Copula-based Sensitivity Analysis for Multi-Treatment Causal Inference with Unobserved Confounding *

Jiajing Zheng
UCSB

Alexander D'Amour
Google Research

Alexander Franks
UCSB

May 15, 2023

Abstract

Recent work has focused on the potential and pitfalls of causal identification in observational studies with multiple simultaneous treatments. Building on previous work, we show that even if the conditional distribution of unmeasured confounders given treatments were known exactly, the causal effects would not in general be identifiable, although they may be partially identified. Given these results, we propose a sensitivity analysis method for characterizing the effects of potential unmeasured confounding, tailored to the multiple treatment setting, that can be used to characterize a range of causal effects that are compatible with the observed data. Our method is based on a copula factorization of the joint distribution of outcomes, treatments, and confounders, and can be layered on top of arbitrary observed data models. We propose a practical implementation of this approach making use of the Gaussian copula, and establish conditions under which causal effects can be bounded. We also describe approaches for reasoning about effects, including calibrating sensitivity parameters, quantifying robustness of effect estimates, and selecting models that are most consistent with prior hypotheses.

Keywords: Observational studies; multiple treatments; sensitivity analysis; copulas, latent confounders; deconfounder

*Jiajing Zheng is a PhD candidate in the Department of Statistics and Applied Probability at the University of Caliornia Santa Barbara (jzheng@pstat.ucsb.edu). Alexander D'Amour is a Research Scientist at Google Research, Cambridge, MA (alexdamour@google.com). Alexander M. Franks is an Assistant Professor of Statistics at the University of California, Santa Barbara (afranks@pstat.ucsb.edu). We thank Steve Yadlowksy, Victor Veitch, and Avi Feller for thoughtful comments and discussion.

1 Introduction

Although it is well-established that treatment effects are not generally identifiable in the presence of unobserved confounding, recent work has focused on whether this challenge can be mitigated when there are multiple simultaneous treatments (Wang and Blei, 2019). Intuitively, dependence among multivariate treatments could provide information about latent confounders, which could in turn be leveraged to facilitate causal inference and identification. This intuition has motivated latent variable approaches such as "the deconfounder", a much discussed approach for estimating causal effects for multiple treatments (Wang and Blei, 2019).

Unfortunately, it was shown that this strategy has limited practical applicability for point identification and estimation of causal effects. For example, D'Amour (2019a) and D'Amour (2019b) note that causal effects are not nonparametrically identifiable in this setting, even when the distribution of latent confounders can be identified. Likewise, Ogburn et al. (2019) and Ogburn et al. (2020) provide several additional counterexamples and detailed rebuttals to previous theoretical results, while Grimmer et al. (2020) argue that the approach cannot consistently outperform naïve regression, even when stringent assumptions are met.

Nonetheless, latent variable–type strategies are used to estimate causal quantities in genomics (Price et al., 2006), computational neuroscience, social science and medicine (Zhang et al., 2019), and time series applications (Bica et al., 2020). Given the practical importance of these qeustions, recent work has focused on stronger identifying assumptions for causal effects in the multi-treatment setting. Miao et al. (2020) propose identifying assumptions involving proxy or negative control variables and in settings when over half of the treatments are assumed to have a null effect (without specifying which treatments are null). Kong et al. (2019) consider identification in a parametric model with binary outcomes.

To date, the literature on multi-treatment causal inference has revolved around a binary question about point identification: can causal effects be identified or not? However, a potentially more productive question is: what information about treatment effects, if any, can be gained from a latent variable model? We propose that sensitivity analysis—which explores a range of causal effects that are consistent with the observed data in the context of a given problem—can be a useful tool to address this question. Specifically, sensitivity analysis can show what can be gained by leveraging latent structure in a given application, even if this (usually) falls short of fully identifying the causal effect of interest.

We focus on settings in which residual dependence between treatments is presumed to be caused by unmeasured confounders. To extend sensitivity analysis to this setting, we replace untestable assumptions about the magnitude of the dependence between each treatment and unmeasured confounders with an assumption about the suitability of a latent variable model. Given a (partially) identifiable latent variable model linking confounders to treatments, we are still left to specify the relationship between unmeasured confounders and the outcome. Here, we suggest using copulas, which completely characterize this confounder-outcome dependence without affecting the model fit. For practical analyses, we focus on characterizing this dependence with Gaussian copulas. Under the Gaussian copula specification, we establish that causal effects which are unbounded under unrestricted sensitivity models are bounded when the latent variable model for the treatments is identifiable. In this sense, we show that appropriately motivated latent variable models can sharpen causal conclusions and provide insights about the implications of specific assumptions, even if they cannot point-identify causal effects.

The paper proceeds as follows. We begin by defining the relevant quantities and notation in Section 2. In Section 3 we illustrate the intuition behind our approach when we can model the treatments via a linear factor model. In Section 4 we describe a more general framework for latent variable sensitivity analysis via a copula factorization, introducing a special case of the more general approach in which we assume confounder-outcome relationships can be characterized by a Gaussian copula. We discuss sensitivity parameter interpretation, calibration, and measures of robustness in Section 5 and, finally, in Section 6 and Section 7 we demonstrate our approach in simulation and with a gene expression dataset recently reanalyzed by Miao et al. (2020).

2 Preliminaries

2.1 Setup

Let T be a random k-vector of treatment variables, Y be a scalar random outcome of interest, and t and y be realizations of the respective random variables. We let U be a random m-vector denoting potential unobserved confounders, and X denote any observed pre-treatment variables. We use the do-calculus framework (Pearl, 2009) and let $f(y \mid do(t))$ denote the density of Y in the population in which we have intervened to assign treatment level t to all units. In general, this is distinct from the observed outcome density, $f(y \mid t)$, which represents the density of the outcome in the subpopulation that received treatment t. These two

densities are the same if and only if there are no confounders (VanderWeele and Shpitser, 2013).

The goal of observational causal inference with multiple treatments is to quantify the effects of different treatments by comparing the intervention distribution at different levels of treatment T (Lechner, 1999; Lopez et al., 2017). In this work we focus on *marginal contrast estimands* (Franks et al., 2019) under arbitrary outcome and treatment distributions. An estimand is a "marginal contrast" if it can be expressed as a function of the marginal distributions of the intervention outcomes, e.g. $\tau(E[v(Y) \mid do(t_1)], E[v(Y) \mid do(t_2)])$ for some functions v and τ. This includes the vast majority of commonly used estimands. For continuous outcomes, our primary estimand is the difference in the population average outcome for treatment $T = t_1$ and the population average outcome given treatment $T = t_2$:

$$\text{PATE}_{t_1,t_2} := E(Y \mid do(t_1)) - E(Y \mid do(t_2)). \tag{1}$$

Here, $v(Y) = Y$ is the identity function and $\tau(a, b) = a - b$. We also consider the difference in the population average outcome receiving treatment t and observed population average outcome, which we denote

$$\text{PATE}_{t,\bullet} := E(Y \mid do(t)) - E(Y), \tag{2}$$

where $E(Y) = \int E(Y \mid t) f(t) dt$ and $\text{PATE}_{t_1,t_2} = \text{PATE}_{t_1,\bullet} - \text{PATE}_{t_2,\bullet}$. By analogy with the PATE, we also define quantile treatment effect for quantile q, as QTE^q ,

$$\text{QTE}^q_{t_1,t_2} := \text{quantile}_q(Y \mid do(t_1)) - \text{quantile}_q(Y \mid do(t_2)), \tag{3}$$

and let $\text{MTE}_{t_1,t_2} = \text{QTE}^{1/2}_{t_1,t_2}$ be the median treatment effect. For binary outcomes, our primary estimand is the causal risk ratio between treatments t_1 and t_2

$$RR_{t_1,t_2} := P(Y = 1 \mid do(t_1))/P(Y = 1 \mid do(t_2)). \tag{4}$$

where $RR_{t,\bullet}$ is defined analogously to (2), as $P(Y = 1 \mid do(t))/P(Y = 1)$, so that we can express $RR_{t_1,t_2} = RR_{t_1,\bullet}/RR_{t_2,\bullet}$. Here $v(Y) = I[Y = 1]$ is the indicator function and $\tau(a, b) = a/b$.

In general, it is difficult to infer PATEs or RRs from observational data since the potential presence of unmeasured confounders, which affect both treatment and outcome, can bias naive estimates. If U were to be observed, the following assumptions would be sufficient to identify the intervention distribution, and

hence the treatment effect:

Assumption 1 (Backdoor Criterion)**.** X and U block all backdoor paths between T and Y so that $f(Y = y \mid do(T) = t, X = x, U = u) = f(Y \mid T = t, X = x, U = u)$ (Pearl, 2009).

Assumption 2 (Positivity)**.** $f(T = t \mid U = u, X = x) > 0$ for all u and x such that $P(U = u, X = x) > 0$.

Assumption 3 (SUTVA)**.** There are no hidden versions of the treatments and there is no interference between units (see Rubin, 1980).

Assumption 2, also called the overlap condition, ensures that every observable level of the potential confounders (U, X) has a positive probability of being observed with any treatment t, and is needed for non-parametric identification of causal effects. Assumption 3 is a standard consistency condition. The focus on this work relates to Assumption 1. By conditioning on U and X, we "block" non-causal paths between the treatments and the outcome, so that any residual dependence between the treatments and the outcome must be induced by the intervention on the treatment (See Figure 1). Under this assumption $f(Y = y \mid do(T = t)) = \int_{\mathcal{X},\mathcal{U}} f(Y = y \mid T = t, X = x, U = u) f(X = x, U = u) dx du$.

However, since U is not observed, and it is not generally true that $f(Y = y \mid do(T) = t, X = x) = f(Y = y \mid T = t, X = x)$, treatment effects are not identifiable without additional assumptions about the influence of U. In this case, a common solution is to conduct a sensitivity analysis which characterizes how the implied causal effects change under different assumptions about U and its relationship to T and Y given X.

2.2 Sensitivity Analysis

There is an extensive literature on assessing sensitivity to violations of unconfoundedness in single treatment models, dating back at least to the work of Cornfield et al. (1959) on the link between smoking and lung cancer. Since then, a wide range of strategies have been proposed for assessing sensitivity to unobserved confounding (e.g. see Greenland, 1996; Gastwirth et al., 1998; Vansteelandt et al., 2006; Imbens, 2003; VanderWeele and Arah, 2011; VanderWeele et al., 2012; Robins et al., 2000; Franks et al., 2019; Cinelli and Hazlett, 2020; Veitch and Zaveri, 2020). The sensitivity analysis approach that we propose in this paper builds on latent confounder approaches to sensitivity analysis. These approaches assert, as in Assumption 1, that unconfoundess would hold if only an additional latent variable U were observed (Rosenbaum and Rubin, 1983; Robins, 1997; Vansteelandt et al., 2006; Daniels and Hogan, 2008). Cinelli and Hazlett (2020) and Cinelli et al. (2019)

In a typical latent confounder analysis, we posit densities $f(u \mid x)$, the marginal density of the latent confounders, $f_{\psi_T}(t \mid x, u)$, the conditional density or probability mass function (PMF) for treatment assignment given all confounders and $f_{\psi_Y}(y \mid x, u, t)$, the outcome density in the treatment arm t. The dependence of Y and T on U is indexed by a vector of sensitivity parameters $\psi = (\psi_Y, \psi_T)$ (e.g. see Imbens, 2003; Dorie et al., 2016). Practitioners can then reason about how assumptions about these parameters translate to different causal conclusions. Often, this is done through *calibration*, by determining reasonable ranges for ψ using analogies about observable associations and through a *robustness* assessment, by examining how strong associations with unobserved confounders must be for conclusions to change. Latent confounder models are usually parameterized so that some specific values of the sensitivity parameters ψ indicate the "no unobserved confounding" case. For example, we can take $\psi_T = 0$ to imply that $f_{\psi_T}(t \mid x, u) = f_{\psi_T}(t \mid x)$ and $\psi_Y = 0$ to imply that $f_{\psi_Y}(y \mid x, u, t) = f_{\psi_Y}(y \mid x, t)$. Then, when either $\psi_T = 0$ or $\psi_Y = 0$, U is not a confounder (VanderWeele and Shpitser, 2013). Without loss of generality, we suppress conditioning on x throughout the remainder of the manuscript, and comment on the role of observed covariates where appropriate.

2.3 Partial Identification and Copula Parameterizations

In this paper, we focus on models for which it is possible to learn the conditional confounder density $f_{\psi_T}(u \mid t)$ from a latent variable model on the multiple treatments. We then explore how the causal effects change under different assumptions about the $U - Y$ relationship, as governed by the sensitivity parameter ψ_Y. To do so, we use a sensitivity parameterization in which the observed data densities are invariant to the choice of sensitivity parameters (Gustafson et al., 2018; Franks et al., 2019). We consider this to be desirable because it implies that the data offer equivalent support to the various causal conclusions admitted by the sensitivity analysis. The model for Y conditional on treatments and potential unobserved confounders can be decomposed into the observed data density and a conditional copula as

$$f_\psi(y \mid u, t) = f(y \mid t) c_{\psi_Y}(F_{Y|t}(y), F_{U|t}^{\psi_T}(u) \mid t) \tag{5}$$

where $F_{Y|t}$ is the CDF of $f(y \mid t)$ and $F_{U|t}$ is the CDF of $f_{\psi_T}(u \mid t)$. c_{ψ_Y} is the conditional copula density, defined on the unit hypercube and parameterized by ψ_Y, which characterizes the joint density of Y and U conditional of $T = t$ after transforming the marginals to uniform random variables (Nelsen, 2007). This factorization holds for all densities (or PMFs) $f(y \mid t)$ and $f_{\psi_T}(u \mid t)$ and any number of treatments, and

thus can be used to characterize the outcome-confounder dependence for any model of the observables.

With Equation 5, we can express the intervention distribution, $f_\psi(y \mid do(t))$, in terms of the observed conditional distribution, $f(y \mid t)$, as:

$$f_\psi(y \mid do(t)) = f(y \mid t) \int c_{\psi_Y}(F_{Y|t}(y), F_{U|t}^{\psi_T}(u) \mid t) f(u) du \tag{6}$$

Note that when either c is the independence copula or $f(u \mid t) = f(u)$ there is no confounding and the integral on the right hand side of (6) evaluates to 1 so that the intervention density is identical to the conditional outcome density, as expected.

Given that we can learn $f_{\psi_T}(u \mid t)$ from a latent variable model, it suffices to explore how causal conclusions change under different assumptions about ψ_Y, which governs the conditional copula, c_{ψ_Y}. We focus primarily on the setting in which this copula is a Gaussian copula which characterizes monotone dependences bewteen unmeasured confounders and the outcome. For the Gaussian copula, we show that the causal effects are partially identified, with the most extreme outcomes acheived when there is perfect dependence between the outcome and unmeasured confounders given the treatment.

Our results contribute to an extensive literature on partial identification (Manski, 2003; Gustafson, 2015), and in particular, approaches to partial identification involving copulas (Tamer, 2010). For partially identified parameters, a common approach is to consider the worst-case bounds under a set of weaker assumptions and show how additional assumptions can further sharpen inferences (Manski, 2003, 2008). Partial identification results have been established in causal settings with instrumental variables (Swanson et al., 2018; Flores and Flores-Lagunes, 2013), causal inference with noisy covariate data (Guo et al., 2022), and for estimation of individual treatment effects (ITEs). A key result from the copula literature, due to Fréchet and Hoeffding, characterizes model-free bounds on the joint CDF of random variables as functions of the marginal CDFS. The Fréchet-Hoeffding bound and other related bounds have been specifically used to bound the distribution of ITEs and other functionals of the joint distribution of potential outcomes (see e.g. Heckman et al., 1997; Fan and Park, 2010; Firpo and Ridder, 2019). Our formulation is fundamentally different from approaches using model-free copula bounds, since we remain focused on marginal contrasts (not joint distributions over potential outcomes) and also focus on parametric copula models for characterizing the dependence between the outcome and potential unmeasured variables.

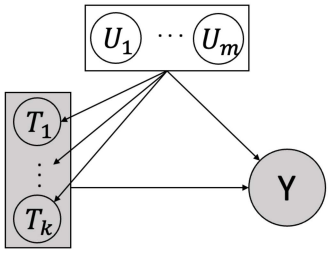


Figure 1: k-vector of treatments, T, m-vector of unmeasured confounders U and a scalar outcome Y. In the linear factor model (8), the treatments are conditionally independent given the unmeasured confounders. We exclude observed covariates, X, from the diagram for simplicity.

3 Confounding Bias in the Linear Factor Model

Before detailing our general copula-based approach, we provide some crucial intuition about our sensitivity analysis in a simple linear Gaussian factor model. The more general approach which we will introduce in the next Section uses the same ideas presented here but relaxes the requirements on the marginal distributions of the treatment and outcome by using copulas. In the linear-Gaussian mdoel, we highlight the following results:

- For causal inference with multiple treatments, we show that the magnitude of the confounding bias for PATE_{t_1,t_2} is bounded. Given standard assumptions for factor model identifiability this bound is identifiable. We characterize how the magnitude of this bound depends on the parameters of the latent confounder model and a scalar sensitivity parameter.

- The confounding bias depends on the treatment contrast. We characterize which treatment contrasts lead to the largest bounds and which treatment contrasts (if any) imply identifiable effects.

- For causal inference with a single treatment, for which the conditional confounder distribution is not identifiable, we cannot identify a bound for confounding bias of PATE_{t_1,t_2} without additional assumptions.

3.1 Sensitivity Bounds in the Linear Gaussian Model

In this section, we establish expressions for confounding bias of PATE_{t_1,t_2} in terms of the parameters of a linear Gaussian model. We assume the following linear structural model:

$$U = \epsilon_u \tag{7}$$

$$T = BU + \epsilon_t \tag{8}$$

$$Y = \tau' T + \gamma' U + \epsilon_y \tag{9}$$

with ϵ_u, ϵ_t and ϵ_y all mean-zero Gaussian, with $\text{Var}(\epsilon_u) = I$, $\text{Var}(\epsilon_y) = \sigma^2$ and $\text{Var}(\epsilon_t) = \Lambda_t$ an arbitrary diagonal matrix. Further, $B \in \mathbb{R}^{k \times m}$, $\tau \in \mathbb{R}^k$, $\gamma \in \mathbb{R}^m$. When either $B = 0$ or $\gamma = 0$, there is no confounding. In the more general framing introduced from Equation (5) in the Section, $\psi_t = \{B\}$ determines the copula defining the T–U relationship and $\psi_Y = \{\gamma\}$ specifies the copula defining Y–U dependence.

Equations (7) and (8) imply that the conditional distribution of the confounder can be expressed as $f_{\psi_t}(u \mid t) \sim N(\mu_{u|t}, \Sigma_{u|t})$, where

$$\mu_{u|t} = B'(BB' + \Lambda_t)^{-1} t \tag{10}$$

$$\Sigma_{u|t} = I - B'(BB' + \Lambda_t)^{-1} B \tag{11}$$

Under model (7)-(9), the intervention distribution has density

$$f(y \mid do(T = t)) \sim N(\tau' t,\ \sigma^2 + \gamma'\gamma). \tag{12}$$

For any t_1, t_2, PATE_{t_1,t_2} is characterized entirely by the regression coefficients τ. The observed outcome distribution can be expressed as

$$f(y \mid T = t) \sim N(\tau'_{\text{naive}} t,\ \sigma^2_{y|t}), \tag{13}$$

where

$$\tau_{\text{naive}} = \tau + \gamma' \mu_{u|t} \tag{14}$$

$$\sigma^2_{y|t} = \sigma^2 + \gamma' \Sigma_{u|t} \gamma. \tag{15}$$

We refer to τ_{naive} as the naive estimate since it naively neglects the effect of unobserved confounders. Equation (15) shows that the observed residual outcome variance, $\sigma^2_{y|t} := \text{Var}(Y \mid T)$, can be decomposed into nonconfounding variation σ^2 and confounding variation, $\gamma'\Sigma_{u|t}\gamma$.

We note that the population average treatment effect and the bias of the naive estimator depends only on the difference between the treatment vectors, $(t_1 - t_2)$. This is the case since the population average treatment effect can be expressed as

$$\text{PATE}_{t_1,t_2} = \tau'(t_1 - t_2) \tag{16}$$

and the confounding bias, $\text{Bias}_{t_1,t_2} = (\tau_{\text{naive}} - \tau)'(t_1 - t_2)$. It is then straightforward to show that the bias is linear in the difference in the confounder means in each treatment group:

$$\text{Bias}_{t_1,t_2} = \gamma'(\mu_{u|t_1} - \mu_{u|t_2}) = \gamma' B'(BB' + \Lambda_t)^{-1}(t_1 - t_2), \tag{17}$$

For the results that follow, it is useful to define the fraction of residual outcome variance explained by confounders as a key quantity:

$$0 \leq R^2_{Y\sim U|T} = \frac{\gamma'\Sigma_{u|t}\gamma}{\sigma^2_{y|t}} \leq 1 \tag{18}$$

This R-squared value can be viewed as a parameter governing the copula c_{ψ_Y} in the general model (5), and plays a central role in sensitivity analysis frameworks such as Cinelli and Hazlett (2020). Using $R^2_{Y\sim U|T}$, we can write a bound on the omitted variable bias of any $PATE_{t_1,t_2}$.

Theorem 1. *Suppose that the observed data is generated by model (7)-(9) with $\Lambda_{t|u} > 0$. Then, $\forall\gamma$,*

$$\gamma'\Sigma_{u|t}\gamma \leq \sigma^2_{y|t} \tag{19}$$

For any given t_1, t_2, we have

$$Bias^2_{t_1,t_2} \leq \sigma^2_{y|t} R^2_{Y\sim U|T} \|\Sigma^{-1/2}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2})\|^2_2. \tag{20}$$

The bound is achieved when γ is colinear with $\Sigma^{-1}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2})$ and is maximized when all the residual outcome variance is due to unmeasured confounders, e.g. $R^2_{Y\sim U|T} = 1$.

Proof. See appendix.

