Conformal Robustness in Prediction-Driven Decision-Making

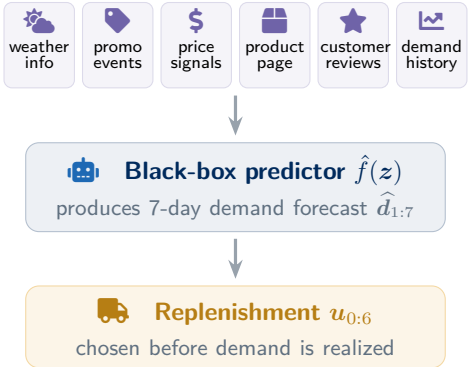
Hansheng Jiang

Rotman School of Management
University of Toronto

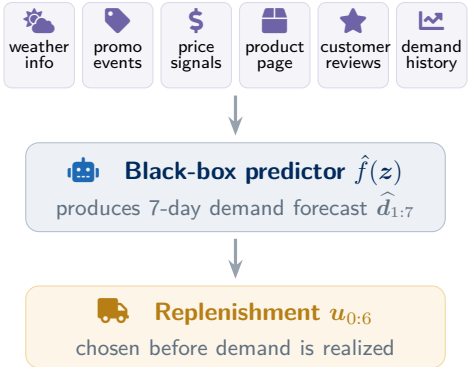
Joint work with
Lingjie Zhao (Tsinghua) and Wei Qi (Tsinghua & McGill)

MSOM 2026
Boston, MA

Motivating example: inventory replenishment

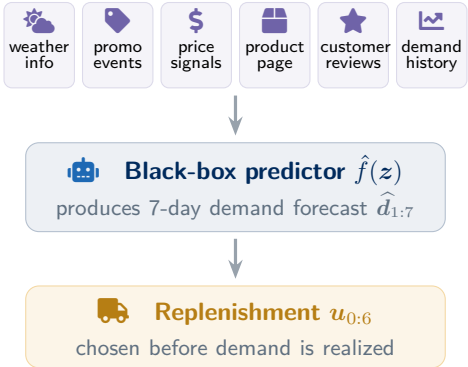


Motivating example: inventory replenishment



Inventory manager

Motivating example: inventory replenishment



“No stockouts with at least 95% probability.”

Reliability request

vs.

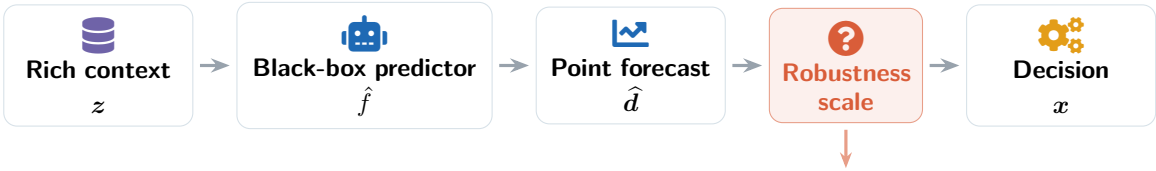
“Keep the inventory costs below $20k.”

Target request

Inventory manager

Point forecasts are powerful but decisions might be fragile

Point forecasts are powerful but decisions might be fragile



Two types of managerial requests

Perishable inventory

Reliability request:

- *"No stockouts with at least 95% probability."*

Target request:

- *"Aim to make the inventory costs below $20k."*

Two types of managerial requests

Perishable inventory

Ride-share dispatching

Reliability request:

- *"No stockouts with at least 95% probability."*
- *"95% of pickups occur within 5 minutes."*

Target request:

- *"Aim to make the inventory costs below $20k."*
- *"Aim to control driver incentive spending below budget."*

Two types of managerial requests

Perishable inventory

Reliability request:

- *"No stockouts with at least 95% probability."*

Target request:

- *"Aim to make the inventory costs below $20k."*

Ride-share dispatching

- *"95% of pickups occur within 5 minutes."*
- *"Aim to control driver incentive spending below budget."*

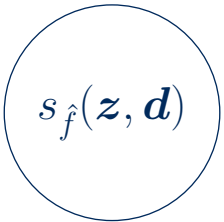
Emergency care

- *"50% of the time doctors are available."*
- *"Aim to keep patient waiting times under 30 minutes."*

One score function supports both views

Fix the predictor $\hat{f}$.

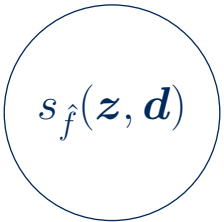
A conformal score function measures the nonnegative discrepancy between the forecast $\hat{f}(\boldsymbol{z})$ and realized outcome $\boldsymbol{d}$.



One score function supports both views

Fix the predictor $\hat{f}$.

A conformal score function measures the nonnegative discrepancy between the forecast $\hat{f}(\boldsymbol{z})$ and realized outcome $\boldsymbol{d}$.



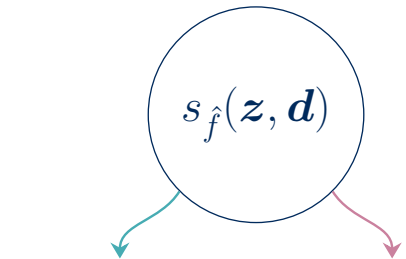
Examples of score choice

- $s_{\hat{f}}(\boldsymbol{z}, \boldsymbol{d}) = \max_j \dfrac{|d_j - \hat{f}_j(\boldsymbol{z})|}{\hat{r}_j}$
- $s_{\hat{f}}(\boldsymbol{z}, \boldsymbol{d}) = \|\widehat{\Sigma}^{-1/2}(\boldsymbol{d} - \hat{f}(\boldsymbol{z}))\|_2$
- $s_{\hat{f}}(\boldsymbol{z}, \boldsymbol{d}) = \sum_j \dfrac{|d_j - \hat{f}_j(\boldsymbol{z})|}{\hat{r}_j}$

One score function supports both views

Fix the predictor $\hat{f}$.

A conformal score function measures the nonnegative discrepancy between the forecast $\hat{f}(\boldsymbol{z})$ and realized outcome $\boldsymbol{d}$.



