Supplementary Materials for "Sensitivity Analysis for Survey Weights"

Erin Hartman and Melody Huang

A Glossary

Table A.1: Glossary for Notation and Sensitivity Parameters

Notation	
S_i	Binary survey inclusion indicator
Y	Outcome of interest
$\mathcal{S}$	Units in the survey ($i : S_i = 1$)
μ	Population mean of Y
$\mathbf{X}$	Set of observable auxiliary variables measured in both the survey and target population
$\phi(\cdot)$	A feature mapping of a given set of covariates
V	Partially observed variable measured in the survey but not the target population
U	Fully unobserved variable
w	Weights estimated using $\mathbf{X}$
w^*	Ideal weights estimated using partially or fully observed confounders
Sensitivity Parameters and Values	
ε	The error in the estimated weights ($w - w^*$)
R^2_ε	Variation in ideal weights explained by the error, ε
$\rho_{\varepsilon,Y}$	Correlation between the error in the weights and the outcome
$\text{var}_\mathcal{S}(w)$	Variance of the estimated weights
$\text{var}_\mathcal{S}(Y)$	Variance of the outcome Y in the survey
b^*	Substantive threshold against which to evaluate robustness
RV_{b^*}	Robustness value at substantive meaningful b^*

B Extended Discussion

B.1 Calibration Weights

In general form, calibration weights are defined in definition B.1 below.

Definition B.1 (Calibration Weights)
Calibration weights w are defined as:

$$\min_w \quad D(w, q) \tag{1}$$

$$\textit{subject to} \quad \sum_{i:S_i=1} w_i f(X_i) = T, \tag{2}$$

$$\sum_{i:S_i=1} w_i = 1, \;\; \textit{and} \;\; 0 \leq w_i \leq 1. \tag{3}$$

where q_i refers to a reference or base weight, commonly defined as unity or using survey design weights, and $D(\cdot,\cdot)$ corresponds to a distance metric which is usually greater for weights that diverge more severely from the base weight.[1] There are many ways to encode moment constraints, $f(X_i)$, with common methods such as "raking" typically using marginal population averages defined using observable characteristics, i.e. $f(X_i) = \frac{1}{N}\sum X_i$. The moment constraints defined by $f(X_i)$ in the target population are encoded in T. See Hartman, Hazlett, and Sterbenz (2021) for a thorough discussion of the choice of population moment constraints in calibration.

Common types of survey weighting correspond to different distance metrics $D(\cdot,\cdot)$, and are closely related to generalized regression estimation (Särndal, 2007). We rely on $D(w,q) = \sum_{i:R=1} w_i log(w_i/q_i)$ as commonly employed in "raking" methods and entropy balancing (Hainmueller, 2012). Other common weighting methods, such as post-stratification and generalized regression estimation, map to alternative distance metrics (Deville and Särndal, 1992). A full review of calibration is beyond the scope of this paper, but readers are directed to Kalton and Flores-Cervantes (2003), Särndal (2007), Wu and Lu (2016), Caughey et al. (2020),or Hartman, Hazlett, and Sterbenz (2021) for a more thorough treatment. The constraints in Equation (3) jointly ensure the weights fall on the simplex; removing this constraint allows for extrapolation beyond the observed data.

An alternative approach to weighting includes inverse propensity score weighting in which weights $w_i \propto \frac{1}{\Pr(S_i=1|\mathbf{X})}$ (Little and Rubin, 2002). These weights make the sample representative, in expectation, on observed characteristics whereas calibration will enforce a break in the relationship between observed characteristics and S_i in every sample (Yiu and Su, 2018). Calibration weights converge asymptotically, with appropriate loss functions, to inverse propensity weights; for example weights estimated using logistic regression are asymptotically equivalent to raking weights using the same covariates (Ben-Michael et al., 2021, e.g. see). Raking, and most calibration estimators, are asymptotically equivalent to generalized regression estimators (Deville and Särndal, 1992).

B.1.1 Sensitivity Analysis for Partial Confounding with Calibration Weights

$$\begin{aligned} \min_{w} \quad & D(w,q) \\ \text{subject to} \quad & \sum_{i\in\mathcal{S}} w_i f(\mathbf{X}_i) = T, \\ & \sum_{i\in\mathcal{S}} w_i f(V_i) = T_V, \\ & \sum_{i\in\mathcal{S}} w_i = 1, \text{ and } 0 \le w_i \le 1. \end{aligned} \tag{4}$$

where Equation 4 is the new constraint for V_i. By varying T_V, which contains the population moment constraint defined by $f(V_i)$, across a plausible range, researchers can re-estimate the survey weights over the range and re-estimate the population mean to assess how the point estimate varies as T_V changes. When using raking on the margins, $T_V = \mathbb{E}(V_i)$. imbalance on a standardized scale (e.g., in terms of z-scores) to help provide a more intuitive understanding of what a large or small degree of imbalance is for a given covariate.

[1]Formally, $D(\cdot,\cdot)$ is a divergence, not a distance, since it is often not always symmetric in its arguments.

