Missing Data in Confounders[*]

Melody Huang[†] Naijia Liu[‡]

Abstract

When confounder data is missing-not-at-random in observational data, standard assumptions are no longer sufficient for identifying the average treatment effect. In practice, researchers often rely on either a complete case estimator, or imputation to estimate the average treatment effect. In the following paper, we demonstrate that despite their popularity, neither approach provides unbiased estimation of the ATE, without additional assumptions that are untenable in practice. We show that the imputation estimators will only be unbiased in settings when we are able to perfectly impute the missing confounder values. Paradoxically, this is only feasible in settings when the missing values can be perfectly explained by the observed data, making it redundant to impute at all. We propose an alternative identification strategy, which allows researchers to leverage a two-stage estimator to consistently estimate the ATE under arguably weaker assumptions. We introduce a suite of validation approaches to evaluate the credibility of the proposed assumptions. We illustrate our framework on a recent study evaluating the impact of government transparency on state legislature and show that results from the proposed two-stage estimator differ significantly from existing missing data estimators.

*** **Preliminary Draft.*****
Please do not share without permission from the corresponding author.

[*] The authors thank Danny Ebanks, Sam Pimentel, Kosuke Imai, Bryant Moy, Kyla Chasalow, Ye Wang, Devin Caughey, Avi Feller, P Aronow, MIT Political methodology lab members, UC Berkeley Methods Workshop participants, Polmeth 2024 participants and the Imai Research Group for helpful comments and discussion. The authors also thank Fan Li and the Harvard Causal Inference Seminar Reading Group for a thought-provoking discussion that inspired this work.

[†] Assistant Professor of Political Science, Yale University, melody.huang@yale.edu, www.melodyyhuang.com

[‡] Assistant Professor of Government, Harvard University, naijialiu@fas.harvard.edu

1 Introduction

Missing data is prevalent in empirical research in social sciences. Issues at the data collection stage, such as data entry mistakes, selective non-response, and record linkage problems, result in datasets with missing values. Political science studies often rely on these incomplete datasets. For example, Lall (2016) found that almost every study in the field of comparative and international political economy published during a five-year period in *International Organization* and *World Politics* suffered from missing values. In this paper, we replicate the study by Harden and Kirkland (2021), which investigates the effect of transparency rules on state legislative outcomes. Unfortunately, the dataset contains substantial missingness among the confounding variables. In the presence of missing *covariate* data, researchers must decide whether to conduct the study using only complete cases or to impute the missing data. Both choices rely on different underlying assumptions that can have different substantive implications for the results (Lall, 2016; King et al., 1998).

The problem of missing data is exacerbated in settings when researchers rely on covariate information to account for potential confounding. Recent methodological advances have allowed for better ways to recover missing covariate data (e.g., Honaker et al., 1999, 2011; Van Buuren and Groothuis-Oudshoorn, 2011). However, despite the significant interest in observational causal inference methods in political science, little work has been done to understand the impact of missing data in pre-treatment confounders. We will show that, problematically, missing data in confounders, and how researchers choose to account for the missingness, can substantially change the research conclusions from an empirical analysis.

To gain an understanding of the magnitude of the problem of missing data in political science, we review the past 5 years of publications in the *American Journal of Political Science* (AJPS) and *American Political Science Review* (APSR).[1] From a keyword search, we found 130 articles in the APSR and 90 articles in the AJPS conducting observational studies with a causal story. Among these, 70 out of 130 articles in APSR and 41 out of 90 articles in AJPS mentioned some form of missing values in their main texts. The remainder of the articles do not offer discussion on missing data, even if their datasets contain missing values. Finally, fewer than 5 articles in AJPS and fewer than 7 articles in APSR disclosed incorporating some form of imputation models into

[1] The five year period is 2019 - 2024. We excluded methodology papers from the count.

their analysis, such as mean imputation and multiple imputation.

In the following paper, we formalize the implications of missing data in confounders in the context of observational studies. We provide three primary contributions. First, we derive the necessary identifying assumptions for existing missing data approaches and find that the two commonly used estimators–complete case and imputation estimators–require untenably strong assumptions that can be difficult to justify in practice. We show that paradoxically, imputation estimators are only unbiased in settings when the missing confounders do not have confounding power, thereby making imputation unnecessary. As such, even under standard imputation assumptions, such as missing-at-random, the imputation estimator will fail to be unbiased.

Second, we propose a novel method for researchers to unbiasedly estimate causal effects in the presence of missing data in confounders. Specifically, we leverage recent work in the statistics literature to propose a new identifying assumption referred to as *outcome-response ignorability* (e.g., Yang et al., 2019). While outcome-response ignorability nonparametrically identifies the average treatment effect (ATE) in the presence of missing confounder data, implementing it in practice is often computationally infeasible, especially when continuous variables are involved. Instead, we propose an alternative estimation approach in the form of a two-stage projection estimator. The estimation procedure involves two stages: (i) obtaining a complete case estimator, and (ii) projecting the result onto imputed data. Informally, the projection estimator will augment the first stage complete case estimator to target potential confounding in the missingness pattern. The two-stage projection estimator relies on an additional estimation assumption, which is that the imputation model correctly recovers the distribution of the missing covariates–an assumption satisfied by missing-at-random. We show that under outcome-response ignorability and missing-at-random, the two-stage projection estimator will consistently recover the ATE. In contrast to the existing complete case and imputation estimators, the projection estimator is able to leverage an informative imputation model to recover the ATE. This is especially advantageous in settings when researchers have taken time to collect covariates that are prognostic of the missingness in their underlying data. Our proposed identification strategy also allows researchers to identify the conditional average treatment effect, thereby enabling researchers to consider subgroup treatment effects within their studies, even in the presence of missing covariate data.

Our third contribution is a suite of diagnostic tools for researchers to evaluate the robustness of the projection estimator. These tools include (1) validation procedures for researchers to evaluate the performance of the underlying imputation model under different missingness mechanisms; (2) diagnostics to evaluate whether the observable implications of the outcome-response ignorability assumption hold; and (3) a sensitivity analysis that allows researchers to consider the robustness of the projection estimator under violations of both outcome-response ignorability and missing-at-random.

The paper is structured as follows. Section 2 provides the notation and assumptions. In Section 3, we formalize the corresponding identifying assumptions needed for the complete case and imputation estimators to be unbiased. In Section 4, we introduce our proposed estimator and discuss the theoretical properties associated with it. Section 5 considers the robustness checks for the proposed estimation approach. In Section 6, we provide different simulated numerical examples of the performance of our approach in comparison to standard methods. In Section 7, We perform a re-analysis of a recent study evaluating the impact of government transparency on state legislature (Harden and Kirkland, 2021). We show that depending on how researchers choose to account for the missingness in the underlying covariates, the substantive takeaway changes from government transparency having no impact to a *positive* impact on reducing policy making.

2 Notation and Assumptions

Define $Y_i(1)$ and $Y_i(0)$ as the potential outcomes under treatment and control, respectively. Define $Z_i \in \{0, 1\}$ as a treatment assignment indicator, such that if $Z_i = 1$, a unit receives treatment, and 0 otherwise. We focus on the setting in which researchers are considering a binary treatment variable. We will invoke the standard assumptions of no interference and consistency (i.e., SUTVA). Furthermore, we assume full compliance, such that units assigned to treatment receive treatment. Define the observed outcomes as $Y := Y_i(1)Z_i + Y_i(0)(1 - Z_i)$. Finally, we define a set of pre-treatment covariates X_i. We will assume that $\{Y_i(1), Y_i(0), Z_i, X_i\}$ is sampled i.i.d. Our estimand of interest throughout the paper is the *average treatment effect*:

$$\tau := \mathbb{E}\left[Y_i(1) - Y_i(0)\right].$$

In a randomized control trial, researchers have the power to randomly assign treatment (i.e., Z_i is randomized). However, in observational studies, treatment assignment is no longer random. A common approach to estimating the treatment effect in this setting is to leverage a conditional ignorability assumption (Rubin, 1976).

Assumption 1 (Conditional Ignorability of Treatment Assignment)

$$Y_i(1), Y_i(0) \perp\!\!\!\perp Z_i \mid X_i$$

Assumption 1 states that given a set of pre-treatment covariates, the confounding effects from selection into treatment will be ignorable.

Researchers also must invoke a positivity, or overlap, assumption.

Assumption 2 (Positivity of Treatment Assignment)

$$0 < \Pr(Z_i = 1 \mid X_i) \leq 1$$

In other words, each unit must have a non-zero probability of receiving treatment.

Leveraging both identifying assumptions, the ATE is identified, and can be estimated using different approaches, such as propensity score weighting, outcome modeling, or doubly robust estimation (e.g., Lunceford and Davidian, 2004; Bang and Robins, 2005; Rosenbaum and Rubin, 1983b). However, in order for Assumption 1 and 2 to be sufficient for unbiased estimation, the set of pre-treatment covariates necessary for Assumption 1 must be fully measured. While existing literature has introduced different approaches to consider omitted variable bias (e.g., Cornfield et al., 1959; Rosenbaum and Rubin, 1983a; Zhao et al., 2019; Cinelli and Hazlett, 2020; Huang and Pimentel, 2022; Dorn and Guo, 2023; Huang and McCartan, 2025, to name a few), a common problem that arises in practice is the presence of missing data in the *observed* covariates X_i.

Recent papers have considered the implications of missing pre-treatment covariate data in the context of randomized control trials (e.g., Zhao and Ding, 2022; Chang et al., 2023). However, within an experimental setting, adjusting for pre-treatment covariates is an efficiency problem, and not necessary for identification (e.g., Zhao et al., 2024). We focus specifically on the more challenging setting of missing data in the pre-treatment covariates X_i in an observational study, in which it is necessary to adjust for the full set of X_i to identify the ATE.

When missingness in the covariates is not completely random, the ATE can no longer be identified without further assumptions. Throughout the paper, we will assume that there is missing data in X_i, but that Y_i and Z_i are fully observed. In other words, we do not consider settings with missing outcome or treatment observations. We refer readers to Zhao et al. (2023) and Huang (2024) for more discussion about these settings. Furthermore, we assume that while there are missing observations in X_i, the set of covariates X_i is otherwise fully measured (i.e., no omitted variable bias). In settings when researchers are also concerned about unobserved confounders, recently introduced sensitivity analyses can be applied in conjunction with the proposed estimation approach to consider sensitivity to omitted variables.

3 Identification Assumptions for Existing Missing Data Methods

For each observation i, define $R_i = \left(R_i^{(1)} \dots R_i^{(p)}\right)$ as a matrix, where each column corresponds to a missingness indicator for a covariate. Let $R_i = \mathbb{1}_p$ denote the complete cases. Define the covariate matrix $X_i^{obs} := X_i \cdot R_i$ as a masked version of the covariate matrix X_i that is observed. Similarly, define $X_i^{mis} := X_i \cdot (1 - R_i)$ is the masked covariate matrix, containing only the missing values of X_i. Finally, define $\tilde{X}_i := X_i^{obs} \cdot R_i + \hat{X}_i \cdot (1 - R_i)$, where $\hat{X}_i$ represents the imputed X_i^{mis} values. Figure 1 provides a visualization of the set-up.

X_i^{obs}	X_i^{mis}	R_i
✓	✓	$(1,1)^\top$
✓	✓	$(1,1)^\top$
✓	✓	$(1,1)^\top$
✓	NA	$(1,0)^\top$
✓	NA	$(1,0)^\top$
✓	NA	$(1,0)^\top$

Figure 1

For the remainder of the manuscript, we assume *positivity in missingness.* [2]

[2] The different estimators considered in this paper do not contend with potential violations of positivity in missingness. We discuss potential extensions and approaches to deal with positivity violations in Section 8.

Assumption 3 (Positivity in $\boldsymbol{R}$) *For $j \in \{1, ..., p\}$,*

$$0 < \Pr(R_i^{(j)} = 1 \mid X_i) \leq 1.$$

Positivity in missingness (Assumption 3) assumes that certain values of a covariate cannot be systematically missing. For example, consider a study in which researchers are controlling for geographic fixed effects by including an indicator for what neighborhood an individual lives in. Assumption 3 rules out settings in which covariates are entirely missing for a given neighborhood. In the study by Harden and Kirkland (2021), which we replicate later, this assumption excludes cases where confounder data from an entire state are missing—for example, when $\Pr(R_i^{(j)} = 1 \mid \text{certain state}) = 0$. In practice, we advise researchers check the observable implications of this assumption by evaluating whether any covariate or combination of covariates can perfectly predict missingness. This is especially important in panel data settings, when there could be units or time periods with systematically missing values. When there are violations of positivity in missingness, researchers must subset their data to the set of observations for which there is positivity in R, or leverage stronger assumptions to account for positivity violations.

In the following section, we will discuss two common estimators—complete case estimator and imputation estimators—often used when there is missing data, and formalize the assumptions needed for the estimators to be unbiased.

3.1 Complete Case Estimators

A common approach used in practice when there are missing values in the underlying confounders is to omit all missing observations, and estimate the ATE across the complete cases. This is also known as *complete case estimators*, or list-wise deletion. In settings when the missingness is completely at random, then the complete case estimator will be an unbiased estimator for the ATE. More concretely, the following assumption must hold.

Assumption 4 (Missing Completely at Random (MCAR))

$$R_i \perp\!\!\!\perp \{Z_i, X_i, Y_i(1), Y_i(0)\}$$

More specifically, MCAR states that missingness cannot impact the treatment assignment process or the outcomes.

In practice, missingness is often confounded with the treatment assignment or the outcomes. For example, consider a study examining the effect of new exposure to polarizing news articles on presidential vote choice (e.g., see Prior, 2013 for a systematic review of common studies related to polarizing media on voter preferences). Since exposure to polarizing news articles is often confounded with existing political ideology (e.g., Tyler et al., 2022), researchers control for individuals' demographic data as well as previous policy preferences, collected from historical large-scale surveys. A well-known challenge in survey data is non-response to more politically charged or sensitive questions—e.g., questions about abortion, the death penalty, or gun control (e.g., Yan, 2021). Missing completely at random would state that whether or not individuals respond to certain policy questions is unrelated to (1) vote choice, (2) whether or not they select into reading polarizing news articles, and (3) how sensitive the underlying policy questions are. Problematically, if individuals who are more likely to read polarizing news articles are also more likely to respond to sensitive questions, then by naively omitting all units with missing observations, we will systematically only include the subset of individuals more likely to read polarizing news articles into the analysis, resulting in a biased estimate.

The bias of a complete case estimator that arises will depend on how strong missingness is related to treatment assignment Z_i, and how strong missingness is related to the underlying treatment effect (i.e., $Y_i(1) - Y_i(0)$). In addition to potential bias, complete case estimators have also been criticized for discarding data, resulting in potentially large amounts of efficiency loss (King et al., 1998). As such, given the restrictiveness of the underlying missing completely at random assumption and potential efficiency loss, the applicability of a complete case estimator is relatively limited.

3.2 Imputation Estimators

Imputation estimators are an alternative to complete case estimators. Researchers first impute the missing data using a chosen imputation model (e.g., Honaker et al., 1999, 2011; Van Buuren and Groothuis-Oudshoorn, 2011; Hollenbach et al., 2014), and then perform estimation across the imputed data $\tilde{X}_i$, often adding in the missingness indicator R_i. Because complete case estimators are only unbiased in settings when missingness is completely at random, imputation estimators are often expected to provide less biased estimates than complete case estimators (e.g., King et al., 2001; Lall, 2016).

A selection on observables assumption, also referred to as *missing-at-random*, is often used to justify imputation estimators.

Assumption 5 (Missing-at-Random)

$$X_i^{mis} \perp\!\!\!\perp R_i \mid X_i^{obs}$$

While Assumption 5 is often invoked to justify imputation estimators, it is insufficient to correctly identify the ATE. Consider the following toy example.

Example 3.1 (Bivariate OLS) *Consider a bivariate covariate setting, where $X := \{X_1, X_2\}$. Assume X_1 is fully observed. Assume X_2 consists of some missing values. In an oracle setting, in which researchers could observe all of X, they would estimate the following regression:*

$$Y = \hat{\tau}_{oracle} Z_i + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\varepsilon}_i^{oracle}$$

Instead, because X_2 has missing values, researchers first impute X_2 using the observed covariate X_1. Define $\tilde{X}_2 := X_2 R_i + \hat{g}(X_1)(1 - R_i)$, where $\hat{g}(X_1)$ represents an imputation model using X_1 to predict missing values of X_2. The following regression is then estimated:

$$Y_i = \hat{\tau}_{impute} Z_i + \hat{\beta}_1^{impute} X_{1i} + \hat{\beta}_2^{impute} \tilde{X}_{2i} + \hat{\beta}_R R_i + \hat{\varepsilon}_i^{impute}.$$

Then, without further assumptions,

$$\hat{\tau}_{impute} = \hat{\tau}_{oracle} + \hat{\beta}_2 \cdot \hat{\delta}_2, \tag{1}$$

where $\hat{\delta}_2 := cov(Z_i^{\perp\{X_{1i}, \tilde{X}_{2i}, R_i\}}, X_{2i}^{\perp\{X_{1i}, \tilde{X}_{2i}, R_i\}}) / var(Z_i^{\perp\{X_{1i}, \tilde{X}_{2i}, R_i\}})$. As such, $\hat{\tau}_{impute} \neq \hat{\tau}_{oracle}$.

Example 3.1 decomposes the bias in a regression estimator from imputing covariates into (1) the residual imbalance across X_2 (represented by $\hat{\delta}_2$), and (2) how much variation X_2 can explain in the outcome Y (i.e., $\hat{\beta}_2$). Importantly, we see from Example 3.1 that in order for the bias of an imputation estimator to be zero, one of the following conditions must hold: (1) $\hat{\beta}_2 = 0$, or (2) $\hat{\delta}_2 = 0$. In order for either of these conditions to hold (i.e., $\hat{\beta}_2 = 0$ or $\hat{\delta}_2 = 0$), the covariate with missing values (i.e., X_2) would have to not be a confounder.

To provide intuition for why, consider both scenarios. If $\hat{\beta}_2 = 0$, then the covariate X_2 is completely unrelated to the outcome. This would imply that X_2 is in fact not a confounder, and did not need to be included in the estimation to begin with. Similarly, if $\hat{\delta}_2 = 0$, the imputed values in X_2 would have to be sufficient to account for the confounding in treatment. Because the imputed values of X_2 are a function of the observed covariate X_1, this implies that X_1 would have been sufficient to account for confounding in treatment assignment.

Thus, Example 3.1 highlights that for the imputed estimator to be unbiased, the covariate with missing values would have to not be a confounder. This is true, *regardless* of if missing-at-random holds. While Example 3.1 focuses on the bias in a regression estimator when using imputation, this intuition is true, regardless of estimation approach.[3] This parallels the findings from the measurement error literature (e.g., Bound et al., 2001), in which we can view the imputed observations as corrupted, or mis-measured, observations.

To formalize, in order for an imputation estimator to be an unbiased estimator for the ATE, the following identifying assumption must hold.

Assumption 6 (Modified Selection on Observables (Rosenbaum and Rubin, 1984))

$$Y_i(1), Y_i(0) \perp\!\!\!\perp Z_i \mid \{X_i^{obs}, \hat{X}_i^{mis}\},$$

where $\hat{X}_i^{mis} := g(X_i^{obs})$,[4] *and represents the imputation model used to predict the missing values of* X_i^{mis}. *This is equivalent to assuming the following:*

$$Y_i(1), Y_i(0) \perp\!\!\!\perp Z_i \mid \{X_i^{obs}, R_i\}.$$

[3] We show in Appendix A that the error in an imputed weighted estimator will equivalently depend on whether the imputed values $\hat{X}_2$ (which are a function of X_1) are sufficient for explaining the confounding across the missing observations.

