Biometrika (2023), **XXX**, X, *pp*. 1–17
Printed in Great Britain

This article has been accepted for publication in *Biometrika*, published by Oxford University Press.

Variance-based sensitivity analysis for weighting estimators results in more informative bounds

BY MELODY HUANG
Yale University, New Haven, Connecticut 06520, U.S.A.
melody.huang@yale.edu

AND SAMUEL D. PIMENTEL
University of California, Berkeley, Berkeley, California 94720, U.S.A.
spi@berkeley.edu

SUMMARY

Weighting methods are popular tools for estimating causal effects, and assessing their robustness under unobserved confounding is important in practice. Current approaches to sensitivity analyses rely on bounding a worst-case error from omitting a confounder. In the following paper, we introduce a new sensitivity model called the *variance-based sensitivity model*, which instead bounds the distributional differences that arise in the weights from omitting a confounder. The variance-based sensitivity model can be parameterized by an R^2 parameter that is both standardized and bounded. We demonstrate, both empirically and theoretically, that the variance-based sensitivity model provides improvements on the stability of the sensitivity analysis procedure over existing methods. We show that by moving away from worst-case bounds, we are able to obtain more interpretable and informative bounds. We illustrate our proposed approach on a study examining blood mercury levels using the National Health and Nutrition Examination Survey (NHANES).

Some key words: causal inference; sensitivity analysis; inverse propensity score weighting

1. INTRODUCTION

In observational studies of causal effects, researchers must address possible confounding effects from non-random treatment assignment. Typically, one relies on pre-treatment covariates either to re-weight units based on propensity of treatment, or model the outcome of interest. In practice, researchers have no way of knowing whether the included covariates fully capture the confounding effects. When confounders are omitted, the resulting estimates will be biased. Sensitivity analyses speak to this concern by allowing researchers to assess the robustness of their estimates to omitted confounders. In a sensitivity analysis, a researcher introduces a parameter describing the amount of unobserved confounding present and redoes the analysis under different parameter values, determining the set of values for which the results of the study will be reversed. The robustness of the study may then be evaluated by reasoning about the plausibility of these values.

In contrast to typical estimands, parameters in sensitivity analysis are inherently unindentifiable, because they are designed to describe an omitted variable. Thus, there exists a trade-off between how complex the sensitivity analysis is, and how informative the sensitivity analysis can

be. For example, Dahabreh et al. (2019) proposed a sensitivity analysis in which researchers can obtain both an adjusted point estimate and the associated uncertainty from omitting a confounder. However, the sensitivity analysis requires researchers to directly model the bias that arises from omitting a confounder. In contrast, Zhao et al. (2019) introduced a sensitivity analysis that only requires one parameter and allows researchers to estimate confidence intervals that account for the unobserved confounder. However, the resulting intervals are often extremely wide, making it difficult to reason about whether or not there is sensitivity from omitting a confounder.

In the following paper, we introduce a new sensitivity model known as the *variance-based sensitivity model*. The proposed sensitivity model constrains distributional differences in the weights that arise from omitting a confounder, and unlike many existing approaches (e.g., Imbens, 2003; Ding & VanderWeele, 2016; Bonvini et al., 2022) does not rely on additional assumptions on the outcome, confounder, or treatment assignment mechanism. We show that the proposed sensitivity analysis can be re-formulated as a bias maximization problem, with a constraint on a weighted average error. Using this re-formulation, we formalize the relationship between the variance-based sensitivity model and alternative sensitivity approaches, which rely on constraining a worst-case error. We demonstrate that by moving away from characterizing bias from the perspective of a worst-case error, variance-based sensitivity analysis can estimate more interpretable, informative, and stable bounds, while retaining the flexibility and generality of existing sensitivity analyses.

2. BACKGROUND

2.1. *Set-Up and Notation*

Consider an observational study with n individuals. Define $Z_i \in \{0,1\}$ as a treatment assignment variable, with $Z_i = 1$ when unit i is assigned to treatment, and 0 otherwise. $Y_i(1), Y_i(0) \in \mathbb{R}$ are potential outcomes, and $\widetilde{X}_i \in \widetilde{\mathcal{X}}$ is a vector of pre-treatment covariates. Let the tuple $(Y_i(1), Y_i(0), \widetilde{X}_i, Z_i)$ for all $i \in \{1, \ldots, n\}$ be independently and identically distributed from an arbitrary joint distribution.

Under the standard SUTVA assumption (i.e., no interference, with treatments identically administered across all units (Rubin, 1980)), which we assume throughout, observed outcomes Y can be written as $Y := Y(1) \cdot Z + Y(0) \cdot (1 - Z)$. For clarity, we focus throughout the paper on estimating the average treatment effect for the treated (ATT):

$$\tau := \mathbb{E}\left\{Y(1) - Y(0) \mid Z = 1\right\}.$$

However, we note the proposed methodology can be extended for a variety of causal effects and alternative missing data settings (see Appendix A.1).

Because treatments are not randomly assigned in an observational study, additional assumptions are needed to estimate the ATT consistently.

Assumption 1 (Conditional Ignorability of Treatment Assignment).

$$Y(1), Y(0) \perp\!\!\!\perp Z \mid \widetilde{X}$$

Assumption 2 (Overlap). For some $0 < \eta \leq 0.5$ and all $x \in \widetilde{\mathcal{X}}$, $\eta < \Pr(Z = 1 \mid \widetilde{X} = x) < 1 - \eta$.

Under Assumption 1, conditioning on pre-treatment covariates $\widetilde{X}$ suffices to remove all confounding between treatment assignment and potential outcomes (for a stronger alternative version of Assumption 1 that may enhance interpretation of sensitivity analyses, see Assumption 1

in Appendix A.2). Assumption 2 requires that no units have probability of treatment too close to zero or one. This is a standard assumption for weighted estimators; as discussed by D'Amour et al. (2021), treatment probabilities arbitrarily close to zero or one cause problems for estimation in the absence of assumptions about the outcome model. In practice, Assumption 2 requires careful checking. When extreme weights do appear present, it may be possible to stabilize or adjust them to render Assumption 2 plausible. For more discussion, see Crump et al. (2009); Fogarty et al. (2016); D'Amour et al. (2021).

A common approach to estimating causal effects is using weighted estimators. Weighted estimators adjust for distributional differences in the pre-treatment covariates $\widetilde{X}$ across the treatment and control groups. For weights $\hat{w}_1, \ldots, \hat{w}_n$, constructed from the observed sample using observed $\widetilde{X}_1, \ldots, \widetilde{X}_n$, the estimator is:

$$\hat{\tau}(\hat{w}_1, \ldots, \hat{w}_n) = \frac{1}{\sum_{i=1}^n Z_i} \sum_{i=1}^n Z_i Y_i - \frac{\sum_{i=1}^n (1 - Z_i) Y_i \hat{w}_i}{\sum_{i=1}^n (1 - Z_i) \hat{w}_i}. \tag{1}$$

When the $\hat{w}_i$s are consistent estimators of the population inverse propensity weights w_i, they provide consistent estimation of the treatment effect under Assumptions 1-2. When Assumption 1 does not hold, the resulting weighted estimate may not converge to the ATT. We propose a sensitivity model to characterize the error induced when it fails. Let $\widetilde{X} = \{X, U\}$ be the minimal separating set, such that both $X \in \mathcal{X}$ and $U \in \mathcal{U}$ are necessary for Assumption 1 to hold, but only values of X are observed and used for estimating the weights (Egami & Hartman, 2019; Bareinboim & Pearl, 2012).

We define

$$w(X_i) = \frac{\Pr(Z_i = 0)}{\Pr(Z_i = 1)} \frac{\Pr(Z_i = 1 \mid X_i)}{\Pr(Z_i = 0 \mid X_i)}, \quad w^*(X_i, U_i) = \frac{\Pr(Z_i = 0)}{\Pr(Z_i = 1)} \frac{\Pr(Z_i = 1 \mid X_i, U_i)}{\Pr(Z_i = 0 \mid X_i, U_i)}.$$

For simplicity, we will often use w_i and w_i^* in place of $w(X_i)$ and $w^*(X_i, U_i)$. We refer to the weights w_i as the *observable* weights — although they are not directly observed they are estimable using observed data — and the weights w_i^* as the *ideal* weights, since they guarantee consistent estimation of the true ATT. Since the (Z_i, X_i, U_i) are independently and identically distributed draws from an infinite superpopulation, $w(X_i)$ and $w^*(X_i, U_i)$ are too, and we often drop the i subscript when discussing an arbitrary draw from this distribution. Finally, we define a population-level estimator $\tau(w) := \mathbb{E}(Y \mid Z = 1) - \mathbb{E}(wY \mid Z = 0)/\mathbb{E}(w \mid Z = 0)$.

Since our sensitivity analysis is based on variances of weights, it is not equipped to address situations in which either the observable or ideal weights are all identical (with zero variance). We will assume this is not the case.

In practice estimates $\hat{w}_i$ for w_i must be constructed from the observed sample. A standard approach is to use a logistic regression to estimate the probability of treatment assignment; recent literature has also introduced semiparametric alternatives in the form of balancing weights for weight estimation, with better finite-sample guarantees (Ben-Michael et al., 2021). Our sensitivity analysis is parameterized using differences between the observable weights and the ideal weights, which are both population quantities, and is not specific to a particular method for estimating weights, but some conditions on the quality of estimation are needed for the inferential guarantees of Section 3.3. We choose to focus on the common logistic regression setting in that section only, while also discussing alternative assumptions for balancing weights.

Throughout the paper, we will use the word 'bias' to refer to the quantity $\tau(w) - \tau(w^*)$. Since stabilized weighted estimators are subject to mild finite sample bias, this quantity is not identical to the difference between $\mathbb{E}[\hat{\tau}(\hat{w})]$ and the population ATT, but they are asymptotically

equivalent and our usage is consistent with alternative sensitivity model approaches (e.g., Tan, 2006; Cinelli & Hazlett, 2020; Zhao et al., 2019; Dorn & Guo, 2023; Zhang & Zhao, 2024). See Appendix B.1 for extended discussion.

2.2. *Related Literature*

A popular approach for assessing the robustness of weighted estimates to omitted confounders uses the marginal sensitivity model (Tan, 2006), in which researchers posit a bound, Λ, on the individual-level multiplicative error in the population weights w:

$$\Lambda^{-1} \leq \frac{w^*(x,u)}{w(x)} \leq \Lambda, \quad \text{for all } x \in \mathcal{X},\ u \in \mathcal{U}. \tag{2}$$

where $\Lambda \geq 1$. Λ represents the largest possible error that can arise from omitting a confounder. Researchers can bound the maximum and minimum bias that arises under a fixed Λ, and use a percentile bootstrap to estimate valid confidence intervals (Aronow & Lee, 2013; Miratrix et al., 2018; Zhao et al., 2019).

In practice, the true Λ is unknown, so to conduct the sensitivity analysis, researchers run it with increasing values of Λ until the estimated confidence intervals contain zero. The minimum Λ value for which the estimated intervals cross zero is denoted as Λ^*. If Λ^* is close to one, even a small amount of error from omitting a confounder could explain a nominally significant effect. On the other hand, if Λ^* is much larger than one, significance of estimated effects is only sensitive to very strong unmeasured confounders.

While the marginal sensitivity model is simple to describe, it often leads to extremely wide intervals in practice, even under a mild degree of confounding. Dorn & Guo (2023) and Nie et al. (2021) propose methods to tighten these intervals, but require additional constraints or parametric outcome models. Furthermore, since the sensitivity parameter in the marginal sensitivity model depends on the worst-case individual impact of omitting a confounder, it can be difficult to reason about in settings where the unobserved confounder may occasionally take on extreme, outlying values. It is natural to ask whether a different approach could lead to narrower intervals under similar assumptions, and provide a more stable and interpretable approach to sensitivity analysis.

We now propose a new sensitivity model, the *variance-based sensitivity model*, which bounds the variance in the ideal weights w^* not explained by the observable weights w. Unlike related frameworks proposed by Hong et al. (2021) and Shen et al. (2011), the variance-based sensitivity model uses a standardized and bounded parameterization of confounding strength, which can help improve transparency and interpretability for applied researchers. Additionally, the aforementioned sensitivity analyses do not engage with how potential confounders may affect inference, focusing solely on movements in the point estimate. In contrast, the variance-based sensitivity model provides a method for estimating valid asymptotic confidence intervals under a fixed level of confounding.

We make a second important contribution by formalizing the connection between variance-based sensitivity analysis and alternative sensitivity approaches. In particular, we demonstrate that the variance-based sensitivity model can be viewed as an optimization under a constrained weighted L_2 norm on the individual-level multiplicative error, in contrast to the marginal sensitivity model's constraint on an L_∞ norm. Moving away from a worst-case error parameterization of the error allows researchers to obtain more informative and stable bounds under the variance-based sensitivity model. The benefits we find for constraining a weighted L_2 norm instead of a worst-case error are conceptually similar to the advantages highlighted in Kallus & Zhou (2018), in which authors consider a constraint on the L_1 norm, and Zhang & Zhao (2024) which introduces an L_2 norm. The latter approach may be considered an alternate type of variance-based

sensitivity analysis, constraining the variance of a different function of the observable and ideal weights than we do. However, our proposed sensitivity model has the additional benefit of a closed-form bias bound, and an interpretable sensitivity parameter in the form of an R^2 value.

2.3. *Running Example: NHANES*

Throughout the paper, we perform a re-analysis of a study presented in Zhao et al. (2018) (as well as Zhao et al., 2019 and Soriano et al., 2023), analyzing the effects of fish consumption on blood mercury levels. More specifically, we use data from the 2013-2014 National Health and Nutrition Examination Survey (NHANES).

Following the original study, we define the outcome of interest as the total blood mercury (in $\log_2$), measured in micrograms per liter. As such, an estimated treated-control outcome difference of one implies that a treated person's total blood mercury is twice that of an individual in control's total blood mercury. The treatment is defined by whether or not individuals consumed more than 12 servings of fish or shellfish in the preceding month. There are 234 total treated units and 873 control units. To account for the non-random treatment assignment, we use the available demographic data for the individuals in the survey, which include variables like gender, age, income, race, education, and smoking history to estimate propensity score weights using a logistic regression. Table 1 reports the raw outcome difference and a weighted estimate based on inverse propensity weights estimated by logistic regression.

Table 1. *Estimated Impact of Fish Consumption*

	Unweighted (DiM)	IPW
Estimated Effect (ATT)	2.37 (0.10)	2.14 (0.12)

**Standard errors reported in parentheses.*

Accounting for the log scale, our estimate suggests that on average, a treated individual who consumes more fish will have around four times the total blood mercury of a control individual.

3. The Variance-Based Sensitivity Model

3.1. *Defining a New Sensitivity Model*

We now introduce the variance-based sensitivity model. Instead of constraining the worst-case, individual-level multiplicative error across the weights, we constrain the variation in the ideal weights w^* not explained by the weights w.

Definition 1 (Variance-Based Sensitivity Model). *Let R^2 be the residual variation in the true weights w^*, not explained by w:*

$$R^2 := 1 - \frac{var\{w(X) \mid Z = 0\}}{var\{w^*(X, U) \mid Z = 0\}}$$

Then, for a fixed $R^2 \in [0, 1)$, we define the variance-based sensitivity model $\sigma(R^2)$:

$$\sigma(R^2) \equiv \left\{ w^* : 1 \leq \frac{var\{w^*(X, U) \mid Z = 0\}}{var\{w(X) \mid Z = 0\}} \leq \frac{1}{1 - R^2} \right\}.$$

The variance-based sensitivity model constrains how different the true weights w^* can be from the observable weights w. This implicitly restricts the residual imbalance in the omitted variable. More formally, we can decompose the true weight w^* into two components: (1) the weight w, and (2) the residual imbalance in U:

$$w^* = \underbrace{\frac{\Pr(Z=0)}{\Pr(Z=1)}\frac{\Pr(Z=1 \mid X)}{1-\Pr(Z=1 \mid X)}}_{(1)} \cdot \underbrace{\frac{\Pr(U \mid X, Z=1)}{\Pr(U \mid X, Z=0)}}_{(2)}, \tag{3}$$

where the imbalance term is a ratio of the conditional probability density function of the omitted variable across the treatment and control groups. The distributional difference between the weights w and the ideal weights w^* will be driven by the imbalance term. If imbalance in U is large, then accounting for the omitted variable results in very different values for w^* and w. Alternatively, if U is relatively balanced, then w^* will be similar to w.

The distributional difference between w and w^* is parameterized by an R^2 value. The R^2 parameter represents the residual variation in the true weights, not explained by the estimated weights. Using Equation (3), we can alternatively interpret the R^2 value as a measure of imbalance in the omitted confounder across the treatment and control group. Because the projection of w^* into X recovers w (i.e., $\mathbb{E}(w^*(X,U) \mid X, Z=0) = w(X)$), the variance of w^* can be decomposed linearly as $\text{var}(w^* \mid Z=0) = \text{var}(w \mid Z=0) + \text{var}(w^* - w \mid Z=0)$ (see Huang, 2024 and Chernozhukov et al., 2022 for more discussion). Thus, the R^2 value will be naturally bounded on the unit interval. $R^2 = 0$ implies that there is *no* imbalance in the omitted confounder between the treatment and control groups. As such, there will be no bias from omitting such a variable. As $R^2 \to 1$, this implies that initial imbalance in observed covariates X is negligible compared to imbalance in the omitted confounder U.

In Section 4, we show that specifying an R^2 value is equivalent to constraining a weighted L_2 norm of the errors w^*/w, in contrast with the marginal sensitivity model, which constrains an L_∞ norm.

3.2. *Constructing Bias Bounds*

We introduce a closed-form representation for a bound on bias over weights in $\sigma(R^2)$. The bias bound is a function of three different components: (1) a correlation bound, which represents the maximum correlation the imbalance in the omitted confounder can have with the outcome of interest; (2) the imbalance (represented by the R^2); and (3) a scaling factor. Theorem 1 formalizes the bound, followed by further discussion of each component.

