Treatment Effect Risk: Bounds and Inference

Emanuel Nussli
Department of Mathematics
ETH Zürich
enussli@student.ethz.ch

Valentin Roth
Department of Mathematics
ETH Zürich
varoth@student.ethz.ch

Abstract

The fundamental problem of causal inference prevents the observation of unit–specific treatment effects. Estimation of treatment effects is therefore performed across individuals. While a positive average treatment effect suggests that the treatment is beneficial on average, there might be adverse effects in the worst–affected sub–population of a given size. To formalize this notion of treatment risk, we adopt the conditional value at risk (CVaR) of individual treatment effects and provide an identifiable lower bound. Existing bounds for the treatment risk using the CVaR rely on the conditional average treatment effect and thus ignore variability in individual treatment effects not explained by observable covariates. To capture variability in the distribution of conditional treatment effects beyond their mean, we use an average of CVaRs of the conditional distributions of treatment effects, given individuals' covariates. We set–identify the resulting bound and study its estimation using distributional regression models. Our results complement the existing literature and investigate a trade–off between bounds depending on two different sources of variability in individual treatment effects. The results thereby motivate studying bounds which combine both sources.

1 Introduction and Motivation

Across disciplines, decision–makers rely on statistical methodology to infer the effects of different treatments on a population of interest. Since single individuals cannot be given several treatment options simultaneously, treatment effects cannot be estimated on the individual level. Thus, treatment effects are estimated across individuals whereby the average treatment effect (ATE) is a widely–used quantity to assess whether treatments have their intended effect.

The ATE captures the central tendency of how individuals react to a given treatment. Yet, due to heterogeneity in the population, individual treatment effects (ITEs) can deviate strongly from their average. For example, when measuring how a certain drug affects the growth of a tumor, an overall negative, statistically significant ATE suggests that tumor growth is slowed down – on average. Single individuals in the tails of the ITE–distribution, however, might experience faster growth after taking the drug and thus are negatively affected by it.

To assess the risk of a treatment negatively impacting some individuals – even though the ATE suggests a positive effect – knowledge of the ITE–distribution is required. As a first step towards this goal, individuals' observable covariates can be used in quantities like the conditional average treatment effect (CATE) to capture some variability in the ITE–distribution. As a measure of treatment risk for the case of binary outcomes, Kallus [2022] proposes the fraction of negatively affected (FNA) and uses the CATE to sharply set–identify the suggested quantity. For the more general setting of real–valued outcomes, Kallus [2023] advocates to quantify a treatment's risk as the average treatment effect in the worst affected sub–population of a given size. Kallus formalizes this risk measure as the conditional value at risk (CVaR) and overcomes its non–identifiability by again bounding it using the CATE. Kallus' bound on the CVaR is tight if all heterogeneity in the individuals' outcomes

Preprint. Under review.

is explained by their covariates. Yet, since the CATE disregards variability from further sources of randomness, the provided bound can perform poorly when there is considerable variability in conditional treatment effects (conditional on observed covariates).

Hence, in this work, we present a complementary lower bound to the one presented in Kallus [2023], which incorporates the aforementioned sources of variability. For this, we likewise adopt the CVaR of the ITE as a measure of treatment risk but do not rely on the CATE to bound it. Instead, we capture aspects of the distribution of the ITE through the conditional distribution of treatment effects given individuals' covariates. Unlike the CATE, this conditional distribution cannot directly be estimated from the data since the dependence structure between both potential outcomes is needed but not identifiable. We therefore identify the bound up to a set whose endpoints are reached in the cases of co– and counter–monotonic potential outcomes. To do so, we use that the CVaR is a coherent risk measure and for further, more fine–grained analysis that it qualifies as a so–called D_2–functional. Further, we show that our bound is tight when the covariates do not explain any variability of the ITE. In addition, we investigate the trade–off between the tightness of Kallus' and our bound, depending on the structure of this variability.

To convey these insights, the remaining paper is structured as follows: Firstly, we introduce the general setup in Section 1.1 and formalize the problem in Section 1.2. We then present our bound and its set–identification in Section 2, where we also provide a simple example with closed–form expressions that give intuition for our results and their relation to Kallus' bound. Afterwards, we show that neither of the bounds dominates the other. In addition, this section elaborates on the estimation of our lower bound. Section 3 discusses how the methodology can be applied to real–world data through an example on Switzerland's COVID–19 facial–mask policy. In Section 4, we summarize our results and provide recommendations for the application of our and Kallus' bound for practitioners. Lastly, we present possible future research directions.

1.1 Setting

In the following, the treatment of the i-th unit is binary, i.e., $A_i \in \{0, 1\}$. We let $Y_i^{(1)}$ be the potential outcome of unit i under treatment, whereas $Y_i^{(0)}$ symbolizes the outcome of unit i under no treatment [Rubin, 1974]. Furthermore, we observe a vector of pre–treatment covariates X_i. The researcher has access to n i.i.d samples $(Y_i^{(A_i)}, A_i, X_i)_{i=1}^n \sim (X, A, Y^{(1)}, Y^{(0)})$ from some joint distribution denoted by $\mathbb{P}$. Thus,

$$A_i \in \{0, 1\},\ X_i \in \mathcal{X} \subseteq \mathbb{R}^p,\ \left(Y_i^{(1)}, Y_i^{(0)}\right) \in \mathbb{R}^2.$$

Using this setup, we can define the three central causal objects used in this paper:

$$\begin{aligned} \delta &:= Y^{(1)} - Y^{(0)} & & F_\delta(\cdot) \\ \delta|X &:= Y^{(1)} - Y^{(0)}|X & \text{with cdf} \quad & F_{\delta|X}(\cdot) \\ \tau(X) &:= \mathbb{E}\left[Y^{(1)} - Y^{(0)}|X\right] & & F_{\tau(X)}(\cdot). \end{aligned} \tag{1}$$

In the following, δ is referred to as the individual treatment effect (ITE), $\delta|X$ as the conditional treatment effect (CTE), and $\tau(X)$ as the conditional average treatment effect (CATE). Additionally, expectations are always taken with respect to $\mathbb{P}$, unless stated otherwise. Throughout, we assume δ to have finite second moment, i.e., $\mathbb{E}(\delta^2) < \infty$.

In addition, we impose the assumption of Unconfoundedness, which states

$$\left(Y^{(1)}, Y^{(0)}\right) \perp\!\!\!\perp A \mid X, \tag{2}$$

meaning that – given X – treatment assignment is not dependent on the potential outcomes. This is guaranteed in randomized trials and can be argued for in observational studies by including all covariates implied as necessary by the (learned or assumed) directed acyclic graph (DAG). This assumption enables the identification of both the ATE and the CATE.

Notation. We use $(u)_+ := \max(0, u)$ for some $u \in \mathbb{R}$. We use $Z \sim F_Z$ to denote a random variable Z distributed according to the cumulative distribution function F_Z. Further, we write $U_{1:n}$ for the collection of $U_1, \ldots, U_n$. To denote convergence in probability, we write $A_n \xrightarrow{p} A$. Correspondingly, we write $A_n \xrightarrow{d} A$ for convergence in distribution. To denote equality of distribution, we write $A \overset{d}{=} B$.

