STAT 530 Project Report

Caleb Skinner

April 29, 2025

Abstract

The Doubly Robust (DR) estimator is consistent if either the outcome model or the propensity score model is specified correctly. If the set of confounders is unknown, then researchers may struggle to specify either model correctly. Recently, Cefalu et al. (2017) leverage advancements in causal Bayesian Model Averaging techniques to account for the uncertainty in model selection of the DR estimator. In this report, I detail the rational, theoretical results, and motivation of this Model Averaged DR (MA-DR) estimator and explore its advantages and disadvantages. In particular, I employ a three-part simulation study to explore the performance of the estimators in the presence of (1) an unknown confounding set, (2) a known confounding set with complex relationships, and (3) instrumental variables and covariates that are weakly associated with the outcome. I find the MA-DR estimator to be a powerful method for identifying the causal effect in real-world data applications.

1 Introduction

In economic, political, and medical situations, researchers are often interested in an exposure's causal effect on an outcome. The potential outcomes framework is helpful for establishing the meaning of "causal effect". Each unit in a study has the potential to receive all levels of an exposure A on a outcome of interest Y. The unit's potential outcomes within all possible worlds are denoted $Y(a)$. The exposure's average causal effect, then, can be written $\Delta = E[Y(1)] - E[Y(0)]$. In observational studies, simple estimates of this causal effect are biased by a set of confounders that impact both the outcome and exposure of a unit. A failure to account for this confounding can produce artificial association between an exposure and outcome and lead to misinformed inference. Fortunately, if a researcher can identify a set of variables that sufficiently accounts for this confounding, then there several methods that can yield unbiased estimates. The outcome modeling approach includes a sufficient set of confounders $\mathbf{X}$ in the outcome model to clear any confounding-induced association between the exposure and outcome. The outcome modeling approach can be estimated by

$$\hat{E}[Y(a)] = \frac{1}{n}\sum_i \hat{E}_{\mathbf{X}}[\hat{E}[Y_i|A_i = a, \mathbf{X}_i]]$$

On the other hand, the exposure modeling approach models each unit's propensity to receive the exposure to standardize across all units (Rosenbaum and Rubin, 1983). In this method, known as inverse probability weighting (IPW), researchers estimate each unit's propensity for exposure, $\pi(X) = P(A = 1|X)$. From here, one estimates the potential outcomes as follows

$$\hat{E}[Y(1)] = \frac{1}{n}\sum_i \left[\frac{A_i Y_i}{\hat{\pi}(\mathbf{X_i})}\right]$$
$$\hat{E}[Y(0)] = \frac{1}{n}\sum_i \left[\frac{(1 - A_i)Y_i}{\hat{\pi}(\mathbf{X_i})}\right]$$

These two methods yield unbiased results if the models are specified correctly. However, if one or both of the models are incorrect, the two methods may yield different results (Zhang and Schaubel, 2012). This may lead to confusion and disputed inferences. The Doubly Robust (DR) estimator incorporates the consistency guarantees from both models, producing an unbiased result if either the outcome model or the propensity score model is specified correctly (Bang and Robins, 2005). It affords the researcher two opportunities to achieve an unbiased estimate. The composition of the DR estimator clearly contains both of the two methods and can be expressed

$$\hat{E}[Y(1)] = \frac{1}{n}\sum_i \left[\frac{A_i Y_i}{\hat{\pi}(\mathbf{X}_i)} - \frac{A_i - \hat{\pi}(\mathbf{X}_i)}{\hat{\pi}(\mathbf{X}_i)}\hat{E}[Y_i|A_i = 1, \mathbf{X}_i]\right]$$
$$\hat{E}[Y(0)] = \frac{1}{n}\sum_i \left[\frac{(1 - A_i)Y_i}{1 - \hat{\pi}(\mathbf{X}_i)} + \frac{A_i - \hat{\pi}(\mathbf{X}_i)}{1 - \hat{\pi}(\mathbf{X}_i)}\hat{E}[Y_i|A_i = 0, \mathbf{X}_i]\right]$$

The difference of the two potential outcome estimates yields the doubly robust estimator $\hat{\Delta}^{DR}$. However, if both models are specified incorrectly, the DR estimator yields mixed results, at best, and may not improve either model. It is essential that the correct set of confounders is identified in either the outcome or exposure model.