This theorem states that the true causal effect lies in the interval

$$\tau'_{naive}(t_1 - t_2) \pm \sqrt{\sigma^2_{y|t} R^2_{Y\sim U|T}} \|\Sigma^{-1/2}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2})\|_2. \tag{21}$$

We refer to the right-hand side of (20) as the "worst-case bias" of the naive estimator. In particular, since τ_{naive} is the midpoint of the ignorance region, it has the minimum worst-case bias over all alternative causal effect estimators. This is consistent with Grimmer et al. (2020) who emphasize that the deconfounder proposed by Wang and Blei (2019) cannot outperform the naive estimator in general.

In the following corollary, we provide additional intuition by establishing the worst-case bias over all possible treatment contrasts in the special case of the homoskedastic factor model, for which $\Lambda_t = \sigma^2_t I$:

Corollary 1.1. *Assume $\Lambda_t = \sigma^2_t I$ and let d_1 be the largest singular value of B. For all t_1, t_2 with $\| (t_1 - t_2) \|_2 = 1$, the squared bias is bounded by*

$$Bias^2_{t_1,t_2} \leq \frac{d_1^2}{(d_1^2 + \sigma_t^2)} \frac{\sigma^2_{y|t}}{\sigma^2_t} R^2_{Y\sim U|T}, \tag{22}$$

with equality when $(t_1 - t_2) = u_1^B$, the first left singular vector of B. When $(t_1 - t_2) \in Null(B')$, the naive estimate is unbiased, that is, $PATE_{t_1,t_2} = \tau'_{naive}(t_1 - t_2)$.
Proof: See Appendix.

The first term in (22), $\frac{d_1^2}{(d_1^2+\sigma_t^2)}$, is the fraction of variance in the first principal component of the causes that can be explained by confounding. The first principal component corresponds to the projection of treatments which is most correlated with confounders, and thus is the causal contrast with the largest ignorance region. We illustrate and discuss some additional insights from Corollary 1.1 in Figure 6, Appendix A.

3.2 Identifiability of Sensitivity Bounds

In the previous subsection, we established bounds on the omitted variable bias, but did not characterize whether the bounds are themselves identifiable. Now, we show that factor model identifiability assumptions (when appropriate) indeed imply that these bounds are identifiable. Thus, multi-treatment inference can yield shaper sensitivity analyses, beyond what is possible when considering inference one treatment at a time.

Crucially, in model (7)-(9), B can be identified (up to rotation) under standard factor model conditions

(Anderson and Rubin, 1956). The parameter γ is not identifiable but can be considered a sensitivity vector that parameterizes the residual correlation between the m-dimensional unobserved confounder U and the outcome Y after conditioning on the treatment vector T. We use results on the identification of B to establish the following proposition.

Proposition 1. *Suppose that the observed data is generated by model (7)-(9). If B is rank m and there remain two disjoint matrices of rank m after deleting any row of B, then $\|\Sigma_{u|t}^{-1/2}(\mu_{u|t_1} - \mu_{u|t_2})\|_2^2$ and $\frac{d_1^2}{(d_1^2+\sigma_t^2)}$ are both identified.*

Proof. See appendix.

As a result, the bounds in both Equations (20) and (22) are identified given $R^2_{Y\sim U|T}$. Further, since $R^2_{Y\sim U|T}$ is itself at most 1, we can indeed identify an upper bound on the omitted variable bias under the assumptions in Proposition 1. Notably, the sufficient conditions in 1 can only be satisfied when the number of confounders is $m \leq (k-1)/2$. This fact can be useful for practitioners to reason about the number of treatments that suffice, relative to the number of unmeasured confounders, to bound the confounding bias.

Proposition 1 also establishes a key distinction between the multi-treamtent and single treatment settings for sensitivity analysis. Specifically, when $k = 1$ (single treatment causal inference), we cannot identify a bound on the omitted variable bias. As shown in Cinelli and Hazlett (2020), in the single treatment case, the squared confounding bias of the PATE can be expressed as

$$Bias^2_{t_1,t_2} = \frac{\sigma^2_{y|t}}{\sigma^2_T}\left(\frac{R^2_{T\sim U}}{1-R^2_{T\sim U}}\right)R^2_{Y\sim U|T} \tag{23}$$

where $\sigma^2_T := BB' + \Lambda_t$ is the marginal variance of the treatment and

$$0 \leq R^2_{T\sim U} = \frac{\sigma^2_T||\mu_{u|t_1} - \mu_{u|t_2}||_2^2}{(t_1 - t_2)^2} \leq 1 \tag{24}$$

is the unidentifed fraction of treatment variance explained by confounders. In the single treatment case, neither $R^2_{T\sim U}$ nor $R^2_{Y\sim U|T=t}$ are identifiable, and since $\frac{R^2_{T\sim U}}{1-R^2_{T\sim U}}$ can be arbitrarily large, the confounding bias is unbounded without additional assumptions. As such, Cinelli and Hazlett (2020) consider a variety of calibration and robustness criteria for reasoning about plausible magnitudes for both $R^2_{Y\sim U|T}$ and $R^2_{T\sim U}$. In contrast, as we show above, in the multiple treatment case, the marginal bounds on the omitted variable bias depend on the sensitivity vector γ only through the fraction of outcome variance explained by confounders given the treatment, $R^2_{Y\sim U|T} = \frac{\gamma'\Sigma_{u|t}\gamma}{\sigma^2_{y|t}}$. In later Sections, we explore how domain knowledge combined with

techniques for calibrating γ and it's magnitude can be used to further sharpen the set of plausible causal conclusions.

4 Sensitivity Analysis via Copula Parameterizations

We now establish a sensitivity parameterization for multiple treatment causal inference in a more general class of models by leveraging the copula parameterization (6). For this class of models, we propose an algorithm for estimating any marginal contrast estimand. We start with the following general structural equation model

$$U = \epsilon_u \tag{25}$$

$$T = h_{\psi_T}(U, \epsilon_t) \tag{26}$$

$$\tilde{Y} = g_{\psi_Y}(T, U, \epsilon_y) \tag{27}$$

$$Y = F^{-1}_{Y|t}(F_{\tilde{Y}|T}(\tilde{Y})) \tag{28}$$

where $F^{-1}_{Y|T}$ is the inverse-CDF of the conditional distribution of Y given $T = t$ and h_{ψ_T} and g_{ψ_Y} are arbitrary functions. In general, neither h_{ψ_T} nor g_{ψ_Y} are identifiable when U is not observed without additional assumptions.

As illustrated in the previous Section, when T is multivariate, we might gain information about ψ_T if we are willing to make assumptions about the class of latent variable models linking the unmeasured confounders to the treatment.

Assumption 4 (Latent variable model identification)**.** The potential confounders are continuous and their distribution given treatments, $f\psi_T(u \mid t)$, is identifiable up to rotation and scale.

In this work, we focus on continuous confounders and leave explorations of discrete latent variable models for future work. Factor model identification is essential for the validity of our sensitivity analysis and will not hold in all settings. However, while a complete discussion of identifiability in latent variable models is outside the scope of this work, there is a broad range of mathematical settings in which this assumption does hold. In the previous Section, we noted the classical result due to Anderson and Rubin (1956) establishing identifiablity conditions for linear factor models. Allman et al. (2009) establish identifiability for many latent class models, including those with limited direct dependence between observations; and Miao et al.

(2020) where weak sufficient conditions are given for similar identifiability in linear models. Barber et al. (2022) consider a set of criterion for establishing when latent variable and Rohe and Zeng (2020) consider identifiability in a broader class of (non-Gaussian) factor models. It is up to the practitioner to decide whether latent variable identifying assumptions are compatible with their applied setting. Finally, in Appendix B we formalize the idea that it is sufficient to identify the latent variable density only up to invertible linear transformations (rotation and scale). To do so we introduce the notion of a "causal equivalence class", which establishes that the substantive causal conclusions do not depend on a particular rotation or scale for the latent confounders.

Given $f(y \mid t)$ (identifiable), $f_{\psi_T}(u \mid t)$ (identifiable by Assumption 4) and c_{ψ_Y} (nonidentifiable, governed by chosen sensitivity parameter ψ_Y) , we can compute the expected value of any function of the outcome under the intervention distribution, $E[v(Y) \mid do(t)] = \int v(y) f(y \mid do(t)) dy$. This can be in turn used to compute any marginal contrast estimand. Applying equation (6), we write this intervention expectation as

$$E[v(Y) \mid do(t)] \ = \int v(y) w_\psi(y,t) f(y \mid t) dy, \tag{29}$$

where $w_\psi(y,t) = \int c_{\psi_Y}(F_{Y|t}(y), F^{\psi_T}_{U|t}(u) \mid t) f_{\psi_T}(u) du$ is the importance weight associated with sampling from the observed data distribution instead of the intervention distribution. In practice, we can approximate the marginal distribution of the unobserved confounder with the mixture density $f_{\psi_T}(u) \approx \frac{1}{n} \sum_i f^{\psi_T}(u \mid t_i)$ where $t_i \in \mathcal{T}$ is the ith observed treatment and $\mathcal{T}$ is the set of all observed treatment vectors. Thus, the importance weight can be approximated as

$$w_\psi(y,t) \approx \frac{1}{|\mathcal{T}|} \sum_{t_i \in \mathcal{T}} \left[\int c_{\psi_Y}(F_{Y|t}(y), F^{\psi_T}_{U|t}(u) \mid t) f_{\psi_T}(u \mid t_i) du \right]. \tag{30}$$

We use this approximation to derive importance sampling algorithm for computing the expected value in Equation (29) for any copula and conditional confounder distributions $f(u \mid t)$ (Appendix A, Algorithm 1). This can in turn be used to compute any marginal contrast estimand, $\tau(E[v(y)|do(t_1)], E[v(y)|do(t_2)])$.

4.1 The Gaussian Copula Sensitivity Parameterization

In order to bridge the gap between the interpretable sensitivity parameterization and established theory in the linear Gaussian model (Section 3) and the more general formulation (25)-(28), we provide additional bounds on the omitted variable bias when the copula in (5) is a Gaussian copula. The Gaussian copula is a

natural choice when the conditional mean of the outcome is plausibly monotone in the conditional means of the confounders. It includes the special limiting case in which the outcome is comonotone (perfect positive dependence) or countermonotone (perfect negative dependence) with confounders. Outcome-confounder monotonocity is often plausible, at least approximately, conditional on each level $T = t$.

Assumption 5 (Gaussian copula)**.** The conditional copula between the outcome and m-dimensional latent confounders given treatments, is a Gaussian copula.

Then, under Assumptions 4 and 5, we can assume without loss of generality

$$f(u \mid t) \sim N(\mu_{u|t}, \Sigma_{u|t}) \tag{31}$$

where $\mu_{u|t}$ and $\Sigma_{u|t}$ can be identified (up to rotation) from a latent variable model on the treatments. We can further rewrite equations (27) - (28) as

$$\tilde{Y} = \gamma_t' U + \epsilon_{\tilde{Y}} \tag{32}$$

$$Y = F_{Y|T}^{-1}(\Phi(\tilde{Y} - \gamma_t' \mu_{u|t})) \tag{33}$$

so that $\psi_Y = \{\gamma_t\}$ and $\epsilon_{\tilde{Y}} \sim N(0, 1 - \gamma_t' \Sigma_{u|t} \gamma_t)$ is chosen so that without loss of generality $\text{Var}(\tilde{Y} \mid T) = 1$.

The Gaussian copula under this model is fully determined by the covariance matrix

$$\text{Cov}([\tilde{Y}, U] \mid T = t) = \begin{bmatrix} 1 & \gamma_t' \Sigma_{u|t} \\ \Sigma_{u|t} \gamma_t & \Sigma_{u|t} \end{bmatrix} \tag{34}$$

with parameters $\psi_T = \{\mu_{u|t}, \Sigma_{u|t} : t \in \mathcal{T}\}$ and $\psi_Y = \{\gamma_t : \mathcal{T}\}$ where $\mathcal{T}$ is the space of all treatments. Given Assumption 4, $\psi_Y = \{\gamma_t\}$ is the sole $m-$dimensional sensitivity vector governing the magnitude of the omitted variable bias. Per Assumption 4, we assume that ψ_T is identified up to invertible linear transformations of U, and explore the range of possible causal effects for different γ_t satisfying $\gamma_t' \Sigma_{u|t} \gamma_t \leq 1$.

In Algorithm 2 (Appendix A), we provide a modification of Algorithm 1 tailored to the Gaussian copula setting. At a high level, we compute a Monte Carlo estimate of $f(y \mid do(t))$ via the following three step procedure: (1) draw a sample from $f(u)$, (2) compute the conditional density of the Gaussianized outcome $f(\tilde{y} \mid u, t)$ via the Gaussian copula and (3) transform $\tilde{y}$ back to original space via the conditional quantile function $F_{Y|t}^{-1}$ (see Figure 7, Appendix A). In the following Sections, we introduce some theoretical insights

about our approach and provide a method for calibrating the magnitude of γ_t and reasoning about its direction.

4.2 Bounds on the Causal Effects in Gaussian Copula Models

Although the causal effects given any set of values γ_t can be inferred from Algorithm 2, when the observed outcome distribution is non-Gaussian, we cannot necessarily express bounds on the PATE_{t_1,t_2} analytically. In fact, there is not even a guarantee that the intervention density $f(y \mid do(T))$ has finite mean without additional assumptions about the outcome density. For quantile estimands, on the other hand, we can identify finite bounds on the causal effect. We state this result specifically for the median treatment effect, in the general setting in which $\Sigma_{u|t}$, and γ_t can vary with treatment level t.

Theorem 2. *Assume model (31) - (33) and that $\sigma_{y|t}$, $\Sigma_{u|t}$, and γ_t can vary with t and assume $\mu_{u|t_1}$ and $\mu_{u|t_2}$ are in the row space of $\Sigma_{u|t_1}$ and $\Sigma_{u|t_2}$ respectively, and Y is continuous. Further, let $\Sigma^{\dagger}_{u|t_1}$ and $\Sigma^{\dagger}_{u|t_2}$ denote the pseudo-inverses of $\Sigma_{u|t_1}$ and $\Sigma_{u|t_2}$ respectively. Then the omitted variable bias for all quantile treatment effects are bounded. The median treatment effect, $MTE_{t_1,t_2} = med\big(Y \mid do(t_1)\big) - med\big(Y \mid do(t_2)\big)$ is in the interval $m_l \leq MTE_{t_1,t_2} \leq m_u$ where*

$$m_l = F^{-1}_{Y|T=t_1}(\Phi(-\parallel (\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1} \parallel_2)) - F^{-1}_{Y|T=t_2}(\Phi(\parallel (\Sigma^{\dagger}_{u|t_2})^{1/2}\mu_{u|t_2} \parallel_2)) \tag{35}$$

$$m_u = F^{-1}_{Y|T=t_1}(\Phi(\parallel (\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1} \parallel_2)) - F^{-1}_{Y|T=t_2}(\Phi(-\parallel (\Sigma^{\dagger}_{u|t_2})^{1/2}\mu_{u|t_2} \parallel_2)) \tag{36}$$

where m_l and m_u are identifiable under Assumptions 4 and 5.

Proof: See Appendix.

When Y is conditionally Gaussian, i.e $F^{-1}_{Y|T=t}$ is the inverse-CDF of a Gaussian random variable for all t, then the mean and median are the same so that $MTE_{t_1,t_2} = \text{PATE}_{t_1,t_2}$ and thus we can use the result from Theorem 2 to bound the bias of the PATE.

Corollary 2.1. *Assume the model (31) - (33) where Y is conditionally Gaussian given treatments, and where $\sigma_{y|t}$, $\Sigma_{u|t}$, and γ_t can vary with t. If $\Sigma_{u|t_1}$ and $\Sigma_{u|t_2}$ are non-invertible, then $Bias_{t_1,t_2}$ is bounded if and only if $\mu_{u|t_1}$ and $\mu_{u|t_2}$ are in the row space of $\Sigma_{u|t_1}$ and $\Sigma_{u|t_2}$ respectively. When bounded,*

$$Bias^2_{t_1,t_2} \leq \left(\sigma_{y|t_1}\sqrt{R^2_{Y\sim U|t_1}}\|(\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1}\|_2 + \sigma_{y|t_2}\sqrt{R^2_{Y\sim U|t_2}}\|(\Sigma^{\dagger}_{u|t_2})^{1/2}\mu_{u|t_2}\|_2\right)^2, \tag{37}$$

with equality when $\gamma_{t_1} \propto \Sigma^{\dagger}_{u|t_1}\mu_{u|t_1}$ *and* $\gamma_{t_2} \propto \Sigma^{\dagger}_{u|t_2}\mu_{u|t_1}$ *and where* $\Sigma^{\dagger}_{u|t_1}$ *and* $\Sigma^{\dagger}_{u|t_2}$ *are the pseudo-inverses of* $\Sigma_{u|t_1}$ *and* $\Sigma_{u|t_2}$. *If* $\Sigma_{u|t_1} = \Sigma_{u|t_2} = \Sigma_{u|t}$ *and* $\gamma_t = \gamma$ *is invariant to* t *(i.e. there are no treatment-confounder interactions), and* $\sigma^2_{y|t_1} = \sigma^2_{y|t_2} = \sigma^2_{y|t}$ *(homoskedastic outcome model) then* $Bias_{t_1,t_2}$ *is bounded if and only if* $\mu_{u|t_1} - \mu_{u|t_2}$ *is in the row space of* $\Sigma_{u|t}$ *and when bounded,*

$$Bias^2_{t_1,t_2} \leq \sigma^2_{y|t} R^2_{Y\sim U|T} \|(\Sigma^{\dagger}_{u|t})^{1/2}(\mu_{u|t_1} - \mu_{u|t_2})\|_2^2. \tag{38}$$

Proof: See Appendix.

As expected Equation (38) takes the same form as Equation (20), but generalizes it in that 1) we do not require that $\Sigma_{u|t}$ has the form in (11) and is only required to be non-negative definite and 2) $\mu_{u|t}$ can be nonlinear in t (i.e. it does not need to follow Equation (10)). As in Section 3, when bounded, the bias is proportional to the norm of the scaled difference in confounder means in the two treatment arms. When there exists an m-vector, q, such that $\text{Var}(q'U \mid T = t) = 0$, then $\Sigma_{u|t}$ is non-invertible because there exists a projection of the confounders that is point identified. Corollary 2.1 says that in this case, the ignorance region for the PATE is bounded if and only if $q'(\mu_{u|t_1} - \mu_{u|t_2}) = 0$. In words, if a projection of the confounders can be identified, then the confounding bias is bounded if and only if the identifiable projection of the confounders has the same value in both treatment arms. This corresponds to observations in D'Amour (2019b) about violations of the positivity assumption (Assumption 2) when confounders are "pinpointed" by the latent variable model. In particular, if the "pinpointed" confounders do not match between t_1 and t_2, this implies that positivity has been violated.

Finally, we note that for binary outcomes, we focus primarily on the risk ratio as the estimands of interest. Interestingly, unlike the ATE and quantile treatment effects, $RR_{t,\bullet}$ and RR_{t_1,t_2} are non-monotone in the magnitude of γ. We discuss this in more detail in Appendix C.3 and provide simulation results with binary outcomes in Section 6. For simplicity, for the remainder of the paper, we focus on settings in which $\gamma_t = \gamma$ does not vary with the level of treatment. This corresponds to a model in which there are no treatment-confounder interactions in the outcome model.

5 Calibration and Robustness

Sensitivity analyses consist of two parts: first, the sensitivity model itself, which specifies a set of data-compatible causal models, indexed by sensitivity parameters; and secondly, exploratory tools for mapping

external assumptions to particular causal models in this set. We now turn to discussing the latter in the context of our proposed model.

In the sensitivity analysis literature so far, two exploratory techniques have been particularly popular in single treatment studies: *calibration*, which maps sensitivity parameter values to interpretable observable or hypothetical quantities; and *robustness analysis*, which characterizes the "strength" of confounding necessary to change the conclusion of a study. Here, we show how to adapt these techniques to our sensitivity model in the multi-treatment setting. In addition, we introduce a third class of tools that are particularly well-suited to the multi-treatment setting, which we call *multiple contrast criteria* (MCCs). MCCs specify aggregate properties of the treatment effects for multiple treatment contrasts that are implied by a single causal model, e.g., the L2 norm of PATEs corresponding to contrasts in each individual treatment variable in T. In many multi-treatment settings, assumptions are often expressed in terms of the aggregates—e.g., in genomics, the idea that the effect of most single nucleotide polymorphisms is small—and we show here how these can be used in conjunction with our sensitivity model to characterize candidate causal models that may be of interest in an application.

5.1 Calibration for a Single Contrast

We begin by describing calibration for γ in our sensitivity model when the focus is on a single treatment contrast, between levels $T = t_1$ and $T = t_2$. The goal is to develop heuristics for specifying "reasonable" values or ranges for γ, e.g., to derive bounds on treatment effects by specifying bounds on the strength or direction of confounding. Following previous work in the single treatment setting, we outline how to calibrate our sensitivity parameter vector γ in terms of a fraction of outcome variance explained by the unobserved confounder. Recall that γ is a vector that parameterizes the residual correlation between the m-dimensional unobserved confounder U and the outcome Y after conditioning on the treatment vector T.

First, we briefly review calibration in single-treatment settings. In latent variable approaches for single treatment sensitivity analysis, the causal effect is identified given two sensitivity parameters: the fraction of outcome variance explained by unobserved confounders, $R^2_{Y\sim U|T}$, and the fraction of treatment variance explained by unobserved confounders, $R^2_{T\sim U}$ (Cinelli and Hazlett, 2020). In a linear model, these two scalar quantities identify the confounding bias (Equation (23)). Neither R-squared value is identifiable and thus many authors have proposed strategies for drawing analogies between these values and other observable or hypothetical quantities (Cinelli et al., 2020; Veitch and Zaveri, 2020; Franks et al., 2019).

We borrow this strategy for calibration in our setting, with some modifications. First, in our setting there is no need to calibrate $R^2_{T\sim U}$, because we have restricted ourselves to a setting in which this is implicitly identified (Assumption 4) This leaves calibration of the outcome-confounder relationship, which in our setting is more complex because it is parameterized by a vector γ[1]. However, we can reparameterize γ in terms of a direction d and an R-squared for interpretable calibration:

$$\gamma = \sigma_{y|t}\sqrt{R^2_{Y\sim U|T}}\Sigma^{-1/2}_{u|t}d, \tag{39}$$

where $d \in \mathbb{S}^{m-1}$ is an m-dimensional unit vector on the $(m-1)$-sphere. We discuss strategies for calibrating both the magnitude and direction separately.

