsymbol{XX}}^{-1} - \Sigma_{\boldsymbol{Xf}}^T\Sigma_{\boldsymbol{XX}}^{-1}||^2\int||\boldsymbol{X}_i||^2 dP_{\boldsymbol{X}_i}(x).
\end{aligned}$$

$\mathrm{Var}(\boldsymbol{X}_i) < \infty$ implies $\mathbb{E}\left[||\boldsymbol{X}_i||^2\right] < \infty$, and since $\hat{\Sigma}_{\boldsymbol{XY}}^T\hat{\Sigma}_{\boldsymbol{XX}}^{-1} - \Sigma_{\boldsymbol{Xf}}^T\Sigma_{\boldsymbol{XX}}^{-1} \to_p 0$, we conclude $\int\left(\hat{h}_n^{(w)}(x) - h^{(w)}(x)\right)^2 dP_{\boldsymbol{X}_i}(x) \to_p 0$. I.e. the conditions of Lemma 3.1 are satisfied, and

therefore

$$\sqrt{n}\left(\hat{\beta}^{\text{ANCOVA}} - \bar{\tau}_{\text{DIR}}\right) \to_d N\left(0, \pi(1-\pi)\left(\text{Var}\left[R_i + Q_i - A_i\right] + \left(\mathbb{E}\left[Q_i\right]\right)^2\right)\right).$$

Since $A_i \perp U_i$, we can write $\text{Var}\left[R_i + Q_i - A_i\right] = \text{Var}\left[R_i + Q_i\right] + \text{Var}\left[A_i\right] - 2\text{Cov}\left[R_i, A_i\right]$, which gives us the desired result.

□

Corollary 3.1. *Consider the ANCOVA model in Equation 3.11 where $X_i \in \mathbb{R}$. Let $\rho_{W_i} = Cov[f(W_i, \pi, X_i, U_i), X_i]$, $\rho_0 = Cov[f(0, \pi, X_i, U_i), X_i]$, and $\rho_0 = Cov[f(0, \pi, X_i, U_i), X_i]$. We can characterize the relationship between $Var\left[\hat{\beta}^{ANCOVA}\right]$ and $Var\left[\hat{\tau}^{HA}\right]$ as follows:*

i. $Var\left[\hat{\beta}^{ANCOVA}\right] < Var\left[\hat{\tau}^{HA}\right]$ *if* $\rho_{W_i} \neq 0$ *and* $\frac{(1-\pi)\rho_1 + \pi\rho_0}{\rho_{W_i}} > \frac{1}{2}$.

ii. $Var\left[\hat{\beta}^{ANCOVA}\right] > Var\left[\hat{\tau}^{HA}\right]$ *if* $\rho_{W_i} \neq 0$ *and* $\frac{(1-\pi)\rho_1 + \pi\rho_0}{\rho_{W_i}} < \frac{1}{2}$.

iii. $Var\left[\hat{\beta}^{ANCOVA}\right] = Var\left[\hat{\tau}^{HA}\right]$ *if either X_i and f are uncorrelated, i.e. $\rho_{W_i} = 0$, or* $\frac{(1-\pi)\rho_1 + \pi\rho_0}{\rho_{W_i}} = \frac{1}{2}$.

Proof of 3.1. When only a single covariate is included,

$$\text{Var}\left[A_i\right] = \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)^2 \frac{\text{Cov}\left[f(W_i, \pi, X_i, U_i; \epsilon_i), X_i\right]^2}{\text{Var}\left[X_i\right]}$$

$$\text{Cov}\left[A_i, R_i\right] = \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right) \frac{\text{Cov}\left[f(W_i, \pi, X_i, U_i; \epsilon_i), X_i\right]}{\text{Var}\left[X_i\right]} \left(\frac{\text{Cov}\left[f(1, \pi, X_i, U_i; \epsilon_i), X_i\right]}{\pi} + \frac{\text{Cov}\left[f(0, \pi, X_i, U_i; \epsilon_i), X_i\right]}{1-\pi}\right).$$

This immediately gives us (iii) of the result: if $\text{Cov}\left[f(W_i, \pi, X_i, U_i; \epsilon_i), X_i\right] = 0$, then $\text{Var}\left[A_i\right] - 2\text{Cov}\left[A_i, R_i\right] = 0$, and therefore $\text{Var}\left[\hat{\tau}^{\text{HA}}\right] - \text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] = 0$.

Now, assume $\text{Cov}\left[f(W_i, \pi, X_i, U_i; \epsilon_i), X_i\right] \neq 0$. Then $\text{Var}\left[A_i\right] > 0$. We know from Proposition 3.2 that $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] - \text{Var}\left[\hat{\tau}^{\text{HA}}\right] = \text{Var}\left[A_i\right] - 2\text{Cov}\left[A_i, R_i\right]$. So, $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] < \text{Var}\left[\hat{\tau}^{\text{HA}}\right]$ if and only if $\text{Var}\left[A_i\right] - 2\text{Cov}\left[A_i, R_i\right] < 0$, which is equivalent to $\frac{\text{Cov}[A_i, R_i]}{\text{Var}[A_i]} > \frac{1}{2}$ when $\text{Var}\left[A_i\right] > 0$. Similarly, $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] > \text{Var}\left[\hat{\tau}^{\text{HA}}\right]$ if and only if $\frac{\text{Cov}[A_i, R_i]}{\text{Var}[A_i]} < \frac{1}{2}$ when $\text{Var}\left[A_i\right] > 0$.

So, consider the ratio $\frac{\text{Cov}[A_i,R_i]}{\text{Var}[A_i]}$.

$$\begin{aligned}\frac{\text{Cov}\left[A_i,R_i\right]}{\text{Var}\left[A_i\right]} &= \left(\frac{1}{\pi}+\frac{1}{1-\pi}\right)^{-1}\text{Cov}\left[f(W_i,\pi,X_i,U_i;\epsilon_i),X_i\right]^{-1} \\ &\qquad \left(\frac{\text{Cov}\left[f(1,\pi,X_i,U_i;\epsilon_i),X_i\right]}{\pi}+\frac{\text{Cov}\left[f(0,\pi,X_i,U_i;\epsilon_i),X_i\right]}{1-\pi}\right) \\ &= \pi(1-\pi)\frac{(1-\pi)\text{Cov}\left[f(1,\pi,X_i,U_i;\epsilon_i),X_i\right]+\pi\text{Cov}\left[f(0,\pi,X_i,U_i;\epsilon_i),X_i\right]}{\pi(1-\pi)\text{Cov}\left[f(W_i,\pi,X_i,U_i;\epsilon_i),X_i\right]} \\ &= \frac{(1-\pi)\text{Cov}\left[f(1,\pi,X_i,U_i;\epsilon_i),X_i\right]+\pi\text{Cov}\left[f(0,\pi,X_i,U_i;\epsilon_i),X_i\right]}{\pi\text{Cov}\left[f(1,\pi,X_i,U_i;\epsilon_i),X_i\right]+(1-\pi)\text{Cov}\left[f(0,\pi,X_i,U_i;\epsilon_i),X_i\right]}\end{aligned}$$

□

Proposition 3.3. *Under Assumptions 3.1, 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7, $\hat{\beta}^{ANCOVA2}$ is consistent for $\bar{\tau}_{DIR}$, and asymptotically normal in that*

$$\sqrt{n}\left(\hat{\beta}^{ANCOVA2}-\bar{\tau}_{DIR}\right)\to_d N\left(0,\pi(1-\pi)\left(Var\left[\hat{\tau}^{HA}\right]+Var[A_i]-2Cov[A_i,R_i]\right)\right),$$

where

$$\begin{aligned}Var\left[\hat{\tau}^{HA}\right] &= Var[R_i+Q_i]+\left(\mathbb{E}\left[Q_i\right]\right)^2, \\ R_i &= \frac{f(1,X_i,U_i,\pi;\epsilon_i)}{\pi}+\frac{f(0,X_i,U_i,\pi;\epsilon_i)}{1-\pi}, \\ A_i &= \left(\frac{1}{\pi}+\frac{1}{1-\pi}\right)\left[\pi\Sigma^{(0)}_{\boldsymbol{X}f}+(1-\pi)\Sigma^{(1)}_{\boldsymbol{X}f}\right]^T\Sigma^{-1}_{\boldsymbol{XX}}\boldsymbol{X}_i, \\ Q_i &= \mathbb{E}\left[\frac{G(U_i,U_j)(f'(1,\pi,\boldsymbol{X}_j,U_j)-f'(0,\pi,\boldsymbol{X}_j,U_j))}{\mathbb{E}\left[G(U_i,U_j)|U_j\right]}\middle|U_i\right]\end{aligned}$$

Proof of 3.3. This proof is similar to the proof of Proposition 3.2. We will show that $\hat{\beta}^{\text{ANCOVA2}}$ is asymptotically equivalent to $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$ with

$$\hat{Y}^{(w,\boldsymbol{w}_{-i})}_{i,n}=\left[\widehat{\Sigma}_{\boldsymbol{X}Y}+\frac{(1-2\pi)}{\pi(1-\pi)}\widehat{\Sigma}_{\boldsymbol{X}YW}\right]^T\hat{\Sigma}^{-1}_{\boldsymbol{XX}}\boldsymbol{X}_i$$

which converges to

$$\widetilde{Y}^{(w)}_i=\left[\pi\Sigma^{(0)}_{\boldsymbol{X}f}+(1-\pi)\Sigma^{(1)}_{\boldsymbol{X}f}\right]^T\Sigma^{-1}_{\boldsymbol{XX}}\boldsymbol{X}_i,$$

Again, we use Lemma 3.1 to get the desired result.

First, recall that we can rewrite $\hat{\beta}^{\text{ANCOVA2}}$ as

$$\begin{aligned}&\hat{\beta}^{\text{ANCOVA2}} \\ &\quad=\left\{1-\frac{n^2}{n_0n_1}(n^{-1}d_2)^TD^{-1}(n^{-1}d_2)\right\}^{-1} \\ &\qquad\left\{\hat{\tau}^{\text{HA}}-\frac{n}{n_0n_1}d_2^TD^{-1}\begin{pmatrix}\widehat{\Sigma}_{\boldsymbol{X}Y}\\ \widehat{\Sigma}_{\boldsymbol{X}YW}\end{pmatrix}\right\},\end{aligned}$$

where

$$d_2 = \left(\left(\sum_{i=1}^n (W_i - \hat{\pi}) \boldsymbol{X}_i \right)^T, \sum_{i=1}^n (W_i - \hat{\pi})^2 (\boldsymbol{X}_i - \bar{\boldsymbol{X}}^T) \right)^T$$

$$\hat{\Sigma}_{\boldsymbol{X}Y} = \frac{1}{n} \sum_{i=1}^n (\boldsymbol{X}_i - \bar{X})(Y_i - \bar{Y})$$

$$\hat{\Sigma}_{\boldsymbol{X}YW} = \frac{1}{n} \sum_{i=1}^n (\boldsymbol{X}_i - \bar{X})(Y_i - \bar{Y})(W_i - \hat{\pi})$$

$$D = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T & \frac{1}{n} \sum_{i=1}^n (W_i - \hat{\pi})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T \\ \frac{1}{n} \sum_{i=1}^n (W_i - \hat{\pi})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T & \frac{1}{n} \sum_{i=1}^n (W_i - \hat{\pi})^2 (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T \end{pmatrix}.$$

The first step in this proof is to show $\left\{ 1 - \frac{n^2}{n_0 n_1} (n^{-1} d_2)^T D^{-1} (n^{-1} d_2) \right\}^{-1} \to_p 1$. To this end, first note that D converges to a diagonal matrix since the off diagonal elements converge to 0 in probability:

$$\frac{1}{n} \sum_{i=1}^n (W_i - \hat{\pi})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T \to_p \mathbb{E}\left[(W_i - \pi)(\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])(\boldsymbol{X}_i - \mathbb{E}[\boldsymbol{X}_i])^T \right],$$

This follows by independence of W_i and the covariates $\boldsymbol{X}_i$. Also by independence of W_i and the covariates, $\frac{1}{n} \sum_{i=1}^n (W_i - \hat{\pi})^2 (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T \to_p \pi(1-\pi)\Sigma_{\boldsymbol{XX}}$, and since $\frac{1}{n} \sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T \to_p \Sigma_{\boldsymbol{XX}}$, we see that

$$D \to_p \widetilde{D} = \begin{pmatrix} \Sigma_{\boldsymbol{XX}} & 0 \\ 0 & \pi(1-\pi)\Sigma_{\boldsymbol{XX}} \end{pmatrix}.$$

Since $\frac{1}{n} d_2 \to_p 0$, $n_1/n \to_p \pi$, and $n_0/n \to_p 1 - \pi$, $\left\{ 1 - \frac{n^2}{n_0 n_1} (n^{-1} d_2)^T D^{-1} (n^{-1} d_2) \right\}^{-1} \to_p 1$.