THEOREM 1 (BIAS BOUND). *Define $\tilde{w} \in \sigma(R^2)$ as a set of possible weights that satisfy Definition 1. Let Bias$\{\tau(w) \mid \tilde{w}\}$ represent the bias of $\tau(w)$, with respect to a weighted estimator using $\tilde{w}$ (i.e., $\tau(w) - \tau(\tilde{w})$). Then, for a fixed $R^2 \in [0,1)$, the maximum bias under $\sigma(R^2)$ (denoted as $\max_{\tilde{w}\in\sigma(R^2)}$ Bias$\{\tau(w) \mid \tilde{w}\}$) can be written as*

$$\max_{\tilde{w}\in\sigma(R^2)} \textit{Bias}\{\tau(w) \mid \tilde{w}\}$$

$$= \underbrace{\sqrt{1 - \textit{cor}(w, Y \mid Z=0)^2}}_{(a)} \sqrt{\underbrace{\frac{R^2}{1-R^2}}_{(b)} \cdot \underbrace{\textit{var}(Y|Z=0) \cdot \textit{var}(w|Z=0)}_{(c)}}, \tag{4}$$

with a bound for the minimum bias given as the negative of Equation (4)*.*

The first component in Theorem 1 (i.e., Equation (4)-(a)) is a correlation bound. The correlation bound constrains how correlated the residual imbalance in the omitted confounder is to the outcome. The bound is a function of $1 - \text{cor}(w, Y \mid Z=0)^2$. As such, if w is highly correlated with the outcomes, the degree to which the residual imbalance in the omitted confounder

can be correlated to the outcome is limited, as is the overall bias. In contrast, when w is relatively uncorrelated with the outcome then the possible correlation between omitted-confounder imbalance and outcome has a much larger range, possibly including one. Dorn & Guo (2023) demonstrated a similar pattern for the marginal sensitivity model, in which bias is maximized when the imbalance from the omitted confounder is maximally correlated with the outcome.

The second component is the residual imbalance in the omitted confounder, which is a function of the R^2 parameter (Equation (4)-(b)). Unlike the correlation bound, for which overall impact is bounded at one, the residual imbalance grows without bound as R^2 approaches one. Cinelli & Hazlett (2020) discuss a similar asymmetry in a different sensitivity model.

Finally, the scaling factor (represented by Equation (4)-(c)) comprises of the variance of the outcomes across the control units and the variance of the estimated weights. These terms represent the overall heterogeneity present in the analysis; greater heterogeneity, if related to selection into treatment, makes it more difficult to recover the true estimated effect and increases the bias bound. The scaling factor is a function of the observed data, and is not related to the omitted confounder. However, a large scaling factor can amplify bias due to an omitted confounder.

Theorem 1 provides a simple and interpretable bias bound for a given R^2 value. However, the bound is not sharp by construction and may overestimate the maximum feasible bias under the variance-based sensitivity model. In brief, the correlation bound component of our bias bounds is loose in general because it assumes perfect correlation between outcomes and errors in weights, although it may not be possible to achieve perfect correlation over the class of ideal weights w^* that balance all functions of observed covariates across groups. Sharp bounds may be characterized as solutions to a variational optimization problem that does not assume perfect correlation with outcomes. In Appendix A.6, we describe this optimization problem and show that it admits a closed-form solution in certain population parameters. Unfortunately, these parameters are much harder to estimate and interpret than the components in Theorem 1. Accordingly, in what follows we focus primarily on the bound in (4), which provides richer insights about the drivers of potential confounding bias and is straightforward to estimate in practice.

3.3. *Constructing Confidence Intervals*

To construct valid asymptotic confidence intervals under the variance-based sensitivity model we follow a percentile bootstrap approach based on Zhao et al. (2019). Our approach is distinct from those in the partial identification literature that require known asymptotic distributions of the boundaries of the partially identified region (Imbens & Manski, 2004; Aronow & Lee, 2013). As in Zhao et al. (2019), it is difficult to characterize these distributions analytically in our sensitivity framework. Instead, the proposed bootstrap approach allows researchers to account for sampling uncertainty without explicitly characterizing the asymptotic distributions of the boundary estimates.

In the interest of clarity and developing a connection with the work of Zhao et al. (2019) and Dorn & Guo (2023), who focus on the logistic case, we prove the validity of the percentile bootstrap approach in Theorem 2 for the case where the observable weights $w(X)$ obey a logistic model in X and are estimated via logistic regression. However, we note that if the data-generating process for the observable weights is not logistic, we can still understand our $w(X)$ as a parametric approximation to the observable weights, and the sensitivity analysis may be viewed as addressing some combination of unobserved confounding and misspecification as in Zhao et al. (2019). In addition, a novel, more technically-involved argument by Soriano et al. (2023) establishes a result analogous to Theorem 2 for the marginal sensitivity analysis under general balancing weights, and we believe it should extend to variance-based sensitivity analysis as well. Informally, this argument requires primarily that the researcher has access to a well-specified

estimator $\widehat{w}(X)$ for $w(X)$, in the sense that estimated weights constructed for a subject with any given covariate value x converge pointwise to $w(X)$ and that bias of $\hat{\tau}(\widehat{w}_1, \ldots, \widehat{w}_n)$ for $\hat{\tau}(w)$ is of order $\sqrt{n}$ (for more detail on when balancing weights obey these conditions see Wager & Athey, 2018; Wang & Zubizarreta, 2020; Ben-Michael et al., 2021).

For a fixed R^2 and a given set of weights $\tilde{w} \in \sigma(R^2)$, we can construct an estimator $\hat{\tau}(\tilde{w}; \hat{w}_1, \ldots \hat{w}_n)$ that takes the original point estimate $\hat{\tau}(\hat{w}_1, \ldots \hat{w}_n)$ and subtracts a finite-sample bias term that is a function of $\tilde{w}$ (see Appendix B for more details).

Using results from Zhao et al. (2019), for any $\tilde{w} \in \sigma(R^2)$, we construct asymptotically valid confidence intervals for the large-sample limit $\tau(\tilde{w})$ using a percentile bootstrap for $\hat{\tau}(\tilde{w}; \hat{w}_1, \ldots \hat{w}_n)$.

$$[L(\tilde{w}), U(\tilde{w})] = \left[Q_{\alpha/2}\left\{ \hat{\tau}^{(b)}(\tilde{w}; \hat{w}_1^{(b)}, \ldots \hat{w}_n^{(b)}) \right\}, Q_{1-\alpha/2}\left\{ \hat{\tau}^{(b)}(\tilde{w}; \hat{w}_1^{(b)}, \ldots \hat{w}_n^{(b)}) \right\} \right], \quad (5)$$

where $\hat{\tau}^{(b)}(\tilde{w}; \hat{w}_1^{(b)}, \ldots \hat{w}_n^{(b)})$ is the adjusted weighted estimator in bootstrap sample $b \in \{1, ..., B\}$, and $Q_\alpha(\cdot)$ denotes the α-th percentile in the bootstrap distribution. Theorem 2 shows that $[L(\tilde{w}), U(\tilde{w})]$ is an asymptotically valid (1-α) confidence interval for $\tau(\tilde{w})$:

THEOREM 2 (VALIDITY OF PERCENTILE BOOTSTRAP). *Suppose* $\Pr(Z = 1 \mid X)$ *is logistic in* X *and that* $\hat{w}_1, \ldots, \hat{w}_n$ *are plug-in estimates based on propensity scores estimated via logistic regression on all observed covariates* X*. Then under mild regularity conditions (see Assumption 2 in the Appendix), for every* $\tilde{w} \in \sigma(R^2)$*:*

$$\limsup_{n\to\infty} \Pr\{\tau(\tilde{w}) < L(\tilde{w})\} \leq \frac{\alpha}{2} \text{ and } \limsup_{n\to\infty} \Pr\{\tau(\tilde{w}) > U(\tilde{w})\} \leq \frac{\alpha}{2},$$

where $L(\tilde{w})$ *and* $U(\tilde{w})$ *are defined as the* $\alpha/2$ *and* $1 - \alpha/2$*-th quantiles of the bootstrapped estimates (i.e., Equation* (5)*).*

Theorem 2 provides a valid interval for any set of weights $\tilde{w}$, but for a given R^2 value, there exist infinitely many choices $\tilde{w} \in \sigma(R^2)$. As such, we apply the union method to construct a conservative $(1 - \alpha)\%$ confidence interval CI(α) valid for any $\tau(\tilde{w})$ with $\tilde{w} \in \sigma(R^2)$:

$$\left[Q_{\frac{\alpha}{2}} \left\{ \inf_{\tilde{w}\in\sigma(R^2)} \hat{\tau}^{(b)} \left(\tilde{w}; \hat{w}_1^{(b)}, \ldots \hat{w}_n^{(b)} \right) \right\}, \; Q_{1-\frac{\alpha}{2}} \left\{ \sup_{\tilde{w}\in\sigma(R^2)} \hat{\tau}^{(b)} \left(\tilde{w}; \hat{w}_1^{(b)}, \ldots \hat{w}_n^{(b)} \right) \right\} \right]. \quad (6)$$

To estimate the extrema of the bootstrapped point estimates inside the quantile functions, we add (subtract) a sample estimate of the population bias bound to obtain the maximum (minimum); for more details see equation (24) in Appendix B.2. As such, estimating valid confidence intervals amounts to a standard percentile bootstrap with the added tweak of subtracting (adding) an estimated bias bound from each bootstrap estimate before calculating the confidence limits.

3.4. *Illustrating the Sensitivity Analysis on NHANES*

To conduct sensitivity analysis, researchers estimate confidence intervals for increasing R^2 values until an estimated confidence interval just contains the null estimate; the corresponding R^2 is denoted as R^2_*. In the running example, we estimate $R^2_* = 0.52$. This implies that if an omitted confounder explaining 52% or more of the variation in the true weights, our estimated effect of fish consumption on blood mercury is no longer significantly different from the expected distribution under the null.

To assess the plausibility of an omitted confounder resulting in an R^2 value of 0.52, we extend formal benchmarking approaches to calibrate possible R^2 values against the strength of observed covariates (e.g., Huang, 2024; Hartman & Huang, 2024; Cinelli & Hazlett, 2020; Hong et al.,

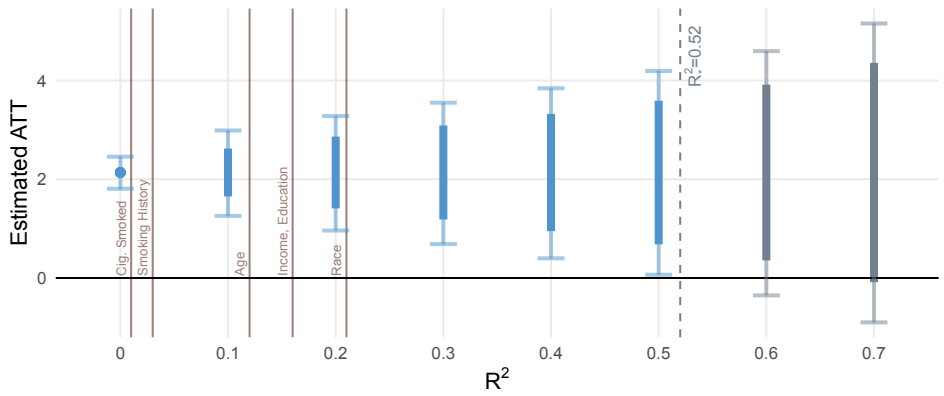

Fig. 1. The solid bars denote the point estimate bounds for a specified R^2 value. The lighter intervals represent the 95% confidence intervals. The benchmarked parameters $\hat{R}^2_{(j)}$ are provided in light brown.

2021; Carnegie et al., 2016; Hsu & Small, 2013). Briefly, we omit each observed covariate j in turn and estimate a corresponding $R^2_{(j)}$ value for each as a guide to the possible impact of omitting an unobserved confounder similar to this observed one, using the following formula (where $\widehat{\text{var}}$ indicates sample variance and $\widehat{w}^{-(j)}$ indicates weights estimated with covariate j omitted):

$$\hat{R}^2_{(j)} = \frac{\hat{R}^2_{-(j)}}{1+\hat{R}^2_{-(j)}}, \quad \text{where } \hat{R}^2_{-(j)} := 1 - \frac{\widehat{\text{var}}(\hat{w}^{-(j)} \mid A=1)}{\widehat{\text{var}}(\hat{w} \mid A=1)}. \tag{7}$$

See Appendix A.4 for a full derivation and additional discussion. Omitting a confounder with equivalent imbalance to covariates like *race* ($\hat{R}^2_{(j)}$ of 0.21), *education* ($\hat{R}^2_{(j)}$ of 0.16), or *income* ($\hat{R}^2_{(j)}$ of 0.16) results in the largest benchmark R^2 values.

We caution that benchmarking should never be used to set threshold values for whether a result is considered 'robust' enough. However, benchmarking allows researchers to consider how strong (or weak) an omitted confounder must be, relative to an observed covariate. In the NHANES example, an omitted confounder would have to explain more than twice as much variance in the true weights as the strongest observed covariate, *race*, is estimated to in order for the R^2 value to be equal the threshold value. While mathematically possible, the plausibility of a confounder resulting in the threshold $R^2_* = 0.52$ value is low. Figure 1 illustrates the sensitivity analysis, as well as benchmarking results.

4. Relationship to the Marginal L_∞ Sensitivity Model

4.1. *Sensitivity Models as an Optimization Problem*

We now examine the relationship between the variance-based sensitivity model and the marginal sensitivity model. We first show that both sensitivity models can be written as norm-constrained optimization problems. More formally, let $\lambda = w^*(X,U)/w(X)$ be the multiplicative error in the weights when covariates are equal to (X,U). The variance-based sensitivity model can then be formulated as a bias maximization problem, given a fixed constraint on a weighted L_2 norm over the distribution of λ.

THEOREM 3 (WEIGHTED L_2 NORM CONSTRAINT). *Define the $L_{2,w}$ norm as follows:*

$$||\lambda||_{2,w}^2 := \begin{cases} \mathbb{E}\left\{\lambda^2 \nu(w) \mid Z = 0\right\} & \textit{if var}(w \mid Z = 0) > 0, \\ \infty & \textit{else} \end{cases},$$

where $\nu(w) := w^2/\mathbb{E}(w^2 \mid Z = 0)$. Then, the variance-based sensitivity model can be written as a norm-constrained optimization problem:

$$\max_{\tilde{w} \in \sigma(R^2)} \textit{Bias}\{\tau(w) \mid \tilde{w}\} \iff \begin{cases} \max\limits_{\tilde{w}} \textit{Bias}\{\tau(w) \mid \tilde{w}\} \\ \textit{s.t. } ||\lambda||_{2,w} \leq \sqrt{\dfrac{k}{1 - R^2}}, \end{cases}$$

where $k := 1 - R^2/\mathbb{E}(w^2 \mid Z = 0)$. See Appendix B for proof and details.

Theorem 3 is especially instructive in concert with the following result from Zhao et al. (2019) showing that the marginal sensitivity model is equivalent to a maximization problem with an L_∞ constraint on λ. Letting $\varepsilon(\Lambda)$ represent the family of all possible values of w^* allowed by constraint (2),

$$\max_{\tilde{w} \in \varepsilon(\Lambda)} \text{Bias}\{\tau(w) \mid \tilde{w}\} \iff \begin{cases} \max\limits_{\tilde{w}} \ \text{Bias}\{\tau(w) \mid \tilde{w}\} \\ \text{s.t. } \Lambda^{-1} \leq ||\lambda||_\infty \leq \Lambda. \end{cases}$$

In short, the key difference between the models is the norms they use to constrain deviation between weights w which marginalize over unobserved U and the ideal weights w^* and the w. In this sense, both sensitivity analyses are "marginal". To minimize confusion, we refer to the marginal sensitivity model as the marginal L_∞ sensitivity model. The distinct constrained-norm representations across the two models provide insight into the benefits expected from the variance-based sensitivity model. Because the marginal L_∞ sensitivity model optimizes over the set of weights defined by a worst-case error, the estimated bounds on the bias always correspond to cases in which *all* units are exposed to this worst-case error. However, in settings when one or two subjects are subject to much larger levels of confounding than others, this can result in an overly pessimistic view of the potential bias (Fogarty & Hasegawa, 2019; Zhao et al., 2019). In contrast, the variance-based sensitivity model is constraining an average weighted error, and thus allows a small number of weights to be exposed to large amounts of error, even at moderate levels of overall confounding.

In practice, researchers care not only about informative bounds, but also about the *utility* of the sensitivity models. In the following two subsections, we examine two cases which showcase that moving away from a worst-case error allows the variance-based sensitivity model to improve upon the marginal L_∞ sensitivity model in stability and interpretability.

4.2. *Infinite Worst-Case Error in Asymptotic Settings*

Since the marginal L_∞ sensitivity model focuses on the maximum individual error, it will be generally invalid for settings in which unobserved confounders drive rare large errors.

Example 1 below illustrates a simple case involving a normally-distributed unobserved confounder and a logit model for treatment (see Jin et al., 2022 for a similar example).

Example 1 (Behavior of Λ for a Logit Model). Assume the true weights follow a logit model in both X and U, but the U is unobserved. The estimable and ideal weights take on the following forms:

$$w = \exp(\gamma^\top X) \qquad w^* = \exp(\gamma^{*\top} X + \beta U)$$

Then let $\hat{\Lambda}$ be the maximum error across our observed sample (i.e., $\hat{\Lambda} := \max_{1\leq i\leq n}\{w^*/w, w/w^*\}$). Assume $\left[X, U\right] \overset{iid}{\sim} MVN(0, I)$. Then $\mathbb{E}(\hat{\Lambda}) \to \infty$ as $n \to \infty$:

$$\lim_{n\to\infty} \frac{\mathbb{E}(\hat{\Lambda})}{\exp\left\{\sqrt{2\nu^2 \log(n)}\right\}} \geq 1,$$

where $\nu^2 = (\gamma^* - \gamma)^\top(\gamma^* - \gamma) + \beta^2$, and the results follow immediately from Wainwright (2019), § 2, pg. 53.

