TAIL-RISK CROSS-ENTROPY FOR ROBUST GENERATIVE DECISION-FOCUSED LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Robust generative *decision-focused learning* (*DFL*) seeks to learn predictive distributions that induce reliable decisions under adverse outcome uncertainty. Existing *Conditional Value-at-Risk* (*CVaR*)-based methods, however, rely on estimated CVaR values or CVaR-optimal decisions as supervision, which depend on the conditional tail distribution and are not directly available from standard single-outcome data. We propose *Tail-Risk Cross-Entropy* (*TRCE*), a decision-focused surrogate that bypasses robust-supervision reconstruction and enables CVaR-aware generative learning directly from single-outcome observations. Our key insight is that, although CVaR is a distribution-level risk measure, the ideal CVaR-weighted objective admits an equivalent sample-wise representation through the Rockafellar–Uryasev (RU) formulation. TRCE learns the corresponding conditional Value-at-Risk (VaR) threshold from observed input–outcome pairs and uses it to construct sample-wise tail-risk signals, eliminating the need for explicit ground-truth CVaR estimates or CVaR-optimal decision labels. We establish finite-sample guarantees connecting model approximation, threshold-estimation, and generalization errors to downstream CVaR decision regret. Experiments under both light- and heavy-tailed uncertainty demonstrate consistent improvements over deterministic and generative DFL baselines.

1 INTRODUCTION

Decision-focused learning (*DFL*) trains predictive models to optimize downstream decision quality rather than predictive accuracy alone by incorporating the underlying optimization problem into the learning objective (Donti et al., 2017; Elmachtoub & Grigas, 2022; Mandi et al., 2020; Wilder et al., 2019). DFL has been applied to a broad range of domains, including supply chain management, transportation, finance, energy systems, healthcare, and recommendation systems (Bengio et al., 2021; Davis et al., 2024; Khorshidi et al., 2024; Mandi et al., 2024; Capitaine et al., 2026; Mandi et al., 2025). Existing approaches include *deterministic* DFL, which learns point predictions for downstream optimization (Mulamba et al., 2021; Mandi et al., 2022; Shah et al., 2022), and *generative* DFL, which learns conditional outcome distributions to capture uncertainty (Donti et al., 2017; Zhao et al., 2026). Most existing methods, however, focus on expected decision performance and do not explicitly account for adverse tail outcomes (Ben-Tal et al., 2009). This is particularly limiting in risk-sensitive applications such as currency portfolio management (Fonseca et al., 2011) and power system optimization (Liang et al., 2024), where rare but costly outcomes can substantially affect decision quality. While robust optimization addresses uncertainty through worst-case optimization (Ben-Tal et al., 2009), it can be overly conservative by focusing exclusively on the most adverse realization.

Conditional Value-at-Risk (*CVaR*) (Rockafellar & Uryasev, 2000; 2002) provides a principled alternative by optimizing expected loss over adverse tail outcomes. Incorporating CVaR into generative DFL, however, creates a fundamental *supervision gap*. CVaR is a conditional distribution-level quantity, whereas standard supervised datasets typically provide only one observed outcome for each input. Consequently, neither the conditional CVaR costs nor the CVaR-optimal decisions required by robust decision-focused objectives are directly available from standard training data.

Recent Gen-DFL (Wang et al., 2026) addresses this gap by reconstructing the missing robust decisions from a learned conditional distribution and using them as supervision in a contrastive loss. It first

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

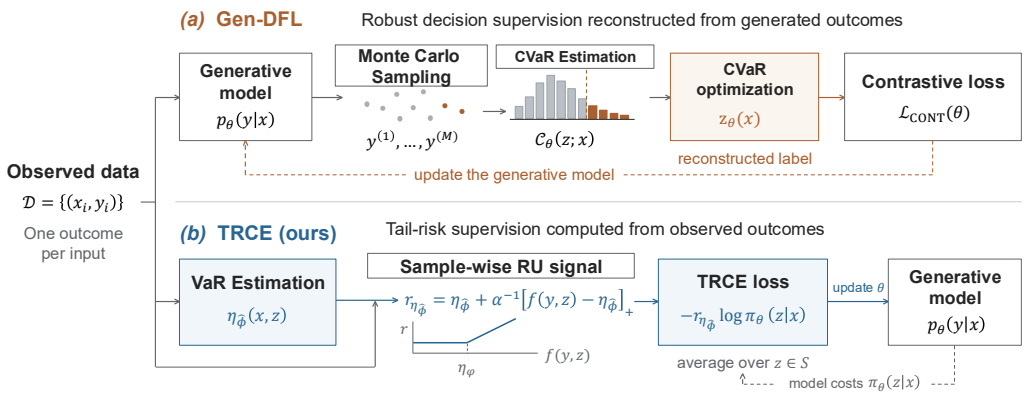

Figure 1: **From robust-supervision reconstruction to sample-wise tail-risk elicitation.** Gen-DFL constructs decision-focused supervision indirectly by generating multiple outcomes from the current model, estimating tail risk CVaR, and solving a sample-based CVaR problem to reconstruct a robust decision label. TRCE instead learns the conditional VaR threshold directly from observed input–outcome pairs and uses the Rockafellar–Uryasev (RU) representation to construct a sample-wise tail-risk signal whose conditional expectation recovers the corresponding true CVaR cost. TRCE therefore removes model-dependent tail-risk estimation and robust-decision reconstruction from the supervision pathway, enabling CVaR-aligned generative learning from single-outcome observations without explicit CVaR estimates or CVaR-optimal decision labels.

draws multiple outcomes from the generative model, estimates CVaR costs from the generated samples, and solves the resulting CVaR optimization problem to infer a robust decision for decision-focused training. While natural, this pipeline requires solving a difficult intermediate estimation problem: the learned distribution must be sufficiently accurate precisely in the low-probability tail that determines CVaR. Errors from distribution approximation and finite-sample tail estimation can therefore propagate into the reconstructed robust decision that subsequently serves as supervision, while repeated tail-risk estimation and robust optimization further increase training cost.

In this paper, we take a different approach. As illustrated in Figure 1, rather than reconstructing robust decision supervision through ground-truth CVaR estimation and optimization for each training input, TRCE directly constructs sample-wise tail-risk supervision from the observed outcomes. Our key observation is that, although true CVaR itself cannot be estimated from a single observed outcome, its role in decision-focused learning can be captured through an expectation of sample-wise tail-risk quantities. Based on this insight, we propose *Tail-Risk Cross-Entropy* (*TRCE*), a decision-focused surrogate for CVaR-aware generative learning from standard single-outcome observations. We begin by constructing an ideal CVaR-weighted cross-entropy objective whose population minimizer preserves the ordering of true CVaR costs across candidate decisions, thereby preserving the downstream robust decision. The remaining challenge is that these CVaR weights are themselves unavailable from individual observations. We resolve this using the Rockafellar–Uryasev (RU) representation of CVaR (Rockafellar & Uryasev, 2000; 2002), which expresses each CVaR cost as the expectation of a tail-risk quantity evaluated at an appropriate *Value-at-Risk* (*VaR*) threshold. We learn this threshold directly from observed input–outcome pairs via quantile regression and use it to construct a sample-wise tail-risk signal whose conditional expectation approximates the corresponding true CVaR cost. This allows the ideal CVaR-weighted objective to be approximated by an expected sample-wise TRCE loss over standard supervised data.

The resulting TRCE objective can therefore be optimized from ordinary input–outcome pairs without explicit ground-truth CVaR estimates or CVaR-optimal decision labels. Rather than reconstructing an entire robust decision through an estimated conditional distribution, TRCE reduces the auxiliary learning problem to estimating a conditional quantile function directly from observed data. This removes repeated robust-decision reconstruction for each training input from the supervision pathway during training. Crucially, this reduction does not weaken the downstream objective: at the population level, the sample-wise TRCE objective recovers the ideal CVaR-weighted criterion, whose optimal generative model induces the desired robust decision. We further establish finite-sample guarantees connecting model approximation, VaR-threshold estimation, and generalization errors to downstream CVaR decision regret, and show that the induced decision approaches the Bayes-optimal robust

decision as these errors vanish. TRCE is model-agnostic and can be naturally combined with conditional generative models such as diffusion models.

Our contributions are summarized as follows: *(i)* We propose *Tail-Risk Cross-Entropy* (*TRCE*), a CVaR-aware decision-focused surrogate that transforms a distribution-level robust objective into a sample-wise learning objective through the RU representation and a learnable VaR threshold. This enables robust generative DFL from standard single-outcome data without explicit CVaR estimates or CVaR-optimal decision labels. *(ii)* We establish finite-sample guarantees for TRCE, characterizing how model approximation, threshold-estimation, and generalization errors propagate to downstream CVaR decision regret. *(iii)* We evaluate TRCE on real-world robust decision-making tasks under both light- and heavy-tailed uncertainty, where it consistently improves downstream CVaR decision quality over deterministic and generative DFL baselines.

2 Robust Generative Decision-Focused Learning

Robust Decision-Making under Outcome Uncertainty. We consider decision-making with covariates $X \in \mathcal{X}$, uncertain outcomes $Y \in \mathcal{Y}$, and a decision space $\mathcal{Z}$. The training data consist of i.i.d. samples $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$ from an unknown joint distribution P, with conditional outcome distribution $p(y \mid x)$. Given a downstream cost $f : \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}^+$, standard expectation-based DFL targets $z^*_{\exp}(x) := \arg\min_{z \in \mathcal{Z}} \mathbb{E}_{Y \sim p(\cdot|x)}[f(Y, z)]$. This objective optimizes average performance but does not explicitly account for adverse tail outcomes.

We instead consider *Conditional Value-at-Risk* (*CVaR*), which measures the expected cost within an adverse tail of the conditional cost distribution. For tail probability $\alpha \in (0, 1]$, define

$$\begin{aligned} \mathcal{C}^*(z; x) &:= \mathrm{CVaR}_\alpha[f(Y, z) \mid X = x] \\ &:= \mathbb{E}[f(Y, z) \mid f(Y, z) \geq \mathrm{VaR}_{1-\alpha}(f(Y, z) \mid X = x),\, X = x]\,, \end{aligned}$$

where $\mathrm{VaR}_{1-\alpha}(f(Y, z) \mid X = x) := \inf\{t \in \mathbb{R} : \mathbb{P}(f(Y, z) \leq t \mid X = x) \geq 1 - \alpha\}$. Here, $\alpha = 1$ recovers the conditional expected cost, while smaller α places greater emphasis on adverse outcomes. The Bayes-optimal robust decision is $z^*(x) := \arg\min_{z \in \mathcal{Z}} \mathcal{C}^*(z; x)$.

Robust Generative DFL. Since $p(y \mid x)$ is unknown, generative DFL approximates it with a conditional model $p_\theta(y \mid x)$, such as a conditional diffusion model. Its model-induced CVaR cost is

$$\mathcal{C}_\theta(z; x) := \mathrm{CVaR}_\alpha^{p_\theta}\big(f(Y, z) \mid X = x\big),$$

which induces the robust decision $z_\theta(x) := \arg\min_{z \in \mathcal{Z}} \mathcal{C}_\theta(z; x)$. The robust decision-focused objective is therefore

$$\theta^* := \arg\min_\theta \mathbb{E}_X \left[\mathcal{C}^*(z_\theta(X); X)\right]. \tag{1}$$

Directly optimizing Eq. (1) is challenging. The induced decision $z_\theta(x)$ involves an $\arg\min$, while its evaluation depends on model-induced CVaR costs determined by the tail of the conditional cost distribution. More fundamentally, the true quantities governing the desired decision, $\mathcal{C}^*(z; x)$, is distribution-level objects, whereas standard supervised data provide only one observed outcome for each input. A practical robust DFL objective must therefore be both *decision-aligned* with the true CVaR criterion and *learnable* from single-outcome supervision.

The Supervision Gap. Recent robust generative DFL methods (Wang et al., 2026) address the non-differentiable downstream decision using a CVaR-based contrastive surrogate. A representative objective is $L_{\mathrm{CONT}}(\theta) := \mathbb{E}_X \sum_{z \in \mathcal{S}} [\mathcal{C}_\theta(z; X) - \mathcal{C}_\theta(z^*(X); X)]$, where $\mathcal{S} \subset \mathcal{Z}$ is a candidate decision set and $z^*(X)$ serves as the target robust decision. The difficulty is that $z^*(x)$ is unavailable under single-outcome supervision: constructing it requires comparing the conditional CVaR costs $\mathcal{C}^*(z; x)$ across candidate decisions, which cannot be directly recovered from a single realization Y_i for each X_i. We refer to this mismatch between the CVaR-based supervision required for robust decision-focused learning, such as ground-truth CVaR costs or CVaR-optimal decisions, and the single-outcome observations available in standard training data as the *supervision gap*.