Now, note that each element of D converges to its limit at a rate of $1/\sqrt{n}$. I.e. $D - \widetilde{D} = O_p(1/\sqrt{n})$. This implies $D^{-1} - \widetilde{D}^{-1} = O_p(1/\sqrt{n})$.

Also, $\frac{n}{n_0 n_1} = o_p(1)$, $\frac{1}{n} d_2 = O_p(1/\sqrt{n})$, and $\hat{\Sigma}_{\boldsymbol{X}Y} = O_p(1/\sqrt{n})$ and $\hat{\Sigma}_{\boldsymbol{X}YW} = O_p(1/\sqrt{n})$ by Proposition 3.6. So,

$$\begin{aligned} &\frac{n}{n_0 n_1} d_2^T D^{-1} \begin{pmatrix} \widehat{\Sigma}_{\boldsymbol{X}Y} \\ \widehat{\Sigma}_{\boldsymbol{X}YW} \end{pmatrix} \\ &\quad = \frac{n}{n_0 n_1} d_2^T \widetilde{D}^{-1} \begin{pmatrix} \widehat{\Sigma}_{\boldsymbol{X}Y} \\ \widehat{\Sigma}_{\boldsymbol{X}YW} \end{pmatrix} + \frac{n}{n_0 n_1} d_2^T (D^{-1} - \widetilde{D}^{-1}) \begin{pmatrix} \widehat{\Sigma}_{\boldsymbol{X}Y} \\ \widehat{\Sigma}_{\boldsymbol{X}YW} \end{pmatrix} \\ &\quad = \frac{n}{n_0 n_1} d_2^T \widetilde{D}^{-1} \begin{pmatrix} \widehat{\Sigma}_{\boldsymbol{X}Y} \\ \widehat{\Sigma}_{\boldsymbol{X}YW} \end{pmatrix} + o_p(1/\sqrt{n}) \end{aligned}$$

Since $\sum_{i=1}^n (W_i - \hat{\pi})^2 (\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T = \sum_{i=1}^n (W_i - \hat{\pi}) \boldsymbol{X}_i^T (1 - 2\hat{\pi})$, we see that

$$
\begin{aligned}
& d_2^T \widetilde{D}^{-1} \begin{pmatrix} \widehat{\Sigma}_{\boldsymbol{X}Y} \\ \widehat{\Sigma}_{\boldsymbol{X}YW} \end{pmatrix} \\
&= \left(\textstyle\sum_{i=1}^n (W_i - \hat{\pi}) \boldsymbol{X}_i^T \quad \sum_{i=1}^n (W_i - \hat{\pi}) \boldsymbol{X}_i^T (1 - 2\hat{\pi}) \right) \begin{pmatrix} \Sigma_{\boldsymbol{XX}}^{-1} & 0 \\ 0 & (\pi(1-\pi))^{-1} \Sigma_{\boldsymbol{XX}}^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\Sigma}_{\boldsymbol{X}Y} \\ \widehat{\Sigma}_{\boldsymbol{X}YW} \end{pmatrix} \\
&= \sum_{i=1}^n (W_i - \hat{\pi}) \boldsymbol{X}_i^T \Sigma_{\boldsymbol{XX}}^{-1} \left(\hat{\Sigma}_{\boldsymbol{X}Y} + \frac{1 - 2\hat{\pi}}{\pi(1-\pi)} \hat{\Sigma}_{\boldsymbol{X}YW} \right) \\
&= \sum_{i=1}^n (W_i - \hat{\pi}) \left(\hat{\Sigma}_{\boldsymbol{X}Y} + \frac{1 - 2\hat{\pi}}{\pi(1-\pi)} \hat{\Sigma}_{\boldsymbol{X}YW} \right)^T \Sigma_{\boldsymbol{XX}}^{-1} \boldsymbol{X}_i.
\end{aligned}
$$

I.e.

$$
\begin{aligned}
& \hat{\beta}^{\text{ANCOVA2}} \\
&= \left\{ 1 - \frac{n^2}{n_0 n_1} (n^{-1} d_2)^T D^{-1} (n^{-1} d_2) \right\}^{-1} \\
&\qquad \left\{ \hat{\tau}^{\text{HA}} - \frac{n}{n_0 n_1} \sum_{i=1}^n (W_i - \hat{\pi}) \left(\hat{\Sigma}_{\boldsymbol{X}Y} + \frac{1 - 2\hat{\pi}}{\pi(1-\pi)} \hat{\Sigma}_{\boldsymbol{X}YW} \right)^T \Sigma_{\boldsymbol{XX}}^{-1} \boldsymbol{X}_i \right\} \\
&\quad + o_p(1/\sqrt{n}),
\end{aligned}
$$

Note that $\hat{\tau}^{\text{HA}} - \frac{n}{n_0 n_1} \sum_{i=1}^n (W_i - \hat{\pi}) \left(\hat{\Sigma}_{\boldsymbol{X}Y} + \frac{1-2\hat{\pi}}{\pi(1-\pi)} \hat{\Sigma}_{\boldsymbol{X}YW} \right)^T \Sigma_{\boldsymbol{XX}}^{-1} \boldsymbol{X}_i$ is exactly of the form of the estimator $\hat{\tau}^{\text{HA}}_{\text{AIPW}}$.

Since we already know that $\hat{\Sigma}_{\boldsymbol{X}Y} \to_p \Sigma_{\boldsymbol{X}f}$, the result follows if $\hat{\Sigma}_{\boldsymbol{X}YW} \to_p \Sigma_{\boldsymbol{X}fW} = \mathbb{E}\left[(\boldsymbol{X_i} - \mathbb{E}[\boldsymbol{X}])(W_i - \pi)(f(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i) - \mathbb{E}[f(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)]) \right]$, since $\Sigma_{\boldsymbol{X}f} = (1 - \pi)\Sigma_{\boldsymbol{X}f}^{(0)} + \pi \Sigma_{\boldsymbol{X}f}^{(1)}$ and $\Sigma_{\boldsymbol{X}fW} = \pi(1-\pi)\left(\Sigma_{\boldsymbol{X}f}^{(1)} - \Sigma_{\boldsymbol{X}f}^{(0)} \right)$, and so

$$
\Sigma_{\boldsymbol{X}f} + \frac{1 - 2\pi}{\pi(1-\pi)} \Sigma_{\boldsymbol{X}fW} = \pi \Sigma_{\boldsymbol{X}f}^{(0)} + (1 - \pi) \Sigma_{\boldsymbol{X}f}^{(1)}
$$

To show that $\hat{\Sigma}_{\boldsymbol{X}YW} \to_p \Sigma_{\boldsymbol{X}fW}$, note that

$$
\hat{\Sigma}_{\boldsymbol{X}YW} = \frac{1}{n} \sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(W_i - \hat{\pi}) Y_i - \frac{1}{n} \sum_{i=1}^n (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(W_i - \hat{\pi}) \bar{Y}.
$$

We have previously shown that $\bar{Y} = \bar{f} + o_p(1)$, where $\bar{f} = \frac{1}{n} \sum_{i=1}^n f(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i)$. Since we can use a Taylor expansion to get

$$
Y_i = f(W_i, \pi, \boldsymbol{X}_i, U_i; \epsilon_i) + \left(\frac{M_i}{N_i} - \pi \right) f'(W_i, \pi^*, \boldsymbol{X}_i, U_i; \epsilon_i),
$$

we see that the first average is

$$\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})(W_i-\hat{\pi})f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)$$
$$+\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})(W_i-\hat{\pi})\left(\frac{M_i}{N_i}-\pi\right)f'(W_i,\pi^*,\boldsymbol{X}_i,U_i;\epsilon_i).$$

Using Hölder's inequality, we see that the expected value of the second average above is $o_p(1)$. Considering the second moment of the same average, and once again applying Hölder's inequality, we also note that the second moment i $o_p(1)$, and therefore

$$\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})(W_i-\hat{\pi})\left(\frac{M_i}{N_i}-\pi\right)f'(W_i,\pi^*,\boldsymbol{X}_i,U_i;\epsilon_i)=o_p(1).$$

So,

$$\hat{\Sigma}_{\boldsymbol{X}YW}=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i-\bar{\boldsymbol{X}})(W_i-\hat{\pi})(f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)-\bar{f})+o_p(1)$$
$$\to_p \mathbb{E}\left[(\boldsymbol{X_i}-\mathbb{E}\left[\boldsymbol{X}\right])(W_i-\pi)(f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)-\mathbb{E}\left[f(W_i,\pi,\boldsymbol{X}_i,U_i;\epsilon_i)\right])\right]$$

□

Corollary 3.2. *If the ANCOVA model in Equation 3.13 includes a single covariate, then* $Var\left[\hat{\beta}^{ANCOVA2}\right] \leq Var\left[\hat{\tau}^{HA}\right]$ *with equality if and only if* $\pi Cov[f(0,\pi,X_i,U_i;\epsilon_i),X_i]+(1-\pi)Cov[f(1,\pi,X_i,U_i;\epsilon_i),X_i]=0$.