1.2 Problem Definition

We are interested in quantifying the risk of treatments as outlined in Section 1. Without loss of generality, we consider the situation where a treatment A is deemed successful if it reduces the outcome Y. Thus, a suitable definition of treatment risk has to summarize the right tail of the individual treatment effect distribution. The conditional value at risk (CVaR) is a natural candidate. For a continuous random variable Z, the CVaR has the following, intuitive definition:

$$\mathrm{CVaR}_\alpha(Z) \coloneqq \mathbb{E}\left[Z \mid Z \geq F_Z^{-1}(1-\alpha)\right], \tag{3}$$

where $F_Z^{-1}(1-\alpha) \coloneqq \inf\{\gamma : F_Z(\gamma) \geq 1-\alpha\}$ is the generalized inverse cdf of Z [Rockafellar and Uryasev, 2002]. Thus, the CVaR of δ at level α formalizes the expected individual treatment effect among the $\alpha \cdot 100\%$–worst affected units. This nicely captures the essence of treatment risk. For general random variables, the Definition in (3) can be generalized as:

$$\mathrm{CVaR}_\alpha(Z) \coloneqq \inf_{\beta \in \mathbb{R}} \Big(\beta + \frac{1}{1-\alpha}\mathbb{E}((Z-\beta)_+)\Big). \tag{4}$$

The CVaR carries a few interesting properties that will be useful later, such as being a coherent risk measure. Specifically, the properties of sub–additivity coupled with co–monotonic additivity are especially desirable in the scenario considered [Acerbi and Tasche, 2002].

However, $\mathrm{CVaR}_\alpha(\delta)$ is not identifiable from $(Y_i^{(A_i)}, A_i, X_i)_{i=1}^n$ due to the fundamental problem of causal inference as we only ever observe either $Y_i^{(1)}$ or $Y_i^{(0)}$ per unit i [Holland, 1986]. This issue can be nicely illustrated by the way the factual outcomes Y_i are generated (assuming causal consistency, see Rubin [1986])

$$Y_i = A_i Y_i^{(1)} + (1-A_i)Y_i^{(0)}. \tag{5}$$

Yet, if one can provide identifiable (tight) lower bounds for $\mathrm{CVaR}_\alpha(\delta)$, then these can be used as a screening tool for treatment risk. Concretely, if the lower bounds flag substantial treatment risk, this implies that the true treatment risk exceeds that risk–level, meaning that the treatment should potentially further be investigated for adverse effects in some sub–populations.

Deriving the lower bound, providing results regarding tightness, and comparing the newly derived bounds to the existing literature will be discussed in Section 2.1. Furthermore, we consider estimation and inference for the proposed lower bounds in Section 2.2.

2 Main Results

In the following, we discuss the proposed lower bound for the CVaR of the individual treatment effect and provide set–identification for the causal estimand in Section 2.1. Secondly, we describe the estimation and ways to perform inference in Section 2.2.

2.1 Causal Estimand and Partial Identification

Firstly, we will motivate our proposed lower bound for $\mathrm{CVaR}_\alpha(\delta)$. Under the assumption of Unconfoundedness (2), one can identify the CATE by using the linearity of the expectation, meaning

$$\begin{aligned}\mathbb{E}\left[Y^{(1)} - Y^{(0)} \mid X\right] &= \mathbb{E}\left[Y^{(1)} \mid X\right] - \mathbb{E}\left[Y^{(0)} \mid X\right] \\ &= \mathbb{E}\left[Y \mid X, A=1\right] - \mathbb{E}\left[Y \mid X, A=0\right],\end{aligned} \tag{6}$$

where the last equality follows from Unconfoundedness [Holland, 1986]. The last term can easily be estimated from the data by, e.g., using a regression–adjustment approach. This approach summarizes the conditional distribution $F_{\delta|X}$ by its implied expectation (which consequentially yields identification thereof). Kallus [2023] suggests using the CVaR of the CATE to lower bound $\mathrm{CVaR}_\alpha(\delta)$.[1]

[1] We transfer Kallus' Theorem 3.1 to our setting in Appendix 5.1, as Kallus works with the left–tail CVaR instead of the right–tail CVaR, as we do.

While Kallus' bound holds with equality when $\tau(X) \coloneqq \mathbb{E}(\delta|X) = \delta$, meaning the observable covariates X explain all variability in the ITE, it disregards all information besides the first moment of $F_{\delta|X}$. Thus, we strive to use the additional information contained in $F_{\delta|X}$. Instead of collapsing $F_{\delta|X}$ to its mean and computing the CVaR of the CATE, we directly summarize the information in the right–tail of $F_{\delta|X}$. Analogously to the CATE in Equation (6), we consider

$$\mathrm{CVaR}_\alpha(\delta|X) \coloneqq \mathrm{CVaR}_\alpha\left(Y^{(1)} - Y^{(0)} \mid X\right), \tag{7}$$

which is defined through the conditional expectation (given X). To then summarize the random variable in Definition (7), we consider the expected value (where the expectation solely runs over X as the only source of randomness). However, the object in Definition (7) is not point–identifiable [Cross and Manski, 2002]. We will revisit this issue after introducing our first Theorem, which governs our proposed lower bound of $\mathrm{CVaR}_\alpha(\delta)$.

Theorem 1. *Let F_δ be the cumulative distribution function of the individual treatment effect and let $F_{\delta|X}$ be the cumulative distribution function of the conditional treatment effect. Then, with $\delta \sim F_\delta$ and $\delta|X \sim F_{\delta|X}$, for any fixed $\alpha \in (0,1)$, we have*

$$\mathrm{CVaR}_\alpha(\delta) \geq \mathbb{E}\,\mathrm{CVaR}_\alpha(\delta|X). \tag{8}$$

Moreover, given any δ–distribution, there exists some (X,δ)–distribution such that Equation (8) holds with equality.

Proof. We will use the fact that for any measurable function h and random variable Z, we have that $\inf_h \mathbb{E}h(Z) \geq \mathbb{E}\inf_h h(Z)$, which is easily proved. Thus, we write – using the Definition (4) of the CVaR, for all $\alpha \in (0,1)$,

$$\begin{aligned}
\mathrm{CVaR}_\alpha(\delta) &= \inf_{\beta\in\mathbb{R}}\left(\beta + \frac{1}{1-\alpha}\mathbb{E}\left((\delta-\beta)_+\right)\right) \\
&= \inf_{\beta\in\mathbb{R}}\left(\beta + \frac{1}{1-\alpha}\mathbb{E}\mathbb{E}\left((\delta-\beta)_+ \mid X\right)\right) \\
&\geq \mathbb{E}\inf_{\beta\in\mathbb{R}}\left(\beta + \frac{1}{1-\alpha}\mathbb{E}\left((\delta-\beta)_+ \mid X\right)\right) \\
&= \mathbb{E}\inf_{\beta\in\mathbb{R}}\left(\beta + \frac{1}{1-\alpha}\mathbb{E}\left((\delta|X-\beta)_+\right)\right).
\end{aligned} \tag{9}$$

We use iterated expectations in the first line and $\inf_h \mathbb{E}h(Z) \geq \mathbb{E}\inf_h h(Z)$ in the second line. As in this instance, h is not constant, $\inf_h \mathbb{E}h(X) = \mathbb{E}\inf_h h(X)$ holds only when the distribution of X is degenerate. Therefore, if $\mathrm{Var}(X) = 0$, the statement holds with equality. □

This means that the object in Definition (7) is a valid lower bound for $\mathrm{CVaR}_\alpha(\delta)$. This brings us back to the issue of identification discussed in Section 1.2, as only a certain class of statistical functionals $\gamma : F_{\delta|X} \longmapsto \gamma(F_{\delta|X})$ is identifiable from the observed $(Y_i^{(A_i)}, A_i, X_i)_{i=1}^n$. Because we do not have access to unit–couplings of $Y^{(1)}, Y^{(0)}$, we will not be able to estimate the joint distribution of the potential outcomes – also not conditional on X.[2] Hence, the only functionals of $F_{\delta|X}$ that are identifiable solely depend on the marginals $F_{Y^{(1)}|X}, F_{Y^{(0)}|X}$.