[4] In many traditional imputation approaches, the imputation function for missing covariates will also take in outcomes Y as an input (i.e., $g(X_i^{obs}, Y)$. In the context of observational causal inference, imputing the covariates with outcome information will result in conditioning on post-treatment information.

Assumption 6 is akin to the modified selection on observables assumption introduced in Rosenbaum and Rubin (1984) and Mayer et al. (2023), and implies that across the fully observed units ($R_i = \mathbb{1}_p$), the full set of covariates are necessary for the selection on observables. However, across the partially observed units $R_i = 0$, only the observed covariates X_i^{obs} is necessary for the selection on observables assumption to hold. This is equivalent to assuming that the observed covariates X_i^{obs} and missingness pattern R_i are sufficient for explaining the confounding effects of selection into treatment.

Consider the earlier example in examining the impact of exposure to polarizing news articles on vote choice, where researchers are concerned about non-response in more sensitive survey questions on policy preferences. Assumption 6 would imply that in settings where individuals have fully responded to all questions, their full set of policy preferences would impact both exposure to news articles and vote choice. However, in settings where individuals have chosen to not respond to the survey questions, their missing policy preferences would no longer impact exposure to news articles and/or vote choice.

The plausibility of Assumption 6 is similarly limited outside of survey settings with sensitive questions. Consider an alternative setting in which researchers are studying the impact of exposure to violence on support for a peace deal (e.g., as studied in Hazlett and Parente, 2023). In these settings, an important pre-treatment variable to control for is historical violent incidents, which can impact both the exposure to violence and support for the peace deal. However, regions exposed to more conflict historically may also have larger amounts of missing covariate data, as high-quality data collection can be challenging. Assumption 6 would suggest that for regions with fully observed covariate data, historical violent incidents would impact exposure to violence; however, when the data is missing, then historical violent incidents will no longer impact present-day exposure to violence.

Problematically, from Assumption 6, we see that the unbiasedness and consistency of an imputation estimator does not depend on the imputation model, or whether missing-at-random holds. Assumption 6 implies that we could estimate the ATE using the observed covariates X^{obs}, effectively ignoring the missingness.[5] In other words, whether or not we impute provides *no* additional

[5] More concretely, consider a setting with only two covariates X_1 and X_2, where X_1 is fully observed, but X_2 has missing values. Then Assumption 6 implies that a valid estimation strategy would be to first estimate the ATE across the subset of units that are complete cases (where both X_1 and X_2 are observed). Then, estimate the ATE across

benefit to identifying the ATE when using imputation.

Relationship with Alternative Imputation Approaches. We will consider two alternative approaches to the imputation approach described above. First, researchers often use *multiple imputation* directly to recover the missing outcome values (Westreich et al., 2015). While certain multiple imputation approaches allow researchers to relax underlying parametric assumptions about the relationship between the outcomes and the covariates, the same underlying assumption as Assumption 6 must hold in order to unbiasedly recover the distribution of $Y_i(1)$ and $Y_i(0)$. In other words, the *observed* covariates must be sufficient to fully recover the missing outcome distributions. Multiple imputation is a powerful tool for recovering missing values in datasets in many settings. However, when applied to observational causal inference—where the goal is to identify causal estimands (such as the ATE) without bias—naively deploying a multiple imputation estimator is insufficient. Second, an alternative approach is the *missing indicator approach.* In this setting, instead of imputing the missing covariates with a particular model, a constant value is used to fill in the missing covariate values. The missing indicator approach can thus be thought of as a special case of the imputation estimator we discuss in the paper.

4 Proposed Method: Two-stage Projection Estimator

In the following section, we propose an alternative approach to estimating the ATE when there is missing data in the underlying confounders. We leverage an alternative identification assumption, *outcome-response ignorability.* While this identification approach has been introduced in recent statistics literature (e.g., Miao and Tchetgen Tchetgen, 2016; Yang et al., 2019; Sun and Liu, 2021), the existing estimation approaches are computationally intensive and often infeasible to leverage in practice. Instead, we propose a two-stage projection estimator, which allows researchers to leverage the benefits of a well-estimated imputation model.

units that have missing X_2 values using only X_1, and combine the estimates together. This effectively bypasses the need to impute the missing values of X_2. This is distinct from using a complete case estimator, which would only estimate the ATE across the complete cases.

4.1 Outcome-Response Ignorability

To begin, we introduce an alternative assumption for identification. The alternative identifying assumption relies on assuming ignorability between the observed outcomes and the missingness indicator R_i, given the treatment assignment indicator Z_i and a set of pre-treatment covariates X_i.

Assumption 7 (Outcome-Response Ignorability)

$$R_i \perp\!\!\!\perp Y_i \mid \{Z_i, X_i\}$$

Assumption 7 was initially introduced in Yang et al. (2019),[6] and implies that given X_i and Z_i, the mean value of the outcome will be equal, across the fully observed cases and cases with missing values (i.e., $\mathbb{E}(Y_i \mid Z_i = z, X_i = x, R_i = \mathbb{1}_p) = \mathbb{E}(Y_i \mid Z_i = z, X_i = x, R_i \neq \mathbb{1}_p) \equiv \mathbb{E}(Y_i \mid Z_i = z, X_i = x)$). Assumption 7 is different from missing completely at random, which assumes that R_i is independent of the outcomes Y_i, unconditionally. Instead, Assumption 7 allows for differential response across the pre-treatment covariates X_i, as well as treatment Z_i.

Consider again the earlier example of the researcher studying the impact of exposure to polarizing news articles on vote choice. Outcome-response ignorability assumes that conditioning on an individual's underlying policy preference–even if unobserved, whether or not an individual responds to the survey policy question will not impact the vote choice. Similarly, in the setting of the researcher studying conflict and the impact of violence on support for peace, Assumption 7 states that conditioning on a region having the same historical violence incidents, whether or not historical violent incidents is missing will not impact the outcome–i.e., support for peace.

To our knowledge, Assumption 7 has not been leveraged in the political science literature for missing data.[7] Assumption 7 allows us to directly identify the conditional average treatment effect, given $X_i = x$, using the complete case data.

Lemma 4.1 (Identification of the CATE under Outcome-Response Ignorability) *Under*

[6] A different strand of literature in statistics introduces a similar assumption, known as a shadow variables assumption to recover the mean of an outcome variable missing not at random (e.g., Miao and Tchetgen Tchetgen, 2016; Miao et al., 2024). This work relies on leveraging an auxiliary variable that can serve as a proxy variable that will affect the missingness in the *outcome variable* through its association with the outcome. We can view outcome-response ignorability as leveraging the missingness patterns in the covariates (i.e., R_i) as a shadow variable.

[7] We searched for this assumption across the top political science journals, and also checked all papers that cited the original Yang et al. (2019) paper.

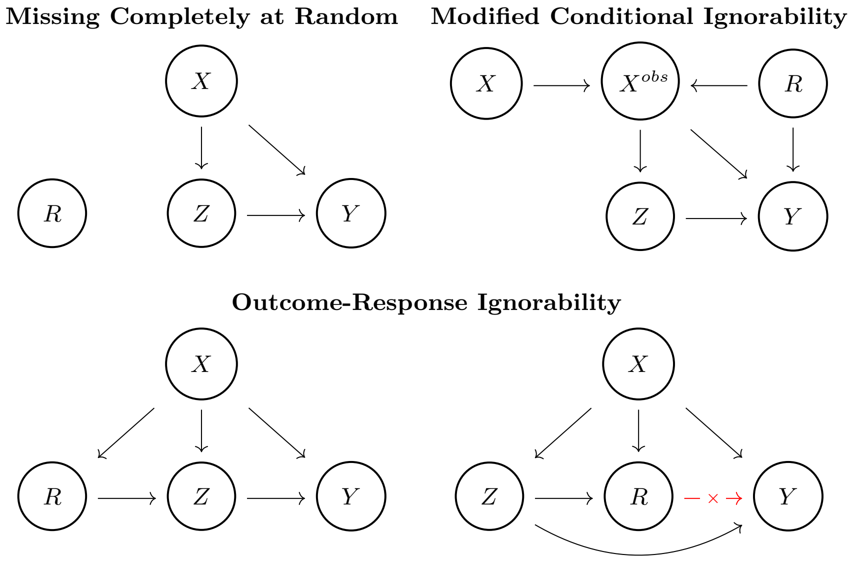

Missing Completely at Random
X
R
Z
Y
Modified Conditional Ignorability
X
X^{obs}
R
Z
Y
Outcome-Response Ignorability
X
R
Z
Y
X
Z
R
Y

Figure 2: Illustration of different identifying assumptions: The upper left panel depicts MCAR, where the missingness R is independent of all other components in the graph. The upper right panel presents modified conditional ignorability, indicating that X influences the process solely through X_{obs} (see the main text for further discussion on the paradox). The lower panels illustrate that outcome response ignorability permits missingness to occur either before or after treatment.

outcome-response ignorability,

$$\tau(X_i) := \mathbb{E}(Y_i(1) - Y_i(0) \mid X_i) = \mathbb{E}(Y_i \mid X_i, Z_i = 1, R_i = \mathbb{1}_p) - \mathbb{E}(Y_i \mid X_i, Z_i = 0, R_i = \mathbb{1}_p).$$

While the CATE is identified under outcome-response ignorability, to identify the ATE, researchers have to marginalize over the distribution of the covariates X_i, which is not fully observed as a result of the missingness. More concretely, the average treatment effect τ can be written as:

$$\tau : = \int \tau(X_i) f(X_i) d\nu,$$

where $f(X_i)$ represents the full distribution of X_i. While $\tau(X_i)$ (i.e., the CATE) is identified directly from outcome-response ignorability, $f(X_i)$ is not observable, as there are missing values of X_i.

Yang et al. (2019) show that $f(X_i)$ can be non-parametrically identified, which then allows researchers to re-weight the complete case observations using a density ratio. However, *estimating* the density in practice is a challenge and requires solving a computationally infeasible set of estimating equations, outside of the setting when the outcomes Y_i and covariates X_i are all discrete (e.g., Kuroki and Pearl, 2014; McCartan et al., 2024). Yang et al. (2019) leverage different basis expansions of the observed covariates X_i to approximate a solution. However, this is computationally challenging, especially in higher dimensional settings; furthermore, it can be difficult to justify if the density approximation is sufficiently valid. We provide more discussion about the relationship between our proposed approach and the non-parametric approach from Yang et al. (2019) in Appendix A.3.

We propose a two-stage projection estimator, which relies on modeling the outcomes across the complete cases, and then uses an imputation model to approximate the density of $f(X_i)$. If the outcome models are correctly specified and the imputation model consistently recovers density of the missing covariate values, the two-stage projection estimator will be a consistent estimator under Assumption 7. Unlike alternative estimation approaches, the two-stage projection estimator can leverage the power of a well-estimated imputation model.

4.2 Two-stage Projection Estimator

In the following subsection, we propose a two-stage projection estimator that allows researchers to recover the ATE under outcome-response ignorability. As an overview, we start by estimating outcome models for both the treatment and control outcomes across the fully observed data to model the relationship between the outcomes and the covariates. This is effectively equivalent to a standard complete case estimator. We then impute the missing covariate data across the subset of units with missing data. In the second stage of the projection estimator, we use the estimated model from the first stage to predict the outcomes across the incomplete cases, using the imputed covariates. We informally refer to this second stage as a 'projection' step. In the case of a linear regression, this corresponds to multiplying the estimated coefficients by the imputed covariate values to obtain predicted outcomes. The same idea extends naturally to more complex machine learning models, where fitted parameters are applied to the imputed covariates to generate predictions.

The projection step target the missingness in the covariates by projecting the estimated model from the first step into the data with missing values. Intuitively, under outcome-response ignorability, the relationship between the outcome and the covariates will be stable across the

Step 1. Split the data into two subsets: (1) the complete cases (i.e., $R_i = \mathbb{1}_p$) and (2) cases with missing X_i observations. Denote these as $\mathcal{S}_1$ and $\mathcal{S}_0$, respectively.

Step 2. Across $\mathcal{S}_1$, estimate outcome models across both the treatment and control units: $\hat{m}_1(X_i; \mathcal{S}_1)$, $\hat{m}_0(X_i; \mathcal{S}_1)$. For simplicity, we will denote $\hat{\tau}(X_i; \mathcal{S}_1) := \hat{m}_1(X_i; \mathcal{S}_1) - \hat{m}_0(X_i; \mathcal{S}_1)$.

Step 3. Estimate an imputation model $\hat{g}$ across $\mathcal{S}_1$ to generate the imputed covariates $\tilde{X}_i$ in $\mathcal{S}_0$.

Step 4. Using the imputed covariates $\tilde{X}_i$ and the estimated outcome models, generate predictions of the conditional average treatment effect across $\mathcal{S}_0$ (i.e., $\hat{\tau}(\tilde{X}_i; \mathcal{S}_1)$).

Step 5. Combine together to compute the projection estimator:

$$\hat{\tau}_{proj} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{m}_1(\tilde{X}_i; \mathcal{S}_1) - \hat{m}_0(\tilde{X}_i; \mathcal{S}_1) \right\}$$

Table 1: Steps for estimating the proposed two-stage projection estimator.

observed and the missing data. The second step thus leverages an imputation model to re-construct the missing data in the pre-treatment covariates. Table 1 summarizes the procedure for estimating the two-stage projection estimator.

The two-stage projection estimator can be written as the sum of the complete case estimator and a projected component:

$$\begin{aligned}\hat{\tau}_{proj} &= \frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{m}_1(\tilde{X}_i; \mathcal{S}_1) - \hat{m}_0(\tilde{X}_i; \mathcal{S}_1) \right\} \\ &= \underbrace{p_R \cdot \sum_{i=1}^{n} \left\{ \hat{m}_1(X_i; \mathcal{S}_1) - \hat{m}_0(X_i; \mathcal{S}_1) \right\} R_i}_{\propto \text{ Complete Case Estimator}} + \underbrace{(1 - p_R) \cdot \sum_{i=1}^{n} \left\{ \hat{m}_1(\tilde{X}_i; \mathcal{S}_1) - \hat{m}_0(\tilde{X}_i; \mathcal{S}_1) \right\} (1 - R_i)}_{\text{Projected Component}},\end{aligned} \tag{2}$$

where p_R is the proportion of complete cases (i.e., $p_R = \mathbb{E}(\mathbb{1}\{R_i = \mathbb{1}_p\})$), $\mathcal{S}_1$ denotes the subset of observations that are completely observed (i.e., $R_i = \mathbb{1}_p$), and $\hat{\tau}(X_i; \cdot)$ represents the estimated treatment effect.

When the outcome models $\{\hat{m}_1(X_i; \mathcal{S}_1), \hat{m}_0(X_i; \mathcal{S}_1)\}$ and the imputation model is consistently estimated, then the projection estimator will be a consistent estimator for the ATE.

Theorem 4.1 (Consistency of the Projection Estimator) *Assume Assumptions 1-3, and 7 (outcome-response ignorability) hold. Then, assume the following estimation assumptions:*

- *Consistent outcome models:* $\mathbb{E}\left\{\hat{m}_z(X_i;\mathcal{S}_1) - m_z(X_i;\mathcal{S}_1)\right\} = o_p(1)$ *for* $z \in \{0,1\}$*, where* $m_z(X_i;\mathcal{S}_1) := \mathbb{E}\left[Y_i \mid X_i, Z_i = z, R_i = \mathbb{1}_p\right]$.
- *Valid imputation model:* $f(\{X_i^{obs}, \hat{X}_i^{mis}\} \mid R_i \neq \mathbb{1}_p) = f(\{X_i^{obs}, X_i^{mis}\} \mid R_i \neq \mathbb{1}_p)$

Then, the projection estimator will be a consistent estimator for the average treatment effect:

$$\hat{\tau}_{proj} \xrightarrow{p} \tau.$$

The projection estimator requires two additional estimation assumptions for consistency. The first estimation assumption is that the outcome model is correctly spceified across the complete cases. Notably, alternative approaches (i.e., complete case estimators or the imputation estimator) require additional estimation assumptions about the consistency of either the propensity score model and outcome model across *all* units–even those with missing confounder data. In contrast, the outcome-response ignorability estimators require consistency of the outcome model only across the complete cases. The second estimation assumption needed is that the imputation model sufficiently approximates the density of the underlying missing covariates. Imputation models like Amelia or MICE will satisfy this condition in settings when missing-at-random holds (Blackwell et al., 2012). In other words, unlike the previously discussed estimators (i.e., complete case estimator, the imputation estimator, and the weighted estimator), the projection estimator is able to benefit from a well-estimated imputation model. Table 2 summarizes the different estimators, and their corresponding identification and estimation assumptions.

To compute the variance of $\hat{\tau}_{proj}$, we must account for both the variation in the outcome model estimation, as well as the uncertainty in the imputation model. If researchers are using a parametric approach to both imputation and outcome modeling, then we can apply standard M-estimation approaches to compute a closed form representation of the variance (Lunceford and Davidian, 2004). However, because many imputation models rely on black-box approaches, for which the closed-form standard errors are not readily available, we recommend researchers use a bootstrap approach to account for uncertainty, which would flexibly capture the variability from imputation, even when the underlying imputation model is complex or nonparametric.

Estimator	**Identification Assumption**	**Estimation Assumptions**
Complete Case	Missing Completely at Random	Consistent propensity score model (across *all* units) (i.e., $\hat{e}(X_i; R_i = \mathbb{1}_p) \xrightarrow{p} \Pr(Z_i = 1 \mid X_i)$) and/or consistent outcome model (across *all* units) (i.e., $\hat{m}_z(X_i; R_i = \mathbb{1}_p) \xrightarrow{p} \mathbb{E}(Y_i \mid X_i, Z_i = z)$)
Imputation Estimator	Modified Selection-on-Observables	Consistent propensity score model (across *all* units) (i.e., $\hat{e}(X_i \cdot R_i + \tilde{X}_i \cdot (1 - R_i)) \xrightarrow{p} \Pr(Z_i = 1 \mid X_i)$) and/or consistent outcome model (across *all* units) (i.e., $\hat{m}_z(X_i \cdot R_i + \tilde{X}_i \cdot (1 - R_i)) \xrightarrow{p} \mathbb{E}(Y_i \mid X_i, Z_i = z)$)
Two-Stage Projection	Outcome-Response Ignorability	1. Consistent outcome models (across complete cases) (i.e., $\hat{m}_1(X_i; \mathcal{S}_1) \xrightarrow{p} \mathbb{E}(Y_i \mid X_i, Z_i = 1, R_i = \mathbb{1}_p)$ and $\hat{m}_0(X_i; \mathcal{S}_1) \xrightarrow{p} \mathbb{E}(Y_i \mid X_i, Z_i = 0, R_i = \mathbb{1}_p)$) 2. Valid imputation model (i.e., $f(\tilde{X}_i \mid R_i \neq \mathbb{1}_p) = f(X_i \mid R_i \neq \mathbb{1}_p)$)

Table 2: We assume across all estimators, conditional ignorability and positivity holds.

Extensions for other estimation approaches. The two-stage projection estimator relies on outcome modeling and an imputation model to recover the ATE. There are alternative estimation approaches that leverage outcome-response ignorability. For example, instead of relying on an imputation model to approximate the density $f(X_i)$ or a non-parametric approach like in Yang et al. (2019), Sun and Liu (2021) propose using different parametric models that target the joint density of treatment assignment and being a complete case. However, this requires being able to consistently model both the propensity of receiving treatment, as well as the probability of being a complete case (i.e., $\Pr(R_i = \mathbb{1}_p \mid X_i, Z_i = z)$), which can be challenging in practice. Furthermore, researchers can combine the weighting approach from Sun and Liu (2021) with the two-stage projection estimator to construct a doubly robust augmented weighted estimator. We provide more discussion in Appendix A.4. We compare the performance of the two-stage projection estimator to these alternative approaches in simulations and find that the weighting estimator can be unstable, resulting in highly variable estimates.