Calibrating the magnitude of γ. For Gaussian outcomes, the magnitude of γ is characterized entirely by $R^2_{Y\sim U|T}$, the partial fraction of outcome variance explained by U given T. When $R^2_{Y\sim U|T} = 0$ there is no unobserved confounding and when $R^2_{Y\sim U|T} = 1$, all the observed residual variance in Y is due to confounding factors. In order to calibrate this magnitude, we adopt an idea proposed by Cinelli and Hazlett (2020) for causal inference with single treatments.

First, we consider calibration when observed covariates, X, are also available and consider the importance of an unmeasured confounder U relative to a measured confounder (or set of confounders), X_j, given all other confounders X_{-j}. Specifically, assume that we believe that $R^2_{Y\sim U|X_{-j},T} \leq \kappa R^2_{Y\sim X_j|X_{-j},T}$, where κ is a user chosen parameter reflecting an upper bound on how much "stronger" U might be than X_j. Then Cinelli and Hazlett (2020) show that this implies

$$R^2_{Y\sim U|X,T} \leq \kappa \frac{R^2_{Y\sim X_j|X_{-j},T}}{1 - R^2_{Y\sim X_j|X_{-j},T}}. \tag{40}$$

We use the right hand side of (40), which is estimable given any choice of κ, to benchmark the fraction of outcome variance explained by unmeasured confounders given observed confounders and treatments.

When there are no measured confounders, we can still use the same strategy as above, by leveraging the presence of multiple treatments to calibrate $R^2_{Y\sim U|T}$. For example, in the context of the example to come in Section 7, where treatments are gene expression levels, we might posit that unmeasured confounders cannot

[1]Unlike the single treatment setting, the confounder-outcome relationship cannot be sufficiently summarized in terms of a scalar $R^2_{Y\sim U|T}$. Each confounder can impact each treatment in different ways.

explain more variation in the outcome than a set of genes T_j, given all other genes T_{-j}. We can compute this quantity, the fraction of variation in Y that can be explained by a specific treatment (or set of treatments), T_j, after controlling for all other treatments T_{-j} as

$$R^2_{Y\sim T_j|T_{-j}} := \frac{R^2_{Y\sim T} - R^2_{Y\sim T_{-j}}}{1 - R^2_{Y\sim T_{-j}}}. \tag{41}$$

and then, analogously to (40), can make the assumption that $R^2_{Y\sim U|T_{-j}} \leq \kappa R^2_{Y\sim T_j|T_{-j}}$. As before, this implies the benchmark $R^2_{Y\sim U|T} \leq \kappa \frac{R^2_{Y\sim T_j|T_{-j}}}{1-R^2_{Y\sim T_j|T_{-j}}}$.

When the observed outcome is non-Gaussian, we calibrate the "implicit R^2", by considering the explained variance of the latent Gaussian outcome, $\tilde{Y}$ in Equation (27). The implicit R^2 of T for model (31) - (33) is defined as $R^2_{\tilde{Y}\sim T} = \frac{\text{Var}(E[\tilde{Y}|T])}{\text{Var}(E[\tilde{Y}|T])+1}$. and the implicit partial R-squared of treatment T_j, $R^2_{\tilde{Y}\sim T_j|T_{-j}}$, is defined analogously to Equation (41). As before, these estimable partial R-squared values can be used to provide a useful comparison for the partial R-squared of potential unobserved confounders, $R^2_{\tilde{Y}\sim U|T}$. For more detail, see Imbens (2003) and Franks et al. (2019) who discuss calibration with implicit R-squared values in logistic regression models. See Veitch and Zaveri (2020) and Cinelli et al. (2020) propose useful graphical summaries for calibration based on these metrics in the single treatment setting.

Choosing the direction of γ. Given a magnitude, we now propose a default method for identifying the direction of γ for a single contrast. The dot product $d'\Sigma^{-1/2}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2})$ corresponds to the projection of the scaled difference in confounder means onto the outcome space. By default, we suggest using the direction which maximizes the squared bias. As shown in Corrolary 2.1, when d is colinear with $\Sigma^{-1/2}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2})$, the confounding bias of the naive estimator for Gaussian outcomes is maximized at

$$|\text{Bias}_{t_1,t_2}| = \sigma_{y|t}\sqrt{R^2_{Y\sim U|T}}\|\Sigma^{-1/2}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2})\|_2, \tag{42}$$

Choosing the direction of the sensitivity vector in this way provides conservative bounds for each contrast of interest. For non-Gaussian outcomes or alternative estimands, there may not be an analytic solution to the direction which maximizes the bias, but we can still compute the direction via numerical optimization.

5.2 Robustness for Individual Contrasts

We now turn to assessing the robustness of conclusions using our sensitivity model, extending work by Cinelli and Hazlett (2020) and VanderWeele and Ding (2017) in the single treatment setting. Specifically, we propose an extension of the robustness value (RV) within our model, which characterizes the minimum strength of confounding needed to change the sign of the treatment effect. As in the previous section, the extension is most straightforward when considering the effect of a single treatment contrast, between levels $T = t_1$ and $T = t_2$.

To review briefly, in single treatment settings, Cinelli and Hazlett define the robustness value as the smallest value of $\max(R^2_{Y\sim U|T}, R^2_{T\sim U})$, needed to change the sign of the effect. A robustness value close to one means the treatment effect maintains the same sign even if nearly all the observed residual variance in the outcome is due to confounding *and* all the residual treatment variance is due to confounding. On the other hand, a robustness value close to zero means that even weak confounding would change the sign of the point estimate. In the multi-treatment setting, we can more precisely characterize the robustness of causal effects, subject to Assumptions 4 and 5.

In single treatment analyses, the smallest value of $\max(R^2_{Y\sim U|T}, R^2_{T\sim U})$ needed to change the sign of the treatment effect is achieved when $R^2_{Y\sim U|T} = R^2_{T\sim U}$. As such, the value of the single treatment robustness value can be misleading when $R^2_{Y\sim U|T}$ is very different from $R^2_{T\sim U}$. In detail, when $R^2_{Y\sim U|T} > R^2_{T\sim U}$, the single-treatment RV will be too conservative. Conversely, when $R^2_{Y\sim U|T} < R^2_{T\sim U}$ the single-treatment RV will overestimate the robustness of the effect. In the multiple treatment setting, Assumptions 4 and 5 imply that for any treatment, the fraction of treatment variance due to confounding is identifiable, which allows us to define the multi-treatment RV as the minimum value of $R^2_{Y\sim U|T}$ needed to explain away the treatment effect of interest, assuming the direction of the sensitivity vector is chosen to maximize the bias. This allows us to more precisely characterize robustness.

When the observed outcomes are Gaussian, the robustness value can be computed in closed form.

Corollary 2.2. *Assume the model (31) - (33) where Y is conditionally Gaussian given treatments. Further, assume a homoskedastic outcome with no interaction between unmeasured confounders and treatments, so that $\sigma_{y|t}$, $\Sigma_{u|t}$, and γ are invariant to the level of t. Then,*

$$RV_{t_1,t_2} = min\left(\frac{(\mu_{y|t_1} - \mu_{y|t_2})^2/\sigma^2_{y|t}}{\|\Sigma^{\dagger}_{u|t})^{1/2}(\mu_{u|t_1} - \mu_{u|t_2})\|^2_2}, 1\right). \tag{43}$$

Proof: Immediate from Corollary 2.1, by setting $Bias_{t_1,t_2}$ *equal to the observed difference in outcomes,* $\mu_{y|t_1} - \mu_{y|t_2}$.

Note that because the bias is bounded, it is possible that for some treatment effects, no matter how much variance in the outcome is due to unmeasured confounding, the sign of the effect will remain the same. In this case, by convention we say anything with an RV of 1 is "robust". The numerator of the robustness value corresponds to the squared difference in mean outcomes under each treatment arm, in units of residual standard deviations. Likewise, the denominator is the squared difference in confounder means in each treatment arm, in units of residual standard deviations. When the (scaled) difference in mean outcomes is larger than the (scaled) difference in unmeasured confounders, the causal effect is robust.

RV metrics for alternative estimands and/or non-Gaussian data can still be computed using the same principle. For example, when the observed outcome is binary, the RV can be computed numerically by solving $RR_{t_1,t_2} = 1$, which corresponds to the minimum strength of confounding needed for the observed risk ratio (RR) to equal to one. We can view this robustness value as a multi-treatment parametric analog of the "E-value" proposed by VanderWeele and Ding (2017).

In our setting, we can also make stronger statements about robustness than in the single treatment setting: under the latent variable model, it is possible to declare an effect robust to *any* level of confounding. In particular, when the latent variable model implies $R^2_{T\sim U} < 1$ (i.e., we have confounder overlap), then even when $R^2_{Y\sim U|T} = 1$, the ignorance region is bounded (Corollary 2.1). When this ignorance region excludes zero, we declare the effect "robust". This operation is consistent with the result in Miao et al. (2020), showing that hypotheses of zero effect can be tested in this setting, even if the treatment effect cannot be identified.

5.3 Multiple Contrast Criteria

So far, we have examined the sensitivity of causal conclusions by exploring the marginal bounds on a treatment contrast in isolation. However, the multi-treatment setting presents opportunities for exploring sensitivity models in new ways. Here we characterize a choice of sensitivity vector γ by concurrently considering its implications for the causal effects of multiple treatment contrasts. Thus, while the sensitivity vector γ that gives the worst-case bias may differ across individual contrasts, here we explore criteria for selecting a single γ which concurrently incorporates implications for multiple treatment contrasts in aggregate. We term these "multiple contrast criteria" or MCCs.

Formally, for a set of treatment contrasts $\mathcal{T}^2 = \{(t_1, t_2)_k\}_{k=1}^K$, and a candidate sensitivity vector γ, let $\mathbf{PATE}_{\mathcal{T}^2}(\gamma)$ be the vector of PATEs implied by the causal model indexed by γ. An MCC is a scalar summary of this treatment effect vector, which we write as $\omega(\mathbf{PATE}_{\mathcal{T}^2}(\gamma))$. An MCC is specified by the set of constrasts $\mathcal{T}^2$ and the summary function ω, both of which can be chosen to meet the needs of a given analysis.

MCCs can be used in many ways, but here we consider how they can be used to search for the causal model that yields the minimum norm treatment effect vector, subject to Assumptions 1-5 and a confounding limit $\mathcal{R}^2$. Specifically, we take ω to be an L_p norm for some p, and consider sensitivity vectors γ_* that satisfy:

$$\gamma_* = \underset{\gamma}{\operatorname{argmin}}\ \omega(\mathbf{PATE}_{\mathcal{T}^2}(\gamma)) \text{ subject to } R^2_{Y\sim U|T}(\gamma) \leq \mathcal{R}^2 \tag{44}$$

where $R^2_{Y\sim U|T,X}(\gamma) = \frac{\gamma' \Sigma_{u|t} \gamma}{\sigma^2_{y|t}}$ is the partial fraction of outcome variance explained by confounding for sensitivity vector γ. Causal models selected in this way are often highly interpretable, in terms of either "worst case" effect sizes or established prior knowledge. For example, we can choose ω to be the L_∞ norm, so that γ_* is the sensitivity vector that minimizes the maximum absolute treatment effect across contrasts. Alternatively, we could choose ω to be the L_1 or L_2 norm of the treatment effects to incorporate prior knowledge that might imply small "typical" effect sizes. We demonstrate how this minimization approach can be used to express prior knowledge about small effects in simulated data in Section 6.2, and how it can be used to evaluate robustness on a real data set in Section 7.

6 Simulation Studies

In this Section, we demonstrate our sensitivity analysis workflow in several numerical simulations. The goal of these simulations is twofold: first, to demonstrate some of the operating characteristics of the approach in settings that are more realistic than the linear Gaussian settings we characterized analytically; and secondly, to show how exploratory tools like calibration, robustness analysis, and MCCs can be used to draw conclusions and choose interesting candidate models.

We consider two broad simulation settings. In the first setting, we construct simulations with non-linear responses to treatment to show how the ignorance regions returned by our method can vary in different scenarios. In the second setting, we construct a simulation that mimics the structure of a Genome Wide Association Study (GWAS). Here, we examine the behavior of our method when a popular approximate latent

variable method—the Variational Auto Encoder (VAE)—is used to estimate the effects of latent confounders, and demonstrate how MCCs can be useful tools for using prior information to choose potentially useful causal models from the set that is compatible with the observed data. In both subsections, we simulate data from the following generating process:

$$U := \epsilon_u, \quad \epsilon_u \sim N(0, I), \tag{45}$$

$$T := h_T(BU + \epsilon_t), \quad \epsilon_t \sim N(0, \sigma_t^2 I) \tag{46}$$

$$Y := h_{Y|T}(g(T) + \gamma' U + \epsilon_y), \quad \epsilon_y \sim N(0, \sigma^2) \tag{47}$$

The functions $h_{Y|T}$ and h_T are chosen according to be either the identity for Gaussian data, or an indicator function for binary data.

6.1 Example with Non-Linear Response Functions

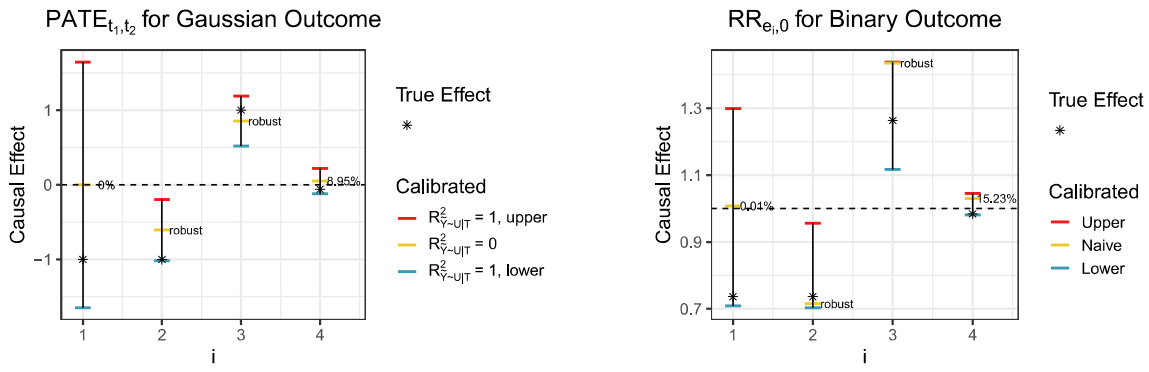
We start by exploring variation in the size of ignorance regions for different contrasts in a simple simulated example with four treatments where the outcome is a nonlinear function of these treatments. We consider two cases: first, a case where Y is Gaussian with $h_{Y|T}(\tilde{Y}) = \tilde{Y}$; and secondly, a case where Y is binary with $h_{Y|T}(\tilde{Y}) = I_{\tilde{Y}>0}$. We aim to estimate the $\text{PATE}_{e_i,0}$ for Gaussian outcome and $\text{RR}_{e_i,0}$ for binary outcome, where e_i denotes the ith canonical vector, i.e. the vector with a 1 in the i-th coordinate and 0's elsewhere.

In both examples, we generate the data with a 1-dimensional latent confounder ($m = 1$), $k = 4$ treatments, $B = [2, 0.5, -0.4, 0.2]$, $\sigma_t^2 = 1$, $\gamma = 2.8$, $\sigma^2 = 1$, $h_T(\tilde{T}) = \tilde{T}$ and

$$g(T) = 3T_1 - T_2 + T_3 I_{T_3>0} + 0.7T_3 I_{T_3\leq 0} - 0.06T_4 - 4T_1^2.$$

Based on the choice of g, contrasts along the jth dimension of T have effects of widely varying magnitude. Based on our choice for B, the worst-case confounding bias also varies significantly across contrasts. For example, the effect of confounding is larger when estimating the treatment of T_1, since the first entry of B has the largest magnitude, meaning T_1 is the feature most correlated with U. In order to demonstrate this in simulation, we first apply probabilistic PCA (PPCA) to estimate the distribution $f(u \mid t)$, and then model $f(y \mid t)$ using Bayesian Additive Regression Tree (BART) with R package BART (McCulloch et al., 2018).

For Gaussian outcomes, the width of the ignorance regions are larger for the treatments most correlated



(a) Gaussian Outcome (b) Binary Outcome

Figure 2: Estimated ignorance region for e_i in case when $\tilde{Y}$ is nonlinear in T. (a) $R^2_{\tilde{Y}\sim U|T} = 0$ denotes the treatment effects estimated based on the observed data only, i.e., under the assumption of no confoundedness. $R^2_{\tilde{Y}\sim U|T} = 1$ and $R^2_{\tilde{Y}\sim U|T} = 1$ correspond to the case when all residual variation in Y is due to the confounding, respectively denoting the upper and lower bounds of the ignorance region. (b) In the binary setting, even though the estimand is a non-linear function of the latent Gaussian outcome, the width of the ignorance region and general robustness pattern is largely consistent with the implications of Corollary 1.1.

with confounders as characterized in Corollary 2.1 (see Figure 2). Since B is a vector, the width of the ignorance region of PATE_{t_1,t_2} can be examined by looking at the dot product between B and the treatment contrasts. The larger the dot product, the wider the ignorance region. As expected, the ignorance region of the treatment effect is widest when $t_1 = e_1$ (RV $\approx 0\%$) and narrowest when $t_1 = e_4$, since $B'e_1$ has the largest magnitude while $B'e_4$ has the smallest. Despite the fact that $t_1 = e_4$ has the smallest ignorance region, it is not robust to confounding because the naive effect is already close to zero (RV = 9%). For the second and third treatment contrasts, estimates are robust to confounders, as their entire ignorance regions exclude 0. These results require the Gaussian copula assumption (Assumption 5), but in the Appendix, we show via simulation that alternative choices for the copula yield results that lie within the worst-case Gaussian bounds for $R^2_{Y\sim U|T} = 1$. In Appendix Figure 9, we include the causal effects implied by some Archimedean copulas as well as an example with a non-monotone copula (e.g. quadratic relationship between U and Y). Thus, while the Gaussian copula will not hold exactly in practice, it is likely a that the Gaussian bounds cover the true causal effect when the true copula is non-Gaussian.

For the simulation with binary outcomes, we compute ignorance regions for the risk ratio. Although we do not have a theoretical result about the ignorance regions of the risk ratio, the general trends in the size of the ignorance region and the robustness of effects are comparable to the Gaussian. Most notably, the treatments with the largest ignorance regions are still those that are most correlated with the confounder.

On the other hand, because the outcome is non-linear in U, the naive estimate is not at the center of the ignorance region (Figure 2b). In fact, the ignorance region is also non-monotone in $R^2_{\tilde{Y}\sim U|T}$ because the variance of the intervention distribution also depends on γ. In this case, one of the endpoints of the ignorance region corresponds to $R^2_{\tilde{Y}\sim U|T} = 1$ but the other does not. We compute the endpoints of the ignorance region numerically (see Appendix C.3 for more details).

6.2 Example with Simulated Genome Wide Association Study

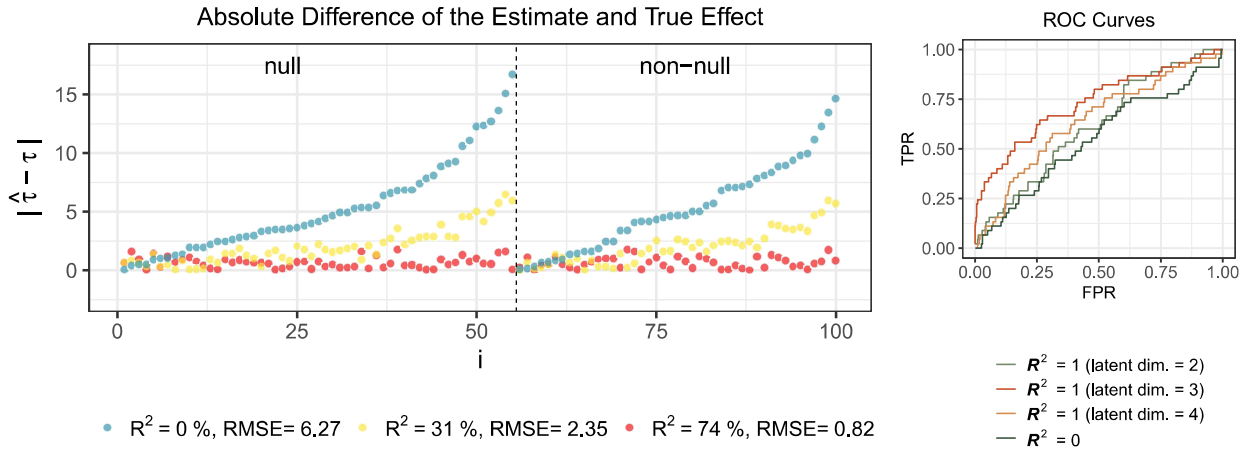
We now explore a slightly more complex setting motivated by applications in biology, particularly in genome wide association studies (GWAS). GWAS investigate the association between hundreds or thousands of genetic features (i.e., single nucleotide polymorphisms, or SNPs) and observable traits (i.e., phenotypes), such as disease status. Despite having "association" in the name, measures of association in GWAS are often adjusted to afford a causal interpretation in which conclusions speak to how a phenotype would change if the genome were intervened upon. For example, most analyses adjust for "population structure", which correspond to broad genetic patterns induced by population dynamics that are often confounded with geography, ancestry, environment, and other lifestyle factors (Price et al., 2006; Song et al., 2015). Wang and Blei (2019) cite this literature as motivation for their work.

Here, we construct a simulated GWAS to demonstrate two properties of our sensitivity analysis method. First, we show that flexible latent variable models can be plugged into our sensitivity model. Secondly, we demonstrate how minimizing multiple contrast criteria (MCC) can be used to select interesting candidate models that conform to broad hypotheses about the nature of genetic effects.