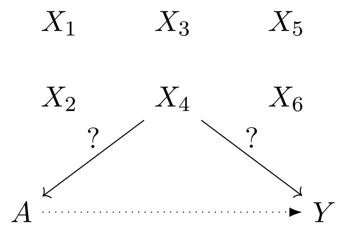

Figure 1: Unknown Set of Confounders DAG

In many observational studies, the task of identifying the confounding set is incredibly challenging (see Figure 1 for a DAG representation of the problem). Even with the assistance of a domain expert or a methodical principle like the disjunctive criterion (see VanderWeele (2019)), the exact confounding set is typically unknown. If the dimension of potential confounders is large, including all confounders in both models is at the very least computationally infeasible. At worst, including excess confounders will add significant noise to the estimate and disrupt inference. At the same time, neglecting a single confounder can bias results.

At first, this appears to be a typical model selection problem. Defining a finite collection of models for the exposure $\mathcal{M}^{\text{em}}$ and outcome $\mathcal{M}^{\text{om}}$, a procedure traverses this space and selects an optimal model from each model class. Classical model selection procedures like stepwise regression or lasso regression reduce the dimensionality of the confounding set by removing confounders with negligible association to the outcome or exposure. These methods fail for two major reasons. First, only one confounding set is selected, and all future analysis rests on the validity of that model. A single selected model fails to account for the uncertainty in the model selection process. As the dimensionality of potential confounders increases, the corresponding model space increases exponentially. In a large model space, these algorithms may be caught in local maximas, and are, thus, liable to excluding important confounders. Second, these model selection methods account for the confounder's association with the exposure and outcome separately. This disjoint approach can create significant bias. For example, including covariates in the exposure model that are not associated with the outcome can diminish the efficiency of the estimator (Rubin, 1997). Simple model selection procedures will include potential confounders in the exposure model even if they possess weak to no association with the outcome. Together, these difficulties complicate inference and make conclusions from model selection methods potentially fallible.

In recent years, several data-driven methods have emerged to account for the uncertainty in this selection of confounders (see Beaudoin and Talbot (2022) for a recent breakdown of available methods). These methods secure an unbiased estimate of the effect of the exposure if the true confounding set is contained within the set of potential confounders. These methods employ both Frequentist and Bayesian methodologies and secure uncertainty quantification from a myriad of the methods. In this report, I will one consider one such method that accounts for model uncertainty and preserves the doubly robust property. Cefalu et al. (2017) propose the model averaged doubly robust (MA-DR) estimator to account for model uncertainty in both the exposure and outcome models. In Section 2, I review the methodology of the MA-DR estimator and identify competing methods. Section 3 compares the MA-DR estimator with these competing methods in a variety of simulation scenarios, and the paper is concluded in Section 5.

2 Review of Methods

2.1 MA-DR Estimator

Cefalu et al. (2017) draw on principles from Bayesian Model Averaging (Hoeting et al., 1999) to account for the uncertainty of a selected exposure and outcome model. Let $\mathcal{M}^{em} = \{\mathcal{M}_1^{em}, \mathcal{M}_2^{em}, \ldots, \mathcal{M}_{M_{em}}^{em}\}$ be the exposure model space and $\mathcal{M}^{om} = \{\mathcal{M}_1^{om}, \mathcal{M}_2^{om}, \ldots, \mathcal{M}_{M_{om}}^{om}\}$ be the outcome model space where M_{em} and M_{om} are number of potential exposure and outcome models, respectively. The DR estimator $\hat{\Delta}_{jk}^{DR}$ is applied to each combination of the j^{th} exposure and k^{th} outcome models. The model averaged doubly robust estimator $\hat{\Delta}_{DR}^{MA}$ takes a weighted average of each model specific DR estimator in the following form

$$\hat{\Delta}_{DR}^{MA} = \sum_{jk} w_{jk} \hat{\Delta}_{jk}^{DR}$$

The weight w_{jk} is the posterior probability of the models $P(\mathcal{M}_j^{em}, \mathcal{M}_k^{om} | \text{Data})$. Crucially, these weights are estimated independently of $\hat{\Delta}_{jk}^{DR}$. The weight depends on the likelihoods and the prior distribution on the model space, so the choice of prior distribution carries substantial impact on the estimator. A simple uniform prior on the model space implies that each model is equally likely a priori, and that the exposure and outcome model are independent. While appealing, this independence gives rise to a disjoint model selection procedure that may exclude important covariates. To avoid this, Cefalu et al. (2017) propose a prior distribution on the model space that links information from the exposure and outcome models through prior model dependence. The objective is to exclude instrumental variables, covariates that are associated with the exposure but not the outcome, from the exposure model. Specifically, the prior odds for the exposure model conditional on the outcome model can be expressed