Proof of 3.2. When only a single covariate and its interaction with treatment assignment are included in the ANCOVA model,

$$\mathrm{Var}\left[A_i\right]=\left(\frac{1}{\pi}+\frac{1}{1-\pi}\right)^2\frac{(\pi\mathrm{Cov}\left[f(0,\pi,X_i,U_i;\epsilon_i),X_i\right]+(1-\pi)\mathrm{Cov}\left[f(1,\pi,X_i,U_i;\epsilon_i),X_i\right])^2}{\mathrm{Var}\left[X_i\right]}$$

$$\begin{aligned}\mathrm{Cov}\left[A_i,R_i\right]&=\left(\frac{1}{\pi}+\frac{1}{1-\pi}\right)\frac{\pi\mathrm{Cov}\left[f(0,\pi,X_i,U_i;\epsilon_i),X_i\right]+(1-\pi)\mathrm{Cov}\left[f(1,\pi,X_i,U_i;\epsilon_i),X_i\right]}{\mathrm{Var}\left[X_i\right]}\\&\qquad\left(\frac{\mathrm{Cov}\left[f(1,\pi,X_i,U_i;\epsilon_i),X_i\right]}{\pi}+\frac{\mathrm{Cov}\left[f(0,\pi,X_i,U_i;\epsilon_i),X_i\right]}{1-\pi}\right)\\&=\left(\frac{1}{\pi}+\frac{1}{1-\pi}\right)\frac{(\pi\mathrm{Cov}\left[f(0,\pi,X_i,U_i;\epsilon_i),X_i\right]+(1-\pi)\mathrm{Cov}\left[f(1,\pi,X_i,U_i;\epsilon_i),X_i\right])^2}{\pi(1-\pi)\mathrm{Var}\left[X_i\right]}\\&=\left(\frac{1}{\pi}+\frac{1}{1-\pi}\right)^2\frac{(\pi\mathrm{Cov}\left[f(0,\pi,X_i,U_i;\epsilon_i),X_i\right]+(1-\pi)\mathrm{Cov}\left[f(1,\pi,X_i,U_i;\epsilon_i),X_i\right])^2}{\mathrm{Var}\left[X_i\right]}\end{aligned}$$

The result follows immediately: since $\mathrm{Var}\left[A_i\right]=\mathrm{Cov}\left[A_i,R_i\right]$, $\mathrm{Var}\left[A_i\right]=2\mathrm{Cov}\left[A_i,R_i\right]$ if and only if $\mathrm{Var}\left[A_i\right]=0$, which happens exactly when $\pi\mathrm{Cov}\left[f(0,\pi,X_i,U_i;\epsilon_i),X_i\right]+(1-\pi)\mathrm{Cov}\left[f(1,\pi,X_i,U_i;\epsilon_i),X_i\right]=0$. When this is not the case, $\mathrm{Var}\left[A_i\right]<2\mathrm{Cov}\left[A_i,R_i\right]$, and so $\mathrm{Var}\left[\hat{\beta}^{\mathrm{ANCOVA2}}\right]<\mathrm{Var}\left[\hat{\tau}^{\mathrm{HA}}_{\mathrm{AIPW}}\right]$.

□

3.D Asymptotic Variances

In general, the variance of the ANCOVA estimator when $X_i \in \mathbb{R}$ satisfy

$$\begin{aligned}\frac{\mathrm{Var}\left[\hat{\beta}^{\mathrm{ANCOVA}}\right]}{\pi(1-\pi)} &= \mathrm{Var}\left[R_i + Q_i - A_i\right] + \mathbb{E}\left[Q_i\right]^2 \\ &= \mathrm{Var}\left[R_i\right] + \mathrm{Var}\left[Q_i\right] + \mathrm{Var}\left[A_i\right] \\ &\quad + 2\mathrm{Cov}\left[R_i, Q_i\right] - 2\mathrm{Cov}\left[R_i, A_i\right] + \mathbb{E}\left[Q_i\right]^2 \\ &= \mathrm{Var}\left[R_i\right] + \mathrm{Var}\left[A_i\right] \\ &\quad + 2\mathrm{Cov}\left[R_i, Q_i\right] - 2\mathrm{Cov}\left[R_i, A_i\right] + \mathbb{E}\left[Q_i^2\right]\end{aligned}$$

since $A_i \perp Q_i$. We can further expand this to

$$\begin{aligned}\frac{\mathrm{Var}\left[\hat{\beta}^{\mathrm{ANCOVA}}\right]}{\pi(1-\pi)} &= \mathrm{Var}\left[R_i\right] + \mathrm{Var}\left[A_i\right] + 2\mathrm{Cov}\left[R_i, Q_i\right] - 2\mathrm{Cov}\left[R_i, A_i\right] + \mathbb{E}\left[Q_i\right]^2 \\ &= \frac{\mathrm{Var}\left[f_i(1)\right]}{\pi^2} + \frac{\mathrm{Var}\left[f_i(0)\right]}{(1-\pi)^2} + 2\frac{\mathrm{Cov}\left[f_i(1), f_i(0)\right]}{\pi(1-\pi)} \\ &\quad + \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)^2 \frac{\left(\mathrm{Cov}\left[f_i(1), X_i\right]\pi + \mathrm{Cov}\left[f_i(0), X_i\right](1-\pi)\right)^2}{\mathrm{Var}\left[X_i\right]} \\ &\quad + 2\left(\frac{\mathrm{Cov}\left[Q_i, f_i(1)\right]}{\pi} + \frac{\mathrm{Cov}\left[Q_i, f_i(0)\right]}{1-\pi}\right) \\ &\quad - 2\left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\frac{\mathrm{Cov}\left[f_i(1), X_i\right]\pi + \mathrm{Cov}\left[f_i(0), X_i\right](1-\pi)}{\mathrm{Var}\left[X_i\right]} \\ &\qquad \cdot \left(\frac{\mathrm{Cov}\left[f_i(1), X_i\right]}{\pi} + \frac{\mathrm{Cov}\left[f_i(0), X_i\right]}{1-\pi}\right) \\ &\quad + \mathbb{E}\left[Q_i^2\right],\end{aligned}$$

where we use the shorthand notation $f_i(w) = f(w, \pi, X_i, U_i; \epsilon_i)$. We can similarly write the asymptotic variance of the Hájek estimator as

$$\begin{aligned}\frac{\mathrm{Var}\left[\hat{\beta}^{\mathrm{ANCOVA}}\right]}{\pi(1-\pi)} &= \mathrm{Var}\left[R_i\right] + 2\mathrm{Cov}\left[R_i, Q_i\right] + \mathbb{E}\left[Q_i\right]^2 \\ &= \frac{\mathrm{Var}\left[f_i(1)\right]}{\pi^2} + \frac{\mathrm{Var}\left[f_i(0)\right]}{(1-\pi)^2} + 2\frac{\mathrm{Cov}\left[f_i(1), f_i(0)\right]}{\pi(1-\pi)} \\ &\quad + 2\left(\frac{\mathrm{Cov}\left[Q_i, f_i(1)\right]}{\pi} + \frac{\mathrm{Cov}\left[Q_i, f_i(0)\right]}{1-\pi}\right) \\ &\quad + \mathbb{E}\left[Q_i^2\right],\end{aligned}$$

While both of these expressions are long, they are useful when calculating the variance for specific graphon models and potential outcomes functions, as we will do below. We will throughout assume the covariace follows a standard normal distribution, i.e. $X_i \sim N(0,1)$, which in particular means $\mathbb{E}\left[X_i\right] = \mathbb{E}\left[X_i^3\right] = 0$, $\mathbb{E}\left[X_i^2\right] = \mathrm{Var}\left[X_i\right] = 1$, and $\mathbb{E}\left[X_i^4\right] = 3$.

3.D.1 Asymptotic Variances under Linear Potential Outcomes Function

Consider the potential outcomes function $f(w,p,x,u;\epsilon) = \beta_W w + \beta_X x + p + \beta_U u + \epsilon$. In this case, $Q_i = 0$ since $f'(1,p,x,u;\epsilon) - f'(0,p,x,u;\epsilon) = 0$. Also,

$$\text{Var}\left[f(w,\pi,X_i,U_i;\epsilon_i)\right] = \beta_X^2 \text{Var}\left[X_i\right] + \beta_U^2 \text{Var}\left[U_i\right] + \text{Var}\left[\epsilon\right] \quad \text{for } w \in \{0,1\},$$

$$\text{Cov}\left[f(1,\pi,X_i,U_i;\epsilon_i), f_i(0,\pi,X_i,U_i;\epsilon_i)\right] = \beta_X^2 \text{Var}\left[X_i\right] + \beta_U^2 \text{Var}\left[U_i\right] + \text{Var}\left[\epsilon\right]$$

$$\text{Cov}\left[f(w,\pi,X_i,U_i;\epsilon_i), X_i\right] = \beta_X \text{Var}\left[X_i\right] \quad \text{for } w \in \{0,1\}.$$

When this is plugged into the equations above, we see

$$\begin{aligned}
&\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right] \\
&= \pi(1-\pi)\left[\frac{\beta_X^2 \text{Var}\left[X_i\right] + \beta_U^2 \text{Var}\left[U_i\right] + \text{Var}\left[\epsilon\right]}{\pi^2(1-\pi)^2}\right. \\
&\quad + \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)^2 \beta_X^2 \text{Var}\left[X_i\right] \\
&\quad \left. - 2\left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)^2 \beta_X^2 \text{Var}\left[X_i\right]\right] \\
&= \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\left(\beta_U^2 \text{Var}\left[U_i\right] + \text{Var}\left[\epsilon_i\right]\right)
\end{aligned}$$

and

$$\begin{aligned}
\text{Var}\left[\hat{\tau}^{\text{HA}}\right] &= \pi(1-\pi)\left(\frac{1}{\pi^2} + \frac{1}{(1-\pi)^2} + 2\frac{1}{\pi(1-\pi)}\right)\left(\beta_X^2 \text{Var}\left[X_i\right] + \beta_U^2 \text{Var}\left[U_i\right] + \text{Var}\left[\epsilon\right]\right) \\
&= \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\left(\beta_X^2 \text{Var}\left[X_i\right] + \beta_U^2 \text{Var}\left[U_i\right] + \text{Var}\left[\epsilon\right]\right).
\end{aligned}$$

3.D.2 Asymptotic Variances under Quadratic Potential Outcomes Function

Consider the potential outcomes function $f(w,p,x,u;\epsilon) = \frac{1}{2}\left(\beta_W w + \beta_X x + p \cdot u\right)^2 + \epsilon_i/5$, and the graphon $G(U_i,U_j) = G_1(U_i,U_j)$. To get to the asymptotic variances, we start by considering some relevant quantities involving Q_i. Since $f_i'(1,p,x,u) - f_i'(0,p,x,u) = \beta_W u$

and $g(U_j) = \frac{2}{5}$,

$$\begin{aligned}
Q_i &= \mathbb{E}\left[G(U_i, U_j)\frac{f'(1, \pi, X_j, U_j; \epsilon_j) - f'(0, \pi, X_j, U_j; \epsilon_j)}{g(U_j)}\middle| U_i\right] \\
&= \frac{5}{2}\beta_W \mathbb{E}\left[U_j \cdot G_1(U_i, U_j)|U_i\right] \\
&= \frac{\beta_W}{2}\mathbb{E}\left[U_j|U_i\right] + \\
&\quad \beta_W\left(\frac{3}{2}\mathbb{1}\left[U_i < \tfrac{1}{3}\right] \cdot \mathbb{E}\left[U_j \cdot \mathbb{1}\left[U_j < \tfrac{1}{3}\right]\right] + \right. \\
&\qquad \frac{3}{2}\mathbb{1}\left[\tfrac{1}{3} \le U_i \le \tfrac{2}{3}\right] \cdot \mathbb{E}\left[U_j \cdot \mathbb{1}\left[\tfrac{1}{3} \le U_j \le \tfrac{2}{3}\right]\right] + \\
&\qquad \left. \frac{3}{2}\mathbb{1}\left[U_i > \tfrac{2}{3}\right] \cdot \mathbb{E}\left[U_j \cdot \mathbb{1}\left[U_j > \tfrac{2}{3}\right]\right]\right).
\end{aligned}$$