Examples of score choice

- $s_{\hat{f}}(\boldsymbol{z}, \boldsymbol{d}) = \max_j \dfrac{|d_j - \hat{f}_j(\boldsymbol{z})|}{\hat{r}_j}$
- $s_{\hat{f}}(\boldsymbol{z}, \boldsymbol{d}) = \|\widehat{\Sigma}^{-1/2}(\boldsymbol{d} - \hat{f}(\boldsymbol{z}))\|_2$
- $s_{\hat{f}}(\boldsymbol{z}, \boldsymbol{d}) = \sum_j \dfrac{|d_j - \hat{f}_j(\boldsymbol{z})|}{\hat{r}_j}$

Conformal score calibration quantifies prediction uncertainty

- After $\hat{f}$ and $s_{\hat{f}}$ are fixed, use an exchangeable calibration dataset $\{(\boldsymbol{Z}_i, \boldsymbol{D}_i)\}_{i=1}^n$ and compute its scores

$$S_i = s_{\hat{f}}(\boldsymbol{Z}_i, \boldsymbol{D}_i), \qquad i = 1, \dots, n.$$

- Sort the scores and select the split-conformal quantile

$$S_{(1)} \le \cdots \le S_{(n)}, \qquad k_\alpha = \lceil (n+1)\alpha \rceil, \qquad \eta_\alpha = S_{(k_\alpha)},$$

with $S_{(n+1)} := +\infty$.

- For a new context $\boldsymbol{z}_{\text{test}}$, form the calibrated score sublevel set

$$\mathcal{U}_\alpha(\boldsymbol{z}_{\text{test}}) = \{\boldsymbol{d} \in \mathcal{D} : s_{\hat{f}}(\boldsymbol{z}_{\text{test}}, \boldsymbol{d}) \le \eta_\alpha\}.$$

Calibration guarantees transfer to decisions

Lemma: calibration–decision transfer

Let $\tilde{\boldsymbol{x}}$ be chosen before observing $\boldsymbol{D}_{\text{test}}$. If every performance requirement is enforced throughout the conformal set,

$$h_m(\tilde{\boldsymbol{x}}, \boldsymbol{d}) \le 0, \quad m = 1, \dots, M, \qquad \forall \boldsymbol{d} \in \mathcal{U}_\alpha(\boldsymbol{Z}_{\text{test}}),$$

then exchangeability and conformal calibration imply

$$\Pr\{h_m(\tilde{\boldsymbol{x}}, \boldsymbol{D}_{\text{test}}) \le 0,\ m = 1, \dots, M\} \ge \alpha.$$

- The calibrated set $\mathcal{U}_\alpha(\boldsymbol{Z}_{\text{test}})$ contains the realized uncertainty with probability at least α.
- Any objective or constraint certificate valid over that set inherits the same reliability level.

Two decision-making strategies

Unknown consumer demand

Unknown ride requests

Unknown patient needs

ConfRO

Conformal Robust Optimization

Choose reliability α

$\Downarrow$

Calibrate uncertainty set

$\mathbb{P}(\boldsymbol{d} \in \mathcal{U}_\alpha(\boldsymbol{z})) \geq \alpha$

$\Downarrow$

Enforce constraints

$\mathbb{P}(a_i(\boldsymbol{x}, \boldsymbol{d}) \leq b_i) \geq \alpha$

Two decision-making strategies

Unknown consumer demand

Unknown ride requests

Unknown patient needs

ConfRO
Conformal Robust Optimization

Choose reliability α

$\Downarrow$

Calibrate uncertainty set

$\mathbb{P}(\boldsymbol{d} \in \mathcal{U}_\alpha(\boldsymbol{z})) \geq \alpha$

$\Downarrow$

Enforce constraints

$\mathbb{P}(a_i(\boldsymbol{x}, \boldsymbol{d}) \leq b_i) \geq \alpha$

ConfRS
Conformal Robust Satisficing

Choose target τ

$\Downarrow$

Limit target violation

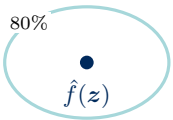
$a(\boldsymbol{x}, \boldsymbol{d}) - \tau \leq k\, \mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})$

$\Downarrow$

Minimize fragility

$\kappa(\tau) = \min_{\boldsymbol{x},\, k \geq 0} k$

ConfRO: decision-maker chooses reliability level α, not a set radius



Reliability level α decides a calibrated score quantile $S_{(\lceil \alpha(n+1) \rceil)}$ and uncertainty set $\mathcal{U}_\alpha(z)$.

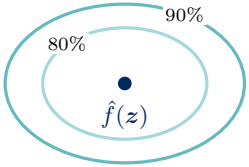
Conformal Robust Optimization (ConfRO) problem:

$$Z_{\text{ConfRO}}(\alpha) = \min_{x} \max_{d \in \mathcal{U}_\alpha(z)} a_0(x, d)$$
$$\text{s.t. } a_i(x, d) \le b_i, \quad i \in [I],$$
$$\forall d \in \mathcal{U}_\alpha(z),$$
$$x \in \mathcal{X}.$$

$$\mathcal{U}_\alpha(z) = \{d \in \mathfrak{D} : s_{\hat{f}}(z, d) \le \eta_\alpha\}.$$

- **Modular:** works with a fixed predictor; no retraining or white-box access.

ConfRO: decision-maker chooses reliability level α, not a set radius



Reliability level α decides a calibrated score quantile $S_{(\lceil \alpha(n+1) \rceil)}$ and uncertainty set $\mathcal{U}_\alpha(z)$.

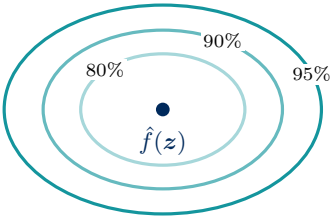
Conformal Robust Optimization (ConfRO) problem:

$$Z_{\text{ConfRO}}(\alpha) = \min_{x} \max_{d \in \mathcal{U}_\alpha(z)} a_0(x, d)$$
$$\text{s.t. } a_i(x, d) \le b_i, \quad i \in [I],$$
$$\forall d \in \mathcal{U}_\alpha(z),$$
$$x \in \mathcal{X}.$$

$$\mathcal{U}_\alpha(z) = \{d \in \mathfrak{D} : s_{\hat{f}}(z, d) \le \eta_\alpha\}.$$

- **Modular:** works with a fixed predictor; no retraining or white-box access.
- **Interpretable:** the input is a reliability level, not an opaque tuning radius.