5 Robustness Checks for the Projection Estimator

While we argue that the projection estimator relies on weaker underlying assumptions than the complete case or imputation estimators, the validity of the projection estimator does rely on both outcome-ignorability and missing-at-random assumptions. In the following section, we provide a suite of diagnostic tools for researchers to use to evaluate the plausibility of these assumptions in practice.

5.1 Validating the Imputation Model

Unlike the alternative estimation approaches, the two-stage projection estimator allows researchers to leverage a well-estimated imputation model. However, this is a double-edged sword. In settings when the imputation model does a poor job recovering the missing covariates, this can mean that the projection estimator will be susceptible to potential bias. To help address this concern, we propose a validation procedure for researchers to evaluate the stability and performance of the underlying imputation model. In particular, we leverage the complete case observations, and construct different common missingness mechanisms. Using the constructed missingness processes, we can then (1) mask the covariate observations across the complete cases, (2) estimate an imputation model, and (3) compare the imputed covariate values with the true covariate values. Different imputation models have different underlying assumptions. The proposed validation procedure is helpful in considering how sensitive the underlying performance of the imputation model is to potential perturbations from the assumed data generating process.

Our proposed procedure is similar to the idea of overimputation (e.g., Blackwell et al., 2012). However, instead of performing a leave-one-out evaluation by sequentially omitting each individual covariate observation, we proposed generating different missingness mechanisms. We recommend researchers, at a minimum, evaluate several common missingness mechanisms, summarized below.

Missing-at-random. The first missingness mechanism researchers should evaluate is missingness at random. If an imputation model is unable to perform well under MAR, then this likely means that the imputation results should be considered with great caution, as even when the underlying MAR assumption holds, the model is unable to correctly recover the missing covariates well. This could occur if the signal to noise ratio is too low.

For each covariate $X_i^{(j)}$ for $j \in \{1, ..., p\}$, construct a generalized linear model for the missingness indicator $\tilde{R}_{(j)}$, using that is a function of the other covariates (i.e., $X_i^{-(j)}$):

$$\Pr(\tilde{R}_{(j)}^{\text{MAR}} \mid X_i^{-(j)}; \alpha^{(j)}) = h^{-1}\left(\sum_{k \neq j} \beta_k X_i^{(k)} + \alpha^{(j)}\right), \tag{3}$$

where $\alpha^{(j)}$ represents the scale parameter and h is a link function. Researchers should calibrate $\alpha^{(j)}$ such that the proportion of simulated missing values matches the true proportion of missing values. A useful approach is to modify the link function to evaluate the impact of distributional assumptions in the underlying imputation model.

Missing-Not-at-Random. While model specification choices can result in biased result, a larger driver of bias in imputation models is violations of the underlying missing-at-random assumption. As such, we recommend researchers generate missingness mechanisms that explicitly violate the missing-at-random assumption to evaluate the performance of the imputation model under different, adversarial missing-not-at-random data generating processes. To generate a missing-not-at-random process, researchers can directly add in the covariate $X_{(j)}$ into the generalized linear model:

$$\Pr(\tilde{R}_{(j)}^{\text{MNAR}} \mid X_i^{-(j)}; \alpha^{(j)}) = h^{-1}\left(\sum_{k \neq j} \beta_k X_i^{(k)} + \gamma \cdot X_i^{(j)} + \alpha^{(j)}\right), \tag{4}$$

where $|\gamma|$ controls how strong the violation in missing-at-random is. We recommend researchers evaluate different values of γ to see the impact of missing-not-at-random on the imputation model's performance. If, for very small values of γ, the imputation model incurs large amounts of error, this implies that there is a large sensitivity in the imputation results.

Near violations of positivity in missingness. For large γ values in Equation (4), researchers can also simulate a missingness mechanism that results in near violations of positivity in missingness. As $|\gamma|$ increases, this implies that for large (or small) values of covariate $X^{(j)}$, the probability of missingness will be larger. For the imputation model to recover missing covariate values in the presence of near violations of positivity in missingness, the imputation model must correctly extrapolate beyond the convex hull of the observed covariates. Simulating near violations of positivity in missingness can be helpful to evaluate the imputation model's ability to extrapolate.

For different missingness mechanisms, researchers can repeatedly mask the observed covariates, and perform imputation. They can then compare the imputed values with the observed covariate values to evaluate the imputation error. Table 3 summarizes. We recommend researchers normalize the estimated error to interpret the relative error from imputation. For example, if considering the mean absolute error, researchers could normalize by the mean covariate value. The normalized mean absolute error would correspond to a percentage error in recovering the covariate values. A normalized mean absolute error of 0.1 would imply that on average, the imputed covariate values are off by about 10% of the average value.[8]

A low estimated imputation error does not necessarily *imply* that researchers have successfully recovered the missing covariate data. However, it helps provide researchers with a sense of the performance of the imputation model. For example, if researchers find that even under missing-at-random, there is a relatively large amount of error in recovering the missing covariate values with the imputation model, this implies that the other covariates are likely not sufficiently explanatory of the missing covariate.

Table 3: Proposed validation procedure for an imputation model

For a chosen missingness process:

Step 1. Subset the dataset to the complete cases (i.e., $R_i = \mathbb{1}_p$), where consistent with Section 4.2, we denote this as $\mathcal{S}_1$.

Step 2. Across $\mathcal{S}_1$, generate missing values for the covariates (i.e., $\tilde{R}_i$).

Step 3. Estimate an imputation model.

Step 4. Compare the imputed covariate values with the true covariate values:

$$\text{Error}(\hat{X}_i, X_i \mid \tilde{R}_i)$$

5.2 Observable implications of Outcome-Response Ignorability

While researchers can never check if outcome response ignorability holds, there are *observable implications* of the assumption that can be evaluated. In particular, under outcome-repsonse ignorability, the average value of Y_i, given X_i and Z_i will be equal across the fully observed cases

[8] Researchers can alternatively normalize by other measures. For example, if they wish to interpret the imputation error with respect to the scale of the original data, they should normalize by the standard deviation of the covariate. In such a setting, a normalized mean absolute error of 1 would imply that the imputation model is likely not capturing much of variability in the covariate values, and is performing similarly to a mean imputation.

and cases with missing values:

$$\mathbb{E}(Y_i \mid Z_i = z, X_i = x, R_i = \mathbb{1}_p) = \mathbb{E}(Y_i \mid Z_i = z, X_i = x, R_i \neq \mathbb{1}_p)$$

Within the treatment and control groups, researchers can compare the projected outcomes with the observed outcomes. If the projected outcomes are centered at the observed outcomes, then this provides credibility in the plausibility of outcome-response ignorability.

If the projected outcomes are not similar to the observed outcomes, then this implies several issues could be present. First, outcome-response ignorability could be violated. As such, there are shifts in the underlying distribution of the outcomes across the missing and fully observed cases. Second, the error in the imputed covariates $\tilde{X}_i$ is resulting in an incorrect projection of the outcomes. As a result, even if outcome-response ignorability holds, since the input of covariates is wrong, and the resulting projected outcome value is incorrect. Finally, the underlying outcome model could be misspecified to begin with, resulting in poor performance.

While we cannot precisely diagnose which scenario is occurring in practice, a helpful check is to compare the goodness-of-fit of the outcome model across the complete cases (i.e., $R_i = \mathbb{1}_p$) and the goodness-of-fit of the outcome model across the projected, incomplete cases (i.e., $R_i \neq \mathbb{1}_p$). If researchers see that within the complete cases, there is a relatively high goodness-of-fit, but the goodness-of-fit deteriorates across the projected, incomplete cases, this indicates that there is likely a shift in the underlying distribution of the outcomes across the missing and fully observed cases, or the imputed covariates are resulting in an incorrect projection. In contrast, if researchers see that the goodness-of-fit is poor in both the complete cases and the incomplete cases, then this is a sign that the underlying outcome model is not sufficiently accounting for enough variation in the outcome process.

In a special setting when the outcome and covariates are all categorical, researchers can directly test for observable implications of violations in outcome-response ignorability (Sjölander and Hägg, 2025). This is done by first computing a set of constraints that are implied by outcome-response ignorability and seeing if the observed data violate the given constraints. If the observed data violate the constraints, this implies that outcome-response ignorability has been violated. We refer readers to Sjölander and Hägg (2025) for more details.

5.3 Sensitivity Analysis

Finally, we propose a sensitivity analysis for the projection estimator that allows researchers to bound the range of possible estimates under violations of outcome-response ignorability and missing-at-random. Sensitivity analyses allow researchers to consider how robust their results are to potential violations in the underlying analysis. If small assumption violations cause large changes in the estimated treatment effect, the result is sensitive; if only large violations matter, the estimate is relatively robust. In the following subsection, we provide an overview of the proposed sensitivity analysis, with technical details in Appendix A.5.

We introduce two parameters that control the following: (1) the amount of imputation error for a given covariate; and (2) how different the relationship is between the outcomes and covariates across the missing subset $R_i \neq \mathbb{1}_p$, in comparison to the complete case subset $R_i = \mathbb{1}_p$. We provide more details below.

Varying the imputation error. As the amount of imputation error increases, this implies a greater deviation from missing-at-random, which will result in a larger amount of error in the projection step of the estimator. One way researchers can calibrate how much imputation error to consider within the sensitivity analysis is to leverage the results from the validation exercise proposed in Section 5.1. In particular, researchers can use the estimated error incurred from a specified missingness mechanism (i.e., small amounts of MNAR) to reason about what is a plausible imputation error for each covariate. They can vary how many times larger the imputation error is, relative to the calibrated imputation error.

Varying the relationship between the outcomes and covariates. To quantify the differences in the relationship between the outcomes and covariates across the complete and incomplete subsets, we consider the deviation between $\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, R_i = \mathbb{1}_p]$ and $\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, R_i \neq \mathbb{1}_p]$:

$$\Gamma(x) := \frac{\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, R_i \neq \mathbb{1}_p]}{\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, R_i = \mathbb{1}_p]} \leq \Gamma. \tag{5}$$

When outcome-response ignorability holds, then $\Gamma = 1$. As Γ deviates further from 1, then this implies there is a greater difference in the relationship between the outcomes and covariates across the two subsets of data, and there is a larger violation in the underlying outcome-response ignorability assumption. In other words, we are unable to learn much about the missing subset of data from the complete cases.

6 Simulations

Across all simulations, we find that under MAR missingness, the projection estimator unbiasedly recovers the ATE and is far more efficient than weighting methods using outcome–response ignorability. Under MNAR, the two-stage estimator shows some bias due to imperfect imputation, but its bias remains far smaller than that of complete-case or standard imputation estimators.

6.1 Set-Up

We provide a summary of the simulation set-up, with additional details in Appendix C. For simplicity, we consider a bivariate setting, with $\{X_i^{(1)}, X_i^{(2)}\}$. We let $X_i^{(1)}$ and $X_i^{(2)}$ both be standard normal random variables, with a set correlation $\text{cor}(X_i^{(1)}, X_i^{(2)}) = \rho$. Define the outcome as

$$Y_i = \tau \cdot Z_i + \sum_{j=1}^{2} \left\{\beta_j X_i^{(j)} + \varphi_j \big(X_i^{(j)} \cdot Z_i\big)\right\} + u_i, \text{ where } u_i \sim N(0,1),$$

and the treatment assignment process as

$$\Pr(Z_i = 1 \mid X_i) = \frac{1}{1 + \exp\left\{-\left(\sum_{j=1}^{2} \gamma_j X_i^{(j)}\right)\right\}}.$$

We assume that $X_i^{(1)}$ is fully observed for all units, while $X_i^{(2)}$ has missing values. The missingness probability is defined as

$$\Pr(R_i = \mathbb{1}_p \mid X_i) = \frac{1}{1 + \exp\left\{-\left(X_i^{(1)} + \alpha X_i^{(2)}\right)\right\}}.$$

We consider two scenarios: (1) $\alpha = 0$, such that missingness in $X_i^{(2)}$ can be fully explained by variation in $X_i^{(1)}$ (i.e., missing-at-random holds), and (2) $\alpha = 1$, where missingness in $X_i^{(2)}$ depends on the values of $X_i^{(2)}$ (i.e., missing-not-at-random). Within each scenario, we vary the correlation between $X_i^{(1)}$ and $X_i^{(2)}$, which proxies how well we are able to recover the missing values in $X_i^{(2)}$ with an imputation model. In general, $\Pr(R_i = 1) = 0.5$ across the simulations, indicating that above half of the observations in $X_i^{(2)}$ are missing.

For each simulation iteration, we estimate seven different estimators: (1) the complete case estimator, (2) the imputation estimator, (3) the missing indicator estimator, (4) a multiple imputation estimator, (5) a parametric, weighted estimator following Yang et al. (2019) and Sun and Liu (2021), (6) the proposed, two-stage projection estimator, and finally (7) an augmented weighted estimator.

For the outcome models used in the complete case, imputation, and two-stage projection estimators, we model the outcomes using an interacted linear regression. For the imputation estimator, we first impute the missing values in $X^{(2)}$, and then estimate the outcome model by including $X^{(1)}$, the imputed $\hat{X}^{(2)}$, and a missingness indicator for whether or not $X^{(2)}$ was observed or not. To impute the missing covariate values, we use a linear regression between $X^{(2)}$ and $X^{(1)}$. The missing indicator estimator uses the same procedure, but mean imputes the missing values in $X^{(2)}$ instead of relying on an imputation model. For the multiple imputation estimator, we separate the data into the treatment and control subsets, and then impute both the missing covariate values, as well as the missing counterfactual outcomes simultaneously, using Amelia (Honaker et al., 2011). This effectively uses a black-box modeling approach without relying on a model specification. The parametric weighted estimator relies on a logistic regression specification, and then utilizes a generalized method of moments framework to solve for the joint probability between missingness and treatment. We provide more details on all of the different estimation approaches used in the simulations in Appendix C.

6.2 Simulation results

There are several key takeaways to highlight from the simulations. First, under missing-at-random, the proposed two-stage projection estimator is unbiased, even for low values of ρ. When the covariate values are missing-not-at-random, the two-stage projection estimator will be biased; however, notably, the bias incurred by the two-stage projection estimator will be lower than the bias in the complete case and imputation estimators. Furthermore, we see that in settings with sufficiently high correlation between $X^{(1)}$ and $X^{(2)}$, even under MNAR, the bias from the two-stage projection estimator is relatively small (i.e., ranging from $0.08 - 0.14$, in settings where $\rho \geq 0.5$).

Second, consistent with the theoretical results in Section 3, we find that irrespective of how high ρ is and whether or not $X_i^{(2)}$ is missing-at-random or not, the imputation estimators (i.e., both the standard imputation estimator, as well as the multiple imputation estimator) are all substantially biased. Interestingly, we see that in settings when the missingness mechanism is missing-at-random, the multiple imputation estimator provides slightly less biased estimates than the imputation estimator. However, in settings when the missingness mechanism is missing-not-at-random, then using the multiple imputation estimator results in larger amounts of bias than a standard imputation estimator. This re-iterates the fact that missing-at-random is not a sufficient assumption to identify the ATE, and that using imputation estimators in practice warrants careful consideration.

Third, we compare the projection estimator to a parametric weighted estimator that also leverages outcome-response ignorability. Unlike the projection estimator, which relies on outcome modeling and an imputation model, the weighted estimator accounts for the distribution shift from missingness. In theory, the weighted estimator should still be consistent in settings when there is missing-not-at-random, but requires researchers consistently estimate the probability of missingness. We find in the simulations that while the weighted estimator is unbiased even in settings when covariate values are missing-not-at-random, there is a large degree of variance inflation from the re-weighting. The instability of the estimator is also amplified in settings when the covariates are correlated to one another. We find that the resulting mean squared error of the weighted estimator is substantially higher than the projection estimator across the different simulation settings. Augmenting the weighted estimator (i.e., the augmented weighted estimator) helps stabilize the resulting estimates substantially, as it leverages information from the outcome model. We find that the augmented weighted estimator performs similarly to the projection estimator, though with slightly more variance, as a result of the weighting.[9]

[9] The similarity in performance between the augmented weighted estimator and the projection estimator arise from the fact that in the two settings considered here, the outcome model is correctly specified. In Appendix C, we consider settings with model misspecification. In those settings, we find that the augmented weighted estimator can

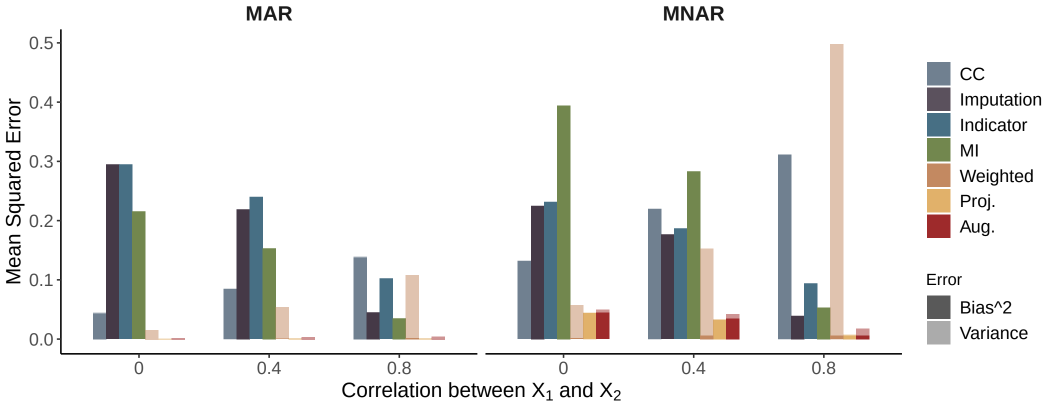


Figure 3: We plot the mean squared error of the different estimators in the simulation study. We decompose the mean squared error into the squared bias and variance associated with each estimator.

In Appendix C, we consider more complex simulation settings, such as settings with model misspecification and alternative data generating processes, and find that the general patterns hold across these different settings. We also evaluate the validity of the bootstrapped standard errors, and find that the proposed two-stage projection estimator is able to obtain nominal coverage.

7 Application: Evaluating Transparency and Political Compromise in State Legislatures

7.1 Background

In a recent study, Harden and Kirkland (2021) studies the influence of transparency laws on partisanship and budgetary changes within state legislatures. The authors are interested in understanding the impact of "sunshine laws", which require government agency meetings to be open to the public. Certain states exempt (partially or fully) legislatures from the sunshine law citing efficiency reasons, while others have these laws in place. The treatment in the study is defined as the presence of an open meeting requirement for state legislatures in a given year. To consider partisanship and the level of budgetary changes, the authors consider a variety of outcomes, such as proportion of bills enacted, budget kurtosis , and polarization. To control for potential confounding, the authors incorporate a set of pre-treatment covariates, which includes the number of bills voted on, level of professionalism, state ideology, governmental ideology, Ranney index, presence of term limits,

provide some protection against outcome model misspecification, if the estimated weights are correctly specified.

state population, the Gross State Product (GSP), and the legislative expenditures.

Many of the pre-treatment covariates exhibit missing values. The authors employ multiple imputation techniques on the data (Honaker et al., 2011), and subsequently present the results using imputed datasets. In the original study, the authors find little to no impact of the presence of transparency laws on the outcomes of interest. In other words, transparency laws do not appear to impact government productivity, or result in higher rates of partisanship. A natural question thus is whether the authors would have arrived at different conclusions, had they employed an alternative missing data estimator. We compare five total estimators: (1) an imputation estimator, following the authors' original specification; (2) a complete case estimator; (3) a multiple imputation estimator; (4) a weighting estimator; (5) our proposed, two-stage projection estimator; and (6) an augmented weighted estimator. For each estimator, we add in the same pre-treatment covariates, as well as time and state-level fixed effects, and estimate standard errors using a non-parametric block bootstrap procedure to account for time and state-level fixed effects. See Appendix D for details on the estimation.