Since $U_j \sim \text{Uniform}[0, 1]$,

$$\mathbb{E}\left[U_j \mathbb{1}\left[U_j \in A\right]\right] = \mathbb{E}\left[U_j|A\right] \cdot P(U_j \in A) = \frac{a_1 + a_2}{2}(a_2 - a_1) = \frac{a_2^2 - a_1^2}{2}.$$

for any interval $A = [a_1, a_2]$ or $A = (a_1, a_2)$. So

$$\begin{aligned}
Q_i &= \beta_W\left(\frac{1}{4} + \frac{1}{12}\mathbb{1}\left[U_i < \tfrac{1}{3}\right] + \frac{1}{4}\mathbb{1}\left[\tfrac{1}{3} \le U_i \le \tfrac{2}{3}\right] + \frac{5}{12}\mathbb{1}\left[U_i > \tfrac{2}{3}\right]\right) \\
&= \beta_W\left(\frac{1}{3}\mathbb{1}\left[U_i < \tfrac{1}{3}\right] + \frac{1}{2}\mathbb{1}\left[\tfrac{1}{3} \le U_i \le \tfrac{2}{3}\right] + \frac{2}{3}\mathbb{1}\left[U_i > \tfrac{2}{3}\right]\right).
\end{aligned}$$

We can now find $\mathbb{E}\left[Q_i^2\right]$. Since any product of indicator functions in Q_i will be constantly 0,

$$\begin{aligned}
\mathbb{E}\left[Q_i^2\right] &= \beta_W^2\left(\frac{1}{9}\mathbb{E}\left[\mathbb{1}\left[U_i < \frac{1}{3}\right]\right] + \frac{1}{4}\mathbb{E}\left[\mathbb{1}\left[\tfrac{1}{3} \le U_i \le \tfrac{2}{3}\right]\right] + \frac{4}{9}\mathbb{E}\left[\mathbb{1}\left[U_i > \tfrac{2}{3}\right]\right]\right) \\
&= \frac{\beta_W^2}{3}\left(\frac{1}{9} + \frac{1}{4} + \frac{4}{9}\right) \\
&= \frac{29\beta_W^2}{108}.
\end{aligned}$$

Next, let us consider the quantities $\text{Var}\left[f_i(0)\right], \text{Var}\left[f_i(1)\right], \text{Cov}\left[f_i(0), f_i(1)\right]$. Remember,

$X_i \perp U_i$, and therefore $\text{Cov}\left[X_i, U_i\right] = 0$.

$$\begin{aligned}
\text{Var}\left[f_i(1)\right] &= \frac{1}{4}\text{Var}\left[\beta_X^2 X_i^2 + 2\beta_W\beta_X X_i + \pi^2 U_i^2 + 2\beta_W \pi U_i + 2\beta_X X_i \pi U_i\right] + \text{Var}\left[\epsilon_i\right] \\
&= \frac{1}{4}\beta_X^4 \text{Var}\left[X_i^2\right] + \beta_W^2\beta_X^2 \text{Var}\left[X_i\right] + \frac{\pi^4}{4}\text{Var}\left[U_i^2\right] + \beta_W^2\pi^2\text{Var}\left[U_i\right] + \beta_X^2\pi^2\text{Var}\left[X_i U_i\right] \\
&\quad + \beta_X^3\beta_W \text{Cov}\left[X_i^2, X_i\right] + \beta_X\pi\text{Cov}\left[X_i^2, X_i U_i\right] \\
&\quad + 2\cdot\beta_W\beta_X^2\pi\text{Cov}\left[X_i, X_i U_i\right] \\
&\quad + \pi^3\beta_W\text{Cov}\left[U_i^2, U_i\right] + \pi^3\beta_X\text{Cov}\left[U_i^2, X_i U_i\right] \\
&\quad + 2\cdot\beta_W\beta_X\pi\text{Cov}\left[U_i, U_i X_i\right] \\
&\quad + \text{Var}\left[\epsilon_i\right]/25.
\end{aligned}$$

Since, by assumptions, $\mathbb{E}\left[X_i\right] = \mathbb{E}\left[X_i^3\right] = 0$, $\mathbb{E}\left[X_i^4\right] = 3$, $U_i \sim \text{Uniform}[0,1]$, and $X_i \perp U_i$, a lot of these covariances are 0:

$$\begin{aligned}
\text{Var}\left[X_i^2\right] &= \mathbb{E}\left[X_i^4\right] - \mathbb{E}\left[X_i^2\right]^2 = 2 \\
\text{Var}\left[U_i^2\right] &= \mathbb{E}\left[U_i^4\right] - \mathbb{E}\left[U_i^2\right]^2 = \frac{1}{5} - \frac{1}{9} = \frac{4}{45} \\
\text{Var}\left[X_i U_i\right] &= \mathbb{E}\left[X_i^2\right]\mathbb{E}\left[U_i^2\right] - \mathbb{E}\left[X_i\right]^2\mathbb{E}\left[U_i\right]^2 = \frac{1}{3} \\
\text{Cov}\left[X_i^2, X_i\right] &= \mathbb{E}\left[X_i^3\right] - \mathbb{E}\left[X_i^2\right]\mathbb{E}\left[X_i\right] = 0, \\
\text{Cov}\left[X_i^2, X_i U_i\right] &= \mathbb{E}\left[X_i^3\right]\mathbb{E}\left[U_i\right] - \mathbb{E}\left[X_i^2\right]\mathbb{E}\left[X_i\right]\mathbb{E}\left[U_i\right] = 0, \\
\text{Cov}\left[X_i, X_i U_i\right] &= \mathbb{E}\left[X_i^2\right]\mathbb{E}\left[U_i\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[U_i\right] = \frac{1}{2} \\
\text{Cov}\left[U_i^2, U_i\right] &= \mathbb{E}\left[U_i^3\right] - \mathbb{E}\left[U_i^2\right]\mathbb{E}\left[U_i\right] = \frac{1}{4} - \frac{1}{3\cdot 2} = \frac{1}{12} \\
\text{Cov}\left[U_i^2, X_i U_i\right] &= \mathbb{E}\left[U_i^3\right]\mathbb{E}\left[X_i\right] - \mathbb{E}\left[U_i^2\right]\mathbb{E}\left[X_i\right]\mathbb{E}\left[U_i\right] = 0.
\end{aligned}$$

We plug these results into the expression above to get

$$\begin{aligned}
\text{Var}\left[f_i(1)\right] = \frac{1}{2}\beta_X^4 + \beta_W^2\beta_X^2 + \frac{\pi^4}{45} + \beta_W^2\frac{\pi^2}{12} + \beta_X^2\frac{\pi^2}{3} \\
+ \beta_W\beta_X^2\pi \\
+ \beta_W\frac{\pi^3}{12} \\
+ \text{Var}\left[\epsilon_i\right]/25.
\end{aligned}$$

The variance of $f(0, \pi, X_i, U_i; \epsilon_i)$ can be obtained similarly, or by realizing it is the same expression where $\beta_W = 0$. I.e.

$$\begin{aligned}
\text{Var}\left[f_i(0)\right] = \frac{1}{2}\beta_X^4 + \frac{\pi^4}{45} + \beta_X^2\frac{\pi^2}{3} \\
+ \text{Var}\left[\epsilon_i\right]/25.
\end{aligned}$$

Now, we turn our attention to $\text{Cov}\left[f_i(1), X_i\right]$ and $\text{Cov}\left[f_i(0), X_i\right]$:

$$\begin{aligned}
\text{Cov}\left[f_i(1), X_i\right] &= \frac{1}{2}\text{Cov}\left[\beta_X^2 X_i^2 + 2\beta_W\beta_X X_i + \pi^2 U_i^2 + 2\beta_W \pi U_i + 2\beta_X X_i \pi U_i, X_i\right] \\
&= \frac{1}{2}\beta_X^2 \text{Cov}\left[X_i^2, X_i\right] + \beta_W\beta_X \text{Var}\left[X_i\right] + \beta_X \pi \text{Cov}\left[X_i U_i, X_i\right] \\
&= \beta_W\beta_X + \beta_X\frac{\pi}{2}
\end{aligned}$$

$$\begin{aligned}
\text{Cov}\left[f_i(0), X_i\right] &= \frac{1}{2}\text{Cov}\left[\beta_X^2 X_i^2 + \pi^2 U_i^2 + 2\beta_X X_i \pi U_i, X_i\right] \\
&= \frac{1}{2}\beta_X^2 \text{Cov}\left[X_i^2, X_i\right] + \beta_X \pi \text{Cov}\left[X_i U_i, X_i\right] \\
&= \beta_X\frac{\pi}{2}.
\end{aligned}$$

Next, consider the covariance of $f(1, \pi, X_i, U_i; \epsilon_i)$ and $f(0, \pi, X_i, U_i; \epsilon_i)$:

$$\begin{aligned}
&\text{Cov}\left[f_i(1), f_i(0)\right] \\
&\quad= \text{Cov}\left[\frac{1}{2}\left(\beta_W^2 + \beta_X^2 X_i^2 + 2\beta_W\beta_X X_i + \pi^2 \cdot U_i^2 + 2(\beta_W + \beta_X X_i)\pi U_i\right) + \epsilon_i/5,\right. \\
&\qquad\left.\frac{1}{2}\left(\beta_X^2 X_i^2 + \pi^2 \cdot U_i^2 + 2\beta_X X_i \pi U_i\right) + \epsilon_i/5\right] \\
&\quad= \frac{1}{4}\beta_X^4 \text{Var}\left[X_i^2\right] + \frac{1}{2}\beta_X^3 \pi \text{Cov}\left[X_i^2, X_i U_i\right] \\
&\qquad+ \frac{1}{2}\beta_W\beta_X^3 \text{Cov}\left[X_i, X_i^2\right] + \beta_W\beta_X^2 \pi \text{Cov}\left[X_i, X_i U_i\right] \\
&\qquad+ \frac{1}{4}\pi^4 \text{Var}\left[U_i^2\right] + \frac{1}{2}\pi^3\beta_X \text{Cov}\left[U_i^2, X_i U_i\right] \\
&\qquad+ \frac{1}{2}\beta_X^3 \pi \text{Cov}\left[X_i U_i, X_i^2\right] + \frac{1}{2}\beta_X \pi^3 \text{Cov}\left[X_i U_i, U_i^2\right] + \beta_X^2 \pi^2 \text{Var}\left[X_i U_i\right] \\
&\qquad+ \text{Var}\left[\epsilon\right]/25 \\
&\quad= \frac{1}{2}\beta_X^4 + \beta_W\beta_X^2\frac{\pi}{2} + \frac{\pi^4}{45} + \beta_X^2\frac{\pi^2}{3} + \text{Var}\left[\epsilon\right]/25
\end{aligned}$$

Finally, we turn to the covariance of Q_i and $f_i(1)$ and $f_i(0)$. Consider the former first:

$$\begin{aligned}
\text{Cov}\left[Q_i, f_i(1)\right] &= \text{Cov}\left[Q_i, \frac{1}{2}\left(\beta_W^2 + \beta_X^2 X_i^2 + 2\beta_W\beta_X X_i + \pi^2 \cdot U_i^2 + 2(\beta_W + \beta_X X_i)\pi U_i\right) + \epsilon_i/5\right] \\
&= \text{Cov}\left[Q_i, \frac{1}{2}\left(\pi^2 \cdot U_i^2 + 2(\beta_W + \beta_X X_i)\pi U_i\right)\right] \\
&= \frac{1}{2}\pi^2 \text{Cov}\left[Q_i, U_i^2\right] + \beta_W \pi \text{Cov}\left[Q_i, U_i\right] + \beta_X \pi \text{Cov}\left[Q_i, X_i U_i\right].
\end{aligned}$$