B.2 Formal benchmarking with relative confounding strength

We now introduce formal benchmarking with relative confounding strength. To begin, define k_σ and k_ρ as follows:

$$k_\sigma = \frac{\text{var}_\mathcal{S}(\varepsilon_i)/\text{var}_\mathcal{S}(w_i^*)}{\text{var}_\mathcal{S}(\varepsilon_i^{-(j)})/\text{var}_\mathcal{S}(w_i^*)}, \qquad k_\rho = \frac{\text{cor}_\mathcal{S}(\varepsilon_i, Y_i)}{\text{cor}_\mathcal{S}(\varepsilon_i^{-(j)}, Y_i)}, \tag{5}$$

where the blue terms are equivalent to the unobserved terms from Theorem 3.1. k_σ compares the relative residual imbalance after accounting for $\mathbf{X}_i$ in the unobserved confounder to the imbalance after accounting for $\mathbf{X}_i^{-(j)}$ for the covariate $\mathbf{X}_i^{(j)}$. Consider, for example, if educational attainment is used to benchmark the fully unobserved confounder of late decision in candidate choice. If researchers believe that late decision in candidate choice, even after adjusting for the other covariates, is more imbalanced than education, after adjusting for all covariates except for education, then $k_\sigma > 1$. However, if researchers believe that there is less residual imbalance in the confounder than the observed covariate, $k_\sigma < 1$. Setting $k_\sigma = 1$ evaluates the case in which the omitted confounder has the same level of residual imbalance as the benchmarked covariate.

Similarly, k_ρ compares how correlated the outcome and the imbalance in the unobserved confounder are, relative to the correlation between the outcome and the imbalance in the benchmarked covariate. If $k_\rho > 1$, then this implies that the unobserved confounder's imbalance can explain more variation in the outcome variable than the benchmarked covariate.

Given k_σ and k_ρ, we can re-write the sensitivity parameters as functions of k_σ and k_ρ and observable quantities:

$$R^2_\varepsilon = \frac{k_\sigma \cdot R_\varepsilon^{2-(j)}}{1 + k_\sigma \cdot R_\varepsilon^{2-(j)}}, \qquad \rho_{\varepsilon,Y} = k_\rho \cdot \rho_{\varepsilon,Y}^{-(j)}, \tag{6}$$

in which $R_\varepsilon^{2-(j)}$ and $\rho_{\varepsilon,Y}$ are defined in Equation 5.

Similar to the MRCS, researchers can solve for the minimum k_σ and k_ρ value for each observed covariate (or subset of covariates) which would result in the parameters taking on the robustness value (i.e., $\rho^2_{\varepsilon,Y} = R^2_\varepsilon = RV_q$). We denote this as k_σ^{min} and k_ρ^{min}.

C Proofs and Derivations

C.1 Corollary C.1

To provide some intuition for ε, consider the following corollary.

Corollary C.1 (Error Decomposition for Inverse Propensity-Score Weights) *The ε_i for IPW weights estimated using $\phi(\mathbf{X}_i)$ can be written as follows:*

$$\varepsilon_i = \underbrace{\frac{P(S_i = 1)}{P(S_i = 1 \mid \phi(\mathbf{X}_i))}}_{\text{Estimated weights } w_i} \underbrace{\left(1 - \frac{P(U_i \mid \phi(\mathbf{X}_i))}{P(U_i \mid \phi(\mathbf{X}_i), S_i = 1)}\right)}_{\text{Residual imbalance in } U_i}$$

More specifically, ε_i is a function of the estimated weights (left), and the residual imbalance in the omitted confounder, after accounting for $\phi(\mathbf{X}_i)$ (right).

When using inverse propensity weights, the estimated weights and the ideal weights are :

$$w_i = \frac{P(S_i = 1)}{P(S_i = 1 \mid \phi(\mathbf{X}_i))} \qquad w_i^* = \frac{P(S_i = 1)}{P(S_i = 1 \mid \phi(\mathbf{X}_i), U_i)}$$

Then the error term is:

$$\begin{aligned}
\varepsilon_i &= w_i - w_i^* \\
&= \frac{P(S_i = 1)}{P(S_i = 1 \mid \phi(\mathbf{X}_i))} - \frac{P(S_i = 1)}{P(S_i = 1 \mid \phi(\mathbf{X}_i), U_i)} \\
&= P(S_i = 1) \cdot \left(\frac{1}{P(S_i = 1 \mid \phi(\mathbf{X}_i))} - \frac{1}{P(S_i = 1 \mid \phi(\mathbf{X}_i), U_i)} \right)
\end{aligned}$$