Existing approaches therefore reconstruct this missing supervision through the learned generative model (Wang et al., 2026): multiple synthetic outcomes are sampled to estimate tail-sensitive

costs and infer a robust decision for decision-focused training. This produces *model-dependent supervision*, whose reliability depends on how accurately the current model captures the conditional tail distribution. Moreover, because the reconstructed robust decision evolves with the generative model, the supervision itself changes during training, while repeated sampling, tail-risk estimation, and robust optimization introduce additional computational overhead.

Our goal is to avoid this reconstruction. Rather than explicitly estimating $\mathcal{C}^*(z;x)$ or constructing $z^*(x)$ as supervision, we seek a decision-aligned objective whose stochastic training signal can be computed directly from observed pairs (X_i, Y_i). The next section shows that the Rockafellar–Uryasev representation makes this possible: although CVaR is a distribution-level quantity, the desired CVaR-weighted learning objective admits an equivalent sample-wise representation, leading to *Tail-Risk Cross-Entropy*.

3 Tail-Risk Cross-Entropy for Robust Generative DFL

We now develop *Tail-Risk Cross-Entropy* (*TRCE*), a decision-focused surrogate that learns CVaR-aligned conditional distributions directly from single-outcome supervision. We first construct an ideal objective using the true conditional CVaR costs and then exploit the Rockafellar–Uryasev (RU) representation to derive an equivalent sample-wise formulation. Learning the required Value-at-Risk (VaR) threshold from observed input–outcome pairs yields a practical solution without explicit conditional CVaR estimates or CVaR-optimal decision labels.

Ideal TRCE: Preserving CVaR Cost Ordering. Recall that $\mathcal{C}_\theta(z;x)$ denotes the CVaR cost induced by $p_\theta(y \mid x)$. For a candidate set $\mathcal{S} \subset \mathcal{Z}$, define the *cost-induced decision distribution* $\pi_\theta(z \mid x) := \mathcal{C}_\theta(z;x)/\sum_{z' \in \mathcal{S}} \mathcal{C}_\theta(z';x)$ over decisions $z \in \mathcal{S}$, where $\mathcal{C}_\theta(z;x) > 0$ since $f(y,z) > 0$. If the true CVaR costs were available, we could define the ideal tail-risk cross-entropy objective over a conditional decision distribution π as

$$L^*_{\mathrm{TRCE}}(\pi) := -\mathbb{E}_X\left[\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}} \mathcal{C}^*(z;X)\log \pi(z \mid X)\right]. \tag{2}$$

The following result characterizes the optimal decision distribution induced by this ideal objective.

Theorem 3.1 (Optimal Decision Distribution)**.** *Assume $\mathcal{C}^*(z;x) > 0$ for all $z \in \mathcal{S}$ and almost every x. Then, for each x, the unique minimizer of the conditional objective in Eq. (2) is*

$$\pi^*(z \mid x) = \frac{\mathcal{C}^*(z;x)}{\sum_{z' \in \mathcal{S}} \mathcal{C}^*(z';x)}, \qquad z \in \mathcal{S}. \tag{3}$$

The parameterized objective is obtained by restricting π to the model-induced distribution π_θ, i.e., $L^*_{\mathrm{TRCE}}(\theta) := L^*_{\mathrm{TRCE}}(\pi_\theta)$. If $\pi_\theta(\cdot \mid x) = \pi^*(\cdot \mid x)$, then for some $\lambda(x) > 0$ independent of z, $\mathcal{C}_\theta(z;x) = \lambda(x)\mathcal{C}^*(z;x), \forall z \in \mathcal{S}$, where $\lambda(x) = \sum_{z' \in \mathcal{S}} \mathcal{C}_\theta(z';x)/\sum_{z' \in \mathcal{S}} \mathcal{C}^*(z';x)$.

Thus, for each fixed x, the model-induced costs $\mathcal{C}_\theta(z;x)$ and true CVaR costs $\mathcal{C}^*(z;x)$ have identical ordering over $z \in \mathcal{S}$ and, in particular, the same robust minimizer. Ideal TRCE therefore targets the relative tail-risk structure required for downstream decision-making rather than exact recovery of absolute CVaR values. The obstacle is that the weights $\mathcal{C}^*(z;x)$ in Eq. (2) are the distribution-level quantities unavailable under single-outcome supervision. We next show that these weights need not be explicitly estimated.

Sample-Wise Tail-Risk Elicitation. For any measurable threshold function $\eta : \mathcal{X} \times \mathcal{S} \to \mathbb{R}$, define the sample-wise RU quantity $r_\eta(x,y,z) := \eta(x,z) + \alpha^{-1}(f(y,z) - \eta(x,z))_+$, where $(u)_+ := \max\{u, 0\}$, and let $\eta^*(x,z) := \arg\min_{\eta\in\mathbb{R}} \mathbb{E}\left[r_\eta(X,Y,z) \mid X = x\right]$.

The following result converts the distribution-level supervision in Eq. (2) into an equivalent sample-wise stochastic objective.

Theorem 3.2 (Sample-Wise Reformulation of TRCE)**.** *For every $x \in \mathcal{X}$ and $z \in \mathcal{S}$, we have*

$$\eta^*(x,z) = \mathrm{VaR}_{1-\alpha}\big(f(Y,z) \mid X = x\big) \quad \textit{and} \quad \mathbb{E}\left[r_{\eta^*}(X,Y,z) \mid X = x\right] = \mathcal{C}^*(z;x).$$

Consequently,

$$L^*_{\mathrm{TRCE}}(\theta) = -\mathbb{E}_{X,Y}\left[\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}} r_{\eta^*}(X,Y,z)\log \pi_\theta(z \mid X)\right]. \tag{4}$$

Theorem 3.2 provides the key reduction behind TRCE. It does *not* imply that conditional CVaR itself can be estimated from a single outcome; rather, it shows that the ideal CVaR-weighted population objective admits an equivalent expectation of sample-wise losses. Given the VaR threshold, ordinary i.i.d. pairs (X, Y) therefore provide unbiased stochastic estimates of the ideal TRCE objective, eliminating the need to explicitly reconstruct CVaR values or CVaR-optimal decision labels.

Learning the VaR Threshold. The oracle threshold $\eta^*(x, z)$ is unknown, so we approximate it with $\eta_\phi(x, z)$ and optimize the RU objective $\mathcal{L}_{\mathrm{VaR}}(\phi) := \mathbb{E}_{X,Y}[\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}} r_{\eta_\phi}(X, Y, z)]$. Because this objective is itself an expectation over paired observations, ϕ can be learned directly from the training data:

$$\widehat{\phi} := \arg\min_{\phi} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}} r_{\eta_\phi}(x_i, y_i, z). \tag{5}$$

Practical TRCE. Replacing η^* in Eq. (4) with $\eta_{\widehat{\phi}}$ yields

$$L_{\mathrm{TRCE}}(\theta; \widehat{\phi}) := -\mathbb{E}_{X,Y}\left[\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}} r_{\eta_{\widehat{\phi}}}(X, Y, z)\log \pi_\theta(z \mid X)\right]. \tag{6}$$

Eq. (6) resolves the supervision gap identified in Section 2: its stochastic training signal depends only on observed input–outcome pairs and the learned threshold function, rather than explicit $\mathcal{C}^*(z; x)$ or $z^*(x)$ supervision. The effect of replacing the oracle threshold with $\eta_{\widehat{\phi}}$ is characterized explicitly in the finite-sample analysis of Section 4.

Integration with Conditional Generative Models. TRCE is agnostic to the specific modeling of $p_\theta(y \mid x)$. In our implementation, we combine it with a conditional diffusion model (DM):

$$\widehat{\theta} := \arg\min_{\theta} L_{\mathrm{DM}}(\theta) + \gamma\,\omega(\theta)\,L_{\mathrm{TRCE}}(\theta; \widehat{\phi}), \tag{7}$$

where $L_{\mathrm{DM}}(\theta) := \mathbb{E}_{(x,y)\sim q, t, \epsilon\sim\mathcal{N}(0,I)} \|\epsilon - \epsilon_\theta(y_t, x, t)\|_2^2$ is the standard diffusion loss for preserving distributional fidelity, with $y_t = \sqrt{\bar{\alpha}_t}\,y + \sqrt{1-\bar{\alpha}_t}$, L_{TRCE} provides tail-risk decision alignment, and $\gamma > 0$ controls their trade-off. We use the detached batch-wise gradient-balancing factor $\omega(\theta) := 10^{-\lfloor \log_{10}(\|\nabla_\theta L_{\mathrm{TRCE}}\|_2 / \|\nabla_\theta L_{\mathrm{DM}}\|_2)\rfloor}$ to balance their relative gradient scales. After training, the learned model induces $\widehat{z}(x) := \arg\min_{z\in\mathcal{Z}} \mathcal{C}_{\widehat{\theta}}(z; x)$.

Candidate Decision Set. Evaluating TRCE over the full discrete space $\mathcal{Z}$ can be intractable. Following standard practice (Mandi et al., 2022; Wang et al., 2026), we maintain $\mathcal{S} \subset \mathcal{Z}$, initialized with sample-optimal decisions $\mathcal{S} \leftarrow \{\arg\min_{z\in\mathcal{Z}} f(y_i, z)\}_{i=1}^n$. During training, we periodically augment $\mathcal{S}$ with decisions induced by the current generative model: for x_i, we draw M outcomes from $p_\theta(\cdot \mid x_i)$ and, with probability $p_{\mathrm{solve}} = 0.1$, solve the corresponding sample-based CVaR problem and add the resulting decision to $\mathcal{S}$. Importantly, these model-induced decisions are used only to improve candidate-space coverage; they never serve as target decision labels or define the TRCE supervision. Algorithm 1, provided at the beginning of the appendix, summarizes the complete TRCE training procedure.

4 Statistical Guarantees for TRCE

Theorem 3.2 shows that the ideal TRCE objective in Eq. (2), which uses the true CVaR costs as weights, admits an equivalent sample-wise reformulation in Eq. (4) that does not require explicit CVaR values. Building on this reformulation, we obtain the practical TRCE objective in Eq. (6) by replacing the unknown VaR thresholds with their estimates learned from standard supervised data. In this section, we analyze the downstream CVaR decision regret of the resulting TRCE learner when each training input is paired with only one observed outcome. Our finite-sample analysis isolates three sources of error—model approximation, finite-sample generalization, and VaR-threshold estimation—and quantifies how they propagate to downstream CVaR decision regret. As these errors vanish, the decision induced by the TRCE minimizer approaches the Bayes-optimal robust decision in terms of CVaR regret.

From TRCE Risk to CVaR Decision Regret. We first establish how excess TRCE risk translates into downstream robust decision quality. Let $L_{\mathrm{TRCE}}(\theta;\phi^*)$ denote the population TRCE objective with oracle threshold $\eta_{\phi^*}(x,z)=\eta^*(x,z)$, and let $\widehat{\mathcal{L}}_{\mathrm{TRCE}}(\theta;\widehat{\phi})$ denote its empirical counterpart constructed from observed single-outcome samples and the learned threshold function.

The practical objective in Eq. (7) combines TRCE with a generative-model loss to preserve distributional fidelity. For the theoretical analysis, we represent this role using the oracle conditional KL regularizer $\mathcal{K}(\theta):=\mathbb{E}_X\left[D_{\mathrm{KL}}(p_\theta(\cdot\mid X)\,\|\,p(\cdot\mid X))\right]$. This term is used only to formalize the distributional-fidelity requirement and is not assumed to be available during training. We study

$$\widehat{\mathcal{J}}_n(\theta;\widehat{\phi}):=\widehat{\mathcal{L}}_{\mathrm{TRCE}}(\theta;\widehat{\phi})+\beta\mathcal{K}(\theta),$$

with $\beta>0$, and let $\widehat{\theta}\in\arg\min_{\theta\in\Theta}\widehat{\mathcal{J}}_n(\theta;\widehat{\phi})$.