Remark 1. *One such functional that solely depends on the marginals is the expectation of course. Yet, it is crucial for risk measures to incorporate the dependencies between the random variables, so looking for a risk measure that fulfills this condition is ill–posed.*

On the bright side, we can use two important properties of the CVaR to provide a lower and an upper bound to the CVaR in Definition (7). Furthermore, we can navigate the interval between the lower and the upper bound depending on the hypothesized, and inherently not observable, dependence structure of the potential outcomes. The first property is the sub–additivity of the CVaR [Acerbi and Tasche, 2002].

[2]Thus, we also cannot estimate the distribution $F_{\delta|X}$ because we need access to the joint distribution for that.

Lemma 1 (Embrechts and Wang [2015]). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an atom–less probability space. Let L^0 be the set of all random variables. Then, for all $\alpha \in (0,1)$, we have*

$$\mathrm{CVaR}_\alpha(X+Y) \leq \mathrm{CVaR}_\alpha(X) + \mathrm{CVaR}_\alpha(Y), \tag{10}$$

for all $X, Y \in L^0$.

This Lemma already hints at how one can provide bounds for CVaR in Definition (7), by directly using the property of sub–additivity. Nevertheless, being able to navigate the identified set by varying the supposed dependence structure of the potential outcomes is crucial. This is partially enabled by the second crucial property of the CVaR.

Lemma 2 (Kusuoka [2001]). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an atom–less probability space. Let X, Y be co–monotonic random variables, i.e., there exists a r.v. Z and non–decreasing functions g, f such that $(X,Y) \overset{d}{=} (g(Z), f(Z))$. Then, for all $\alpha \in (0,1)$, we have*

$$\mathrm{CVaR}_\alpha(X+Y) = \mathrm{CVaR}_\alpha(X) + \mathrm{CVaR}_\alpha(Y), \tag{11}$$

Using Lemmata 1 and 2 and theory on the joint distribution of the potential outcomes for certain functionals of the joint distribution called D_2–functionals (to which the CVaR belongs), we establish the identified set of $\mathrm{CVaR}_\alpha(Y^{(1)} - Y^{(0)}|X)$ [Fan and Park, 2010]. Further, we show that the lower bound is attained when, both conditional on X, the potential outcomes are co–monotonic, and the upper bound is attained when the potential outcomes are counter–monotonic.

Remark 2. *Note that (conditional on X), counter–monotonic potential outcomes specify the case where, e.g., low values of $Y^{(0)}$ (=good) are coupled with high values of $Y^{(1)}$ (=bad). This encodes the most pessimistic assumption about the unobserved dependence and the treatment in general and thus marks the upper bound of the risk in this approach. The converse holds for co–monotonicity, which encapsulates the most optimistic perspective.*

We are now ready to state the formal result.

Theorem 2. *Let $F_{Y^{(1)}|X}$ and $F_{Y^{(0)}|X}$ be the cumulative distribution functions of the two X–conditional potential outcomes. Denoting $Y^{(1)}|X \sim F_{Y^{(1)}|X}$ and $Y^{(0)}|X \sim F_{Y^{(0)}|X}$ and $\delta|X \sim F_{\delta|X}$, for any fixed $\alpha \in (0,1)$, we have*

$$\mathbb{E}\,\mathrm{CVaR}_\alpha(\delta|X) \in \Big[\psi_{lb}, \psi_{ub}\Big], \tag{12}$$

where

$$\begin{aligned}
\psi_{lb} = \max\Big(&\mathbb{E}\,\mathrm{CVaR}_\alpha(Y^{(1)}|X) - \mathbb{E}\,\mathrm{CVaR}_\alpha(Y^{(0)}|X), \\
&\mathbb{E}\,\mathrm{CVaR}_\alpha(-Y^{(0)}|X) - \mathbb{E}\,\mathrm{CVaR}_\alpha(-Y^{(1)}|X)\Big), \\
\psi_{ub} = &\mathbb{E}\,\mathrm{CVaR}_\alpha(Y^{(1)}|X) + \mathbb{E}\,\mathrm{CVaR}_\alpha(-Y^{(0)}|X).
\end{aligned}$$

Furthermore, the upper bound ψ_{ub} is reached when $Y^{(1)}$ and $Y^{(0)}$ are X–conditional counter–monotonic random variables. The lower bound ψ_{lb} is reached if $Y^{(1)}$ and $Y^{(0)}$ are X–conditional co–monotonic random variables. In either case, it holds that $\mathrm{CVaR}_\alpha(\delta) \geq \varphi_{lb}$.

Proof. See Proof 5.2 in Appendix 5.2. □

Remark 3. *In addition to the lower– and upper bounds presented in Theorem 2, we suggest additionally estimating $\mathrm{CVaR}_\alpha(\delta|X)$ under the assumption of X–conditional independence of the potential outcomes. This will present the middle–ground between the extreme assumptions of co– and counter–monotonicity and should help guide the practitioner. To implement this approach, derive the distribution of $Y^{(1)} - Y^{(0)} \mid X$ via*

$$F_{Y^{(1)}-Y^{(0)}|X} = \int F_{Y^{(1)}|X}(y_1 + y_0)\,\mathrm{d}F_{Y^{(0)}|X}(y_0). \tag{13}$$

Then, use the derived cumulative distribution function, compute the CVaRs, and perform the aggregation as usual. We denote this approach by ψ_{id}.

Now, we have introduced three new quantities ψ_{lb}, ψ_{ub} and ψ_{id} that are identifiable from the data because they solely depend on the X–conditional marginal distributions of the potential outcomes – and not on the joint distribution (see Equation (21) in Section 2.2 for more details on the identifiability of the marginals). Furthermore, we discussed the lower–bound proposed in Kallus' paper. To provide some intuition for these results, let us consider a simple example where we have closed–form expressions for all terms involved.

Example 1. *Assume linear Gaussian generating equations for the potential outcomes $Y^{(1)}, Y^{(0)}$. Concretely, let*

$$Y^{(j)} = \alpha_j + X\beta_j + \epsilon_j,\ j = 0, 1, \tag{14}$$

where $X, \epsilon_1, \epsilon_0$ are jointly Gaussian (with variances $\sigma_X^2, \sigma_1^2, \sigma_0^2$) and α, β are non–random. For more details, consider the Appendix 5.3. By elementary properties of Gaussians, we can determine the distribution of the potential outcomes and utilize the closed–form expression for the CVaR_α of Gaussians. For some $B \sim \mathcal{N}(\mu, \sigma^2)$, one gets

$$\mathrm{CVaR}_\alpha(B) = \mu + \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}\sigma, \tag{15}$$

where ϕ, Φ are the density– and the cumulative distribution functions of a standard–normal Gaussian. After some calculations, we obtain the three important CVaRs: the truth, our approach (the theoretical one assuming access to $F_{\delta|X}$), and Kallus' approach.

$$\begin{aligned}
\mathrm{CVaR}_\alpha(\delta) &= \mu_\delta + \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}\sqrt{(\beta_1-\beta_0)^2\sigma_X^2 + \sigma_1^2 + \sigma_0^2 - 2\rho\sigma_1\sigma_0} \\
\mathbb{E}\,\mathrm{CVaR}_\alpha(\delta|X) &= \mu_\delta + \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}\sqrt{\sigma_1^2 + \sigma_0^2 - 2\rho\sigma_1\sigma_0} \\
\mathrm{CVaR}_\alpha(\tau(X)) &= \mu_\delta + \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}\sqrt{(\beta_1-\beta_0)^2\sigma_X^2},
\end{aligned} \tag{16}$$

where ρ is the correlation between the potential outcomes, conditional on X. Here, ρ captures the trade–off between co– and counter–monotonicity.