Applying Bayes' Rule:

$$\begin{aligned}
&= P(S_i = 1) \cdot \left(\frac{1}{P(S_i = 1|\phi(\mathbf{X}_i))} - \frac{P(U_i|\phi(\mathbf{X}_i))}{P(U_i|\phi(\mathbf{X}_i), S_i = 1) \cdot P(S_i = 1|\phi(\mathbf{X}_i))} \right) \\
&= \frac{P(S_i = 1)}{P(S_i = 1|\phi(\mathbf{X}_i))} \cdot \left(1 - \frac{P(U_i \mid \phi(\mathbf{X}_i)}{P(U_i \mid \phi(\mathbf{X}_i), S_i = 1)} \right)
\end{aligned}$$

In cases when the omitted confounder U_i is binary, we can rewrite the expression as a function of the conditional average of U_i:

$$\varepsilon_i = \frac{P(S_i = 1)}{P(S_i = 1|\phi(\mathbf{X}_i))} \cdot \left(1 - \frac{\mathbb{E}(U_i \mid \phi(\mathbf{X}_i)}{\mathbb{E}(U_i \mid \phi(\mathbf{X}_i), S_i = 1)} \right)$$

C.2 Theorem 3.1

$$\begin{aligned}
\text{Bias}(\hat{\mu}) &= \mathbb{E}(\hat{\mu}) - \mu \\
&= \mathbb{E}\left(\sum_{i \in \mathcal{S}} w_i Y_i \right) - \mu \\
&= \mathbb{E}\left(\sum_{i \in \mathcal{S}} w_i Y_i \right) - \mathbb{E}\left(\sum_{i \in \mathcal{S}} w_i^* Y_i \right) && \text{by conditional ignorability assumption} \\
&= \mathbb{E}_{\mathcal{S}}(w_i Y_i) - \mathbb{E}_{\mathcal{S}}(w_i^* Y_i) && \text{by linearity in expectation} \\
&= \mathbb{E}_{\mathcal{S}}((w_i - w_i^*) \cdot Y_i) \\
&= \mathbb{E}_{\mathcal{S}}((w_i - w_i^*) \cdot Y_i) - \mathbb{E}_{\mathcal{S}}(w_i - w_i^*)\mathbb{E}_{\mathcal{S}}(Y_i) && \text{by construction, } \mathbb{E}(w_i) = \mathbb{E}(w_i^*) \\
&= \mathbb{E}_{\mathcal{S}}(\varepsilon_i \cdot Y_i) - \mathbb{E}_{\mathcal{S}}(\varepsilon_i) \cdot \mathbb{E}_{\mathcal{S}}(Y_i) \\
&= \text{cov}_{\mathcal{S}}(\varepsilon_i, Y_i) \\
&= \text{cor}_{\mathcal{S}}(\varepsilon_i, Y_i) \cdot \sqrt{\text{var}_{\mathcal{S}}(Y_i) \cdot \text{var}_{\mathcal{S}}(\varepsilon_i)} \qquad (7)
\end{aligned}$$

Defining $R^2 = \frac{\text{var}_{\mathcal{S}}(\varepsilon_i)}{\text{var}_{\mathcal{S}}(w_i^*)}$ and noting that $\text{var}_{\mathcal{S}}(w_i^*)$ can be written as $\text{var}_{\mathcal{S}}(w_i) + \text{var}_{\mathcal{S}}(\varepsilon_i)$:

$$\begin{aligned}
\text{var}_{\mathcal{S}}(\varepsilon_i) &= \text{var}_{\mathcal{S}}(w_i^*) - \text{var}_{\mathcal{S}}(w_i) \\
&= \text{var}_{\mathcal{S}}(w_i^*) \underbrace{\left(1 - \frac{\text{var}_{\mathcal{S}}(w_i)}{\text{var}_{\mathcal{S}}(w_i^*)} \right)}_{\equiv R_\varepsilon^2}
\end{aligned}$$

Then, noting that $\text{var}_{\mathcal{S}}(w_i^*) = \text{var}_{\mathcal{S}}(w_i)/(1 - R_\varepsilon^2)$

$$= \text{var}_{\mathcal{S}}(w_i) \cdot \frac{R_\varepsilon^2}{1 - R_\varepsilon^2} \qquad (8)$$

Substituting Equation (8) into Equation (7) recovers the results from Theorem 3.1.

Remark. It is worth noting that the bias formula and subsequent sensitivity analyses derived are implicitly treating the estimated weights w and the ideal weights w^* as fixed. This is mathematically equivalent to looking at the asymptotic bias of the weighted estimators (see Huang (2022)). However, we can extend the same framework and bias expressions for the finite-sample case, in which we define the bias as the error in the estimated weights $\hat{w}$ and a set of oracle weights $\hat{w}^*$ (i.e., see Cinelli and Hazlett (2020) and Soriano et al. (2021) as examples). We provide extensions for relaxing this assumption in Section 6.