In this simulation, we generate data with high-dimensional binary treatments (SNPs), and set the true causal effects to be mostly small, with a small fraction of treatments having effects of larger magnitudes. The simulation is then designed so that unobserved confounding biases estimates for each of these treatment effects, obscuring the difference between large and small effects. To generate data, we follow the template in Equations 45–47. We generate data with $m = 3$ latent confounders and $k = 500$ treatments, $T \in \{0, 1\}^k$, where $T_j = 1$ if the the jth site shows a deviation from the baseline sequence (i.e., the presence of at least one minor allele). We set the response function $g(T) = \tau' T$ to be linear in the treatments (a common assumption in GWAS), and set the outcome Y to be Gaussian by setting $h_Y(\tilde{Y}) = \tilde{Y}$. We focus on estimating

$$\frac{1}{n}\sum_{i=1}^{n} PATE_{t_i^j, t_i^{-j}} \text{ for all } j = 1, \cdots, k, \tag{48}$$

where t_i^j and t_i^{-j} correspond to the i^{th} observed treatment vector with the j^{th} SNP set to be 1 and 0 respectively. Note that since $g(T)$ is linear in T, $\frac{1}{n}\sum_{i=1}^{n} PATE_{t_i^j, t_i^{-j}} = \tau_j$, the j^{th} element of τ. We generate τ from a two component mixture with 90% of the coefficients from a Uniform$(-0.1,\ 0.1)$ (small effects) and 10% from a Uniform$(-2,\ 2)$ (large effects). We assume that there are $m = 3$ latent confounders.



(a) L1 minimizing effects

(b) Non-null classification

Figure 3: Causal inference with 500 binary treatments with $k = 3$ latent confounders. The omitted confounder bias of the naive estimates are large for both the null and non-null effects due to unmeasured confounding. (a) The absolute difference between the true effects and the inferred minimum L1-norm treatment effects shown for fifty randomly chosen small effects ("null" contrasts) and all large effects ("non-null" contrasts) for three different limits on the magnitude of confounding, $\mathcal{R}^2 \in \{0, 0.3, 1.0\}$. When $\mathcal{R}^2 = 1$, the overall L1 minimizer of the treatment effects is achieved for the sensitivity vector which explains $R^2_{Y\sim U|T} = 74\%$ of the residual outcome variance. (b) We construct a simple non-null classifier from minimum L1 treatment effects with $\mathcal{R}^2 = 1$ and naive effects ($\mathcal{R}^2 = 0$). The blue curve represents the ROC curves from the naive estimates and the green, yellow and red curves represents the L1 minimizer of the treatment effect estimates for inferred confounder models with dimensions $\hat{k} \in \{2, 3, 4\}$. The area under the curve (AUC) for the naive estimates is 0.54, whereas the AUC for the L1-minimized estimates are 0.61 ($\hat{k} = 2$), 0.73 ($\hat{k} = 3$) and 0.64 ($\hat{k} = 4$).

We consider a model for the observed data with two components, paying special attention to the latent confounder model. In particular, we model the conditional distribution of confounders given treatment $f(u \mid t)$ using a variational autoencoder (VAE), which is a popular, flexible neural network–based approximate latent variable model. This model is particularly appropriate because it yields an approximate Gaussian conditional distribution $f(u \mid t)$, even for discrete T as we have here. (We discuss latent confounder inference with VAEs in more detail in Appendix C.2.) We fit the observed outcome model $f(y \mid t)$ using a simple linear regression, ignoring confounding, which corresponds to the setting in which $R^2_{Y\sim U|T} = 0$.

Worst-Case Ignorance Regions. With this simulation setup, we first examine whether the ignorance regions contain the true causal effects. Importantly, because the VAE is an approximate latent variable model, and we are currently ignoring estimation uncertainty, it is not immediate that the ignorance regions should be valid. We find that, even using our plug-in approach, the worst case ignorance regions cover 498 out of 500 of the true treatment effects. In all cases, the worst case bounds communicate substantial fundamental uncertainty about the true treatment effects (See Appendix Figure 11).

Finding Candidate Models with MCCs. Investigators often have strong hypotheses about the aggregate properties of SNP treatment effects. For example, while some phenotypes can be predominantly explained by only a small number of SNPs, other phenotypes may be more plausibly described by the omnigenic hypothesis, which suggests that some observable effects must be explained by the sum of many small effects across many SNPs (Boyle et al., 2017). Here, we show that some of these aggregate hypotheses can be formalized in terms of MCCs, and in these cases, the MCC minimization procedure from Section 5.3 can be used to find useful candidate causal models that align with these hypothesis while being fully consistent with the observed data.

To motivate candidate model selection, we consider the use case of estimating effect sizes from a single coherent model, under the hypothesis that the median effect size is small. Specifically, we formalize this hypothesis by defining a MCC $\omega(\mathbf{PATE}_{\mathcal{T}^2}(\gamma))$ to be the L_1 norm of the effects of each contrast $\mathcal{T}^2 = \{(t_i^j, t_i^{-j}) : i \in (1, ..., n)\}$ for all treatments $j = 1, \cdots, k$. We then select the model that minimizes this criterion by selecting γ subject to different allowed levels of confounding $R^2_{Y\sim U|T}$.

In Figure 3a, we plot the the resulting coefficients estimates for three values of $\mathcal{R}^2$: 0 (naive effects), 0.3 and 1. Because the true effects are much smaller in magnitude than the naïve effects, the RMSE of the estimates decreases as we increase $R^2_{Y\sim U|T}$, although all effects are equally compatible with the observed data. In this simulation, the L1 norm of naive estimates is approximately 2525 and the norm of the true effects is drastically smaller at approximately 75.

Models selected using this MCC minimization procedure are also useful for the coarser goal of separating small and large effects. From the naive regression, the coefficients are overdispersed to the true causal effects and the true small coefficients are practically indistinguishable from true large coefficients. Meanwhile, models chosen with the MCC minimization procedure provide more useful signal. To formalize this, we consider a classifier that separates large and small effects using the magnitude of the inferred coefficients

as the classification score. In Figure 3b we plot the receiver operating characteristic (ROC) curves for the classifiers based on the naive estimates as well as the overall $L1$ minimizer of the treatment effects ($\mathcal{R}^2 = 1$, i.e. no limit on the value $R^2_{Y\sim U|T}$).

Importantly, the difference in conditional confounder means, $\mu_{u|t_i^j} - \mu_{u|t_i^{-j}}$, varies between non-null and null contrasts. This leads to a larger reduction in the relative magnitude of the null effects for models chosen through MCC minimization, accentuating the differences between large and small treatment effects (See Appendix Figure 12). For models selected by MCC minimization, the area under the ROC curve (AUC) increases from 0.54 (almost no ability to distinguish small and large treatments) to 0.72 ($\hat{k} = 3$, red curve). The selected model achieves nearly 25% true positive rate without accruing any false positives. Naturally, the classifier performance is the best when we fit a latent variable model with the correct number of latent factors, although the classifier based on latent variable models of dimensions $\hat{k} = 2$ and $\hat{k} = 4$ still outperform classification from naive effects. In the Discussion, we note how this approach relates to, and complements recent identification results for a similar setting in Miao et al. (2020).

7 Analysis of Mouse Obesity Data

In this Section, we apply our sensitivity analysis to mice obesity data generated by Wang et al. (2006) and Ghazalpour et al. (2006), and compiled into a single dataset by Lin et al. (2015). The data consists of body weight and gene expression levels for 17 genes in each of 227 mice, and the goal is to estimate the causal effect of the gene expression levels on mouse weight. In gene expression datasets like this one, batch effects can induce confounding when the batches are correlated with outcomes. This problem has motivated several approaches for removing sources of potential unwanted variation prior to analysis (Gagnon-Bartsch and Speed, 2012; Listgarten et al., 2010; Leek and Storey, 2007). Miao et al. (2020) analyze the mouse obesity dataset in the context of the multiple treatment problem, under the assumption that at least half of the true treatments have no causal effect on the outcome. Here, we provide a complementary analysis, and explore the broader set of causal effects that are compatible with the observed expression data.

First, we use the linear treatment and outcome model, Equations (7)-(9), to model the data. To represent the possible relationship between treatments and confounders we fit a linear factor model, which is commonly used to characterize the unmeasured confounding in gene expression studies (Gagnon-Bartsch and Speed, 2012), using the `factanal` method. From the scree plot of the singular values of the gene expression matrix,

we find that there are two singular values which exceed the rest, which suggests that an $m = 2$ confounder model is a reasonable choice (Appendix Figure 13). We then fit a Bayesian linear regression model of mouse weight on gene expression levels using the default prior distributions from the `rstanarm` package (Goodrich et al., 2020). In Appendix Table 1, we report the posterior mean for the observed regression coefficients, $\tau_{\text{pm}}^{\text{naïve}}$, as well as the endpoint of the 95% posterior credible interval closest to zero, $\tau_{\text{endpt}}^{\text{naïve}}$, for genes whose 95% posterior credible interval excludes zero. We also report the robustness value, RV, in terms of the percentage of outcome variance explained by confounding needed for the true causal effect to change sign, using $\tau_{\text{endpt}}^{\text{naïve}}$ for a more conservative measure of robustness that accounts for estimation uncertainty. Only five genes are found to be significantly different from zero without confounding. The significance of two genes, Sirpa and Avpr1a, are extremely sensitive to confounding in that confounders would only need to explain less than 2% of the residual outcome variance to change the sign of the effect. In contrast, while the gene Gstm2 does not have the largest magnitude of $\tau_{\text{pm}}^{\text{naïve}}$ among the significant genes, it is the most robust to confounding in the two-factor model (RV=80%).

We also use the MCC approach to examine the treatment effects with the smallest L1 and L2 norm, and compare these results to the results from Miao et al. (2020), who use robust linear regression to infer multiple causal effects under a sparsity assumption. We apply the MCC criteria to the causal effects associated with all 17 genes, to identify how small the causal effects can be in aggregate[2]. In Figure 4, we show how these additional identifying assumptions still lead to solution vectors inside the worst-case ignorance regions. To accommodate both estimation uncertainty and uncertainty due to confounding, we construct the ignorance region from the endpoints of the 95% posterior credible interval of the naive effects.

While similar in spirit, the L1 and L2 MCC methods are distinct from the null treatments approach of Miao et al. (2020) in that the MCC solutions encourage small causal effects across *all* treatments, and thus identify the solution for which the entire gene expression profile causes the smallest change in mouse weight. This MCC approach tends to reduce the number of causal outliers. In contrast, the null treatments assumption can accommodate some genes with significantly larger causal effects (e.g. Fam105a), as long as at least half of the treatments are true null genes.

Our sensitivity analysis can also be applied with more complex non-linear models. We demonstrate this using Bayesian Additive Regression Trees (BART), a method that has previously been applied for estimating (heterogeneous) causal effects in the presence of observed confounders in the single-treatment settings (Hill,

[2]See (Zheng et al., 2022) for an example in which prior knowledge is used to apply a similar shrinkage criteria to only a subset of the genes, which are *a priori* thought to have little to no effect on mouse obesity.

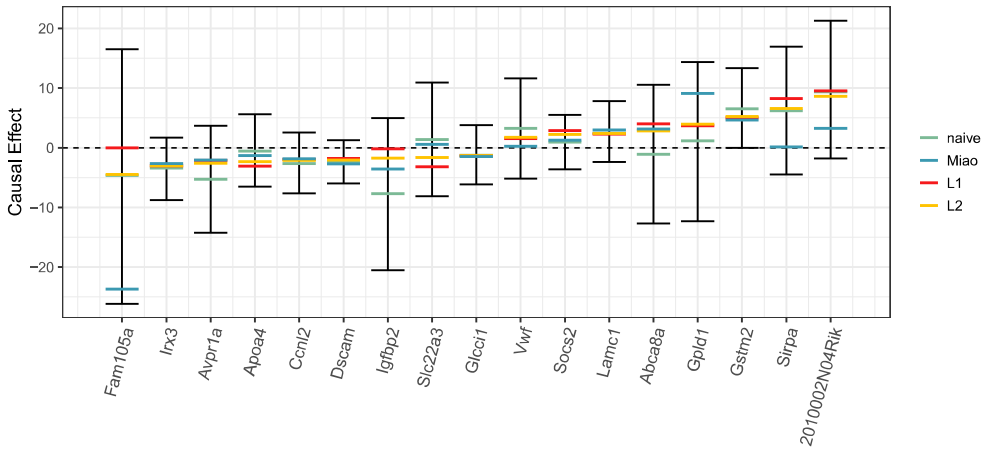


Figure 4: Possible causal effects for a one standard deviation change in expression level on mouse obesity. Black bars define the support of causal effects consist with the observed data under the linear treatment and outcome model, Equations (7)-(9). To accommodate both estimation uncertainty and uncertainty due to confounding, we construct this region from the endpoints of the 95% posterior credible interval of the naive effects. Inside the ignorance region we plot the naive treatment coefficients, the L1 and L2 minimizing MCC solutions, and the results of the null treatment approach from Miao et al. (2020).

2011; Hahn et al., 2020). Here, we use BART to infer non-linearities in the causal effects across multiple treatments while characterizing robustness to unobserved confounding. As our estimand, we consider the population average treatment effect of changing gene j from the median level to the qth quantile:

$$\tau_j^q = E[Y \mid do(t_j^q)] - E[Y \mid do(t_j^{0.5})]$$

where t_j^q denotes the treatment vector with all treatments assigned to the median level in the observed population except for the jth treatment which is assigned to the qth quantile. This is a useful set of estimands when the outcome is nonlinear in the level of the exposure, precisely because it reveals such nonlinearities.

Using BART, we found only one gene, Igfbp2, had 95% posterior credible regions for t_j^q which excluded zero for at least one q under the no unobserved confounding assumption. In Figure 5 we show the posterior 95% region as a function of the expression quantile, q, for different values of $R^2_{Y\sim U|T}$. With this particular dataset, we can see that the estimation uncertainty is fairly large relative to the uncertainty induced by 2-factor shared confounding across the multiple-treatments. For all values of q ¡ 0.7, t_j^q is not significantly different from zero, even without confounding ($R^2_{Y\sim U|T} = 0$), whereas for $q \geq 0.75$, t_j^q is significantly negative even if all the residual outcome variance was explained by shared confounding ($R^2_{Y\sim U|T} = 1$). As such, we

might conclude that high levels of Igfbp2 have an effect on mouse weight, but there is no robust difference in mouse weight for average and low levels of Igfbp2.

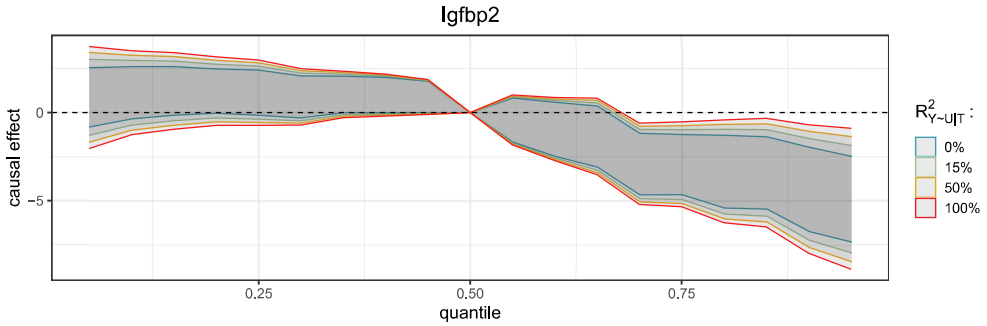


Figure 5: Causal effect of Igfbp2 expression level on mouse obesity, relative to the median level, inferred using BART. We use an m=2 in the factor model to infer the distribution of latent confounders and find that there is no discernible effect of Ifgbp2 on mouse obesity when expressed at low levels, but for levels above the 0.75 quantile, Igfbp2 reduces mouse weight and is robust to confounding.

8 Discussion

In this paper, we introduced a framework for sensitivity analysis with multiple treatments which provides further context to the growing literature on the challenges of inference in this setting. Unlike previous work, we emphasize the importance of carefully defined estimands and show that bounds on the magnitude of confounding bias depend on the particular estimands of interest. Our work also provides a practical solution to characterizing and calibrating the robustness of causal effects across multiple treatments in the presence of unobserved confounding. Code to replicate all analyses is available (Zheng, 2021b) and an R package implementing our methodology is also available and in active development (Zheng, 2021a). In addition to the GWAS simulation and gene expression dataset analyzed in this paper, we also include a reanalysis of the TMDB 5000 Movie Dataset (Kaggle, 2017) in Appendix F. This data was extensively analyzed by Wang and Blei (2019) and Grimmer et al. (2020), where the goal is to infer the causal effect of an actor's presence in a movie on revenue.

In this work, we focused primarily on partial identification results and less on practical issues pertaining to estimation. In this regard, generalizations based on joint inference of the treatment and outcome models should be explored. Joint inference is essential for accounting for both estimation uncertainty and uncertainty due to unobserved confounding. This is particularly important for the multiple contrast criteria which, as described, does not incorporate parameter uncertainty into the objective function. Future work exploring

frequentist strategies for constructing confidence intervals, perhaps leverage bootstrap methods. In a follow up to this paper, Zheng et al. (2022) consider uncertainty quantification in multi-treatment inference in the Bayesian paradigm. Here, they view MCC criteria as Bayesian priors and consider how such shrinkage priors influence posterior estimates of treatment effects. They also consider additional constraints on the calibration criteria, by considering the role of negative control exposures (NCE) (Shi et al., 2020). A set of NCEs are a subset of causes that were known a priori to have zero (or bounded) causal effect on the outcome, and can be considered a degenerate prior on some treatment effects. Such additional constraints can further shrink the bounds on all causal contrasts.

Finally, it is worth further exploring the relationship between inference with multiple treatments and inference with multiple outcomes. In Zheng et al. (2022) consider causal inference with a scalar treatment and no outcomes, but do not consider settings with both multiple treatments and multiple outcomes. The ideas in this work can be combined with the strategies applied to multi-outcome causal inference, to yield even more informative bounds on causal effects. We leave exploration of these extensions to future work.

References

Allman, E. S., C. Matias, and J. A. Rhodes (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics 37*(6A), 3099–3132.

Anderson, T. W. and H. Rubin (1956). Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry*, Volume 3.5, pp. 111–150. University of California Press.

Barber, R. F., M. Drton, N. Sturma, and L. Weihs (2022). Half-trek criterion for identifiability of latent variable models. *The Annals of Statistics 50*(6), 3174–3196.

Bica, I., A. M. Alaa, C. Lambert, and M. van der Schaar (2020). From real-world patient data to individualized treatment effects using machine learning: Current and future methods to address underlying challenges. *Clinical Pharmacology & Therapeutics*.

Boyle, E. A., Y. I. Li, and J. K. Pritchard (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell 169*(7), 1177–1186.

Cinelli, C., J. Ferwerda, and C. Hazlett (2020). sensemakr: Sensitivity analysis tools for ols in r and stata. *Submitted to the Journal of Statistical Software*.

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

Cinelli, C., D. Kumor, B. Chen, J. Pearl, and E. Bareinboim (2019). Sensitivity analysis of linear structural causal models. In *ICML*.

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

D'Amour, A. (2019a). Comment: Reflections on the deconfounder. *Journal of the American Statistical Association 114*(528), 1597–1601.

D'Amour, A. (2019b). On multi-cause approaches to causal inference with unobserved counfounding: Two cautionary failure cases and a promising alternative. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 3478–3486.

Daniels, M. J. and J. W. Hogan (2008). *Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis*. CRC Press.

Dorie, V., M. Harada, N. B. Carnegie, and J. Hill (2016). A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Statistics in Medicine 35*(20), 3453–3470.

Everett, B. (2013). *An introduction to latent variable models*. Springer Science & Business Media.

Fan, Y. and S. S. Park (2010). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory 26*(3), 931–951.

Firpo, S. and G. Ridder (2019). Partial identification of the treatment effect distribution and its functionals. *Journal of Econometrics 213*(1), 210–234.

Flores, C. A. and A. Flores-Lagunes (2013). Partial identification of local average treatment effects with an invalid instrument. *Journal of Business & Economic Statistics 31*(4), 534–545.

Franks, A. M., A. D'Amour, and A. Feller (2019). Flexible sensitivity analysis for observational studies without observable implications. *Journal of the American Statistical Association*, 1–33.

Gagnon-Bartsch, J. A. and T. P. Speed (2012). Using control genes to correct for unwanted variation in microarray data. *Biostatistics 13*(3), 539–552.

Gastwirth, J. L., A. M. Krieger, and P. R. Rosenbaum (1998). Dual and simultaneous sensitivity analysis for matched pairs. *Biometrika 85*(4), 907–920.

Gavish, M. and D. L. Donoho (2014). The optimal hard threshold for singular values is $4/\sqrt{3}$. *Information Theory, IEEE Transactions on 60*(8), 5040–5053.

Ghazalpour, A., S. Doss, B. Zhang, S. Wang, C. Plaisier, R. Castellanos, A. Brozell, E. E. Schadt, T. A. Drake, A. J. Lusis, et al. (2006). Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS genetics 2*(8), e130.

Ghosh, P., M. S. M. Sajjadi, A. Vergari, M. Black, and B. Scholkopf (2020). From variational to deterministic autoencoders. In *International Conference on Learning Representations*.

Goodrich, B., J. Gabry, I. Ali, and S. Brilleman (2020). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.21.1.

Gopalan, P., J. M. Hofman, and D. M. Blei (2013). Scalable recommendation with poisson factorization. *arXiv preprint arXiv:1311.1704*.

Greenland, S. (1996). Basic methods for sensitivity analysis of biases. *International journal of epidemiology 25*(6), 1107–1116.

Grimmer, J., D. Knox, and B. M. Stewart (2020). Naive regression requires weaker assumptions than factor models to adjust for multiple cause confounding. *arXiv preprint arXiv:2007.12702*.

Guo, W., M. Yin, Y. Wang, and M. Jordan (2022). Partial identification with noisy covariates: A robust optimization approach. In *Conference on Causal Learning and Reasoning*, pp. 318–335. PMLR.

Gustafson, P. (2015). *Bayesian inference for partially identified models: Exploring the limits of limited data*. Chapman and Hall/CRC.