Before deriving the regret bound, we first identify three sources of error. The generalization error under the oracle threshold is defined as $\varepsilon_g:=\sup_{\theta\in\Theta}\left|\widehat{\mathcal{L}}_{\mathrm{TRCE}}(\theta;\phi^*)-L_{\mathrm{TRCE}}(\theta;\phi^*)\right|$, while the error induced by replacing the oracle threshold with $\eta_{\widehat{\phi}}$ is $\varepsilon_\phi:=\sup_{\theta\in\Theta}\left|\widehat{\mathcal{L}}_{\mathrm{TRCE}}(\theta;\widehat{\phi})-\widehat{\mathcal{L}}_{\mathrm{TRCE}}(\theta;\phi^*)\right|$. Let L^*_{TRCE} denote the minimum population TRCE risk over the unrestricted class of conditional distributions, and define $\overline{\mathcal{J}}(\theta):=L_{\mathrm{TRCE}}(\theta;\phi^*)-L^*_{\mathrm{TRCE}}+\beta\mathcal{K}(\theta)$. The model-class approximation error is $\varepsilon_a:=\inf_{\theta\in\Theta}\overline{\mathcal{J}}(\theta)$. In particular, $\varepsilon_a=0$ whenever Θ contains the true conditional distribution.

Theorem 4.1 (CVaR Decision Regret)**.** *Suppose that the downstream cost is uniformly bounded: $0<f(y,z)\le B$ for all $y\in\mathcal{Y}$ and $z\in\mathcal{Z}$. Then the robust decision $\widehat{z}(x)$ induced by $\widehat{\theta}$ satisfies*

$$\mathbb{E}_X\left[\mathcal{C}^*(\widehat{z}(X);X)-\mathcal{C}^*(z^*(X);X)\right]\le c_1\left(\varepsilon_a+2\varepsilon_g+2\varepsilon_\phi\right)^{1/2},$$

where $c_1>0$ depends only on problem constants, including B, α, and β.

Theorem 4.1 is the main calibration result: controlling TRCE approximation, generalization, and threshold-estimation errors is sufficient to control downstream CVaR decision regret.

VaR-Threshold Estimation from Single-Outcome Data. The sample-wise TRCE representation depends on the oracle threshold $\eta^*(x,z)$, the conditional $(1-\alpha)$-quantile of the cost distribution. We next quantify the error incurred by replacing it with $\eta_{\widehat{\phi}}(x,z)$. Let $F_{x,z}$ denote the conditional CDF of $f(Y,z)$ given $X=x$.

Assumption 4.2 (Local Quantile Identifiability)**.** *There exist constants $\kappa>0$ and $r_0>0$ such that, for every (x,z) and every t satisfying $|t-\eta^*(x,z)|\le r_0$, we have $|F_{x,z}(t)-(1-\alpha)|\ge\kappa\,|t-\eta^*(x,z)|$.*

This is a standard local identifiability condition for conditional quantiles; for example, it holds when the conditional density of $f(Y,z)$ is bounded away from zero near its $(1-\alpha)$-quantile.

Define the RU function class $\mathcal{G}_{\mathrm{RU}}:=\{(x,y)\mapsto\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}}r_{\eta_\phi}(x,y,z):\phi\in\Phi\}$, and let $\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}})$ denote its expected Rademacher complexity. We further define the threshold approximation error $\varepsilon_a^\Phi:=\inf_{\phi\in\Phi}\mathcal{L}_{\mathrm{VaR}}(\phi)-\mathcal{L}_{\mathrm{VaR}}(\phi^*)$, where $\mathcal{L}_{\mathrm{VaR}}(\phi)$ is the population RU risk associated with η_ϕ.

Theorem 4.3 (Threshold Estimation Error)**.** *Suppose that $0<f(y,z)\le B$ and Assumption 4.2 holds. Then, with probability at least $1-\delta$,*

$$\varepsilon_\phi\le c_2\left[\varepsilon_a^\Phi+4\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}})+\frac{2B}{\alpha}\sqrt{\frac{\log(2/\delta)}{2n}}\right]^{1/2},$$

where $c_2>0$ depends only on the boundedness and local-identifiability constants and on $|\mathcal{S}|$.

Theorem 4.3 bridges the oracle sample-wise formulation and practical TRCE. Rather than estimating CVaR values directly, TRCE only requires learning the corresponding conditional quantile. Its threshold-estimation error is explicit and vanishes as the threshold class becomes sufficiently expressive and its statistical complexity decreases with sample size.

Generalization from Single-Outcome Samples. We next control the discrepancy between empirical and population TRCE risk. For each $z \in \mathcal{S}$, define $\mathcal{G}_z := \{x \mapsto \mathcal{C}_\theta(z; x) : \theta \in \Theta\}$, and let $\mathfrak{R}_n(\mathcal{G}_z)$ denote its expected Rademacher complexity.

Proposition 4.4 (TRCE Generalization)**.** *Suppose that $0 < b \le f(y, z) \le B$ for all $y \in \mathcal{Y}$ and $z \in \mathcal{S}$. Then, with probability at least $1 - \delta$,*

$$\varepsilon_g \le \frac{4B}{\alpha b} \sup_{z \in \mathcal{S}} \mathfrak{R}_n(\mathcal{G}_z) + \frac{B}{\alpha} \log\left(\frac{|\mathcal{S}|B}{b}\right) \sqrt{\frac{\log(2/\delta)}{2n}}.$$

Proposition 4.4 establishes uniform convergence of TRCE from standard i.i.d. pairs (X_i, Y_i). No repeated realizations are required for a fixed input, and the generalization error vanishes whenever the complexity of the induced CVaR function class is appropriately controlled.

Consistency of Robust Decision Learning. Combining Theorem 4.1, Theorem 4.3, and Proposition 4.4 yields the following consistency result.

Corollary 4.5 (Decision Consistency of TRCE)**.** *Let Assumption 4.2 hold. If, as $n \to \infty$, $\varepsilon_a \to 0$, $\varepsilon_a^\Phi \to 0$, $\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}}) \to 0$, $\sup_{z \in \mathcal{S}} \mathfrak{R}_n(\mathcal{G}_z) \to 0$, then $\mathbb{E}_X\left[\mathcal{C}^*(\widehat{z}(X); X) - \mathcal{C}^*(z^*(X); X)\right] \to 0$.*

Corollary 4.5 completes the statistical argument. Although CVaR is a conditional distribution-level functional, its decision-relevant learning objective can be optimized from ordinary single-outcome observations through TRCE. Threshold estimation and finite-sample generalization contribute explicit, vanishing error terms, and controlling them together with model approximation is sufficient for convergence to the Bayes-optimal CVaR decision—without observing true CVaR values or CVaR-optimal decisions during training.

5 Experiments

We evaluate whether TRCE can learn CVaR-optimal decisions from single-outcome supervision and whether its advantage persists across different tail regimes. We focus on three questions: (i) whether TRCE improves downstream CVaR decision quality over deterministic and generative DFL baselines; (ii) whether its advantage remains robust as the objective becomes increasingly tail-sensitive; and (iii) whether the gains arise from the proposed tail-risk alignment rather than a particular generative backbone or hyperparameter choice.

Datasets and Stochastic Construction. We evaluate on three real-world decision-making benchmarks: `Portfolio` optimization (Wang et al., 2020), `Power` scheduling (Donti et al., 2017), and `Web` advertising (Wilder et al., 2019). Detailed formulations are provided in Appendix B.1.

Each input x_i is associated with a ground-truth outcome y_i^*, from which we construct stochastic outcomes as $y_i = y_i^* + \varepsilon_i$, where ε_i is noise perturbation. Crucially, for every training input we draw only *one* perturbation and retain only $y_i = y_i^* + \varepsilon_i$ as supervision, yielding $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$ with exactly one observed outcome per input. This matches the single-outcome setting studied in Sections 2–4. Let $\sigma_j := \mathrm{Std}(\{y_{i,j}^*\}_{i=1}^n)$ and $\sigma = (\sigma_1, \ldots, \sigma_K)$. For light-tailed uncertainty, we use $y_i = y_i^* + \rho \cdot \sigma \odot \epsilon_i$, $\epsilon_i \sim \mathcal{N}(0, I)$, with $\rho \in \{0.1, 0.5, 1\}$. For heavy-tailed uncertainty, we use $y_i = y_i^* + \sqrt{(\nu-2)/\nu} \cdot \sigma \odot \epsilon_i$, $\epsilon_i \sim t_\nu$, with $\nu \in \{3, 5, 10\}$. The variance normalization isolates tail shape, with smaller ν corresponding to heavier tails.

Evaluation Protocol and Baselines. At test time, each method observes only x and outputs $\widehat{z}(x)$. We measure CVaR decision regret, $\mathrm{Regret}(\widehat{z}) := \frac{1}{n_t} \sum_{x_i \in D_t} [\mathcal{C}^*(\widehat{z}(x_i); x_i) - \mathcal{C}^*(z^*(x_i); x_i)]$, where D_t is the test dataset and lower is better. Ground-truth CVaR quantities are used *only for evaluation*: for each test input, we draw multiple independent outcomes from the same perturbation distribution to accurately approximate $\mathcal{C}^*(z; x_i)$ and $z^*(x_i)$. Repeated outcomes are never available during training.

We compare against pairwise and listwise ranking (Mandi et al., 2022), NCE (Mulamba et al., 2021), predict-then-optimize (PTO), SPO+ (Elmachtoub & Grigas, 2022), Diff-DRO (Ma et al., 2024), Gen-DFL (Wang et al., 2026), and Diff-DFL (Zhao et al., 2026). All methods use identical train/test splits and the same evaluation protocol.

Main Results. Tables 1–3 show that TRCE achieves the lowest CVaR decision regret in all 18 dataset–uncertainty configurations at $\alpha = 0.1$. Its advantage holds across Gaussian and Student-t uncertainty and across three substantially different downstream problems, indicating consistent robustness to both task structure and tail regime.

Table 1: CVaR decision regret ($\times 100$) on `Portfolio` under light- and heavy-tailed uncertainty.

	Gaussian			Student-t		
Method	$\rho = 0.1$	$\rho = 0.5$	$\rho = 1$	$\nu = 3$	$\nu = 5$	$\nu = 10$
Pairwise	2.52±0.12	2.23±0.11	1.98±0.17	2.01±0.11	1.89±0.10	2.04±0.14
Listwise	2.61±0.24	2.30±0.17	1.95±0.21	1.95±0.12	1.84±0.14	2.10±0.19
NCE	2.51±0.13	2.29±0.14	1.93±0.14	1.98±0.13	1.84±0.13	2.01±0.24
PTO	4.19±0.27	3.64±0.18	2.62±0.17	2.87±0.22	2.67±0.19	2.77±0.20
SPO+	2.54±0.13	2.20±0.22	1.74±0.06	1.79±0.06	1.77±0.08	1.79±0.08
Diff-DRO	2.71±0.22	1.97±0.20	1.79±0.13	1.89±0.16	1.86±0.16	2.00±0.10
Gen-DFL	1.92±0.25	2.10±0.12	1.41±0.23	1.89±0.10	1.80±0.12	2.06±0.20
Diff-DFL	3.05±0.38	2.70±0.29	2.76±0.35	2.90±0.25	2.69±0.33	2.83±0.23
TRCE (Ours)	**1.31±0.22**	**0.91±0.13**	**1.28±0.15**	**1.38±0.08**	**1.49±0.14**	**1.47±0.40**

Table 2: CVaR decision regret on `Power` under light- and heavy-tailed uncertainty.

	Gaussian			Student-t		
Method	$\rho = 0.1$	$\rho = 0.5$	$\rho = 1$	$\nu = 3$	$\nu = 5$	$\nu = 10$
Pairwise	59.09±13.76	69.79±13.21	55.42±2.040	92.99±21.65	83.44±10.69	72.29±18.46
Listwise	40.01±11.83	61.81±8.350	56.55±9.430	87.62±5.880	78.18±16.55	69.79±9.740
NCE	29.46±15.19	74.06±14.45	63.73±11.83	100.2±18.08	87.03±13.94	78.13±21.09
PTO	80.50±17.58	93.31±24.53	98.37±11.29	128.2±15.97	123.7±18.68	110.8±18.56
SPO+	25.51±7.890	56.98±10.41	78.55±9.490	97.80±19.06	86.92±12.63	72.91±14.91
Diff-DRO	40.22±11.04	38.55±12.23	39.63±8.240	72.96±16.78	69.77±14.12	61.71±13.32
Gen-DFL	28.32±2.410	26.64±3.240	32.02±5.350	48.16±6.020	38.95±6.210	31.25±6.840
Diff-DFL	60.31±10.52	78.40±9.430	76.15±8.280	88.91±13.22	81.75±10.28	75.77±9.190
TRCE (Ours)	**12.19±2.590**	**7.050±1.650**	**10.96±4.600**	**24.07±0.990**	**19.09±1.700**	**13.27±4.780**

Table 3: CVaR decision regret ($\times 100$) on `Web` under light- and heavy-tailed uncertainty.