C.3 Corollary 3.1

The results of Corollary 3.1 follow results from Huang (2022), who show that for inverse propensity weights, projecting the ideal weights w_i^* into the space of observed covariates $\phi(\mathbf{X}_i)$ recovers the estimated weights w_i (i.e., $\mathbb{E}(w_i^* \mid \phi(\mathbf{X}_i)) = w_i$). As such:

$$\begin{aligned}
\text{var}_{\mathcal{S}}(\varepsilon_i) &= \text{var}_{\mathcal{S}}(w_i - w_i^*) \\
&= \text{var}_{\mathcal{S}}(w_i) + \text{var}_{\mathcal{S}}(w_i^*) - 2\text{cov}_{\mathcal{S}}(w_i, w_i^*) \\
&= \text{var}_{\mathcal{S}}(w_i) + \text{var}_{\mathcal{S}}(w_i^*) - 2\left(\mathbb{E}_{\mathcal{S}}(w_i \cdot w_i^*) - \mathbb{E}_{\mathcal{S}}(w_i)\mathbb{E}_{\mathcal{S}}(w_i^*)\right)
\end{aligned}$$

Because $\mathbb{E}_{\mathcal{S}}(w_i) = \mathbb{E}_{\mathcal{S}}(w_i^*) = 1$ and by Law of Iterated Expectation:

$$= \text{var}_{\mathcal{S}}(w_i) + \text{var}_{\mathcal{S}}(w_i^*) - 2\left(\mathbb{E}_{\mathcal{S}}(\mathbb{E}_{\mathcal{S}}(w_i \cdot w_i^* | \phi(\mathbf{X}_i))) - \mathbb{E}_{\mathcal{S}}(w_i)^2\right)$$

Because $\mathbb{E}_{\mathcal{S}}(w_i^* | \phi(\mathbf{X}_i)) = w_i$:

$$\begin{aligned}
&= \text{var}_{\mathcal{S}}(w_i) + \text{var}_{\mathcal{S}}(w_i^*) - 2\left(\mathbb{E}_{\mathcal{S}}(w_i^2) - \mathbb{E}_{\mathcal{S}}(w_i)^2\right) \\
&= \text{var}_{\mathcal{S}}(w_i) + \text{var}_{\mathcal{S}}(w_i^*) - 2\text{var}_{\mathcal{S}}(w_i) \\
&= \text{var}_{\mathcal{S}}(w_i^*) - \text{var}_{\mathcal{S}}(w_i)
\end{aligned}$$

As such, several immediate properties follow. First, by definition of variance, $\text{var}_{\mathcal{S}}(\varepsilon_i) = \text{var}_{\mathcal{S}}(w_i^*) - \text{var}_{\mathcal{S}}(w_i) \geq 0$, which implies that $\text{var}_{\mathcal{S}}(w_i^*) \geq \text{var}_{\mathcal{S}}(w_i)$. Second, defining $R^2_\varepsilon := \text{var}_{\mathcal{S}}(\varepsilon_i)/\text{var}_{\mathcal{S}}(w_i^*)$, we see that R^2_ε is naturally bound on the interval [0,1].

D Partially Observed Confounders

D.1 Sensitivity Analysis for Partially Observed Confounders with Inverse Propensity Weighting

We now consider an example where V is binary and researchers estimate inverse propensity score weights. In this case, we can rewrite the results of Corollary C.1 as follows.

Example D.1 (IPW Weights with Binary V.) *When V is binary, we can rewrite the error term, ϵ, for IPW weights from Corollary C.1 as a function of sample and population proportions of V conditional on $\phi(\mathbf{X})$ as*

$$\varepsilon_i = \underbrace{\frac{P(S_i = 1)}{P(S_i = 1 \mid \phi(\mathbf{X})_i)}}_{:= w_i} \cdot \left(1 - \frac{\mathbb{E}(V_i \mid \phi(\mathbf{X})_i)}{\mathbb{E}(V_i \mid S_i = 1, \phi(\mathbf{X})_i)}\right).$$

The key takeaway from Example D.1 is that, because V is observed across the sample, the denominator $\mathbb{E}(V_i \mid S_i = 1, \phi(\mathbf{X})_i)$ can be directly estimated from the observed data, reducing the problem to a single sensitivity parameter. The sensitivity analysis requires researchers to posit reasonable values of the unobservable $\mathbb{E}(V_i \mid \phi(\mathbf{X})_i)$, denoted in blue, estimate ε, and adjust the point estimate via Theorem 3.1. This is similar to the sensitivity analysis for partial confounders proposed by Nguyen et al. (2017), which uses an outcome model based approach.

Calibration only requires that researchers reason about the unconditional population quantity, since it will implicitly solve for the weights that meet the balancing constraints without needing to know the conditional expectations. However, in some simple cases it may be possible for researchers to reason about the conditional means, for example if the estimated weights only include a handful of covariates with a limited number of levels. In general, though, this is a difficult quantity to reason about without rich auxiliary data, and if the researcher had such rich auxiliary data, they would likely not have a partially observed confounder.

D.2 Detection of Partially Observed Confounders

While researchers can often posit partially observed confounders based on theoretical considerations, sometimes researchers are unsure if there are observed covariates that may be partially observed confounders. In this section we propose a method for detecting partially observed confounders that can be used for determining the existence of a partially observed confounder or for confirming if a theoretically relevant covariate is a partially observed confounder. These variables can then be used in the one parameter sensitivity analysis described in Section 4.