The latter term is 0: $\text{Cov}\left[Q_i, X_i U_i\right] = \mathbb{E}\left[Q_i U_i\right]\mathbb{E}\left[X_i\right] - \mathbb{E}\left[Q_i\right]\mathbb{E}\left[U_i\right]\mathbb{E}\left[X_i\right] = 0$. We will

tackle $\mathrm{Cov}\left[Q_i, U_i\right]$ next:

$$\begin{aligned}
\mathrm{Cov}\left[Q_i, U_i\right] &= \frac{\beta_W}{3}\mathrm{Cov}\left[\mathbb{1}\left[U_i < \tfrac{1}{3}\right], U_i\right] \\
&\quad + \frac{\beta_W}{2}\mathrm{Cov}\left[\mathbb{1}\left[\tfrac{1}{3} \le U_i \le \tfrac{2}{3}\right], U_i\right] \\
&\quad + \frac{2\beta_W}{3}\mathrm{Cov}\left[\mathbb{1}\left[U_i > \tfrac{2}{3}\right], U_i\right] \\
&= \frac{\beta_W}{3}\left(\mathbb{E}\left[U_i \mathbb{1}\left[U_i < \tfrac{1}{3}\right]\right] - \frac{1}{6}\right) \\
&\quad + \frac{\beta_W}{2}\left(\mathbb{E}\left[U_i \mathbb{1}\left[\tfrac{1}{3} < U_i < \tfrac{2}{3}\right]\right] - \frac{1}{6}\right) \\
&\quad + \frac{2\beta_W}{3}\left(\mathbb{E}\left[U_i \mathbb{1}\left[U_i > \tfrac{2}{3}\right]\right] - \frac{1}{6}\right) \\
&= \frac{-\beta_W}{27} + 0 + \frac{5\beta_W}{27} \\
&= \frac{4\beta_W}{27}.
\end{aligned}$$

Final term needed for $\mathrm{Cov}\left[Q_i, f_i(1)\right]$ is $\mathrm{Cov}\left[Q_i, U_i^2\right]$. For this, note

$$\begin{aligned}
\mathbb{E}\left[U_i^2 \mathbb{1}\left[a \le U_i \le b\right]\right] &= \mathbb{E}\left[U_i^2 | a \le U_i \le b\right](b-a) \\
&= \mathbb{E}\left[\widetilde{U}_i^2\right](b-a) \qquad \text{where } \widetilde{U}_i \sim \mathrm{Uniform}[a,b] \\
&= \frac{a^2 + ab + b^2}{3}(b-a).
\end{aligned}$$

With this,

$$\begin{aligned}
\mathrm{Cov}\left[Q_i, U_i^2\right] &= \frac{\beta_W}{3}\mathrm{Cov}\left[\mathbb{1}\left[U_i < \tfrac{1}{3}\right], U_i^2\right] \\
&\quad + \frac{\beta_W}{2}\mathrm{Cov}\left[\mathbb{1}\left[\tfrac{1}{3} \le U_i \le \tfrac{2}{3}\right], U_i^2\right] \\
&\quad + \frac{2\beta_W}{3}\mathrm{Cov}\left[\mathbb{1}\left[U_i > \tfrac{2}{3}\right], U_i^2\right] \\
&= \frac{\beta_W}{3}\left(\frac{1}{3^4} - \frac{1}{9}\right) \\
&\quad + \frac{\beta_W}{2}\left(\frac{7}{3^4} - \frac{1}{9}\right) \\
&\quad + \frac{2\beta_W}{3}\left(\frac{19}{3^4} - \frac{1}{9}\right) \\
&= \frac{-8\beta_W}{3^5} - \frac{\beta_W}{3^4} + \frac{20\beta_W}{3^5} \\
&= \frac{\beta_W}{27}.
\end{aligned}$$

So,

$$\begin{aligned}\text{Cov}\left[Q_i, f_i(1)\right] &= \frac{1}{2}\pi^2\text{Cov}\left[Q_i, U_i^2\right] + \beta_W\pi\text{Cov}\left[Q_i, U_i\right] + \beta_X\pi\text{Cov}\left[Q_i, X_iU_i\right] \\ &= \beta_W\frac{\pi^2}{54} + \beta_W^2\frac{\pi}{27}.\end{aligned}$$

Similar calculations lead us to

$$\text{Cov}\left[Q_i, f_i(0)\right] = \beta_W\frac{\pi^2}{54}.$$

Putting all this together, we arrive at

$$\begin{aligned}&\frac{\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right]}{\pi(1-\pi)} \\ &= \frac{\frac{1}{2}\beta_X^4 + \beta_W^2\beta_X^2 + \frac{\pi^4}{45} + \beta_W^2\frac{\pi^2}{12} + \beta_X^2\frac{\pi^2}{3} + \beta_W\beta_X^2\pi + \beta_W\frac{\pi^3}{12} + \text{Var}\left[\epsilon_i\right]/25}{\pi^2} \\ &\quad + \frac{\frac{1}{2}\beta_X^4 + \frac{\pi^4}{45} + \beta_X^2\frac{\pi^2}{3} + \text{Var}\left[\epsilon_i\right]/25}{(1-\pi)^2} \\ &\quad + 2\frac{\frac{1}{2}\beta_X^4 + \beta_W\beta_X^2\frac{\pi}{2} + \frac{\pi^4}{45} + \beta_X^2\frac{\pi^2}{3} + \frac{\text{Var}[\epsilon]}{25}}{\pi(1-\pi)} \\ &\quad + \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)^2\beta_X^2\pi^2\left(\beta_W + \frac{1}{2}\right)^2 \\ &\quad + 2\left(\frac{\beta_W\frac{\pi^2}{54} + \beta_W^2\frac{\pi}{27}}{\pi} + \frac{\beta_W\frac{\pi^2}{54}}{1-\pi}\right) \\ &\quad - 2\left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)^2\beta_X^2\pi\left(\beta_W + \frac{1}{2}\right)\cdot\left(\beta_W(1-\pi) + \pi/2\right) \\ &\quad + 29\beta_W/108,\end{aligned}$$

We can characterize the relationship between $\text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right]$ and $\text{Var}\left[\hat{\tau}^{\text{HA}}\right]$ using Corollary 3.1. The result is as follows:

1. $\text{Var}\left[\hat{\tau}^{\text{HA}}\right] = \text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right]$ if one of the following hold:
   i. $\beta_X = 0$
   ii. $\beta_W = -0.5$,
   iii. $\beta_W \neq \frac{1}{6}$ and $\pi = \frac{4\beta_W}{6\beta_W - 1}$
2. $\text{Var}\left[\hat{\tau}^{\text{HA}}\right] > \text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right]$ if $\beta_X \neq 0$ and one of the following hold:
   i. $\beta_W < -0.5$ and $\pi < \frac{4\beta_W}{6\beta_W - 1}$
   ii. $\beta_W \in (-0.5, 0)$ and $\pi > \frac{4\beta_W}{6\beta_W - 1}$
   iii. $\beta_W > 0.5$ and $\pi < \frac{4\beta_W}{6\beta_W - 1}$
   iv. $\beta_W \in (0, 0.5]$
3. $\text{Var}\left[\hat{\tau}^{\text{HA}}\right] < \text{Var}\left[\hat{\beta}^{\text{ANCOVA}}\right]$ if $\beta_X \neq 0$ and one of the following hold:

i. $\beta_W < -0.5$ and $\pi > \frac{4\beta_W}{6\beta_W - 1}$
ii. $\beta_W \in (-0.5, 0)$ and $\pi < \frac{4\beta_W}{6\beta_W - 1}$
iii. $\beta_W > 0.5$ and $\pi > \frac{4\beta_W}{6\beta_W - 1}$

We can do a similar analysis for $\widehat{\beta}^{\text{ANCOVA2}}$ following Corollary 3.2, and get that $\text{Var}\left[\widehat{\beta}^{\text{ANCOVA2}}\right] = \text{Var}\left[\widehat{\tau}^{\text{HA}}\right]$ if $\beta_X = 0$ or $\pi = \frac{2\beta_W}{2\beta_W - 1}$. Otherwise, $\text{Var}\left[\widehat{\beta}^{\text{ANCOVA}}\right] < \text{Var}\left[\widehat{\tau}^{\text{HA}}\right]$.

Chapter 4

Conclusions

We have studied two examples of causal methods from two different areas of research that have very similar aims, but apply to very different scenarios and are at different levels of maturity.

In Chapter 2, we tackled the problem of partial identification of average treatment effects from two-sample Mendelian randomization studies. While the area of nonparametric bounds is relatively well studied, the application to MR studies is still relatively new, and the application to two-sample MR studies completely unexplored. Chapter 2 illustrated the usefulness of such bounds in the two-sample setting, and in particular highlighted their limitations. Because of the two-sample study design, crucial information linking the exposure and the outcome is missing, which results in that the bounds can be up to twice as wide as is the case in one-sample studies. The width of the bounds is still upper bounded by the strength of the instrumental variable, but when using genetic markers for instruments, the strength is often very low. In general, we see that nonparametric bounds from two-sample MR data are very wide, and therefore not very informative. This is not to say that bounds do not have a role to play in two-sample MR studies. The falsification inequalities can be used as a minimal check of the IV assumptions, and if prior knowledge about the effect of the exposure on the outcome is available, this might inform further constraints that could lead to narrower bounds.

The research area of causal effect estimation when interference is not negligible considered in Chapter 3 is still rapidly developing. While much of the early work imposed the assumption of partial interference, more recent work has shifted to various flavors of general interference. The assumption of observations being partitioned into independent groups has been replaced by various assumptions on the generation and/or growth of the interference graph which often has lead to asymptotic results. We considered asymptotic theory based on random graphon models, which allowed us to get a somewhat simple and interpretable expression for the asymptotic variance of the popular AIPW and ANCOVA estimators.

The conclusions drawn from the work in Chapter 3 are best framed in a practical setting. In a broader perspective, we have started to tackle the question of covariate adjustment as a way to improve precision when general interference is present. As has been noted by others, the presence of interference tends to inflate the variance of estimators. Therefore variance reduction techniques will play a crucial role in the analysis of data where interference is not negligible. While we in Chapter 3 considered the simpler case of a randomized controlled trial, it seems likely that most results can be extended to observational studies. This suggests that the AIPW and ANCOVA estimators are still suitable for the estimation of causal effects.

A more immediate consequence of this work is that analyzes already performed using an AIPW or ANCOVA estimator where interference was wrongfully assumed negligible still yields unbiased and consistent estimates assuming that the covariates adjusted for are in fact independent. However, inference based on these analyzes should be treated with skepticism as the traditional approaches to variance estimation, whether that be OLS based or bootstrap based, yield inconsistent results. In fact, the problem of variance estimation when interference is strong is still largely an unsolved problem. We do take one step in the right direction here, as we see how the problem simplifies under the sharp null hypothesis of no direct effect. In this scenario, we find an asymptotically consistent variance estimator, but show that this performs rather poorly in finite samples. The poor performance is consistent in the sense that it largely underestimates the variance. There seems to be room for improvement here, and a finite sample correction might be possible. Other non-sharp null

hypotehses might also be testable. We leave this as an area for future research.

References

Abraham, Miriam, Salem Alramadhan, Carmen Iniguez, Liesbeth Duijts, Vincent W. V. Jaddoe, Herman T. Den Dekker, Sarah Crozier, Keith M. Godfrey, Peter Hindmarsh, Torstein Vik, Geir W. Jacobsen, Wojciech Hanke, Wojciech Sobala, Graham Devereux, and Steve Turner (2017). "A Systematic Review of Maternal Smoking during Pregnancy and Fetal Measurements with Meta-Analysis". In: *PloS One* 12.2, e0170946. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0170946. pmid: 28231292.