We check for missingness patterns in the covariates across the time and state fixed effects. We find that the state of Nebraska is missing all observations for some of the pre-treatment covariates, indicating a violation in positivity of missingness (see Appendix D for more discussion). As a result, we restrict our study population to exclude the state of Nebraska. As a result, we do not expect the findings to exactly replicate the estimates from the original Harden and Kirkland (2021) paper. However, we note that the substantive conclusions from the imputation estimator in our re-analysis matches the substantive conclusion from the Harden and Kirkland (2021) paper.

7.2 Results

From our analysis, we see that the choice of estimator results in different substantive takeaways. For example, using the same model specification as the original authors (i.e., an imputation estimator with ordinary least squares), we find that there is little to no impact of government transparency in government productivity (as proxied by the proportion of bills enacted by the state legislature) and polarization. In particular, both imputation and the complete case estimators result in a null estimate. In contrast, the projection estimator results in a positive estimate of the impact of government transparency on bill enactment. Similarly, for the outcome of *polarization*, while the imputation estimator resulted in a null estimate, the complete case estimator and the projection estimator estimate a statistically significantly negative impact, implying that government transparency could result in a decrease in polarization in state legislatures. When looking at *kurtosis*, the original study found an increase in the budget kurtosis from government transparency. The projection estimator similarly estimates a positive impact, though the estimated effect is attenuated towards zero. Interestingly, the results from using the projection estimator imply a stronger result than the original paper: there is actually an *improvement* in government productivity and a *decrease* in polarization from enacting government transparency laws.

Because each estimator relies on different assumptions about missing pre-treatment covariates, the substantive conclusions can vary widely. The complete-case estimator requires MCAR, which is unlikely given clear time-patterned missingness (Figures 8 and 9 in appendix). The imputation estimator assumes that complete cases contain all confounders of transparency laws, while incomplete cases require only observed covariates to be confounded. In contrast, the projection, weighted, and augmented weighted estimators rely on outcome–response ignorability, which in this setting means the outcome is independent of missingness conditional on transparency status and full covariates.

The credibility of each estimator thus depends on the plausibility of the underlying assumptions for each estimator. In the following subsection, we walk through the proposed robustness checks for the projection estimator and find that the results are relatively robust to potential violations in the underlying assumptions.

7.3 Robustness checks

To evaluate the sensitivity of the projection estimator to potential violations of the underlying assumptions, we illustrate the proposed robustness checks. See Appendix D for details.

Validating the imputation model. We begin by evaluating the performance of the imputation model. We simulate three different missingness mechanisms for each covariate: (1) missing-at-random; (2) missing-not-at-random (with a slight dependency on the missing covariate values); (3) missing-not-at-random (with a larger dependency on the missing covariate values). For each missingness mechanism, we calibrate the proportion of total missing values to match the true

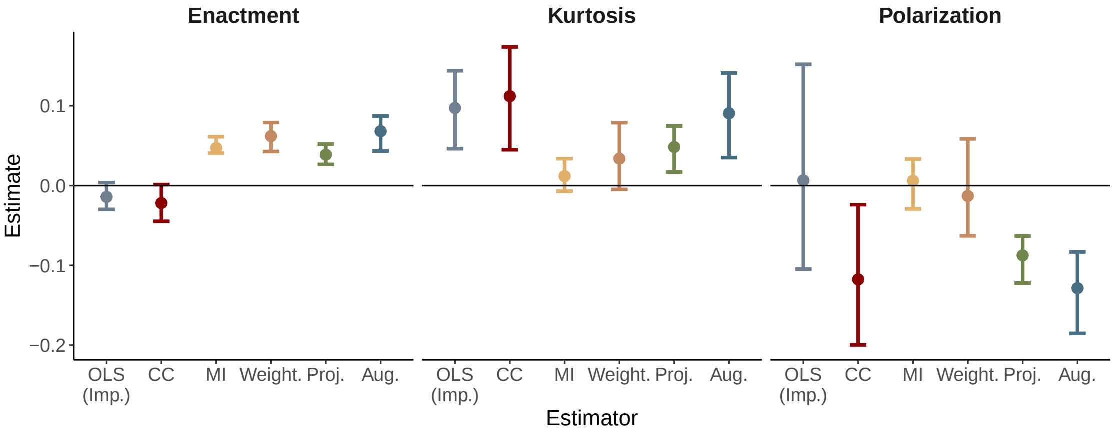


Figure 4: Estimated impact of transparency laws on (1) proportion of bills enacted, (2) budget kurtosis, and (3) polarization. We see that the estimated impact changes depending on the estimator used to account for the missing data in the observed confounders.

proportion of missing values. See Table 8 for the full validation results. As expected, when the missingness mechanism is missing-at-random, the overall error in recovering the covariate values is relatively small, around 0.1 to 0.5 normalized mean absolute error. As we allow for greater violations of missing-at-random by inducing a dependency with the underlying covariate value, the error increases. Interestingly, certain covariates, such as the multidimensional measure of legislature professionalism (Bowen and Greene, 2014), suffer from high rates of error, even when the missingness mechanism is missing-at-random. This is likely because the other covariates included in the model are unable to sufficiently explain the variation in the multidimensional measure, resulting in a large amount of imputation error even under the missingness mechanism is missing-at-random. From examining variable importance measures, we notice that the outcome *kurtosis* has greater dependence on legislature professionalism, which may indicate some degree of sensitivity in the estimated result to potential imputation error.

Observable implications of Outcome-Response Ignorability. We compare the projected values from the underlying outcome model to the observed outcomes in the incomplete cases $R_i \neq \mathbb{1}_p$. The R^2 value for the projected subset is not substantially different from the complete cases. For the outcomes of *proportion of bills enacted* and *polarization*, the R^2 values are substantially higher, ranging around 0.3-0.45. When examining *kurtosis* as an outcome, we see that the R^2 values across both the complete cases and projected subset are relatively low (i.e., around 0.02). This likely implies that the underlying outcome model may not be sufficiently accounting for enough variation in the outcome process.

Sensitivity analysis. We now formally evaluate the sensitivity of the projection estimator to violations in outcome-response ignorability and potential imputation error. We start by setting a constraint on the amount of imputation error present. To do so, we leverage the estimated imputation errors from the validation exercise above to calibrate what a reasonable amount of imputation error there could be for the different covariates. We then vary the parameter Γ, which controls how different the relationship between the outcome and the covariates are between the fully observed complete cases ($R_i = \mathbb{1}_p$) and the missing cases ($R_i \neq \mathbb{1}_p$), and solve for the range of possible values that could exist, given a fixed amount of imputation error and Γ. In general, for the outcomes of *proportion of bills enacted* and *polarization*, the relationship between the covariates and the outcomes would have to change signs for the results to lose statistical significance (represented by $\Gamma < 0$ values).

We also compare the partial identification bounds with the estimates from the imputation estimator. Recall, the projection estimator estimated a positive impact on *proportion of bills enacted* from enacting transparency laws. In contrast, the imputation estimator estimated a negative impact, albeit not statistically significant. We can examine the Γ value for which the partial identification bounds contains the estimate from the imputation estimator. We see that for a fixed amount of imputation error, the relationship between the covariates and outcomes across the $R \neq \mathbb{1}_p$ cases would have to be about twice as large, in the opposite direction, than the estimated relationship across the fully observed cases to recover the same substantive result as the imputation estimator. We observe similar findings for the outcome of *polarization.*

In contrast, for the outcome of *kurtosis*, there is a greater sensitivity to both potential imputation error and violations in outcome-response ignorability. This is in line with our findings from the imputation validation exercise and examining the observable implications of outcome-response ignorability, in which we see a greater dependence on the outcome model for kurtosis on covariates that have a greater degree of imputation error. A Γ value less than -0.5 or greater than 1 would result in a statistically insignificant result. We can compare the bounds with the results from the imputation estimator. In order for the partial identification bounds to contain the

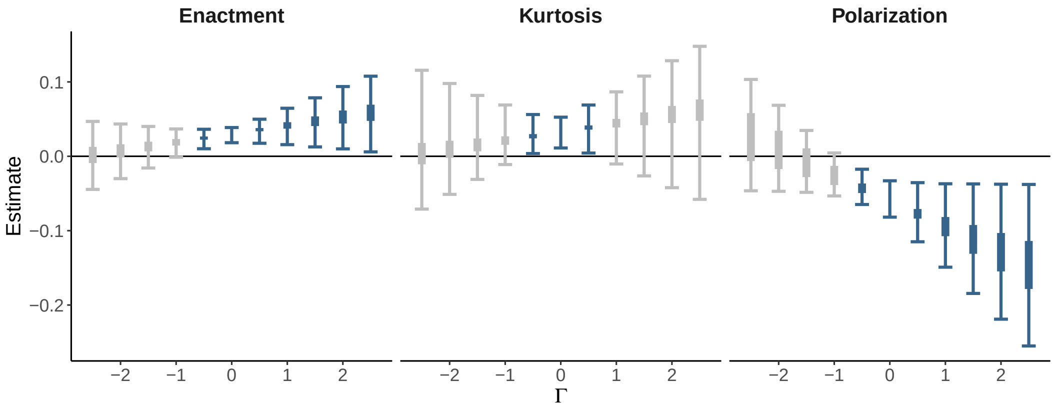


Figure 5: At $\Gamma = 1$, this implies that there is the same relationship between the outcome and covariates across the incomplete cases as the complete cases (i.e., outcome-response ignorability holds). As a result, the bounds represent the range of possible ATE estimates, assuming a fixed amount of imputation error, but no violation in outcome-response ignorability. At $\Gamma = 0$, this implies that there is *no* relationship between the outcomes and the covariates across the incomplete cases. As a result, there is no dependency on the imputation error in the missing covariates.

estimate from the imputation estimator, Γ would have to be greater than 2—i.e., the relationship between the covariates and the outcomes across the incomplete subset would have to be over two times as strong as the estimated relationship across the complete subset. We conclude that while the estimated effect from the projection estimator is relatively sensitive to potential violations in the underlying assumptions, there is a relatively large degree of robustness to *underestimating* the impact of transparency laws on kurtosis.

8 Conclusion

We conclude with a few directions for future work. First, throughout this paper, we have assumed that researchers only have missing confounder values. In other words, the outcomes and treatment assignment indicators are fully observed. In practice, when covariate values are missing, it is also likely that there is outcome and treatment information as well. Future work could account for missingness in outcomes and treatment in observational settings. Furthermore, extensions of this framework should consider missing data in settings when researchers are using instrumental variables to estimate a local average treatment effect, as well as settings with multi-valued and/or continuous treatments. We anticipate that in settings when researchers have continuous treatments, outcome-response ignorability can still be used to identify the treatment effect; however, there may be additional complexities in estimating the conditional average treatment effect for a continuous treatment. Furthermore, throughout the paper, we are currently agnostic as to how researchers perform the outcome modeling. Specific estimation approaches, like tree-based algorithms using surrogate splits, can better accommodate missing data in the covariates. Future work should explore whether there are more robust approaches that can leverage outcome-response ignorability while relaxing estimation assumptions.

Second, a closely related strain of literature considers observations that are not necessarily missing, but subject to measurement error. This can be due to actual measurement issues (i.e., data coding errors, enumerator fixed effects, to name a few), or response biases that arise from respondents masking their true answers on surveys (i.e., social desirability bias) or failing attention checks. While we have not considered settings in which there are measurement errors in the underlying covariates, an interesting future avenue of research could extend the proposed identification and estimation strategy to consider how to account for settings in which there is measurement error in the underlying covariates.

Finally, we have assumed positivity holds throughout. An interesting avenue of future work should consider what happens under violations of positivity of missingness. Recent work in causal inference has introduced sensitivity analyses to consider overlap violations–i.e., settings in which specific subsets of units are systematically missing from a study (Huang, 2024). A similar framework could be considered for positivity violations in missing data.

References

Bang, H. and J. M. Robins (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics 61*(4), 962–973.

Blackwell, M., J. Honaker, and G. King (2012). Multiple overimputation: A unified approach to measurement error and missing data. *URL: http://gking. harvard. edu/files/gking/files/measure. pdf*.

Bound, J., C. Brown, and N. Mathiowetz (2001). Measurement error in survey data. In *Handbook of econometrics*, Volume 5, pp. 3705–3843. Elsevier.

Bowen, D. C. and Z. Greene (2014). Should we measure professionalism with an index? a note on theory and practice in state legislative professionalism research. *State Politics & Policy Quarterly 14*(3), 277–296.

Boyd, S., S. P. Boyd, and L. Vandenberghe (2004). *Convex optimization*. Cambridge university press.

Chang, C.-R., Y. Song, F. Li, and R. Wang (2023). Covariate adjustment in randomized clinical trials with missing covariate and outcome data. *Statistics in Medicine 42*(22), 3919–3935.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. Technical report, National Bureau of Economic Research.

Cinelli, C. and C. Hazlett (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82*(1), 39–67.

Cornfield, J., W. Haenszel, E. C. Hammond, A. M. Lilienfeld, M. B. Shimkin, and E. L. Wynder (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer institute 22*(1), 173–203.

Dorn, J. and K. Guo (2023). Sharp sensitivity analysis for inverse propensity weighting via quantile balancing. *Journal of the American Statistical Association 118*(544), 2645–2657.

Harden, J. J. and J. H. Kirkland (2021). Does transparency inhibit political compromise? *American journal of political science 65*(2), 493–509.

Hazlett, C. and F. Parente (2023). From "is it unconfounded?" to "how much confounding would it take?": Applying the sensitivity-based approach to assess causes of support for peace in colombia. *The Journal of Politics 85*(3), 1145–1150.

Hollenbach, F., N. W. Metternich, S. Minhas, and M. D. Ward (2014). Fast & easy imputation of missing social science data. *arXiv preprint arXiv:1411.0647*.

Honaker, J., A. Joseph, G. King, K. Scheve, and N. Singh (1999). Amelia: A program for missing data. *Department of Government, Harvard University*.

Honaker, J., G. King, and M. Blackwell (2011). Amelia ii: A program for missing data. *Journal of statistical software 45*, 1–47.

Huang, M. (2024). Overlap violations in external validity. *arXiv preprint arXiv::2403.19504*.

Huang, M. and C. McCartan (2025). Relative bias under imperfect identification in observational causal inference. *arXiv preprint arXiv:2507.23743*.

Huang, M. and S. D. Pimentel (2022). Variance-based sensitivity analysis for weighting estimators result in more informative bounds. *arXiv preprint arXiv:2208.01691*.

King, G., J. Honaker, A. Joseph, and K. Scheve (1998). List-wise deletion is evil: what to do about missing data in political science. In *Annual Meeting of the American Political Science Association, Boston*, Volume 52.

King, G., J. Honaker, A. Joseph, and K. Scheve (2001). Analyzing incomplete political science data: An alternative algorithm for multiple imputation. *American political science review 95*(1), 49–69.

Kuroki, M. and J. Pearl (2014). Measurement bias and effect restoration in causal inference. *Biometrika 101*(2), 423–437.

Lall, R. (2016). How multiple imputation makes a difference. *Political Analysis 24*(4), 414–433.

Lunceford, J. K. and M. Davidian (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in medicine 23*(19), 2937–2960.

Mayer, I., J. Josse, and T. Group (2023). Generalizing treatment effects with incomplete covariates: Identifying assumptions and multiple imputation algorithms. *Biometrical Journal*, 2100294.

McCartan, C., R. Fisher, J. Goldin, D. E. Ho, and K. Imai (2024). Estimating racial disparities when race is not observed. Technical report, National Bureau of Economic Research.

Miao, W., L. Liu, Y. Li, E. J. Tchetgen Tchetgen, and Z. Geng (2024). Identification and semiparametric efficiency theory of nonignorable missing data with a shadow variable. *ACM/JMS Journal of Data Science 1*(2), 1–23.

Miao, W. and E. J. Tchetgen Tchetgen (2016). On varieties of doubly robust estimators under missingness not at random with a shadow variable. *Biometrika 103*(2), 475–482.

Prior, M. (2013). Media and political polarization. *Annual review of political science 16*(1), 101–127.

Rosenbaum, P. R. and D. B. Rubin (1983a). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological) 45*(2), 212–218.

Rosenbaum, P. R. and D. B. Rubin (1983b). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Rosenbaum, P. R. and D. B. Rubin (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association 79*(387), 516–524.

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*(3), 581–592.

Sinha, A., H. Namkoong, R. Volpi, and J. Duchi (2017). Certifying some distributional robustness with principled adversarial training. *arXiv preprint arXiv:1710.10571*.

Sjölander, A. and S. Hägg (2025). Testable implications of outcome-independent missingness not at random in covariates. *Biometrika 112*(2), asaf009.

Sun, Z. and L. Liu (2021). Semiparametric inference of causal effect with nonignorable missing confounders. *Statistica Sinica 31*(4), 1669–1688.

Tibshirani, J., S. Athey, R. Friedberg, V. Hadad, D. Hirshberg, L. Miner, E. Sverdrup, S. Wager, M. Wright, and M. J. Tibshirani (2018). Package 'grf'. *Comprehensive R Archive Network*.

Tyler, M., J. Grimmer, and S. Iyengar (2022). Partisan enclaves and information bazaars: Mapping selective exposure to online news. *The Journal of Politics 84*(2), 1057–1073.

Van Buuren, S. and K. Groothuis-Oudshoorn (2011). mice: Multivariate imputation by chained equations in r. *Journal of statistical software 45*, 1–67.

Westreich, D., J. K. Edwards, S. R. Cole, R. W. Platt, S. L. Mumford, and E. F. Schisterman (2015). Imputation approaches for potential outcomes in causal inference. *International journal of epidemiology 44*(5), 1731–1737.

Yan, T. (2021). Consequences of asking sensitive questions in surveys. *Annual Review of Statistics and Its Application 8*(1), 109–127.

Yang, S., L. Wang, and P. Ding (2019). Causal inference with confounders missing not at random. *Biometrika 106*(4), 875–888.

Zhao, A. and P. Ding (2022). To adjust or not to adjust? estimating the average treatment effect in randomized experiments with missing covariates. *Journal of the American Statistical Association*, 1–11.

Zhao, A., P. Ding, and F. Li (2023). Covariate adjustment in randomized experiments with missing outcomes and covariates. *arXiv preprint arXiv:2311.10877*.

Zhao, A., P. Ding, and F. Li (2024). Covariate adjustment in randomized experiments with missing outcomes and covariates. *Biometrika*, asae017.

Zhao, Q., D. S. Small, and B. B. Bhattacharya (2019). Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 81*(4), 735–761.