Gustafson, P., L. C. McCandless, et al. (2018). When is a sensitivity parameter exactly that? *Statistical Science 33*(1), 86–95.

Hahn, P. R., J. S. Murray, and C. M. Carvalho (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects (with discussion). *Bayesian Analysis 15*(3), 965–1056.

Hao, W., M. Song, and J. D. Storey (2015). Probabilistic models of genetic variation in structured populations applied to global human studies. *Bioinformatics 32*(5), 713–721.

Heckman, J. J., J. Smith, and N. Clements (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies 64*(4), 487–535.

Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics 20*(1), 217–240.

Hofert, M. (2008). Sampling archimedean copulas. *Computational Statistics & Data Analysis 52*(12), 5163–5174.

Horn, R. (1985). *Matrix analysis*. Cambridge Cambridgeshire New York: Cambridge University Press.

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review 93*(2), 126–132.

Kaggle (2017, Sep). Tmdb 5000 movie dataset. data retrieved from Kaggle, https://www.kaggle.com/tmdb/tmdb-movie-metadata.

Kong, D., S. Yang, and L. Wang (2019). Multi-cause causal inference with unmeasured confounding and binary outcome. *arXiv preprint arXiv:1907.13323*.

Lechner, M. (1999, December). Identification and Estimation of Causal Effects of Multiple Treatments Under the Conditional Independence Assumption. IZA Discussion Papers 91, Institute of Labor Economics (IZA).

Leek, J. T. and J. D. Storey (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS genetics 3*(9), e161.

Lin, W., R. Feng, and H. Li (2015). Regularization methods for high-dimensional instrumental variables regression with an application to genetical genomics. *Journal of the American Statistical Association 110*(509), 270–288.

Listgarten, J., C. Kadie, E. E. Schadt, and D. Heckerman (2010). Correction for hidden confounders in the genetic analysis of gene expression. *Proceedings of the National Academy of Sciences 107*(38), 16465–16470.

Lopez, M. J., R. Gutman, et al. (2017). Estimation of causal effects with multiple treatments: a review and new ideas. *Statistical Science 32*(3), 432–454.

Lopez, R., P. Boyeau, N. Yosef, M. I. Jordan, and J. Regier (2020). Decision-making with auto-encoding variational bayes. *arXiv preprint arXiv:2002.07217*.

Louizos, C., U. Shalit, J. M. Mooij, D. Sontag, R. Zemel, and M. Welling (2017). Causal effect inference with deep latent-variable models. In *Advances in Neural Information Processing Systems*, pp. 6446–6456.

Manski, C. F. (2003). *Partial identification of probability distributions*, Volume 5. Springer.

Manski, C. F. (2008). *Identification for prediction and decision*. Harvard University Press.

Mardia, K. V., J. T. Kent, and J. M. Bibby (1980). *Multivariate analysis*. Academic press.

McCulloch, R., R. Sparapani, R. Gramacy, C. Spanbauer, and M. Pratola (2018). *BART: Bayesian Additive Regression Trees*. R package version 1.9.

Miao, W., W. Hu, E. L. Ogburn, and X. Zhou (2020). Identifying effects of multiple treatments in the presence of unmeasured confounding.

Minka, T. P. (2001). Automatic choice of dimensionality for pca. In *Advances in neural information processing systems*, pp. 598–604.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Ogburn, E. L., I. Shpitser, and E. J. Tchetgen Tchetgen (2019). Comment on "blessings of multiple causes". *Journal of the American Statistical Association 114*(528), 1611–1615.

Ogburn, E. L., I. Shpitser, and E. J. Tchetgen Tchetgen (2020). Counterexamples to" the blessings of multiple causes" by wang and blei. *arXiv preprint arXiv:2001.06555*.

Pearl, J. (2009). *Causality*. Cambridge university press.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics 38*(8), 904–909.

Pu, Y., Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin (2016). Variational autoencoder for deep learning of images, labels and captions. In *Advances in neural information processing systems*, pp. 2352–2360.

Robins, J. M. (1997). Non-response models for the analysis of non-monotone non-ignorable missing data. *Statistics in Medicine 16*(1), 21–37.

Robins, J. M., A. Rotnitzky, and D. O. Scharfstein (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical models in epidemiology, the environment, and clinical trials*, pp. 1–94. Springer.

Rohe, K. and M. Zeng (2020). Vintage factor analysis with varimax performs statistical inference. *arXiv preprint arXiv:2004.05387*.

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

Rubin, D. B. (1980). Comment. *Journal of the American Statistical Association 75*(371), 591–593.

Shi, X., W. Miao, and E. T. Tchetgen (2020). A selective review of negative control methods in epidemiology. *arXiv preprint arXiv:2009.05641*.

Song, M., W. Hao, and J. D. Storey (2015). Testing for genetic associations in arbitrarily structured populations. *Nature genetics 47*(5), 550–554.

Swanson, S. A., M. A. Hernán, M. Miller, J. M. Robins, and T. S. Richardson (2018). Partial identification of the average treatment effect using instrumental variables: review of methods for binary instruments, treatments, and outcomes. *Journal of the American Statistical Association 113*(522), 933–947.

Tamer, E. (2010). Partial identification in econometrics. *Annu. Rev. Econ. 2*(1), 167–195.

Tipping, M. E. and C. M. Bishop (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61*(3), 611–622.

VanderWeele, T. J. and O. A. Arah (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*, 42–52.

VanderWeele, T. J. and P. Ding (2017, July). Sensitivity analysis in observational research: Introducing the e-value. *Annals of Internal Medicine 167*(4), 268.

VanderWeele, T. J., B. Mukherjee, and J. Chen (2012). Sensitivity analysis for interactions under unmeasured confounding. *Statistics in medicine 31*(22), 2552–2564.

VanderWeele, T. J. and I. Shpitser (2013). On the definition of a confounder. *Annals of statistics 41*(1), 196.

Vansteelandt, S., E. Goetghebeur, M. G. Kenward, and G. Molenberghs (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica 16*(3), 953–979.

Veitch, V. and A. Zaveri (2020). Sense and sensitivity analysis: Simple post-hoc analysis of bias due to unobserved confounding. *arXiv preprint arXiv:2003.01747*.

Wang, S., N. Yehya, E. E. Schadt, H. Wang, T. A. Drake, and A. J. Lusis (2006). Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS genetics 2*(2), e15.

Wang, Y. and D. M. Blei (2019). The blessings of multiple causes. *Journal of the American Statistical Association 114*(528), 1574–1596.

Zhang, L., Y. Wang, A. Ostropolets, J. J. Mulgrave, D. M. Blei, and G. Hripcsak (2019). The medical deconfounder: Assessing treatment effects with electronic health records. *arXiv preprint arXiv:1904.02098*.

Zheng, J. (2021a). Copsens: Copula-based sensitivity analysis method for unobserved confounding in multi-treatment inference. https://github.com/JiajingZ/CopSens.

Zheng, J. (2021b). Replication code for "copula-based sensitivity analysis for observational multi-treatment causal inference". https://github.com/JiajingZ/CopulaSensitivity.

Zheng, J., A. D'Amour, and A. Franks (2022). Bayesian inference and partial identification in multi-treatment causal inference with unobserved confounding. In *International Conference on Artificial Intelligence and Statistics*, pp. 3608–3626. PMLR.

Zheng, J., J. Wu, A. D'Amour, and A. Franks (2022). Sensitivity to unobserved confounding in studies with factor-structured outcomes. *arXiv preprint arXiv:2208.06552*.

A Theory

A.1 General Contrast Estimation Algorithm

Algorithm 1 Marginal Contrast Estimation for Arbitrary Copulas

function ComputeMean(t, ψ)
  for $k = 1, 2, \ldots, M$ **do**
    Sample y_k from $f(y \mid t)$
    for $i = 1, 2, \ldots, n$ **do**
      for $j = 1, 2, \ldots, N$ **do**
        Sample u_{ij} from $f(u \mid t_i)$
        Compute $c_{ij} \leftarrow c_\psi(y_k, u_{ij} \mid t)$
      end for
    end for
    Compute $w_k \leftarrow \frac{1}{nN} \sum_{ij} c_{ij}$
  end for
  return $\frac{1}{M} \sum_k \nu(y_k) w_k$
end function
Return $\tau(\text{ComputeMean}(t_1, \psi), \text{ComputeMean}(t_2, \psi)$

A.2 Contrast Estimation Algorithm with Gaussian Copulas

Suppose that the observed data is generated by model (31) - (33):

$$\begin{aligned} f(u \mid t) &\sim N(\mu_{u|t}, \Sigma_{u|t}), \\ \tilde{Y} &= \gamma_t' U + \epsilon_{\tilde{Y}}, \\ Y &= F_{Y|T}^{-1}(\Phi(\tilde{Y} - \gamma_t' \mu_{u|t})), \end{aligned}$$

given sensitivity vectors γ_t, any marginal contrast estimand can be estimated by Algorithm 2.

Algorithm 2 Marginal Contrast Estimation with Gaussian Copulas.

function ComputeMean(t, γ)
  for $i = 1, 2, \ldots, n$ **do**
    $\mu_i \leftarrow \gamma'(\mu_{u|t_i} - \mu_{u|t})$
    for $j = 1, 2, \ldots, nSim$ **do**
      Sample $\tilde{y}_{ij}$ from $N(\mu_i, 1)$
      $y_{ij} \leftarrow F_{Y|t}^{-1}(\Phi(\tilde{y}_{ij}))$
    end for
  end for
  return $\frac{1}{n} \sum_{ij} v(y_{ij})$
end function
Return $\tau(\text{ComputeMean}(t_1, \gamma), \text{ComputeMean}(t_2, \gamma))$

Derivation of Algorithm 2

Since we have Equation (29) and (30), furthermore, we can write

$$E[v(Y) \mid do(t)] \;= \iint v(y) f(y \mid \tilde{y}) w_\psi(\tilde{y}, t) f(\tilde{y} \mid t) d\tilde{y} dy, \tag{49}$$

where $w_\psi(\tilde{y}, t) \approx \frac{1}{|\mathcal{T}|} \sum_{t_i \in \mathcal{T}} \left[\int c_\psi(F_{\tilde{Y}|t}(\tilde{y}), F_{U|t}(u) \mid t) f(u \mid t_i) du \right]$. To verify Algorithm 2, we only need to show that

$$\int f(\tilde{y} \mid t, u) f(u \mid t_i) du \sim N(\gamma'(\mu_{u|t_i} - \mu_{u|t}),\ 1), \tag{50}$$

where $f(\tilde{y} \mid t, u) = f(\tilde{y} \mid t) c_\psi(F_{\tilde{Y}|t}(\tilde{y}), F_{U|t}(u) \mid t)$. Based on model (31) - (33), we have

$$f(u \mid t_i) \sim N(\mu_{u|t_i},\ \Sigma_{u|t}), \tag{51}$$
$$f(\tilde{y} \mid t, u) \sim N(\gamma_t'(u - \mu_{u|t}),\ 1 - \gamma_t' \Sigma_{u|t} \gamma_t). \tag{52}$$

By integrating out the U,

$$\int f(\tilde{y} \mid t, u) f(u \mid t_i) du = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(y - \gamma_t'(\mu_{u|t_i} - \mu_{u|t}))^2}{2} \right\}. \tag{53}$$

A.3 Additional Figures Illustrating the Proposed Sensitivity Analysis

We illustrate some key insights from Corollary 1.1 in Figure 6 where we display the worst-case bias as a function of the treatment contrasts, $(t_1 - t_2)$. In this illustration, we assume that $(t_1 - t_2)$ lies on a plane spanned by u_1^B and n_0^B, an arbitrary vector in the null space of B. We let $\theta = \arccos((t_1 - t_2)' u_1^B)$ be the angle of $(t_1 - t_2)$ relative to u_1^B, Figure 6a. Figure 6b depicts the bias as function of θ for different values of $R^2_{Y \sim U|T}$. When $(t_1 - t_2)$ is in the null space of B, PATE_{t_1,t_2} is identified because the confounder distributions are identical in the two treatment arms, i.e. there is no confounding for this particular contrast. When $(t_1 - t_2)$ is colinear with u_1^B the scaled difference in means of u is largest, which implies the largest worst-case bias for the treatment effect, Figure 6c (left). Even when PATE_{t_1,t_2} is identified, we emphasize that $\text{PATE}_{t_1,\cdot}$ and $\text{PATE}_{t_2,\cdot}$ are both biased, since the distribution of confounders in the treatment arm differs from the distribution of confounder in the superpopulation, Figure 6c (right). As noted by others, identification of PATE_{t_1,t_2} for $(t_1 - t_2)$ in the null space of B arises due to bias cancellation in intervention means of the two treatment arms (Grimmer et al., 2020).

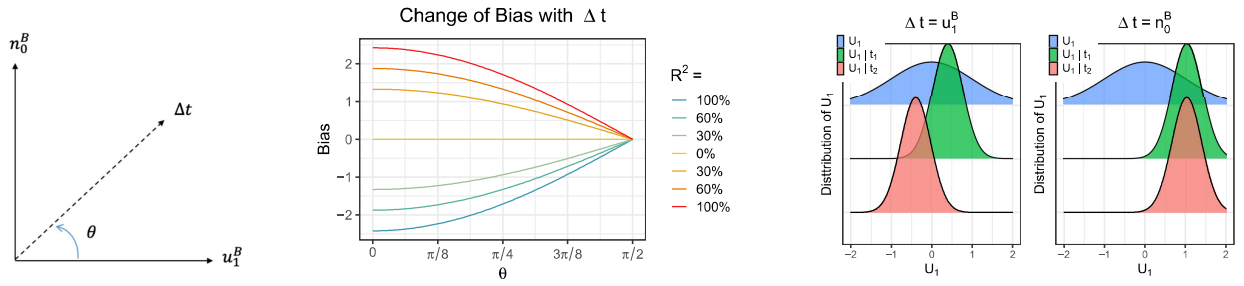
A.4 Proof of Theorem 1

Theorem 1. *Suppose that the observed data is generated by model (7)-(9):*

$$\begin{aligned} U &= \epsilon_u \\ T &= BU + \epsilon_t \\ Y &= \tau' T + \gamma' U + \epsilon_y \end{aligned}$$

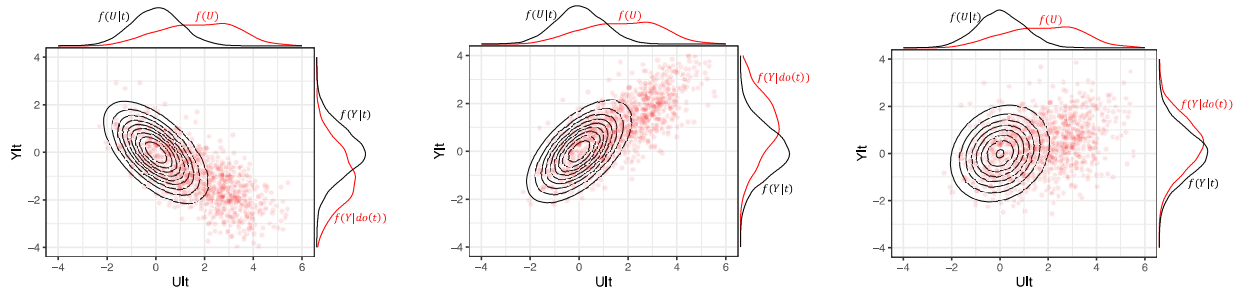
with $\Lambda_{t|u} > 0$. Then, $\forall \gamma$ satisfying Assumptions 1 and 2,

$$\gamma' \Sigma_{u|t} \gamma \leq \sigma^2_{y|t} \tag{54}$$



(a) $\Delta t := t_1 - t_2$ (b) Estimation Bias of $\text{PATE}_{(t_1 - t_2)}$ (c) Distribution of U_1

Figure 6: Illustration of Corollary 1.1. (a) We parameterize $(t_1 - t_2)$ with θ, the angle between n_0^B, a vector in the null space of B, and u_1^B, the first left singular vector of B. (b) The confounding bias of naive estimates of PATE_{t_1,t_2} changes with θ and depends on $R^2_{Y\sim U|T}$. (c) Confounder densities in different populations. The blue, green, red densities denote distributions of U_1 in the observed population, the subpopulation receiving t_1 and the subpopulation receiving treatment t_2 respectively. Observed data estimates of PATE_{t_1,t_2} are unbiased when $(t_1 - t_2) = n_0^B$, since the confounder distributions are the same in two treatment arms. However, observed data estimates of $\text{PATE}_{t_1,\cdot}$ and $\text{PATE}_{t_2,\cdot}$ are biased since in general the superpopulation distribution of the confounder is different.



(a) $\sqrt{R^2_{\tilde{Y}\sim U|T}} = 0.6$ (b) $\sqrt{R^2_{\tilde{Y}\sim U|T}} = 0.6$ (opposite sign) (c) $\sqrt{R^2_{\tilde{Y}\sim U|T}} = 0.2$

Figure 7: Differences between observed and intervention densities as a function of the fraction of outcome variance explained by a single confounder. The black contours depict the conditional Gaussian copula, $c_\gamma(F_{Y|t}(y), F_{U|t}(u) \mid t)$ whereas red points represent samples from the joint distribution, $f(y, u \mid do(t)) \propto f(y \mid t)c_\gamma(F_{Y|t}(y), F_{U|t}(u) \mid t)f(u)$. We visualize the shift in the outcome density for different conditional correlations and note that smaller values of $R^2_{Y\sim U|T}$ imply smaller biases in the outcome despite large imbalance in the distribution of U.

For any given t_1, t_2, we have

$$Bias^2_{t_1,t_2} \leq \sigma^2_{y|t} R^2_{Y\sim U|T} \parallel \Sigma^{-1/2}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2}) \parallel^2_2 . \tag{55}$$

The bound is achieved when γ is colinear with $\Sigma^{-1}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2})$ and is maximized when all the residual outcome variance is due to unmeasured confounders, e.g. $R^2_{Y\sim U|T} = 1$.

Proof. Under model (7)-(9), the variance of the observed outcome equals

$$\begin{aligned} \sigma^2_{y|t} &:= Var(Y \mid T) \\ &= \sigma^2 + \gamma'(I - B'(BB' + \Lambda_t)^{-1}B)\gamma \\ &= \sigma^2 + \gamma'\Sigma_{u|t}\gamma, \end{aligned} \tag{56}$$

where $\gamma'\Sigma_{u|t}\gamma$ corresponds to the confounding variation, and σ^2 stands for the non-confounding variation in the residual of observed outcome. Hence, the fraction of confounding variation in the residual of Y, $R^2_{Y\sim U|T}$, can be expressed in terms of equation (18), which produces a constrain for γ (equation (19)) that the confounding variation in the residual of Y, $\gamma'\Sigma_{u|t}\gamma$, should not be larger than $\sigma^2_{y|t}R^2_{Y\sim U|T}$ for a given level of $R^2_{Y\sim U|T}$.

Let

$$Z := \Sigma^{1/2}_{u|t}\gamma, \tag{57}$$

then the omitted variable bias in equation (17) can be written as

$$Bias_{t_1,t_2} = Z'\Sigma^{-1/2}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2}),$$

where $Z'Z \leq \sigma^2_{y|t}R^2_{Y\sim U|T}$, implied by inequality (19).

Therefore,

$$\begin{aligned} \text{Bias}^2_{t_1,t_2} &= Z'\Sigma^{-1/2}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2})(\mu_{u|t_1} - \mu_{u|t_2})'\Sigma^{-1/2}_{u|t}Z & (58) \\ &\leq \sigma^2_{y|t}R^2_{Y\sim U|T} \parallel \Sigma^{-1/2}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2}) \parallel^2_2, & (59) \end{aligned}$$

where the bounds are reached when $Z = \frac{\sqrt{\sigma^2_{y|t}R^2_{Y\sim U|T}}\Sigma^{-1/2}_{u|t}(\mu_{u|t_1}-\mu_{u|t_2})}{\|\Sigma^{-1/2}_{u|t}(\mu_{u|t_1}-\mu_{u|t_2})\|_2}$, i.e., γ is colinear with the $\Sigma^{-1}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2})$ inferred by the relationship defined in equation (57).

A.5 Proof of Proposition 1

Proposition 1 *Suppose that the observed data is generated by model (7)-(9):*

$$\begin{aligned} U &= \epsilon_u, \\ T &= BU + \epsilon_t, \\ Y &= \tau'T + \gamma'U + \epsilon_y. \end{aligned}$$

If B is rank m and there remain two disjoint matrices of rank m after deleting any row of B, then $\|\Sigma^{-1/2}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2})\|^2_2$ and $\frac{d^2_1}{(d^2_1+\sigma^2_t)}$ are both identified.

Proof. With model (7)-(9), the conditional distribution of confounder U

$$f_B(u \mid t) \sim N(\mu_{u|t},\ \Sigma_{u|t}), \tag{60}$$

where $\mu_{u|t} := B'(BB' + \Lambda_t)^{-1}t$, $\Sigma_{u|t} := I - B'(BB' + \Lambda_t)^{-1}B$.