$$\frac{P(\mathcal{M}_j^{em}|\mathcal{M}_k^{om})}{P(\mathcal{M}_1^{em}|\mathcal{M}_k^{om})} = \begin{cases} 1, & \text{if } \mathcal{M}_j^{em} \subset \mathcal{M}_k^{om} \\ \tau, & \text{otherwise} \end{cases}$$

where $\tau \in [0, 1]$ determines the linkage between the two models. With $\tau = 1$, the prior reverts to the simple uniform prior, but with $\tau = 0$, the set of potential confounders in the exposure model is restricted to those included in the outcome model.

The MA-DR estimator possess favorable theoretical justification that extend those of the DR estimator. Cefalu et al. (2017) prove the asymptotic consistency of the Model Averaged Estimator provided that the true model $\mathcal{M}_{true}$ is contained within the model space $\mathcal{M}$ and sufficient regularity conditions are met. This result extends the robustness property over the entire model space of exposure and outcome models.

In practice, it is difficult to produce exact estimates of the posterior model probabilities. The authors deviate from a fully Bayesian methodology and recommend approximating these posterior model probabilities with the Bayesian Information Criterion (Schwarz, 1978). Moreover, as the dimensionality of the model space increases, it becomes infeasible to evaluate this approximation for every model. This burden is mitigated by a Markov chain Monte Carlo algorithm that searches the model space. Despite it's Bayesian formulation, the MA-DR estimator relies on bootstrapping techniques for uncertainty quantification. In moderate situations, this is feasible, but with large sets of confounders, the computational burden can be immense.

2.2 Foundational Estimators

Much of the inspiration for the MA-DR estimator originates in the Bayesian Adjusting for Confounding (BAC) prior proposed by Wang et al. (2012). The BAC estimator also leverages Bayesian Model Averaging (BMA) techniques and utilizes a dependent prior to extend consistency over a class of models. For years, it was understood that the traditional BMA assigned significant weight to models excluding important confounders. If a single confounder is omitted from the outcome model, the estimate will be biased. The BAC prior resolves this dilemma with a dependent prior structure that forces potential confounders that are strongly associated with the exposure into the outcome model.

While the MA-DR estimator and BAC estimator share strong resemblances, the prior distributions imposed on the model space differ in strategy significantly. The MA-DR approach aims to improve efficiency by removing instrumental variables from the exposure model, while the BAC approach ensures that all confounders associated with the outcome are included in the outcome model. These objectives present a tradeoff. For example, removing instruments from the exposure model may incidentally remove covariates that are weakly associated with the outcome. At times, these estimators will produce contradicting results, and I will explore these differences in Section 3.3.

The model averaged DR estimator is preceded by another DR estimator robust over a model space. The Multiply Robust Estimator (Han and Wang, 2013) define model spaces for the exposure and outcome model and propose a method similar to that of inverse probability weighting. The authors create subject weights with a series of constraints on the maximum likelihood estimates of each potential model. They guarantee a consistent estimator if the true exposure or outcome model is present in one of these two model spaces. Unfortunately, the method lacks available code or means to apply it in a simulation. Still, the method pioneers a DR estimator consistent over a model space, so it deserves a reference.

3 Simulations

I conduct three simulations to illustrate the strengths and weaknesses of the MA-DR estimator in a finite sample. Within these simulations, I compare the results of the MA-DR estimator, BAC estimator, and various stepwise model selected estimators. The model selected estimators include the estimator from the model selected outcome model (MS-O), the IPW estimator from model selected exposure model (MS-E), and model selected DR estimator (MS-DR). As a reference, I display the results of the "gold standard" (GS) estimator, an estimator obtained using the true outcome model. Since this estimator mimics the data generating mechanism, it yields unbiased and efficient results. I use the `madr` package for the MA-DR estimator and the `BAC` package for the BAC estimator.