A Extensions

A.1 Imputation with Weighted Estimators

Example A.1 (Weighting) *Consider the same bivariate setting as Example 3.1. Then, define the oracle and imputed propensity scores as follows:*

$$e_{oracle}(X_i) = \Pr(Z_i = 1 \mid X_1, X_2), \qquad e_{impute}(\tilde{X}_i) = \Pr(Z_i = 1 \mid X_1, \tilde{X}_2, R_i).$$

Then, the propensity scores estimated with the imputed data (i.e., $e_{impute}(\tilde{X}_i)$) can be written as a function of e_{oracle}, as well as a multiplicative error:

$$\begin{aligned} e_{impute}(X_i) &= \Pr(Z_i = 1 \mid X_1, \tilde{X}_2, R_i) \\ &= \frac{\Pr(R_i \mid X_1, \tilde{X}_2, Z_i = 1) \cdot \Pr(Z_i = 1 \mid X_1, \tilde{X}_2)}{\Pr(R_i \mid X_1, \tilde{X}_2)} \\ &= \underbrace{\Pr(Z_i = 1 \mid X_1, X_2)}_{e_{oracle}(X_i)} \cdot \frac{\Pr(R_i \mid X_1, \tilde{X}_2, Z_i = 1)}{\Pr(R_i \mid X_1, \tilde{X}_2)} \cdot \frac{\Pr(Z_i = 1 \mid X_1, \tilde{X}_2)}{\Pr(Z_i = 1 \mid X_1, X_2)}. \end{aligned}$$

In other words, without invoking further assumptions, $e_{impute}(X_i) \neq e_{oracle}(X_i)$.

If missing-at-random holds, then we expect $\Pr(R_i \mid X_1, \tilde{X}_2, Z_i = 1) = \Pr(R_i \mid X_1, \tilde{X}_2) = \Pr(R_i = \mathbb{1}_p \mid X_1)$. However, in order for $e_{impute}(X_i)$ to be equal to $e_{oracle}(X_i)$, then $\Pr(Z_i = 1 \mid X_1, \tilde{X}_2) = \Pr(Z_i = 1 \mid X_1, X_2)$. This will only occur if X_2 is independent of Z_i, conditional on the fully observed X_1. In other words, X_2 is not a confounder. We provide two examples of when this could occur. First, if X_2 is erroneous (i.e., $Z_i \perp\!\!\!\perp X_2$). Second, consider the setting in which X_2 can be perfectly explained by X_1. Paradoxically, when considering imputation, we are trying to impute the missing values as well as possible with the observed covariates. What this example highlights is that in settings when we can successfully impute missing values well, we may not need to impute at all.

A.2 Bias under Missing-Not-at-Random

In practice, we may not believe that missing-at-random holds–i.e., the observed covariates may be insufficient to account for the missingess pattern R_i. However, in the following section, we demonstrate that even in the presence of MNAR, we expect the projection estimator to have less bias in recovering the ATE than the existing methods. Furthermore, we propose a partial identification approach that allows researchers to bound the range of possible ATE estimates under violations of MAR.

To begin, the following corollary formalizes the bias of a projection estimator under MNAR.

Corollary A.1 (Bias of Projection Estimator under Missing-Not-at-Random) *When the missing confounders are missing-not-at-random, the bias of the projection estimator can be written as follows:*

$$\begin{aligned} \mathbb{E}(\hat{\tau}_{proj}) - \tau &= \Pr(R_i \neq \mathbb{1}_p) \left(\int \hat{\tau}(\tilde{X}_i) f(\tilde{X}_i \mid R_i \neq \mathbb{1}_p) d\nu(\tilde{X}_i) - \int \hat{\tau}(X_i) f(X_i \mid R_i \neq \mathbb{1}_p) d\nu(X_i) \right) \\ &= \Pr(R_i \neq \mathbb{1}_p) \Big\{ \mathbb{E}\big[\hat{\tau}(\tilde{X}_i) \mid R_i \neq \mathbb{1}_p\big] - \mathbb{E}\big[\hat{\tau}(X_i) \mid R_i \neq \mathbb{1}_p\big] \Big\}. \end{aligned}$$

Informally, in settings when the missing covariates are not very important for treatment effect heterogeneity, then bias will be minimized. Alternatively, if we are able to recover the missing values well with the observed covariates, then the bias will also be minimized.

We can compare the bias of the projection estimator under missing-not-at-random with the bias of the alternative estimators.

Theorem A.1 (Relative Reduction in Bias under Missing-Not-at-Random) *Under outcome-response ignorability, the bias from the projection estimator will be less than, or equal to, the bias from an imputation estimator:*

$$Bias(\hat{\tau}_{proj}) \leq Bias(\hat{\tau}_{impute}),$$

even under missing-not-at-random. Furthermore, if the following holds:

$$\int \Big(\hat{\tau}_{cc}(\tilde{X}_i) f(\tilde{X}_i \mid R_i \neq \mathbb{1}_p) - \hat{\tau}_{cc}(X_i) f(X_i \mid R_i \neq \mathbb{1}_p)\Big)\, d\nu(X_i)$$
$$\leq \int \left(\hat{\tau}_{cc}(X_i) f(X_i \mid R_i = \mathbb{1}_p) - \hat{\tau}_{cc}(X_i) f(X_i \mid R_i \neq \mathbb{1}_p)\right) d\nu(X_i),$$

then the bias from using the projection estimator will be less than the complete case estimator (i.e., $Bias(\hat{\tau}_{proj}) \leq Bias(\hat{\tau}_{cc})$).

Intuitively, we expect that the bias from $\hat{\tau}_{proj}$ will be less than the bias from the complete case estimator, because $f(\tilde{X}_i \mid R_i = 0)$ is at least an approximation of the missing values. In contrast, the complete case bias is driven by how close the distribution of the complete cases $f(X_i \mid R_i = \mathbb{1}_p)$ are to the missing (i.e., $f(X_i \mid R_i \neq \mathbb{1}_p)$). Since $f(\tilde{X}_i \mid R_i \neq \mathbb{1}_p)$ leverages the *observed* data (i.e., X_i^{obs}) across $R_i \neq \mathbb{1}_p$), we expect that as long if the imputation model can account for some variation in the missing covariates X_i^{mis}, there will be an improvement in bias over the complete case estimator.

Practical Considerations. In practice, researchers may have a covariate with substantially more missingness than the other covariates. In such a setting, if using a complete case estimator, to minimize precision loss, researchers often omit this covariate entirely from estimation. To assess the impact of omitting such a covariate, researchers may leverage tools from the omitted variable bias literature.

A.3 Relationship with Yang et al. (2019)

Recall from Section 4, while outcome-response ignorability directly allows for the identification of the CATE (i.e., $\tau(X_i)$), the distribution of the covariates $f(X_i)$ is not still not directly observable. In the following subsection, we walk through the non-parametric identification result from Yang et al. (2019), as well as the approximation approach they propose.

To begin, note that we can write the joint density of $f(X_i, Z_i, Y_i)$ as a function of the observed distribution:

$$f(X_i, Z_i, Y_i) = \frac{f(X_i, Z_i, Y_i, R_i = \mathbb{1}_p)}{f(R_i = \mathbb{1}_p \mid X_i, Z_i, Y_i)}.$$

$f(X_i, Z_i, Y_i, R_i = \mathbb{1}_p)$ is observable, as it relies on only the complete cases $R_i = \mathbb{1}_p$. As such, recovering $f(R_i = \mathbb{1}_p \mid X_i, Z_i, Y_i)$ allows us to recover $f(X_i, Y_i, Z_i)$. To start, define $\varphi_{rz}(X_i)$ as:

$$\varphi_{rz}(X_i) = \frac{\Pr(R_i = r \mid Z_i = z, X_i, Y_i)}{\Pr(R_i = \mathbb{1}_p \mid Z_i = z, X_i, Y_i)} = \frac{\Pr(R_i = r \mid Z_i = z, X_i)}{\Pr(R_i = \mathbb{1}_p \mid Z_i = z, X_i)},$$

where

$$\Pr(R_i = \mathbb{1}_p \mid Z_i = z, X_i) = \frac{1}{1 + \sum_{r' \neq \mathbb{1}_p} \varphi_{r'z}(X)},$$

and $r \in \mathcal{R}$, which is the possible missingness patterns present in the covariates X. $\varphi_{rz}(X)$ then allows us to identify $\Pr(R_i = \mathbb{1}_p \mid X_i, Z_i, Y_i)$. Let X_r represent the observed part of X. Then, following Theorem 1 from Yang et al. (2019), identification follows from solving the following integral equation:

$$f(Z = z, X_r, Y, R' = r) = \int \varphi_{rz}(X) f(Z = z, X, Y, R' = \mathbb{1}_p) d\nu(X_r). \tag{6}$$

Equation (6) leverages the different missingness patterns in the data to identify $\varphi_{rz}(X)$. While Equation (6) allows for the identification of $\varphi_{rz}(X)$, and by extension, the ATE, *solving* the integral equation in practice is often infeasible. In particular, Yang et al. (2019) show that while the population-level estimating equation has a unique solution, for a given sample, there is no guarantee that there will be a unique solution. Furthermore, directly solving Equation (6) with the consistent estimators of the densities$f(Z = z, X_r, Y, R' = r)$ and $f(Z = z, X, Y, R' = \mathbb{1}_p)$ does not necessarily yield a consistent estimator of $\varphi_{rz}(X)$ (Yang et al., 2019).

Approximating the solution. The estimator proposed in Yang et al. (2019) works by first estimating the CATE via standard outcome-based modeling approaches, and then estimating the probability that a covariate value is observed, $\pi_z(x) = \Pr(R = 1 \mid Z = z, X = x)$. The authors start by estimating how the distribution of the outcome differs between units with missing covariates ($R \neq \mathbb{1}_p$) and units with observed covariates ($R = \mathbb{1}_p$). This difference is summarized by a density ratio, which compares the two outcome distributions $r_z(y) = \frac{f_{Y|Z=z,R=0}(y)}{f_{Y|Z=z,R=1}(y)}$, which measures how the outcome distribution differs between units with missing and observed covariates. The key idea is that this ratio can be approximated by regressing $r_z(y)$ onto a set of conditional moment functions

$$H(y) \;=\; \mathbb{E}[\phi(X) \mid Y = y, Z = z, R = 1],$$

where $\phi(X)$ is a chosen basis (e.g., polynomials or splines). From this regression, they obtain a sieve approximation

$$\xi_z(x) \approx \phi(x)^\top \beta, \qquad \pi_z(x) = \frac{1}{1 + \xi_z(x)}.$$

In practice, even solving the approximated solution by estimating these objects is difficult. Each component—the outcome density estimates, the conditional-moment regressions for $H(y)$, the projection of r_z onto H, and the regularization used to stabilize the problem—introduces sampling error. These errors also interact: for example, changing the flexibility of the spline basis for $H(y)$ changes the $L^2(f_1)$ projection of r_z, so improving one step can worsen another. Because the procedure is not doubly robust, small misspecifications in any part can propagate directly into the final estimate $\widehat{\tau}$.

The challenges are amplified in high-dimensional, continuous settings. Representing functions of p covariates with degree-D tensor-product bases requires

$$K = \sum_{d=0}^{D} \binom{p+d-1}{d} = \binom{p+D}{D},$$

which grows combinatorially. Large K leads to unstable regressions, strong shrinkage, and high computational cost. Second, estimating the conditional-moment functions $H(y)$ well typically requires flexible smoothing and many interaction terms, which increases variance and effectively forces the use of cross-fitting to avoid overfitting. Third, estimating the ratio of two outcome densities is sensitive to the tails: if the distributions for $R = \mathbb{1}_p$ and $R \neq \mathbb{1}_p$ do not have sufficient overlap, the ratio becomes unstable and needs to be truncated or re-normalized, which in turn introduces additional bias.

For these reasons, even with careful tuning—orthonormal bases, trimmed grids, self-normalized ratios, and cross-fitting—the estimator can remain statistically unstable and computationally expensive when p is moderate or large. In contrast to the Yang et al. (2019) approach, the two-stage projection estimator relies on an imputation model to approximate the density $f(X_i)$. The weighted estimator considered in the simulations is a parametric alternative to the sieve-based approximations used in Yang et al. (2019), and is discussed in the following subsection.

A.4 Alternative Estimation Approaches

We provide an overview of alternative estimation approaches that similarly leverage outcome-response ignorability. In particular, we focus on two additional approaches. The first is weighting, where researchers model the density ratio between the fully observed and incomplete cases (i.e., $f(X_i)/f(X_i \mid Z_i = z, R_i = \mathbb{1}_p)$).

A.4.1 Weighting

To begin, we introduce a weighting estimator. Bias from using only the complete cases arises from distributional differences in the covariates across the fully observed and incomplete cases, as represented by the density ratio $f(X_i)/f(X_i \mid Z_i = z, R_i = \mathbb{1}_p)$. The weighting estimator thus constructs weights that directly target the density ratio.

To begin, we re-write the density ratio as the inverse conditional probability of being fully observed and being assigned to treatment $Z_i = z$:

$$\frac{f(X_i)}{f(X_i \mid Z_i = z, R_i = \mathbb{1}_p)} = \frac{\Pr(R_i = \mathbb{1}_p, Z_i = z)}{\Pr(R_i = \mathbb{1}_p, Z_i = z \mid X_i)}.$$

Then, define the weights $w_z(X_i)$ for $z \in \{0, 1\}$ as:

$$w_z(X_i) = \frac{1}{\Pr(R_i = \mathbb{1}_p, Z_i = z \mid X_i)}.$$

Because the joint probability depends on the full set of X_i, we leverage a decomposition, first introduced in Sun and Liu (2021), which allows us to re-write $\Pr(R_i = \mathbb{1}_p, Z_i = z \mid X_i)$ as a function of (1) the propensity of receiving treatment assignment $Z_i = z$ across the complete cases (i.e., $e(X_i; R_i = \mathbb{1}_p) := \Pr(Z_i = 1 \mid X_i, R_i = \mathbb{1}_p)$), and (2) the probability of being a fully observed case (i.e., $\pi_r(X_i, Z_i) := \Pr(R_i = \mathbb{1}_p \mid Z_i = z, X_i)$).

Usefully, $e(X_i; R_i = \mathbb{1}_p)$ depends only on the complete cases. As such, researchers can employ standard propensity score approaches across the complete cases to estimate probability of receiving treatment. To estimate $\pi_r(X_i, Z_i)$, researchers can solve a set of estimating equations that balance the probability of being a fully observed case across the treatment assignments and outcomes. Helpfully, the estimating equations only require having covariate data across the fully observed cases $R_i = \mathbb{1}_p$.

Following Sun and Liu (2021), we represent the specified model with $\phi(Z_i, X_i; \gamma)$, where γ represents the set of parameters estimated for some model $\phi(\cdot)$. Then, with the specified model, we solve the following estimation equation:

$$\mathbb{E}\left[\left(\frac{R_i}{\phi(Z_i, X_i; \gamma)} - 1\right) h(Z_i, Y_i)\right] = 0,$$

where $h(Z_i, Y_i)$ is a differentiable, vector function of Z_i and Y_i.

As a concrete example, consider the setting where researchers assume the probability of being a complete case follows a logistic function. Then:

$$\phi(Z_i, X_i; \gamma) = \frac{1}{1 + \exp(\gamma_\alpha + \gamma_z Z + \gamma_x^\top X_i)},$$

where $\gamma = (\gamma_\alpha, \gamma_z, \gamma_x) \in \mathbb{R}^{p+2}$, and $h(Z_i, Y_i) = (1, Z_i, Y_i)^\top$. The estimating equations can then be written as:

$$\begin{cases} \mathbb{E}\left[R_i \cdot \left\{1 + \exp(\gamma_\alpha + \gamma_z Z + \gamma_x^\top X_i)\right\}\right] = 1 \\ \mathbb{E}\left[R_i Z_i \cdot \left\{1 + \exp(\gamma_\alpha + \gamma_z Z + \gamma_x^\top X_i)\right\}\right] = \mathbb{E}[Z_i] \\ \mathbb{E}\left[R_i Y_i \left\{1 + \exp(\gamma_\alpha + \gamma_z Z + \gamma_x^\top X_i)\right\}\right] = \mathbb{E}[Y_i] \end{cases}$$

With estimates $\hat{e}(X_i; R_i = \mathbb{1}_p)$ and $\hat{\pi}_r(X_i, Z_i)$, researchers can then estimate $\{\hat{w}_1(X_i), \hat{w}_0(X_i)\}$. The weights allow researchers to account for the distributional differences in the covariates X_i, across the fully observed cases and the missing cases:

$$\hat{\tau}_w := \frac{1}{\sum_{i=1}^n Z_i R_i} \sum_{i=1}^n Z_i R_i Y_i \hat{w}_1(X_i) - \frac{1}{\sum_{i=1}^n (1 - Z_i) R_i} \sum_{i=1}^n (1 - Z_i) R_i Y_i \hat{w}_0(X_i).$$

Then, assuming $\hat{e}(X_i; R_i = \mathbb{1}_p)$ and $\hat{\pi}_r(X_i, Z_i)$ are consistent estimates of the true probabilities, the weighted estimator will be a consistent estimator for the ATE.

Theorem A.2 (Consistency of the Weighted Estimator) *Assume Assumptions 1-3, and 7 (outcome-response ignorability) hold. Furthermore, assume the following estimation assumptions:*

- *Consistent propensity score model:* $\hat{e}(X_i; R_i = \mathbb{1}_p) \xrightarrow{p} \Pr(Z_i = 1 \mid R_i = \mathbb{1}_p, X_i)$
- *Consistent complete case probability model:* $\hat{\pi}_r(X_i, Z_i) \xrightarrow{p} \Pr(R_i = \mathbb{1}_p \mid Z_i, X_i)$

Then, $\hat{\tau}_w$ will be a consistent estimator for the ATE:

$$\hat{\tau}_w \xrightarrow{p} \tau.$$

To perform inference, researchers can employ a standard sandwich estimator, which will provide conservative estimates of the underlying uncertainty in $\hat{\tau}_w$.

A.4.2 Details on Augmented Weighted Estimator

An alternative approach to estimation is to use an augmented weighted estimator:

$$\begin{aligned}\hat{\tau}_{aug} =& \frac{1}{\sum_{i=1}^n Z_i R_i} \sum_{i=1}^n Z_i R_i \hat{w}_1(X_i)\left(Y_i - \hat{m}_1(X_i;\mathcal{S}_1)\right) - \frac{1}{\sum_{i=1}^n (1-Z_i) R_i} \sum_{i=1}^n (1-Z_i) R_i \left(Y_i - \hat{m}_0(\tilde{X}_i;\mathcal{S}_1)\right)\hat{w}_0(X_i) \\ &+ \frac{1}{n}\sum_{i=1}^n \left\{\hat{m}_1(\tilde{X}_i;\mathcal{S}_1) - \hat{m}_0(\tilde{X}_i;\mathcal{S}_1)\right\},\end{aligned}$$

where $\hat{m}_1, \hat{m}_0, \hat{w}_1, \hat{w}_0$ correspond to the estimated outcome models and estimated weights, respectively. The augmented weighted estimator allows researchers to combine both the weighting estimator with the outcome model estimated in the two-stage projection approach. We establish the consistency of the augmented weighted estimator.

Theorem A.3 (Consistency of the Augmented Weighted Estimator) *Assume Assumptions 1-3, and 7 (outcome-response ignorability) hold. Then, with a valid imputation model, if either the propensity score model and complete case probability model or if the outcome models are consistently estimated, the augmented weighted estimator will be a consistent estimator for the ATE:*

$$\hat{\tau}_{aug} \xrightarrow{p} \tau.$$

Notably, the augmented weighted estimator will have properties of doubly robustness. This means that the weights can be misspecified, *or* the outcome models can be misspecified, and the augmented weighted estimator will still consistently recover the ATE. However, while the weights or the outcome model can be misspecified, the doubly robustness does not extend towards the assumption that the imputation model can consistently recover the density of the missing covariates. This is analogous to the property highlighted in Sun and Liu (2021), who show that doubly robustness depends crucially that a density ratio can be parametrically modeled correctly.

A.5 Details on the Sensitivity Analysis

To begin, we introduce the following assumption, which constrains the amount of error that can occur when recovering the missing covariate values.