The basic issue in Example 1 is that the unbounded support of the unobserved confounder leads to unbounded maximum individual error in the population; although in any finite sample the maximum individual error is bounded, the marginal L_∞ sensitivity model cannot hold for any finite Λ in the population. This divergence occurs, *regardless* of the confounding strength of the omitted confounder (represented by β).

In addition to concerns about the model's validity, this characteristic of the marginal L_∞ sensitivity model makes it difficult to interpret the sensitivity parameter. With the aid of benchmarking researchers can understand the worst-case errors arising from omitting observed covariates, but reasoning about whether it is *plausible* for such an error to arise from an omitted variable amounts to reasoning about whether it is plausible for potential outliers to occur. As the sample size increases, researchers must account for the increasing chance of an outlier.

In contrast, we can derive the R^2 value for the variance-based sensitivity model under the same setting as Example 1. It is a function of the relative strength of the omitted confounder, and does not depend on the sample size.

Example 2 (Behavior of R^2 for a Logit Model). Consider the same setting as Example 1. Then, the R^2 value can be written as follows:

$$R^2 = 1 - \frac{\exp(\gamma^\top\gamma) - 1}{\exp(\gamma^{*\top}\gamma^* + \beta^2) - 1} \cdot \frac{\exp(\gamma^\top\gamma)}{\exp(\gamma^{*\top}\gamma^* + \beta^2)}.$$

Example 1 provides one setting in which Λ will be infinitely large, regardless of the confounding strength of the omitted variable. More generally, the following corollary shows that under any setting when the error from omitting a confounder can take on values that are arbitrarily small or large, the marginal L_∞ sensitivity model will be invalid, while the variance-based sensitivity model will remain valid. Furthermore, when the outcomes are also unbounded, the bounds estimated under the marginal L_∞ sensitivity model will be infinitely wide. Thus, in sufficiently large samples, the variance-based sensitivity model necessarily produces narrower bounds.

COROLLARY 1. *Consider the set of confounders, in which for all $\delta > 0$, $\Pr(w^*/w < \delta) > 0$, or $\Pr(w^*/w > \delta) > 0$. Then, the marginal L_∞ sensitivity model will no longer be valid. Furthermore, if the outcomes are unbounded, the size of the bounds under the marginal L_∞ sensitivity model will diverge in probability to infinity. Therefore, for sufficiently large n, the variance-based sensitivity model will produce narrower bounds.*

Importantly, the variance-based sensitivity model can be flexibly applied, regardless of the underlying distribution of the omitted confounder, and the observed data generating process. Furthermore, the interpretation of the sensitivity parameter R^2 remains intrinsically tied to reasoning about the confounding strength of the omitted variable, in contrast to reasoning about potential outliers.

4.3. *Limited Overlap in Finite-Samples*

Section 4.2 demonstrated cases in which the marginal L_∞ sensitivity model leads to unappealingly large intervals; however, paradoxically, in slightly different settings, the marginal L_infty sensitivity model can lead to intervals that are unappealingly short. While the intervals constructed by Zhao et al. (2019) under the marginal L_∞ sensitivity model are asymptotically valid, in smaller datasets they are also susceptible to a particularly egregious form of finite-sample bias. This bias, which arises when empirical outcome distributions within treatment and control groups do not overlap sufficiently, leads to substantial undercoverage. In these settings, the variance-based sensitivity model will tend to return wider intervals, but maintain nominal coverage.

The key to this phenomenon is a property of the marginal L_∞ sensitivity model, referred to as *sample boundedness*. Sample boundedness implies that even at infinitely large Λ values, the worst-case bounds under the marginal L_∞ sensitivity model approach, but cannot exceed, the range of the observed control outcomes. Sample boundedness follows automatically from the form of estimator (1), which relies on a convex combination of observed control outcomes to impute an expected potential outcome; the convex combination must lie within the range of observed control outcomes.

In contrast, the variance-based sensitivity model is not inherently sample bounded. In settings with relatively large amounts of confounding, the marginal L_∞ sensitivity model will have narrower intervals than the variance-based sensitivity model, since as R^2 increases towards one, the estimated bounds under the variance-based sensitivity model will be adequately wide. However, sample boundedness may prohibit the construction of valid confidence intervals in the absence of a key implicit assumption on the distribution of the unobserved potential outcomes.

Example 3 (Misleading Optimism from Sample Boundedness). Consider the following population of 4 units, with the following potential outcomes, treatment assignment, and the estimated probability of treatments for each unit:

i	$Y_i(0)$	$Y_i(1)$	$\hat{P}(Z_i = 1)$	Z_i
1	-10	-10	0.1	0
2	5	5	0.2	0
3	10	10	0.9	1
4	20	20	0.95	1

The true ATT is zero, but the estimated ATT is equal to 14.6, so substantial confounding is present. However, since the sample bounds for the ATT are the interval $[10, 25]$, no value of Λ can produce an estimated interval (under the marginal L_∞ sensitivity model) containing zero, erroneously suggesting the presence of a true effect highly robust to substantial confounding.

While this example is somewhat contrived, it highlights the problems with sample boundedness if the potential outcome ranges in the two groups have limited overlap, which may occur when potential outcomes are strongly correlated with the probability of treatment. For a formal characterization of this outcome overlap condition, see Appendix A.7.

When there exists limited outcome overlap, estimated intervals from the marginal L_∞ sensitivity model may be misleadingly optimistic, especially for dramatic levels of potential confounding. In contrast, intervals constructed under the variance-based sensitivity model, which are not sample bounded, are not affected. Figure 2 illustrates the behavior and coverage rates of both sets of sensitivity models under varying amounts of outcome overlap and sample sizes in an empirical example, described in greater detail in Appendix A.7.

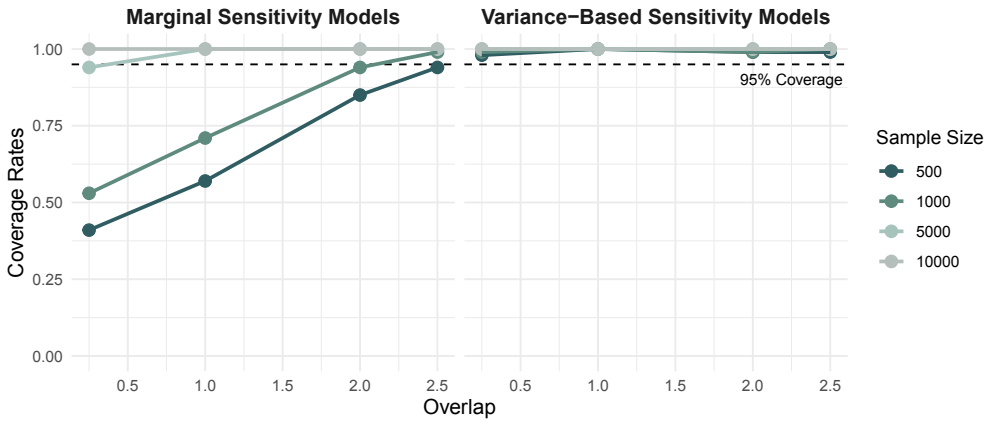

Fig. 2. Coverage rates for the MSM and the VBM, assuming an oracle bias setting when researchers have full knowledge of the true underlying sensitivity parameter.

We note that sample boundedness is not necessarily a negative feature in the context of *estimation*. The bias-variance tradeoff of using a stabilized weighted estimator has been extensively studied (e.g., Robins et al., 2007). However, in the context of a sensitivity analysis, in which we are explicitly interested in examining the potential bias that can arise under varying levels of confounding, the relative impact of the finite-sample bias can be greatly exacerbated. As such, imposing sample boundedness can lead to misleading conclusions, and potential issues with outcome overlap should be considered carefully when interpreting results.

4.4. *Comparing the bounds: Illustration on NHANES*

While constrained-norm representations provide intuition for why the variance-based sensitivity model may obtain narrower bounds than the marginal L_∞ sensitivity model, in practice it is difficult to directly compare the bounds estimated under the two models. The two approaches use two different parameters, each of which indexes a different class of possible true data distributions; in the formal language of Section 4.1, the set of weights $\sigma(R^2)$ is neither contained by nor fully contains the set of weights $\varepsilon(\Lambda)$ in general. This makes it hard to find an appropriate mapping between a posited value for R^2 and a comparable value for Λ.

However, for any given true distribution of weights w^* we may find the smallest set $\varepsilon(\Lambda)$ that contains it (with parameter Λ_0) and similarly we may find the smallest set $\sigma(R^2)$ that contains it (with parameter R^2_0). Comparing the width of the confidence intervals for the variance-based model under R^2_0 and the marginal L_∞ model under Λ_0 is meaningful because it shows how tightly each model can bound bias under a common distribution. Intuitively, we anticipate that if the worst-case error Λ_0 is much larger than the average weighted error, R^2_0, the variance-based sensitivity model will result in narrower bounds. If the difference in the worst-case error and average weighted error is not very large, then there will not be much improvement in the estimated bounds from using the variance-based sensitivity model.

In practice, researchers do not have access to the true weight w^*. However, benchmarking helps to make the comparison between models more concrete: for each covariate used in the benchmarking procedure, a benchmarked R^2 value (given by $R^2_{(j)}$ in (7)) and a benchmarked Λ value (as detailed in Soriano et al. (2023)) can each be produced, reflecting the R^2_0 and Λ_0 values necessary to capture the impact of this variable had it remained unobserved. Since empirical

researchers already frequently use benchmarking to interpret results of sensitivity analyses, the resulting comparison between sensitivity models is both natural and highly relevant if it can lead to shorter intervals in practice.

We now conduct benchmarking for the variance-based models and the marginal L_∞ sensitivity model in our running example, estimate the corresponding bounds and intervals under both approaches, and compare their widths. For the marginal L_∞ model, we estimate both the conservative, standard bounds introduced in Zhao et al. (2019), as well as the sharp bounds, obtained using quantile balancing (Dorn & Guo, 2023). Figure 3 visualizes the results. We see that for each of the covariates, omitting a confounder like any of the observed covariates would result in substantially wider bounds under the marginal L_∞ sensitivity model than the variance-based model. This is true, even when comparing the sharp bounds under marginal L_∞ sensitivity model with the conservative bounds under the variance-based model. This suggests that the improvements we observe from using the variance-based approach is present, even when accounting for improvements in sharpness.

The relative improvement is most apparent when looking at the benchmarking results for *education* and *race*. In particular, we see that while the average error from omitting a variable like *education* or *race* is relatively low, the maximum error is large. The marginal L_∞ sensitivity model, which assumes such a maximal error could occur in the unobserved confounder for all data points, thus produces much wider intervals than the variance-based model, which is much less responsive to individual outliers.

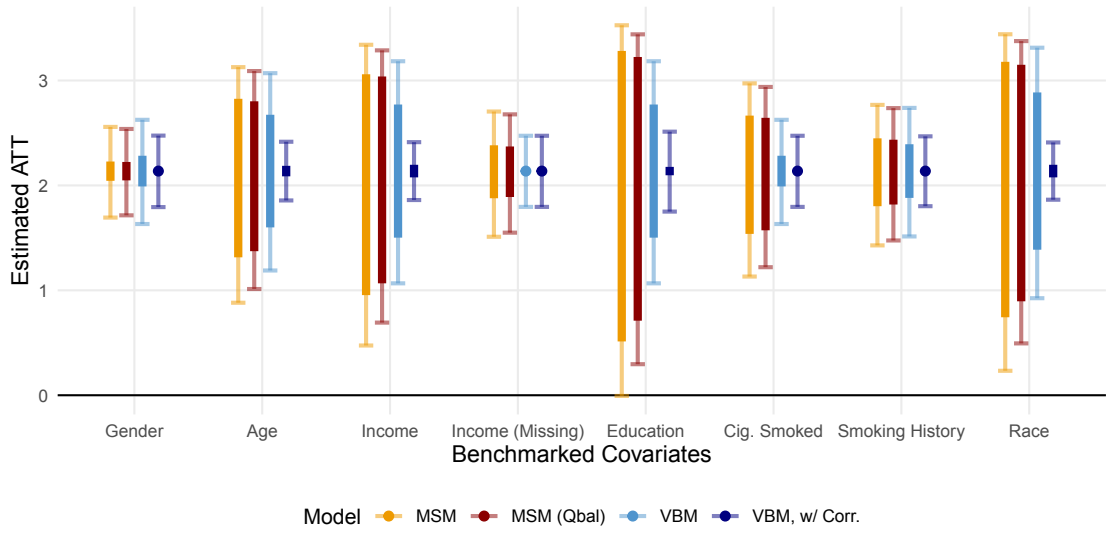

Fig. 3. Estimated intervals under MSM and VBM. From left to right: MSM with conservative bounds, MSM with sharp bounds, VBM with conservative bounds, and VBM with a less conservative correlation bound.

We also estimate intervals (and bounds) under the variance-based sensitivity model using a relaxed correlation bound. In particular, we choose the correlation bound by benchmarking an optional correlation parameter, giving the correlation between the outcome and the imbalance in an omitted confounder, to the observed correlation between the outcome and each observed covariate. (See Appendix A.5 for more details.) By accounting for the relationship between the confounder and the outcome, we are able to obtain much narrower intervals. In particular, we

see that even in cases where a potential omitted confounder is highly imbalanced (e.g., omitting a confounder like *age* results in a benchmarked R^2 value of 0.12, and a benchmarked Λ value of 2.2), the overall bias that occurs from omitting it may be relatively low if the imbalance is largely unrelated to the outcome. By considering this additional dimension of the bias–which can be easily done using the variance-based sensitivity model– researchers are able to better characterize the types of confounders that may lead to large amounts of bias and obtain a more holistic understanding of the sensitivity in their estimated effects.

5. CONCLUSION

We suggest several directions for future work. First, throughout the paper, we focused our discussion on comparing the variance-based sensitivity model to the marginal L_∞ sensitivity model, due to the relative popularity of the marginal L_∞ sensitivity model. However, there exist many alternative approaches to conducting sensitivity analysis by constraining distributional divergences (e.g., Jin et al., 2022; Bertsimas et al., 2022). The variance-based sensitivity model can be viewed as a special case of a distributionally-constrained sensitivity model. For example, Jin et al. (2022) propose constraining f-divergences for a general set of functions. The variance-based sensitivity model corresponds to the setting in which researchers are constraining a quadratic f (i.e., $f(x) = x^2$). Exploring the implications of these different constraints could lead to a broad unified sensitivity framework, helping contextualize a wider variety of different sensitivity methods with their own strengths and weaknesses.

Second, numerous extensions of the marginal L_∞ sensitivity model for alternative, more complex settings in causal inference have been proposed — e.g., Rosenman & Owen (2021) for experimental design, Bonvini et al. (2022) for time-varying treatment effects, and Kallus & Zhou (2018) for policy learning. An interesting line of future research could extend the variance-based sensitivity model for these settings, where we anticipate similar benefits and advantages to the ones highlighted in this paper. Additionally, while we focused on a choice between bounding a weighted average error and bounding a worst-case error, future work could incorporate both constraints in the same study. We anticipate that this would result in further narrowing of sensitivity bounds.

Third, it is natural to ask what factors under a researcher's control at the design stage may influence the degree of robustness to unmeasured bias exhibited under the variance-based sensitivity analysis. While the closed form for the bias bound already provides insights in this direction, developing a metric akin to design sensitivity for matched studies (Rosenbaum, 2004, 2010) would inform how to design weighting estimators for maximum robustness.

ACKNOWLEDGEMENTS

The authors would like to thank David Bruns-Smith, Angela Zhou, Erin Hartman, Kevin Guo, Alex Franks, Avi Feller, Peng Ding, Nicole Pashley, Cyrus Samii, and the three anonymous reviewers for the helpful comments and feedback. Furthermore, the authors would like to thank the Berkeley Casual & Friends Working Group, the Online Causal Inference Seminar, and PolMeth 2022 for the feedback. Part of this work was completed while Melody Huang was supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2146752. Samuel D. Pimentel is supported by the National Science Foundation under Grant No. 2142146. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

ARONOW, P. M. & LEE, D. K. (2013). Interval estimation of population means under unknown but bounded probabilities of sample selection. *Biometrika* **100**, 235–240.

BAREINBOIM, E. & PEARL, J. (2012). Controlling selection bias in causal inference. In *Artificial Intelligence and Statistics*. PMLR.

BEN-MICHAEL, E., FELLER, A., HIRSHBERG, D. A. & ZUBIZARRETA, J. R. (2021). The balancing act for causal inference. *arXiv preprint arXiv:2110.14831* .

BERTSIMAS, D., IMAI, K. & LI, M. L. (2022). Distributionally robust causal inference with observational data. *arXiv preprint arXiv:2210.08326* .

BONVINI, M., KENNEDY, E., VENTURA, V. & WASSERMAN, L. (2022). Sensitivity analysis for marginal structural models. *arXiv preprint arXiv:2210.04681* .

CARNEGIE, N. B., HARADA, M. & HILL, J. L. (2016). Assessing sensitivity to unmeasured confounding using a simulated potential confounder. *Journal of Research on Educational Effectiveness* **9**, 395–420.

CHERNOZHUKOV, V., CINELLI, C., NEWEY, W., SHARMA, A. & SYRGKANIS, V. (2022). Long story short: Omitted variable bias in causal machine learning. Tech. rep., National Bureau of Economic Research.

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

CRUMP, R. K., HOTZ, V. J., IMBENS, G. W. & MITNIK, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96**, 187–199.