ConfRO: decision-maker chooses reliability level α, not a set radius



Reliability level α decides a calibrated score quantile $S_{(\lceil \alpha(n+1) \rceil)}$ and uncertainty set $\mathcal{U}_\alpha(\boldsymbol{z})$.

Conformal Robust Optimization (ConfRO) problem:

$$Z_{\text{ConfRO}}(\alpha) = \min_{\boldsymbol{x}} \max_{\boldsymbol{d} \in \mathcal{U}_\alpha(\boldsymbol{z})} a_0(\boldsymbol{x}, \boldsymbol{d})$$
$$\text{s.t. } a_i(\boldsymbol{x}, \boldsymbol{d}) \leq b_i, \quad i \in [I],$$
$$\forall \boldsymbol{d} \in \mathcal{U}_\alpha(\boldsymbol{z}),$$
$$\boldsymbol{x} \in \mathcal{X}.$$

$$\mathcal{U}_\alpha(\boldsymbol{z}) = \{\boldsymbol{d} \in \mathfrak{D} : s_{\hat{f}}(\boldsymbol{z}, \boldsymbol{d}) \leq \eta_\alpha\}.$$

- **Modular:** works with a fixed predictor; no retraining or white-box access.
- **Interpretable:** the input is a reliability level, not an opaque tuning radius.
- **Decision-ready:** score geometry can reflect tractability and economically relevant deviations.

What determines the decision-efficiency loss in ConfRO?

Proposition: decision-sensitive efficiency bound (informal)

For linear/convex robust optimization with a metric-type conformal score function,

$$\Delta_{\text{decision}} \leq \underbrace{\eta_\alpha}_{\substack{\text{calibrated} \\ \text{radius}}} \times \underbrace{\lambda^R}_{\substack{\text{problem} \\ \text{sensitivity}}} \times \underbrace{L_{\mathfrak{s}}(\boldsymbol{x}^\star)}_{\substack{\text{score–decision} \\ \text{alignment}}} .$$

What determines the decision-efficiency loss in ConfRO?

Proposition: decision-sensitive efficiency bound (informal)

For linear/convex robust optimization with a metric-type conformal score function,

$$\Delta_{\text{decision}} \leq \underbrace{\eta_\alpha}_{\substack{\text{calibrated} \\ \text{radius}}} \times \underbrace{\lambda^R}_{\substack{\text{problem} \\ \text{sensitivity}}} \times \underbrace{L_{\mathfrak{s}}(\boldsymbol{x}^\star)}_{\substack{\text{score–decision} \\ \text{alignment}}} .$$

- Calibrated radius: better forecasts concentrate calibration scores near zero and reduce η_α.

What determines the decision-efficiency loss in ConfRO?

Proposition: decision-sensitive efficiency bound (informal)

For linear/convex robust optimization with a metric-type conformal score function,

$$\Delta_{\text{decision}} \leq \underbrace{\eta_\alpha}_{\substack{\text{calibrated} \\ \text{radius}}} \times \underbrace{\lambda^R}_{\substack{\text{problem} \\ \text{sensitivity}}} \times \underbrace{L_{\mathfrak{s}}(\boldsymbol{x}^\star)}_{\substack{\text{score–decision} \\ \text{alignment}}} .$$

- Calibrated radius: better forecasts concentrate calibration scores near zero and reduce η_α.
- Problem sensitivity: λ^R is the marginal objective value of relaxing the robust constraint.

What determines the decision-efficiency loss in ConfRO?

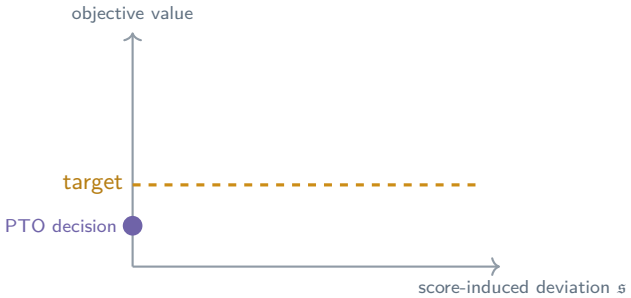
Proposition: decision-sensitive efficiency bound (informal)

For linear/convex robust optimization with a metric-type conformal score function,

$$\Delta_{\text{decision}} \leq \underbrace{\eta_\alpha}_{\substack{\text{calibrated} \\ \text{radius}}} \times \underbrace{\lambda^R}_{\substack{\text{problem} \\ \text{sensitivity}}} \times \underbrace{L_{\mathfrak{s}}(\boldsymbol{x}^\star)}_{\substack{\text{score–decision} \\ \text{alignment}}} .$$

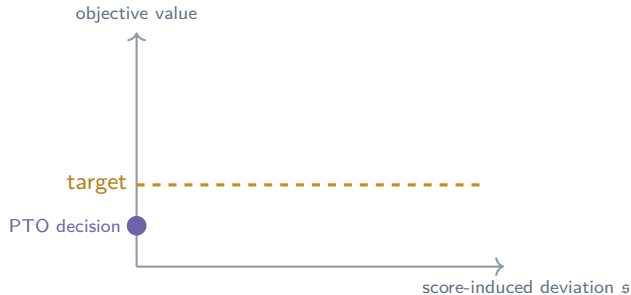
- Calibrated radius: better forecasts concentrate calibration scores near zero and reduce η_α.
- Problem sensitivity: λ^R is the marginal objective value of relaxing the robust constraint.
- Score-decision alignment: better geometry protects economically relevant directions and lowers $L_{\mathfrak{s}}(\boldsymbol{x}^\star)$.

Target satisficing versus decision fragility



* We consider a minimization problem, so lower objective values are better.