Aldous, David J. (Dec. 1981). "Representations for Partially Exchangeable Arrays of Random Variables". In: *Journal of Multivariate Analysis* 11.4, pp. 581–598. ISSN: 0047259X. DOI: 10.1016/0047-259X(81)90099-3. URL: https://linkinghub.elsevier.com/retrieve/pii/0047259X81900993 (visited on 05/01/2023).

Angrist, Joshua D and Alan B Krueger (Nov. 1, 2001). "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments". In: *Journal of Economic Perspectives* 15.4, pp. 69–85. ISSN: 0895-3309. DOI: 10.1257/jep.15.4.69. URL: https://pubs.aeaweb.org/doi/10.1257/jep.15.4.69 (visited on 04/30/2021).

Aronow, Peter M. and Cyrus Samii (Dec. 1, 2017). "Estimating Average Causal Effects under General Interference, with Application to a Social Network Experiment". In: *The Annals of Applied Statistics* 11.4. ISSN: 1932-6157. DOI: 10.1214/16-AOAS1005. URL: https://projecteuclid.org/journals/annals-of-applied-statistics/volume-11/issue-4/Estimating-average-causal-effects-under-general-interference-with-application-to/10.1214/16-AOAS1005.full (visited on 10/25/2021).

Athey, Susan, Dean Eckles, and Guido W. Imbens (Jan. 2, 2018). "Exact P-Values for Network Interference". In: *Journal of the American Statistical Association* 113.521, pp. 230–240. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.2016.1241178. URL: https://www.tandfonline.com/doi/full/10.1080/01621459.2016.1241178 (visited on 11/15/2023).

Balke, Alexander and Judea Pearl (Sept. 1, 1997). "Bounds on Treatment Effects from Studies with Imperfect Compliance". In: *Journal of the American Statistical Association* 92.439, pp. 1171–1176. ISSN: 0162-1459. DOI: 10.1080/01621459.1997.10474074. URL: https://doi.org/10.1080/01621459.1997.10474074 (visited on 02/05/2020).

Bolthausen, Erwin, Edwin Arend Perkins, Aad W. van der Vaart, and Pierre Bernard (2002). *Lectures on Probability Theory and Statistics*. Lecture Notes in Mathematics 1781. Berlin Heidelberg New York [etc.]: Springer. ISBN: 978-3-540-43736-9.

Bound, John, David A. Jaeger, and Regina M. Baker (June 1995). "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak". In: *Journal of the American Statistical Association* 90.430, pp. 443–450. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1995.10476536. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476536 (visited on 11/14/2023).

Bowden, Jack, George Davey Smith, Philip C. Haycock, and Stephen Burgess (May 2016). "Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator". eng. In: *Genetic Epidemiology* 40.4, pp. 304–314. ISSN: 1098-2272. DOI: 10.1002/gepi.21965.

Bowden, Jack, Fabiola Del Greco M, Cosetta Minelli, George Davey Smith, Nuala A. Sheehan, and John R. Thompson (Dec. 2016). "Assessing the Suitability of Summary Data for Two-Sample Mendelian Randomization Analyses Using MR-Egger Regression: The Role of the I2 Statistic". en. In: *International Journal of Epidemiology* 45.6, pp. 1961–1974. ISSN: 0300-5771. DOI: 10.1093/ije/dyw220.

Burgess, Stephen (June 1, 2014). "Sample Size and Power Calculations in Mendelian Randomization with a Single Instrumental Variable and a Binary Outcome". In: *International*

Journal of Epidemiology 43.3, pp. 922–929. ISSN: 0300-5771. DOI: 10.1093/ije/dyu005. URL: https://academic.oup.com/ije/article/43/3/922/761826 (visited on 11/22/2020).

Burgess, Stephen, Adam Butterworth, and Simon G. Thompson (Nov. 2013). "Mendelian Randomization Analysis with Multiple Genetic Variants Using Summarized Data". eng. In: *Genetic Epidemiology* 37.7, pp. 658–665. ISSN: 1098-2272. DOI: 10.1002/gepi.21758.

Burgess, Stephen and Jeremy A. Labrecque (Oct. 2018). "Mendelian Randomization with a Binary Exposure Variable: Interpretation and Presentation of Causal Estimates". In: *European Journal of Epidemiology* 33.10, pp. 947–952. ISSN: 0393-2990, 1573-7284. DOI: 10.1007/s10654-018-0424-6.

Burgess, Stephen, Robert A. Scott, Nicholas J. Timpson, George Davey Smith, Simon G. Thompson, and EPIC-InterAct Consortium (July 2015). "Using Published Data in Mendelian Randomization: A Blueprint for Efficient Identification of Causal Risk Factors". eng. In: *European Journal of Epidemiology* 30.7, pp. 543–552. ISSN: 1573-7284. DOI: 10.1007/s10654-015-0011-z.

Burgess, Stephen, Dylan S. Small, and Simon G. Thompson (Oct. 2017). "A Review of Instrumental Variable Estimators for Mendelian Randomization". en. In: *Statistical Methods in Medical Research* 26.5, pp. 2333–2355. ISSN: 0962-2802. DOI: 10.1177/0962280215597579.

Burgess, Stephen and Simon G. Thompson (2012). "Improving Bias and Coverage in Instrumental Variable Analysis with Weak Instruments for Continuous and Binary Outcomes". In: *Statistics in Medicine* 31.15, pp. 1582–1600. ISSN: 1097-0258. DOI: 10.1002/sim.4498. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4498 (visited on 11/29/2020).

– (Mar. 2015). *Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation*. English. 1st edition. Boca Raton: Chapman and Hall/CRC. ISBN: 978-1-4665-7317-8.

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet (2015a). *Replication Data for: Social Networks and the Decision to Insure*. Version 1. ICPSR - Interuniversity Consortium for Political and Social Research. DOI: 10.3886/E113593V1. URL: https://www.openicpsr.org/openicpsr/project/113593/version/V1/view (visited on 10/16/2023).

– (Apr. 1, 2015b). "Social Networks and the Decision to Insure". In: *American Economic Journal: Applied Economics* 7.2, pp. 81–108. ISSN: 1945-7782, 1945-7790. DOI: 10.1257/app.20130442. URL: https://pubs.aeaweb.org/doi/10.1257/app.20130442 (visited on 05/04/2023).

Cheng, Jing and Dylan S. Small (2006). "Bounds on Causal Effects in Three-Arm Trials with Non-Compliance". In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 68.5, pp. 815–836. ISSN: 1369-7412.

Chin, Alex (Mar. 6, 2019a). *Central Limit Theorems via Stein's Method for Randomized Experiments under Interference*. arXiv: 1804.03105 [math, stat]. URL: http://arxiv.org/abs/1804.03105 (visited on 11/15/2023). preprint.

– (Sept. 25, 2019b). "Regression Adjustments for Estimating the Global Treatment Effect in Experiments with Interference". In: *Journal of Causal Inference* 7.2, p. 20180026. ISSN: 2193-3685, 2193-3677. DOI: 10.1515/jci-2018-0026. URL: https://www.degruyter.com/document/doi/10.1515/jci-2018-0026/html (visited on 10/16/2023).

Choi, David (July 3, 2017). "Estimation of Monotone Treatment Effects in Network Experiments". In: *Journal of the American Statistical Association* 112.519, pp. 1147–1155. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.2016.1194845. URL: https://www.tandfonline.com/doi/full/10.1080/01621459.2016.1194845 (visited on 11/15/2023).

– (Aug. 29, 2023). *New Estimands for Experiments with Strong Interference*. arXiv: 2107.00248 [stat]. URL: http://arxiv.org/abs/2107.00248 (visited on 11/15/2023). preprint.

Cholesterol Treatment Trialists' (CTT) Collaborators (2005). "Efficacy and Safety of Cholesterol-Lowering Treatment: Prospective Meta-Analysis of Data from 90 056 Participants in 14 Randomised Trials of Statins". In: *The Lancet* 366.9493, pp. 1267–1278. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(05)67394-1. URL: http://www.sciencedirect.com/science/article/pii/S0140673605673941.

– (Aug. 2012). "The Effects of Lowering LDL Cholesterol with Statin Therapy in People at Low Risk of Vascular Disease: Meta-Analysis of Individual Data from 27 Randomised Trials". In: *The Lancet* 380.9841, pp. 581–590. ISSN: 01406736. DOI: 10.1016/S0140-

6736(12)60367-5. URL: https://linkinghub.elsevier.com/retrieve/pii/S0140673612603675 (visited on 12/03/2020).

Cornfield, Jerome, William Haenszel, E. Cuyler Hammond, Abraham M. Lilienfeld, Michael B. Shimkin, and Ernst L. Wynder (Jan. 1959). "Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions". en. In: *JNCI: Journal of the National Cancer Institute* 22.1, pp. 173–203. ISSN: 0027-8874. DOI: 10.1093/jnci/22.1.173.

Cox, D. R. (1958). *Planning of Experiments*. Planning of Experiments. Oxford, England: Wiley, p. 308.

Cribari-Neto, Francisco, Nancy Lopes Garcia, and Klaus L. P. Vasconcellos (2000). "A Note on Inverse Moments of Binomial Variates". In: *Brazilian Review of Econometrics* 20.2.

Davey Smith, George and Shah Ebrahim (Feb. 2003). "'Mendelian Randomization': Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease?" en. In: *International Journal of Epidemiology* 32.1, pp. 1–22. ISSN: 0300-5771. DOI: 10.1093/ije/dyg070.

Davies, Neil M., Michael V. Holmes, and George Davey Smith (July 2018). "Reading Mendelian Randomisation Studies: A Guide, Glossary, and Checklist for Clinicians". en. In: *BMJ* 362. ISSN: 0959-8138, 1756-1833. DOI: 10.1136/bmj.k601.

Didelez, Vanessa and Nuala Sheehan (Aug. 2007). "Mendelian Randomization as an Instrumental Variable Approach to Causal Inference". eng. In: *Statistical Methods in Medical Research* 16.4, pp. 309–330. ISSN: 0962-2802. DOI: 10.1177/0962280206077743.

Diemer, Elizabeth W., Jeremy Labrecque, Henning Tiemeier, and Sonja A. Swanson (Jan. 2020). "Application of the Instrumental Inequalities to a Mendelian Randomization Study With Multiple Proposed Instruments". en-US. In: *Epidemiology* 31.1, p. 65. ISSN: 1044-3983. DOI: 10.1097/EDE.0000000000001126.

Egami, Naoki and Eric J. Tchetgen Tchetgen (Sept. 4, 2021). *Identification and Estimation of Causal Peer Effects Using Double Negative Controls for Unmeasured Network Confounding*. arXiv: 2109.01933 [stat]. URL: http://arxiv.org/abs/2109.01933 (visited on 11/15/2023). preprint.

Emmenegger, Corinne, Meta-Lina Spohn, Timon Elmer, and Peter Bühlmann (Sept. 4, 2023). *Treatment Effect Estimation with Observational Network Data Using Machine Learning*. DOI: 10.48550/arXiv.2206.14591. arXiv: 2206.14591 [math, stat]. URL: http://arxiv.org/abs/2206.14591 (visited on 09/13/2023). preprint.