First, I construct a set of potential confounders with varying associations with the exposure and outcome. This simulation replicates the simulation settings in Cefalu et al. (2017), and adds the BAC estimator. Second, I introduce a complex setting where the set of confounders is known, but the relationships between the confounders and the outcome and exposure are complex and non-linear. Lastly, I put forth two scenarios specifically designed to parse the advantages and disadvantages of the MA-DR and BAC estimators.

In each of the simulations, I specify the number of potential confounders p and samples n.

Unless otherwise specified, I generate the data as follows: the potential confounders $X_1, ..., X_p \overset{\text{iid}}{\sim} N(0, 1)$; the exposure $A \sim \text{Bernoulli}(p = \text{expit}(X\alpha))$; and the outcome $Y \sim N(\Delta A + X\beta, \sigma^2)$. I consider several values for the confounders level of association with the exposure α and the confounders level of association with the outcome β. For all simulations, I consider a true effect of $\Delta = 1$.

Within each scenario, I produce 500 data sets and compute the relevant estimates. Of the sample of 500 estimates, I compute the mean estimate and mean absolute error (ℓ_1 error) of each estimate (with respect to the true effect of 1). In practice, one would utilize bootstrap sampling to construct confidence intervals around each of the estimates. This would facilitate comparisons in coverage rates and efficiency. Unfortunately, for computational reasons, this exercise is impractical for a simulation of this magnitude.

3.1 Unknown Confounding Sets

In modern data applications, the set of potential confounders can be of substantial size. An appropriate estimator must be able to identify the true set of confounders even in small samples. To analyze these conditions, Cefalu et al. (2017) consider a space with 100 potential confounders and 200 samples. Unfortunately, this is not computationally feasible, so I set $p = 15$ covariates with $n = 200$ samples and variance $\sigma^2 = 1$. I consider four scenarios of values of α and β. In each case, the first five covariates have non-zero parameters, and the next ten are entirely noise. The table of parameter values is located in the Appendix Table 4.

I produce 500 data sets and compute the MA-DR, BAC, MS-O, MS-E, and MS-DR estimates for each scenario. In Table 1, I report the the mean estimated effect and mean absolute error of each estimate. As expected, the Gold Standard for each scenario produces a mean near 1 and the smallest ℓ_1 error of the estimators. In Scenario 1, all potential confounders are purely noise, so each of the methods perform well. However, in Scenarios 2-4, more complex confounding structures produce bias within the model selected methods. In each scenario, the mean of each method is far from 1. It is also worth noting that MS-DR is always outperformed by MS-E. The MA-DR and BAC estimators, however, are able to preserve estimates centered around 1. The two model averaging estimates produce similar means and error, and outperform the model selection methods in each of these scenarios.

Method	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
	Mean	ℓ_1 error	Mean	ℓ_1 error	Mean	ℓ_1 error	Mean	ℓ_1 error
GS	1.003	0.081	0.993	0.083	0.995	0.081	1.000	0.083
MS-E	0.803	0.199	0.931	0.120	0.855	0.214	1.433	0.450
MS-O	1.005	0.109	1.314	0.335	1.253	0.391	2.572	1.570
MS-DR	1.005	0.109	1.314	0.335	1.223	0.349	2.572	1.570
MA-DR	1.003	0.109	0.999	0.120	0.993	0.112	1.001	0.125
BAC	1.003	0.110	0.996	0.121	0.994	0.134	1.001	0.122

Table 1: Unknown Confounding Sets Simulation - Results

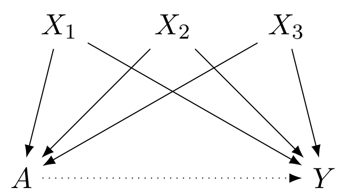

Figure 2: Complex Simulation DAG

3.2 Complex Confounding Relationships

Often, the set of confounders is known to a researcher, but the exact relationship between the confounders and the exposure and outcome is unknown or uncertain. For example, some confounders interact with each other to form complex relationships with an outcome. In other situations, confounders may possess a quadratic or piecewise relationship with an exposure or outcome. In response, it is common for analysts to add more terms to fit the data well. To mimic these circumstances, I develop a simulation scenario with $p = 3$ known confounders and $n = 500$ samples. A researcher would draw a DAG as shown in Figure 2, but the DAG is unable to express the nature of the relationships. I generate the treatment $A \sim$ Bernoulli $(p = \text{expit}\left(X_1X_2 + \mathbb{1}(X_3 > 1) + 2\mathbb{1}(X_3 > 2)\right))$ and outcome $Y \sim N\left(A + \frac{1}{2}X_1 + \frac{1}{2}X_2 - X_1X_2 + \mathbb{1}(X\right.$ Clearly, both the exposure and outcome are generated by complex terms.