Assumption 8 (Constrained Imputation Error) *For $X_i^{(j)} \in \mathcal{X}_{\rangle mis}$, where $j \in \{1, ..., |\mathcal{X}_{\rangle mis}|\}$,*

$$||\hat{X}_i^{(j)} - X_i^{(j)}||_p \leq L_j$$

The p-norm chosen corresponds to the error metric we are constraining. For example, if $p = 1$, this constrains the average absolute error of the imputation model. If we constrain $p = 2$, this constrains the root mean squared error. The assumption constrains the error for each individual covariate within the set of missing covariates.

Definition A.1 (Imputation Uncertainty Set)

$$\Theta(L;p) = \bigcup_{j=1}^{|X_i^{mis}|} \left\{X_i^{mis(j)} \mid ||\hat{X}_i^{mis(j)} - X_i^{mis(j)}||_p \leq L_j\right\}$$

$\Theta(L;p)$ represents the set of all possible X_i^{mis} values that exist, given a constraint on the imputation error. Each covariate j can have its own constraint. In practice, researchers can choose to set a single constraint across all covariates, which would serve as a global constraint on the entire set of missing covariates. However, this can be overly conservative in settings when researchers are able to impute many covariates well, but do a poor job imputing one covariate. For ease of notation, we will denote the setting in which researchers are choosing an L_∞ norm error on the imputation error constraint as $\Theta(L)$, suppressing the p.

We also allow for potential violations in outcome-response ignorability by evaluating the ratio between the CATE for the incomplete cases (i.e., $R_i \neq \mathbb{1}_p$) and the complete cases.

Assumption 9 (Violation in Outcome-Response Ignorability) *For all $x \in \mathcal{X}$:*

$$\Gamma(x) := \frac{\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, R_i \neq \mathbb{1}_p]}{\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, R_i = \mathbb{1}_p]} \leq \Gamma.$$

Then, for a fixed set of covariates, we can define the CATE uncertainty set as $\varepsilon(\Gamma, X_i)$.

Definition A.2 (CATE Uncertainty Set)

$$\varepsilon(\Gamma, X_i) = \left\{ \tau(X_i) : \frac{\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, R_i \neq \mathbb{1}_p]}{\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i = x, R_i = \mathbb{1}_p]} \leq \Gamma \right\}$$

Then, accounting for both imputation error and violations in outcome-response ignorability, the CATE uncertainty set is written as $\varepsilon(\Gamma; \Theta(L))$. As a result, for a fixed $\varepsilon(\Gamma; \Theta(L))$, the partially identified region is defined as follows:

$$\tau \in \left[\inf_{\varepsilon(\Gamma;\Theta(L))} \tau(\tilde{X}_i), \quad \sup_{\varepsilon(\Gamma;\Theta(L))} \tau(\tilde{X}_i) \right]$$

To estimate the range of possible values, that the ATE can take on, given the imputation uncertainty set, we must solve the following optimization problem:

$$\begin{aligned} \min_{\varepsilon(\Gamma;\Theta(L))} / \max \ & \frac{1}{n} \sum_{i=1}^{n} \hat{\tau}(\tilde{X}_i) \cdot \gamma_i, \\ \text{s.t. } & ||\tilde{X}_i^{(j)} - \hat{X}_i^{(j)}||_p \leq L_j \text{ for all } j, \\ & \tilde{X}_i^{(j)} - \hat{X}_i^{(j)} = 0 \text{ if } R_i^{(j)} = 1 \\ & \gamma_i \leq \Gamma \text{ for all } R_i \neq \mathbb{1}_p \\ & \gamma_i = 1 \text{ for all } R_i = \mathbb{1}_p \end{aligned} \tag{7}$$

Given mild regularity conditions on the estimated treatment effect heterogeneity model, we can employ existing optimization methods to efficiently estimate the range of possible estimates (Boyd et al., 2004; Sinha et al., 2017). By solving Problem (7), we are generating the set of adversarial, missing X_i that would result in the largest (and smallest) possible ATE estimates. As such, we expect the bounds to be sharp by construction.

There are connections to the distributionally robust optimization literature. We can generalize the proposed partial identification approach to account for uncertainty sets constructed from constraining distributional divergences between the imputed $\tilde{X}_i$ and true X_i (i.e., ϕ-divergences, MMD, etc.) In particular, Equation (7), which constrains p-norm balls around $\tilde{X}_i$ is a special case of constrained optimization, constraining the Wasserstein distance between the distribution of the imputed $\tilde{X}_i$ and true X_i. However, we focus on the p-norm constraints for several reasons. First, the constraints are relatively straightforward to interpret. Second, Wasserstein-style constraints allow for covariate shift outside of the support of the observed data. This is important for our specific setting, because we are worried that the support of the observed covariates may not necessarily contain the full support of the covariates.

B Proofs

Following Yang et al. (2019), Assumption 7 allows us to identify the conditional average treatment effect. We provide the formal lemma for completeness.

Lemma B.1 (Nonparametric Identification of $\tau(X_i)$ Across Observed Data)
Under Assumptions 1-2 and 7, $\tau(X_i) := \mathbb{E}(Y_i(1) - Y_i(0) \mid X_i = x)$ can be identified:

$$\begin{aligned}\tau_{cc}(X) &= \mathbb{E}(Y \mid Z = 1, X = x, R = \mathbb{1}_p) - \mathbb{E}(Y \mid Z = 0, X = x, R = \mathbb{1}_p)\\ &= \mathbb{E}(Y(1) \mid Z = 1, X = x, R = \mathbb{1}_p) - \mathbb{E}(Y(0) \mid Z = 0, X = x, R = \mathbb{1}_p)\end{aligned}$$

By Assumption 7 (Outcome-Response Ignorability):

$$= \mathbb{E}(Y(1) \mid Z = 1, X = x) - \mathbb{E}(Y(0) \mid Z = 0, X = x)$$

By Assumption 1 (Conditional Ignorability of Treatment Assignment):

$$\equiv \tau(X)$$

B.1 Nonparametric Identification under Imputation

$$\begin{aligned}&\mathbb{E}(Y(z) \mid Z = z, X_1, \tilde{X}_2)\\ =&\mathbb{E}(Y(z) \mid Z = z, X_1, \tilde{X}_2, R_i = \mathbb{1}_p) \Pr(R = \mathbb{1}_p \mid Z = z, X_1, \tilde{X}_2) +\\ &\mathbb{E}(Y(z) \mid Z = z, X_1, \tilde{X}_2, R \neq \mathbb{1}_p) \Pr(R \neq \mathbb{1}_p \mid Z = z, X_1, \tilde{X}_2)\\ =&\mathbb{E}(Y(z) \mid Z = z, X_1, X_2, R = \mathbb{1}_p) \Pr(R = \mathbb{1}_p \mid Z = z, X_1, \tilde{X}_2) +\\ &\mathbb{E}(Y(z) \mid Z = z, X_1, g(X_1), R \neq \mathbb{1}_p) \Pr(R \neq \mathbb{1}_p \mid Z = z, X_1, \tilde{X}_2)\\ =&\underbrace{\mathbb{E}(Y(z) \mid Z_i = z, X_1, X_2, R = \mathbb{1}_p)}_{=\mathbb{E}(Y(z) \mid X_1, X_2, R = \mathbb{1}_p)} \Pr(R = \mathbb{1}_p \mid Z = z, X_1, \tilde{X}_2) +\\ &\mathbb{E}(Y(z) \mid Z = z, X_1, R_i = 0) \Pr(R \neq \mathbb{1}_p \mid Z = z, X_1, \tilde{X}_2)\end{aligned}$$

If $Y(z) \perp\!\!\!\perp Z \mid X_1, R \neq \mathbb{1}_p$ and $Y(z) \perp\!\!\!\perp Z \mid X_1, X_2, R = \mathbb{1}_p$, then:

$$\begin{aligned}=&\mathbb{E}(Y(z) \mid X_1, X_2, R_i = \mathbb{1}_p) \Pr(R = \mathbb{1}_p \mid Z = z, X_1, \tilde{X}_2) +\\ &\mathbb{E}(Y(z) \mid X_1, R \neq \mathbb{1}_p) \Pr(R \neq \mathbb{1}_p \mid Z = z, X_1, \tilde{X}_2)\\ =&\mathbb{E}(Y(z) \mid X_1, \tilde{X}_2)\end{aligned}$$

As such, under the modified selection on observables assumption, we can identify the conditional ATE from imputation:

$$\mathbb{E}(Y \mid Z = 1, X_1, \tilde{X}_2) - \mathbb{E}(Y \mid Z = 0, X_1, \tilde{X}_2)$$

$$
\begin{aligned}
&= \mathbb{E}(Y(1) \mid Z = 1, X_1, \tilde{X}_2) - \mathbb{E}(Y(0) \mid Z = 0, X_1, \tilde{X}_2) \\
&= \mathbb{E}(Y(1) - Y(0) \mid X_1, \tilde{X}_2)
\end{aligned}
$$

B.2 Proof of Theorem 4.1

We will show that $\hat{\tau}_{proj}$ is a consistent estimator for the ATE. To begin, we apply law of large numbers, such that $\hat{\tau}_{proj} \xrightarrow{p} \mathbb{E}(\hat{\tau}_{proj})$. Then:

$$
\begin{aligned}
\mathbb{E}(\hat{\tau}_{proj}) &= \mathbb{E}\Bigg(\underbrace{\frac{1}{n}\sum_{i=1}^{n} \{\hat{m}_1(X_i;\mathcal{S}_1) - \hat{m}_0(X_i;\mathcal{S}_1)\} R_i}_{\equiv(1)\text{ Complete Case Estimator}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} \left\{\hat{m}_1(\tilde{X}_i;\mathcal{S}_1) - \hat{m}_0(\tilde{X}_i;\mathcal{S}_1)\right\}(1 - R_i)}_{(2)\text{ Projected Component}} \Bigg) \\
&= \mathbb{E}\left(\{\hat{m}_1(X_i;\mathcal{S}_1) - \hat{m}_0(X_i;\mathcal{S}_1)\} R_i\right) + \mathbb{E}\left(\left\{\hat{m}_1(\tilde{X}_i;\mathcal{S}_1) - \hat{m}_0(\tilde{X}_i;\mathcal{S}_1)\right\}(1 - R_i)\right) \\
&= \underbrace{\mathbb{E}\left(\{m_1(X_i;\mathcal{S}_1) - m_0(X_i;\mathcal{S}_1)\} R_i\right)}_{(a)} + \underbrace{\mathbb{E}\left(\left\{m_1(\tilde{X}_i;\mathcal{S}_1) - m_0(\tilde{X}_i;\mathcal{S}_1)\right\}(1 - R_i)\right)}_{(b)} \\
&\qquad + \mathbb{E}\left\{\hat{m}_1(X_i;\mathcal{S}_1) - m_1(X_i;\mathcal{S}_1)\right\} - \mathbb{E}\left\{\hat{m}_0(X_i;\mathcal{S}_1) - m_0(X_i;\mathcal{S}_1)\right\}
\end{aligned} \tag{8}
$$

For ease of notation , we will define $\tau(X_i, \mathcal{S}_1) := m_1(X_i;\mathcal{S}_1) - m_0(X_i;\mathcal{S}_1)$. From Equation (8)-(a), it follows immediately from Assumptions 1-2, and Assumption 7:

$$
\mathbb{E}\left(\tau(X_i;\mathcal{S}_1)R_i\right) = \mathbb{E}(\tau(X_i;\mathcal{S}_1) \mid R_i = \mathbb{1}_p) \cdot \Pr(R_i = \mathbb{1}_p) = \mathbb{E}_X(\tau(X_i) \mid R_i = \mathbb{1}_p) \cdot \Pr(R_i = \mathbb{1}_p).
$$

Similarly, for Equation (8)-(b), we can write $\mathbb{E}(\tau(\tilde{X}_i;\mathcal{S}_1)\cdot(1-R_i)) = \mathbb{E}(\tau(\tilde{X}_i;\mathcal{S}_1) \mid R_i \neq \mathbb{1}_p)\cdot\Pr(R_i \neq \mathbb{1}_p)$. Then:

$$
\begin{aligned}
\mathbb{E}(\tau(\tilde{X}_i;\mathcal{S}_1) \mid R_i \neq \mathbb{1}_p) &= \int \tau(X_i) f(\tilde{X}_i \mid R_i \neq \mathbb{1}_p) \\
&= \int \tau(X_i) f(\{X_i^{obs}, \hat{X}_i^{mis}\} \mid R_i \neq \mathbb{1}_p)
\end{aligned}
$$

Under the assumption of a valid imputation model (i.e., $f(\{X_i^{obs}, \hat{X}_i^{mis}\} \mid R_i \neq \mathbb{1}_p) = f(\{X_i^{obs}, X_i^{mis}\} \mid R_i \neq \mathbb{1}_p)$):

$$
\begin{aligned}
&= \int \tau(X_i) f(\{X_i^{obs}, X_i^{mis}\} \mid R_i \neq \mathbb{1}_p) \\
&= \mathbb{E}_X[\tau(X_i) \mid R_i \neq \mathbb{1}_p]
\end{aligned}
$$

Notably, the assumption of a valid imputation model will hold under MAR:

$$
f(\{X_i^{obs}, \hat{X}_i^{mis}\} \mid R_i \neq \mathbb{1}_p) = f(R_i \neq \mathbb{1}_p \mid X_i^{obs}, \hat{X}_i^{mis}) f(\hat{X}_i^{mis} \mid X_i^{obs}) f(X_i^{obs}) \cdot \frac{1}{f(R_i \neq \mathbb{1}_p)}
$$

Under MAR ($X_i^{mis} \perp\!\!\!\perp R_i \mid X_i^{obs}$):

$$
= \frac{f(R_i \neq \mathbb{1}_p \mid X_i^{obs}, X_i^{mis}) f(X_i^{mis} \mid X_i^{obs}, R_i = \mathbb{1}_p) f(X_i^{obs})}{f(R_i \neq \mathbb{1}_p)}
$$

$$
\begin{aligned}
&= \frac{f(X_i^{obs}, X_i^{mis} \mid R_i \neq \mathbb{1}_p) f(X_i^{mis} \mid X_i^{obs}, R_i = \mathbb{1}_p) f(X_i^{obs})}{f(X_i^{obs}, X_i^{mis})} \\
&= \frac{f(X_i^{obs}, X_i^{mis} \mid R_i \neq \mathbb{1}_p) f(X_i^{mis} \mid X_i^{obs}, R_i = \mathbb{1}_p)}{f(X_i^{mis} \mid X_i^{obs})} \\
&= \frac{f(X_i^{obs}, X_i^{mis} \mid R_i \neq \mathbb{1}_p) f(X_i^{mis} \mid X_i^{obs}, R_i = \mathbb{1}_p)}{f(X_i^{mis} \mid X_i^{obs}, R_i \neq \mathbb{1}_p)} \\
&= f(X_i^{obs}, X_i^{mis} \mid R_i \neq \mathbb{1}_p)
\end{aligned}
$$

Combining together, and under our assumption that $\mathbb{E}\left\{\hat{m}_z(X_i; \mathcal{S}_1) - m_z(X_i; \mathcal{S}_1)\right\} = o_p(1)$ for $z \in \{0, 1\}$:

$$
\mathbb{E}(\hat{\tau}_{proj}) = \mathbb{E}_X(\tau(X_i)) + o_p(1) \to \tau \text{ as } n \to \infty.
$$

As such, we have shown $\hat{\tau}_{proj} \xrightarrow{p} \tau$.

B.3 Proof of Theorem A.2

Proof: We begin by showing that with the oracle weights $w_1(X_i)$ and $w_0(X_i)$, the weighted estimator will provide unbiased estimates of the ATE.

$$
\begin{aligned}
&\mathbb{E}\left[Z_i R_i Y_i w_1(X_i)\right] \\
&\quad = \mathbb{E}_X\left[\mathbb{E}\left[Z_i R_i Y_i(1) w_1(X_i) \middle| X_i\right]\right] \\
&\quad = \mathbb{E}_X\left[\mathbb{E}\left[Y_i(1) w_1(X_i) \middle| X_i, Z_i = 1, R_i = 1\right] \Pr(Z_i = 1, R_i = 1 \mid X_i)\right] \\
&\quad = \mathbb{E}_X\left[\mathbb{E}(Y_i(1) \mid X_i, Z_i = 1, R_i = 1) \cdot \frac{\Pr(Z_i = 1, R_i = 1 \mid X_i)}{\Pr(Z_i = 1, R_i = 1 \mid X_i)}\right] \\
&\quad = \mathbb{E}_X\left[\mathbb{E}(Y_i(1) \mid X_i, Z_i = 1, R_i = 1)\right] \\
&\quad = \mathbb{E}_X\left[\mathbb{E}(Y_i(1) \mid X_i, Z_i = 1)\right] && \text{(Outcome-response ignorability)} \\
&\quad = \mathbb{E}_X\left[\mathbb{E}(Y_i(1) \mid X_i)\right] && \text{(Conditional ignorability)} \\
&\quad \equiv \mathbb{E}(Y_i(1))
\end{aligned}
$$

We can similarly show that $\mathbb{E}\left[(1 - Z_i) R_i Y_i w_0(X_i)\right] = \mathbb{E}(Y_i(0))$. Then, it follows immediately:

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{n_1} \sum_{i=1}^n Z_i R_i Y_i w_1(X_i) - \frac{1}{n_0} \sum_{i=1}^n (1 - Z_i) R_i Y_i w_0(X_i)\right] \\
&\quad = \mathbb{E}_X\left[\mathbb{E}(Y_i(1) - Y_i(0) \mid X_i)\right] \\
&\quad \equiv \tau
\end{aligned}
$$

Now,

$$
\hat{\tau}_w := \frac{1}{\sum_{i=1}^n Z_i R_i} \sum_{i=1}^n Z_i R_i Y_i \hat{w}_1(X_i) - \frac{1}{\sum_{i=1}^n (1 - Z_i) R_i} \sum_{i=1}^n (1 - Z_i) R_i Y_i \hat{w}_0(X_i).
$$

$$
\hat{\tau}_w = \frac{1}{\sum_{i=1}^n Z_i R_i} \sum_{i=1}^n Z_i R_i Y_i \hat{w}_1(X_i) - \frac{1}{\sum_{i=1}^n (1 - Z_i) R_i} \sum_{i=1}^n (1 - Z_i) R_i Y_i \hat{w}_0(X_i)
$$

$$
\begin{aligned}
=&\frac{1}{\sum_{i=1}^n Z_iR_i}\sum_{i=1}^n Z_iR_iY_iw_1(X_i) - \frac{1}{\sum_{i=1}^n(1-Z_i)R_i}\sum_{i=1}^n(1-Z_i)R_iY_iw_0(X_i)+\\
&\frac{1}{\sum_{i=1}^n Z_iR_i}\sum_{i=1}^n Z_iR_iY_i\left\{\hat{w}_1(X_i)-w_1(X_i)\right\} - \frac{1}{\sum_{i=1}^n(1-Z_i)R_i}\sum_{i=1}^n(1-Z_i)R_iY_i\left\{\hat{w}_0(X_i)-w_0(X_i)\right\}
\end{aligned}
$$

Since $\hat{w}_z(X_i) - w_z(X_i) = o_p(1)$ for $z \in \{0,1\}$, we can apply Weak Law of Large Numbers, which directly shows $\hat{\tau}_w \xrightarrow{p} \tau$. □

B.4 Proof of Theorem A.3

Proof: We will now show that $\hat{\tau}_{aug}$ is a consistent estimator for τ if ...