The reader observes that our approach can capture the risk of treatment conditional on X, which confirms our intuition, as we compute the CVaR based on $F_{\delta|X}$. However, we miss the heterogeneity in the potential outcomes introduced by X as we take the expectation over X, thus disregarding the tails and their information about treatment risk. For Kallus' approach, we find the exact opposite behavior. This is also what we would expect because conditional on X, he ignores the risk information by taking the expectation. However, as he computes the CVaR over X, he incorporates the variability in the potential outcomes introduced via variability in X. This trade–off can be nicely described using the law of total variance, i.e.,

$$\mathrm{Var}(\delta) = \mathbb{E}\Big[\mathrm{Var}(\delta|X)\Big] + \mathrm{Var}\Big[\mathbb{E}(\delta|X)\Big]. \tag{17}$$

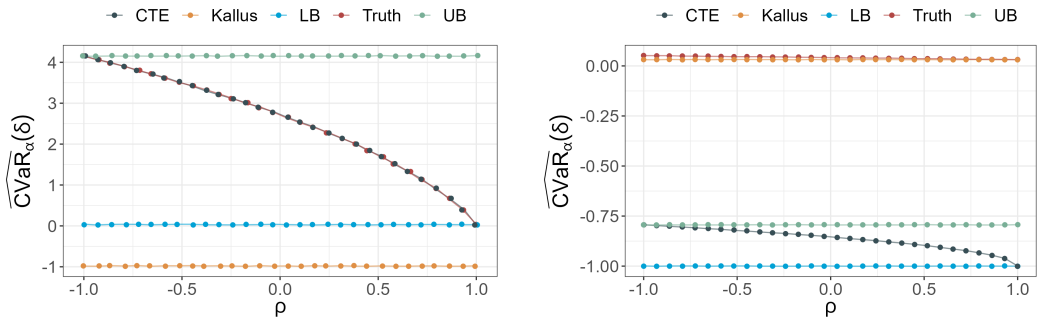
Note that this does not hold for the CVaR, however, it is still an instructive decomposition of the tail–behavior of δ. We see that our approach captures the risk associated with the first term on the RHS in the variance decomposition (17) well, while Kallus's approach captures the risk encoded in the second term of the RHS.

We visualize the described behavior of the two bounds in the case of linear Gaussian generating equations in Figure 1. We see that our approach is tight in regime a) whereas Kallus' approach is tight in regime b). Furthermore, we see that $\mathbb{E}\,\mathrm{CVaR}_\alpha(\delta|X)$ reaches ψ_{ub} if $\rho = -1$ and ψ_{lb} if $\rho = 1$. Note that $\rho = -1 \Rightarrow$ counter–monotonicity and vice versa for $\rho = 1$. It is thus an illustration for the second statement of Theorem 2.

Remark 4. *Observe that the linear Gaussian case in Example 1 allows us to construct another bound by combining our and Kallus' approach. The resulting lower bound is always as good or better than both and is given as:*

$$\mathrm{Comb}(\delta) = \mu_\delta + \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}\sqrt{(\beta_1-\beta_0)^2\sigma_X^2 + \sigma_1^2 + \sigma_0^2 - 2\sigma_0\sigma_1}. \tag{18}$$

It is worth noting that this bound equals $\mathrm{CVaR}_\alpha(\delta)$ up to the value of ρ. Here, a choice of $\rho = 1$ assures the quantity to be a true lower bound of $\mathrm{CVaR}_\alpha(\delta)$. We provide further insights on the validity of this bound in the Appendix 5.3.



(a) (b)

Figure 1: The bounds in Section 2.1 as well as the truth depending on the X–conditional correlation ρ between the potential outcomes. CTE denotes $\mathbb{E}\,\mathrm{CVaR}_\alpha(\delta|X)$, Kallus denotes $\mathrm{CVaR}_\alpha(\tau(X))$ and Truth denotes $\mathrm{CVaR}_\alpha(\delta)$. Furthermore, LB refers to ψ_{lb} from Theorem 2 and UB to ψ_{ub} from said Theorem. In Figure (a), we have $\sigma_X^2 \approx 0$. In Figure (b), we consider $\sigma_1^2 = \sigma_0^2 \approx 0$.

We can gain two insights from Example 1. Firstly, the case of linear Gaussian generating equations already informally shows that, while both are a lower bound for $\mathrm{CVaR}_\alpha(\delta)$, neither Kallus' nor our lower bound introduced in Theorem 1 always dominates the respective other. Thus, neither is generally preferable. We formalize this result in Corollary 1.

Corollary 1. *Let $F_{\delta|X}$ be the cumulative distribution function of the conditional treatment effect. Now for $\delta|X \sim F_{\delta|X}$, $\tau(X) = \mathbb{E}(\delta|X)$ and any $\alpha \in (0,1)$,*

$$\mathrm{CVaR}_\alpha(\mathbb{E}(\delta|X)) \leq \mathbb{E}\,\mathrm{CVaR}_\alpha(\delta|X) \tag{19}$$

cannot hold for all (X,δ)–distributions. The statement with the inequality reversed is also true.

Proof. See Proof 5.2 in Appendix 5.2 □

Secondly, the example shows that better lower bounds such as $\mathrm{Comb}(\delta)$ that dominate Kallus' and our lower bound can be constructed in special cases. Note, however, that generalizing the combined bound introduced in Remark 4 beyond the linear Gaussian case is not straightforward.

In summary, in this Section, we have presented a theoretical lower bound for $\mathrm{CVaR}_\alpha(\delta)$, derived set–identification thereof and thereby also constructed an identifiable lower bound. In the following Section 2.2, we present a method of estimation and statistical inference for the quantities we presented.

2.2 Estimation and Inference

To estimate the proposed bounds, note that they are all statistical functionals of the conditional cumulative distribution functions $F_{Y^{(1)}|X}$ and $F_{Y^{(0)}|X}$. Concretely, taking ψ_{ub} as an example, we get

$$\psi_{ub} = \mathbb{E}\,\mathrm{CVaR}_\alpha(F_{Y^{(1)}|X}) + \mathbb{E}\,\mathrm{CVaR}_\alpha(F_{-Y^{(0)}|X}), \tag{20}$$

by writing ψ_{ub} from Theorem 2 in the spirit of a statistical functional γ, where $\gamma(G,F) = \mathbb{E}\,\mathrm{CVaR}_\alpha(G) + \mathbb{E}\,\mathrm{CVaR}_\alpha(F)$. Thus, as a first step, we need to gain access to high–quality estimates of $F_{Y^{(1)}|X}$ and $F_{Y^{(0)}|X}$. To do so, note that under Unconfoundedness (2), we have (for $t \in \mathbb{R}$)

$$\begin{aligned} \mathbb{P}(Y^{(1)} \leq t \mid X = x) &= \mathbb{P}(Y \leq t \mid X = x, A = 1) \\ \mathbb{P}(Y^{(0)} \leq t \mid X = x) &= \mathbb{P}(Y \leq t \mid X = x, A = 0). \end{aligned} \tag{21}$$

Hence, the first task reduces to the estimation of conditional cumulative distribution functions. We propose to use distributional random forests to accomplish this task due to a) the intuitive construction of the estimates and b) their theoretical properties. By Corollary 5 of Cevid et al. [2022], we get

$$\widehat{F}_{Y|X=x,A=a}(t) \xrightarrow{p} F_{Y|X=x,A=a}(t), \tag{22}$$

for all continuity points $t \in \mathbb{R}$ of the population version. Further, convergence in probability also holds for the generalized inverse cdf.