Freedman, David A. (Feb. 2008). "On Regression Adjustments to Experimental Data". In: *Advances in Applied Mathematics* 40.2, pp. 180–193. ISSN: 01968858. DOI: 10.1016/j.aam.2006.12.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S019688580700005X (visited on 11/13/2023).

Halloran, M. Elizabeth and Claudio J. Struchiner (Mar. 1995). "Causal Inference in Infectious Diseases". In: *Epidemiology* 6.2, pp. 142–151. ISSN: 1044-3983. URL: https://journals.lww.com/epidem/Abstract/1995/03000/Causal_Inference_in_Infectious_Diseases.10.aspx (visited on 03/02/2022).

Hemani, Gibran, Jie Zheng, Benjamin Elsworth, Kaitlin H. Wade, Valeriia Haberland, Denis Baird, Charles Laurin, Stephen Burgess, Jack Bowden, Ryan Langdon, Vanessa Y. Tan, James Yarmolinsky, Hashem A. Shihab, Nicholas J. Timpson, David M. Evans, Caroline Relton, Richard M. Martin, George Davey Smith, Tom R. Gaunt, and Philip C. Haycock (May 2018). "The MR-Base platform supports systematic causal inference across the human phenome". In: *eLife* 7. Ed. by Ruth Loos, e34408. ISSN: 2050-084X. DOI: 10.7554/eLife.34408. URL: https://doi.org/10.7554/eLife.34408.

Hernández, Adrián V., Ewout W. Steyerberg, Isabella Butcher, Nino Mushkudiani, Gillian S. Taylor, Gordon D. Murray, Anthony Marmarou, Sung C. Choi, Juan Lu, J. Dik F. Habbema, and Andrew I.R. Maas (Sept. 2006). "Adjustment for Strong Predictors of Outcome in Traumatic Brain Injury Trials: 25% Reduction in Sample Size Requirements in the IMPACT Study". In: *Journal of Neurotrauma* 23.9, pp. 1295–1303. ISSN: 0897-7151, 1557-9042. DOI: 10.1089/neu.2006.23.1295. URL: http://www.liebertpub.com/doi/10.1089/neu.2006.23.1295 (visited on 11/13/2023).

Hernan, Miguel A. and James M. Robins (Sept. 30, 2023). "Causal Inference: What If". In.

Holmes, Michael V., Folkert W. Asselbergs, Tom M. Palmer, Fotios Drenos, Matthew B. Lanktree, Christopher P. Nelson, Caroline E. Dale, Sandosh Padmanabhan, Chris Finan,

Daniel I. Swerdlow, Vinicius Tragante, Erik P.A. van Iperen, Suthesh Sivapalaratnam, Sonia Shah, Clara C. Elbers, Tina Shah, Jorgen Engmann, Claudia Giambartolomei, Jon White, Delilah Zabaneh, Reecha Sofat, Stela McLachlan, on behalf of the UCLEB consortium, Pieter A. Doevendans, Anthony J. Balmforth, Alistair S. Hall, Kari E. North, Berta Almoguera, Ron C. Hoogeveen, Mary Cushman, Myriam Fornage, Sanjay R. Patel, Susan Redline, David S. Siscovick, Michael Y. Tsai, Konrad J. Karczewski, Marten H. Hofker, W. Monique Verschuren, Michiel L. Bots, Yvonne T. van der Schouw, Olle Melander, Anna F. Dominiczak, Richard Morris, Yoav Ben-Shlomo, Jackie Price, Meena Kumari, Jens Baumert, Annette Peters, Barbara Thorand, Wolfgang Koenig, Tom R. Gaunt, Steve E. Humphries, Robert Clarke, Hugh Watkins, Martin Farrall, James G. Wilson, Stephen S. Rich, Paul I.W. de Bakker, Leslie A. Lange, George Davey Smith, Alex P. Reiner, Philippa J. Talmud, Mika Kivimäki, Debbie A. Lawlor, Frank Dudbridge, Nilesh J. Samani, Brendan J. Keating, Aroon D. Hingorani, and Juan P. Casas (Mar. 1, 2015). "Mendelian Randomization of Blood Lipids for Coronary Heart Disease". In: *European Heart Journal* 36.9, pp. 539–550. ISSN: 1522-9645, 0195-668X. DOI: 10.1093/eurheartj/eht571. URL: https://academic.oup.com/eurheartj/article/36/9/539/506920 (visited on 12/03/2020).

Hu, Yuchen, Shuangning Li, and Stefan Wager (Nov. 29, 2022). "Average Direct and Indirect Causal Effects under Interference". In: *Biometrika* 109.4, pp. 1165–1172. ISSN: 0006-3444, 1464-3510. DOI: 10.1093/biomet/asac008. URL: https://academic.oup.com/biomet/article/109/4/1165/6524620 (visited on 05/01/2023).

Hudgens, Michael G. and M. Elizabeth Halloran (June 2008). "Toward Causal Inference With Interference". In: *Journal of the American Statistical Association* 103.482, pp. 832–842. ISSN: 0162-1459. DOI: 10.1198/016214508000000292. pmid: 19081744.

Kojevnikov, Denis, Vadim Marmer, and Kyungchul Song (June 2021). "Limit Theorems for Network Dependent Random Variables". In: *Journal of Econometrics* 222.2, pp. 882–908. ISSN: 03044076. DOI: 10.1016/j.jeconom.2020.05.019. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304407620302402 (visited on 11/15/2023).

Kurz, Christoph F. (Feb. 2022). "Augmented Inverse Probability Weighting and the Double Robustness Property". In: *Medical Decision Making* 42.2, pp. 156–167. ISSN: 0272-989X.

DOI: 10.1177/0272989X211027181. pmid: 34225519. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8793316/ (visited on 10/19/2023).

Lawlor, Debbie A., Roger M. Harbord, Jonathan A. C. Sterne, Nic Timpson, and George Davey Smith (Apr. 2008). "Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology". eng. In: *Statistics in Medicine* 27.8, pp. 1133–1163. ISSN: 0277-6715. DOI: 10.1002/sim.3034.

Leung, Michael P. (May 2020). "Treatment and Spillover Effects Under Network Interference". In: *The Review of Economics and Statistics* 102.2, pp. 368–380. ISSN: 0034-6535, 1530-9142. DOI: 10.1162/rest_a_00818. URL: https://direct.mit.edu/rest/article/102/2/368-380/96745 (visited on 11/29/2022).

Li, Shuangning and Stefan Wager (Aug. 2022). "Random Graph Asymptotics for Treatment Effect Estimation under Network Interference". In: *The Annals of Statistics* 50.4, pp. 2334–2358. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/22-AOS2191. (Visited on 03/28/2023).

Lin, Winston (Mar. 1, 2013). "Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman's Critique". In: *The Annals of Applied Statistics* 7.1. ISSN: 1932-6157. DOI: 10.1214/12-AOAS583. URL: https://projecteuclid.org/journals/annals-of-applied-statistics/volume-7/issue-1/Agnostic-notes-on-regression-adjustments-to-experimental-data--Reexamining/10.1214/12-AOAS583.full (visited on 11/13/2023).

Liu, Lan, Michael G. Hudgens, and Sylvia Becker-Dreps (Dec. 1, 2016). "On Inverse Probability-Weighted Estimators in the Presence of Interference". In: *Biometrika* 103.4, pp. 829–842. ISSN: 0006-3444. DOI: 10.1093/biomet/asw047. URL: https://doi.org/10.1093/biomet/asw047 (visited on 02/02/2021).

Liu, Lan, Michael G. Hudgens, Bradley Saul, John D. Clemens, Mohammad Ali, and Michael E. Emch (Jan. 2019). "Doubly Robust Estimation in Observational Studies with Partial Interference". In: *Stat* 8.1. ISSN: 2049-1573, 2049-1573. DOI: 10.1002/sta4.214. URL: https://onlinelibrary.wiley.com/doi/10.1002/sta4.214 (visited on 10/25/2021).

Loh, Po-Ru, Gaurav Bhatia, Alexander Gusev, Hilary K. Finucane, Brendan K. Bulik-Sullivan, Samuela J. Pollack, Schizophrenia Working Group of Psychiatric Genomics Consortium, Teresa R. de Candia, Sang Hong Lee, Naomi R. Wray, Kenneth S. Kendler,

Michael C. O'Donovan, Benjamin M. Neale, Nick Patterson, and Alkes L. Price (Dec. 2015). "Contrasting Genetic Architectures of Schizophrenia and Other Complex Diseases Using Fast Variance-Components Analysis". In: *Nature Genetics* 47.12, pp. 1385–1392. ISSN: 1546-1718. DOI: 10.1038/ng.3431. pmid: 26523775.

Lovász, László and Balázs Szegedy (Nov. 1, 2006). "Limits of Dense Graph Sequences". In: *Journal of Combinatorial Theory, Series B* 96.6, pp. 933–957. ISSN: 0095-8956. DOI: 10.1016/j.jctb.2006.05.002. URL: https://www.sciencedirect.com/science/article/pii/S0095895606000517 (visited on 07/15/2022).

Lunceford, Jared K. and Marie Davidian (Oct. 15, 2004). "Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study". In: *Statistics in Medicine* 23.19, pp. 2937–2960. ISSN: 02776715. DOI: 10.1002/sim.1903. URL: https://onlinelibrary.wiley.com/doi/10.1002/sim.1903 (visited on 04/28/2023).

Manolio, Teri A. (July 8, 2010). "Genomewide Association Studies and Assessment of the Risk of Disease". In: *New England Journal of Medicine* 363.2. Ed. by W. Gregory Feero and Alan E. Guttmacher, pp. 166–176. ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMra0905980. URL: http://www.nejm.org/doi/10.1056/NEJMra0905980 (visited on 12/31/2021).

Manski, Charles F. (1990). "Nonparametric Bounds on Treatment Effects". In: *The American Economic Review* 80.2, pp. 319–323. ISSN: 0002-8282.

– (July 1993). "Identification of Endogenous Social Effects: The Reflection Problem". In: *The Review of Economic Studies* 60.3, p. 531. ISSN: 00346527. DOI: 10.2307/2298123. URL: https://academic.oup.com/restud/article-lookup/doi/10.2307/2298123 (visited on 11/15/2023).

McFowland III, Edward and Cosma Rohilla Shalizi (June 17, 2021). "Estimating Causal Peer Influence in Homophilous Social Networks by Inferring Latent Locations". arXiv: 1607.06565 [physics, stat]. URL: http://arxiv.org/abs/1607.06565 (visited on 09/28/2021).

Millard, Louise A. C., Marcus R. Munafò, Kate Tilling, Robyn E. Wootton, and George Davey Smith (Oct. 31, 2019). "MR-pheWAS with Stratification and Interaction: Searching for the Causal Effects of Smoking Heaviness Identified an Effect on Facial Aging". In:

PLOS Genetics 15.10, e1008353. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1008353. URL: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1008353 (visited on 11/22/2020).

Neyman, Jerzy (1923). "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." Trans. by D. M. Dabrowska and T. P. Speed. In: *Statistical Science* 5.4, pp. 465–472. ISSN: 0883-4237. JSTOR: 2245382. URL: https://www.jstor.org/stable/2245382 (visited on 05/04/2022).

Nj, Timpson, Greenwood Cmt, Soranzo N, Lawson Dj, and Richards Jb (Dec. 11, 2017). "Genetic Architecture: The Shape of the Genetic Contribution to Human Traits and Disease." In: *Nature reviews. Genetics* 19.2, pp. 110–124. ISSN: 1471-0056, 1471-0064. DOI: 10.1038/nrg.2017.101. pmid: 29225335. URL: http://europepmc.org/article/MED/29225335 (visited on 11/29/2020).