DAHABREH, I. J., ROBINS, J. M., HANEUSE, S. J., SAEED, I., ROBERTSON, S. E., STUART, E. A. & HERNÁN, M. A. (2019). Sensitivity analysis using bias functions for studies extending inferences from a randomized trial to a target population. *arXiv preprint arXiv:1905.10684* .

DING, P. & VANDERWEELE, T. J. (2016). Sensitivity analysis without assumptions. *Epidemiology (Cambridge, Mass.)* **27**, 368.

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

D'AMOUR, A., DING, P., FELLER, A., LEI, L. & SEKHON, J. (2021). Overlap in observational studies with high-dimensional covariates. *Journal of Econometrics* **221**, 644–654.

EGAMI, N. & HARTMAN, E. (2019). Covariate selection for generalizing experimental results. *arXiv preprint arXiv:1909.02669* .

FOGARTY, C. B. & HASEGAWA, R. B. (2019). Extended sensitivity analysis for heterogeneous unmeasured confounding with an application to sibling studies of returns to education. *The Annals of Applied Statistics* **13**, 767–796.

FOGARTY, C. B., MIKKELSEN, M. E., GAIESKI, D. F. & SMALL, D. S. (2016). Discrete optimization for interpretable study populations and randomization inference in an observational study of severe sepsis mortality. *Journal of the American Statistical Association* **111**, 447–458.

HARTMAN, E. & HUANG, M. (2024). Sensitivity analysis for survey weights. *Political Analysis* **32**, 1–16.

HONG, G., YANG, F. & QIN, X. (2021). Did you conduct a sensitivity analysis? a new weighting-based approach for evaluations of the average treatment effect for the treated. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **184**, 227–254.

HSU, J. Y. & SMALL, D. S. (2013). Calibrating sensitivity analyses to observed covariates in observational studies. *Biometrics* **69**, 803–811.

HUANG, M. (2024). Sensitivity analysis in the generalization of experimental results. *Journal of the Royal Statistical Society: Series A (Statistics in Society) (Forthcoming)* .

IMBENS, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review* **93**, 126–132.

IMBENS, G. W. & MANSKI, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica* **72**, 1845–1857.

JIN, Y., REN, Z. & ZHOU, Z. (2022). Sensitivity analysis under the f-sensitivity models: Definition, estimation and inference. *arXiv preprint arXiv:2203.04373* .

KALLUS, N. & ZHOU, A. (2018). Confounding-robust policy improvement. *Advances in neural information processing systems* **31**.

MIRATRIX, L. W., WAGER, S. & ZUBIZARRETA, J. R. (2018). Shape-constrained partial identification of a population mean under unknown probabilities of sample selection. *Biometrika* **105**, 103–114.

NIE, X., IMBENS, G. & WAGER, S. (2021). Covariate balancing sensitivity analysis for extrapolating randomized trials across locations. *arXiv preprint arXiv:2112.04723* .

ROBINS, J., SUED, M., LEI-GOMEZ, Q. & ROTNITZKY, A. (2007). Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. *Statistical Science* **22**, 544–559.

ROSENBAUM, P. R. (2004). Design sensitivity in observational studies. *Biometrika* **91**, 153–164.

ROSENBAUM, P. R. (2010). Design sensitivity and efficiency in observational studies. *Journal of the American Statistical Association* **105**, 692–702.

ROSENMAN, E. T. & OWEN, A. B. (2021). Designing experiments informed by observational studies. *Journal of Causal Inference* **9**, 147–171.

RUBIN, D. B. (1980). Discussion of 'Randomization analysis of experimental data: The Fisher randomization test comment' by Basu. *Journal of the American Statistical Association* **75**, 591–593.

SHEN, C., LI, X., LI, L. & WERE, M. C. (2011). Sensitivity analysis for causal inference using inverse probability weighting. *Biometrical Journal* **53**, 822–837.

SORIANO, D., BEN-MICHAEL, E., BICKEL, P. J., FELLER, A. & PIMENTEL, S. D. (2023). Interpretable sensitivity analysis for balancing weights. *Journal of the Royal Statistical Society Series A: Statistics in Society* **186**, 707–721.

TAN, Z. (2006). A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association* **101**, 1619–1637.

WAGER, S. & ATHEY, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* **113**, 1228–1242.

WAINWRIGHT, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, vol. 48. Cambridge University Press.

WANG, Y. & ZUBIZARRETA, J. R. (2020). Minimal dispersion approximately balancing weights: asymptotic properties and practical considerations. *Biometrika* **107**, 93–105.

ZHANG, Y. & ZHAO, Q. (2024). Sharp bounds and semiparametric inference in l_∞-and l_2-sensitivity analysis for observational studies. *arXiv preprint arXiv:2211.04697* .

ZHAO, Q., SMALL, D. S. & BHATTACHARYA, B. B. (2019). Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **81**, 735–761.

ZHAO, Q., SMALL, D. S. & ROSENBAUM, P. R. (2018). Cross-screening in observational studies that test many hypotheses. *Journal of the American Statistical Association* **113**, 1070–1084.

Biometrika (2023), **XXX**, X, *pp*. 1–22
Printed in Great Britain

Supplementary Materials for "Variance-based sensitivity analysis for weighting estimators results in more informative bounds"

BY MELODY HUANG
Yale University, New Haven, Connecticut 06520, U.S.A.
melody.huang@yale.edu

AND SAMUEL D. PIMENTEL
University of California, Berkeley, Berkeley, California 94720, U.S.A.
spi@berkeley.edu

A. ADDITIONAL DISCUSSION

A.1. *Missingness*

In the main manuscript, the estimand of interest is the average treatment effect, across the treated. However, we note that the sensitivity framework introduced can be applied to more general settings, in which we consider missingness conditionally at random:

$$Y \perp\!\!\!\perp A \mid \mathcal{X}$$

This provides a very flexible framework to consider many settings of interest. Table 1 summarizes several settings of interest, along with the associated conditional ignorability assumption to be relaxed by sensitivity analysis.

Table 1. *Summary of different common missingness settings.*

Setting	Missingness Indicator	Ignorability Statement
Survey Response	R (Response)	$Y \perp\!\!\!\perp R \mid \mathcal{X}$
Internal Validity	Z (Treatment Assignment)	$Y(1), Y(0) \perp\!\!\!\perp Z \mid \mathcal{X}$
External Validity	S (Inclusion in Experimental Sample)	$Y(1) - Y(0) \perp\!\!\!\perp S \mid \mathcal{X}$

A.2. *Parametric Assumption of Conditional Ignorability*

In practice, when researchers estimate weights, they are implicitly assuming a parametric version of Assumption 1. Following Hartman et al. (2021), we formalize the parametric version of Assumption 1:

Assumption 1 (Linear ignorability in $\phi(X)$). There exists a feature mapping $\phi(\cdot)$ for X such that we can write the outcome Y as follows:

$$Y = \phi(X)^\top \beta + \delta$$

In addition, we can write $\Pr(Z = 1 \mid X)$ as follows for some function $g(\cdot) : \mathbb{R} \mapsto [0, 1]$:

$$Pr(Z = 1|X) = g\left\{\phi(X)^\top \theta + \eta\right\},$$

where $\delta \perp\!\!\!\perp \eta$.

Linear ignorability in $\phi(X)$ implies that the part of the outcome that is orthogonal to $\phi(X)$ is independent to the part of the treatment assignment process that is orthogonal to $\phi(X)$.

The distinction between the non-parametric version of conditional ignorability (i.e., Assumption 1) and the parametric version (i.e., Assumption 1) is consequential for classifying that types of violations that may lead to omitted variable bias. Under the non-parametric version of conditional ignorability, only variables that are fully unobserved (or omitted) will result in bias. However, under Assumption 1, in addition to including all of the correct variables, the choice of feature mapping also matters. For example, if researchers only include first-order moments in their weights estimation, then $\phi(X) = X$. However, if the true feature map necessary for linear ignorability to hold also includes higher-order terms or non-linear interactions between covariates, then using only the first-order moments will result in bias (Huang et al., 2022). As such, omitted variables in such a setting would also include any transformations of existing covariates that have not been explicitly accounted for in the estimated weights. We refer readers to Hartman & Huang (2024) for more discussion about the two assumptions in the context of sensitivity analysis. We note that the proposed sensitivity framework is valid, regardless of which version of conditional ignorability researchers are interested in using.

A.3. *Bounding a Confounder's Relationship with the Outcome*

Previous literature has highlighted two characteristics of the imbalance term in Equation (3) that affect the bias from omitting a variable: (1) the overall magnitude of the the imbalance term, and (2) the relationship between the imbalance term to the outcomes (e.g., Huang, 2024; Hong et al., 2021; Cinelli & Hazlett, 2020; Shen et al., 2011). Like the marginal sensitivity model, the variance-based sensitivity model constrains the overall magnitude of the imbalance term, and implicitly assumes that the imbalance is maximally correlated with the outcome. In settings when researchers wish to account for this additional characteristic of the imbalance term, the variance-based sensitivity model can be easily extended to allow researchers to bound the relationship between the imbalance and the outcome. In particular, unlike the marginal sensitivity model, in which researchers must solve a linear programming problem to identify the extrema, there exists a closed-form solution for the bias bounds under the variance-based sensitivity model. As such, researchers can choose to evaluate the bias bounds and associated confidence intervals using less conservative values of the correlation bound.

While amplification approaches allowing researchers to examine the relationship between the outcomes and the confounder for a fixed level of imbalance exist for alternative sensitivity models, many of these methods require introducing additional complexities. (For example, Rosenbaum & Silber (2009) requires invoking parametric assumptions on the outcomes.) In contrast, the variance-based sensitivity model allow researchers to easily incorporate additional information about the confounder to directly bound the relationship between the outcome and the imbalance in an omitted confounder. We provide recommendations for alternative bounds that researchers can use in Appendix A.5, as well as benchmarking procedure that allows researchers to use obesrved covariate data to estimate plausible correlation bounds.

A.4. *Benchmarking for the Variance-based Sensitivity Model*

Previous works have suggested the use of benchmarking to help assess the plausibility of sensitivity parameter values (Huang, 2024; Hartman & Huang, 2024; Cinelli & Hazlett, 2020; Hong et al., 2021; Carnegie et al., 2016; Hsu & Small, 2013). To perform benchmarking, researchers sequentially omit different observed covariates and re-estimate the weights. They can then calculate the error that arises from omitting each covariate and directly estimate the corre-

sponding sensitivity parameters (or bound for the corresponding sensitivity parameters). These quantities are benchmarks in the sense that they describe the degree of bias that may be occurred by omitting an unobserved covariate similar to the associated observed covariate (in a sense we make formal below). They can enhance interpretability of sensitivity parameter values by giving concrete examples of the kinds of variables that might be associated with them.

We propose a formal benchmarking procedure for the variance-based sensitivity model. To begin, let there be p total observed covariates (i.e., $X \in \mathbb{R}^{n \times p}$). Then for the j-th covariate, where $j \in \{1, ..., p\}$, we define the benchmarked weights $w^{-(j)}$ as the population weights defined using all covariates except for the j-th covariate. Following Huang (2024) and Hartman & Huang (2024), we consider an unobserved confounder that is similar in strength to the jth covariate in the following sense:

$$\frac{\text{var}(w^* - w \mid A = 1)}{\text{var}(w - w^{-(j)} \mid A = 1)} = 1. \tag{1}$$

In words, our omitted confounder exhibits imbalance relative to all observed covariates (as measured by the variance of the error in the weights when it is omitted) identical to that exhibited by the jth observed covariate relative to all other observed covariates.

For an unobserved confounder satisfying (1), we can write the associated R^2 value in terms of $w^{-(j)}$:

$$\begin{aligned}
R^2 &= 1 - \frac{\text{var}(w \mid A = 1)}{\text{var}(w^* \mid A = 1)} \\
&= \frac{\text{var}(w^* - w \mid A = 1)}{\text{var}(w^* \mid A = 1)} \\
&= \frac{\text{var}(w \mid A = 1) - \text{var}(w^{-(j)} \mid A = 1)}{\text{var}(w^* \mid A = 1)} \\
&= \frac{\text{var}(w \mid A = 1) - \text{var}(w^{-(j)} \mid A = 1)}{\text{var}(w|A = 1) + \big\{\text{var}(w \mid A = 1) - \text{var}(w^{-(j)} \mid A = 1)\big\}}
\end{aligned}$$

Dividing both the numerator and denominator by $\text{var}(w \mid A = 1)$:

$$\begin{aligned}
&= \frac{\{\text{var}(w \mid A = 1) - \text{var}(w^{-(j)} \mid A = 1)\}/\text{var}(w \mid A = 1)}{1 + \{\text{var}(w \mid A = 1) - \text{var}(w^{-(j)} \mid A = 1)\}/\text{var}(w \mid A = 1)} \\
&= \frac{R^2_{-(j)}}{1 + R^2_{-(j)}}.
\end{aligned}$$

We use the notation $R^2_{(j)}$ to denote this R^2 value computed under assumption (1) for covariate j, and $\hat{R}^2_{(j)}$ to denote its sample estimate given in Equation (7); the term "benchmarked R^2" is also used to refer to both quantities.

This particular form for the benchmarked R^2 is specific to the relationship between covariate j and the unobserved confounder that we posited in equation (1). For example, if we had assumed instead that $\text{var}(w^* - w|A = 1)/\text{var}(w^*|A = 1)$ and $\text{var}(w - w^{-(j)}|A = 1)/\text{var}(w|A = 1)$ were identical, $R^2_{-(j)}$ would be the appropriate benchmarked R^2 value. We prefer our approach over this particular alternate assumption because $\text{var}(w|A = 1)$ is not a reliable estimate of the baseline variation in the true weights w^* relative to which R^2 is defined (for related discussion see Cinelli & Hazlett, 2020, §6.2 and Huang, 2024, §4.3). However, in principle our

benchmarking approach could be generalized to many other kinds of relationships between an unobserved confounder and an observed covariate j.

When interpreting the benchmarking results, it is important to consider that the magnitude of the benchmarked R^2 values is determined by the *residual* imbalance. More concretely, we consider the variables *income* and *education* in the running example. We expect that both income and education will be predictive of individuals' propensity for fish consumption. However, omitting income alone may not result in a very large R^2 value, because by balancing education, we have implicitly controlled for some of the imbalance in income. The benchmarked R^2 parameter thus represents the setting in which researchers have omitted a variable that, when controlling for all the other observed variables, has the same amount of residual imbalance as income after controlling for education. In cases when researchers wish to consider omitting a variable similar to a set of collinear variables, they can omit subsets of variables and perform the same benchmarking exercise.

Benchmarking can also be used to assess the plausibility of the event $R^2 \geq R^2_*$. More specifically, we can directly compare the benchmarked $\hat{R}^2_{(j)}$ values for $j \in \{1, ..., p\}$ with the estimated R^2_* to see how much more or less imbalanced an omitted confounder must be, relative to an observed covariate, in order to result in an R^2 value equal to R^2_*. We refer to this as the *minimum relative imbalance* (MRI):

$$\text{MRI}(j) = \frac{R^2_*}{\hat{R}^2_{(j)}}.$$

If the MRI is small (i.e., $\text{MRI}(j) < 1$), the omitted confounder need not be very imbalanced, relative to the j-th covariate, in order to make a null result plausible. In contrast, if the MRI is large (i.e., $\text{MRI}(j) > 1$), then the omitted confounder must be more imbalanced than the j-th observed covariate to make a null result plausible.

Benchmarking offers an opportunity for researchers to incorporate their substantive understanding into the sensitivity analysis and provides much-needed interpretability for the sensitivity framework. In particular, when researchers have strong priors about which underlying observed variables control the treatment assignment mechanism, formal benchmarking is very useful for reasoning about the plausibility of an omitted confounder strong enough to explain observed results in the absence of a true effect.

A.5. *Moving Away from Worst-Case Correlation Bounds*

Theorem 1 allows researchers to calculate the maximum bias that can occur for a fixed R^2. This is done by assuming the correlation between the imbalance in the omitted confounder is maximally correlated with the outcome. This can be conservative in practice. We provide several recommendations for researchers who may wish to relax this bound. Doing so can result in narrower bounds, at the cost of having to reason about an additional parameter. Throughout this section, we will refer to the correlation bound as ρ^*, such that the maximum bias is written as:

$$\rho^* \cdot \sqrt{\frac{R^2}{1 - R^2} \cdot \text{var}(Y \mid A = 1) \cdot \text{var}(w \mid A = 1)}$$

We suggest several different approaches for researchers to estimate less conservative bounds.

Approach 1: Estimating Bounds using Relative Correlation. Applying the results from Huang (2024) (Lemma 3.2, Equation 22), we can decompose the correlation between the imbalance and the outcome into a function of the R^2 value, the correlation between the estimated

weights and the outcomes, and the correlation between the true weights and the outcomes:

$$\text{cor}(w, Y \mid A=1)\sqrt{\frac{1-R^2}{R^2}} - \text{cor}(w^*, Y \mid A=1) \cdot \sqrt{\frac{1}{R^2}} \tag{2}$$

As such, an intuitive way to evaluate bounds for the correlation term is to posit a bound for the correlation between the true weights and the outcomes by a relative scaling constant k:

$$k := \frac{\text{cor}(w^*, Y \mid A=1)}{\text{cor}(w, Y \mid A=1)},$$

where k represents how many more times correlated the true weights are to the outcomes, relative to the estimated weights. k will be naturally upper-bounded at $1/\text{cor}(w, Y)$. Using Equation (2), researchers can then obtain a new upper bound for ρ^*:

$$\rho^* \leq \frac{\text{cor}(w, Y \mid A=1)}{\sqrt{R^2}} \left(\sqrt{1-R^2} - k\right)$$

It is worth noting that the correlation bound will change, depending on the R^2 parameter.