Given access to the estimates of the conditional cumulative distribution functions described in Statement (22), we need a method to estimate the bounds, i.e., construct an estimator for $\mathbb{E}\,\mathrm{CVaR}_\alpha(\cdot)$ and compose the resulting terms according to Theorem 2. We employ the plug–in principle by considering the tail–expectation interpretation of the CVaR described in Definition (3), and estimating it with its empirical counterpart. This leads to the following Lemma, which governs the theoretical properties of the described estimator. The following results are taken from Zwingmann and Holzmann [2016], while adapting them to our scenario.

Lemma 3 (Zwingmann and Holzmann [2016])**.** *Let $\widehat{q}_{1-\alpha,x,a} \coloneqq \inf\{\gamma : \widehat{F}_{Y|X=x,A=a}(\gamma) \geq 1-\alpha\}$ be the empirical $(1-\alpha)$–quantile estimated using the distributional random forest. Furthermore, let $n' \coloneqq |\{i \in \{1,\dots,n\} : X_i = x, A_i = a\}|$. Then, we get*

$$\frac{1}{\alpha n'} \sum_{i=1}^{n'} Y_i \mathbb{1}_{Y_i \geq \widehat{q}_{1-\alpha,x,a}} \xrightarrow{p} \mathrm{CVaR}_\alpha(Y \mid X = x, A = a). \tag{23}$$

This implies that the plug–in approach constitutes a consistent estimator. Note that for Lemma 3, the nuisance parameter $q_{1-\alpha,x,a}$ needs to be consistently estimated, which is guaranteed by Corollary 5 presented by Cevid et al. [2022]. Thus, we know that the most central building block of our bounds can be estimated sufficiently well. However, as we are interested in $\mathbb{E}\,\mathrm{CVaR}_\alpha(Y \mid X, A = a)$ with $a \in \{0,1\}$, we further need that (informal)

$$\widehat{\mathbb{E}}_X \frac{1}{\alpha n'} \sum_{i=1}^{n'} Y_{i,j} \mathbb{1}_{Y_{i,j} \geq \widehat{q}_{1-\alpha,x,a}} \xrightarrow{p} \mathbb{E}\,\mathrm{CVaR}_\alpha(Y \mid X = x, A = a). \tag{24}$$

This follows directly from the Law of Large Numbers as we draw i.i.d samples from X. We denote the LHS of Statement 24 by $\widehat{\mathbb{E}}_X \widehat{\mathrm{CVaR}}_\alpha(Y|X = x, A = a)$.

To summarize, we show how to construct well–behaved estimators for the central quantities in Theorem 2. Concretely, we have

$$\begin{aligned} \widehat{\mathbb{E}}_X \widehat{\mathrm{CVaR}}_\alpha(Y|X = x, A = 1) &\xrightarrow{p} \mathbb{E}\,\mathrm{CVaR}_\alpha(Y^{(1)} \mid X = x) \\ \widehat{\mathbb{E}}_X \widehat{\mathrm{CVaR}}_\alpha(Y|X = x, A = 0) &\xrightarrow{p} \mathbb{E}\,\mathrm{CVaR}_\alpha(Y^{(0)} \mid X = x). \end{aligned} \tag{25}$$

Hence, to construct both ψ_{lb} and ψ_{ub}, we suggest replacing the population quantities in the lower– and upper bounds by their empirical counterparts described in Statement 24. We provide more details on the estimation and inference in the Appendix 5.4, including the limiting distribution of the proposed estimator, which allows us to perform statistical inference.

In the following Section 3, we provide an empirical application illustrating the proposed concepts.

3 Application

To provide a guide on how to use the proposed concepts, we briefly work through an empirical example. We consider the estimation of the impact of the strict facial–mask policy on the COVID–19 dynamics in Switzerland.

During several weeks between July 2020 and December 2020, the cantons of Switzerland could choose to adopt the government–determined facial–mask policy (mandatory facial–mask wearing on public transport) or a strict facial–mask policy (mandatory facial–mask wearing in public transport and in all public or shared spaces where social distancing was not possible). We use data collected by Nussli et al. [2024] and employ their causal assumptions.

Figure 2 briefly lists the driving variables in the causal model, which are grouped into categories. Bold–face letters denote collections of variables. The data is observed at the cantonal $\times$ weekly level.

The DAG on the right–hand side visualizes the assumed causal structure of COVID–19 dynamics under policy interventions. Ultimately, we are interested in the treatment effect of the strict facial mask policy ($M_{i,t}$) on the outcome of the pandemic ($Y_{i,t+\ell}$). To do so, we estimate the four quantities

$M_{i,t}$: facial–mask policy indicator (baseline or strict policy)

$Y_{i,t+\ell}$: pandemic state (effective reproductive number; leaded to account for the reporting delay)

$Y_{i,t'}$: information about pandemic situation (past effective reproductive number)

$B_{i,t}$: social distancing behavior (growth rate of monetary transactions)

$H_{i,t}$: holiday indicator

$\boldsymbol{P}_{i,t}$: all other COVID–19 related non–pharmaceutical policies

$\boldsymbol{D}_i$: demographic variables

$\boldsymbol{W}_{i,t}$: weather variables

$\boldsymbol{U}_i^1$: general canton–specific factors

$\boldsymbol{U}_t^2$: general time–specific factors

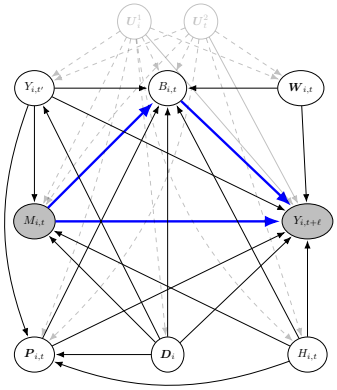


Figure 2: Left: the variables assumed to be driving the dynamics of COVID–19. Right: The DAG visualizing the dependencies between the variables. The blue edges represent the direct effect and the indirect effect of the facial–mask policy. The dashed edges represent unobserved dependencies.

($\psi_{lb}, \psi_{ub}, \psi_{id}$ and Kallus' bound) using the methodology described in Section 2.2.[3] The results are visualized in Figure 3. In this application, a negative treatment effect implies that the strict facial–mask policy was successful in limiting the spread of COVID–19. Depending on the specification, Nussli et al. [2024] estimate average treatment effects (ATE) ranging from -0.29 to -0.16.