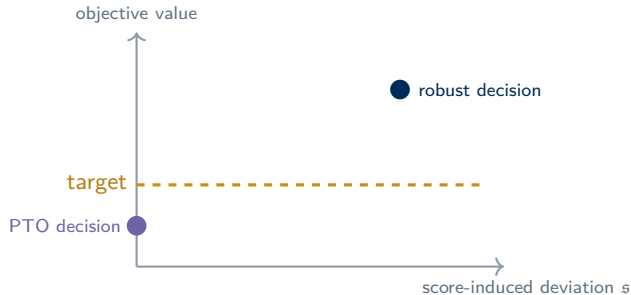
Target satisficing versus decision fragility



A decision is fragile when its guaranteed objective value deviates farther from the target (larger target violation) as prediction error grows (larger score-induced deviation).

* We consider a minimization problem, so lower objective values are better.

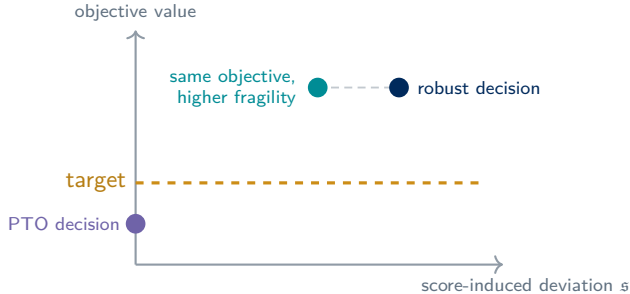
Target satisficing versus decision fragility



A decision is fragile when its guaranteed objective value deviates farther from the target (larger target violation) as prediction error grows (larger score-induced deviation).

* We consider a minimization problem, so lower objective values are better.

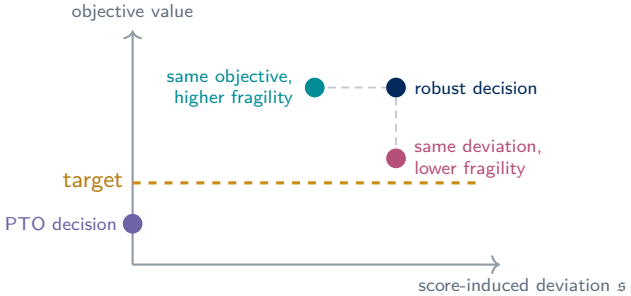
Target satisficing versus decision fragility



A decision is fragile when its guaranteed objective value deviates farther from the target (larger target violation) as prediction error grows (larger score-induced deviation).

* We consider a minimization problem, so lower objective values are better.

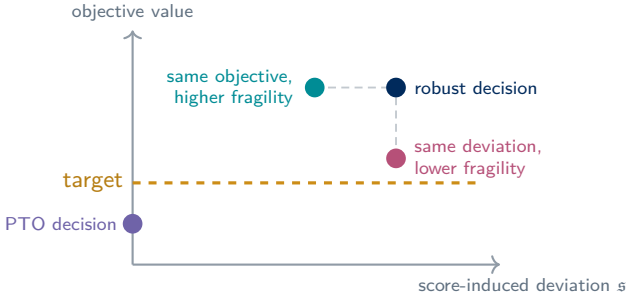
Target satisficing versus decision fragility



A decision is fragile when its guaranteed objective value deviates farther from the target (larger target violation) as prediction error grows (larger score-induced deviation).

* We consider a minimization problem, so lower objective values are better.

Target satisficing versus decision fragility



A decision is fragile when its guaranteed objective value deviates farther from the target (larger target violation) as prediction error grows (larger score-induced deviation).

$$\text{decision fragility} \approx \frac{\text{target violation}}{\text{score-induced deviation } s}$$

* We consider a minimization problem, so lower objective values are better.

ConfRS: a new *post-prediction* satisficing formulation

Pointwise form

$$\kappa^\star(\tau) = \min_{\boldsymbol{x}\in\mathcal{X},\, k\geq 0} k$$
$$\text{s.t. } a(\boldsymbol{x},\boldsymbol{d}) - \tau \leq k\,\mathfrak{s}(\boldsymbol{d},\widehat{\boldsymbol{d}}), \qquad \forall \boldsymbol{d}\in\mathcal{D}$$

Equivalent worst-case form

$$\kappa^\star(\tau) = \min_{\boldsymbol{x}\in\mathcal{X},\, k\geq 0} k$$
$$\text{s.t. } \sup_{\boldsymbol{d}\in\mathcal{D}}\Big\{a(\boldsymbol{x},\boldsymbol{d}) - k\,\mathfrak{s}(\boldsymbol{d},\widehat{\boldsymbol{d}})\Big\} \leq \tau$$

* For fixed context z, let $\hat{\boldsymbol{d}} = \hat{f}(z)$ and introduce the score-induced deviation $\mathfrak{s}(\boldsymbol{d},\hat{\boldsymbol{d}}) := s_{\hat{f}}(z,\boldsymbol{d})$.

ConfRS and DRS robustify different objects

Conformal Robust Satisficing (ConfRS)

Forecast $\hat{\boldsymbol{d}} = \hat{f}(\boldsymbol{z})$ → Realization $\boldsymbol{d}$

Conformal score deviation $\mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})$

Instance-level fragility

$$\frac{a(\boldsymbol{x}, \boldsymbol{d}) - \tau}{\mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})}$$

ConfRS and DRS robustify different objects

Conformal Robust Satisficing (ConfRS)

Forecast $\hat{\boldsymbol{d}} = \hat{f}(\boldsymbol{z})$ → Realization $\boldsymbol{d}$

Conformal score deviation $\mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})$

Instance-level fragility

$$\frac{a(\boldsymbol{x}, \boldsymbol{d}) - \tau}{\mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})}$$

Distributionally RS (DRS)

Reference $\hat{P}$ → Truth P

Wasserstein distance $\Delta(P, \hat{P})$

Distribution-level fragility

$$\frac{\mathbb{E}_P[a(\boldsymbol{x}, \boldsymbol{d})] - \tau}{\Delta(P, \hat{P})}$$

ConfRS and DRS robustify different objects

Conformal Robust Satisficing (ConfRS)

Forecast $\hat{\boldsymbol{d}} = \hat{f}(\boldsymbol{z})$ → Realization $\boldsymbol{d}$

Conformal score deviation $\mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})$

Instance-level fragility

$$\frac{a(\boldsymbol{x}, \boldsymbol{d}) - \tau}{\mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})}$$

Distributionally RS (DRS)

Reference $\hat{P}$ → Truth P

Wasserstein distance $\Delta(P, \hat{P})$

Distribution-level fragility

$$\frac{\mathbb{E}_P[a(\boldsymbol{x}, \boldsymbol{d})] - \tau}{\Delta(P, \hat{P})}$$

ConfRS adds a modular post-prediction robustness layer to any fitted black-box predictor.