From Anderson and Rubin (1956), we know that, if B is rank m and there remain two disjoint matrices of rank m after deleting any row of B, then we B is identifiable up to rotations from the right. Let $\tilde{B} = BA$ for an arbitrary positive matrix A such that the observed treatments are consistent with $Cov(T) = BB' + \Lambda_t = \tilde{B}\tilde{B}' + \Lambda_t$, implying that $AA' = I$ and A is an orthogonal matrix. With $\tilde{B}$, we have the conditional distribution of confounder U as

$$f_{\tilde{B}}(u \mid t) \sim N(\tilde{\mu}_{u|t},\ \tilde{\Sigma}_{u|t}), \tag{61}$$

where

$$\tilde{\mu}_{u|t} := \tilde{B}'(\tilde{B}\tilde{B}' + \Lambda_t)^{-1}t = A'\mu_{u|t}, \tag{62}$$

$$\tilde{\Sigma}_{u|t} := I - \tilde{B}'(\tilde{B}\tilde{B}' + \Lambda_t)^{-1}\tilde{B} = A'\Sigma_{u|t}A. \tag{63}$$

With equation (62) and (63), we have

$$\begin{aligned}
\|\tilde{\Sigma}_{u|t}^{-1/2}(\tilde{\mu}_{u|t_1} - \tilde{\mu}_{u|t_2})\|_2^2 &= (\tilde{\mu}_{u|t_1} - \tilde{\mu}_{u|t_2})'\tilde{\Sigma}_{u|t}^{-1/2}\tilde{\Sigma}_{u|t}^{-1/2}(\tilde{\mu}_{u|t_1} - \tilde{\mu}_{u|t_2}) && (64)\\
&= (A'\mu_{u|t_1} - A'\mu_{u|t_2})'\tilde{\Sigma}_{u|t}^{-1}(A'\mu_{u|t_1} - A'\mu_{u|t_2}) && (65)\\
&= (\mu_{u|t_1} - \mu_{u|t_2})A(A'\Sigma_{u|t}A)^{-1}A'(\mu_{u|t_1} - \mu_{u|t_2}) && (66)\\
&= (\mu_{u|t_1} - \mu_{u|t_2})\Sigma_{u|t}^{-1}(\mu_{u|t_1} - \mu_{u|t_2})' && (67)\\
&= \|\Sigma_{u|t}^{-1/2}(\mu_{u|t_1} - \mu_{u|t_2})\|_2^2, && (68)
\end{aligned}$$

Therefore, $\|\Sigma_{u|t}^{-1/2}(\mu_{u|t_1} - \mu_{u|t_2})\|_2^2$ is identified.

For $\frac{d_1^2}{(d_1^2+\sigma_t^2)}$, it depends on the largest singular value of B, which is rotation-invariant and therefore identified.

A.6 Proof of Corollary 1.1

Corollary 1.1 *Assume $\Lambda_t = \sigma_t^2 I$ and let d_1 be the largest singular value of B. For all t_1, t_2 with $\|(t_1 - t_2)\|_2 = 1$, the squared bias is bounded by*

$$Bias_{t_1,t_2}^2 \leq \frac{d_1^2}{(d_1^2 + \sigma_t^2)} \frac{\sigma_{y|t}^2}{\sigma_t^2} R_{Y\sim U|T}^2, \tag{69}$$

with equality when $(t_1 - t_2) = u_1^B$, the first left singular vector of B. When $(t_1 - t_2) \in Null(B')$, the naive estimate is unbiased, that is, $PATE_{t_1,t_2} = \tau'_{naive}(t_1 - t_2)$.

Proof. Suppose that the matrix B has the singular value decomposition,

$$B = UDV',$$

where the diagonal entries of D are the singular values of B in descending order. Then, we can write

$$(\mu_{u|t_1} - \mu_{u|t_2}) = VD(D^2 + \sigma_t^2 I)^{-1}U'(t_1 - t_2), \tag{70}$$

and

$$\Sigma_{u|t}^{-1} = V[I + \frac{1}{\sigma_t^2} D^2]V'. \tag{71}$$

By plugging Equation (70) and (71) into the result of theorem 1, we have

$$\text{Bias}_{t_1,t_2}^2 \leq \frac{\sigma_{y|t}^2}{\sigma_t^2} R_{Y\sim U|T}^2 \parallel VD(\sigma_t^2 I + D^2)^{-1/2} U'(t_1 - t_2) \parallel_2^2, \tag{72}$$

where, according to Rayleigh quotient (Horn, 1985), the squared L2 norm reaches its maximum, $\frac{d_1^2}{(d_1^2+\sigma_t^2)}$, when $(t_1 - t_2)$ equals the first column of U, i.e., the first left singular vector of B.

Therefore, we have

$$\text{Bias}_{t_1,t_2}^2 \leq \frac{d_1^2}{(d_1^2 + \sigma_t^2)} \frac{\sigma_{y|t}^2}{\sigma_t^2} R_{Y\sim U|T}^2. \tag{73}$$

A.7 Proof of Theorem 2

Theorem 2 *Assume model (31) - (33):*

$$\begin{aligned} f(u \mid t) &\sim N(\mu_{u|t}, \Sigma_{u|t}) \\ \tilde{Y} &= \gamma_t' U + \epsilon_{\tilde{Y}} \\ Y &= F_{Y|T}^{-1}(\Phi(\tilde{Y} - \gamma_t' \mu_{u|t})) \end{aligned}$$

and that $\sigma_{y|t}$, $\Sigma_{u|t}$, and γ_t can vary with t and assume $\mu_{u|t_1}$ and $\mu_{u|t_2}$ are in the row space of $\Sigma_{u|t_1}$ and $\Sigma_{u|t_2}$ respectively, and Y is continuous. Then the omitted variable bias for all quantile treatment effects are bounded. The median treatment effect, $MTE_{t_1,t_2} = med\big(Y \mid do(t_1)\big) - med\big(Y \mid do(t_2)\big)$ is in the interval $m_l \leq MTE_{t_1,t_2} \leq m_u$ where

$$m_l = F_{Y|T=t_1}^{-1}(\Phi(- \parallel (\Sigma_{u|t_1}^{\dagger})^{1/2} \mu_{u|t_1} \parallel_2)) - F_{Y|T=t_2}^{-1}(\Phi(\parallel (\Sigma_{u|t_2}^{\dagger})^{1/2} \mu_{u|t_2} \parallel_2)) \tag{74}$$

$$m_u = F_{Y|T=t_1}^{-1}(\Phi(\parallel (\Sigma_{u|t_1}^{\dagger})^{1/2} \mu_{u|t_1} \parallel_2)) - F_{Y|T=t_2}^{-1}(\Phi(- \parallel (\Sigma_{u|t_2}^{\dagger})^{1/2} \mu_{u|t_2} \parallel_2)) \tag{75}$$

where m_l and m_u are identifiable under Assumptions 4 and 5.

Proof. In equation (33),

$$Y = F_{Y|T}^{-1}(\Phi(\tilde{Y} - \gamma_t' \mu_{u|t})),$$

where $F_{Y|T}^{-1}$ and Φ are both non-decreasing monotone functions, so the composition of these functions is also monotone non-decreasing. The quantile of a monotone function of a random variable is same as the monotone function of the quantile of the random variable. As such, we can first consider the quantiles in the space of $\tilde{Y}$, i.e., $(\tilde{Y} - \gamma_t' \mu_{u|t}) = \gamma_t'(U - \mu_{u|t}) + \epsilon_{\tilde{Y}}$, then the observed and intervention distribution respectively have densities:

$$f\big((\tilde{Y} - \gamma_t' \mu_{u|t}) \mid T = t\big) \sim N(0,\ 1) \tag{76}$$

$$f\big((\tilde{Y} - \gamma_t' \mu_{u|t}) \mid do(T = t)\big) \sim N(-\gamma_t' \mu_{u|t},\ 1 + \gamma_t'(\Sigma_u - \Sigma_{u|t})\gamma_t) \tag{77}$$

To show that the omitted variable bias for all quantile treatment effects are bounded, it is suffice to show that $\text{E}\big((\tilde{Y} - \gamma_t' \mu_{u|t}) \mid do(T = t)\big)$ is bounded when $\mu_{u|t}$ is in the row space of $\Sigma_{u|t}$.

Let $Z := \Sigma_{u|t}^{1/2}\gamma_t$, with $Z'Z \leq 1$ as a consequence of the constraint on γ_t. Then, we have

$$\| \mathrm{E}\big((\tilde{Y} - \gamma_t'\mu_{u|t}) \mid do(T=t)\big) \|_2^2 = \| \gamma_t'\mu_{u|t} \|_2^2 \tag{78}$$

$$= Z'(\Sigma_{u|t}^{\dagger})^{1/2}\mu_{u|t}\mu_{u|t}'(\Sigma_{u|t}^{\dagger})^{1/2}Z \tag{79}$$

$$\leq \| (\Sigma_{u|t}^{\dagger})^{1/2}\mu_{u|t} \|_2^2 \,. \tag{80}$$

where the bounds are reached when Z is colinear with $(\Sigma_{u|t}^{\dagger})^{1/2}\mu_{u|t}$, i.e., γ is colinear with the $\Sigma_{u|t}^{\dagger}\mu_{u|t}$.

Suppose that $\Sigma_{u|t}$ has the eigendecomposition,

$$\Sigma_{u|t} = Q\Lambda Q', \tag{81}$$

where Q is the square $s \times s$ matrix whose jth column is the eigenvector q_j of $\Sigma_{u|t}$, and Λ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues, $\Lambda_{jj} = \lambda_j$, in descending order. If $\Sigma_{u|t}$ is non-invertible and has rank p ($p \leq s$), we have $\lambda_j = 0$ for $j = p+1, \cdots, s$. when $\mu_{u|t}$ is in the row space of $\Sigma_{u|t}$, it can be expressed as a linear combination of q_j, $\sum_{j=1}^{p} a_j q_j$, $a_j \in \mathbb{R}$, then we have $\mathrm{E}\big((\tilde{Y} - \gamma_t'\mu_{u|t}) \mid do(T=t)\big)$ be bounded as

$$\| \mathrm{E}\big((\tilde{Y} - \gamma_t'\mu_{u|t}) \mid do(T=t)\big) \|_2^2 \leq \| (\Sigma_{u|t}^{\dagger})^{1/2}\mu_{u|t} \|_2^2 \tag{82}$$

$$= \| Q(\Lambda^{\dagger})^{1/2}Q' \sum_{j=1}^{p} a_j q_j \|_2^2, \tag{83}$$

$$= \sum_{i=1}^{s} (\sum_{j=1}^{p} a_j \lambda_j^{-\frac{1}{2}} Q_{ij})^2. \tag{84}$$

Since $(\tilde{Y} - \gamma_t'\mu_{u|t}) \mid do(T=t)$ follows a Gaussian distribution, thus we have $\mathrm{med}\big((\tilde{Y} - \gamma_t'\mu_{u|t}) \mid do(T=t)\big) = \mathrm{E}\big((\tilde{Y} - \gamma_t'\mu_{u|t}) \mid do(T=t)\big) = -\gamma_t'\mu_{u|t}$.

Plugging t_1 and t_2 into inequality (80), we have

$$-\| (\Sigma_{u|t_1}^{\dagger})^{1/2}\mu_{u|t_1} \|_2 \leq \mathrm{med}\big(\tilde{Y} - \gamma_t'\mu_{u|t_1} \mid do(T=t_1)\big) \leq \| (\Sigma_{u|t_1}^{\dagger})^{1/2}\mu_{u|t_1} \|_2, \tag{85}$$

$$-\| (\Sigma_{u|t_2}^{\dagger})^{1/2}\mu_{u|t_2} \|_2 \leq \mathrm{med}\big(\tilde{Y} - \gamma_t'\mu_{u|t_2} \mid do(T=t_2)\big) \leq \| (\Sigma_{u|t_2}^{\dagger})^{1/2}\mu_{u|t_2} \|_2 \,. \tag{86}$$

Since both $F_{Y|T}^{-1}$ and Φ are monotonously non-decreasing functions, MTE_{t_1,t_2} would reach its smallest when $\mathrm{med}\big(\tilde{Y} - \gamma_t'\mu_{u|t_1} \mid do(T=t_1)\big)$ being its smallest and $\mathrm{med}\big(\tilde{Y} - \gamma_t'\mu_{u|t_2} \mid do(T=t_2)\big)$ being its largest, i.e.,

$$MTE_{t_1,t_2} \geq m_l := F_{Y|T=t_1}^{-1}(\Phi(-\| (\Sigma_{u|t_1}^{\dagger})^{1/2}\mu_{u|t_1} \|_2)) - F_{Y|T=t_2}^{-1}(\Phi(\| (\Sigma_{u|t_2}^{\dagger})^{1/2}\mu_{u|t_2} \|_2)), \tag{87}$$

Conversely, MTE_{t_1,t_2} would reach its largest when $\mathrm{med}\big(\tilde{Y} - \gamma_t'\mu_{u|t_1} \mid do(T=t_1)\big)$ being its largest and $\mathrm{med}\big(\tilde{Y} - \gamma_t'\mu_{u|t_2} \mid do(T=t_2)\big)$ being its smallest, i.e.,

$$MTE_{t_1,t_2} \leq m_u := F_{Y|T=t_1}^{-1}(\Phi(\| (\Sigma_{u|t_1}^{\dagger})^{1/2}\mu_{u|t_1} \|_2)) - F_{Y|T=t_2}^{-1}(\Phi(-\| (\Sigma_{u|t_2}^{\dagger})^{1/2}\mu_{u|t_2} \|_2)). \tag{88}$$

A.8 Proof of Corollary 2.1

Corollary 2.1 *Assume the model (31) - (33):*

$$
\begin{aligned}
f(u \mid t) &\sim N(\mu_{u|t}, \Sigma_{u|t}) \\
\tilde{Y} &= \gamma_t' U + \epsilon_{\tilde{Y}} \\
Y &= F^{-1}_{Y|T}(\Phi(\tilde{Y} - \gamma_t' \mu_{u|t}))
\end{aligned}
$$

where Y is conditionally Gaussian given treatments, and where $\sigma_{y|t}$, $\Sigma_{u|t}$, and γ_t can vary with t. If $\Sigma_{u|t_1}$ and $\Sigma_{u|t_2}$ are non-invertible, then $Bias_{t_1,t_2}$ is bounded if and only if $\mu_{u|t_1}$ and $\mu_{u|t_2}$ are in the row space of $\Sigma_{u|t_1}$ and $\Sigma_{u|t_2}$ respectively. When bounded,

$$
Bias^2_{t_1,t_2} \leq \left(\sigma_{y|t_1} \sqrt{R^2_{Y\sim U|t_1}} \|(\Sigma^{\dagger}_{u|t_1})^{1/2} \mu_{u|t_1}\|_2 + \sigma_{y|t_2} \sqrt{R^2_{Y\sim U|t_2}} \|(\Sigma^{\dagger}_{u|t_2})^{1/2} \mu_{u|t_2}\|_2 \right)^2, \tag{89}
$$

with equality when $\gamma_{t_1} \propto \Sigma^{\dagger}_{u|t_1}\mu_{u|t_1}$ and $\gamma_{t_2} \propto \Sigma^{\dagger}_{u|t_2}\mu_{u|t_1}$ and where $\Sigma^{\dagger}_{u|t_1}$ and $\Sigma^{\dagger}_{u|t_2}$ are the pseudo-inverses of $\Sigma_{u|t_1}$ and $\Sigma_{u|t_2}$. If $\Sigma_{u|t_1} = \Sigma_{u|t_2} = \Sigma_{u|t}$ and $\gamma_t = \gamma$ is invariant to t (i.e. there are no treatment-confounder interactions), and $\sigma^2_{y|t_1} = \sigma^2_{y|t_2} = \sigma^2_{y|t}$ (homoskedastic outcome model) then $Bias_{t_1,t_2}$ is bounded if and only if $\mu_{u|t_1} - \mu_{u|t_2}$ is in the row space of $\Sigma_{u|t}$ and when bounded,

$$
Bias^2_{t_1,t_2} \leq \sigma^2_{y|t} R^2_{Y\sim U|T} \|(\Sigma^{\dagger}_{u|t})^{1/2}(\mu_{u|t_1} - \mu_{u|t_2})\|_2^2. \tag{90}
$$

Proof. Under model (31) - (33) with Gaussian outcomes, and $\sigma_{y|t}$, $\Sigma_{u|t}$, γ_t varying with t, we have

$$
\text{PATE}_{t_1,t_2} = (\mu_{y|t_1} - \mu_{y|t_2}) - (\sigma_{y|t_1}\gamma_{t_1}'\mu_{u|t_1} - \sigma_{y|t_2}\gamma_{t_2}'\mu_{u|t_1}), \tag{91}
$$

and

$$
\text{Bias}_{t_1,t_2} = \sigma_{y|t_1}\gamma_{t_1}'\mu_{u|t_1} - \sigma_{y|t_2}\gamma_{t_2}'\mu_{u|t_1} \tag{92}
$$

with sensitivity parameter γ_{t_1} and γ_{t_2} satisfying constraints $\gamma_{t_1}'\Sigma_{u|t_1}\gamma_{t_1} \leq R^2_{Y\sim U|t_1}$ and $\gamma_{t_2}'\Sigma_{u|t_2}\gamma_{t_2} \leq R^2_{Y\sim U|t_2}$ respectively, since $\gamma_{t_1}'\Sigma_{u|t_1}\gamma_{t_1}$ and $\gamma_{t_2}'\Sigma_{u|t_2}\gamma_{t_2}$ correspond to the confounding variations and should not be larger than a given level of $R^2_{Y\sim U|T}$.

may need to change $R^2_{\tilde{Y}}$ accordingly with the main text

Let

$$
Z_1 := \Sigma^{1/2}_{u|t_1}\gamma_{t_1}, \tag{93}
$$

$$
Z_2 := \Sigma^{1/2}_{u|t_2}\gamma_{t_2}, \tag{94}
$$

then the ommitted variable bias can be written as

$$
\text{Bias}_{t_1,t_2} = \sigma_{y|t_1} Z_1'(\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1} - \sigma_{y|t_2} Z_2'(\Sigma^{\dagger}_{u|t_2})^{1/2}\mu_{u|t_2} \tag{95}
$$

with $Z_1'Z_1 \leq R^2_{Y\sim U|t_1}$ and $Z_2'Z_2 \leq R^2_{Y\sim U|t_2}$. Then,

$$
\begin{aligned}
|\ \text{Bias}_{t_1,t_2}\ | &\leq \sigma_{y|t_1} \mid Z_1'(\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1} \mid + \sigma_{y|t_2} \mid Z_2'(\Sigma^{\dagger}_{u|t_2})^{1/2}\mu_{u|t_2} \mid & (96) \\
&\leq \sigma_{y|t_1}\sqrt{R^2_{Y\sim U|t_1}} \parallel (\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1} \parallel_2 + \sigma_{y|t_2}\sqrt{R^2_{Y\sim U|t_2}} \parallel (\Sigma^{\dagger}_{u|t_2})^{1/2}\mu_{u|t_2} \parallel_2 & (97)
\end{aligned}
$$

where the bounds are reached when Z_1 and Z_2 are respectively colinear with $(\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1}$ and $(\Sigma^{\dagger}_{u|t_2})^{1/2}\mu_{u|t_2}$, i.e., γ_{t_1} and γ_{t_2} are respectively colinear with the $\Sigma^{\dagger}_{u|t_1}\mu_{u|t_1}$ and $\Sigma^{\dagger}_{u|t_2}\mu_{u|t_2}$.

To show that Bias_{t_1,t_2} is bounded if and only if $\mu_{u|t_1}$ and $\mu_{u|t_2}$ are in the row space of $\Sigma_{u|t_1}$ and $\Sigma_{u|t_2}$ respectively, we can first show that the first term in Equation (97), $\sigma_{y|t_1}\sqrt{R^2_{Y\sim U|t_1}} \parallel (\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1} \parallel_2$, is bounded if and only if $\mu_{u|t_1}$ is in the row space of $\Sigma_{u|t_1}$. Suppose that $\Sigma_{u|t_1}$ has the eigendecomposition,

$$\Sigma_{u|t_1} = Q\Lambda Q', \tag{98}$$

where Q is the square $s \times s$ matrix whose jth column is the eigenvector q_j of $\Sigma_{u|t_1}$, and Λ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues, $\Lambda_{jj} = \lambda_j$, in descending order. If $\Sigma_{u|t_1}$ is non-invertible and has rank p ($p \leq s$), we have $\lambda_j = 0$ for $j = p+1, \cdots, s$.

On the one hand, when $\mu_{u|t_1}$ is in the row space of $\Sigma_{u|t_1}$, it can be expressed as a linear combination of q_j, $\sum_{j=1}^{p} a_j q_j$, $a_j \in \mathbb{R}$. Then, we have the squared first term in Equation (97) as

$$\sigma^2_{y|t_1} R^2_{Y\sim U|t_1} \parallel (\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1} \parallel_2^2, \tag{99}$$

$$= \sigma^2_{y|t_1} R^2_{Y\sim U|t_1} \parallel Q(\Lambda^{\dagger})^{1/2}Q' \sum_{j=1}^{p} a_j q_j \parallel_2^2, \tag{100}$$

$$= \sigma^2_{y|t_1} R^2_{Y\sim U|t_1} \sum_{i=1}^{s} (\sum_{j=1}^{p} a_j \lambda_j^{-\frac{1}{2}} Q_{ij})^2, \tag{101}$$

where Q_{ij} denotes the element at the ith row and jth column of matrix Q, and $\Lambda^{\dagger}$ is the pseudo-inverse of Λ by taking the reciprocal of each its non-zero element on the diagonal, leaving the zeros in place.

On the other hand, when $\sigma_{y|t_1}\sqrt{R^2_{Y\sim U|t_1}} \parallel (\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1} \parallel_2$ is bounded, let's assume that $\mu_{u|t_1}$ is not in the row space of $\Sigma_{u|t_1}$, say $\mu_{u|t_1} = q_s$. Since $\lambda_s = 0$, $\lambda_s^{-1/2} = \infty$, resulting in $\parallel (\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1} \parallel_2$ equaling ∞, which contradicts the condition that $\sigma_{y|t_1}\sqrt{R^2_{Y\sim U|t_1}} \parallel (\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1} \parallel_2$ is bounded.

Similarly, we can show that the second term in Equation (97), $\sigma_{y|t_2}\sqrt{R^2_{Y\sim U|t_2}} \parallel (\Sigma^{\dagger}_{u|t_2})^{1/2} \quad \mu_{u|t_2} \parallel_2$, is bounded if and only if $\mu_{u|t_2}$ is in the row space of $\Sigma_{u|t_2}$ by expressing $\Sigma_{u|t_2}$ in its eigendecomposition. Altogether, we can see that Bias_{t_1,t_2} is bounded if and only if $\mu_{u|t_1}$ and $\mu_{u|t_2}$ are in the row space of $\Sigma_{u|t_1}$ and $\Sigma_{u|t_2}$ respectively.