We consider the data setting in which researchers can posit a sampling set–all variables related to sample selection–either through data driven methods or from theory, but this set is only partially observed, i.e. every variable is measured in the survey sample and all but one variable is measured in the target population. Throughout this section, we will denote $\mathbf{X}^S$ as the sampling set, and $\mathbf{X}$ as the entire set of observed variables. We assume, without loss of generality, that there is only one partially observed covariate, V and return to this assumption below.[2] Because V is not measured in the target population, we cannot rely on an appropriate feature expansion $\phi(\mathbf{X}^S)$ of the full sampling set to construct survey weights.

We adapt an estimation technique from Egami and Hartman (2021). The method, described in detail in Appendix D.2, uses the survey data to estimate a Markov Random Field (Yang et al., 2014) (MRF), or an undirected graph, over the variables in $\{\mathbf{X}^S, \mathbf{X} \not\subset \mathbf{X}^S, Y\}$. We then determine if there is a set of variables that render the outcome and all variables in the sampling set $\mathbf{X}^S$ conditionally ignorable, thus justifying Assumption 1, but which does not include the partially observed confounder. If no such set exists, then V is a partially observed confounder. In this case, we suggest researchers use the sensitivity analysis for partially observed confounders described in Section 4. When there are multiple partially observed confounders, the algorithm can also be updated to determine the set that contains the smallest number of partial confounders as an additional constraint.

To begin, we first formalize the idea that if a separating set consisting of fully observed covariates exist, then this set may be used in lieu of the partially observed selection set to identify the population mean.

Theorem D.1 (Identification of Population Mean from Survey Sample (Egami and Hartman, 2021)) *When $\mathbf{X}^S$ is known and partially observed in the survey sample, for units*

[2] Formally, this means $\mathbf{X}^S \subseteq \{\mathbf{X} \cup V\}$ and $\{\mathbf{X} \cap V\} = \varnothing$, with $\mathbf{X}^S$ measured for all units with $S_i = 1$. When all covariates in $\mathbf{X}^S$ are fully observed, i.e. $\mathbf{X}^S \subseteq \mathbf{X}$, our proposal reduces to a variable selection method for constructing survey weights.

$S_i = 1$, *consider a set* $\mathbf{W}$ *that is fully observed, then under Assumption 1 (replacing* $\mathbf{X}$ *with* $\mathbf{W}$*):*

$$Y_i \perp\!\!\!\perp \mathbf{X}_i^S \mid \mathbf{W}_i, S_i = 1 \implies Y_i \perp\!\!\!\perp S_i \mid \mathbf{W}_i \tag{9}$$

Proof:

$$
\begin{aligned}
& Y_i \perp\!\!\!\perp \mathbf{X}_i^S \mid \mathbf{W}_i, S_i = 1 && && (10)\\
\implies & Y_i \perp\!\!\!\perp \mathbf{X}_i^S \mid \mathbf{W}_i, S && \text{by ignorability} && (11)\\
& \text{Note: } Y_i \perp\!\!\!\perp S_i \mid \mathbf{W}_i, \mathbf{X}_i^S && \text{By definition of Sampling Set} && (12)\\
\implies & Y_i \perp\!\!\!\perp S_i, \mathbf{X}_i^S \mid \mathbf{W}_i && \text{Combining Equations (11) and (12)} && (13)\\
& && \text{using Intersection principle in Pearl (2000)} && \\
\implies & Y_i \perp\!\!\!\perp S_i \mid \mathbf{W}_i && \square && (14)
\end{aligned}
$$

The intuition behind Theorem D.1 is that if a separating set $\mathbf{W}$ renders all variables in the sampling set conditionally ignorable to the outcome within the survey sample, Assumption 1 holds and $\mathbf{W}$ can be used to identify and estimate the population mean. Because conditional ignorability need only hold across the survey sample, we can directly estimate using just the survey sample data whether or not there exists a $\mathbf{W}$ that blocks all paths between the outcome Y and the sampling set $\mathbf{X}^S$. When $\mathbf{W}$ does not exist, this implies that the partially observed V is necessary for unbiased identification of the population mean.

To estimate such separating sets $\mathbf{W}$, we use the algorithm described in Egami and Hartman (2021), outlined in Table D.2. We first estimate a Markov Random Field (MRF) over all variables in $\mathbf{X}^S$ (some of which are partially observed), all additional covariates that are fully observed $\mathbf{X}$, and the outcome Y. MRFs are statistical models that encode conditional independence structures of variables using graph separation rules. We estimate the MRFs using mixed graphical models (Yang et al., 2014; Haslbeck and Waldorp, 2020), which allow for a mixture of continuous and categorical covariates. Figure D.1 contains an estimated MRF for our application to the 2020 U.S. Presidential survey conducted by ABC News/Washington Post for Michigan.