ConfRS yields a well-defined fragility measure

For a target-violation function $v : \mathcal{D} \to \mathbb{R}$, define the *conformal fragility measure*

$$\rho(v) = \inf\{k \geq 0 : \ v(\boldsymbol{d}) \leq k\, \mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}}),\ \forall \boldsymbol{d} \in \mathcal{D}\}$$

Theorem: conformal fragility measure

Suppose $\mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}}) \geq 0$, $\mathfrak{s}(\widehat{\boldsymbol{d}}, \widehat{\boldsymbol{d}}) = 0$, and $\mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}}) > 0$ for $\boldsymbol{d} \neq \widehat{\boldsymbol{d}}$. Then ρ is lower semicontinuous and satisfies:

- *Monotonicity:* if $v_1(\boldsymbol{d}) \geq v_2(\boldsymbol{d})$ for all $\boldsymbol{d} \in \mathcal{D}$, then $\rho(v_1) \geq \rho(v_2)$.
- *Positive homogeneity:* for any $\lambda \geq 0$, $\rho(\lambda v) = \lambda \rho(v)$.
- *Subadditivity:* $\rho(v_1 + v_2) \leq \rho(v_1) + \rho(v_2)$.
- *Pro-robustness:* if $v(\boldsymbol{d}) \leq 0$ for all $\boldsymbol{d} \in \mathcal{D}$, then $\rho(v) = 0$.
- *Anti-fragility:* if $v(\widehat{\boldsymbol{d}}) > 0$, then $\rho(v) = +\infty$.

ConfRS enjoys data-driven target certificates

Let $(\boldsymbol{x}^\star, k^\star)$ solve ConfRS at target τ, and let $\widehat{F}_n$ be the empirical CDF of held-out calibration scores.

Proposition: target-violation tail bound

Under i.i.d. calibration and test observations, for any violation margin $\Delta > 0$ and confidence $\varepsilon \in (0, 1)$,

$$\Pr\{a(\boldsymbol{x}^\star, \boldsymbol{D}_{\text{test}}) > \tau + \Delta\} \leq 1 - \widehat{F}_n(\Delta/k^\star) + \sqrt{\frac{\ln(1/\varepsilon)}{2n}}$$

with probability at least $1 - \varepsilon$ over calibration.

Remark. An finite-sample upper bound on the target holds naturally with ConfRS, requiring only exchangeability data

$$\Pr\{a(\boldsymbol{x}^\star, \boldsymbol{D}_{\text{test}}) \leq \tau + k^\star \eta_\alpha\} \geq \alpha.$$

What connects reliability and target-oriented robustness?

	ConfRO	ConfRS
Input	Reliability level α	Acceptable target τ
Decision goal	Minimize worst-case cost over $\mathcal{U}_\alpha(\boldsymbol{z})$	Minimize fragility at target τ
Robustness object	$\mathcal{U}_\alpha(\boldsymbol{z}) = \{\boldsymbol{d} : s(\boldsymbol{z}, \boldsymbol{d}) \leq \eta_\alpha\}$	$\inf\{k \geq 0 : \ v(\boldsymbol{d}) \leq k\, \mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})\}$
Guarantee	Reliability via conformal-set coverage	Target-violation certificate via fragility
Managerial question	"How much reliability do I want?"	"How fragile is the target?"

What connects reliability and target-oriented robustness?

	ConfRO	ConfRS
Input	Reliability level α	Acceptable target τ
Decision goal	Minimize worst-case cost over $\mathcal{U}_\alpha(\boldsymbol{z})$	Minimize fragility at target τ
Robustness object	$\mathcal{U}_\alpha(\boldsymbol{z}) = \{\boldsymbol{d} : s(\boldsymbol{z}, \boldsymbol{d}) \leq \eta_\alpha\}$	$\inf\{k \geq 0 : \ v(\boldsymbol{d}) \leq k\, \mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})\}$
Guarantee	Reliability via conformal-set coverage	Target-violation certificate via fragility
Managerial question	"How much reliability do I want?"	"How fragile is the target?"

What connects reliability and target-oriented robustness?

	ConfRO	ConfRS
Input	Reliability level α	Acceptable target τ
Decision goal	Minimize worst-case cost over $\mathcal{U}_\alpha(\boldsymbol{z})$	Minimize fragility at target τ
Robustness object	$\mathcal{U}_\alpha(\boldsymbol{z}) = \{\boldsymbol{d} : s(\boldsymbol{z}, \boldsymbol{d}) \leq \eta_\alpha\}$	$\inf\{k \geq 0 : \ v(\boldsymbol{d}) \leq k\, \mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})\}$
Guarantee	Reliability via conformal-set coverage	Target-violation certificate via fragility
Managerial question	"How much reliability do I want?"	"How fragile is the target?"

Why this is important?

- *Tools extension:* transfers existing tools and guarantees between the two paradigms, broadening both theoretical and algorithmic scope.

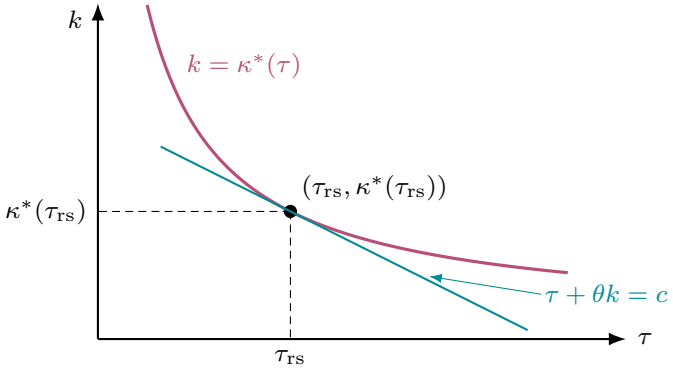
What connects reliability and target-oriented robustness?