To start, we will show that under a valid imputation model, if $\hat{w}_z(X_i) - w_z(X_i) = o_p(1)$, then $\hat{\tau}_{aug} \xrightarrow{p} \tau$. We can re-write $\hat{\tau}_{aug}$ as follows:

$$
\begin{aligned}
\hat{\tau}_{aug} =&\frac{1}{\sum_{i=1}^n Z_iR_i}\sum_{i=1}^n Z_iR_i\hat{w}_1(X_i)\left(Y_i-\hat{m}_1(X_i;\mathcal{S}_1)\right) - \frac{1}{\sum_{i=1}^n(1-Z_i)R_i}\sum_{i=1}^n(1-Z_i)R_i\left(Y_i-\hat{m}_0(\tilde{X}_i;\mathcal{S}_1)\right)\hat{w}_0(X_i)\\
&+\frac{1}{n}\sum_{i=1}^n\left\{\hat{m}_1(\tilde{X}_i;\mathcal{S}_1)-\hat{m}_0(\tilde{X}_i;\mathcal{S}_1)\right\}\\
=&\frac{1}{\sum_{i=1}^n Z_iR_i}\sum_{i=1}^n Z_iR_iY_i\hat{w}_1(X_i) - \frac{1}{\sum_{i=1}^n(1-Z_i)R_i}\sum_{i=1}^n(1-Z_i)R_iY_i\hat{w}_0(X_i)\\
&+\frac{1}{n}\sum_{i=1}^n\left(\hat{m}_1(\tilde{X}_i;\mathcal{S}_1)-\hat{m}_0(\tilde{X}_i;\mathcal{S}_1)\right)\\
&-\left(\frac{1}{\sum_{i=1}^n Z_iR_i}\sum_{i=1}^n Z_iR_i\hat{m}_1(X_i;\mathcal{S}_1)\hat{w}_1(X_i) - \frac{1}{\sum_{i=1}^n(1-Z_i)R_i}\sum_{i=1}^n(1-Z_i)R_i\hat{m}_0(X_i;\mathcal{S}_1)\hat{w}_0(X_i)\right)\\
=&\,\hat{\tau}_w+\frac{1}{n}\sum_{i=1}^n\left(\hat{m}_1(\tilde{X}_i;\mathcal{S}_1)-\hat{m}_0(\tilde{X}_i;\mathcal{S}_1)\right)\\
&-\underbrace{\left\{\frac{1}{\sum_{i=1}^n Z_iR_i}\sum_{i=1}^n Z_iR_i\hat{m}_1(X_i;\mathcal{S}_1)\hat{w}_1(X_i) - \frac{1}{\sum_{i=1}^n(1-Z_i)R_i}\sum_{i=1}^n(1-Z_i)R_i\hat{m}_0(X_i;\mathcal{S}_1)\hat{w}_0(X_i)\right\}}_{(*)}.
\end{aligned}
$$

Then, under the assumption that $\hat{w}_z(X_i) - w_z(X_i) = o_p(1)$, we can re-write $(*)$ as:

$$
\begin{aligned}
&\frac{1}{\sum_{i=1}^n Z_iR_i}\sum_{i=1}^n Z_iR_i\hat{m}_1(X_i;\mathcal{S}_1)\hat{w}_1(X_i) - \frac{1}{\sum_{i=1}^n(1-Z_i)R_i}\sum_{i=1}^n(1-Z_i)R_i\hat{m}_0(X_i;\mathcal{S}_1)\hat{w}_0(X_i)\\
&\quad=\frac{1}{\sum_{i=1}^n Z_iR_i}\sum_{i=1}^n Z_iR_i\hat{m}_1(X_i;\mathcal{S}_1)w_1(X_i) - \frac{1}{\sum_{i=1}^n(1-Z_i)R_i}\sum_{i=1}^n(1-Z_i)R_i\hat{m}_0(X_i;\mathcal{S}_1)w_0(X_i)\\
&\qquad+\frac{1}{\sum_{i=1}^n Z_iR_i}\sum_{i=1}^n Z_iR_i\hat{m}_1(X_i;\mathcal{S}_1)\{\hat{w}_1(X_i)-w_1(X_i)\}\\
&\qquad-\frac{1}{\sum_{i=1}^n(1-Z_i)R_i}\sum_{i=1}^n(1-Z_i)R_i\hat{m}_0(X_i;\mathcal{S}_1)\{\hat{w}_0(X_i)-w_0(X_i)\}
\end{aligned}
$$

Then, taking the expectation of the term:

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{\sum_{i=1}^n Z_i R_i}\sum_{i=1}^n Z_i R_i \hat{m}_1(X_i;\mathcal{S}_1)\hat{w}_1(X_i) - \frac{1}{\sum_{i=1}^n (1-Z_i)R_i}\sum_{i=1}^n (1-Z_i)R_i\hat{m}_0(X_i;\mathcal{S}_1\hat{w}_0(X_i)\right]\\
=&\mathbb{E}\left[\hat{m}_1(X_i;\mathcal{S}_1) - \hat{m}_0(X_i;\mathcal{S}_1)\right] + o_p(1),
\end{aligned}$$

where the final equality follows from a similar argument to Theorem A.2. Then, under a valid imputation model, $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \left(\hat{m}_1(\tilde{X}_i;\mathcal{S}_1) - \hat{m}_0(\tilde{X}_i;\mathcal{S}_1)\right)\right] = \mathbb{E}\left[\hat{m}_1(X_i;\mathcal{S}_1) - \hat{m}_0(X_i;\mathcal{S}_1)\right]$. As a result, we have shown:

$$\begin{aligned}
\mathbb{E}\left[\hat{\tau}_{aug}\right] &= \mathbb{E}\left[\hat{\tau}_w\right] + \mathbb{E}\left[\hat{m}_1(X_i;\mathcal{S}_1) - \hat{m}_0(X_i;\mathcal{S}_1)\right] + o_p(1) - \mathbb{E}\left[\hat{m}_1(X_i;\mathcal{S}_1) - \hat{m}_0(X_i;\mathcal{S}_1)\right]\\
&= \mathbb{E}\left[\hat{\tau}_w\right] + o_p(1)\\
&= \tau + o_p(1)
\end{aligned}$$

As such, we have shown $\hat{\tau}_{aug} \xrightarrow{p} \tau$.

Now, assume a valid imputation model and $\hat{m}_z(X_i;\mathcal{S}_1) - m_z(X_i;\mathcal{S}_1) = o_p(1)$ for $z \in \{0,1\}$. Then,

$$\begin{aligned}
&\mathbb{E}\left[\frac{1}{\sum_{i=1}^n Z_i R_i}\sum_{i=1}^n Z_i R_i \hat{w}_1(X_i)\left(Y_i - \hat{m}_1(X_i;\mathcal{S}_1)\right)\right]\\
&= \mathbb{E}\left[\mathbb{E}\left\{\frac{1}{\sum_{i=1}^n Z_i R_i}\sum_{i=1}^n Z_i R_i \hat{w}_1(X_i)\left(Y_i - \hat{m}_1(X_i;\mathcal{S}_1)\right) \;\middle|\; X_i\right\}\right]\\
&= \mathbb{E}\left[\frac{1}{\sum_{i=1}^n Z_i R_i}\sum_{i=1}^n \frac{\hat{w}_1(X_i)}{w_1(X_i)}\mathbb{E}\left\{Y_i - \hat{m}_1(X_i;\mathcal{S}_1) \mid X_i\right\}\right]\\
&= \mathbb{E}\left[\frac{1}{\sum_{i=1}^n Z_i R_i}\sum_{i=1}^n \frac{\hat{w}_1(X_i)}{w_1(X_i)}\mathbb{E}\left\{Y_i - \hat{m}_1(X_i;\mathcal{S}_1) \mid X_i, Z_i = 1, R_i = 1\right\}\right]\\
&= \mathbb{E}\left[\frac{1}{\sum_{i=1}^n Z_i R_i}\sum_{i=1}^n \frac{\hat{w}_1(X_i)}{w_1(X_i)}\left\{\mathbb{E}\left[Y_i \mid X_i, Z_i = 1, R_i = \mathbb{1}_p\right] - \hat{m}_1(X_i;\mathcal{S}_1)\right\}\right]\\
&= o_p(1)
\end{aligned}$$

We can similarly show $\mathbb{E}\left[\frac{1}{\sum_{i=1}^n (1-Z_i)R_i}\sum_{i=1}^n (1-Z_i)R_i\hat{w}_0(X_i)\left(Y_i - \hat{m}_0(X_i;\mathcal{S}_1)\right)\right] = o_p(1)$. Then,

$$\begin{aligned}
\mathbb{E}\left[\hat{\tau}_{aug}\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \hat{m}_1(\tilde{X}_i;\mathcal{S}_1) - \hat{m}_0(\tilde{X}_i;\mathcal{S}_1)\right] + o_p(1)\\
&= \tau + o_p(1),
\end{aligned}$$

where the final equality follows from applying Theorem 4.1. As such, we have shown $\hat{\tau}_{aug} \xrightarrow{p} \tau$.

□

B.5 Example 3.1

Proof: Applying Frisch-Waugh-Lovell (FWL) theorem,

$$
\begin{aligned}
\hat{\tau}_{impute} &= \frac{\text{cov}(Z_i^{\perp\{X_1,\tilde{X}_2,R_i\}}, Y^{\perp\{X_1,\tilde{X}_2,R_i\}})}{\text{var}(Z_i^{\perp\{X_1,\tilde{X}_2,R_i\}})} \\
&= \frac{\text{cov}(Z_i^{\perp\{X_1,\tilde{X}_2,R_i\}}, \hat{\tau} Z_i^{\perp\{X_1,\tilde{X}_2,R_i\}} + \hat{\beta}_2 X_2^{\perp\{X_1,\tilde{X}_2,R_i\}})}{\text{var}(Z_i^{\perp\{X_1,\tilde{X}_2,R_i\}})} \\
&= \hat{\tau}_{oracle} + \hat{\beta}_2 \frac{\text{cov}(Z_i^{\perp\{X_1,\tilde{X}_2,R_i\}}, X_2^{\perp\{X_1,\tilde{X}_2,R_i\}})}{\text{var}(Z_i^{\perp\{X_1,\tilde{X}_2,R_i\}})}
\end{aligned}
$$

□

C Additional simulation results

C.1 Details on estimation

We provide details on how we construct each estimator in the simulation study on the fololwing table.

Estimator	**Estimation Details**
Complete Case	Subset to complete cases $R = 1$, and regress outcomes with treatment, X_1, X_2. $Y \sim Z \cdot (X_1 + X_2) \mid R = 1$
Imputation	Impute missing values in $X_2 \to \hat{X}_2$ $\hat{X}_2 = g(X_1)$, where $g(X_1) : X_2 \sim X_1 \mid R = 1$ Regress outcomes with treatment, X_1, $\hat{X}_2$, and missingness indicator. $Y \sim Z \cdot (X_1 + \hat{X}_2 + R)$
Missing Indicator	Impute missing values in $X_2 \to \hat{X}_2$ $\hat{X}_2 = \text{mean}(X_2 \mid R = 1)$ Regress outcomes with X_1, $\hat{X}_2$, interacted with treatment and missingness indicator. $Y \sim Z \cdot R \cdot (X_1 + \hat{X}_2)$
Multiple Imputation	Impute X_2, $Y(1)$, and $Y(0)$ simultaneously using Amelia. $\hat{Y}_1 = g_{\textsc{Amelia}}(Y \cdot Z, X_1, X_2)$, $\hat{Y}_0 = g_{\textsc{Amelia}}(Y \cdot (1 - Z), X_1, X_2)$
Weighting	Subset to complete cases $R = 1$ Model probability of treatment across complete cases using a logistic model. $\hat{e}(X_i; R_i = 1) = \text{logit}^{-1}(X_1 + X_2 \mid R = 1)$ Model propensity of being a complete case via GMM. $\hat{\pi}_r(X_i, Z_i) = \text{logit}^{-1}(X_1 + X_2 + Z)$ Re-weight complete cases.
Projection	Subset to complete cases $R = 1$. Model both treatment/control outcomes using a linear regression: $\hat{m}_1 : Y \sim X_1 + X_2 \mid R = 1, Z = 1$ $\hat{m}_0 : Y \sim X_1 + X_2 \mid R = 1, Z = 0$ Impute missing values in $X_2 \to \hat{X}_2$ $\hat{X}_2 = g(X_1)$, where $g(X_1) : X_2 \sim X_1 \mid R = 1$ Predict treatment/control outcomes across $R = 0$ cases using $\hat{m}_1, \hat{m}_0, \hat{X}_2$: $\hat{Y}_1 = \hat{m}_1(X_1, \hat{X}_2), \hat{Y}_0 = \hat{m}_0(X_1, \hat{X}_2)$
Augmented	Follow procedure for both weighting and projection estimators to estimate outcome models and weights. Residualize the outcomes across the complete cases. Re-weight the residuals, and augment with the outcome models.

C.2 Model Misspecification

We update the outcome data generation process to the following:

$$Y_i = \tau \cdot Z_i + \sum_{j=1}^{2} \left\{ \beta_j X_i^{(j)} + \varphi_j \big(X_i^{(j)} \cdot Z_i \big) + \delta_j \left(X_i^{(j)} \right)^2 \right\} + u_i, \text{ where } u_i \sim N(0, 1),$$

where we have added in an additional higher-order term, controlled by δ_1, δ_2. We consider four different scenarios, in which we toggle α. toggling α and δ. In particular, when $\alpha = 0$, this implies missingness in $X_i^{(2)}$ can be fully explained by variation in $X_i^{(1)}$ (i.e., missing-at-random holds). When $\alpha \neq 1$, this means missingness in $X_i^{(2)}$ depends on the values of $X_i^{(2)}$ (i.e., missing-not-at-random). Similarly, when $\delta = 0$, this implies that the outcome is a linear function of the covariates X. However, when $\delta \neq 0$, this implies that the outcome will contain non-linearities in the covariates. We summarize the different settings in Table 4.

Scenario	Description	Parameters
1	MAR, correct outcome model specification	$\delta = 0$, $\alpha = 0$
2	MAR, incorrect outcome model specification	$\delta \neq 0$, $\alpha = 0$
3	MNAR, correct outcome model specification	$\delta = 0$, $\alpha \neq 0$
4	MNAR, incorrect outcome model specification	$\delta \neq 0$, $\alpha \neq 0$

Table 4: Simulation parameters

Because we do not include higher-order terms in the models, this means that in Scenarios 2 and 4 (where the outcome includes a second-order of $X^{(1)}$), the models will be misspecified. Scenarios 1 and 3, where the outcome is a function of only the first-order covariate values, correspond to the simulation settings presented in the main manuscript.

We visualize the performance of the estimators in Figure 6, which displays the mean squared error of each estimator in each scenario. We see that the general patterns from the main manuscript (i.e., Scenario 1 and 3) hold for Scenario 2 and 4. However, we see that the projection estimator incurs more bias, as a result of the outcome model being misspecified. We see that the augmented weighted estimator, which is doubly robust, is still unbiased in Scenario 2, as the weights are correctly specified. In Scenario 4, because the imputation model does not exactly recover the density of the missing covariates, both the projection estimator and the augmented weighted estimator are still biased. However, the augmented weighted estimator has much lower bias in comparison. Furthermore, in all of these settings, the projection and augmented weighted estimator outperform the weighted estimator on an MSE basis due to the variance inflation from the re-weighting.

C.3 Simulating Modified Conditional Ignorability

We also consider an alternative alternative data generating process where we generate the outcomes such that the modified version of conditional ignorability holds. More specifically, we generate the outcomes as a piece-wise function:

$$Y_i = \begin{cases} \tau \cdot Z_i + \sum_{j=1}^{2} \left\{\beta_j X_i^{(j)} + \varphi_j \left(X_i^{(j)} \cdot Z_i\right)\right\} + u_i & \text{if } R_i = 1 \\ \tau \cdot Z_i + \beta_1 X_i^{(1)} + \varphi_1 Z_i \cdot X_i^{(1)} + u_i & \text{if } R_i = 0 \end{cases},$$

where $u_i \sim N(0, 1)$. The treatment assignment and missingness indicator are generated following the set-up in Section 6.1. We vary how correlated $X_i^{(1)}$ and $X_i^{(2)}$ are to each other, and then evaluate the performance of the estimators across these settings.

The imputation estimator and the missing indicator estimator perform well in settings when there is no correlation between $X_i^{(1)}$ and $X_i^{(2)}$. Under correlated covariate settings, the imputation estimator and missing indicator estimator incur a small amount of bias, which likely arises from estimation error due to the correlations between the covariates. In contrast, the multiple imputation estimator performs poorly when there is no correlation–likely because it struggles to recover the counterfactual outcome values. In contrast, when there is high correlation, the multiple imputation estimator is effectively unbiased. Notably, the projection estimator is unbiased in all of this setting. Table 5 summarizes the results.

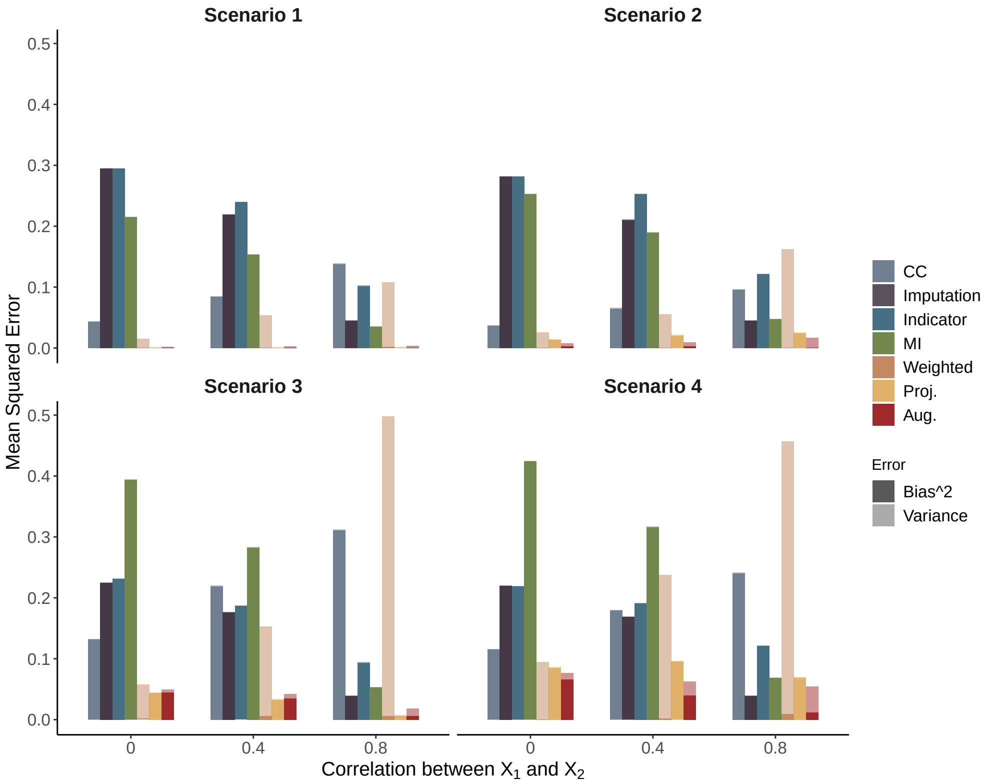
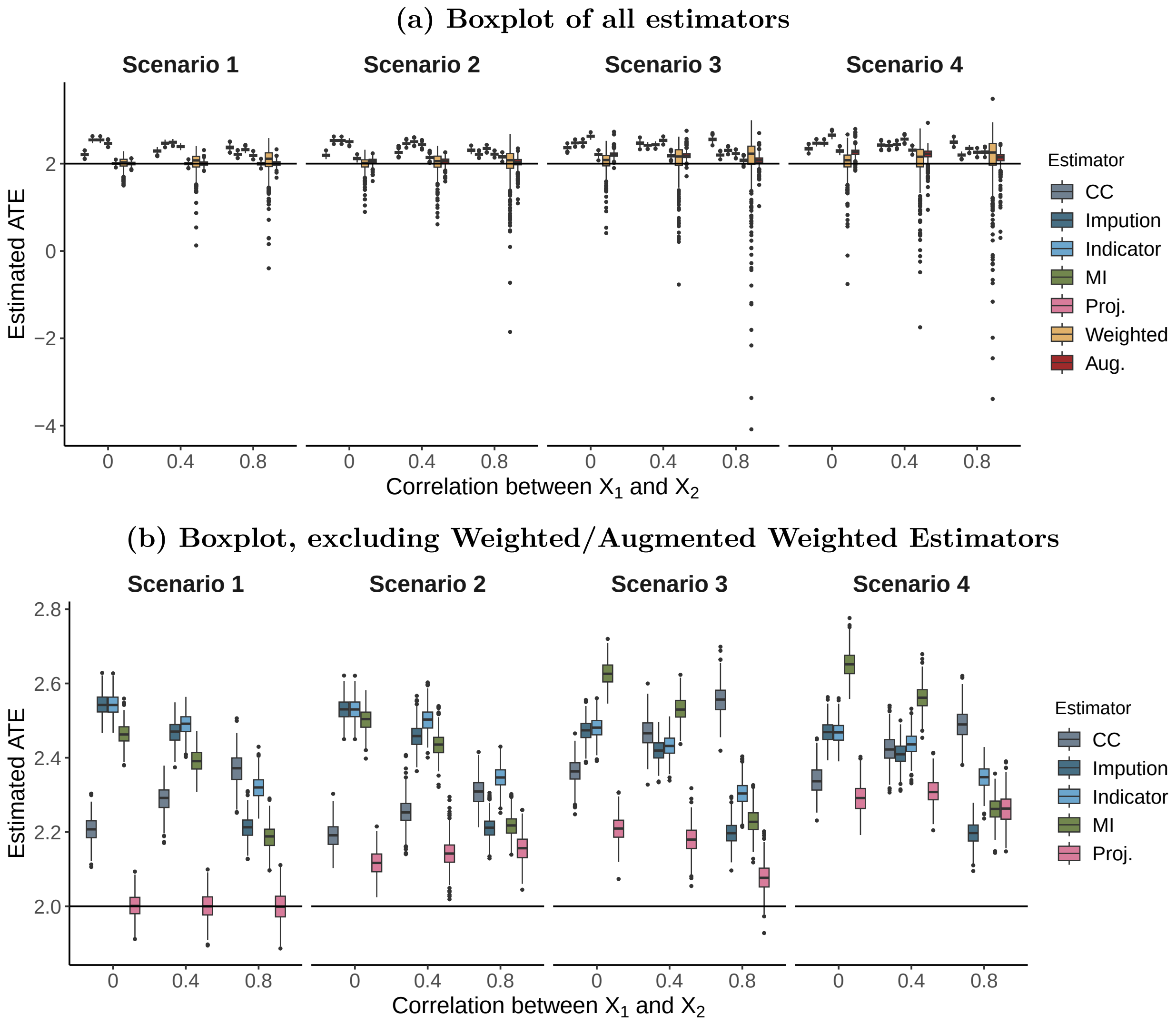