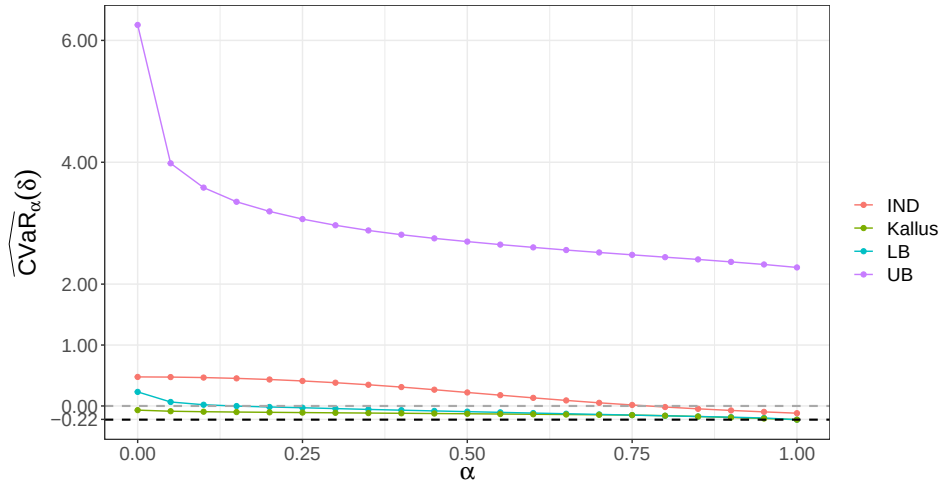


Figure 3: The four bounds of $\text{CVaR}_\alpha(\delta)$ described above. IND refers to ψ_{id}, Kallus to Kallus' approach, LB to ψ_{lb} and UB to ψ_{ub}. The black dashed line at -0.22 represents the ATE.

In this example, we see that ψ_{lb} and Kallus' bound are very close for all values of α. Furthermore, the value under the assumption of X–conditional independent potential outcomes is also rather close to them, albeit higher as is necessary. The upper bound is much larger, thus symbolizing that the strict facial–mask policy was not a good treatment if the potential outcomes were counter–monotonic in reality, conditional on X. Again, as described in Remark 3, this corresponds to the most pessimistic

[3]Note that we first have to apply the within–operator to remove the effects of $\boldsymbol{U}_i^1, \boldsymbol{U}_t^2$. See Nussli et al. [2024] for more details on the estimation procedure.

assumption regarding the treatment possible. Furthermore, note that – as mentioned in Theorem 2 – ψ_{ub} is not necessarily a lower bound for $\mathrm{CVaR}_\alpha(\delta)$, except for the case of X–conditional counter–monotonic potential outcomes.

In empirical applications where domain knowledge is available, we suggest that this knowledge should be used to navigate the space between X–conditional co– and counter–monotonic potential outcomes. In this application, where all canton– or time–specific effects are captured through X, it is reasonable to assume that observations (at the level of week $\times$ canton) with high potential outcomes $Y^{(0)}$ also have high values of $Y^{(1)}$. This encodes the researchers' assumption that units with high values under no treatment also have high values once treated. Taking breast cancer as an example, it is very reasonable to impose that units with high $Y^{(0)}$ (severe breast cancer) would also have high $Y^{(1)}$ because chemotherapy is not more efficient for heavily ill patients. These considerations let us navigate towards the lower bound, which encodes co–monotonicity of the potential outcomes.

Lastly, remember that for $\alpha = 1$, the CVaR and the expectation coincide. Thus, for $\alpha = 1$, Kallus' approach and our lower bound recover the ATE. This shows that Figure 3 nicely conveys the information available, as bounds for the risk of treatment for varying α are displayed while also communicating the ATE at $\alpha = 1$. Here, the implied ATE is -0.22, which is consistent with the results of Nussli et al. [2024].

4 Conclusion

In this paper, we study the risk that the worst affected sub–population of a given size might face – despite the average treatment effect potentially indicating that the treatment is beneficial. To do so, we formalize treatment risk as the conditional value at risk of individual treatment effects.

Since the ITE is not observable, we first use the information on treatment risk encoded in the conditional distributions of individual treatment effects, given the individual's observable covariates, to provide a theoretical lower bound on the CVaR of the ITE. Second, we derive set–identification for our theoretical bound where the boundaries correspond to the cases of conditionally co– and counter–monotonic potential outcomes. In addition, we provide a lower bound under the assumption of conditional independence of potential outcomes.

Besides these two main results, we study our theoretical lower bound and its set–identification using an intuitive linear Gaussian example and describe the trade–off between our approach and an existing lower bound from the literature. We then formalize the notion that our lower bound complements the existing one and that neither of them is strictly better than the other one. Further, we study the estimation and statistical inference of the quantities we propose and show that our bounds can be estimated consistently using distributional random forests. Finally, we apply our results to assess the treatment risk of a strict facial–mask policy using COVID–19 data from Switzerland. The application to real world data shows that our bound and its set–identification – in addition to the existing bound – provides practitioners with a rich picture when evaluating treatment risk.

Based on this finding, we advise practitioners to use the set identification of our bound, its version for conditionally independent potential outcomes, and the existing bound. Given all, practitioners then know that, up to estimation error, the treatment risk must be above the maximum of the existing bound and the lower limit of our identified set. In case researchers have additional knowledge on the conditional dependence structure between potential outcomes, they can further orient themselves within our identified set – from conditional co–monotonicity via independence to counter–monotonicity of the potential outcomes.

Despite these advances, the trade–off between our and the existing bound suggests that neither is tight when both sources of variability discussed previously are large. In our analysis of the linear Gaussian case, we constructed a lower bound that incorporates both sources and therefore dominates the other bounds. Investigating its generalization might therefore be a promising direction for future research.

5 Appendix

5.1 External Theorem

Theorem (Theorem 3.1., Kallus [2023])**.**

$$\mathrm{CVaR}_\alpha(\delta) \geq \mathrm{CVaR}_\alpha(\tau(X)). \tag{26}$$

Moreover, given any X–distribution and integrable $\tau : \mathcal{X} \to \mathbb{R}$, some (X, δ)–distribution has the given X–marginal, $\tau(X) = \mathbb{E}(\delta|X)$, and Eq. (26) holding with equality.

5.2 Omitted Proofs

Proof Theorem 2

Proof. Firstly, we prove that $\mathbb{E}\,\mathrm{CVaR}_\alpha(\delta|X) \in \left[\psi_{lb}, \psi_{ub}\right]$. Following that, we show that the lower– and upper bounds are reached at X–conditional co– and counter–monotonicity of the potential outcomes. For the first part, we directly use the sub–additivity of the CVaR presented in Lemma 1. For the lower bound, consider

$$\begin{aligned} \mathrm{CVaR}_\alpha(Y^{(1)}|X) &= \mathrm{CVaR}_\alpha(Y^{(1)} - Y^{(0)} + Y^{(0)}|X) \\ &\leq \mathrm{CVaR}_\alpha(Y^{(1)} - Y^{(0)}|X) + \mathrm{CVaR}_\alpha(Y^{(0)}|X). \end{aligned} \tag{27}$$

This implies that

$$\mathrm{CVaR}_\alpha(Y^{(1)}|X) - \mathrm{CVaR}_\alpha(Y^{(0)}|X) \leq \mathrm{CVaR}_\alpha(Y^{(1)} - Y^{(0)}|X). \tag{28}$$

Now we want to employ the same method but for $-Y^{(0)}$ instead. We write

$$\begin{aligned} \mathrm{CVaR}_\alpha(-Y^{(0)}|X) &= \mathrm{CVaR}_\alpha(-Y^{(1)} + Y^{(1)} - Y^{(0)}|X) \\ &\leq \mathrm{CVaR}_\alpha(-Y^{(1)}|X) + \mathrm{CVaR}_\alpha(Y^{(1)} - Y^{(0)}|X). \end{aligned} \tag{29}$$

Thus, we get

$$\mathrm{CVaR}_\alpha(-Y^{(0)}|X) - \mathrm{CVaR}_\alpha(-Y^{(1)}|X) \leq \mathrm{CVaR}_\alpha(Y^{(1)} - Y^{(0)}|X). \tag{30}$$

The upper bound follows directly by sub–additivity as

$$\mathrm{CVaR}_\alpha(Y^{(1)} - Y^{(0)}|X) = \mathrm{CVaR}_\alpha(Y^{(1)} + (-Y^{(0)})|X) \tag{31}$$

$$\leq \mathrm{CVaR}_\alpha(Y^{(1)}|X) + \mathrm{CVaR}_\alpha(-Y^{(0)}|X). \tag{32}$$

Now we conclude the first part by taking expectations on both sides and choosing the max of both expressions for the lower bound (of the Inequalities (28) and (30)). This is justified as they both have to hold and we seek a tight lower bound.