Lastly, let's consider the case when $\Sigma_{u|t_1} = \Sigma_{u|t_2}$ and $\gamma_t = \gamma$, the constraints on the sensitivity parameters can be unified as

$$\gamma'\Sigma_{u|t}\gamma \leq R^2_{Y\sim U|T} \tag{102}$$

where $R^2_{Y\sim U|T}$ denotes the fraction of confounding variation in residual variance of $\tilde{Y}$ conditional on T [3].

Let

$$Z := \Sigma^{1/2}_{u|t}\gamma, \tag{103}$$

then the omitted variable bias can be written as

$$\text{Bias}_{t_1,t_2} = \sigma_{y|t} Z'(\Sigma^{\dagger}_{u|t})^{1/2}(\mu_{u|t_1} - \mu_{u|t_2}), \tag{104}$$

where $Z'Z \leq R^2_{Y\sim U|T}$, implied by inequality (102).

[3] $R^2_{\tilde{Y}\sim U|T}$ coincides with $R^2_{\tilde{Y}\sim U}$ here, but we use notation $R^2_{\tilde{Y}\sim U|T}$ for consistency.

Therefore,

$$\text{Bias}^2_{t_1,t_2} = \sigma^2_{y|t} Z'(\Sigma^{\dagger}_{u|t})^{1/2}(\mu_{u|t_1} - \mu_{u|t_2})(\mu_{u|t_1} - \mu_{u|t_2})'(\Sigma^{\dagger}_{u|t})^{1/2} Z \tag{105}$$

$$\leq \sigma^2_{y|t} R^2_{Y\sim U|T} \parallel (\Sigma^{\dagger}_{u|t})^{1/2}(\mu_{u|t_1} - \mu_{u|t_2}) \parallel_2^2, \tag{106}$$

where the bounds are reached when Z is colinear with $(\Sigma^{\dagger}_{u|t})^{1/2}(\mu_{u|t_1} - \mu_{u|t_2})$, i.e., γ is colinear with the $\Sigma^{\dagger}_{u|t}(\mu_{u|t_1} - \mu_{u|t_2})$.

Similarly as above of showing $\sigma_{y|t_1}\sqrt{R^2_{Y\sim U|t_1}} \parallel (\Sigma^{\dagger}_{u|t_1})^{1/2}\mu_{u|t_1} \parallel_2$ is bounded if and only if $\mu_{u|t_1}$ is in the row space of $\Sigma_{u|t_1}$, we can show that Bias_{t_1,t_2} in Equation (105) is bounded if and only if $\mu_{u|t_1} - \mu_{u|t_2}$ is in the row space of $\Sigma_{u|t}$. We leave out the details here to avoid redundancy.

B Causal Equivalence

B.1 Multiple Treatments and Causal Equivalence Classes

In this Section, we formally clarify why we latent variable identification is only required up to invertible linear transformations (Assumption 4. We say that ψ_T is identified up to a *causal equivalence class* when, for any value of ψ_T that is compatible with the observed data distribution, the set of possible causal effects, as indexed by ψ_Y, is invariant to the particular value of ψ_T in the equivalence class.

Definition 1 (Causal equivalence class)**.** *$[\psi_T]$ is a causal equivalence class of ψ_T if and only if for any $\tilde{\psi}_T$ in $[\psi_T]$, then, for every ψ_Y there exists a $\tilde{\psi}_Y$ such that $f_{\psi_Y,\psi_T}(y \mid do(T = t)) = f_{\tilde{\psi}_Y,\tilde{\psi}_T}(y \mid do(T = t))$ for all y, t.*

For the purposes of sensitivity analysis, when ψ_T is identified up to a causal equivalence class, we can assume that ψ_T is point-identified at a particular value within the class $[\psi_T]$ without loss of generality. Crucially, the copula-based formulation enables valid sensitivity analysis without observable implications, even in these cases where ψ_T is restricted by the observed data. In this case, the outcome-confounder copula c_{ψ_Y} remains the lone degree of freedom in the sensitivity model. As we will show, this restriction can induce qualitatively different sensitivity regions compared too cases where ψ_T is unrestricted. For example, sensitivity regions can be bounded, even without additional restrictions on ψ_Y.

We consider causal equivalence under the Gaussian copula model, where as a reminder, under Assumptions 4 and 5, we have

$$f(u \mid t) \sim N(\mu_{u|t}, \Sigma_{u|t}) \tag{107}$$

where $\mu_{u|t}$ and $\Sigma_{u|t}$ can be identified (up to rotation) from a latent variable model on the treatments. We can further rewrite equations (27) - (28) as

$$\tilde{Y} = \gamma_t' U + \epsilon_{\tilde{Y}} \tag{108}$$

$$Y = F^{-1}_{Y|T}(\Phi(\tilde{Y} - \gamma_t'\mu_{u|t})) \tag{109}$$

so that $\psi_Y = \{\gamma_t\}$ and $\epsilon_{\tilde{Y}} \sim N(0, 1 - \gamma_t'\Sigma_{u|t}\gamma_t)$ is chosen so that without loss of generality $\text{Var}(\tilde{Y} \mid T) = 1$. The Gaussian copula under this model is fully determined by the covariance matrix

$$\text{Cov}([\tilde{Y}, U] \mid T = t) = \begin{bmatrix} 1 & \gamma_t'\Sigma_{u|t} \\ \Sigma_{u|t}\gamma_t & \Sigma_{u|t} \end{bmatrix} \tag{110}$$

with parameters $\psi_T = \{\mu_{u|t}, \Sigma_{u|t} : t \in \mathcal{T}\}$ and $\psi_Y = \{\gamma_t : \mathcal{T}\}$ where $\mathcal{T}$ is the space of all treatments. Given Assumption 4, $\psi_Y = \{\gamma_t\}$ is the sole $m-$dimensional sensitivity vector governing the magnitude of

the omitted variable bias. The following theorem establishes that the class of ψ_T defined by all invertible linear transformations of U is a causal equivalance class.

Theorem 3. *Assume model (107) - (109). Let* $[\psi_T] = \{\tilde{\psi}_T = \{A\mu_{u|t}, A\Sigma_{u|t}A : t \in \mathcal{T}\} : A \in \mathcal{S}^+\}$ *where* $\mathcal{S}^+$ *is the space of symmetric positive definite matrices. Then* $[\psi_T]$ *is a causal equivalence class.*

Proof. The intervention distribution for $\tilde{Y}$ is defined as

$$f_\psi(\tilde{y} \mid do(t)) = \int \left[\int f_{\psi_Y}(\tilde{y} \mid t, u) f_{\psi_T}(u \mid \tilde{t}) du \right] f(\tilde{t}) d\tilde{t} \tag{111}$$

where $\psi_Y = \gamma$ and $\psi_T = \{\mu_{u|t}, \Sigma_{u|t}\}$. Then, $\int f_{\gamma_t}(\tilde{y} \mid t, u) f_{\psi_T}(u \mid \tilde{t}) du \sim N(\gamma_t' \mu_{u|\tilde{t}},\ 1)$ for any γ_t such that $\gamma_t' \Sigma_{u|t} \gamma_t \leq 1$ (see Equation (53)). Let $\tilde{\psi}_T = \{A\mu_{u|t}, A\Sigma_{u|t}A\} \in [\psi_T]$ where $A \in \mathcal{S}^+$ is a positive definite matrix and assume $\tilde{\psi}_Y = \tilde{\gamma}_t$. Then,

$$\int f_{\tilde{\gamma}_t}(\tilde{y} \mid t, u) f_{\tilde{\psi}_T}(u \mid \tilde{t}) du \sim N(\tilde{\gamma}_t' A \mu_{u|\tilde{t}},\ 1). \tag{112}$$

Let $\tilde{\gamma}_t = A^{-1}\gamma_t$ be a bijective mapping from γ_t to $\tilde{\gamma}_t$. For any γ_t and positive definite A, we have $\tilde{\gamma}_t' A \Sigma_{u|t} A \tilde{\gamma}_t = \gamma_t' \Sigma_{u|t} \gamma_t \leq 1$ so that $\tilde{\gamma}_t$ is a valid copula parameter. In addition, $\int f_{\gamma_t}(\tilde{y} \mid t, u) f_{\psi_T}(u \mid \tilde{t}) du = \int f_{\tilde{\gamma}_t}(\tilde{y} \mid t, u) f_{\tilde{\psi}_T}(u \mid \tilde{t}) du$, which implies $f_{\gamma_t, \psi_T}(\tilde{y} \mid do(t)) = f_{\tilde{\gamma}_t, \tilde{\psi}_T}(\tilde{y} \mid do(t))$. Since Y is a deterministic function of $\tilde{Y}$, this implies $f_{\gamma_t, \psi_T}(y \mid do(t)) = f_{\tilde{\gamma}_t, \tilde{\psi}_T}(y \mid do(t))$. Therefore, $[\psi_T]$ is a causal equivalence class.

The gist of the proof is that for any invertible linear transformation, A, of U, the copula parameterized by $\tilde{\gamma}_t = A^{-1}\gamma_t$ yields equivalent causal effects in the reparameterized coordinates of U as γ does in the original confounder coordinates. See Miao et al. (2020) for more general theory about causal equivalence in larger class of latent variable models.

C Modeling Choice Details

C.1 Identification and Inference in the Factor Model

Here, we briefly elaborate on identifiability of the probabilistic principal components model, which is a prerequisite for our multi-cause sensitivity analysis. Identifiability under various factor model assumptions is well studied and has a long history in the literature (Mardia et al., 1980; Everett, 2013). In the specific probabilistic principal components model 8, Tipping and Bishop (1999) provide a maximum likelihood solution for inferring the latent confounder parameters conditional on m. Many procedures are available for selecting the appropriate value of m, using for example Bayesian model selection techniques (Minka, 2001) or large p, small n asymptotics Gavish and Donoho (2014).

When the observed outcome distribution is non-Gaussian, we cannot necessarily express PATE_{t_1,t_2} analytically. In particular, for non-Gaussian Y, when $f(u \mid t_1) \sim f(u \mid t_2)$ the average treatment effect among the t_1- and t_2-treated units is unconfounded, but the bias of PATE_{t_1,t_2} may be nonzero since $f(u \mid t) \nsim f(u)$. The causal effects, however, can still be calculated using Algorithm 1.

In the following, we would like to further elucidates important situations in which we cannot bound the omitted variable bias due to non-identifiability of the factor model. Again, we focus on the rotated treatments $\tilde{T} \sim N(0, \Delta + \sigma_t^2 I)$ and highlight two simple situations in which we cannot bound the the causal effects. First, when B is rank k, i.e. there exist $m = k$ independent confounders, Δ has no non-zero entries on the diagonal and thus we cannot identify either Δ nor $\text{Cov}(\epsilon_t) = \sigma_t^2 I_k$, only their sum. Second, if $\text{Cov}(\epsilon_t)$ is an unknown arbitrary diagonal matrix (as opposed to a matrix proportional to the identity), then $\text{Cov}(\epsilon_t)$ is not

distinguishable from Δ. In both of these cases, the worst-case bias is unbounded since the non-confounding variation of the treatment assignment, $\text{Cov}(\epsilon_t)$, can be arbitrarily small. In such settings, we can still apply approaches used in single cause sensitivity analysis, by specifying both Ψ_Y and Ψ_T; when the factor model is not identifiable, Ψ_T must be chosen as a true parameter, e.g. by bounding the fraction of treatment variation due to confounding, $R^2_{T\sim U}$.

C.2 Confounder Inference with Variational Autoencoders

Probabilistic Principal Component Analysis should only be used when the treatments are approximately Gaussian treatments. For binary and other general treatment distributions, more sophisticated probabilistic latent variables models are required. Examples of such latent variable models include models for count data like the logistic factor analysis (Hao et al., 2015) and Poisson factor analysis methods (Gopalan et al., 2013). Unfortunately, these models imply posteriors which are non-Gaussian, violating Assumption 5.

As such, for general treatment distributions, our approach is to infer a conditional Gaussian latent variable model using a variational autoencoder (VAE). VAEs have been extremely popular in machine learning, in particular for generating low dimensional representations of complex inputs like images (Pu et al., 2016) but more recently have been used in scientific and decision-making applications (Lopez et al., 2020) and in applications to causal inference (Louizos et al., 2017). A VAE consists of a prior distribution, $f(u)$, typically for the low-dimensional latent variables, a stochastic encoder, and a stochastic decoder. In our application, the inferred stochastic decoder, $\hat{f}_\theta(t \mid u)$, is a non-linear map from latent confounders to a distribution over causes. Together, the prior distribution for u and the decoder imply a posterior confounder distribution, $\hat{f}(u \mid t)$.

In practice, inference for the true posterior is intractable and so a variational approximation, called the encoder, $q_\phi(u \mid t)$, is used in place of the true posterior. Typically the encoder is chosen to be a normal distribution with mean and variance which are non-linear functions of the input, $q_\phi = N(\mu_\phi(t), \sigma^2_\phi(t))$. A crucial question is that how well the Gaussian encoder approximates the true posterior; improving the variational approximation to the true latent variable posterior is an area of active research. In this work, we follow a common strategy of using the encoder learned by the VAE as the proposal distribution in an importance sampler (Lopez et al., 2020).

Specifically, we apply a variant of the Constant-Variance Variational Autoencoder (CV-VAE) (Ghosh et al., 2020) to infer the conditional confounder distribution, $f(u \mid t) \sim N(\mu_{u|t}, \Sigma_{u|t})$, in which $\Sigma_{u|t}$ does not depend on the level of t. We use the importance sampling to improve estimates of the conditional mean $\mu_{u|t}$, and posterior variance, $\Sigma_{u|t}$. While this approach only yields an approximation to the true posterior, we demonstrate the practical effectiveness of this approach in Sections 6 and in Appendix F.

C.3 Binary Outcomes

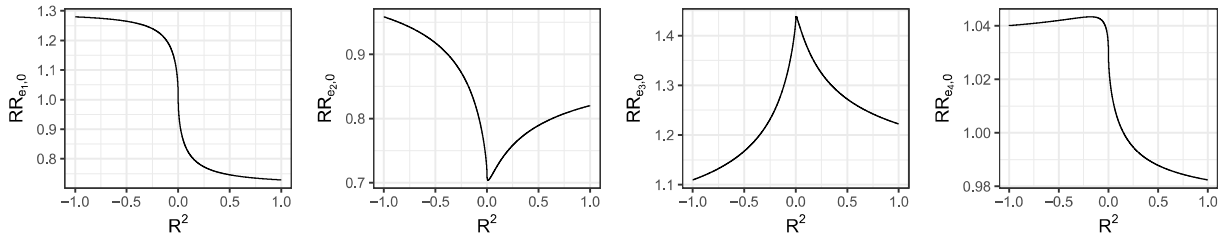
For binary outcomes with the risk ratio estimand:

$$RR_{t,\bullet} = \sum_{t_i \in \mathcal{T}} \Phi\left(\Phi^{-1}(\mu_{y|t}) + \gamma'(\mu_{u|t_i} - \mu_{u|t})\right) \Big/ Pr(Y = 1), \tag{113}$$

which implies that

$$RR_{t_1,t_2} = \sum_{t_i \in \mathcal{T}} \Phi\left(\Phi^{-1}(\mu_{y|t_1}) + \gamma'(\mu_{u|t_i} - \mu_{u|t_1})\right) \Big/ \sum_{t_i \in \mathcal{T}} \Phi\left(\Phi^{-1}(\mu_{y|t_2}) + \gamma'(\mu_{u|t_i} - \mu_{u|t_2})\right), \tag{114}$$

where $\gamma'\Sigma_{u|t}\gamma \leq \sigma^2_{\tilde{y}|t} R^2_{\tilde{Y}\sim U|T}$. We can numerically explore values of RR_{t_1,t_2} within the valid domain of γ, and calculate the corresponding implicit partial R-squared by $R^2_{\tilde{Y}\sim U|T} = \frac{\gamma'\Sigma_{u|t}\gamma}{\sigma^2_{\tilde{y}|t}}$. To calculate the robustness value, we only need to find the value of $R^2_{\tilde{Y}\sim U|T}$ for which the corresponding $RR_{t_1,t_2} = 1$. Noticeably, RR_{t_1,t_2} is not monotone in $R^2_{\tilde{Y}\sim U|T}$, since the variance of intervention distribution also depends on γ. This is evident in the simulation in Section 6 where we fit the observed outcome model by probit regression and the valid range for scalar γ is $[-\frac{1}{\sigma_{u|t}}, \frac{1}{\sigma_{u|t}}]$. We visualize the non-monotone relationship between RR_{t_1,t_2} and $R^2_{\tilde{Y}\sim U|T}$ in Figure 8.



(a) $RR_{e_1,0}$ vs R^2 (b) $RR_{e_2,0}$ vs R^2 (c) $RR_{e_3,0}$ vs R^2 (d) $RR_{e_4,0}$ vs R^2

Figure 8: RR_{t_1,t_2} is non-monotone in $R^2_{\tilde{Y}\sim U|T}$. Positive values of R^2 indicates that U is positively correlated with $\tilde{Y}$, and negative values of R^2 means that U is negatively correlated with $\tilde{Y}$.

D Additional Simulation Results

D.1 Robustness of the Gaussian Copula Assumption

Here, we explore how alternative copula specifications affect the results. We use the observed data generating process defined in Section 6.1.

We explore three classes of copulas: Clayton copulas, a quadratic dependence copula and a gaussian mixture copula. The Clayton copula is one of the well-studied Archimedian copulas (Hofert, 2008). For the quadratic depenndence, we consider the copula implied by the relationship $f(y \mid t) = F^{-1}_{Y|T}(a(u-b)^2)$. Finally, we consider a Gaussian mixture (G.M.) based copula, $f(y \mid t) = F^{-1}_{Y|T}(h_\epsilon(u))$ where $h_\epsilon(u) = a_1 u + b_1$ with probability $1/2$ and $h_\epsilon(u) = a_2 u + b_2$ with probability $1/2$. Examples of these copula densities are presented in Figure 10.

Most imporantly, for all of the alternative copulas we explored, the PATE was inside the bounds implied by the Gaussian copula (Figure 9). This suggest that the Gaussian bounds are quite robust to a range of different types of dependence.

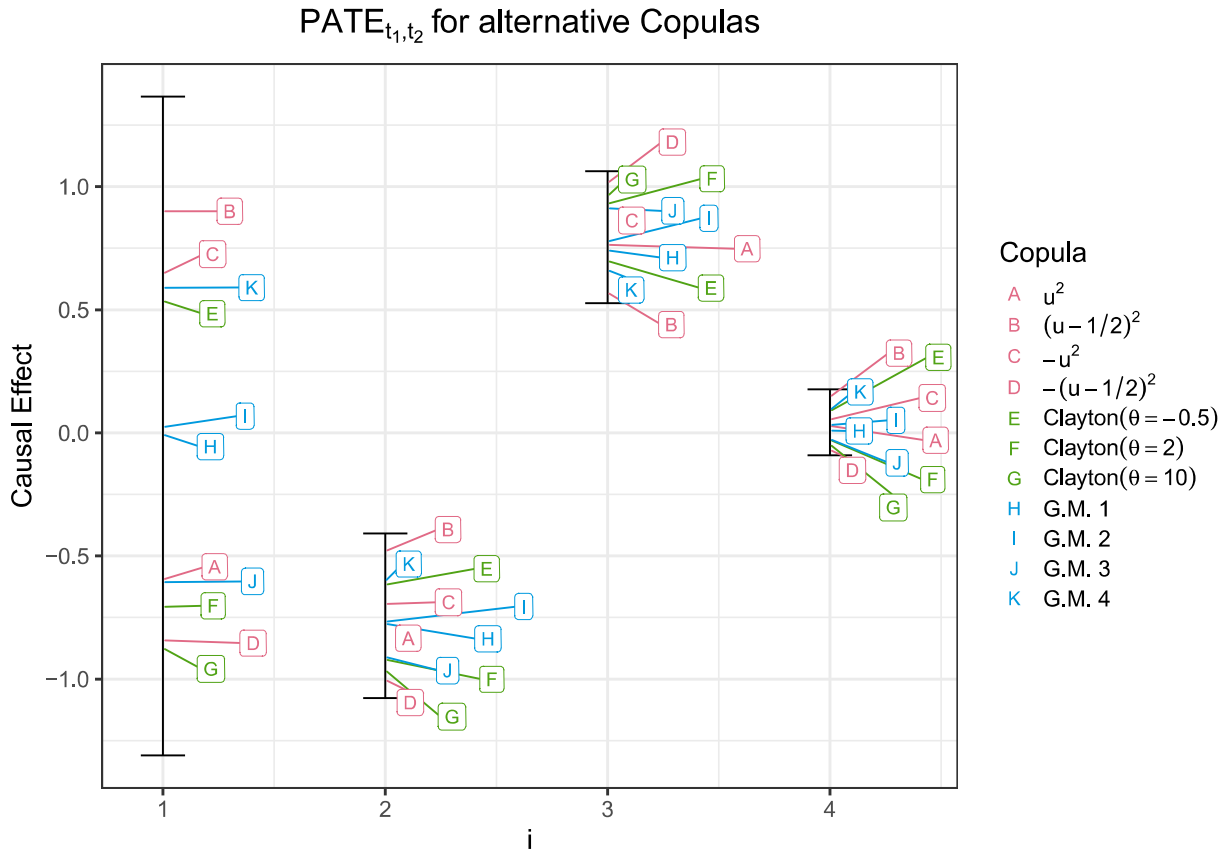


Figure 9: Robustness to alternative copulas. Black intervals are the ignorance regions for the Gaussian copula ($R^2_{Y\sim U|T} = 1$). We consider quadratic non-monotone copulas (pink), Clayton copulas (green) (Hofert, 2008), and four Gaussian mixture-based copulas. All alternative specifications lie within the bounds defined by the Gaussian coupla.