Approach 2: Benchmarking the Correlation Term. In practice, researchers may also perform benchmarking to estimate what may be plausible correlation values. More specifically, researchers can calculate the error from omitting the j-th covariate and evaluate the correlation between the residual imbalance in the j-th covariate and the outcome, using this as the upper bound for ρ^*:

$$\rho^*_{(j)} \leq \widehat{\text{cor}}(w - w^{-(j)}, Y \mid A=1)$$

Evaluating the bias at $\rho^*_{(j)}$ and $\hat{R}^2_{(j)}$ provides researchers with an estimate of the bias if they omitted a confounder with residual imbalance that is (1) equivalent in magnitude as the residual imbalance of the j-th covariate, and (2) equivalently as correlated with the outcome as the residual imbalance of the j-th covariate. Researchers can then estimate the associated confidence intervals by fixing both the correlation term and R^2.

A.6. *Considering Sharp Bounds*

To construct sharp bounds under the variance-based sensitivity model, we leverage the norm representation of the variance-based sensitivity model, introduced in Theorem 3, and solve the following optimization problem:

$$\min_{\lambda} \Big/ \max_{\lambda} \; \mathbb{E}\left(\lambda w Y \mid A=1\right)$$

$$\text{s.t. } ||\lambda||_{2,w} \leq \sqrt{\frac{k}{1-R^2}}, \qquad \text{(Constraint from VBM)} \tag{3a}$$

$$\mathbb{E}\left(\lambda \mid X=x, A=1\right) = 1, \qquad \text{(Density Constraint)} \tag{3b}$$

$$\lambda > 0. \qquad \text{(Positivity Constraint)} \tag{3c}$$

We provide details about each constraint. (3a) enforces the weighted L_2 norm constraint, imposed by the variance-based sensitivity model (see Theorem 3). The constraint in (3b) constrains the λ values chosen so that the resulting implied weights (λw) balance arbitrary functions of the observed covariates X.

This constraint is discussed in detail (in the context of the marginal L_∞ sensitivity model) by Dorn & Guo (2023), who refer to it as "data compatibility," and by Bruns-Smith & Zhou

(2023) and Kallus & Zhou (2020), who refer to it as the "density constraint," Finally, (3c) must be imposed, to ensure that the constructed ideal weights λw are non-negative.

The bound proposed in Theorem 1 fails to be sharp because it does not fully account for the density constraint and because it relies on a bound for correlations between errors in weights and outcomes that may not be sharp. In the following subsections, we will discuss how to solve for the optimization problem in (3) and provide more intuition for the looseness of the bias bound.

Solving for Sharp Bounds. In the following subsection, we discuss how to solve the optimization problem (3). Our argument closely follows a strategy used by Zhang & Zhao (2024) for a related problem. To be clear about what terms depend on which variables, we write λ as $\lambda(X, U)$.

To begin, we note that constraints (3b)-(3c) condition on the event $X = x$ while constraint (3a) and the objective function do not. We first seek to work with a problem that conditions on $X = x$ throughout. First, we remove constraint (3a) by moving it into the objective function.

$$\min_{\lambda} \ \frac{1}{2}\mathbb{E}\left\{\lambda^2(x,U)w^2(x) \mid X = x, A = 1\right\} + \theta\mathbb{E}\left\{\lambda(x,U)w(x)Y \mid X = x, A = 1\right\} \tag{4a}$$

$$\text{s.t. } \mathbb{E}\left\{\lambda(x,U) \mid X = x, A = 1\right\} = 1 \tag{4b}$$

$$\lambda(x,u) \geq 0 \quad \forall\, u \in \mathcal{U} \tag{4c}$$

In this new penalized form of the problem a penalty parameter $\theta > 0$ replaces R^2, but solving it is guaranteed to produce a solution that is also optimal for problem (3) for a particular value of R^2 that can be computed post hoc. For discussion of closely-related transformations to optimization problems, see Pimentel & Kelz (2020) and Zhang & Zhao (2024, Prop. 2).

Since we do not restrict how $\lambda(X, U)$ varies with X, this problem is separable, and we can find its solution by taking a version of the objective function that conditions on $X = x$, solving it separately for each x, and combining. Here is the conditional version of the problem:

$$\min_{\lambda} \ \frac{1}{2}\mathbb{E}\left\{\lambda^2(x,U)w^2(x) \mid X = x, A = 1\right\} + \theta\mathbb{E}\left\{\lambda(x,U)w(x)Y \mid X = x, A = 1\right\} \tag{5a}$$

$$\text{s.t. } \mathbb{E}\left\{\lambda(x,U) \mid X = x, A = 1\right\} = 1 \tag{5b}$$

$$\lambda(x,u) \geq 0 \quad \forall\, u \in \mathcal{U} \tag{5c}$$

Using (5), we take the Lagrangian:

$$\begin{aligned}\mathcal{L} =& \frac{1}{2}\mathbb{E}\left\{\lambda^2(x,U)w^2(x) \mid X = x, A = 1\right\} + \theta\mathbb{E}\left\{\lambda(x,U)w(x)Y \mid X = x, A = 1\right\} \\ &+ \theta_{x,2}\left[1 - \mathbb{E}\left\{\lambda(x,U) \mid X = x, A = 1\right\}\right] - \mathbb{E}\left\{\theta_{U,3} \cdot \lambda(x,U) \mid X = x, A = 1\right\}\end{aligned}$$

Taking the functional derivative:

$$\lambda(x,U)w^2(x) + \theta w(x)Y - \theta_{x,2} - \theta_{U,3} = 0$$

. From complementary slackness we get the following.

- Scenario 1: $\theta_{U,3} = 0, \lambda(x,U) > 0$. Then,

$$\lambda(x,U) = \frac{\theta_{x,2} - \theta Y w(x)}{w^2(x)} > 0 \implies \theta_{x,2} - \theta Y w(x) > 0. \tag{6}$$

- Scenario 2: $\theta_{U,3} > 0$. Then, $\lambda(x,U) = 0$, and we have the following:

$$\theta_{U,3} = \theta w(x)Y - \theta_{x,2} > 0 \tag{7}$$

This implies that $\theta_{x,2} - \theta Y w(x) < 0$.

Combining Equation (6) and (7) we have the following:

$$\lambda_*(x,U) = \frac{\theta_{x,2} - \theta Y w(x)}{w^2(x)} \mathbb{1}\{\theta_{x,2} - \theta Y w(x) > 0\}$$

Re-arranging the terms:

$$= \frac{\theta}{w(x)} \left\{ \frac{\theta_{x,2}}{\theta w(x)} - Y \right\} \mathbb{1}\left\{ \frac{\theta_{x,2}}{\theta w(x)} - Y > 0 \right\}. \tag{8}$$

The solution in Equation (8) depends on an unknown variable $\theta_{x,2}$. To solve for $\theta_{x,2}$, we can exploit the fact that the density constraint requires that $\mathbb{E}(\lambda(x,U) \mid X = x) = 1$ for all $x \in \mathcal{X}$. Thus, we arrive an estimating equation, which can be used to solve for $\theta_{x,2}$:

$$\begin{aligned}
\mathbb{E}\left[\frac{\theta}{w(x)} \left\{ \frac{\theta_{x,2}}{\theta w(x)} - Y \right\} \mathbb{1}\left\{ \frac{\theta_{x,2}}{\theta w(x)} - Y > 0 \right\} \middle| X = x, A = 1 \right] &= 1 \\
\frac{\theta}{w(x)} \mathbb{E}\left[\left\{ \frac{\theta_{x,2}}{\theta w(x)} - Y \right\} \mathbb{1}\left\{ \frac{\theta_{x,2}}{\theta w(x)} - Y > 0 \right\} \middle| X = x, A = 1 \right] &= 1 \\
\mathbb{E}\left[\left\{ \frac{\theta_{x,2}}{\theta w(x)} - Y \right\} \mathbb{1}\left\{ \frac{\theta_{x,2}}{\theta w(x)} - Y > 0 \right\} \middle| X = x, A = 1 \right] - \frac{w(x)}{\theta} &= 0
\end{aligned} \tag{9}$$

The estimating equation in Equation (9) is reasonable since the left-hand side is negative as $\theta_{x,2} \to -\infty$, positive for sufficiently large $\theta_{x,2}$, and non-decreasing in $\theta_{x,2}$.

In summary, equation (8) provides a closed-form representation for the optimal solution to problem (3) in terms of population parameters including $\theta_{x,2}$. At a high level, this formula tells us that the worst-case weights $\lambda(x,U)$ are based on an x-specific outcome threshold ($\theta_{x,2}/\theta w(x)$); observations with outcomes larger than this threshold receive zero weight, while observations with outcome below the threshold receive weights that increase linearly in magnitude with distance below the threshold.

Unfortunately, in practice, estimating this sharp bound from observed sample data can be challenging. In particular, solving the estimating equation to determine the x-specific threshold requires an estimate of the underlying density function $f(Y \mid X, A = 1)$. Existing estimation methods for similar problems rely either on strong parametric assumptions or on nonparametric density estimation procedures and associated extensive mathematical argument tailored to the specific problem to obtain sufficient asymptotic convergence guarantees (e.g., Jordan et al., 2022; Zhang & Zhao, 2024). Deriving estimators for this bound under sufficiently general assumptions and validating their finite-sample performance is beyond our scope here. Furthermore, even if derived we are not convinced that such estimators would be practically appealing to investigators working on substantive problems in medical and social sciences. For example, in their version of L_2 sensitivity analysis, Zhang & Zhao (2024) suggest either using a Nadaraya-Watson kernel estimator to estimate the conditional density of Y or assuming that outcomes are drawn from an additive regression model with normally-distributed errors. Both approaches seem unattractive in practice compared to estimating the bounds in (4), which can be done with simple statistical models and without strong distributional assumptions on the outcome variable.

Interrogating looseness in the bias bound. Our proposed bias bound relies on bounding the correlation between the error in the weights $(w - w^*)$ and the outcome Y. The gap between the proposed bias bound and the sharp bound arises from the handling of this quantity.

We provide the full derivation of the bound for reference:

$$\text{cor}(w - w^*, Y \mid A = 1)$$

Denote $C^{\perp B} := C - \text{cor}(C, B) \cdot B$ for random variables C and B. Then:

$$\begin{aligned}
&= \text{cor}(w - w^*, w \mid A = 1) \cdot \text{cor}(Y, w \mid A = 1) + \text{cov}\{(w - w^*)^{\perp w}, Y^{\perp w} \mid A = 1\} \\
&= \text{cor}(w - w^*, w \mid A = 1) \cdot \text{cor}(Y, w \mid A = 1) \\
&\quad + \text{cor}\{(w - w^*)^{\perp w}, Y^{\perp w} \mid A = 1\} \cdot \sqrt{\text{var}\{(w - w^*)^{\perp w} \mid A = 1\}\text{var}(Y^{\perp w} \mid A = 1)} \\
&= \text{cor}(w - w^*, w \mid A = 1) \cdot \text{cor}(Y, w \mid A = 1) \\
&\quad + \text{cor}\{(w - w^*)^{\perp w}, Y^{\perp w} \mid A = 1\} \cdot \sqrt{1 - \text{cor}^2(w, w - w^* \mid A = 1)}\sqrt{1 - \text{cor}^2(w, Y \mid A = 1)}
\end{aligned}$$

From the density constraint, $\text{cor}(w - w^*, w \mid A = 1) = 0$ and $(w - w^*)^{\perp w} = w - w^*$:

$$\begin{aligned}
&= \text{cor}(w - w^*, Y^{\perp w} \mid A = 1) \cdot \sqrt{1 - \text{cor}^2(w, Y \mid A = 1)} \\
&\leq \sqrt{1 - \text{cor}^2(w, Y \mid A = 1)}
\end{aligned}$$

The bound fails to be sharp in general because there may not exist a choice of weights w^* such that the density constraint is satisfied and $\text{cor}(w - w^*, Y^{\perp w} \mid A = 1) = 1$. This gap is particularly evident if we consider the optimal solution to the sharp bound problem in equation (8). We can write the error term $w - w^*$ as:

$$w - w^* = w(X) - \theta \left\{ \frac{\theta_{x,2}}{\theta w(X)} - Y \right\} \mathbb{1} \left\{ \frac{\theta_{x,2}}{\theta w(X)} - Y > 0 \right\},$$

In order for $w - w^*$ to be perfectly correlated with $Y^{\perp w}$, $w - w^*$ would need to be a linear transformation of $Y^{\perp w}$; however, this will be generally infeasible, given the thresholding from $\mathbb{1}\{\theta_{x,2}/\theta w(X) - Y > 0\}$.

We can quantify the size of the gap using our representation for the sharp bounds. In particular, define

$$\xi^* = \text{cor}\left[w(X) - \theta \left\{ \frac{\theta_{x,2}}{\theta w(X)} - Y \right\} \mathbb{1} \left\{ \frac{\theta_{x,2}}{\theta w(X)} - Y > 0 \right\}, Y^{\perp w} \mid A = 1 \right].$$

Then, the gap between the conservative, closed-form bias bound proposed in Theorem 1 and the sharp bias bound is:

$$(1 - \xi^*) \cdot \sqrt{1 - \text{cor}^2(w, Y \mid A = 1)}\sqrt{\frac{R^2}{1 - R^2}\text{var}(Y \mid A = 1) \cdot \text{var}(w \mid A = 1)} \tag{10}$$

The gap thus depends on how much slippage there is between ξ^* and the upper bound of 1 imposed by the conservative bias bound. Furthermore, Equation (10) highlights that the conservative bias bound can, at most, be twice as large as the sharp bounds. A more detailed characterization of the looseness in the bound would seem to require a thorough understanding of how $\theta_{x,2}$ covaries with X and Y; as $\theta_{x,2}$ is defined as the solution to an estimating equation, such analysis is nontrivial and we leave it to future work.

A.7. *Extended discussion for sample boundedness*

PROPOSITION 1 (NECESSARY CONDITION FOR VALIDITY OF SAMPLE BOUNDS). *Define $\mathcal{A}$ as the set of all observed Y_i values across the sample $A_i = 1$. For the true weighted mean to be estimable under sample boundedness (i.e., $\mathbb{E}(Y \mid A = 0) \in [\min_{i:A=1} Y, \max_{i:A=1} Y]$), the expectation of the outcomes not contained in the sample range must be constrained by the following:*

$$\mathbb{E}(Y \mid A = 0, Y \notin \mathcal{A}) \in \left[\frac{1}{1 - p_{\mathcal{A}}} \min_{i:A_i=1} Y_i - \frac{p_{\mathcal{A}}}{1 - p_{\mathcal{A}}} \max_{i:A_i=1} Y_i,\ \frac{1}{1 - p_{\mathcal{A}}} \max_{i:A_i=1} Y_i - \frac{p_{\mathcal{A}}}{1 - p_{\mathcal{A}}} \min_{i:A_i=1} Y_i \right],$$

where $p_{\mathcal{A}} := \Pr(Y \in \mathcal{A} \mid A = 0)$ represents the proportion of unobserved outcomes that fall within the observed sample range.

The bound specified above represents how much overlap there must exist in the observed and unobserved potential outcomes. The bound is a function of (1) the proportion of unobserved units with outcomes that are outside the range of outcomes across the observed sample units (i.e., $1 - p_{\mathcal{A}} = \Pr(Y \notin \mathcal{A} \mid A = 0)$), and (2) the sample bounds. If a small proportion of the outcomes in the unobserved population fall outside the sample bounds, then the bound will be relatively wide. However, if a large proportion of outcomes in the unobserved population fall outside the sample bounds, then the bound will be more narrow.

We also simulate the behavior of both sensitivity models under varying amounts of overlap.

Example 1 (Coverage Rates in Limited Outcome Overlap Settings). Define the treatment assignment mechanism as a logit model, and the outcome model as a linear model:

$$\begin{cases} \Pr(Z_i = 1 \mid \mathcal{X}) \propto \frac{\exp(\gamma_1 X_{i,1} + \gamma_2 X_{i,2} + \beta U_i)}{1 + \exp(\gamma_1 X_{i,1} + \gamma_2 X_{i,2} + \beta U_i)}, \\ Y_i = \gamma_1 X_{i,1} + \gamma_2 X_{i,2} + \beta U_i + v_i, \end{cases}$$

where $X_{i,1}, X_{i,2}$ and U_i are standard normal random variables, and $v_i \sim N(0, \sigma_v^2)$. v_i represents a noise parameter that controls for how much outcome overlap there is. When σ_v^2 is large, then there is increased overlap between the treatment and control groups, as the treatment probability is less correlated with the outcome.

We vary $\sigma_v^2 \in \{0, 0.1, 0.25, 1, 2, 2.5\}$, and set $\gamma_1 = 2.5$, $\gamma_2 = 5$, and $\beta = 1$. For each iteration of the simulation, we assume that researchers omit U_i, and estimate confidence intervals using both the marginal sensitivity model and the variance-based sensitivity model, using the true sensitivity parameters. We visualize the coverage rates across simulations in Figure 2. We see that even in low overlap scenarios and small sample sizes, the variance-based sensitivity model have nominal coverage. However, the marginal sensitivity model struggles to achieve nominal coverage in limited overlap settings.

Example 1 highlights that in small sample settings and limited overlap, the marginal sensitivity model fails to obtain nominal coverage, *even with the true* Λ *value*. In contrast, the variance-based sensitivity model consistently has nominal coverage.

We see that within finite-sample settings, the marginal sensitivity model may obtain narrower bounds than the variance-based sensitivity model, due to their inherent sample boundedness. However, these narrower bounds risk not being valid in settings with smaller sample size and limited outcome overlap, and can risk large amounts of under-coverage. Thus, the estimated confidence intervals under the variance-based sensitivity model are technically wider, but appropriately so, providing at least nominal coverage, even in cases with severely limited outcome overlap.

B. PROOFS AND DERIVATIONS

B.1. *Proof of Theorem 1*

Proof. We will start by deriving the bias bounds.