	ConfRO	ConfRS
Input	Reliability level α	Acceptable target τ
Decision goal	Minimize worst-case cost over $\mathcal{U}_\alpha(\boldsymbol{z})$	Minimize fragility at target τ
Robustness object	$\mathcal{U}_\alpha(\boldsymbol{z}) = \{\boldsymbol{d} : s(\boldsymbol{z}, \boldsymbol{d}) \leq \eta_\alpha\}$	$\inf\{k \geq 0 : \ v(\boldsymbol{d}) \leq k\, \mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})\}$
Guarantee	Reliability via conformal-set coverage	Target-violation certificate via fragility
Managerial question	"How much reliability do I want?"	"How fragile is the target?"

Why this is important?

- *Tools extension:* transfers existing tools and guarantees between the two paradigms, broadening both theoretical and algorithmic scope.
- *Philosophical insight:* reveals an intrinsic link between maximizing utility under robustness and minimizing fragility around a target.

Equivalence geometry: ConfRO supports the ConfRS frontier



Interpretation

A ConfRS target selects a point on the fragility frontier; the matched ConfRO radius is the slope of the supporting hyperplane.

Equivalence begins with an exact score-dual representation

For score radius θ, define

$$\mathcal{D}(\theta) = \{\boldsymbol{d} \in \mathcal{D} : \mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}}) \leq \theta\}.$$

Assumption: strong duality of score-constrained robust optimization

For every $\boldsymbol{x} \in \mathcal{X}$ and every score radius θ of interest,

$$\sup_{\boldsymbol{d} \in \mathcal{D}(\theta)} a(\boldsymbol{x}, \boldsymbol{d}) = \inf_{k \geq 0} \left\{ k\theta + \sup_{\boldsymbol{d} \in \mathcal{D}} (a(\boldsymbol{x}, \boldsymbol{d}) - k\mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})) \right\}.$$

Equivalence begins with an exact score-dual representation

For score radius θ, define

$$\mathcal{D}(\theta) = \{\boldsymbol{d} \in \mathcal{D} : \mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}}) \leq \theta\}.$$

Assumption: strong duality of score-constrained robust optimization

For every $\boldsymbol{x} \in \mathcal{X}$ and every score radius θ of interest,

$$\sup_{\boldsymbol{d} \in \mathcal{D}(\theta)} a(\boldsymbol{x}, \boldsymbol{d}) = \inf_{k \geq 0} \left\{ k\theta + \sup_{\boldsymbol{d} \in \mathcal{D}} (a(\boldsymbol{x}, \boldsymbol{d}) - k\mathfrak{s}(\boldsymbol{d}, \widehat{\boldsymbol{d}})) \right\}.$$

The robust radius θ becomes a linear price on score-measured fragility.

Every ConfRO solution has a target-oriented interpretation

Proposition: target-oriented interpretation of ConfRO

Assume strong duality. For any optimal solution $(\boldsymbol{x}^\star, k^\star, \tau^\star)$ of

$$\min_{\boldsymbol{x},k,\tau} \ \tau + k\theta \quad \text{s.t.} \quad \sup_{\boldsymbol{d}\in\mathcal{D}} \{a(\boldsymbol{x},\boldsymbol{d}) - k\mathfrak{s}(\boldsymbol{d},\widehat{\boldsymbol{d}})\} \le \tau,$$

the pair $(\boldsymbol{x}^\star, k^\star)$ is optimal for ConfRS at target $\tau^\star$.

Every ConfRO solution has a target-oriented interpretation

Proposition: target-oriented interpretation of ConfRO

Assume strong duality. For any optimal solution $(\boldsymbol{x}^\star, k^\star, \tau^\star)$ of

$$\min_{\boldsymbol{x},k,\tau} \ \tau + k\theta \quad \text{s.t.} \quad \sup_{\boldsymbol{d}\in\mathcal{D}} \{a(\boldsymbol{x},\boldsymbol{d}) - k\mathfrak{s}(\boldsymbol{d},\widehat{\boldsymbol{d}})\} \le \tau,$$

the pair $(\boldsymbol{x}^\star, k^\star)$ is optimal for ConfRS at target $\tau^\star$.

ConfRO implicitly chooses an acceptable target and then minimizes fragility at that target.

Reliability and target levels map through the fragility frontier

Let $\kappa^\star(\tau)$ be the optimal ConfRS value. The score-radius formulation of ConfRO reduces to

$$Z_{\text{ConfRO}}(\theta) = \min_{\tau \in [\underline{\tau}, \overline{\tau}]} \{\tau + \theta \kappa^\star(\tau)\}.$$

Theorem: reliability-target translation (informal)

Assume convexity and strong duality. Fix an interior target τ_{rs} and choose a negative subgradient $\beta(\tau_{\text{rs}}) \in \partial \kappa^\star(\tau_{\text{rs}})$. Set

$$\theta^\star = -1/\beta(\tau_{\text{rs}}).$$

Then every optimal ConfRS decision at target τ_{rs} is optimal for ConfRO at radius $\theta^\star$. If the scalar minimizer is unique, the two optimal decision sets coincide.

Fragility is the marginal cost of robustness

Robust value: $V(\theta) := \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{d}\in\mathfrak{D}:\, \mathfrak{s}(\boldsymbol{d},\widehat{\boldsymbol{d}})\le\theta} a(\boldsymbol{x},\boldsymbol{d})$.

Target minimizers: $T(\theta) := \arg\min_{\tau\in[\underline{\tau},\bar{\tau}]}\{\tau+\theta\kappa^\star(\tau)\}$.