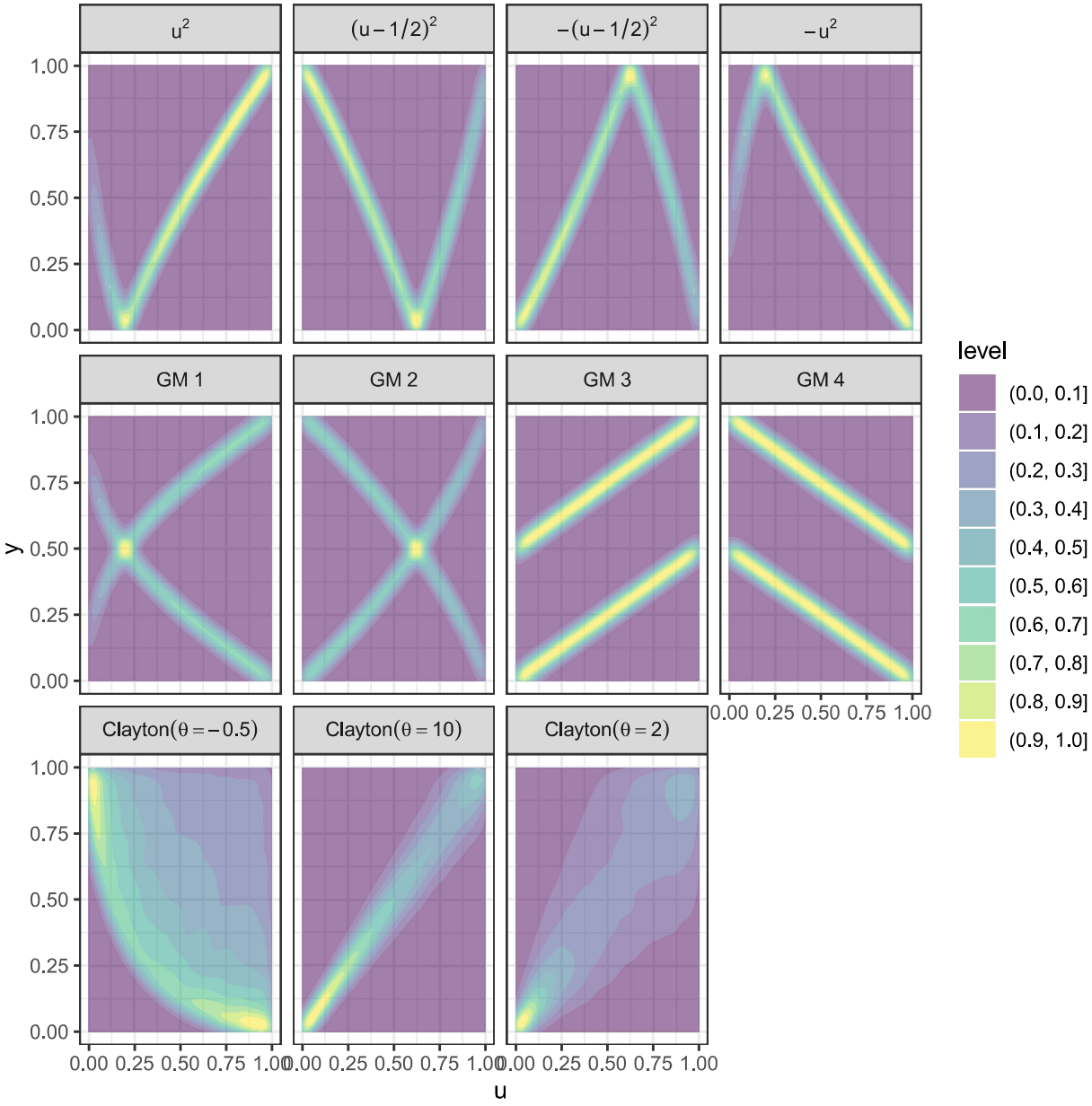


Figure 10: Conditional copula densities associated with effects in Figure 9 for $t = (0, 0, 0, 0)$

D.2 Additional Results from Simulation in Sparse Effects Setting

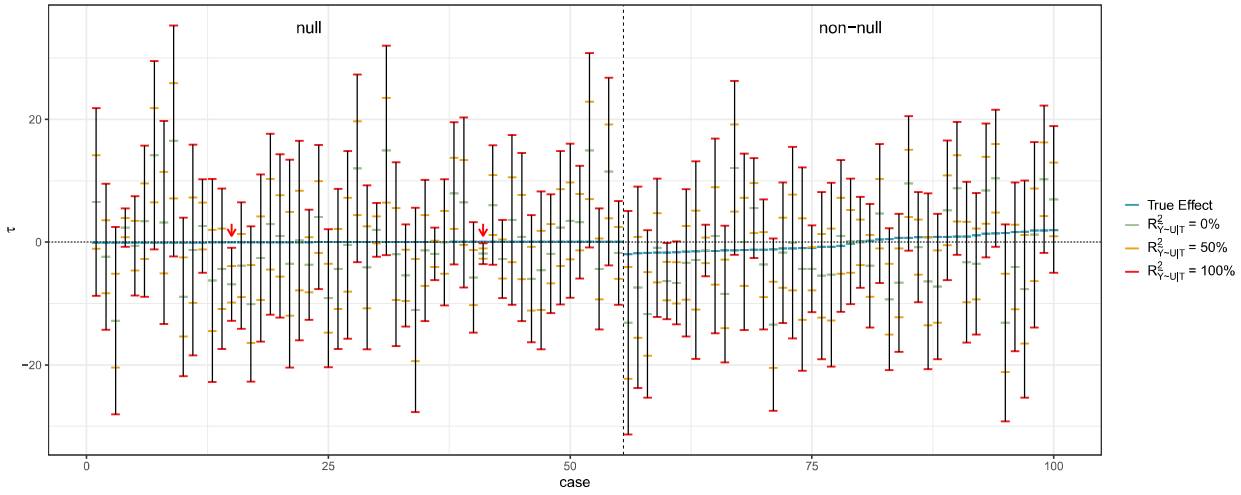
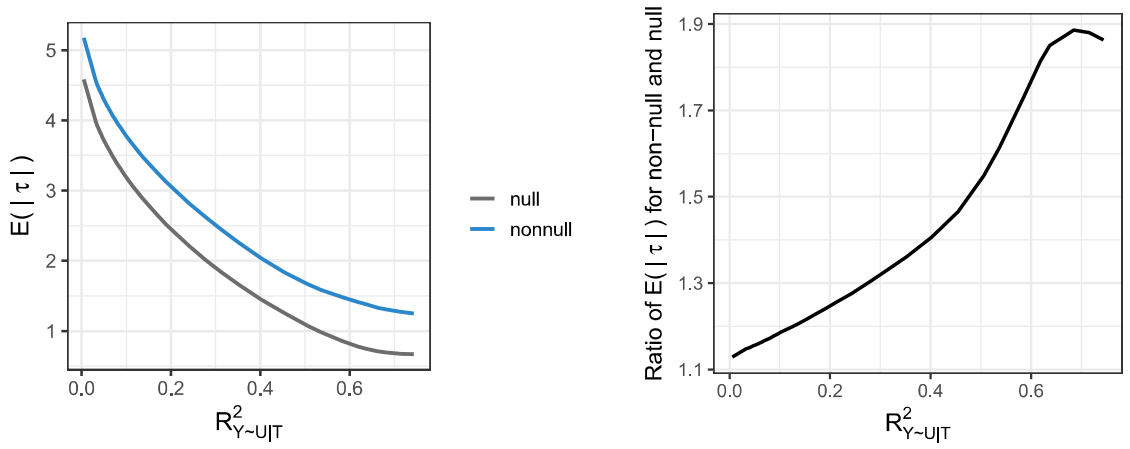


Figure 11: Worst-case ignorance regions for 55 randomly chosen null effects (left) and all 45 non-null effects (right) ordered by the magnitude of true effects in each group. Two red arrows indicate non-null treatments for which the worst-case ignorance region does not cover the true effect. This appears to be due to estimation error in the outcome model, more so than with the VAE.



(a) (b)

Figure 12: Change in $E(|\tau|)$ for the L1-minimized estimates as a function of $R^2_{Y\sim U|T}$, separated by null and non-null effects. (a) The magnitude of effects decreases with $\mathcal{R}^2$, with a larger relative decrease for null contrasts. (b) The relative magnitude of non-null and null effects increases with $\mathcal{R}^2$ in general. The magnitude of non-null effects can be as large as 1.9 times the null effects when $\mathcal{R}^2$ is large.

E Addition Results From the Mouse Obesity Analysis

Table 1: Point estimates and robustness values for the effect of gene expression on mouse obesity. Only genes which have 95% posterior credible intervals which exclude zero under the no unobserved confounding assumption are included in the table. The first two columns correspond to the posterior mean estimate of the regression coefficients under no unobserved confounding, $\tau_{\text{pm}}^{\text{naïve}}$, as well as the endpoint of the 95% posterior credible interval closest to zero for these coefficients. The third column is the robustness value based on the interval endpoint.

Gene	$\tau_{\text{pm}}^{\text{naïve}}$	$\tau_{\text{endpt}}^{\text{naïve}}$	$RV(\%)$
2010002N04Rik	9.35	1.98	29
Gstm2	6.54	2.02	80
Sirpa	6.17	0.30	1
Avpr1a	-5.30	-0.48	2
Igfbp2	-7.67	-3.95	17

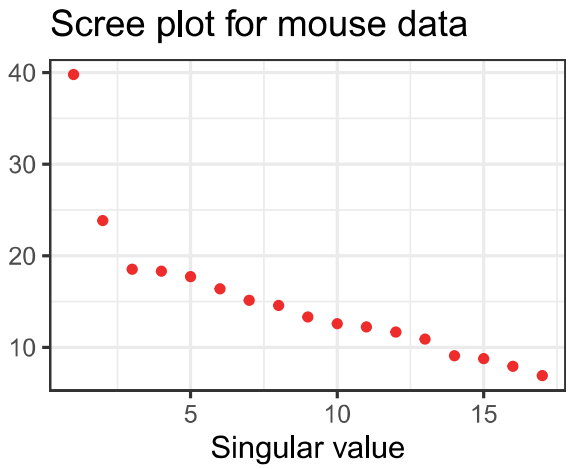


Figure 13: Scree plot for gene expression values in the mouse obesity dataset. Three singular values are singificnatly larger than the rest, so we use a use a 3 factor model to estimate conditional confounder distributions.

F A Reanalysis of the Actor Case Study

In this section, we compare our approach to other recent analyses of the TMDB 5000 Movie Dataset (Kaggle, 2017) which was analyzed extensively by Wang and Blei (2019) and Grimmer et al. (2020). The dataset consists of 5000 movies and their corresponding revenue, budget, genre and the identities of the lead cast members. Following Wang and Blei, we focus on estimating the causal effect of an actor's presence on the movie's log revenue. We let Y denote the log revenue and $T_i = (T_{i1}, \cdots, T_{ik})$ encode the movie cast, where the binary random variable $T_{ij} \in \{0, 1\}$ indicates whether actor j appeared in the movie i and $T_i \in \mathcal{T} = \{T_1, ..., T_n\}$. We also let $\mathcal{T}^j$ denote the set of all movies T_i for which $T_{ij} = 1$. We define the estimand of interest, η_j, as the the total log revenue contributed by actor j:

$$\eta_j := \sum_{t_i \in \mathcal{T}^j} \text{PATE}_{t_i, t_i^{-j}} \tag{115}$$

where t_i^{-j} corresponds to the observed treatment vector for movie i excluding actor j. This estimand is a non-parametric generalization of the regression coefficient τ_j, which was targeted in the analysis in Wang and Blei (2019). Specifically, under the assumption that log-revenue is linear in the cast indicators, η_j reduces to $n_j \tau_j$, the effect of actor j scaled by the number of movies they appeared in, where τ_j are the regression coefficients for actor j. Our estimand is well-defined without this linearity assumption.

We regress the log revenue on cast indicators to estimate actor effects, τ_j^{naive}, under an assumption of no unobserved confounding. In order to demonstrate the applicability of our sensitivity analysis, we explicitly induce unobserved confounding by excluding observed confounders. We validate our analysis, by comparing calibrated effect estimates when the confounders are excluded to estimates when the confounder is included. Most importantly we exclude the movie's budget which we estimate to be the largest known source of confounding (computed using Equation (41), see Appendix Figure 15a)[4].

For simplicity, we model the observed outcome distribution with a linear regression, although other more flexible outcome models (e.g. `BART`) can also be used. As in the previous Section, we use a VAE to infer a Gaussian conditional confounder distribution, $f(u \mid t) \sim N(\mu_{u|t}, \Sigma_{u|t})$ (See Appendix, Section C.2).

Results. Since our focus is on confounding not estimation, in order to limit the influence of estimation uncertainty we subset the data to the $k = 327$ actors who participated in at least twenty movies. This reduces the total number of movies to 2439. We fit the VAE to the treatments and use cross-validation to identify the appropriate latent confounder dimension, which we inferred to be $\hat{m} = 20$ (See Appendix Figure 15b). We then plot the worst-case ignorance region for the causal effect on log revenue as a function of $R^2_{Y \sim U|T}$ for the 46 actors with significant regression coefficients in the naive regression (Figure 14, top). Eight actors in the observed data regression have significantly negative coefficients, whereas 38 actors have significant positive coefficients. However, the worst-case ignorance regions for each actor are all very wide and include zero, which suggests that none of the effects are robust to confounding. In Table 2 of the Appendix we include robustness values for these actors. Leonardo DiCaprio has the largest robustness value at 36%, with the majority of the other actors well below 20%. For reference, the log budget, which was explicitly excluded from our causal analysis, explains about 30% of the variance in log revenue (Appendix Figure 15a). In other words, none of the causal effects are robust at a level which matches the variance explained by the most important excluded confounder.

The worst case ignorance regions depicted in top panel of Figure 14 correspond to a different choice of γ for each actor. We can also explore the robustness of causal effects under a single model by applying an appropriate MCC. Specifically, we search for a "worst case" candidate model by finding the sensitivity vector, γ_*, that implies the smallest L2 norm of the regression coefficients, τ. In this conservative model, the minimum L2 norm of the treatment coefficients is 4.4, down from 7.6 for the naive coefficients. In addition, 40

[4]For illustrative purposes, we can assume that the budget is pre-treatment, meaning that the budget is decided prior to selecting the cast, which may be a dubious assumption in actuality.

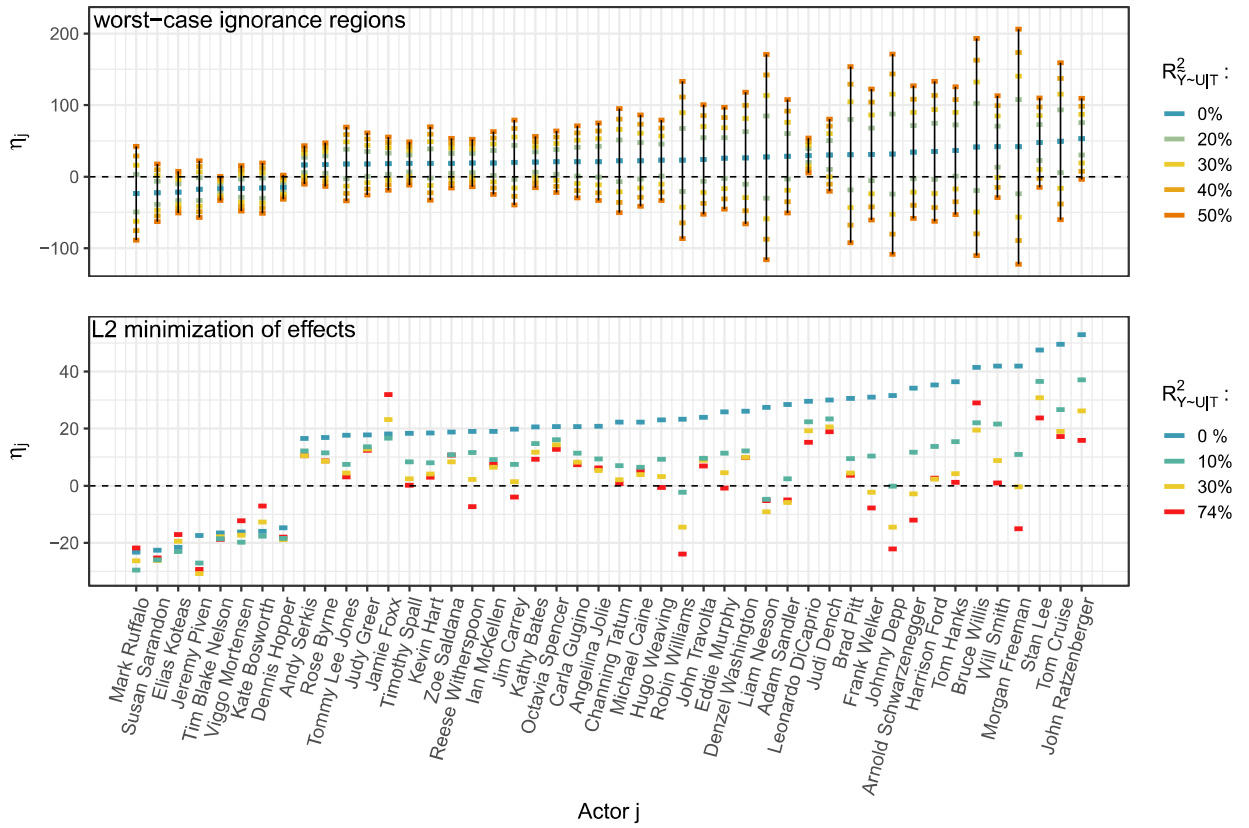
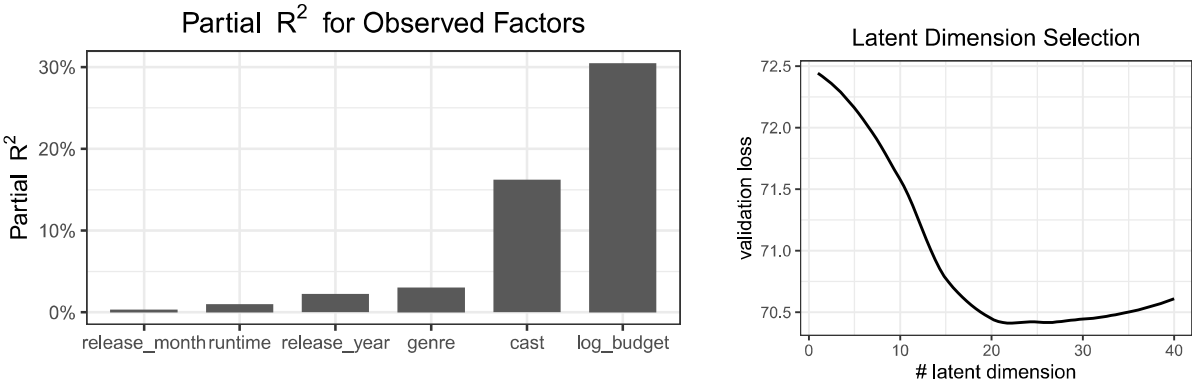


Figure 14: Estimated total log revenue contributed by a given actor. Top: worst-case ignorance region for each actor on a case by case basis. The blue points correspond to $R^2_{Y\sim U|T} = 0$, i.e. the naive estimates. Robustness values can be found in Appendix Table 2. Bottom: Treatment effects for candidate models chosen with the L2 minimzing multiple contrast criterion (MCC). The color correspond to $\mathcal{R}^2$, the limit on the fraction of outcome variance explained by confounding.

out of 46 actors have coefficients that are smaller in magnitude than the magnitudes of the naive coefficients (Figure 14, bottom). For this candidate model, it turns out that $\gamma'_* E[U|T = t]$ is significantly correlated, albeit weakly, with budget (Spearman's rank correlation = 0.2, p-value $< 2e$-16). Thus, the conservative model correctly attributes part of the outcome variation induced by the known excluded confounder to unobserved confounding.

Table 2: Robustness Value for Significant Actors

	Effect	$RV_{mean}(\%)$	$RV_{limit}(\%)$
John Ratzenberger	52.91	21.79	9.36
Tom Cruise	49.48	5.09	1.59
Stan Lee	47.53	14.43	4.16
Morgan Freeman	41.95	1.63	0.26
Will Smith	41.87	8.61	2.64
Bruce Willis	41.44	1.86	0.3
Tom Hanks	36.40	4.16	0.79
Harrison Ford	35.26	3.25	0.6
Arnold Schwarzenegger	34.13	3.39	0.55
Johnny Depp	31.53	1.27	0.08
Frank Welker	30.99	2.86	0.34
Brad Pitt	30.61	1.54	0.09
Judi Dench	30.01	8.87	1.41
Leonardo DiCaprio	29.52	36.47	7.35
Adam Sandler	28.39	3.22	0.19
Liam Neeson	27.40	0.91	0.03
Denzel Washington	26.09	2.01	0.11
Eddie Murphy	25.82	3.30	0.27
John Travolta	23.96	2.45	0.13
Robin Williams	23.26	1.12	0.03
Hugo Weaving	23.02	4.20	0.12
Michael Caine	22.23	3.03	0.11
Channing Tatum	22.22	2.33	0.06
Angelina Jolie	20.81	3.72	0.09
Carla Gugino	20.74	4.24	0.13
Octavia Spencer	20.68	5.78	0.13
Kathy Bates	20.58	8.25	0.17
Jim Carrey	19.79	2.79	0
Ian McKellen	19.02	4.70	0
Reese Witherspoon	18.97	8.17	0.27
Zoe Saldana	18.79	7.33	0.12
Kevin Hart	18.41	3.21	0.05
Timothy Spall	18.40	9.39	0.25
Jamie Foxx	18.15	6.02	0.01
Judy Greer	17.76	4.17	0.01
Tommy Lee Jones	17.67	2.96	0
Rose Byrne	16.85	7.62	0.05
Andy Serkis	16.53	9.48	0.05
Dennis Hopper	-14.79	19.32	0
Kate Bosworth	-16.01	5.20	0.04
Viggo Mortensen	-16.24	6.50	0.02
Tim Blake Nelson	-16.58	24.78	0.09
Jeremy Piven	-17.45	4.86	0.06
Elias Koteas	-21.64	13.87	1.06
Susan Sarandon	-22.57	7.90	0.24
Mark Ruffalo	-23.29	3.17	0.11



(a) (b)

Figure 15: (a) Estimated partial R^2 for observed confounders using method described in section 5.1. Budget is the most dominant variable, which can explain significantly higher variation in outcome Y. (b) Latent confounder dimension selection, based on the reconstruction loss on the validation set.