	Gaussian			Student-t		
Method	$\rho = 0.1$	$\rho = 0.5$	$\rho = 1$	$\nu = 3$	$\nu = 5$	$\nu = 10$
Pairwise	0.63±0.23	2.23±0.29	4.10±0.40	4.06±0.33	4.44±0.48	4.34±0.60
Listwise	0.61±0.35	1.83±0.20	3.69±0.38	3.56±0.31	4.12±0.19	3.84±0.24
NCE	4.51±0.10	4.85±0.17	5.54±0.24	5.49±0.30	5.87±0.47	5.38±0.55
PTO	3.21±0.73	4.23±0.52	6.59±0.71	7.22±0.88	7.24±0.87	7.20±0.68
SPO+	1.04±0.28	2.61±0.61	5.16±0.85	5.11±0.42	5.24±0.47	5.02±0.50
Diff-DRO	0.89±0.31	1.52±0.53	2.65±0.34	4.01±0.19	2.39±0.27	2.61±0.30
Gen-DFL	0.55±0.10	0.99±0.14	1.19±0.26	3.60±0.12	1.89±0.19	1.81±0.22
Diff-DFL	4.21±0.81	4.45±0.62	5.01±0.78	5.20±0.31	4.96±0.91	5.41±0.80
TRCE (Ours)	**0.20±0.10**	**0.42±0.11**	**0.85±0.05**	**2.69±0.73**	**1.53±0.16**	**1.13±0.14**

The comparison with generative baselines is particularly informative. Diff-DFL models conditional uncertainty but does not explicitly align the learned distribution with the downstream CVaR objective. Gen-DFL incorporates tail risk by reconstructing robust decisions from an estimated conditional distribution. TRCE instead constructs its decision-focused signal directly from observed input–outcome pairs through the learned RU threshold. Across the reported settings, TRCE reduces regret by up to 73.5% relative to Gen-DFL and 95.2% relative to Diff-DFL. These results support the central claim of TRCE: distribution-level tail-risk decisions can be learned effectively from single-outcome supervision without reconstructing CVaR-optimal decision labels.

Results under Different Tail Probabilities α. We vary the CVaR tail probability α to examine increasing levels of risk sensitivity on Portfolio. Figure 2 shows that TRCE consistently achieves lower decision regret across the tested values. Its advantage persists as α decreases and the objective concentrates on increasingly adverse outcomes, indicating that sample-wise tail-risk alignment remains effective in more risk-sensitive regimes.

Effect of Decision-Focused Training. The parameter γ in Eq. (7) controls the strength of TRCE relative to distributional training. As shown in Figure 3, removing TRCE ($\gamma = 0$) substantially degrades downstream decision quality on Portfolio, whereas nonzero TRCE weights consistently reduce CVaR regret. Performance remains stable over a broad range of γ, indicating that the gains

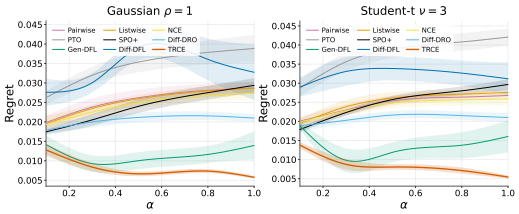

Figure 2: Effect of CVaR tail probability α on Portfolio.

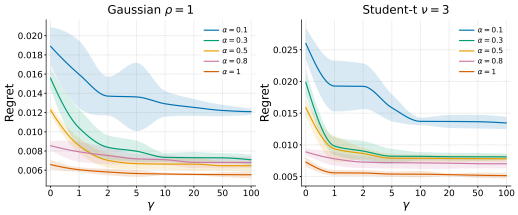

Figure 3: Sensitivity to the parameter γ on Portfolio.

do not depend on fine-tuning a narrow operating point. These results show that distributional fidelity alone is insufficient and that explicit tail-risk decision alignment contributes directly to the downstream improvement. Full hyperparameter settings are reported in Appendix B.3.

Generative-Backbone Ablation. We next test whether the gains of TRCE arise from its objective rather than from the choice of generative model. We compare TRCE and the Gen-DFL contrastive objective using both conditional flows and diffusion models on Portfolio, swapping the generative backbones while retaining the respective decision-focused objectives. As shown in Figure 4, TRCE achieves lower CVaR regret under both backbones. This controlled comparison indicates that the improvement is driven primarily by the tail-risk supervision mechanism rather than a particular generative architecture.

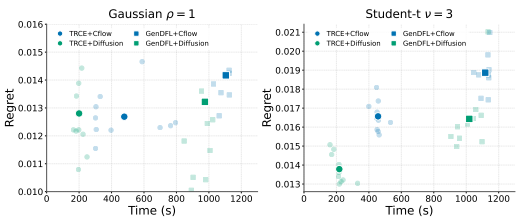

Figure 4: Performance of TRCE and Gen-DFL on the Portfolio task with conditional flow and diffusion backbones.

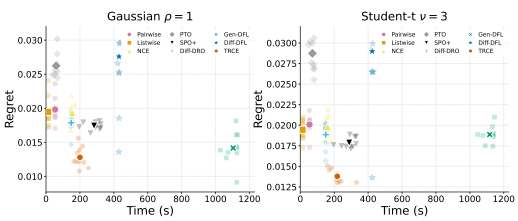

Figure 5: Runtime–regret trade-off for Portfolio.

Runtime–Regret Trade-off. Figure 5 compares computational cost and robust decision quality on Portfolio. Among generative methods, TRCE achieves both lower runtime and lower CVaR regret than Gen-DFL and Diff-DFL. This advantage is consistent with its shorter supervision pathway: TRCE does not require repeated robust-decision optimization to construct its decision-focused training signal. Although generative training remains more expensive than deterministic baselines, TRCE provides a favorable trade-off between computational cost and robust decision quality.

6 Conclusion

We study robust generative decision-focused learning under *single-outcome supervision*, where the distribution-level tail-risk quantities required for CVaR-aware decision learning are not directly observed. We propose *Tail-Risk Cross-Entropy* (*TRCE*), which uses the Rockafellar–Uryasev representation to replace robust-supervision reconstruction with an equivalent sample-wise learning objective whose population expectation recovers the desired CVaR-weighted criterion. By learning the conditional VaR threshold directly from observed input–outcome pairs, TRCE enables decision-aligned generative learning without explicit CVaR estimates or CVaR-optimal decision labels. We further establish finite-sample guarantees connecting model approximation, VaR-threshold estimation, and generalization errors to downstream CVaR decision regret. Experiments across three robust decision-making tasks under both light- and heavy-tailed uncertainty consistently improve CVaR decision quality over deterministic and generative DFL baselines. Overall, our results show that distribution-level robust decisions can be learned from standard single-outcome data by eliciting sample-wise, decision-relevant tail-risk supervision rather than reconstructing robust decisions through an estimated conditional distribution.

AI USE STATEMENT

In this work, we used generative AI tools to assist with intermediate mathematical derivations and provide ingredients for proving mathematical claims, assist with the writing and verification of proofs, provide feedback on research methodology and experimental design, and assist with the interpretation and presentation of experimental results. We did not use generative AI tools to generate synthetic datasets, develop the theoretical models or conceptual framework of the proposed method, formulate mathematical claims, propose or refine research hypotheses, implement the proposed method, assist with translation, or clean or reformat datasets. Qualitative and thematic data analysis was not applicable to this work.

Additionally, we used generative AI tools for brainstorming, summarizing and discussing relevant literature, suggesting experimental parameters, creating or editing auxiliary software code, drafting and restructuring parts of the manuscript, improving the clarity and readability of the writing, suggesting paper structure and presentation, and assisting with reference formatting.

All AI-assisted content was critically reviewed and revised by the authors. In particular, all AI-assisted mathematical derivations and proofs were independently checked by the authors, experimental choices and interpretations were validated against the actual experimental results, and AI-assisted code was reviewed and tested before use. The research questions, theoretical and conceptual framework, mathematical claims, final methodological choices, theoretical results, and experimental conclusions were determined and verified by the authors. We take responsibility for the final content of this work, including text, claims, code, and other artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work develops a general methodology for robust decision-focused learning under outcome uncertainty and does not involve new data collection, human-subject studies, or the release of personally identifiable information. The proposed framework may be applicable to decision-making problems in potentially high-stakes domains. While optimizing tail risk can improve robustness to adverse outcomes, it does not by itself guarantee fairness, safety, or compliance with application-specific requirements. In practice, the choice of cost function, CVaR risk level, data, and downstream constraints may substantially affect the resulting decisions and should therefore be determined with appropriate domain knowledge and human oversight. We encourage practitioners to evaluate potential distributional biases, model misspecification, and application-specific risks before deployment.

REPRODUCIBILITY STATEMENT

We provide the methodological, theoretical, and experimental details needed to reproduce our results in the appendix and supplementary materials. Algorithm 1 presents the complete TRCE training procedure, including candidate-set construction, VaR learning, and decision-focused fine-tuning. Section A provides complete proofs and additional theoretical results, including the sample-wise reformulation, CVaR decision-regret and finite-sample generalization guarantees, conditional VaR estimation and calibration analysis, and the resulting end-to-end decision-consistency guarantee. Section B provides the experimental details required for reproduction, including the decision tasks and robust formulations, baseline methods, training configurations, and differentiation through reparameterized generative sampling. We additionally provide the implementation code in the supplementary materials.

REFERENCES

Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, 2009.

Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. Machine learning for combinatorial optimization: A methodological tour d'horizon. *European Journal of Operational Research*, 290(2):405–421, 2021.

Aymeric Capitaine, Maxime Haddouche, Eric Moulines, Michael I Jordan, Etienne Boursier, and Alain Durmus. Online decision-focused learning. In *The Fourteenth International Conference on Learning Representations*, 2026.

Andrew M. Davis, Shawn Mankad, Charles J. Corbett, and Elena Katok. OM Forum—The best of both worlds: Machine learning and behavioral science in operations management. *Manufacturing & Service Operations Management*, 26(5):1605–1621, September 2024.

Priya L. Donti, Brandon Amos, and J. Zico Kolter. Task-based end-to-end model learning in stochastic optimization. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Proceedings of the 31st International Conference on Neural Information Processing Systems*, volume 30 of *NeurIPS'17*, pp. 5490–5500, Red Hook, NY, USA, 2017. Curran Associates, Inc.

Adam N. Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1): 9–26, January 2022.

Raquel Fonseca, Steve Zymler, Wolfram Wiesemann, and Berc Rustem. Robust optimization of currency portfolios. *The Journal of Computational Finance*, 15(1):3–30, 2011.

Hadi A. Khorshidi, Marzieh Soltanolkottabi, Richard Allmendinger, and Uwe Aickelin. Editorial: Special issue on operations research and machine learning. *INFOR: Information Systems and Operational Research*, 62(2):183–185, 2024.

Jinhao Liang, Wenqian Jiang, Chenbei Lu, and Chenye Wu. Joint chance-constrained unit commitment: Statistically feasible robust optimization with learning-to-optimize acceleration. *IEEE Transactions on Power Systems*, 39(5):6508–6521, 2024.

Xutao Ma, Chao Ning, and Wenli Du. Differentiable distributionally robust optimization layers. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 33880–33901. PMLR, 21–27 Jul 2024.

Jayanta Mandi, Emir Demirović, Peter J. Stuckey, and Tias Guns. Smart predict-and-optimize for hard combinatorial optimization problems. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, volume 34, pp. 1603–1610. AAAI Press, 2020.

Jayanta Mandi, Víctor Bucarey, Maxime Mulamba Ke Tchomba, and Tias Guns. Decision-focused learning: Through the lens of learning to rank. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 14935–14947. PMLR, 17–23 Jul 2022.

Jayanta Mandi, James Kotary, Senne Berden, Maxime Mulamba, Víctor Bucarey, Tias Guns, and Ferdinando Fioretto. Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. *Journal of Artificial Intelligence Research*, 81:1623–1701, September 2024.

Jayanta Mandi, Ali İrfan Mahmutoğulları, Senne Berden, and Tias Guns. Minimizing surrogate losses for decision-focused learning using differentiable optimization. In *Proceedings of the 28th European Conference on Artificial Intelligence (ECAI)*, volume 413 of *Frontiers in Artificial Intelligence and Applications*, pp. 3888–3895, Bologna, Italy, 2025. IOS Press.