Theorem: marginal cost of robustness (informal)

Fix $\beta_{\mathrm{rs}} \in \partial\kappa^\star(\tau_{\mathrm{rs}})$ with $\beta_{\mathrm{rs}} < 0$, and set $\theta_{\mathrm{rs}} = -1/\beta_{\mathrm{rs}}$ and $\alpha_{\mathrm{rs}} = F(\theta_{\mathrm{rs}})$. Then

$$\lim_{h\downarrow 0} \frac{V(\theta_{\mathrm{rs}}+h)-V(\theta_{\mathrm{rs}})}{h} = \min_{\tau\in T(\theta_{\mathrm{rs}})} \kappa^\star(\tau), \qquad \lim_{h\downarrow 0} \frac{V(\theta_{\mathrm{rs}})-V(\theta_{\mathrm{rs}}-h)}{h} = \max_{\tau\in T(\theta_{\mathrm{rs}})} \kappa^\star(\tau).$$

If F is differentiable and increasing near θ_{rs}, with $f_S(\theta_{\mathrm{rs}}) = F'(\theta_{\mathrm{rs}}) > 0$, and V is differentiable, then

$$\partial_\alpha V(F^{-1}(\alpha))\big|_{\alpha=\alpha_{\mathrm{rs}}} = \kappa^\star(\tau_{\mathrm{rs}}) f_S(\theta_{\mathrm{rs}})^{-1}.$$

Fragility is the marginal cost of robustness

Robust value: $V(\theta) := \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{d}\in\mathfrak{D}:\, \mathfrak{s}(\boldsymbol{d},\widehat{\boldsymbol{d}})\le\theta} a(\boldsymbol{x},\boldsymbol{d})$.

Target minimizers: $T(\theta) := \arg\min_{\tau\in[\underline{\tau},\bar{\tau}]}\{\tau+\theta\kappa^\star(\tau)\}$.

Theorem: marginal cost of robustness (informal)

Fix $\beta_{\mathrm{rs}} \in \partial\kappa^\star(\tau_{\mathrm{rs}})$ with $\beta_{\mathrm{rs}} < 0$, and set $\theta_{\mathrm{rs}} = -1/\beta_{\mathrm{rs}}$ and $\alpha_{\mathrm{rs}} = F(\theta_{\mathrm{rs}})$. Then

$$\lim_{h\downarrow 0} \frac{V(\theta_{\mathrm{rs}}+h)-V(\theta_{\mathrm{rs}})}{h} = \min_{\tau\in T(\theta_{\mathrm{rs}})} \kappa^\star(\tau), \qquad \lim_{h\downarrow 0} \frac{V(\theta_{\mathrm{rs}})-V(\theta_{\mathrm{rs}}-h)}{h} = \max_{\tau\in T(\theta_{\mathrm{rs}})} \kappa^\star(\tau).$$

If F is differentiable and increasing near θ_{rs}, with $f_S(\theta_{\mathrm{rs}}) = F'(\theta_{\mathrm{rs}}) > 0$, and V is differentiable, then

$$\partial_\alpha V(F^{-1}(\alpha))\big|_{\alpha=\alpha_{\mathrm{rs}}} = \kappa^\star(\tau_{\mathrm{rs}}) f_S(\theta_{\mathrm{rs}})^{-1}.$$

$\kappa^\star(\tau_{\mathrm{rs}})$ is the marginal cost of expanding the score radius.

The slope of $V \circ F^{-1}$ reflects both decision fragility and prediction accuracy. Reliability is costly when $\kappa^\star(\tau_{\mathrm{rs}})$ is large or $f_S(\theta_{\mathrm{rs}})$ is small.

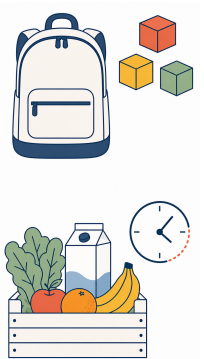
Numerical study

Evaluation logic

A good method should meet coverage, avoid excessive conservatism, and improve realized decision quality.

Experiment overview

- Synthetic benchmark: robust fractional knapsack
  - Tests empirical coverage and realized utility.
  - Validates the ConfRO–ConfRS map.
  - Compares ConfRS with distributionally RS .
- Real-data case study: multiperiod perishable inventory in e-grocery
  - Converts deep-learning demand forecasts into calibrated uncertainty sets.
  - Evaluates cost, coverage, score geometry, and residual scaling.



Synthetic data: robust fractional knapsack

Consider a n-item fractional knapsack problem with context-dependent item utilities $\boldsymbol{c}$ and a budget constraint B.

Item utilities $\boldsymbol{c}$ are unknown and a prediction is available using context $\boldsymbol{z}$.

Utility maximization formulated as a minimization problem:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & a(\boldsymbol{x}, \boldsymbol{c}) := -\boldsymbol{c}^\top \boldsymbol{x}, \\ \text{s.t.} \quad & \boldsymbol{x} \in [0,1]^n, \quad \boldsymbol{p}^\top \boldsymbol{x} \leq B. \end{aligned}$$

Synthetic data: robust fractional knapsack

Consider a n-item fractional knapsack problem with context-dependent item utilities $\boldsymbol{c}$ and a budget constraint B.

Item utilities $\boldsymbol{c}$ are unknown and a prediction is available using context $\boldsymbol{z}$.

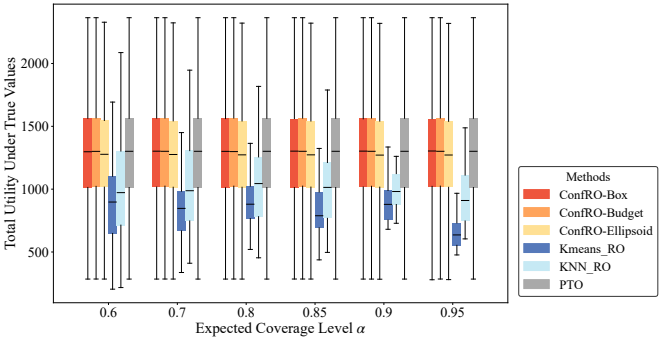
Utility maximization formulated as a minimization problem:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & a(\boldsymbol{x}, \boldsymbol{c}) := -\boldsymbol{c}^\top \boldsymbol{x}, \\ \text{s.t.} \quad & \boldsymbol{x} \in [0,1]^n, \quad \boldsymbol{p}^\top \boldsymbol{x} \leq B. \end{aligned}$$