Once we estimate the MRF, we can solve for the separating set as a constrained, linear programming problem, in which we optimize for a separating set of smallest size.[3] We constrain the optimization problem such that (1) that all paths between Y and the variables in $\mathbf{X}^S$ are blocked by the candidate separating set (i.e., $\mathbf{Pd} \geq 1$, in Equation 15), and (2) that the candidate separating set does not contain the partially observed variable V (i.e., $\mathbf{v}^\top \mathbf{d} = \mathbf{0}$, in Equation 15), where $\mathbf{v}$ encodes which variables are observed in the target population. If the optimization problem returns a valid separating set, then this implies that partially observed confounding is not a concern. However, if no feasible set exists, then this implies that given the fully observed variables available to the researcher for constructing survey weights, there is no set of weights that can recover the population mean without bias due to V.

D.2.1 Running Example: 2020 U.S. Presidential Election

We return now to our running example. We assume our fully observed covariates include $\mathbf{X} = \{$*Age*, *Gender*, *Race/Ethnicity*, *Education Level*, *Party ID*, *Born-Again Christian*$\}$. As in Section 1.1, these demographic variables are commonly used to construct survey weights. Existing literature indicates that political interest is also correlated with propensity to respond to surveys; however it

[3] Alternative loss functions include the set that yields the least dispersion in weights or the lowest variance in the point estimate.

Estimating Separating Sets (Egami and Hartman, 2021)	
Step 1:	Using mixed graphical models, estimate a Markov Random Field over $\{\mathbf{X}^S \cup \mathbf{X}\}$ and the outcome Y. Define q as the total number of covariates. Store all simple paths from Y to $\mathbf{X}^S$ from the estimated MRF in a matrix $\mathbf{P}$.
Step 2:	Define $\mathbf{v}$ to be a q-dimensional vector that encodes which variables are partially observed, where $\mathbf{v}_j = 1$ if the jth variable is only measured in the survey sample and not the target population, and 0 otherwise.
Step 3:	Solve the following linear programming problem: $$\min_{\mathbf{d}} \sum_{j=1}^{q} d_j \quad \text{s.t., } \mathbf{Pd} \geq \mathbf{1} \text{ and } \mathbf{v}^\top \mathbf{d} = \mathbf{0} \tag{15}$$ *If a fully observed, valid separating set* $\mathbf{W}$ *is found:* ↳ Use construct survey weights using estimated sparating set $\mathbf{W}$. *If no fully observed separating set exists:* ↳ Partial confounders exist. Estimate weights with observed covariates and conduct sensitivity analysis for partially observed confounders (Section 4).

Table D.2: Summary of the algorithm for detecting partially observed confounders. We refer readers to Egami and Hartman (2021) for more details on how to estimate the MRF.

is not commonly incorporated into survey weights because it is not available in target populations data. Therefore, we let V ={*Political Interest*}, encoded as whether individuals are "very closely" following the upcoming 2020 U.S. Presidential Election. We assume that the sampling set includes this partially observed variable as well as these common demographic variables, i.e. $\mathbf{X}^S$ = {*Age*, *Gender*, *Race/Ethnicity*, *Education*, *Party ID*, *Born-Again Christian*, $\boxed{\textit{Political Interest}}$}, where the box denotes that V is partially observed. The algorithm seeks to determine if there is a set of variables in $\mathbf{X}$ that renders political interest and the outcome conditionally independent, thus justifying Assumption 1.

Figure D.1 presents the undirected graph estimated for the ABC News/Washington Post poll. The algorithm returns that political interest is not a partially observed confounder, as it can be rendered conditionally independent using our weighting variables. This is visually evident by the fact that the outcome is not connected to political interest. If researchers choose to use only fully observed covariates in the construction of the survey weights, they need not worry about confounding from this political interest variable.

E Extended Results for the Empirical Application

In our main analysis, we use party identification as a weighting variable, as it is the strongest predictor of vote choice in American politics and it is measured in our target population defined by the CES. However, whether or not to weight on party identification can be controversial. In this section, we conduct our analysis for Michigan without using party identification in our weights. We then use party identification as a partially observed confounder to assess sensitivity. Unsurprisingly, we see that the results are quite sensitive to the exclusion of such a strong predictor of the outcome. This is confirmed in the bias contour plots, as well.

Figure D.1: MRF of ABC News/Washington Post poll conducted in late October 2020 in Michigan. White nodes indicate fully observed covariates, red nodes indicate partially observed covariates, and the gray node is the outcome. Lines between nodes indicate a conditional correlation exists, and the width of the line indicates the strength of the correlation.

E.1 Partially Observed Confounding

To begin, we treat party identification as a partially observed variable. In this section, we assess sensitivity to identification as a Democrat and Republican separately, for the purpose of visual exposition. One could also weight on the full three-way category, and present the results as a heat map, which would result in similar findings.