To begin, we can decompose the bias of a weighted estimator as follows:

$$\begin{aligned}
\text{Bias}\{\tau(w)\} &= \tau(w) - \tau(w^*) \\
&= \frac{\mathbb{E}(wY \mid A=1)}{\mathbb{E}(w \mid A=1)} - \frac{\mathbb{E}(w^*Y \mid A=1)}{\mathbb{E}(w^* \mid A=1)} \\
&\text{Because } w \text{ and } w^* \text{ are centered at mean 1:} \\
&= \mathbb{E}(wY \mid A=1) - \mathbb{E}(w^*Y \mid A=1) \qquad (11) \\
&= \mathbb{E}\{(w-w^*)\cdot Y \mid A=1\} \\
&\text{By construction, } \mathbb{E}(w \mid A=1) = \mathbb{E}(w^* \mid A=1)\text{:} \\
&= \mathbb{E}\{(w-w^*)\cdot Y \mid A=1\} - \mathbb{E}(w-w^* \mid A=1)\cdot\mathbb{E}(Y \mid A=1) \\
&= \text{cov}(w-w^*, Y \mid A=1) \\
&= \text{cor}(w-w^*, Y \mid A=1)\cdot\sqrt{\text{var}(w-w^* \mid A=1)\cdot\text{var}(Y \mid A=1)} \qquad (12)
\end{aligned}$$

This is similar to the derivation provided in Shen et al. (2011) and Hong et al. (2021). However, we will go a step further to amplify the term, $\text{var}(w-w^* \mid A=1)$, into an R^2 value and the variance of the estimated weights. To do so, we extend the results from Huang (2024), which examined the bias in the context of an external validity setting, and thus, focused on re-weighting an individual-level treatment effect τ. We instead apply the results to a general missingness setting, in which we are re-weighting outcomes Y. We re-write the variance of the error in the weights in Equation (12) as a function of the R^2 parameter and the variance of the estimated weights, providing the following bias decomposition:

$$\text{Bias}\{\hat{\tau}(w)\} = \text{cor}(w-w^*, Y \mid A=1)\cdot\sqrt{\frac{R^2}{1-R^2}\cdot\text{var}(Y \mid A=1)\cdot\text{var}(w \mid A=1)},$$

where R^2 is defined in Definition 1. Because we are fixing $R^2 \in [0,1)$,and $\text{var}(Y \mid A=1)\cdot\text{var}(w \mid A=1)$ are directly estimable from the data, to maximize the bias, we must maximize the correlation term.

Applying Lemma 3.1 from Huang (2024), the error in the weights (i.e., $w-w^*$) is orthogonal to the estimated weights w (i.e., $\text{cov}(w-w^*, w \mid A=1)=0$). Then, applying the recursive formula of partial correlation, we obtain the following bounds for the correlation:

$$-\sqrt{1-\text{cor}(w, Y \mid A=1)^2} \leq \text{cor}(w-w^*, Y \mid A=1) \leq \sqrt{1-\text{cor}(w, Y \mid A=1)^2} \qquad (13)$$

Thus, Equation 4 in Theorem 1 directly follows.

It is important to note that we are implicitly comparing the population quantity $\tau(w)$ with $\tau(w^*)$. This quantity differs from the population bias, given by the difference between $\hat{\tau}(\hat{w})$ and the population ATT. In particular, because we are using a stabilized weighted estimator, there will be finite-sample error in recovering the ATT, even with the ideal weights w^*. However, the finite-sample error is of order $o(1/n)$, and will be dominated by the bias incurred from omitting a confounder from the weights (see Miratrix et al. (2013), Rosenbaum et al. (2010) for more discussion) so that the two notions of bias are asymptotically equivalent. In settings when researchers are considering a non-stabilized weighted estimator (i.e., a Horvitz-Thompson style

estimator) that is unbiased in finite samples, then the two notions of bias agree exactly.

Relaxing the Assumption that Weights have Mean 1. In the manuscript, we define the weights as standard inverse propensity score weights, normalized to mean 1. In practice, researchers often employ non-centered propensity score weights. In the following subsection, we show that whether or not the weights are normalized is inconsequential for the proposed sensitivity model. To begin, consider the following weights w' and w'^*:

$$w'(X) := \frac{\Pr(Z=1 \mid X)}{\Pr(Z=0 \mid X)}, \quad \text{and } w'^*(X,U) = \frac{\Pr(Z=1 \mid X,U)}{\Pr(Z=0 \mid X,U)}.$$

w' and w'^* can be normalized to be mean 1 by scaling by the sum of the weights: $w = \kappa w'$, where $\kappa = 1/\mathbb{E}(w' \mid Z=0)$. Furthermore, because $\mathbb{E}(w' \mid Z=0) = \mathbb{E}(w'^* \mid Z=0)$, the scaling factor used to center both sets of weights at 1 will be equivalent.

Because we are using a stabilized weighted estimator (i.e., the primary estimator of interest in the paper), point estimates do not depend on whether the centered weights (i.e., w and w^*) or the non-centered weights (w' and w'^*) are employed:

$$\begin{aligned}
\frac{\sum_{i=1}^n (1-Z_i)Y_i w_i}{\sum_{i=1}^n (1-Z_i) w_i} &= \frac{\sum_{i=1}^n (1-Z_i)Y_i \kappa \cdot w'_i}{\sum_{i=1}^n (1-Z_i)\kappa w'_i} \\
&= \frac{\kappa \cdot \sum_{i=1}^n (1-Z_i)Y_i w'_i}{\kappa \sum_{i=1}^n (1-Z_i) w'_i} \\
&\equiv \frac{\sum_{i=1}^n (1-Z_i)Y_i w'_i}{\sum_{i=1}^n (1-Z_i) w'_i}
\end{aligned}$$

Furthermore, the bias of a stabilized weighted estimator using non-normalized weights (i.e., w') will be equivalent to the bias of a stabilized weighted estimator using the weights normalized to mean 1 (i.e., w):

$$\begin{aligned}
\text{Bias}\{\tau(w')\} &= \tau(w') - \tau(w'^*) \\
&= \frac{\mathbb{E}(w'Y \mid A=1)}{\mathbb{E}(w' \mid A=1)} - \frac{\mathbb{E}(w'^*Y \mid A=1)}{\mathbb{E}(w'^* \mid A=1)} \\
&= \frac{\kappa \cdot \mathbb{E}(w'Y \mid A=1)}{\kappa \cdot \mathbb{E}(w' \mid A=1)} - \frac{\kappa \cdot \mathbb{E}(w'^*Y \mid A=1)}{\kappa \cdot \mathbb{E}(w'^* \mid A=1)} \\
&= \frac{\mathbb{E}(wY \mid A=1)}{\mathbb{E}(w \mid A=1)} - \frac{\mathbb{E}(w^*Y \mid A=1)}{\mathbb{E}(w^* \mid A=1)} \\
&\equiv \text{Bias}\{\tau(w)\}
\end{aligned}$$

We can also derive the bias of the weighted estimator with non-normalized weights:

$$\begin{aligned}
\text{Bias}\{\tau(w')\} &= \tau(w') - \tau(w'^*) \\
&= \frac{\mathbb{E}(w'Y \mid A=1)}{\mathbb{E}(w' \mid A=1)} - \frac{\mathbb{E}(w'^*Y \mid A=1)}{\mathbb{E}(w'^* \mid A=1)} \\
&= \mathbb{E}(\kappa w'Y \mid A=1) - \mathbb{E}(\kappa w'^*Y \mid A=1) \\
&= \text{cov}(\kappa(w' - w'^*), Y \mid A=1) \\
&= \kappa \cdot \text{cor}(w' - w'^*, Y \mid A=1) \sqrt{\text{var}(w' - w'^* \mid A=1) \cdot \text{var}(Y \mid A=1)}
\end{aligned}$$

The bias expression with respect to the non-normalized weights w' and w'^* is identical to the bias expression with normalized weights, up to scaling by κ. Thus, by constructing the weights to be centered at mean 1, we bypass the need to track the scaling factor κ explicitly. □

B.2. *Proof of Theorem 2*

Proof. To begin, let $\hat{\tau}(\tilde{w})$ be the estimator obtained by plugging some set of weights $\tilde{w} \in \sigma(R^2)$ into Equation (1). We present a finite-sample analogue of the bias decomposition in Equation (12) in terms of $\hat{\tau}(\widetilde{w})$ and $\hat{\tau}(w)$. We denote with a subscript n quantities that are computed across a fixed sample n (rather than across the infinite population).

$$\begin{aligned}
\hat{\tau}(\hat{w}) - \hat{\tau}(\tilde{w}) &= \frac{1}{\sum_{i\in\mathcal{A}} \hat{w}_i} \sum_{i\in\mathcal{A}} \hat{w}_i Y_i - \frac{1}{\sum_{i\in\mathcal{A}} \tilde{w}_i} \sum_{i\in\mathcal{A}} \tilde{w}_i Y_i \\
&= \frac{1}{n}\left\{\sum_{i\in\mathcal{A}} (\hat{w}_i - \tilde{w}_i) Y_i\right\} \\
&= \frac{1}{n}\left\{\sum_{i\in\mathcal{A}} (\hat{w}_i - \tilde{w}_i) Y_i\right\} - \frac{1}{n}\left\{\sum_{i\in\mathcal{A}} (\hat{w}_i - \tilde{w}_i)\right\}\left(\sum_{i\in\mathcal{A}} Y_i\right) \qquad (14) \\
&= \text{cov}_n(\hat{w}_i - \tilde{w}_i, Y_i \mid A_i = 1) \\
&= \text{cor}_n(\hat{w}_i - \tilde{w}_i, Y_i \mid A_i = 1) \cdot \sqrt{\text{var}_n(\hat{w}_i \mid A_i = 1) \cdot \frac{\hat{R}^2_{\tilde{w}}}{1 - \hat{R}^2_{\tilde{w}}} \cdot \text{var}_n(Y_i \mid A_i = 1)} \\
&= \hat{\theta}_{\tilde{w}} \cdot \hat{\rho} \sqrt{\text{var}_n(\hat{w}_i \mid A_i = 1) \cdot \frac{\hat{R}^2_{\tilde{w}}}{1 - \hat{R}^2_{\tilde{w}}} \cdot \text{var}_n(Y_i \mid A_i = 1)} \qquad (15)
\end{aligned}$$

The result in line (14) follows from assuming, without loss of generality, that $\hat{w}_i$ and $\tilde{w}_i$ are normalized to be mean 1, and $\hat{R}^2_{\tilde{w}} := 1 - \text{var}_n(\hat{w}_i)/\text{var}_n(\tilde{w}_i)$. The final line, in which $\hat{\rho} = \sqrt{1 - \text{cor}_n(\hat{w}_i, Y_i \mid A_i = 1)^2}$ and $\hat{\theta}_{\tilde{w}} = \frac{\text{cor}_n(\hat{w}_i - \tilde{w}_i, Y_i \mid A_i = 1)}{\sqrt{1 - \text{cor}_n(\hat{w}_i, Y_i \mid A_i = 1)^2}} \in [-1, 1]$, follows from a finite-sample version of Equation (13).

We construct our estimator for $\tau(\tilde{w})$ by taking Equation (15) and replacing the finite-sample quantities $\hat{R}^2_{\tilde{w}}$ and $\hat{\theta}_{\tilde{w}}$ with the population quantities $R^2_{\tilde{w}} = 1 - \text{var}(w)/\text{var}(\tilde{w})$ and $\theta_{\tilde{w}} = \frac{\text{cor}(w - \tilde{w}, Y \mid A=1)}{\sqrt{1 - \text{cor}(w, Y \mid A=1)^2}}$. We then subtract this bias estimate from $\hat{\tau}(\hat{w})$:

$$\begin{aligned}
\hat{\tau}(\tilde{w}; \hat{w}_1, ..., \hat{w}_n) &:= \hat{\tau}(\hat{w}) - \theta_{\tilde{w}} \cdot \hat{\rho} \sqrt{\text{var}_n(\hat{w}_i \mid A_i = 1) \frac{R^2_{\tilde{w}}}{1 - R^2_{\tilde{w}}} \cdot \text{var}_n(Y_i \mid A_i = 1)} \\
&= \hat{\tau}(\hat{w}) - \text{Bias}_n\{\hat{\tau}(\hat{w}) \mid \tilde{w}\}. \qquad (16)
\end{aligned}$$

While it may seem odd to construct an estimator using population quantities such as $R^2_{\tilde{w}}$ and $\theta_{\tilde{w}}$, we note that in practice we never calculate $\hat{\tau}(\tilde{w})$ directly, and instead use Equation (6) to construct union bounds for confidence intervals derived from infinitely many values of $\tilde{w}$; these union bounds are estimable without knowledge of any particular $R^2_{\tilde{w}}$ or $\theta_{\tilde{w}}$.

We now consider bootstrapped estimates $\hat{\tau}^{(b)}(\tilde{w}; \hat{w}_1^{(b)}, ..., \hat{w}_n^{(b)})$:

$$\begin{aligned}
&\hat{\tau}^{(b)}(\tilde{w}; \hat{w}_1^{(b)}, ..., \hat{w}_n^{(b)}) \\
&\quad = \hat{\tau}^{(b)}(\hat{w}^{(b)}) - \text{Bias}_n \left\{ \hat{\tau}^{(b)}(\hat{w}^{(b)}) \mid \tilde{w} \right\} \\
&\quad = \hat{\tau}^{(b)}(\hat{w}^{(b)}) - \theta_{\tilde{w}} \cdot \hat{\rho}^{(b)} \sqrt{\text{var}_n(\hat{w}_i^{(b)} \mid A_i = 1) \frac{R^2_{\tilde{w}}}{1 - R^2_{\tilde{w}}} \cdot \text{var}_n(Y_i^{(b)} \mid A_i = 1)},
\end{aligned}$$

where $\hat{w}^{(b)}$ represents the vector of bootstrapped weights, $\hat{\tau}^{(b)}(\hat{w}^{(b)})$ represents the weighted estimator, estimated across the b-th bootstrap sample $\{Y_i^{(b)}, Z_i^{(b)}, X_i^{(b)}\}_{i=1}^n$. Because $\theta_{\tilde{w}}$ and $R^2_{\tilde{w}}$ are fixed (across bootstrap samples), the components that drive variation across bootstrap samples are: $\hat{\tau}^{(b)}(\hat{w}^{(b)})$, $\text{var}_n(\hat{w}_i^{(b)} \mid A_i = 1)$, $\text{var}_n(Y_i^{(b)} \mid A_i = 1)$, and $\hat{\rho}^{(b)}$ (which is a function of $\text{cor}_n(\hat{w}_i^{(b)}, Y_i^{(b)} \mid A_i = 1)$).

An overview of the proof is as follows. Following Zhao et al. (2019), we will use a Z-estimation framework. In particular, we will add in three additional parameters: $\hat{\mu}_w^2$, $\hat{\mu}_Y$, $\hat{\mu}_Y^2$, which represent the second order moment of the weights, the average of the outcomes, and the second order moment of the outcomes, respectively. Then, we will invoke the asymptotic normality of bootstrapped Z-estimators. In the following proof, we will show the validity of the percentile bootstrap in the case that researchers are using inverse propensity score weights; however, we note that researchers can invoke the results in Soriano et al. (2023) to show validity of the results for balancing weights.

To begin, define μ_w as the expectation of the weights:

$$\mu_w = \mathbb{E}(Aw) \equiv \mathbb{E}[A \cdot \{1 + \exp(-\beta X)\}].$$

Then, we define μ as:

$$\mu = \frac{\mathbb{E}[AY\{1 + \exp(-\beta^\top X)\}]}{\mu_w}.$$

Define $\mu_w^2 = \mathbb{E}(Aw^2)$ and $\sigma_Y^2 = \mathbb{E}(AY^2)$ as the second moment of the weights and the outcomes, respectively. Then, we define the vector $\theta = \left(\mu, \mu_w, \beta, \mu_w^2, \mu_Y, \mu_Y^2\right)^\top \in \Theta$. Define the function $Q : 0, 1 \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d+5}$, where for $t = (a, x^\top, y) \in \{0, 1\} \times \mathbb{R}^d \times \mathbb{R}$:

$$Q(t \mid \theta) = \begin{pmatrix} Q_1(t|\theta) \\ Q_2(t|\theta) \\ Q_3(t|\theta) \\ Q_4(t|\theta) \\ Q_5(t|\theta) \\ Q_6(t|\theta) \end{pmatrix} := \begin{pmatrix} \left\{ a - \frac{\exp(\beta^\top x)}{1+\exp(\beta^\top x)} \right\} x \\ \mu_w - a \left\{ 1 + \exp(-\beta^\top x) \right\} \\ \mu_w \mu - ay \left\{ 1 + \exp(-\beta^\top x) \right\} \\ \mu_w^2 - a \left\{ 1 + \exp(-\beta^\top x) \right\}^2 \\ \mu_y - ay \\ \mu_y^2 - ay^2 \end{pmatrix} \tag{17}$$

Finally, we define $\Phi(\theta)$ as:

$$\Phi(\theta) = \int Q(t|\theta) d\mathbb{P}(t),$$

where $T = (A, X^\top, AY)^\top \sim \mathbb{P}$, where $\mathbb{P}$ represents the true distribution generating the data. It is simple to see that $\Phi(\theta^*) = 0$, when θ^* is equal to the true parameter values. Then, the Z-

estimates $\hat{\theta}$:

$$\Phi_n(\hat{\theta}) := \frac{1}{n}\sum_{i=1}^{n} Q(T_i|\hat{\theta}) \tag{18}$$

$$= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\left\{A_i - \frac{\exp(\hat{\beta}^\top X_i)}{1+\exp(\hat{\beta}^\top X_i)}\right\} X_i \\ \hat{\mu}_w - \frac{1}{n}\sum_{i=1}^{n} A_i\left\{1+\exp(-\hat{\beta}^\top X_i)\right\} \\ \hat{\mu}_w\mu - \frac{1}{n}\sum_{i=1}^{n} A_i Y_i\left\{1+\exp(-\hat{\beta}^\top X_i)\right\} \\ \hat{\mu}_w^2 - \frac{1}{n}\sum_{i=1}^{n} A_i\left\{1+\exp(-\hat{\beta}^\top X_i)\right\}^2 \\ \hat{\mu}_y - \frac{1}{n}\sum_{i=1}^{n} A_i Y_i \\ \hat{\mu}_y^2 - \frac{1}{n}\sum_{i=1}^{n}(A_i Y_i^2) \end{pmatrix} = 0 \tag{19}$$

We define $\Sigma := \mathbb{E}\{Q(t \mid \theta)Q(t \mid \theta)^\top\}$. We will invoke the following regularity conditions, consistent with Zhao et al. (2019).