Maxime Mulamba, Jayanta Mandi, Michelangelo Diligenti, Michele Lombardi, Victor Bucarey, and Tias Guns. Contrastive losses and solution caching for predict-and-optimize. In Zhi-Hua Zhou (ed.), *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2833–2840. International Joint Conferences on Artificial Intelligence Organization, August 2021.

Quandl. WIKI various end-of-day data. `https://www.quandl.com/data/WIKI`, 2022.

R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000.

R.Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.

Sanket Shah, Kai Wang, Bryan Wilder, Andrew Perrault, and Milind Tambe. Decision-focused learning without differentiable optimization: learning locally optimized decision losses. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Proceedings of the 36th International Conference on Neural Information Processing Systems*, volume 35 of *NeurIPS '22*, pp. 1320–1332, Red Hook, NY, USA, 2022. Curran Associates, Inc.

Kai Wang, Bryan Wilder, Andrew Perrault, and Milind Tambe. Automatically learning compact quality-aware surrogates for optimization problems. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Proceedings of the 34th International Conference on Neural Information Processing Systems*, volume 33 of *NeurIPS '20*, pp. 9586–9596, Red Hook, NY, USA, 2020. Curran Associates, Inc.

Prince Zizhuang Wang, Shuyi Chen, Jinhao Liang, Ferdinando Fioretto, and Shixiang Zhu. Gen-DFL: Decision-focused generative learning for robust decision making. In *The Fourteenth International Conference on Learning Representations*, 2026.

Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI)*, volume 33, pp. 1658–1666. AAAI Press, 2019.

Yahoo! Webscope. Yahoo! Webscope dataset: ydata-ysm-advertiser-bids-v1.0. `https://webscope.sandbox.yahoo.com/`, 2007.

Zihao Zhao, Christopher Yeh, Lingkai Kong, and Kai Wang. Diffusion-DFL: Decision-focused diffusion models for stochastic optimization. In *The Fourteenth International Conference on Learning Representations*, 2026.

APPENDIX

The appendix provides the complete TRCE training procedure, full proofs and additional theoretical results, and supplementary experimental details. Algorithm 1 first summarizes the end-to-end training procedure, including candidate-set construction, conditional VaR learning, and decision-focused fine-tuning.

Section A provides the theoretical analysis. Section A.1 proves the sample-wise reformulation of the ideal TRCE objective; Section A.2 derives the CVaR decision-regret guarantee; Section A.3 establishes finite-sample generalization bounds; Section A.4 analyzes estimation and calibration of the VaR threshold; and Section A.5 combines these results into an end-to-end decision-consistency guarantee. Section B provides supplementary experimental details, including the decision tasks and robust formulations in Section B.1, compared methods in Section B.2, training configurations in Section B.3, and differentiation through reparameterized generative sampling in Section B.4.

Algorithm 1 TRCE Training with a Conditional Diffusion Model

Pre-train $p_\theta(y \mid x)$ using L_{DM}.
Initialize the candidate set $\mathcal{S} = \{\arg\min_{z\in\mathcal{Z}} f(y_i, z)\}_{i=1}^n$.
Learn $\eta_{\widehat{\phi}}$ by minimizing Eq. (5).
while not converged **do**
  for each mini-batch $\mathcal{B} \subset \mathcal{D}$ **do**
    for each $(x_i, y_i) \in \mathcal{B}$ **do**
      With probability $p_{\mathrm{solve}} = 0.1$, augment $\mathcal{S}$:
      Draw $\{y_i^{(m)}\}_{m=1}^M \sim p_\theta(\cdot \mid x_i)$.
      Compute

$$\widehat{z}_\theta(x_i) := \arg\min_{z\in\mathcal{Z}} \widehat{\mathrm{CVaR}}_\alpha\big(\{f(y_i^{(m)}, z)\}_{m=1}^M\big).$$

      Update $\mathcal{S} \leftarrow \mathcal{S} \cup \{\widehat{z}_\theta(x_i)\}$.
    end for
    Compute L_{TRCE} on $\mathcal{B}$ using Eq. (6).
    Compute the combined objective in Eq. (7).
    Update θ by gradient descent.
  end for
end while

The model-induced robust decisions in Algorithm 1 are used only to improve coverage of the candidate set $\mathcal{S}$; TRCE supervision itself is constructed from observed input–outcome pairs and the learned conditional VaR threshold.

A PROOFS AND ADDITIONAL THEORETICAL RESULTS

This section provides the complete theoretical analysis underlying the main results. We first establish the sample-wise RU reformulation, then connect excess TRCE risk to downstream CVaR regret, control the two finite-sample errors arising from generalization and threshold estimation, and finally combine them into an end-to-end consistency guarantee.

A.1 PROOF OF SECTION 3

Proof of Theorem 3.1. For a fixed x, let $Z^*(x) := \sum_{z\in\mathcal{S}} \mathcal{C}^*(z; x)$. Using $\mathcal{C}^*(z; x) = Z^*(x)\pi^*(z \mid x)$, the conditional objective can be written as For a fixed x, let $Z^*(x) := \sum_{z\in\mathcal{S}} \mathcal{C}^*(z; x)$. By Eq. (3),

we have $\mathcal{C}^*(z;x) = Z^*(x)\pi^*(z \mid x)$. Therefore, the conditional objective can be rewritten as

$$\begin{aligned}
-\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}}\mathcal{C}^*(z;x)\log\pi(z \mid x) &= -\frac{Z^*(x)}{|\mathcal{S}|}\sum_{z\in\mathcal{S}}\pi^*(z \mid x)\log\pi(z \mid x) \\
&= \frac{Z^*(x)}{|\mathcal{S}|}\left[-\sum_{z\in\mathcal{S}}\pi^*(z \mid x)\log\pi^*(z \mid x) + \sum_{z\in\mathcal{S}}\pi^*(z \mid x)\log\frac{\pi^*(z \mid x)}{\pi(z \mid x)}\right] \\
&= \frac{Z^*(x)}{|\mathcal{S}|}\left[H\big(\pi^*(\cdot \mid x)\big) + \mathrm{KL}\big(\pi^*(\cdot \mid x)\,\|\,\pi(\cdot \mid x)\big)\right].
\end{aligned}$$

Since $Z^*(x) > 0$ and the KL divergence is nonnegative, with equality if and only if $\pi(\cdot \mid x) = \pi^*(\cdot \mid x)$, the unique minimizer is given by Eq. (3). □

Proof of Theorem 3.2. Fix an arbitrary input–decision pair (x, z). By the Rockafellar–Uryasev representation of CVaR,

$$\mathcal{C}^*(z;x) = \mathrm{CVaR}_\alpha\big[f(Y,z) \mid X = x\big] = \inf_{\eta\in\mathbb{R}}\big\{\eta + \alpha^{-1}\mathbb{E}\big[\big(f(Y,z)-\eta\big)_+ \mid X = x\big]\big\}.$$

Define the sample-wise RU function $r_\eta(x,y,z) := \eta + \alpha^{-1}\big(f(y,z)-\eta\big)_+$. Then

$$\mathcal{C}^*(z;x) = \inf_{\eta\in\mathbb{R}}\mathbb{E}\left[r_\eta(x,Y,z) \mid X = x\right]. \tag{A.1}$$

Let $\eta^*(x,z) \in \arg\min_{\eta\in\mathbb{R}}\mathbb{E}\left[r_\eta(x,Y,z) \mid X = x\right]$. Substituting the minimizing threshold into Eq. (A.1) gives

$$\mathcal{C}^*(z;x) = \mathbb{E}\left[r_{\eta^*}(x,Y,z) \mid X = x\right]. \tag{A.2}$$

To characterize $\eta^*(x,z)$, define $\Psi_{x,z}(\eta) := \eta + \alpha^{-1}\mathbb{E}[(f(Y,z)-\eta)_+ \mid X = x]$. Since $\Psi_{x,z}$ is convex in η, its optimality condition $0 \in \partial\Psi_{x,z}(\eta^*(x,z))$ is precisely the conditional quantile condition. Hence every minimizer is a conditional $(1-\alpha)$-quantile of $f(Y,z)$ given $X = x$. Under the quantile convention adopted in the main text, we may choose $\eta^*(x,z) = \mathrm{VaR}_{1-\alpha}\big(f(Y,z) \mid X = x\big)$. If the conditional distribution has an atom at the quantile, the RU minimizer need not be unique; any minimizer satisfies Eq. (A.2) and yields the same CVaR value.

We now establish the sample-wise TRCE representation. Let $g : \mathcal{X}\times\mathcal{S} \to \mathbb{R}$ be any measurable function for which the following expectations are finite. By the tower property and Eq. (A.2),

$$\begin{aligned}
\mathbb{E}_{X,Y}\left[\sum_{z\in\mathcal{S}} r_{\eta^*}(X,Y,z)g(X,z)\right] &= \mathbb{E}_X\left[\sum_{z\in\mathcal{S}} g(X,z)\mathbb{E}\left[r_{\eta^*}(X,Y,z) \mid X\right]\right] \\
&= \mathbb{E}_X\left[\sum_{z\in\mathcal{S}}\mathcal{C}^*(z;X)g(X,z)\right].
\end{aligned}$$

Setting $g(X,z) := -\frac{1}{|\mathcal{S}|}\log\pi_\theta(z \mid X)$ yields

$$-\mathbb{E}_X\left[\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}}\mathcal{C}^*(z;X)\log\pi_\theta(z \mid X)\right] = -\mathbb{E}_{X,Y}\left[\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}} r_{\eta^*}(X,Y,z)\log\pi_\theta(z \mid X)\right].$$

Therefore, the ideal CVaR-weighted population objective admits an exact sample-wise stochastic representation over ordinary i.i.d. observations (X,Y). This does not imply that conditional CVaR itself is estimable from a single observation; rather, individual observations provide valid stochastic signals for optimizing the population CVaR-aligned objective. □

A.2 Decision-Regret Analysis of Regularized TRCE

We next establish the central calibration step: small excess regularized TRCE risk implies small downstream CVaR decision regret. The analysis is stated over the candidate set $\mathcal{S}$; when $z^*(x) \in \mathcal{S}$ almost surely, it coincides with regret over the full decision space.

Let $p^*(y \mid x)$ denote the true conditional distribution of Y given $X = x$. For each $z \in \mathcal{S}$, define $\mathcal{C}^*(z;x) := \mathrm{CVaR}_{\alpha,Y\sim p^*(\cdot|x)}\big[f(Y,z)\big]$, and, for a conditional generative model $p_\theta(y \mid x)$, $\mathcal{C}_\theta(z;x) := \mathrm{CVaR}_{\alpha,Y\sim p_\theta(\cdot|x)}\big[f(Y,z)\big]$. The corresponding true and learned robust decisions are $z^*_{\mathcal{S}}(x) \in \arg\min_{z\in\mathcal{S}} \mathcal{C}^*(z;x)$ and $\widehat{z}_{\mathcal{S}}(x) \in \arg\min_{z\in\mathcal{S}} \mathcal{C}_{\widehat{\theta}}(z;x)$, respectively. Define the conditional distribution discrepancy $\mathcal{K}(\theta) := \mathbb{E}_X\left[D_{\mathrm{KL}}\left(p_\theta(\cdot \mid X) \,\|\, p^*(\cdot \mid X)\right)\right]$. We consider the population regularized objective $\mathcal{J}(\theta) := L_{\mathrm{TRCE}}(\theta;\phi^*) + \beta\mathcal{K}(\theta)$, where $\beta > 0$ matches the distributional-fidelity regularization used in the main analysis.