I produce 500 data sets and compute the MA-DR, BAC, MS-O, MS-E, and MS-DR estimates. For each method, I display results for potential confounding sets of $\mathbf{X} = \{X_1, X_2, X_3\}$ and $\mathbf{X}_{\text{extra}} = \{X_1, X_2, X_3, X_1^2, X_2^2, X_3^2, X_1 \cdot X_2, X_1 \cdot X_3, X_2 \cdot X_3\}$ in Table 2. Figure 3 displays the mean estimates and 95% intervals of the estimates across the 500 replications. All methods suffer when extra terms are not added to the model space. Interestingly, the performance of these models are ordered roughly in terms of the complexity of the method. The simplest procedure, IPW, performs the best, and the model averaged methods perform the worst. When extra terms are added, the

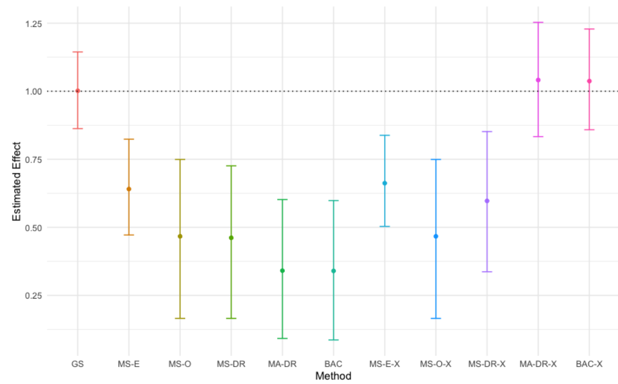

Figure 3: Complex Relationship Estimates and Intervals

estimates of the two model averaged methods improve dramatically, while the model selected estimates are poor. These results lead to two major conclusions. First, the BAC and MA-DR estimators are not guaranteed to give better estimates than the model selected estimators if the true confounders are not in the confounding set. Second, the BAC and MA-DR methods are vastly better suited than model selection techniques to approximate complex confounding relationships.

Method	Simple		Extra Terms	
	Mean	ℓ_1 error	Mean	ℓ_1 error
GS	1.002	0.058	1.002	0.058
MS-E	0.641	0.359	0.662	0.338
MS-O	0.467	0.533	0.467	0.533
MS-DR	0.462	0.538	0.597	0.403
MA-DR	0.341	0.659	1.041	0.092
BAC	0.340	0.660	1.037	0.085

Table 2: Complex Relationships Simulation - Results

3.3 MA-DR vs BAC

Despite their unique constructions, the MA-DR and BAC estimators have performed very similarly in the simulation contexts thus far. This last simulation emphasizes the differences between the methods and identifies a situation where each is superior to the other. First, I introduce significant instrumental variables into the model space. Including these variables into the model decreases efficiency (Rubin, 1997) and may introduce bias (Pearl, 2010). The MA-DR estimator is better organized to prohibit instrumental variables from entering the exposure and outcome models than the BAC estimator. In this first scenario, I consider a space of $p = 10$ potential confounders with $n = 200$ samples. Three covariates are instruments, three are confounders, and four are noise variables. The outcome is generated with $\sigma^2 = 4$. Second, I introduce several variables weakly associated with the outcome and strongly associated with the exposure. Excluding these variables from the outcome model biases the estimate. The BAC estimator forces these variables into the outcome model, but the MA-DR estimator may unintentionally exclude them. I place the exact coefficients of each model in the Appendix Table 5.

I produce 500 data sets and place the estimates in Table 3. Overall, the simulation results fit the intuition. The MA-DR estimate yields much smaller ℓ_1 error than the BAC estimate when instruments are introduced into the model space. Conversely, the BAC produces a smaller ℓ_1 error in the presence of important confounders that are weakly associated with an outcome.