Assumption 2 (Regularity Conditions). Assume that the parameter space Θ is compact, and that θ is in the interior of Θ. Furthermore, (Y, X) satisfies the following:

1. $\mathbb{E}(Y^4) < \infty$
2. $\det\left[\mathbb{E}\left\{\frac{\exp(\beta^\top X)}{(1+\exp(\beta^\top X))^2} XX^\top\right\}\right] > 0$
3. $\forall$ compact subsets $S \subset \mathbb{R}^d$, $\mathbb{E}\{\sup_{\beta\in S}\exp(\beta^\top X)\} < \infty$

To show asymptotic normality of bootstrapped Z-estimators, we must first verify that $\dot{\Phi}_0$ and Σ are well-defined.

$$\dot{\Phi}_0 = \mathbb{E}\left\{\nabla_{\theta=\theta_0} Q(T|\theta)\right\}$$

$$= \begin{pmatrix} 0 & 0 & -\mathbb{E}\left\{\frac{\exp(\beta_0^\top X)}{1+\exp(\beta_0^\top X)^2} XX^\top\right\} & 0\,0\,0 \\ 0 & 1 & \mathbb{E}\{AX^\top \exp(-\beta_0^\top X)\} & 0\,0\,0 \\ \mu_w & \mu & \mathbb{E}[AYX^\top\{\exp(\beta_0^\top X)\}] & 0\,0\,0 \\ 0 & 0 & \mathbb{E}[AX^\top\{\exp(\beta_0^\top X) + \exp(-2\beta_0^\top X)\}] & 1\,0\,0 \\ 0 & 0 & 0 & 0\,1\,0 \\ 0 & 0 & 0 & 0\,0\,1 \end{pmatrix}$$

By Leibniz Formula:

$$\left|\det(\dot{\Phi}_0)\right| = \left|\det\begin{pmatrix} 0 & 0 & -\mathbb{E}\left\{\frac{\exp(\beta_0^\top X)}{1+\exp(\beta_0^\top X)^2} XX^\top\right\} \\ 0 & 1 & \mathbb{E}\{AX^\top\exp(-\beta_0^\top X)\} \\ \mu_w & \mu & \mathbb{E}[AYX^\top\{\exp(\beta_0^\top X)\}] \end{pmatrix} \det\begin{pmatrix} 1\,0\,0 \\ 0\,1\,0 \\ 0\,0\,1 \end{pmatrix}\right|$$

$$= \mu_w \left|\det \mathbb{E}\left[\frac{\exp(\beta_0^\top X)}{\{1+\exp(\beta_0^\top X)\}^2} XX^\top\right]\right| > 0,$$

which follows by regularity condition (2). As such, $\dot{\Phi}_0$ is invertible. Furthermore, by regularity condition (1), $\Sigma < \infty$.

As such, we simply need to verify the three conditions for asymptotic normality of bootstrapped Z-estimators:

1. The class of functions $t \to Q(t|\theta) : \theta \in \Theta$ is $\mathbb{P}$-Glivenko-Cantelli.
2. $||\Phi(\theta)||_1$ is strictly positive outside every open neighborhood of θ_0.

3. The class of functions is $\mathbb{P}$-Donsker, and $\mathbb{E}((Q(T|\theta_n) - Q(T|\theta_0))^2) \to 0$ whenever $||\theta_n - \theta_0||_1 \to 0$.

It is worth noting that the first three parameters (μ, μ_w, β) are special cases from Zhao et al. (2019), in which we do not perform any shifting in the weights (i.e., $h(x, y) = 0$). We will then show that the three conditions still hold after additionally accounting for the last three parameters. The proof for each condition is provided below.

Condition 1: The class of functions $t \to Q(t|\theta) : \theta \in \Theta$ is $\mathbb{P}$-Glivenko-Cantelli.

$$||Q(t|\theta) \leq ||Q_1(t|\theta)||_1 + \sum_{b=2}^{5} |Q_b(t|\theta)|$$

Zhao et al. (2019) show that $||Q_1(t|\theta)||_1 + |Q_2(t|\theta)| + |Q_3(t|v)|$ is bounded as a function of x, y, and some absolute constant M_1:

$$||Q_1(t|\theta)||_1 + |Q_2(t|\theta)| + |Q_3(t|v)| \leq ||x||_1 + |y| + \exp(-\beta^\top x)(1 + |y|) + M_1.$$

As such, all that is left to show is to show that $|Q_4(t|\theta)| + |Q_5(t|\theta)| + |Q_6(t|\theta)|$ is finite. To begin:

$$\begin{aligned}
|Q_4(t|\theta)| &= |\mu_w^2 - a\{1 + \exp(-\beta^\top x)^2\}| \\
&\leq \mu_w^2 + \{1 + \exp(-\beta^\top x)\}^2 \\
|Q_5(t|\theta)| &= |\mu_y - ay| \\
&\leq |\mu_y| + |y| \\
|Q_6(t|\theta)| &= |\mu_y^2 - ay^2| \\
&\leq \mu_y^2 + |y^2|
\end{aligned}$$

As such:

$$|Q_4(t|\theta)| + |Q_5(t|\theta)| + |Q_6(t|\theta)| \leq M_2 + (1 + \exp(-\beta^\top x))^2 + |y| + |y^2|,$$

where M_2 is some absolute constant. As such, where M is an absolute constant:

$$||Q(t|\theta||_1 \leq ||x||_1 + 2|y| + |y^2| + \exp(-\beta^\top x)(1 + |y|) + \{1 + \exp(-\beta^\top x)\}^2 + M,$$

where $M < \infty$ by regularity condition (1). Therefore, $\mathbb{E}(\sup_{\theta\in\Theta} ||Q(t|\theta)||_1) < \infty$, and $\{t \to Q(t|\theta) : \theta \in \Theta\}$ is $\mathbb{P}$-Glivenko-Cantelli.

Condition 2: $||\Phi(\theta)||_1$ is strictly positive outside every open neighborhood of θ_0.
Following Zhao et al. (2019), we fix some $\varepsilon > 0$. If $||\beta - \beta_0||_1 > \varepsilon/M$, then it is trivial to show that $||\Phi(\theta)||_1 > 0$. Zhao et al. (2019) show that when $||\beta - \beta_0||_1 \leq \varepsilon/M$, if $|\mu_w - \mu_{w,0}| > \varepsilon/4K$, where $K = \sup_{\theta\in\Theta} |\mu| \in (0, \infty)$, then $||\Phi(\theta)||_1 > 0$. Furthermore, when $||\beta - \beta_0||_1 \leq \varepsilon/M$ and $|\mu_w - \mu_{w,0}| \leq \varepsilon/4K$ and $|\mu - \mu_0| > \varepsilon/2\mu_w$, then $||\Phi(\theta)||_1 > 0$.

Thus, we must show for the remaining 3 parameters that when $||\mu_w^2 - \mu_{w,0}^2||$, $||\mu_y - \mu_{y,0}||$, or $||\mu_y^2 - \mu_{y,0}^2||_1$ are greater than some ε, $||\Phi(\theta)||_1 > 0$. Assume $|||\beta - \beta_0||_1 \leq \varepsilon/M$. Then:

$$\begin{aligned}\left|\mathbb{E}\left\{A\exp(-\beta^\top X)^2 + A\exp(-\beta_0^\top X)^2\right\}\right| &= \left|\mathbb{E}\left\{A\exp(-2\beta^\top X) + A\exp(-2\beta_0^\top X)\right\}\right| \\ &\leq \left|\mathbb{E}\left\{\exp(-2\beta^\top X - 2\beta_0^\top X)\right\}\right| \\ &\leq 2||\beta - \beta_0||_\infty \mathbb{E}\left\{||X||_1 \exp(-t^*)\right\} \text{ for } t^* \in [\beta_0^\top X, \beta^\top X] \\ &\leq 2 \cdot \frac{\varepsilon}{64K} = \frac{\varepsilon}{32K}\end{aligned}$$

As such, if $||\mu_w^2 - \mu_{w,0}^2|| > \varepsilon/32K$:

$$||\Phi(\theta)||_1 \geq \left|\mu_w^2 - \mu_{w,0}^2 + \mathbb{E}\left\{A\exp(-\beta^\top X)^2 + A\exp(-\beta_0^\top X)^2\right\}\right| > 0 \tag{20}$$

For the final two parameters, it is worth noting that there is no dependency on the other parameter estimates. As such, regardless of whether the other parameters are smaller than some ϵ, if $||\mu_y - \mu_{y,0}||_1 > \varepsilon$:

$$\begin{aligned}||\Phi(\theta)||_1 &\geq |\mu_y - \mathbb{E}(AY) - \{\mu_{y,0} - \mathbb{E}(AY)\}| \\ &= |\mu_y - \mu_{y,0}| > 0\end{aligned} \tag{21}$$

Similarly, if $||\mu_y^2 - \mu_{y,0}^2|| > \varepsilon$

$$\begin{aligned}||\Phi(\theta)||_1 &\geq \left|\mu_y^2 - \mathbb{E}(AY^2) - \{\mu_{y,0}^2 - \mathbb{E}(AY^2)\}\right| \\ &= \left|\mu_{y,0}^2 - \mu_y\right| > 0\end{aligned} \tag{22}$$

As such, combining Equation (20), (21), (22), as well as the results from Zhao et al. (2019), we have shown that for all $\delta > 0$, $\inf\{||\Phi(\theta)||^2 : ||\theta - \theta_0||_1 > \delta\} > 0$.

Condition 3: The class of functions is $\mathbb{P}$-Donsker, and $\mathbb{E}[\{Q(T|\theta_n) - Q(T|\theta_0)\}^2] \to 0$ whenever $||\theta_n - \theta_0||_1 \to 0$.

From Zhao et al. (2019), we obtain a bound for the first three terms (i.e., $Q_1(t|\theta)$, $Q_2(t|\theta)$, and $Q_3(t|\theta)$). Furthermore, consistent with Zhao et al. (2019), for some $a, b \in \mathbb{R}$, and some constant

$C > 0$, if $a \leq C \cdot b$, then we write $a \lesssim b$. Then, for the 4th term:

$$
\begin{aligned}
&|Q_4(t|\theta_2) - Q_4(t|\theta_1)| \\
&\quad = |\mu_{w,2}^2 - a\{1+\exp(-\beta_2^\top x)\}^2 - (\mu_{w,1}^2 - a\{1+\exp(-\beta_1^\top x)\}^2| \\
&\quad \leq |\mu_{w,2}^2 - \mu_{w,1}^2| + |\{1+\exp(-\beta_2^\top x)\}^2 - \{1+\exp(-\beta_1^\top x)\}^2| \\
&\quad = |\mu_{w,2}^2 - \mu_{w,1}^2| + \Big|2\big\{\exp(-\beta_2^\top x) - \exp(-\beta_1^\top x)\big\} + \exp(-2\beta_2^\top x) - \exp(-2\beta_1^\top x)\Big| \\
&\quad \leq |\mu_{w,2}^2 - \mu_{w,1}^2| + \Big|2\big\{\exp(-\beta_2^\top x) - \exp(-\beta_1^\top x)\big\}\Big| + \Big|\exp(-2\beta_2^\top x) - \exp(-2\beta_1^\top x)\Big| \\
&\quad \lesssim |\mu_{w,2}^2 - \mu_{w,1}^2| + 2||\beta_2 - \beta_1||_2||x||_2 \sup_{\beta\in\Theta} \exp(-\beta^\top x) + ||2\beta_2 - 2\beta_1||_2||x||_2 \sup_{\beta\in\Theta} \exp(-2\beta^\top x) \\
&\quad = |\mu_{w,2}^2 - \mu_{w,1}^2| + 2||\beta_2 - \beta_1||_2||x||_2 \sup_{\beta\in\Theta} \exp(-\beta^\top x) + 2||\beta_2 - \beta_1||_2||x||_2 \sup_{\beta\in\Theta} \exp(-2\beta^\top x) \\
&\quad = |\mu_{w,2}^2 - \mu_{w,1}^2| + 2||\beta_2 - \beta_1||_2||x||_2 \sup_{\beta\in\Theta} \exp(-\beta^\top x) \left\{1 + \sup_{\beta\in\Theta} \exp(-\beta^\top x)\right\} \\
&\quad \lesssim M_4(x)\left(|\mu_{w,2}^2 - \mu_{w,1}^2| + ||\beta_2 - \beta_1||_1\right)
\end{aligned}
$$

Finally, for the 5th and 6th terms:

$$
\begin{aligned}
&|Q_5(t|\theta_2) - Q_5(t|\theta_1)| \\
&\quad = |\mu_{y,2} - ay - (\mu_{y,1} - ay)| \\
&\quad = |\mu_{y,2} - \mu_{y,1}| \\
&|Q_6(t|\theta_2) - Q_6(t|\theta_1)| \\
&\quad = |\mu_{y,2}^2 - ay^2 - (\mu_{y,1}^2 - ay^2)| \\
&\quad \leq |\mu_{y,2}^2 - \mu_{y,1}^2|
\end{aligned}
$$

Combining results with Zhao et al. (2019), we see that:

$$
||Q(t|\theta_2) - Q(t|\theta_1)||_1 = \sum_{b=1}^{6} ||Q_b(t|\theta_2) - Q_b(t|\theta_1)|| \lesssim M(x,y)||\theta_2 - \theta_1||_1
$$

Since $\mathbb{E}(M(X,Y)^2) < \infty$, we have shown that the class of functions is $\mathbb{P}$-Donsker, and furthermore, that whenever $||\theta_n - \theta_0||_1 \to 0$, $\mathbb{E}\big[(Q(t|\theta_n) - Q(t|\theta_0))^2\big] \to 0$.

Then, by invoking Kosorok (2008), Theorem 10.16:

$$
\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N\left(0, \dot{\Phi}_0^{-1}\Sigma\dot{\Phi}_0\right), \quad \text{and} \quad \sqrt{n}(\hat{\theta}^{(b)} - \theta) \xrightarrow{d} N\left(0, \dot{\Phi}_0^{-1}\Sigma\dot{\Phi}_0\right), \tag{23}
$$

As such, applying the delta method and results from Appendix C3 in Zhao et al. (2019) concludes the proof. □

Remark 1. As discussed in Section 3.3, an interval for an individual $\tilde{w}$ is not ever computed or used in practice. Instead, all such intervals are combined into a union bound, which (in combination with Theorem 1) gives rise to the following estimated confidence interval:

$$
\left[Q_{\frac{\alpha}{2}}\left\{\hat{\tau}^{(b)}(\hat{w}_1^{(b)}, \dots \hat{w}_n^{(b)}) - \widehat{\text{Bias}}_{\max}^{(b)}\right\}, Q_{1-\frac{\alpha}{2}}\left\{\hat{\tau}^{(b)}(\hat{w}_1^{(b)}, \dots \hat{w}_n^{(b)}) + \widehat{\text{Bias}}_{\max}^{(b)}\right\}\right] \tag{24}
$$

where

$$\widehat{\text{Bias}}_{\max}^{(b)} = \sqrt{1 - \text{cor}_n(\hat{w}^{(b)}, Y^{(b)} \mid Z = 0)^2}\sqrt{\frac{R^2}{1 - R^2} \cdot \text{var}_n(Y^{(b)}|Z = 0) \cdot \text{var}_n(\hat{w}^{(b)}|Z = 0)}.$$

B.3. *Proof of Theorem 3 (Weighted L_2 Analog)*

Proof. Define $\lambda := w^*/w$. From Huang (2024), Lemma 3.1, we can decompose $\text{var}(w - w^* \mid A = 1)$ as the difference in the variance of the ideal weights and the estimated weights (i.e., $\text{var}(w - w^* \mid A = 1) = \text{var}(w^* \mid A = 1) - \text{var}(w \mid A = 1)$). Then:

$$\begin{aligned}
&\text{var}(w - w^* \mid A = 1) \\
&\quad = \text{var}(w^* \mid A = 1) - \text{var}(w \mid A = 1) \\
&\quad = \text{var}(\lambda \cdot w \mid A = 1) - \text{var}(w \mid A = 1) \\
&\quad = \mathbb{E}(\lambda^2 \cdot w^2 \mid A = 1) - \underbrace{\mathbb{E}(\lambda \cdot w \mid A = 1)^2}_{\equiv \mathbb{E}(w^*|A=1)^2 = 1} - \text{var}(w \mid A = 1) \\
&\quad = \text{cov}(\lambda^2, w^2 \mid A = 1) + \mathbb{E}(\lambda^2 \mid A = 1)\mathbb{E}(w^2 \mid A = 1) - 1 - \text{var}(w \mid A = 1) \\
&\quad = \text{cov}(\lambda^2, w^2 \mid A = 1) + \mathbb{E}(\lambda^2 \mid A = 1)\text{var}(w \mid A = 1) + \mathbb{E}(\lambda^2 \mid A = 1) - 1 - \text{var}(w \mid A = 1) \\
&\quad = \text{cov}(\lambda^2, w^2 \mid A = 1) + (\mathbb{E}(\lambda^2 \mid A = 1) - 1) \cdot (\text{var}(w \mid A = 1) + 1) \\
&\quad = \text{cov}(\lambda^2, w^2 \mid A = 1) + (\mathbb{E}(\lambda^2 \mid A = 1) - 1) \cdot \mathbb{E}(w^2 \mid A = 1)
\end{aligned}$$

As such,

$$\implies \frac{\text{var}(w - w^* \mid A = 1)}{\mathbb{E}(w^2 \mid A = 1)} = \frac{\text{cov}(\lambda^2, w^2 \mid A = 1)}{\mathbb{E}(w^2 \mid A = 1)} + (\mathbb{E}(\lambda^2 \mid A = 1) - 1)$$

Re-arranging the terms:

$$\begin{aligned}
&\frac{\text{cov}(\lambda^2, w^2 \mid A = 1)}{\mathbb{E}(w^2 \mid A = 1)} + \mathbb{E}(\lambda^2 \mid A = 1) \\
&= 1 + \frac{\text{var}(w - w^* \mid A = 1)}{\mathbb{E}(w^2 \mid A = 1)} \\
&= 1 + \frac{\mathbb{E}(w^2 \mid A = 1) - \mathbb{E}(w \mid A = 1)^2}{\mathbb{E}(w^2 \mid A = 1)} \cdot \frac{R^2}{1 - R^2} \\
&= 1 + \frac{R^2}{1 - R^2} - \frac{\mathbb{E}(w \mid A = 1)^2}{\mathbb{E}(w^2 \mid A = 1)} \cdot \frac{R^2}{1 - R^2} \\
&= \frac{1}{1 - R^2} - \underbrace{\frac{\mathbb{E}(w \mid A = 1)^2}{\mathbb{E}(w^2 \mid A = 1)}}_{1/\mathbb{E}(w^2|A=1)} \cdot \frac{R^2}{1 - R^2} \\
&= \frac{1}{1 - R^2} \underbrace{\left(1 - \frac{R^2}{\mathbb{E}(w^2 \mid A = 1)}\right)}_{:=k}
\end{aligned}$$

By setting R^2, we are also setting the value for $\frac{\text{cov}(\lambda^2, w^2|A=1)}{\mathbb{E}(w^2|A=1)} + \mathbb{E}(\lambda^2 \mid A = 1)$.