Ogburn, Elizabeth L., Oleg Sofrygin, Iván Díaz, and Mark J. Van Der Laan (Dec. 12, 2022). "Causal Inference for Social Network Data". In: *Journal of the American Statistical Association*, pp. 1–15. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.2022.2131557. URL: https://www.tandfonline.com/doi/full/10.1080/01621459.2022.2131557 (visited on 11/15/2023).

Ogburn, Elizabeth L. and Tyler J. VanderWeele (Nov. 1, 2014). "Causal Diagrams for Interference". In: *Statistical Science* 29.4. ISSN: 0883-4237. DOI: 10.1214/14-STS501. URL: https://projecteuclid.org/journals/statistical-science/volume-29/issue-4/Causal-Diagrams-for-Interference/10.1214/14-STS501.full (visited on 11/15/2023).

Ramsahai, Roland R. (Mar. 2012). "Causal Bounds and Observable Constraints for Non-Deterministic Models". In: *J. Mach. Learn. Res.* 13, pp. 829–848. ISSN: 1532-4435.

Richardson, Thomas S. and James M. Robins (Aug. 2014). "ACE Bounds; SEMs with Equilibrium Conditions". In: *Statistical Science* 29.3, pp. 363–366. ISSN: 0883-4237. DOI: 10.1214/14-STS485. arXiv: 1410.0470.

Richardson, Tom G., Eleanor Sanderson, Tom M. Palmer, Mika Ala-Korpela, Brian A. Ference, George Davey Smith, and Michael V. Holmes (Mar. 23, 2020). "Evaluating the Relationship between Circulating Lipoprotein Lipids and Apolipoproteins with Risk of

Coronary Heart Disease: A Multivariable Mendelian Randomisation Analysis". In: *PLOS Medicine* 17.3. Ed. by Daniel J. Rader, e1003062. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1003062. URL: https://dx.plos.org/10.1371/journal.pmed.1003062 (visited on 12/03/2020).

Robins, James M. (1989). "The Analysis of Randomized and Nonrandomized AIDS Treatment Trials Using A New Approach to Causal Inference in Longitudinal Studies". In: *Health Service Research Methodology: A Focus On AIDS*, pp. 113–159. URL: https://cdn1.sph.harvard.edu/wp-content/uploads/sites/343/2013/03/nchsr.pdf (visited on 11/01/2020).

Robins, James M. and Andrea Rotnitzky (Mar. 1995). "Semiparametric Efficiency in Multivariate Regression Models with Missing Data". In: *Journal of the American Statistical Association* 90.429, pp. 122–129. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1995.10476494. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476494 (visited on 04/28/2023).

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao (Mar. 1995). "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data". In: *Journal of the American Statistical Association* 90.429, pp. 106–121. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1995.10476493. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476493 (visited on 04/28/2023).

Rosenbaum, Paul R. (June 1987). "Model-Based Direct Adjustment". In: *Journal of the American Statistical Association* 82.398, pp. 387–394. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1987.10478441. URL: http://www.tandfonline.com/doi/abs/10.1080/01621459.1987.10478441 (visited on 05/16/2023).

– (2007). "Interference between Units in Randomized Experiments". In: *Journal of the American Statistical Association* 102.477, pp. 191–200. ISSN: 0162-1459. JSTOR: 27639831. URL: http://www.jstor.org/stable/27639831 (visited on 02/23/2022).

Ross, Sir Ronald (Feb. 1916). "An Application of the Theory of Probabilities to the Study of a Priori Pathometry.—Part I". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 92.638, pp. 204–230. ISSN:

0950-1207, 2053-9150. DOI: 10.1098/rspa.1916.0007. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.1916.0007 (visited on 11/18/2023).

Rotnitzky, Andrea, Q. Lei, M. Sued, and James M. Robins (June 1, 2012). "Improved Double-Robust Estimation in Missing Data and Causal Inference Models". In: *Biometrika* 99.2, pp. 439–456. ISSN: 0006-3444, 1464-3510. DOI: 10.1093/biomet/ass013. URL: https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/ass013 (visited on 05/16/2023).

Rubin, Donald B. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." In: *Journal of Educational Psychology* 66.5, pp. 688–701. ISSN: 0022-0663. DOI: 10.1037/h0037350. URL: http://content.apa.org/journals/edu/66/5/688 (visited on 11/09/2020).

– (1980). "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment". In: *Journal of the American Statistical Association* 75.371, pp. 591–593. ISSN: 0162-1459. DOI: 10.2307/2287653.

– (1990). "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies". In: *Statistical Science* 5.4, pp. 472–480. JSTOR: 2245383. URL: http://www.jstor.org/stable/2245383.

Sävje, Fredrik, Peter M. Aronow, and Michael G. Hudgens (Apr. 1, 2021). "Average Treatment Effects in the Presence of Unknown Interference". In: *The Annals of Statistics* 49.2. ISSN: 0090-5364. DOI: 10.1214/20-AOS1973. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-49/issue-2/Average-treatment-effects-in-the-presence-of-unknown-interference/10.1214/20-AOS1973.full (visited on 03/20/2023).

Schmitt, J., A. Seidler, T. L. Diepgen, and A. Bauer (Feb. 2011). "Occupational Ultraviolet Light Exposure Increases the Risk for the Development of Cutaneous Squamous Cell Carcinoma: A Systematic Review and Meta-Analysis". In: *The British Journal of Dermatology* 164.2, pp. 291–307. ISSN: 1365-2133. DOI: 10.1111/j.1365-2133.2010.10118.x. pmid: 21054335.

Shi, Huwenbo, Gleb Kichaev, and Bogdan Pasaniuc (July 7, 2016). "Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data". In: *The American*

Journal of Human Genetics 99.1, pp. 139–153. ISSN: 0002-9297. DOI: 10.1016/j.ajhg.2016.05.013. URL: http://www.sciencedirect.com/science/article/pii/S0002929716301483 (visited on 11/29/2020).

Slob, Eric A. W. and Stephen Burgess (2020). "A Comparison of Robust Mendelian Randomization Methods Using Summary Data". In: *Genetic Epidemiology* 44.4, pp. 313–329. ISSN: 1098-2272. DOI: 10.1002/gepi.22295. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/gepi.22295 (visited on 12/08/2020).

Smoking, Surgeon General's Advisory Committee on and United States. Health (1964). *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Public Health Service Publication. U.S. Department of Health, Education, and Welfare, Public Health Service. URL: https://books.google.com/books?id=yPtqAAAAMAAJ.

Sobel, Michael E. (2006). "What Do Randomized Studies of Housing Mobility Demonstrate?: Causal Inference in the Face of Interference". In: *Journal of the American Statistical Association* 101.476, pp. 1398–1407. ISSN: 0162-1459. JSTOR: 27639760. URL: https://www.jstor.org/stable/27639760 (visited on 11/25/2023).

Splawa-Neyman, Jerzy (1923). "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." Trans. by D. M. Dabrowska and T. P. Speed. In: *Statistical Science* 5.4. Translated in 1990., pp. 465–472. ISSN: 0883-4237.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo (Oct. 2002). "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments". In: *Journal of Business & Economic Statistics* 20.4, pp. 518–529. ISSN: 0735-0015. DOI: 10.1198/073500102288618658.

Stock, James H. and Motohiro Yogo (2005). *Testing for Weak Instruments in Linear IV Regression*. New York: Cambridge University Press, pp. 80–108. URL: http://www.economics.harvard.edu/faculty/stock/files/TestingWeakInstr_Stock%5C%2BYogo.pdf.

Sussman, Daniel L., Minh Tang, and Carey E. Priebe (Jan. 2014). "Consistent Latent Position Estimation and Vertex Classification for Random Dot Product Graphs". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.1, pp. 48–57. ISSN: 0162-

8828, 2160-9292. DOI: 10.1109/TPAMI.2013.135. URL: http://ieeexplore.ieee.org/document/6565321/ (visited on 10/09/2023).

Swanson, Sonja A. (Jan. 2017). "Commentary: Can We See the Forest for the IVs? Mendelian Randomization Studies with Multiple Genetic Variants". In: *Epidemiology* 28.1, pp. 43–46. ISSN: 1044-3983. DOI: 10.1097/EDE.0000000000000558. URL: https://journals.lww.com/00001648-201701000-00007 (visited on 12/01/2020).

Swanson, Sonja A., Miguel A. Hernán, Matthew Miller, James M. Robins, and Thomas S. Richardson (Apr. 2018). "Partial Identification of the Average Treatment Effect Using Instrumental Variables: Review of Methods for Binary Instruments, Treatments, and Outcomes". en. In: *Journal of the American Statistical Association* 113.522, pp. 933–947. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.2018.1434530.

Tchetgen, Eric J. Tchetgen and Tyler J. VanderWeele (Feb. 2012). "On Causal Inference in the Presence of Interference". In: *Statistical methods in medical research* 21.1, pp. 55–75. ISSN: 0962-2802. DOI: 10.1177/0962280210386779. pmid: 21068053. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4216807/ (visited on 02/02/2021).

Tsiatis, Anastasios A., Marie Davidian, Min Zhang, and Xiaomin Lu (Oct. 2008). "Covariate Adjustment for Two-Sample Treatment Comparisons in Randomized Clinical Trials: A Principled yet Flexible Approach". In: *Statistics in Medicine* 27.23, pp. 4658–4677. ISSN: 02776715, 10970258. DOI: 10.1002/sim.3113. (Visited on 12/22/2022).

U.S. Food and Drug Administration. Center for Drug Evaluation and Research (May 2023). "Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products - Guidance for Industry". In.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK ; New York, NY, USA: Cambridge University Press. ISBN: 978-0-521-49603-2.

Verbanck, Marie, Chia-Yen Chen, Benjamin Neale, and Ron Do (May 2018). "Detection of Widespread Horizontal Pleiotropy in Causal Relationships Inferred from Mendelian Randomization between Complex Traits and Diseases". In: *Nature Genetics* 50.5, pp. 693–

698. ISSN: 1061-4036, 1546-1718. DOI: 10.1038/s41588-018-0099-7. URL: http://www.nature.com/articles/s41588-018-0099-7 (visited on 12/08/2020).

Wang, Linbo and Eric Tchetgen Tchetgen (Jan. 2018). "Bounded, Efficient and Multiply Robust Estimation of Average Treatment Effects Using Instrumental Variables". In: *arXiv:1611.09925 [stat]*. arXiv: 1611.09925 [stat].

Yang, Jian, Beben Benyamin, Brian P. McEvoy, Scott Gordon, Anjali K. Henders, Dale R. Nyholt, Pamela A. Madden, Andrew C. Heath, Nicholas G. Martin, Grant W. Montgomery, Michael E. Goddard, and Peter M. Visscher (July 2010). "Common SNPs Explain a Large Proportion of the Heritability for Human Height". In: *Nature Genetics* 42.7, pp. 565–569. ISSN: 1546-1718. DOI: 10.1038/ng.608. pmid: 20562875.

Zhao, Qingyuan, Jingshu Wang, Gibran Hemani, Jack Bowden, and Dylan S. Small (June 2020). "Statistical Inference in Two-Sample Summary-Data Mendelian Randomization Using Robust Adjusted Profile Score". EN. In: *Annals of Statistics* 48.3, pp. 1742–1769. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/19-AOS1866.