In the following, we will describe when ψ_{lb}, ψ_{ub} are reached. Firstly, we employ the work of Fan and Park [2010]. They provide bounds for certain functionals of both the ITE– and the CTE–distribution. For the class of so–called D_2–functionals, they show in their Lemma 2.2

$$\gamma(F_{\delta'|X=x}) \leq \gamma(F_{\delta|X=x}) \leq \gamma(F_{\delta''|X=x}), \tag{33}$$

where δ' denotes X–conditional co–monotonic potential outcomes whereas δ'' denotes X–conditional counter–monotonic potential outcomes. Thus, we need to show that the CVaR is indeed a D_2–functional and lastly use the co–monotonic additivity of the CVaR described in Lemma 2.

To do so, we briefly introduce the notion of second–order stochastic dominance, as it is needed to define D_2–functionals.

Definition 1. *Consider two random variables A, B with cumulative distribution functions F, G such that $\int x\,\mathrm{d}F = \int x\,\mathrm{d}G$. We say that A second–order stochastically dominates B ($F \succsim_{sso} G$) iff*

$$\int_a^t G(x)\,\mathrm{d}x \geq \int_a^t F(x)\,\mathrm{d}x, \tag{34}$$

for all $t \geq 0$.

Now we are ready to introduce the notion of a D_2–functional.

Definition 2. *A functional $\gamma : F \longmapsto \gamma(F)$ is called a D_2–functional if the following implication holds*

$$F \succsim_{sso} G \implies \gamma(F) \geq \gamma(G). \tag{35}$$

Thus, we are now prepared to show that the CVaR is a D_2-functional.

Lemma 4. *The conditional value at risk is a D_2–functional.*

Proof. Let F, G be cumulative distribution functions such that $F \succsim_{sso} G$. By Acerbi's integral formulation of the CVaR [Acerbi and Tasche, 2002], with $X \sim F, Y \sim G$, we get that

$$\begin{aligned} \mathrm{CVaR}_\alpha(X) &= \frac{1}{1-\alpha}\int_\alpha^1 F^{-1}(\beta)d\beta \\ \mathrm{CVaR}_\alpha(Y) &= \frac{1}{1-\alpha}\int_\alpha^1 G^{-1}(\beta)d\beta \end{aligned} \tag{36}$$

Now, let us ignore the term outside of the integral, choose t from Inequality (34) to be 1 and $a = \alpha$. Because $F(\cdot)$ and $G(\cdot)$ are increasing, the Inequality (34) implies that $\mathrm{CVaR}_\alpha(X) \geq \mathrm{CVaR}_\alpha(Y)$. This proves the claim. □

Finally, we use Lemma 2 (co–monotonic additivity of the CVaR) to derive $\mathrm{CVaR}_\alpha(\delta'|X)$ with $\delta'|X \sim F_{\delta'|X}$ as well as $\mathrm{CVaR}_\alpha(\delta''|X)$ with $\delta''|X \sim F_{\delta''|X}$. If, conditional on X, $Y^{(1)}$ and $Y^{(0)}$ are counter–monotonic, then $Y^{(1)}$ and $-Y^{(0)}$ are co–monotonic. Hence $\mathrm{CVaR}_\alpha(\delta''|X) = \mathrm{CVaR}_\alpha(Y''^{(1)}|X) + \mathrm{CVaR}_\alpha(-Y''^{(0)}|X)$. Secondly, for the lower bound, if conditional on X, $Y^{(1)}$ and $Y^{(0)}$ are co–monotonic, then $Y^{(1)} - Y^{(0)}$ and $Y^{(0)}$ are co–monotonic. Thus, we can again use co–monotonic additivity. This proves the overall claim. □

Proof Corollary 1

Proof. Notice that the left and the right–hand side in Equation (19) can be obtained from the general definition of $\mathrm{CVaR}_\alpha(\delta)$ using iterated expectations and one inequality each:

$$\begin{aligned} \mathrm{CVaR}_\alpha(\delta) &= \inf_{\beta\in\mathbb{R}}\left(\beta + \frac{1}{1-\alpha}\mathbb{E}\mathbb{E}\left((\delta-\beta)_+ \mid X\right)\right) \\ &\geq \mathbb{E}\inf_{\beta\in\mathbb{R}}\left(\beta + \frac{1}{1-\alpha}\mathbb{E}\left((\delta-\beta)_+|X\right)\right), \end{aligned} \tag{37}$$

$$\begin{aligned} \mathrm{CVaR}_\alpha(\delta) &= \inf_{\beta\in\mathbb{R}}\left(\beta + \frac{1}{1-\alpha}\mathbb{E}\mathbb{E}\left((\delta-\beta)_+ \mid X\right)\right) \\ &\geq \inf_{\beta\in\mathbb{R}}\left(\beta + \frac{1}{1-\alpha}\mathbb{E}(\mathbb{E}(\delta|X)-\beta)_+\right). \end{aligned} \tag{38}$$

Here, as shown in the proof of Theorem 1, the inequality employed in Statement (37) is $\inf_h \mathbb{E}h(Z) \geq \mathbb{E}\inf_h h(Z)$. The inequality used for Kallus' lower bound as stated in Statement (38) is Jensen's inequality applied to the convex $(\cdot)_+$. In both cases the bound only holds with equality when the inequality does.

Given this information, let us first show that Inequality (19) does not always hold. Suppose $\mathrm{Var}(\delta|X) = 0$ and $\mathrm{Var}(X) \neq 0$, then

$$\mathrm{CVaR}_\alpha(\mathbb{E}(\delta|X)) = \mathrm{CVaR}_\alpha(\delta) > \mathbb{E}\,\mathrm{CVaR}_\alpha(\delta|X). \tag{39}$$

where the equality is due to tightness of Kallus' bound for determinisitc $\delta|X$ and the strict inequality is due to $\inf_h \mathbb{E}h(X) \neq \mathbb{E}\inf_h h(X)$ since h is not constant and X is not degenerate.

For the reversed inequality, we have to show that $\mathrm{CVaR}_\alpha(\mathbb{E}(\delta|X)) \geq \mathbb{E}\,\mathrm{CVaR}_\alpha(\delta|X)$ is not true for some (X, δ)–distribution. Assume $\mathrm{Var}(X) = 0$ and $\mathrm{Var}(\delta|X) \neq 0$, then

$$\mathbb{E}\,\mathrm{CVaR}_\alpha(\delta|X) = \mathrm{CVaR}_\alpha(\delta) > \mathrm{CVaR}_\alpha(\mathbb{E}(\delta|X)). \tag{40}$$

where the equality is due to tightness of our bound by degeneracy of the distribution of X and as a result, $\inf_h \mathbb{E}h(X) = \mathbb{E}\inf_h h(X)$. Kallus' bound, however, cannot hold with equality, since $\mathbb{E}(\delta|X) \neq \delta$.