Method	Scenario 1		Scenario 2	
	Mean	ℓ_1 error	Mean	ℓ_1 error
MA-DR	1.024	0.255	1.369	0.419
BAC	1.102	0.359	1.108	0.347

Table 3: MA-DR vs BAC Simulation - Results

4 Conclusions

In this report, I demonstrate the difficulty of selecting a model in the presence of unknown sets of confounders. In these circumstances, the MA-DR estimator provides a data-driven estimate of the causal effect that is consistent if the true model is present in a specified space of models. The MA-DR estimator takes a weighted average of the DR estimates over a model space, utilizing BMA techniques and a dependent prior distribution on the model space to efficiently assign model weights. The MA-DR estimator extends the robustness property of the DR estimator over an entire class of models. I demonstrate that this estimator is preferable to model selected estimators in situations with unknown sets of confounders. Moreover, in settings with known confounders and complex confounding relationships, I show that the MA-DR estimator is able to approximate the true confounding relationships with significantly less bias than the model selection methods. The MA-DR estimator is similar to another Bayesian Model Averaging technique, the BAC estimator. They perform similarly in most simulations, but I demonstrate that the MA-DR estimator is more robust to the presence of instruments and the BAC estimator is robust to confounders that weakly associated with the outcome.

These techniques are conducive to real-world applications. Cefalu et al. (2017) and Wang et al. (2012) extend their methods to medical and environmental data sets with unknown sets of confounders. These potential confounding sets number 33 and 164, respectively. With sets of confounders around this size, the computational burden for these methods is tenable. We cannot know the true causal effect in these real data applications, but theoretical results and simulations provide assurance that the estimators provide unbiased estimates of the causal effect. Thus, the MA-DR estimator presents itself as a viable option for researchers searching for a data-driven method to identify a set of confounders and adjust for confounding.

References

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

Beaudoin, C. and Talbot, D. (2022). A generalized double robust bayesian model averaging approach to causal effect estimation with application to the study of osteoporotic fractures. *Journal of Causal Inference*, 10(1):335–371.

Cefalu, M., Dominici, F., Arvold, N., and Parmigiani, G. (2017). Model averaged double robust estimation. *Biometrics*, 73(2):410–421.

Han, P. and Wang, L. (2013). Estimation with missing data: beyond double robustness. *Biometrika*, 100(2):417–430.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical Science*, 14(4):382–417.

Pearl, J. (2010). *On a class of bias-amplifying variables that endanger effect estimates*. AUAI Press, Arlington, Virginia, USA.

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

Rubin, D. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, 127(8):757–763.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.

VanderWeele, T. (2019). Principles of confounder selection. *European Journal of Epidemiology*, 34(1):211–219.

Wang, C., Parmigiani, G., and Dominici, F. (2012). Bayesian effect estimation accounting for adjustment uncertainty. *Biometrics*, 68(3):661–671.

Zhang, M. and Schaubel, D. (2012). Double-robust semiparametric estimator for differences in restricted mean lifetimes in observational studies. *Biometrics*, 68(4):999–1009.

5 Appendix

	Scenario	$\boldsymbol{\alpha}$ (Exposure)	$\boldsymbol{\beta}$ (Outcome)
1	No confounding	(0.4, 0.3, 0.2, 0.1, 0)	(0, 0, 0, 0, 0)
2	Mild confounding	(0.5, 0.5, 0.1, 0, 0)	(0.5, 0, 1, 0.5, 0)
3	Strong outcome	(0.1, 0.1, 1, 1, 1)	(2, 2, 0, 0, 0)
4	Strong confounding	(0.5, 0.4, 0.3, 0.2, 0.1)	(0.5, 1, 1.5, 2, 2.5)

Table 4: Unknown Confounding Sets Simulation - Coefficients

Scenario	Scenario 1 - Instruments	Scenario 2 - Weak outcome
$\boldsymbol{\alpha}$ (Exposure)	(1, 2, 3, 0.5, 0.5, 0.5, 0, 0, 0, 0)	(0.5, 1.0, 1.5, 2.0, 2.5, 0, 0, 0, 0, 0)
$\boldsymbol{\beta}$ (Outcome)	(0, 0, 0, 0.5, 1.0, 1.5, 0, 0, 0, 0)	(0.1, 0.2, 0.2, 0.2, 0.1, 0, 0, 0, 0, 0)

Table 5: MA-DR vs BAC Simulation - Coefficients