We now re-write $\frac{\text{cov}(\lambda^2, w^2|A=1)}{\mathbb{E}(w^2|A=1)} + \mathbb{E}(\lambda^2 \mid A=1)$ as a weighted sum:

$$
\begin{aligned}
&\mathbb{E}(\lambda^2 \mid A=1) + \frac{\text{cov}(\lambda^2, w^2 \mid A=1)}{\mathbb{E}(w^2 \mid A=1)} \\
=&\mathbb{E}\left(\lambda^2 \mid A=1\right) + \frac{1}{\mathbb{E}(w^2)}\mathbb{E}\left[\{\lambda^2 - \mathbb{E}(\lambda^2 \mid A=1)\}\{w^2 - \mathbb{E}(w^2 \mid A=1)\} \mid A=1\right] \\
=&\mathbb{E}\left(\lambda^2 \mid A=1\right) + \frac{\mathbb{E}[\lambda^2\{w^2 - \mathbb{E}(w^2 \mid A=1)\} \mid A=1]}{\mathbb{E}(w^2 \mid A=1)} \\
&- \mathbb{E}(\lambda^2 \mid A=1) \cdot \underbrace{\frac{\mathbb{E}\{w^2 - \mathbb{E}(w^2 \mid A=1) \mid A=1\}}{\mathbb{E}(w^2 \mid A=1)}}_{:=0} \\
=&\mathbb{E}\left[\lambda^2 \left\{1 + \frac{w^2 - \mathbb{E}(w^2 \mid A=1)}{\mathbb{E}(w^2 \mid A=1)}\right\} \middle| A=1\right] \\
=&\mathbb{E}\{\lambda^2 \nu(w) \mid A=1\},
\end{aligned}
$$

where $\nu(w) := w^2/\mathbb{E}(w^2 \mid A=1)$. As such, we can define the $L_{2,w}$ norm as follows:

$$
||\lambda||_{2,w}^2 := \begin{cases} \mathbb{E}\{\lambda^2 \cdot \nu(w) \mid A=1\} & \text{if var}(w) > 0 \\ \infty & \text{else} \end{cases}
$$

We will show that $L_{2,w}$ meets the criteria for being a semi-norm.

1. Triangle Inequality: $||\lambda_1 + \lambda_2||_{2,w} \leq ||\lambda_1||_{2,w} + ||\lambda_2||_{2,w}$

$$
\begin{aligned}
&||\lambda_1 + \lambda_2||_{2,w}^2 \\
&\quad = \mathbb{E}\left\{(\lambda_1 + \lambda_2)^2 \cdot \nu(w) \mid A=1\right\} \\
&\quad = \mathbb{E}\left\{\lambda_1^2 \nu(w) \mid A=1\right\} + \mathbb{E}\left\{\lambda_2^2 \nu(w) \mid A=1\right\} + 2\mathbb{E}\left\{\lambda_1 \lambda_2 \nu(w) \mid A=1\right\} \\
&\quad = \left(||\lambda_1||_{2,w} + ||\lambda_2||_{2,w}\right)^2
\end{aligned}
$$

2. Absolute homogeneity:

$$
\begin{aligned}
||k \cdot \lambda||_{2,w} &= \sqrt{\mathbb{E}\{(k \cdot \lambda)^2 \cdot \nu(w) \mid A=1\}} \\
&= k\sqrt{\mathbb{E}[\lambda^2 \cdot \nu(w) \mid A=1]} \\
&= k \cdot ||\lambda||_{2,w}
\end{aligned}
$$

While we have assumed that w and w^* are centered at mean 1, this assumption is not necessary for the results of this theorem. More specifically, because both w and w^* are centered at the same value, the scaling factor used to normalize both sets of weights will be identical. Because λ represents a multiplicative error, the scaling factor will cancel out.

B.4. *Example 1*

Proof. The multiplicative error between w_i^* and w_i is written as:

$$
\frac{w_i^*}{w_i} = \frac{\exp({\gamma^*}^\top X_i + \beta U_i)}{\exp(\gamma^\top X_i)} = \exp\{(\gamma^* - \gamma)^\top X_i + \beta U_i\},
$$

and $\hat{\Lambda}$ is defined as the maximum:

$$\hat{\Lambda} = \max_{1 \leq i \leq n} \exp\{|(\gamma^* - \gamma)^\top X_i + \beta U_i|\}$$

We will show that $\mathbb{E}(\Lambda) \to \infty$, as $n \to \infty$. To begin, define V_i as:

$$V_i := (\gamma^* - \gamma)^\top X_i + \beta U_i$$

Because X_i and U_i are normally distributed, V_i will be normally distributed, with mean 0, and variance $\nu^2 := (\gamma^* - \gamma)^\top + \beta^2$. Let $V^{(1)}, ..., V^{(n)}$ be the ordered set of V such that $V^{(1)} \leq ... \leq V^{(n)}$. Without loss of generality, assume $|V^{(n)}| \geq |V^{(1)}|$. Then, $\mathbb{E}(\hat{\Lambda}) = \mathbb{E}\{\exp(|V^{(n)}|)\}$. Using Jensen's inequality, the expectation of $\hat{\Lambda}$ may be lower bounded:

$$\mathbb{E}(\hat{\Lambda}) = \mathbb{E}\{\exp(|V^{(n)}|)\} \geq \exp\{\mathbb{E}(|V^{(n)}|)\}$$

Then, we may invoke a well-studied result that for any set of n normally distributed random variables (Wainwright (2019), § 2, pg. 53):

$$\lim_{n \to \infty} \frac{\mathbb{E}(V^{(n)})}{\sqrt{2\nu^2 \log(n)}} = 1$$

Because $\mathbb{E}(|V^{(n)}|) \geq \mathbb{E}(V^{(n)})$, as $n \to \infty$, $\mathbb{E}(\hat{\Lambda}) \to \infty$. □

B.5. *Proof of Example 2*

Proof. Because $[X_i, U_i] \overset{iid}{\sim} MVN(0, I)$, w_i and w_i^* both are lognormal random variables, by definition, the variance of w_i^* is: $\{\exp({\gamma^*}^\top \gamma^* + \beta^2) - 1\} \cdot \exp\{\hat{\gamma}^{*\top} \gamma + \beta^2\}$, and similarly, the variance of w_i is: $\{\exp(\gamma^\top \gamma) - 1\} \cdot \exp(\gamma^\top \gamma)$. Then, the result of the example immediately follows, using $R^2 := 1 - \text{var}(w \mid A = 1)/\text{var}(w^* \mid A = 1)$.

B.6. *Proof of Corollary 1*

Proof. We formalize a condition under which the variance-based sensitivity model will result in narrower bounds than the marginal L_∞ sensitivity model when considering the true distribution of weights w^*, and the corresponding set $\varepsilon(\Lambda_0)$ and $\sigma(R_0^2)$ that contains it. We derive a sufficient condition for the variance-based sensitivity model to produce strictly narrower bounds.

To begin, the length of the point estimate bounds under the variance-based sensitivity model $\sigma(R_0^2)$ is equal to two times the bias bound:

$$\begin{aligned} &\max_{\tilde{w} \in \sigma(R_0^2)} \tau(\tilde{w}) - \min_{\tilde{w} \in \sigma(R_0^2)} \tau(\tilde{w}) \\ &\qquad = 2 \cdot \sqrt{1 - \text{cor}(w_i, Y_i \mid A_i = 1)^2} \cdot \sqrt{\frac{R_0^2}{1 - R_0^2} \cdot \text{var}(w_i \mid A_i = 1) \cdot \text{var}(Y_i \mid A_i = 1)} \end{aligned}$$

The length of the population point estimate bounds under the marginal sensitivity model is represented by $\psi(\Lambda_0)$:

$$\psi(\Lambda_0) := \max_{\tilde{w} \in \varepsilon(\Lambda_0)} \frac{\mathbb{E}(Y\tilde{w} \mid Z = 0)}{\mathbb{E}(\tilde{w} \mid Z = 0)} - \min_{\tilde{w} \in \varepsilon(\Lambda_0)} \frac{\mathbb{E}(Y\tilde{w} \mid Z = 0)}{\mathbb{E}(\tilde{w} \mid Z = 0)}$$

Thus, we want to solve for the R^2 value such that the following inequality holds:

$$2 \cdot \sqrt{1 - \text{cor}(w_i, Y_i \mid A_i = 1)^2} \cdot \sqrt{\frac{R_0^2}{1 - R_0^2} \cdot \text{var}(w_i \mid A_i = 1) \cdot \text{var}(Y_i \mid A_i = 1)} \leq \psi(\Lambda_0)$$

Solving for the R_0^2 value:

$$\begin{aligned}
\frac{R_0^2}{1-R_0^2} &\leq \frac{\psi(\Lambda_0)^2/4}{\{1-\text{cor}(w_i, Y_i \mid A_i=1)^2\} \cdot \text{var}(w_i \mid A_i=1) \cdot \text{var}(Y_i \mid A_i=1)} \\
R_0^2 &\leq \frac{\psi(\Lambda_0)^2/4\{1-\text{cor}(w_i, Y_i \mid A_i=1)^2\}\text{var}(w_i \mid A_i=1)\text{var}(Y_i \mid A_i=1)}{1+\psi(\Lambda_0)^2/4\{1-\text{cor}(w_i, Y_i \mid A_i=1)^2\}\text{var}(w_i \mid A_i=1)\text{var}(Y_i \mid A_i=1)} \\
&= \underbrace{\frac{\psi(\Lambda_0)^2}{4\{1-\text{cor}(w_i, Y_i \mid A_i=1)^2\}\text{var}(w_i \mid A_i=1)\text{var}(Y_i \mid A_i=1)+\psi(\Lambda_0)^2}}_{g(\psi(\Lambda_0);Y_i,w_i)},
\end{aligned}$$

We will use results from Resnick (2008), who show that for a sequence of random variables drawn i.i.d., the maximum of the sequence will converge in probability towards the upper bound of the support. We provide the derivation for completeness. Let $\{W_i\}_{i=1}^n$ be drawn i.i.d. from a distribution F. Then by i.i.d.:

$$\Pr(W^{(n)} \leq w) = \Pr\left(\bigcap_{i=1}^{n}\{W_i \leq w\}\right) = F_Y^n(w),$$

where $W^{(1)} \leq ... \leq W^{(n)}$. Then define $w_0 = \sup\{w : F_W(w) < 1\}$. Then for any $w' < w_0$, $P(W^{(n)} \leq w') = F_W^n(w') \longrightarrow 0$, since $F_W(w') < 1$. As such, $W^{(n)}$ converges in probability, to w_0. We note that the same result can be applied for the minima of $\{W_i\}_{i=1}^n$ by using $-\{W_i\}_{i=1}^n$.

Now, define $\lambda_i := w_i^*/w_i$. We have restricted the set of plausible λ_i such that $\liminf\{\lambda : 1-F_\lambda(\lambda) < 1\} = 0$, or $\limsup\{\lambda : F_\lambda(\lambda) < 1\} \to \infty$. First consider the setting for $\liminf\{\lambda : 1-F_\lambda(\lambda) < 1\} = 0$. We can apply the results from above to show that for a sequence of random $\lambda_1, ..., \lambda_n$, the minimum of the sequence will converge in probability towards zero. Because $\Lambda = \max_{1\leq i\leq n}\{\lambda_i, 1/\lambda_i\}$, this implies that Λ will diverge in probability towards infinity. Similarly, for $\limsup\{\lambda : F_\lambda(\lambda) < 1\} \to \infty$, the maximum of the sequence will diverge in probability towards infinity, which implies that Λ will diverge in probability towards infinity. As such, the marginal sensitivity model will be invalid.

Applying Continuous Mapping Theorem and sample boundedness, the length of the point estimate bounds under the marginal sensitivity model ($\psi(\Lambda_0)$) will be equal to the range of the observed control outcomes. Thus, if the outcomes Y_i are unbounded as well (i.e., $F_Y(y) < 1$ for all $y \in \mathbb{R}$.), $\psi(\Lambda_0)$ will diverge in probability to infinity.

Thus, we have shown that $\psi(\Lambda_0)$ will diverge in probability to infinity. Applying Continuous Mapping Theorem again, $g(\psi(\Lambda_0); Y_i, w_i) \xrightarrow{p} 1$. Because the R^2 parameter is less than 1 by definition, for sufficiently large n, the variance-based sensitivity model will produce narrower bounds, which concludes the proof.

REFERENCES

BRUNS-SMITH, D. & ZHOU, A. (2023). Robust fitted-q-evaluation and iteration under sequentially exogenous unobserved confounders. *arXiv preprint arXiv:2302.00662* .

CARNEGIE, N. B., HARADA, M. & HILL, J. L. (2016). Assessing sensitivity to unmeasured confounding using a simulated potential confounder. *Journal of Research on Educational Effectiveness* **9**, 395–420.

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

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

HARTMAN, E., HAZLETT, C. & STERBENZ, C. (2021). Kpop: A kernel balancing approach for reducing specification assumptions in survey weighting. *arXiv preprint arXiv:2107.08075* .

HARTMAN, E. & HUANG, M. (2024). Sensitivity analysis for survey weights. *Political Analysis* **32**, 1–16.

HONG, G., YANG, F. & QIN, X. (2021). Did you conduct a sensitivity analysis? a new weighting-based approach for evaluations of the average treatment effect for the treated. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **184**, 227–254.

HSU, J. Y. & SMALL, D. S. (2013). Calibrating sensitivity analyses to observed covariates in observational studies. *Biometrics* **69**, 803–811.

HUANG, M. (2024). Sensitivity analysis in the generalization of experimental results. *Journal of the Royal Statistical Society: Series A (Statistics in Society) (Forthcoming)* .

HUANG, M. Y., VEGETABILE, B. G., BURGETTE, L. F., SETODJI, C. & GRIFFIN, B. A. (2022). Higher moments matter for optimal balance weighting in causal estimation. *Epidemiology (Cambridge, Mass.)* .

JORDAN, M. I., WANG, Y. & ZHOU, A. (2022). Data-driven influence functions for optimization-based causal inference. *arXiv preprint arXiv:2208.13701* .

KALLUS, N. & ZHOU, A. (2020). Confounding-robust policy evaluation in infinite-horizon reinforcement learning. *Advances in neural information processing systems* **33**, 22293–22304.

KOSOROK, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer.

MIRATRIX, L. W., SEKHON, J. S. & YU, B. (2013). Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**, 369–396.

PIMENTEL, S. D. & KELZ, R. R. (2020). Optimal tradeoffs in matched designs comparing us-trained and internationally trained surgeons. *Journal of the American Statistical Association* **115**, 1675–1688.

RESNICK, S. I. (2008). *Extreme values, regular variation, and point processes*, vol. 4. Springer Science & Business Media.

ROSENBAUM, P. R., ROSENBAUM, P. & BRISKMAN (2010). *Design of observational studies*, vol. 10. Springer.

ROSENBAUM, P. R. & SILBER, J. H. (2009). Amplification of sensitivity analysis in matched observational studies. *Journal of the American Statistical Association* **104**, 1398–1405.

SHEN, C., LI, X., LI, L. & WERE, M. C. (2011). Sensitivity analysis for causal inference using inverse probability weighting. *Biometrical Journal* **53**, 822–837.

SORIANO, D., BEN-MICHAEL, E., BICKEL, P. J., FELLER, A. & PIMENTEL, S. D. (2023). Interpretable sensitivity analysis for balancing weights. *Journal of the Royal Statistical Society Series A: Statistics in Society* **186**, 707–721.

WAINWRIGHT, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, vol. 48. Cambridge University Press.

ZHANG, Y. & ZHAO, Q. (2024). Sharp bounds and semiparametric inference in l_∞-and l_2-sensitivity analysis for observational studies. *arXiv preprint arXiv:2211.04697* .

ZHAO, Q., SMALL, D. S. & BHATTACHARYA, B. B. (2019). Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **81**, 735–761.