Proof of Theorem 4.1. By definition of the generalization and threshold-estimation errors, for every $\theta \in \Theta$,

$$\big|\widehat{\mathcal{J}}_n(\theta;\widehat{\phi}) - \mathcal{J}(\theta)\big| \le \big|\widehat{\mathcal{L}}_{\mathrm{TRCE}}(\theta;\widehat{\phi}) - \widehat{\mathcal{L}}_{\mathrm{TRCE}}(\theta;\phi^*)\big| + \big|\widehat{\mathcal{L}}_{\mathrm{TRCE}}(\theta;\phi^*) - L_{\mathrm{TRCE}}(\theta;\phi^*)\big|.$$

Hence,

$$\sup_{\theta\in\Theta}\big|\widehat{\mathcal{J}}_n(\theta;\widehat{\phi}) - \mathcal{J}(\theta)\big| \le \varepsilon_\phi + \varepsilon_g. \tag{A.3}$$

Let $\Delta_n := \varepsilon_\phi + \varepsilon_g$. For any comparator $\theta^\dagger \in \Theta$, empirical optimality of $\widehat{\theta}$ gives $\widehat{\mathcal{J}}_n(\widehat{\theta};\widehat{\phi}) \le \widehat{\mathcal{J}}_n(\theta^\dagger;\widehat{\phi})$. Combining this inequality with Eq. (A.3),

$$\mathcal{J}(\widehat{\theta}) \le \widehat{\mathcal{J}}_n(\widehat{\theta};\widehat{\phi}) + \Delta_n \le \widehat{\mathcal{J}}_n(\theta^\dagger;\widehat{\phi}) + \Delta_n \le \mathcal{J}(\theta^\dagger) + 2\Delta_n.$$

Taking the infimum over $\theta^\dagger \in \Theta$ yields

$$\mathcal{J}(\widehat{\theta}) \le \inf_{\theta\in\Theta} \mathcal{J}(\theta) + 2\varepsilon_g + 2\varepsilon_\phi. \tag{A.4}$$

Let L^*_{TRCE} denote the Bayes-optimal population TRCE risk and define $\overline{\mathcal{J}}(\theta) := \mathcal{J}(\theta) - L^*_{\mathrm{TRCE}}$. By definition of the approximation error ε_a, Eq. (A.4) implies $\overline{\mathcal{J}}(\widehat{\theta}) \le \varepsilon_a + 2\varepsilon_g + 2\varepsilon_\phi$. Since $L_{\mathrm{TRCE}}(\widehat{\theta};\phi^*) - L^*_{\mathrm{TRCE}} \ge 0$, we obtain

$$\mathcal{K}(\widehat{\theta}) \le (\varepsilon_a + 2\varepsilon_g + 2\varepsilon_\phi)/\beta. \tag{A.5}$$

We next translate this distribution discrepancy into CVaR decision regret. Fix arbitrary x and z, and write $P_x := p_{\widehat{\theta}}(\cdot \mid x)$ and $Q_x := p^*(\cdot \mid x)$. By the RU representation,

$$\mathcal{C}_{\widehat{\theta}}(z;x) = \inf_{\eta\in\mathbb{R}}\big\{\eta + \alpha^{-1}\mathbb{E}_{P_x}\left[(f(Y,z)-\eta)_+\right]\big\}, \quad \mathcal{C}^*(z;x) = \inf_{\eta\in\mathbb{R}}\big\{\eta + \alpha^{-1}\mathbb{E}_{Q_x}\left[(f(Y,z)-\eta)_+\right]\big\}.$$

Under $0 \le f(Y,z) \le B$, the optimization over η may be restricted to $[0,B]$. For $q_{z,\eta}(y) := (f(y,z)-\eta)_+$, we have $0 \le q_{z,\eta}(y) \le B$, and therefore $|\mathbb{E}_{P_x}[q_{z,\eta}(Y)] - \mathbb{E}_{Q_x}[q_{z,\eta}(Y)]| \le B \cdot \mathrm{TV}(P_x, Q_x)$. Using $|\inf_\eta F(\eta) - \inf_\eta G(\eta)| \le \sup_\eta |F(\eta) - G(\eta)|$, we obtain

$$\sup_{z\in\mathcal{S}}\big|\mathcal{C}_{\widehat{\theta}}(z;x) - \mathcal{C}^*(z;x)\big| \le (B/\alpha)\cdot \mathrm{TV}(P_x, Q_x).$$

By optimality of $\widehat{z}_{\mathcal{S}}(x)$ under $\mathcal{C}_{\widehat{\theta}}$, $\mathcal{C}_{\widehat{\theta}}(\widehat{z}_{\mathcal{S}}(x);x) \le \mathcal{C}_{\widehat{\theta}}(z^*_{\mathcal{S}}(x);x)$. Hence,

$$\mathcal{C}^*(\widehat{z}_{\mathcal{S}}(x);x) - \mathcal{C}^*(z^*_{\mathcal{S}}(x);x) \le 2\sup_{z\in\mathcal{S}}\big|\mathcal{C}_{\widehat{\theta}}(z;x) - \mathcal{C}^*(z;x)\big| \le (2B/\alpha)\cdot \mathrm{TV}(P_x, Q_x).$$

Pinsker's inequality gives $\mathrm{TV}(P_x, Q_x) \le \sqrt{(1/2)\cdot D_{\mathrm{KL}}(P_x \| Q_x)}$. Therefore,

$$\mathcal{C}^*(\widehat{z}_{\mathcal{S}}(x);x) - \mathcal{C}^*(z^*_{\mathcal{S}}(x);x) \le (2B/\alpha)\cdot\sqrt{(1/2)\cdot D_{\mathrm{KL}}\left(p_{\widehat{\theta}}(\cdot \mid x)\| p^*(\cdot \mid x)\right)}.$$

Taking expectation over X and applying Jensen's inequality,

$$\mathbb{E}_X\left[\mathcal{C}^*(\widehat{z}_{\mathcal{S}}(X);X) - \mathcal{C}^*(z^*_{\mathcal{S}}(X);X)\right] \le (\sqrt{2}B/\alpha)\cdot\sqrt{\mathcal{K}(\widehat{\theta})}. \tag{A.6}$$

Combining Eq. (A.6) and Eq. (A.5) yields

$$\mathbb{E}_X\left[\mathcal{C}^*(\widehat{z}_{\mathcal{S}}(X);X) - \mathcal{C}^*(z^*_{\mathcal{S}}(X);X)\right] \le (B/\alpha)\cdot\sqrt{(2/\beta)\cdot(\varepsilon_a + 2\varepsilon_g + 2\varepsilon_\phi)}. \tag{A.7}$$

When $z^*(X) \in \mathcal{S}$ almost surely, $z^*_{\mathcal{S}}(X) = z^*(X)$ and the bound applies to the global robust decision. □

Corollary A.1 (Decision consistency)**.** *Suppose $\varepsilon_a \to 0$, $\varepsilon_g \to 0$, $\varepsilon_\phi \to 0$, while α and β remain bounded away from zero. Then $\mathbb{E}_X\left[\mathcal{C}^*(\widehat{z}_{\mathcal{S}}(X);X) - \mathcal{C}^*(z^*_{\mathcal{S}}(X);X)\right] \to 0$. Thus, vanishing approximation, generalization, and threshold-estimation errors suffice for downstream CVaR decision consistency over the candidate set, without requiring parameter consistency of the conditional generative model.*

Realizable case. If there exists $\theta^0 \in \Theta$ such that $p_{\theta^0}(y \mid x) = p^*(y \mid x)$ almost surely and θ^0 attains the Bayes-optimal population TRCE risk, then $\varepsilon_a = 0$, and Eq. (A.7) reduces to

$$\mathbb{E}_X \left[\mathcal{C}^*(\widehat{z}_{\mathcal{S}}(X); X) - \mathcal{C}^*(z^*_{\mathcal{S}}(X); X)\right] \leq (B/\alpha) \cdot \sqrt{(4/\beta) \cdot (\varepsilon_g + \varepsilon_\phi)}.$$

A.3 Generalization of the TRCE Objective

We next make the generalization term ε_g explicit. The main technical step is to control the complexity of the normalized model-induced CVaR costs appearing in the TRCE objective.

Theorem A.2 (Rademacher complexity of normalized log-costs)**.** *Let $\mathcal{S}$ be finite and for $z \in \mathcal{S}$, define $\mathcal{G}_z := \{x \mapsto \mathcal{C}_\theta(z; x) : \theta \in \Theta\}$. Assume $\mathcal{C}_\theta(z; x) \geq b > 0$ for every $\theta \in \Theta$, $x \in \mathcal{X}$, and $z \in \mathcal{S}$. Recall that $\pi_\theta(z \mid x) = \mathcal{C}_\theta(z; x)/\sum_{z' \in \mathcal{S}} \mathcal{C}_\theta(z'; x)$, define $h_\theta(z; x) := -\log \pi_\theta(z \mid x)$ and $\mathcal{H}_z := \{x \mapsto h_\theta(z; x) : \theta \in \Theta\}$. Then, for every sample $S_n = (x_1, \ldots, x_n)$,*

$$\widehat{\mathfrak{R}}_n(\mathcal{H}_z) \leq \frac{1}{b}\widehat{\mathfrak{R}}_n(\mathcal{G}_z) + \frac{1}{|\mathcal{S}|b} \sum_{z' \in \mathcal{S}} \widehat{\mathfrak{R}}_n(\mathcal{G}_{z'}). \tag{A.8}$$

Consequently,

$$\sup_{z \in \mathcal{S}} \widehat{\mathfrak{R}}_n(\mathcal{H}_z) \leq \frac{2}{b} \sup_{z \in \mathcal{S}} \widehat{\mathfrak{R}}_n(\mathcal{G}_z) \qquad \textit{and} \qquad \sup_{z \in \mathcal{S}} \mathfrak{R}_n(\mathcal{H}_z) \leq \frac{2}{b} \sup_{z \in \mathcal{S}} \mathfrak{R}_n(\mathcal{G}_z). \tag{A.9}$$

Proof. For any $z \in \mathcal{S}$, $h_\theta(z; x) = -\log \mathcal{C}_\theta(z; x) + \log\big(\sum_{z' \in \mathcal{S}} \mathcal{C}_\theta(z'; x)\big)$. Define $\mathcal{A}_z := \{x \mapsto -\log \mathcal{C}_\theta(z; x) : \theta \in \Theta\}$ and $\mathcal{Q} := \{x \mapsto \sum_{z' \in \mathcal{S}} \mathcal{C}_\theta(z'; x) : \theta \in \Theta\}$. By subadditivity,

$$\widehat{\mathfrak{R}}_n(\mathcal{H}_z) \leq \widehat{\mathfrak{R}}_n(\mathcal{A}_z) + \widehat{\mathfrak{R}}_n(\log \circ \mathcal{Q}).$$

Since $\mathcal{C}_\theta(z; x) \geq b$, the map $u \mapsto -\log u$ is $1/b$-Lipschitz on the relevant range. By contraction, $\widehat{\mathfrak{R}}_n(\mathcal{A}_z) \leq (1/b) \cdot \widehat{\mathfrak{R}}_n(\mathcal{G}_z)$. Moreover, $\sum_{z' \in \mathcal{S}} \mathcal{C}_\theta(z'; x) \geq |\mathcal{S}|b$, so $u \mapsto \log u$ is $1/(|\mathcal{S}|b)$-Lipschitz on the range of $\mathcal{Q}$. Hence, $\widehat{\mathfrak{R}}_n(\log \circ \mathcal{Q}) \leq [1/(|\mathcal{S}|b)] \cdot \widehat{\mathfrak{R}}_n(\mathcal{Q})$. Finally, $\widehat{\mathfrak{R}}_n(\mathcal{Q}) \leq \sum_{z' \in \mathcal{S}} \widehat{\mathfrak{R}}_n(\mathcal{G}_{z'})$, which proves Eq. (A.8). Taking the supremum over z, and expectation over the sample, respectively, give the inequalities in Eq. (A.9). □

Corollary A.3 (Uniform boundedness of normalized log-costs)**.** *If $0 < b \leq f(y, z) \leq B$, then, for all θ, x, and $z \in \mathcal{S}$, $0 \leq h_\theta(z; x) \leq H := \log(|\mathcal{S}|B/b)$.*

Proof. Boundedness of f implies $b \leq \mathcal{C}_\theta(z; x) \leq B$. Therefore, $\pi_\theta(z \mid x) \geq b/(|\mathcal{S}|B)$, and hence $h_\theta(z; x) = -\log \pi_\theta(z \mid x) \leq \log(|\mathcal{S}|B/b)$. The lower bound follows from $\pi_\theta(z \mid x) \leq 1$. □

Proof of Proposition 4.4. For the oracle threshold ϕ^*, define

$$\ell_{\mathrm{TRCE}}(\theta, \phi^*; x, y) := \frac{1}{|\mathcal{S}|} \sum_{z \in \mathcal{S}} r_{\phi^*}(x, y, z) h_\theta(z; x).$$

Under $0 \leq f(y, z) \leq B$ and $0 \leq \eta_{\phi^*}(x, z) \leq B$, $0 \leq r_{\phi^*}(x, y, z) \leq B/\alpha$. Together with Corollary A.3, $0 \leq \ell_{\mathrm{TRCE}}(\theta, \phi^*; x, y) \leq BH/\alpha$. Let $\mathcal{F}_{\mathrm{TRCE}} := \{(x, y) \mapsto \ell_{\mathrm{TRCE}}(\theta, \phi^*; x, y) : \theta \in \Theta\}$. Standard Rademacher uniform convergence implies that, with probability at least $1 - \delta$,

$$\varepsilon_g \leq 2\mathfrak{R}_n(\mathcal{F}_{\mathrm{TRCE}}) + \frac{BH}{\alpha}\sqrt{\frac{\log(2/\delta)}{2n}}. \tag{A.10}$$

Since r_{ϕ^*} is independent of θ and bounded by B/α, contraction and subadditivity yield

$$\mathfrak{R}_n(\mathcal{F}_{\mathrm{TRCE}}) \leq \frac{B}{\alpha|\mathcal{S}|} \sum_{z \in \mathcal{S}} \mathfrak{R}_n(\mathcal{H}_z) \leq \frac{B}{\alpha} \sup_{z \in \mathcal{S}} \mathfrak{R}_n(\mathcal{H}_z).$$

Applying Theorem A.2, $\mathfrak{R}_n(\mathcal{F}_{\mathrm{TRCE}}) \leq \frac{2B}{\alpha b} \sup_{z \in \mathcal{S}} \mathfrak{R}_n(\mathcal{G}_z)$. Substitution into Eq. (A.10) gives

$$\varepsilon_g \leq \frac{4B}{\alpha b} \sup_{z \in \mathcal{S}} \mathfrak{R}_n(\mathcal{G}_z) + \frac{B}{\alpha} \log\left(\frac{|\mathcal{S}|B}{b}\right) \sqrt{\frac{\log(2/\delta)}{2n}}, \tag{A.11}$$

which finishes the proof. □

Bounded costs alone do not imply vanishing generalization error; the complexity of the model-induced CVaR class must also decrease with sample size. In particular, if $\sup_{z\in\mathcal{S}} \mathfrak{R}_n(\mathcal{G}_z) = O(n^{-1/2})$, then Eq. (A.11) yields the standard $O(n^{-1/2})$ rate up to problem-dependent constants.