Figure 6: Simulation results across four different scenarios (described in Table 4). While the general paterns from the main manuscript hold, we additionally see that in settings where there is outcome model misspecification, the augmented weighted estimator helps mitigate the bias that arises in the projection estimator, which solely relies on outcome modeling.



(a) Boxplot of all estimators

(b) Boxplot, excluding Weighted/Augmented Weighted Estimators

Figure 7: Boxplots of the different estimators across the different simulation scenarios (described in Table 4). In facet (b) of the plot, we exclude the weighted and augmented weighted estimators, as there is substantially more variance in the resulting estimates.

	No Correlation		Medium Correlation		High Correlation	
	Bias	s.d.	Bias	s.d.	Bias	s.d.
Complete Case	0.21	0.04	0.29	0.04	0.37	0.04
Imputation	0.00	0.02	0.04	0.03	0.05	0.03
Missing Indicator	0.00	0.02	0.04	0.02	0.08	0.03
Multiple Imputation	0.10	0.03	0.07	0.03	0.00	0.03
Weighted	-0.02	0.17	0.02	0.27	0.11	0.30
Projection	0.00	0.04	0.00	0.04	0.00	0.04
Augmented Weighted	0.00	0.05	-0.00	0.06	-0.00	0.06

Table 5: Performance of the estimators under modified conditional ignorability.

C.4 Coverage Evaluation

We adopt the same simulation setup in Section 6, under Scenario 1 (i.e., MAR). We generate 100 datasets, with 1000 bootstrap iterations for each dataset, and compute the coverage as the proportion of times the estimated percentile bootstraps provide coverage of the oracle average treatment effect. Because the imputation estimator is so biased, we see that it does not provide adequate coverage. In contrast, the projection estimator has at least nominal coverage, with the percentile confidence intervals often being conservative.

Correlation	Imputation	Projection
0	0	1
0.2	0	1
0.4	0	1
0.6	0	1
0.8	0	1
1	1	0.98

Table 6: Coverage for imputation and projection estimator under MAR. We adopted the same simulation setup in section 6. The results are calculated for 100 datasets, with 1000 bootstrap each.

D Extended Empirical Results

D.1 Additional Details on Estimation

Details on fixed effects. We include fixed effects corresponding to time and state. For the outcome of proportion of bills enacted, we include a time fixed effect for the decade that the observation corresponds to (i.e., pre-1975, 1976-1985, 1986-1995, 1996- 2005, and 2006 onward). For the outcome of *polarization*, we include a time fixed effect for every five year period (i.e., pre-1995, 1996-2000, 2001-2005, and 2006 onward). Finally, for *kurtosis*, we include a time fixed effect for every five year period (i.e., pre-1985, 1986-1990, 1991-1995, 1996-2000, 2001-2005, and 2006 onward).

We also examine the percentage of missingness across the different fixed effects. See Figure 8 and 9 for a visualization. Notably, when looking at the percentage of missingness across states, we see that the state of Nebraska is missing certain covariates 100% of the time (and in some cases, missing all of the covariate values). Because this is a violation of positivity, we remove

Nebraska from the analysis. Looking at the percentage of missingness across the time fixed effects, we see that there is a greater proportion of missing values for more recent time periods. From the visualizations, it is clear that the missingness is not completely at random. As a result, we expect that methods like a complete case estimator will produce biased estimates.

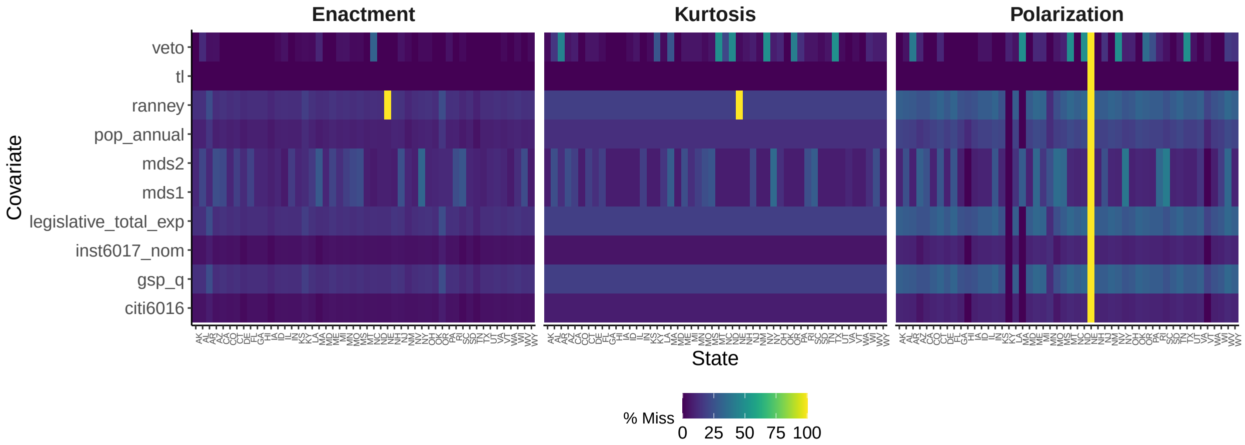


Figure 8: Heat maps corresponding to the percentage of missing covariate values, sorted by state-level fixed effects. We see that across all three outcome measures, the Ranney Index is fully missing for the state of Nebraska (abbreviated 'NE'). For the outcome of polarization, all covariate values are fully missing for Nebraska.

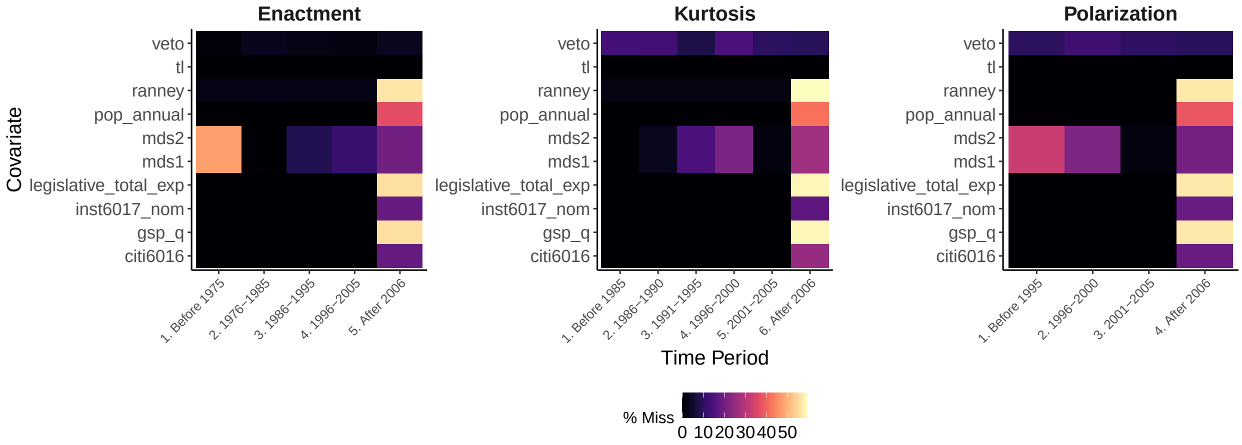


Figure 9: Heat maps corresponding to the percentage of missing covariate values, sorted by time-level fixed effects. We see there is a greater proportion of missing values for the time period corresponding to after 2006 for all three outcomes.

Details on implementation. We provide details on the implementation of each of the estimators evaluated in the empirical application.

- **Imputation estimator:** the imputation estimator is estimated in the same way as Harden and Kirkland (2021). We first impute the missing covariates using Amelia (Honaker et al., 2011). We simulate five datasets from Amelia. Across each of the imputed datasets, we run a regression including the covariates and state/time fixed effects, as well as a missingness indicator for whether or not the covariate was initially observed. We then average the resulting estimator for the final estimate.

- **Complete case estimator:** Across the complete cases, we run a regression including the covariates and state/time fixed effects.
- **Multiple imputation estimator:** We use Amelia to directly impute the missing counterfactual outcome values. This simultaneously imputes the missing covariate values, as well as the missing counterfactual outcomes. We generate five simulated datasets from Amelia, and then directly compute the ATE using the imputed counterfactual outcomes, and then average the resulting multiple imputation estimator across the five simulated datasets for the final estimate.
- **Weighted estimator:** Across the complete cases, we model the probability of treatment across the complete cases using a logistic model, and model the propensity of being a complete case via GMM, using the available pre-treatment covariates.
- **Projection estimator:** Across the complete cases, we estimate a causal forest, distilled into a linear model (e.g., leveraging a best linear predictor approach from Chernozhukov et al., 2018). Then, we impute the missing covariates using Amelia (Honaker et al., 2011). We simulate five datasets from Amelia. Across each of the imputed datasets, we use the estimated model to predict the overall ATE.
- **Augmented:** Using the weights estimated in the same way as the weighted estimator and the outcome model estimated in the projection estimator, we combine the two together for the augmented weighted estimator.

D.2 Evaluating the Imputation Model Performance

We simulate three different missingness mechanisms for each covariate: (1) missing-at-random; (2) missing-not-at-random (with a slight dependency on the missing covariate values); (3) missing-not-at-random (with a larger dependency on the missing covariate values). For each missingness mechanism, we calibrate the proportion of total missing values to match the true proportion of missing values. See Table 8 for the full validation results. To construct the missing-at-random missingness mechanism, we begin by estimating a logistic regression using the missing values in each covariate, and the other covariates:

$$\Pr(\tilde{R}^{\text{MAR}}_{(j)} \mid X_i^{-(j)}; \alpha^{(j)}) = \text{logit}\left(\sum_{k \neq j} \hat{\beta}_k X_i^{(k)} + \alpha^{(j)}\right)$$

We then add in dependency in the underlying logit function to the covariate value for scenarios (2) and (3):

$$\Pr(\tilde{R}^{\text{MNAR}}_{(j)} \mid X_i^{-(j)}; \alpha^{(j)}) = \text{logit}\left(\sum_{k \neq j} \hat{\beta}_k X_i^{(k)} + \gamma X_i^{(j)} + \tilde{\alpha}^{(j)}\right)$$

where for MNAR (low), we set $\gamma = 0.5$, and for MNAR (high), we set $\gamma = 1$. For both settings, we set $\tilde{\alpha}^{(j)}$ to ensure the proportion of the missing values for each covariate matches the observed missingness proportions. We simulate the missingness mechanisms across 100 iterations and compute the mean absolute error for each covariate. The full results are provided in Table 8.

In particular, we are most worried about covariates that explain the most variation in the outcomes. We examine the variable importance associated with each of the models for the three outcomes. Figure 10 provides a visualization.

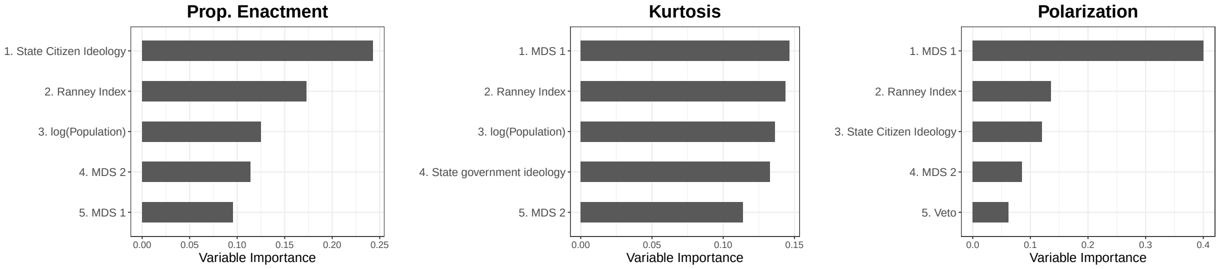


Figure 10: We plot the top five covariates that have the greatest variable importance, as proxied by the number of times they were split on within the causal random forest (Tibshirani et al., 2018).

D.3 Observable Implications of Outcome-Response Ignorability.

We report the R^2 values estimated across the observed treatment and control groups.

	Complete Case $R_i = \mathbb{1}_p$		Projected Subset $R_i \neq \mathbb{1}_p$	
Outcome	$Z = 0$	$Z = 1$	$Z = 0$	$Z = 1$
Proportion of Bills Enacted	0.32	0.22	0.20	0.15
Polarization	0.32	0.18	0.23	0.11
Kurtosis	0.01	0.02	0.01	0.01

Table 7: R^2 values for the regressions fit across the observed treatment and control groups.

D.4 Additional Tables

Covariate	Prop. Missing	MAR	MNAR (Low)	MNAR (High)
Outcome: Proportion of Bills Enacted				
State Citizen Ideology	0.04	0.35 (0.04)	0.35 (0.04)	0.35 (0.03)
Gross State Product	0.14	0.83 (0.09)	1.09 (0.17)	1.18 (0.23)
State Gov. Ideology	0.04	0.31 (0.04)	0.33 (0.03)	0.33 (0.03)
Total Expenditure	0.14	1.25 (0.73)	3.99 (2.44)	5.61 (4.4)
MDS 1	0.14	21.65 (1.28)	19.36 (0.69)	19.06 (0.48)
MDS 2	0.14	17.14 (0.94)	17.25 (0.66)	17.11 (0.43)
State Population	0.09	0.48 (0.02)	0.51 (0.02)	0.53 (0.02)
Ranney Index	0.15	0.13 (0.01)	0.14 (0.01)	0.14 (0.01)
Num. of Bills Vetoed	0.02	1.55 (0.52)	1.09 (0.2)	1.03 (0.11)
Outcome: Kurtosis				
State Citizen Ideology	0.08	0.31 (0.04)	0.3 (0.03)	0.32 (0.03)
Gross State Product	0.18	0.6 (0.06)	0.83 (0.15)	0.92 (0.18)
State Gov. Ideology	0.05	0.29 (0.03)	0.31 (0.03)	0.32 (0.03)
Total Expenditure	0.18	0.71 (0.35)	4.01 (3.72)	5.23 (2.92)
MDS 1	0.14	11.18 (0.88)	9.95 (0.6)	9.75 (0.38)
MDS 2	0.14	309.44 (22.18)	317.09 (18.06)	306.16 (12.27)
State Population	0.13	0.4 (0.03)	0.44 (0.03)	0.47 (0.02)
Ranney Index	0.20	0.12 (0)	0.12 (0.01)	0.13 (0.01)
Num. of Bills Vetoed	0.11	1.35 (0.21)	1.05 (0.09)	0.99 (0.04)
Outcome: Polarization				
State Citizen Ideology	0.09	0.29 (0.03)	0.29 (0.03)	0.3 (0.02)
Gross State Product	0.27	0.27 (0.06)	0.44 (0.15)	0.61 (0.17)
State Gov. Ideology	0.09	0.3 (0.03)	0.3 (0.03)	0.3 (0.03)
Total Expenditure	0.27	0.9 (0.93)	3.38 (4.06)	4.6 (4.04)
MDS 1	0.16	5.72 (0.52)	5.16 (0.32)	5.02 (0.21)
MDS 2	0.16	5.71 (0.48)	6.11 (0.42)	6.05 (0.29)
State Population	0.18	0.24 (0.03)	0.28 (0.03)	0.3 (0.03)
Ranney Index	0.27	0.09 (0)	0.1 (0.01)	0.1 (0.01)
Num. of Bills Vetoed	0.10	1.23 (0.13)	1.1 (0.09)	1.06 (0.08)

Table 8: We report the normalized mean absolute average error for imputing missing values for each covariate under three different missingness mechanisms.

Outcome	Estimator	Estimate	CI (Low)	CI (High)
Enactment	Complete Case	-0.02	-0.04	0.00
Enactment	Imputation (OLS)	-0.01	-0.03	0.00
Enactment	Multiple Imputation	0.05	0.04	0.06
Enactment	Projection	0.04	0.03	0.05
Enactment	Weighted	0.06	0.05	0.08
Enactment	Augmented	0.07	0.04	0.08
Polarization	Complete Case	-0.12	-0.20	-0.02
Polarization	Imputation (OLS)	0.01	-0.11	0.15
Polarization	Multiple Imputation	0.01	-0.03	0.03
Polarization	Weighted	0.00	-0.08	0.05
Polarization	Projection	-0.09	-0.13	-0.07
Polarization	Augmented	-0.12	-0.18	-0.08
Kurtosis	Complete Case	0.11	0.05	0.17
Kurtosis	Imputation (OLS)	0.10	0.05	0.15
Kurtosis	Multiple Imputation	0.01	-0.01	0.03
Kurtosis	Weighted	0.03	0.00	0.08
Kurtosis	Projection	0.05	0.02	0.08
Kurtosis	Augmented	0.09	0.04	0.14

Table 9: Corresponding point estimates to Figure 4.