We begin by assessing sensitivity to the proportion of Democrats. The weighted survey average of proportion of Democrats is 0.29 using our survey weights that exclude party identification. We vary the target population proportion of Democrats and re-estimate the vote margin of a Biden victory in Michigan. Consistent with what we expect, we see that if the population proportion of Democrats is larger than 0.29, then this means that by omitting Democrat from our weights, we have underestimated the vote margin of a Biden victory. In contrast, if the population proportion of Democrats is less than 0.29, then by omitting Democrat from the weights, we will have overestimated the vote margin of a Biden victory. We repeat this analysis for the proportion of Republicans, and find consistent results. We visualize the results in Figure E.2. We have truncated the x-axis to a "reasonable" range for these sensitivity parameters, assuming that no less than 25% or more than 40% of the population identifies for each party. It is worth noting that in our target population, the true proportion of Democrats is 35% and Republicans is 31%. If a researcher can reason that the proportion of Democrats is most likely understated, then these results would indicate the poll most likely understates the true margin.

E.2 Fully Unobserved Confounding

We now conduct our sensitivity analysis for fully unobserved confounding. We begin by computing the robustness value, at a threshold of $b^* = 0$ (see Table E.3). We see that in contrast to the robustness value reported in the main text, by not accounting for party identification, the robustness value drops from 0.11 to 0.03. In other words, a confounder that results in an error that explains 3% of the variation in the ideal weights and the outcome will be sufficiently strong to reduce our

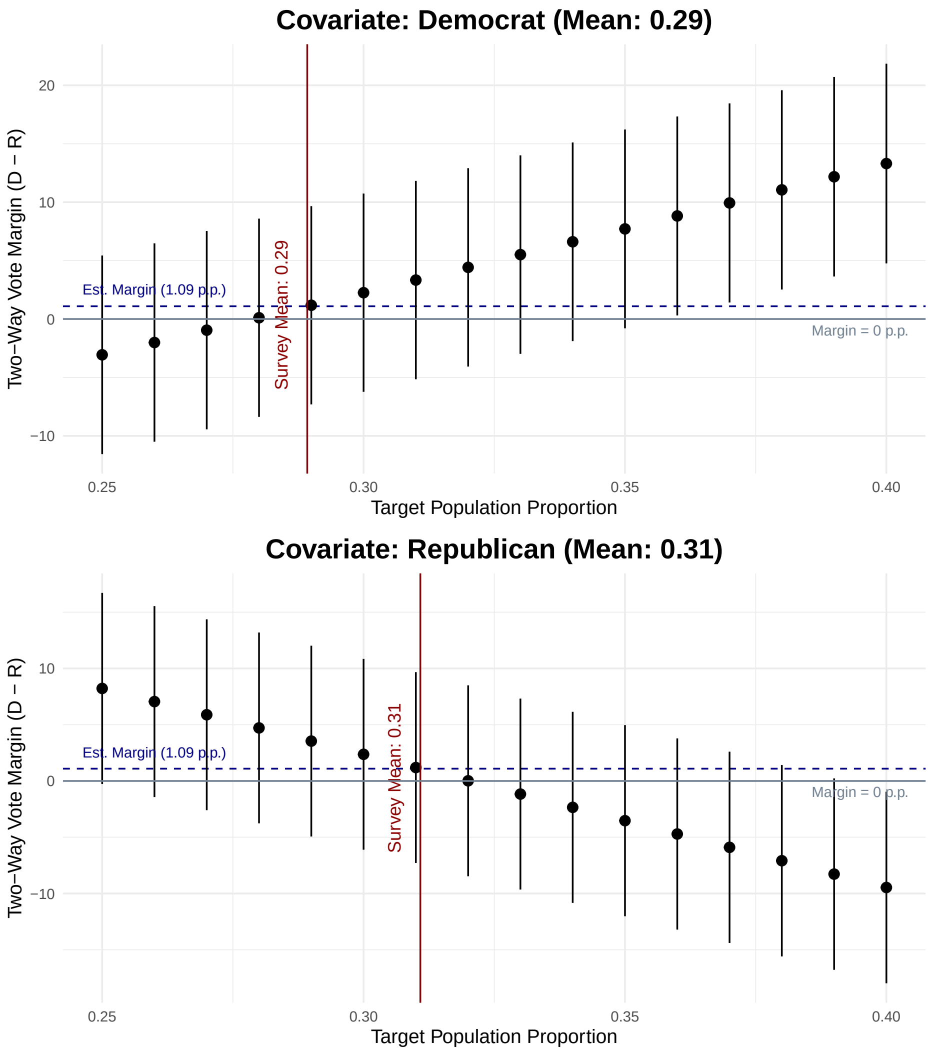

Figure E.2: Partial confounding plots. We vary the target population of the partially observed covariates (i.e., individuals who identify as Democrats or Republicans) and plot the resulting estimates.

estimated vote margin to zero. This is consistent with what we expect, as we know that party identification controls for a lot, and by not accounting for it, our estimate may be more sensitive to potential confounders.