□

5.3 Details Linear Gaussian Generating Equations

Let us make the overall setting of Example 1 more precise. The linear Gaussian generating equations for the potential outcomes $Y^{(1)}, Y^{(0)}$ are given as:

$$Y^{(j)} = \alpha_j + X\beta_j + \epsilon_j,\ j = 0, 1. \tag{41}$$

Here, we assume $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and independence of ϵ_0, ϵ_1. The noise terms are jointly normally distributed with the following structure:

$$\begin{pmatrix}\epsilon_0\\ \epsilon_1\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_0^2 & \rho\sigma_0\sigma_1\\ \rho\sigma_0\sigma_1 & \sigma_1^2\end{pmatrix}\right).$$

By our linearity assumption, δ looks as follows:

$$\delta \coloneqq Y^{(1)} - Y^{(0)} = \alpha_1 - \alpha_0 + (\beta_1 - \beta_0)X + \epsilon_1 - \epsilon_0. \tag{42}$$

The quantities $\delta, \delta|X$ and $\tau(X)$ can be easily derived from δ. Note, that they are all linear combinations of Gaussian random variables and therefore themselves have a Gaussian distribution. These Gaussians are determined by their means and variances. Define $\mu_\delta := \alpha_1 - \alpha_0 + (\beta_1 - \beta_0)\mu_X$,

$$\begin{aligned}\mathbb{E}(\delta) &= \mathbb{E}(\tau(X)) = \mu_\delta\\ \mathbb{E}(\delta|X) &= \alpha_1 - \alpha_0 + (\beta_1 - \beta_0)X\\ \mathrm{Var}(\delta) &= \underbrace{(\beta_1 - \beta_0)^2\sigma_X^2}_{\mathrm{Var}(\tau(X))} + \underbrace{\sigma_0^2 + \sigma_1^2 - 2\rho\sigma_0\sigma_1}_{\mathrm{Var}(\delta|X)}.\end{aligned}$$

Given these means and variances and therefore the distributions of $\delta, \delta|X$ and $\tau(X)$, we can compute the $\mathrm{CVaR}_\alpha(\delta), \mathbb{E}\,\mathrm{CVaR}_\alpha(\delta|X)$ and $\mathrm{CVaR}_\alpha(\tau(X))$ using the closed–form expression for the CVaR of a Gaussian random variable introduced in Equation (15). The resulting quantities are the ones displayed in Equations (16).

When it comes to proving the validity of the combined bound stated in Remark 4, all that needs to be shown is that $\mathrm{CVaR}_\alpha(\delta) \geq \mathrm{Comb}(\delta)$. Under close inspection, we see that both only differ in the scaling factor of $\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}$. It remains to show that

$$\sqrt{(\beta_1 - \beta_0)^2\sigma_X^2 + \sigma_1^2 + \sigma_0^2 - 2\sigma_0\sigma_1} \overset{!}{\leq} \sqrt{(\beta_1 - \beta_0)^2\sigma_X^2 + \sigma_1^2 + \sigma_0^2 - 2\rho\sigma_0\sigma_1}.$$

This reduces to $2\rho\sigma_0\sigma_1 \leq 2\sigma_0\sigma_1$. Since, $\sigma_0, \sigma_1 \geq 0$, this is maximized and holds with equality when $\rho = 1$. Thus, $\mathrm{Comb}(\delta)$ is indeed a valid lower bound.

Remark 5. *Due to $\sqrt{\cdot}$ being a concave function, the linear Gaussian example already shows that $\mathrm{Comb}(\delta)$ is not generalized by $\mathrm{CVaR}_\alpha(\mathrm{CVaR}_\alpha(\delta|X))$, which might appear to be an interesting way of combining Kallus' and our approach at first sight. However, it is thus not a lower bound of $\mathrm{CVaR}_\alpha(\delta)$.*

5.4 Details Estimation and Inference

In addition to the consistency results provided in Section 2.2, Zwingmann and Holzmann [2016] prove that the described estimator of the CVaR (denoted by $\widehat{\mathbb{E}}_X\widehat{\mathrm{CVaR}_\alpha}$) has a Gaussian limiting distribution. More concretely, and adapting their result to our situation, we get

$$\sqrt{n}\Big(\widehat{\mathbb{E}}_X\widehat{\mathrm{CVaR}_\alpha} - \mathbb{E}_X\,\mathrm{CVaR}_\alpha(Y \mid X = x, A = a)\Big) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{n}\,\mathrm{Var}\left(\mathbb{1}_{Y \geq q_\alpha}\frac{q_\alpha - Y}{1-\alpha}\right)\right). \tag{43}$$

Note that the variance in Statement (43) can be estimated using the plug–in principle. However, our bounds are composed of two such terms each (one for $A = 1$, one for $A = 0$). As an example, remember that the upper bound is given by

$$\psi_{ub} = \underbrace{\mathbb{E}\,\mathrm{CVaR}_\alpha(Y^{(1)}|X)}_{(1)} + \underbrace{\mathbb{E}\,\mathrm{CVaR}_\alpha(-Y^{(0)}|X)}_{(2)}. \tag{44}$$

If one is willing to impose that the limiting distributions of (1) and (2) are independent, then we have asymptotic normality of the estimator of ψ_{ub} (the same holds for ψ_{lb}). This follows from basic properties of Gaussian random variables.

Further, note that the same ideas can be employed to estimate Kallus‘ bound using the distributional random forest, see Cevid et al. [2022] for more details on the estimation of the CATE. Then, for the estimation of the CVaR, the same ideas previously described in Section 2.2 apply again. The condition for consistency of the CVaR is again the consistency of the estimation of the CATE, which is governed in Cevid et al. [2022].

References

Carlo Acerbi and Dirk Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26:1487–1503, 2002. ISSN 03784266. doi: 10.1016/S0378-4266(02)00283-2.

Domagoj Cevid, Loris Michel, Jeffrey Näf, Peter Bühlmann, and Nicolai Meinshausen. Distributional random forests: Heterogeneity adjustment and multivariate distributional regression. *Journal of Machine Learning Research*, 23, 2022. ISSN 15337928.

Philip J. Cross and Charles F. Manski. Regressions, short and long, 2002. ISSN 00129682.

Paul Embrechts and Ruodu Wang. Seven proofs for the subadditivity of expected shortfall. *Dependence Modeling*, 3(1):000010151520150009, 2015. doi: doi:10.1515/demo-2015-0009.

Yanqin Fan and Sang Soo Park. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26, 2010. ISSN 02664666. doi: 10.1017/S0266466609990168.

Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81 (396):945–960, 1986. doi: 10.1080/01621459.1986.10478354.

Nathan Kallus. What's the harm? sharp bounds on the fraction negatively affected by treatment. *Proceedings of the 36th Conference on Neural Information Processing Systems*, 2022.

Nathan Kallus. Treatment effect risk: Bounds and inference. *Management Science*, 69, 2023. ISSN 15265501. doi: 10.1287/mnsc.2023.4819.

Shigeo Kusuoka. *On law invariant coherent risk measures*, chapter 4, pages 83–95. Springer, Tokyo, 2001. doi: 10.1007/978-4-431-67891-5_4.

Emanuel Nussli, Simon Hediger, Meta-Lina Spohn, and Marloes Maathuis. The effect of a strict facial-mask policy on the spread of covid–19 in switzerland during the early phase of the pandemic. *Accepted*, 2024.

R. Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26, 2002. ISSN 03784266. doi: 10.1016/S0378-4266(02)00271-6.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 1974. ISSN 00220663. doi: 10.1037/h0037350.

Donald B. Rubin. Statistics and causal inference: Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, 81, 1986. ISSN 01621459. doi: 10.2307/2289065.

Tobias Zwingmann and Hajo Holzmann. Asymptotics for the expected shortfall. *Arxiv preprint*, 2016.