A.4 Estimation of the Conditional VaR Threshold

We now quantify the auxiliary estimation problem introduced by the sample-wise reformulation. Rather than estimating conditional CVaR directly, TRCE learns the corresponding conditional VaR threshold by empirical minimization of the RU objective:

$$\widehat{\phi} \in \arg\min_{\phi\in\Phi} \widehat{\mathcal{R}}_n(\phi) := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}} r_\phi(x_i, y_i, z), \tag{A.12}$$

where $r_\phi(x, y, z) = \eta_\phi(x, z) + \alpha^{-1}\big(f(y, z) - \eta_\phi(x, z)\big)_+$. Define the population RU risk $\mathcal{R}(\phi) := \mathbb{E}_{X,Y}[\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}} r_\phi(X, Y, z)]$. Let ϕ^* denote an oracle threshold function satisfying $\eta_{\phi^*}(x, z) \in \arg\min_\eta \mathbb{E}\left[r_\eta(x, Y, z) \mid X = x\right]$. Thus, $\eta_{\phi^*}(x, z)$ is a conditional $(1-\alpha)$-quantile of $f(Y, z)$ given $X = x$. Assume throughout this subsection that $0 \le f(y, z) \le B$, $0 \le \eta_\phi(x, z) \le B$. Then $0 \le r_\phi(x, y, z) \le B/\alpha$. Define $\mathcal{G}_{\mathrm{RU}} := \{(x, y) \mapsto \frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}} r_\phi(x, y, z) : \phi \in \Phi\}$.

Theorem A.4 (Finite-sample RU excess risk)**.** *With probability at least* $1-\delta$*,*

$$\mathcal{R}(\widehat{\phi}) - \mathcal{R}(\phi^*) \le \varepsilon_a^\Phi + 4\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}}) + \frac{2B}{\alpha}\sqrt{\frac{\log(2/\delta)}{2n}}, \tag{A.13}$$

where $\varepsilon_a^\Phi := \inf_{\phi\in\Phi} \mathcal{R}(\phi) - \mathcal{R}(\phi^*)$ *is the approximation error of the threshold class.*

Proof. Let $\phi^\dagger \in \arg\min_{\phi\in\Phi} \mathcal{R}(\phi)$. By empirical optimality, $\widehat{\mathcal{R}}_n(\widehat{\phi}) \le \widehat{\mathcal{R}}_n(\phi^\dagger)$, and therefore $\mathcal{R}(\widehat{\phi}) - \mathcal{R}(\phi^\dagger) \le 2\sup_{\phi\in\Phi} |\widehat{\mathcal{R}}_n(\phi) - \mathcal{R}(\phi)|$. Since every function in $\mathcal{G}_{\mathrm{RU}}$ is bounded in $[0, B/\alpha]$, standard Rademacher uniform convergence gives, with probability at least $1-\delta$,

$$\sup_{\phi\in\Phi} \left|\widehat{\mathcal{R}}_n(\phi) - \mathcal{R}(\phi)\right| \le 2\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}}) + \frac{B}{\alpha}\sqrt{\frac{\log(2/\delta)}{2n}}.$$

Adding $\mathcal{R}(\phi^\dagger) - \mathcal{R}(\phi^*) = \varepsilon_a^\Phi$ proves Eq. (A.13). □

In fact, the RU complexity can be related directly to that of the threshold network. For $\mathcal{E}_z := \{x \mapsto \eta_\phi(x, z) : \phi \in \Phi\}$, the map $\eta \mapsto \eta + \alpha^{-1}(f(y, z) - \eta)_+$ is $L_\alpha := \max\{1, (1-\alpha)/\alpha\} \le \alpha^{-1}$ Lipschitz. Hence, $\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}}) \le \frac{L_\alpha}{|\mathcal{S}|}\sum_{z\in\mathcal{S}} \mathfrak{R}_n(\mathcal{E}_z)$. If $\sup_{z\in\mathcal{S}} \mathfrak{R}_n(\mathcal{E}_z) \le \mathfrak{R}_n(\mathcal{E})$, then $\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}}) \le L_\alpha \mathfrak{R}_n(\mathcal{E}) \le \alpha^{-1}\mathfrak{R}_n(\mathcal{E})$.

A.4.1 Calibration of the RU Threshold

For each (x, z), define $F_{x,z}(t) := \mathbb{P}\left(f(Y, z) \le t \mid X = x\right)$ and $\eta^*(x, z) := \eta_{\phi^*}(x, z)$.

Theorem A.5 (RU threshold calibration)**.** *Suppose Assumption 4.2 holds and* $|\eta_\phi(x, z) - \eta^*(x, z)| \le r_0$ *for the threshold functions under consideration. Then*

$$\mathcal{R}(\phi) - \mathcal{R}(\phi^*) \ge \frac{\kappa}{2\alpha}\mathbb{E}_X\left[\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}}\big(\eta_\phi(X, z) - \eta^*(X, z)\big)^2\right]. \tag{A.14}$$

Consequently,

$$\mathbb{E}_X\left[\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}}\big(\eta_{\widehat{\phi}}(X, z) - \eta^*(X, z)\big)^2\right] \le \frac{2\alpha}{\kappa}\big[\mathcal{R}(\widehat{\phi}) - \mathcal{R}(\phi^*)\big]. \tag{A.15}$$

Proof. Fix (x, z) and define $\mathcal{R}_{x,z}(\eta) := \eta + \alpha^{-1}\mathbb{E}\left[(f(Y, z) - \eta)_+ \mid X = x\right]$. At continuity points of $F_{x,z}$, $(\partial/\partial\eta)\mathcal{R}_{x,z}(\eta) = [F_{x,z}(\eta) - (1-\alpha)]/\alpha$. Since $\eta^*(x, z)$ is the conditional $(1-\alpha)$-quantile, $F_{x,z}(\eta^*(x, z)) = 1 - \alpha$ under the local regularity condition. For $\eta \ge \eta^*(x, z)$,

$$\mathcal{R}_{x,z}(\eta) - \mathcal{R}_{x,z}(\eta^*(x, z)) = \frac{1}{\alpha}\int_{\eta^*(x,z)}^{\eta} \left[F_{x,z}(t) - (1-\alpha)\right] dt.$$

Assumption 4.2 gives $F_{x,z}(t)-(1-\alpha) \geq \kappa\big(t-\eta^*(x,z)\big)$, and therefore $\mathcal{R}_{x,z}(\eta)-\mathcal{R}_{x,z}(\eta^*(x,z)) \geq \frac{\kappa}{2\alpha}(\eta-\eta^*(x,z))^2$. The case $\eta < \eta^*(x,z)$ is analogous. Averaging over X and $z \in \mathcal{S}$ proves Eq. (A.14); rearranging and setting $\phi = \widehat{\phi}$ gives Eq. (A.15). □

Combining Theorems A.4 and A.5, with probability at least $1-\delta$,

$$\mathbb{E}_X\left[\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}}\big(\eta_{\widehat{\phi}}(X,z)-\eta^*(X,z)\big)^2\right] \leq \frac{2\alpha}{\kappa}\left[\varepsilon_a^{\Phi}+4\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}})+\frac{2B}{\alpha}\sqrt{\frac{\log(2/\delta)}{2n}}\right].$$

A.4.2 Threshold-Induced TRCE Error

Recall $\varepsilon_\phi := \sup_{\theta\in\Theta}|\widehat{\mathcal{L}}_{\mathrm{TRCE}}(\theta;\widehat{\phi})-\widehat{\mathcal{L}}_{\mathrm{TRCE}}(\theta;\phi^*)|$.

Proof of Theorem 4.3. Since the RU function is L_α-Lipschitz in its threshold, $|r_{\widehat{\phi}}(x,y,z) - r_{\phi^*}(x,y,z)| \leq L_\alpha \cdot |\eta_{\widehat{\phi}}(x,z)-\eta_{\phi^*}(x,z)|$. Moreover, by Corollary A.3, $0 \leq h_\theta(z;x) \leq H$. Hence, for any $\theta\in\Theta$,

$$\big|\widehat{\mathcal{L}}_{\mathrm{TRCE}}(\theta;\widehat{\phi})-\widehat{\mathcal{L}}_{\mathrm{TRCE}}(\theta;\phi^*)\big| \leq \frac{HL_\alpha}{n|\mathcal{S}|}\sum_{i=1}^{n}\sum_{z\in\mathcal{S}}\Big|\eta_{\widehat{\phi}}(x_i,z)-\eta_{\phi^*}(x_i,z)\Big|. \tag{A.16}$$

Define

$$\widehat{e}_{\phi,1} := \frac{1}{n|\mathcal{S}|}\sum_{i=1}^{n}\sum_{z\in\mathcal{S}}\Big|\eta_{\widehat{\phi}}(x_i,z)-\eta_{\phi^*}(x_i,z)\Big|,\ \widehat{e}_{\phi,2} := \left[\frac{1}{n|\mathcal{S}|}\sum_{i=1}^{n}\sum_{z\in\mathcal{S}}\big(\eta_{\widehat{\phi}}(x_i,z)-\eta_{\phi^*}(x_i,z)\big)^2\right]^{1/2}.$$

The right-hand side of Eq. (A.16) is independent of θ, so $\varepsilon_\phi \leq HL_\alpha\widehat{e}_{\phi,1}$. By Cauchy–Schwarz, $\varepsilon_\phi \leq HL_\alpha\widehat{e}_{\phi,2} \leq (H/\alpha)\cdot\widehat{e}_{\phi,2}$. For the corresponding population quantity,

$$e_{\phi,2} := \left(\mathbb{E}_X\left[\frac{1}{|\mathcal{S}|}\sum_{z\in\mathcal{S}}\big(\eta_{\widehat{\phi}}(X,z)-\eta_{\phi^*}(X,z)\big)^2\right]\right)^{1/2},$$

Theorems A.4 and A.5 imply

$$e_{\phi,2} \leq \left(\frac{2\alpha}{\kappa}\left[\varepsilon_a^{\Phi}+4\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}})+\frac{2B}{\alpha}\sqrt{\frac{\log(2/\delta)}{2n}}\right]\right)^{1/2}.$$

Combining this population control with the corresponding empirical-to-population deviation yields, up to problem-dependent constants,

$$\varepsilon_\phi \lesssim \frac{H}{\sqrt{\alpha\kappa}}\left[\varepsilon_a^{\Phi}+4\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}})+\frac{2B}{\alpha}\sqrt{\frac{\log(2/\delta)}{2n}}\right]^{1/2}, \tag{A.17}$$

which finishes the proof. □

If $\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}}) = O(\mathfrak{C}_\Phi/\sqrt{n})$, then Eq. (A.17) gives

$$\varepsilon_\phi = O_p\left(\frac{H}{\sqrt{\alpha\kappa}}\left[\sqrt{\varepsilon_a^{\Phi}}+\frac{\sqrt{\mathfrak{C}_\Phi}+\alpha^{-1/2}\log^{1/4}(1/\delta)}{n^{1/4}}\right]\right).$$

In the realizable threshold case $\varepsilon_a^{\Phi}=0$, $\varepsilon_\phi = O_p(n^{-1/4})$, up to problem-dependent constants.