To assess the plausibility of such a killer confounder, we turn to benchmarking (see Table E.4 for results). We see that by omitting party identification from our analysis, omitting variables with equivalent confounding strength to gender or whether or not an individual is a born-again Christian would result in enough confounding to overturn our estimated result of a Biden victory in Michigan. Similarly, we see that omitting a confounder like age or race would also result in a large amount of bias. Note that, without weighting on party identification, the correlation of the errors from omitting these variables is stronger than the original analysis, where the error from omitting these variables was less strong because party identification explained much of the variation with the outcome.

We also generate the bias contour plot (see Figure E.3).

	Weighted Estimate	SE	$RV_{b^*=0}$
Two-Way Vote Margin (D-R)	1.09	3.71	0.03

Table E.3: Sensitivity summary for Michigan, without accounting for party identification.

Variable	$\hat{R}^2_\varepsilon$	$\hat{\rho}_{\varepsilon,Y}$	MRCS	Est. Bias
Age	0.29	0.04	1.43	0.76
Education	0.31	0.04	1.10	1.00
Gender	0.07	-0.13	-0.92	-1.19
Race	0.12	0.07	1.21	0.90
Born Again	0.11	0.15	0.63	1.74

Table E.4: Benchmarking results for Michigan, without accounting for party identification.

References

Ben-Michael, Eli et al. (2021). "The Balancing Act in Causal Inference". In: *arXiv:2110.14831*.

Caughey, Devin et al. (2020). *Target Estimation and Adjustment Weighting for Survey Nonresponse and Sampling Bias*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press.

Cinelli, Carlos and Chad Hazlett (2020). "Making sense of sensitivity: Extending omitted variable bias". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.1, pp. 39–67.

Deville, Jean-Claude and Carl-Erik Särndal (1992). "Calibration Estimators in Survey Sampling". In: *Journal of the American Statistical Association* 87.418, pp. 376–382.

Egami, Naoki and Erin Hartman (2021). "Covariate selection for generalizing experimental results: Application to a large-scale development program in Uganda". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 184.4, pp. 1524–1548.

Hainmueller, Jens (2012). "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies". In: *Political analysis* 20.1, pp. 25–46.

Hartman, Erin, Chad Hazlett, and Ciara Sterbenz (2021). "Kpop: A kernel balancing approach for reducing specification assumptions in survey weighting". In: *arXiv:2107.08075*.

Haslbeck, Jonas and Lourens J Waldorp (2020). "mgm: Estimating Time-Varying Mixed Graphical Models in High-Dimensional Data". In: *Journal of Statistical Software* 93.8, pp. 1–46.

Huang, Melody (2022). "Sensitivity Analysis in the Generalization of Experimental Results". In: *arXiv:2202.03408*.

Kalton, Graham and Ismael Flores-Cervantes (2003). "Weighting Methods". In: *Journal of Official Statistics* 19.2, pp. 81–97.

Little, Roderick JA and Donald B Rubin (2002). *Statistical analysis with missing data*. John Wiley & Sons.

Nguyen, Trang Quynh et al. (2017). "Sensitivity analysis for an unobserved moderator in RCT-to-target-population generalization of treatment effects". In: *The Annals of Applied Statistics* 11.1, pp. 225–247.

Pearl, J (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Särndal, Carl-Erik (2007). "The calibration approach in survey theory and practice". In: *Survey Methodology* 33.2, pp. 99–119.

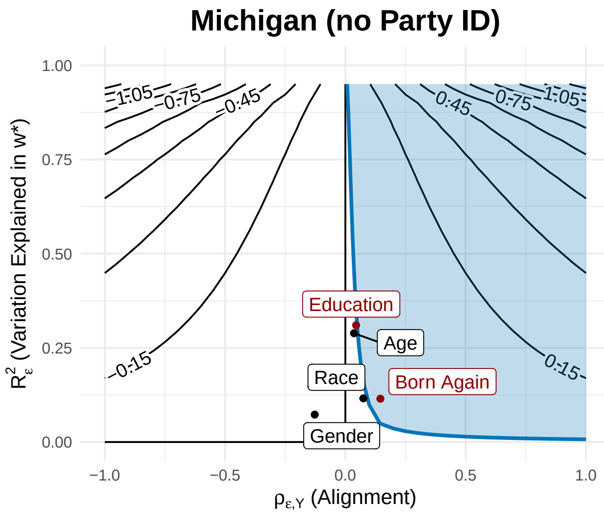

Figure E.3: Bias contour plot for Michigan, not accounting for party identification.

Soriano, Dan et al. (2021). "Interpretable sensitivity analysis for balancing weights". In: *arXiv:2102.13218*.

Wu, Changbao and Wilson W Lu (2016). "Calibration Weighting Methods for Complex Surveys". In: *International Statistical Review* 84.1, pp. 79–98.

Yang, Eunho et al. (2014). "Mixed graphical models via exponential families". In: *Artificial Intelligence and Statistics*, pp. 1042–1050.

Yiu, Sean and Li Su (2018). "Covariate association eliminating weights: a unified weighting framework for causal effect estimation". In: *Biometrika* 105.3, pp. 709–722. ISSN: 0006-3444. DOI: 10.1093/biomet/asy015.