A.5 Combined Decision-Regret Guarantee

We finally combine the preceding results into the complete statistical pathway from single-outcome supervision to robust decision quality: conditional VaR estimation controls threshold-induced TRCE error, generalization controls empirical TRCE risk, and Theorem 4.1 converts these errors into downstream CVaR decision regret. Combining Theorem 4.1 with Eq. (A.17) gives

$$\begin{aligned}&\mathbb{E}_X\left[\mathcal{C}^*(\widehat{z}_{\mathcal{S}}(X);X)-\mathcal{C}^*(z^*_{\mathcal{S}}(X);X)\right]\\ &\quad\lesssim \frac{B}{\alpha}\left(\frac{2}{\beta}\left(\varepsilon_a+2\varepsilon_g+\frac{2H}{\sqrt{\alpha\kappa}}\left[\varepsilon_a^{\Phi}+4\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}})+\frac{2B}{\alpha}\sqrt{\frac{\log(2/\delta)}{2n}}\right]^{1/2}\right)\right)^{1/2}.\end{aligned} \tag{A.18}$$

If $\varepsilon_g = O_p(n^{-1/2})$, $\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}}) = O(n^{-1/2})$, $\varepsilon_a = \varepsilon_a^{\Phi} = 0$, then $\varepsilon_\phi = O_p(n^{-1/4})$, and Eq. (A.18) yields $\mathbb{E}_X\left[\mathcal{C}^*(\widehat{z}_{\mathcal{S}}(X); X) - \mathcal{C}^*(z^*_{\mathcal{S}}(X); X)\right] = O_p(n^{-1/8})$. This conservative rate results from two successive worst-case square-root conversions—from RU excess risk to threshold error and from distribution discrepancy to decision regret—and is not an intrinsic statistical limit of TRCE. If a sharper threshold estimator satisfies $e_{\phi,2} = O_p(n^{-1/2})$, then $\varepsilon_\phi = O_p(n^{-1/2})$ and the resulting regret improves to $\mathbb{E}_X\left[\mathcal{C}^*(\widehat{z}_{\mathcal{S}}(X); X) - \mathcal{C}^*(z^*_{\mathcal{S}}(X); X)\right] = O_p(n^{-1/4})$.

Suppose $\varepsilon_a \to 0$, $\varepsilon_a^{\Phi} \to 0$, $\varepsilon_g \to 0$, $\mathfrak{R}_n(\mathcal{G}_{\mathrm{RU}}) \to 0$, while α, β, and the calibration constant κ remain bounded away from zero. Under Assumption 4.2, $\mathbb{E}_X\left[\mathcal{C}^*(\widehat{z}_{\mathcal{S}}(X); X) - \mathcal{C}^*(z^*_{\mathcal{S}}(X); X)\right] \to 0$. Thus, TRCE consistently recovers the Bayes-optimal CVaR decision over the candidate set from ordinary single-outcome observations without requiring true CVaR values, CVaR-optimal decision labels, or parameter consistency of the conditional generative model. When the candidate set contains the global Bayes-optimal decision, this coincides with consistency over the full decision space.

B Additional Experimental Details

This section provides the task formulations, baseline implementations, and training configurations underlying Section 5. Unless otherwise stated, all methods use identical data splits, stochastic perturbations, and evaluation protocols. Consistent with the learning setting in the main text, each training input is paired with only one observed outcome, while repeated outcomes are used only at test time to accurately estimate the ground-truth CVaR objective and decision regret.

B.1 Real-World Decision-Making Tasks

Portfolio Optimization. We consider stochastic portfolio optimization following Wang et al. (2020) with $m = 50$ candidate stocks. For stock i on a given trading day, $x_i \in \mathbb{R}^F$ contains features constructed from historical prices and returns, while $y_i \in \mathbb{R}$ denotes its next-period return. Features and returns are extracted from the Quandl WIKI dataset (Quandl, 2022).

Let $\Sigma \in \mathbb{R}^{m\times m}$ denote the covariance matrix estimated exclusively from training returns, with a diagonal ridge term for numerical stability. Given a conditional outcome distribution, the portfolio is obtained by solving

$$\min_{z\in[0,1]^m,\, \sum_{i=1}^m z_i = 1} \mathrm{CVaR}_{\alpha, Y\sim p_\theta(\cdot|x)}[-Y^\top z + \tau z^\top \Sigma z],$$

where $\tau = 0.1$ controls the quadratic risk penalty. The objective therefore balances adverse-return protection against covariance-weighted portfolio exposure. To ensure positive CVaR costs, we shift the original costs by subtracting the minimum realized cost over all training outcomes and candidate decisions, $\min_{(x,y)\in D, z\in\mathcal{Z}} f(y,z)$, and adding a small constant 10^{-5}, where D denotes the training set.

Observations are indexed chronologically by trading day, so we use a chronological training/validation/test split to avoid temporal leakage. Each day contributes exactly one input–outcome pair, and Σ is estimated once from the training partition and then fixed.

Power Scheduling. We consider day-ahead power scheduling following Donti et al. (2017); Zhao et al. (2026). The dataset contains more than eight years of real electrical-grid observations. The input x contains the previous day's hourly demand, historical temperatures, next-day temperature forecasts, lagged and rolling statistics, calendar indicators, and yearly sinusoidal features. The uncertain outcome $y = (y_1, \ldots, y_{24})$ is the next-day hourly demand profile.

The operator chooses a 24-hour generation schedule $z \in \mathbb{R}^{24}$ subject to $|z_i - z_{i-1}| \le c_r$, $c_r = 0.5$. Writing $[v]_+ := \max\{v, 0\}$, the robust scheduling problem is

$$\min_{z\in\mathbb{R}^{24},\, |z_i - z_{i-1}|\le c_r} \mathrm{CVaR}_{\alpha, Y\sim p_\theta(\cdot|x)}\left[\sum_{i=1}^{24} \gamma_s (y_i - z_i)_+ + \gamma_e (z_i - y_i)_+ + 0.5(y_i - z_i)^2\right],$$

where $\gamma_s = 50$ and $\gamma_e = 0.5$ are the shortage and excess-generation penalties. The strong asymmetry $\gamma_s \gg \gamma_e$ makes adverse under-generation events disproportionately costly, providing a natural testbed for tail-risk-aware decision learning.

Web Advertising. We consider the budget-allocation benchmark of Wilder et al. (2019) with $m = 5$ candidate websites and $K = 10$ users. For website i and user j, $y_{ij} \in [0, 1]$ denotes the click-through rate (CTR), with $y_i = (y_{ij})_{j=1}^{K} \in [0, 1]^K$, $y = (y_i)_{i=1}^{m}$. The CTR outcomes are constructed from the Yahoo! Webscope advertising dataset (Yahoo! Webscope, 2007). Following standard decision-focused learning benchmarks, inputs are generated through a fixed random linear transformation, $x_i = Ay_i$, $A \in \mathbb{R}^{K\times K}$. The allocation is obtained by solving

$$\min_{z\in[0,1]^m,\, \sum_{i=1}^{m} z_i=1} \mathrm{CVaR}_{\alpha, Y\sim p_\theta(\cdot|x)} \left[\sum_{j=1}^{K} \prod_{i=1}^{m} (1 - z_i y_{ij}) \right].$$

For each user, the product term represents the probability of unsuccessful exposure across all websites; minimizing its CVaR therefore emphasizes adverse allocation outcomes.

B.2 Compared Methods

We compare TRCE against representative methods spanning deterministic DFL, robust optimization, and generative DFL.

Pairwise and Listwise Ranking. We use the ranking-based DFL objectives of Mandi et al. (2022), which encourage correct decision-cost orderings through pairwise or listwise ranking losses. They provide deterministic decision-aware baselines without explicitly modeling conditional outcome uncertainty or tail risk.

NCE. Noise-Contrastive Estimation (NCE) follows the contrastive DFL formulation of Mulamba et al. (2021), learning from relative decision quality by contrasting preferred and non-preferred candidate decisions.

PTO. Predict-then-Optimize (PTO) first minimizes a standard prediction loss and then solves the downstream optimization problem using the predicted outcome. It serves as a non-decision-focused deterministic reference.

SPO+. SPO+ (Elmachtoub & Grigas, 2022) optimizes a convex surrogate of downstream decision regret and provides a canonical deterministic DFL baseline.

Diff-DRO. Diff-DRO (Ma et al., 2024) incorporates distributional robustness through differentiable distributionally robust optimization layers. It provides a robust optimization baseline distinct from TRCE's explicit CVaR-aligned distribution learning.

Gen-DFL. Gen-DFL (Wang et al., 2026) is the closest robust generative DFL baseline. It learns a conditional generative model and estimates tail-sensitive quantities from generated samples to construct robust decision supervision for a contrastive objective. TRCE instead constructs sample-wise tail-risk supervision from observed outcomes through the learned RU threshold, without using CVaR-optimal decision labels as training supervision.

Diff-DFL. Diff-DFL (Zhao et al., 2026) performs decision-focused distribution learning with a conditional diffusion model. While it captures conditional outcome uncertainty, it does not explicitly align the learned distribution with the downstream CVaR objective targeted by TRCE.

All methods use the same data partitions and stochastic outcome constructions and are evaluated under the same ground-truth CVaR regret protocol described in Section 5. This isolates the effect of the learning objective from differences in data construction or evaluation.

B.3 Training Details

TRCE training follows Algorithm 1. We first pretrain the conditional generative model with the standard diffusion objective. We then learn the conditional VaR threshold $\eta_\phi(x, z)$ from single-outcome training pairs by minimizing the empirical RU objective in Eq. (A.12). Finally, we fine-tune the generative model using the combined diffusion and TRCE objective in Eq. (7). Crucially, repeated

outcomes for the same training input are not introduced at any stage. Table 4 summarizes the task-specific configurations. Hyperparameters are selected on the validation partition and fixed before test evaluation.

Table 4: Dataset statistics and training configurations.

Configuration	`Portfolio`	`Power`	`Web`
Train/validation/test	640/80/80	2042/256/256	480/60/60
Input dimension	$50 \times 28 = 1400$	149	50
Outcome dimension	50	24	50
Evaluation scenarios per input	80	80	80
Balancing parameter γ	10	10	10
Epochs (pretrain/VaR/fine-tune)	700/100/60	700/100/60	700/100/60
Learning rate (pretrain/VaR/fine-tune)	$10^{-3}/10^{-3}/10^{-4}$	$10^{-3}/10^{-3}/10^{-5}$	$10^{-3}/10^{-3}/10^{-5}$
Weight decay	10^{-5}	10^{-6}	10^{-5}
Monte Carlo samples	80	80	80
Diffusion steps (forward/sampling)	100/100	50/50	100/100

Epochs correspond to diffusion pretraining, VaR-threshold learning, and TRCE fine-tuning, respectively.

We use the same number of Monte Carlo samples across tasks for model-induced CVaR estimation and final evaluation. Diffusion sampling steps are task-specific, while the remaining optimization hyperparameters are selected independently on each validation set. Sensitivity to the CVaR tail probability α and TRCE coefficient γ is reported in Section 5.

B.4 Differentiation through Generative Sampling

For a reparameterizable conditional generative model, including the diffusion models used in our experiments, a generated outcome can be written as $y^{(m)} = R_\theta\big(\epsilon^{(m)}, x\big)$, $\epsilon^{(m)} \sim p(\epsilon)$, where R_θ denotes the differentiable sampling transformation. Consequently, gradients of model-induced quantities such as $\mathcal{C}_\theta(z; x)$ can propagate through generated outcomes to θ. Reparameterization alone, however, does not make the original nested decision-focused objective in Eq. (1) tractable: its gradient would additionally require differentiation through $z_\theta(x) = \arg\min_{z \in \mathcal{Z}} \mathcal{C}_\theta(z; x)$, which is generally non-differentiable for discrete decisions. TRCE avoids this requirement by aligning model-induced CVaR costs over the candidate set through the differentiable normalized-cost objective $\pi_\theta(z \mid x)$. Reparameterization therefore provides gradients through the generative model, while TRCE avoids differentiation through the downstream discrete optimizer.