Methods compared

- ConfRO with Box, Ellipsoid, and Budget scores
- Local KNN/K-Means RO
- Context-agnostic Ellipsoid RO
- PTO as the point-forecast

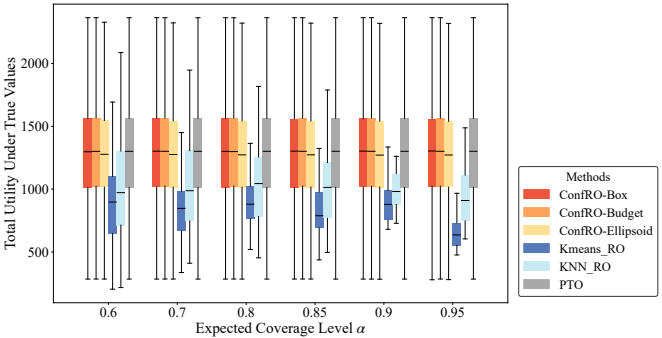
ConfRO achieves robustness without sacrificing objective value



Realized coverages under $\alpha = 0.90$

Method	Realized coverage	Objective
ConfRO-Bo	**0.89**	**-1309**
ConfRO-E	**0.89**	**-1293**
ConfRO-Bu	**0.89**	**-1311**
KNN	0.85	-918
K-Means	0.74	-1051
Ellips.	0.89	**0**
PTO	–	-1311

ConfRO achieves robustness without sacrificing objective value



Realized coverages under $\alpha = 0.90$

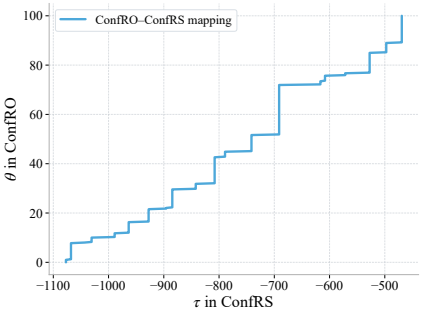
Method	Realized coverage	Objective
ConfRO-Bo	**0.89**	**-1309**
ConfRO-E	**0.89**	**-1293**
ConfRO-Bu	**0.89**	**-1311**
KNN	0.85	-918
K-Means	0.74	-1051
Ellips.	0.89	**0**
PTO	–	-1311

Comparison highlights

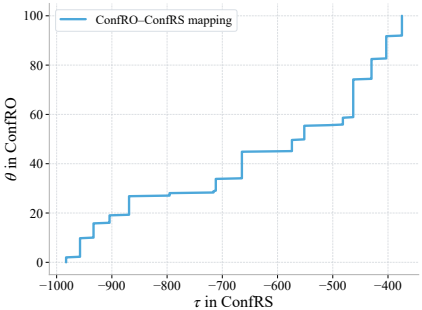
- ConfRO closely tracks nominal coverage while K-Means undercovers at high levels.
- ConfRO improves realized utility over KNN/K-Means by 23.2%–97.7%.
- The context-agnostic Ellipsoid baseline degenerates to zero utility due to over-conservatism.

ConfRO–ConfRS correspondence is numerically validated

With $\mathfrak{s}(\boldsymbol{c}, \hat{\boldsymbol{c}}) = \|(\boldsymbol{c} - \hat{\boldsymbol{c}})/\hat{\boldsymbol{r}}\|_1$, the affine objective, convex feasible set, and strong duality satisfy the equivalence conditions.



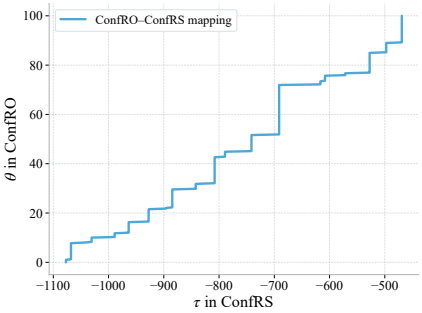
Instance 1



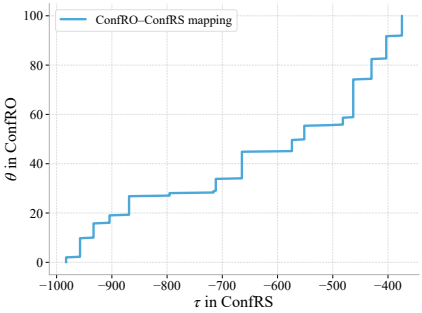
Instance 2

ConfRO–ConfRS correspondence is numerically validated

With $\mathfrak{s}(\boldsymbol{c}, \hat{\boldsymbol{c}}) = \|(\boldsymbol{c} - \hat{\boldsymbol{c}})/\hat{\boldsymbol{r}}\|_1$, the affine objective, convex feasible set, and strong duality satisfy the equivalence conditions.



Instance 1



Instance 2

Why step-like behavior occurs?

The mapping is not one-to-one! A single decision can remain optimal for an interval of targets τ, producing a step-like map to radius θ.

ConfRS vs. DRS

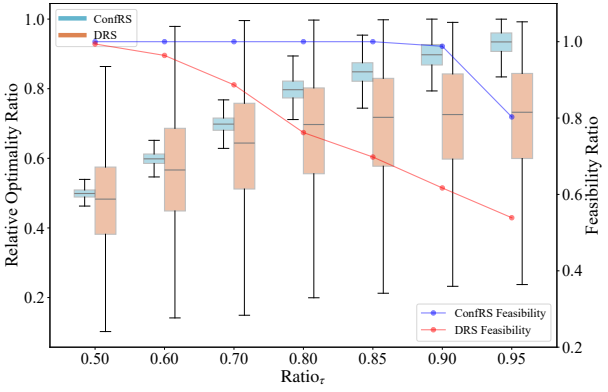
Experiment setup: $S = 40$ historical samples, 1,000 test instances, and unscaled score $\mathfrak{s}(\boldsymbol{c}, \hat{\boldsymbol{c}}) = \|\boldsymbol{c} - \hat{\boldsymbol{c}}\|_1$. Targets are $\tau = \rho_\tau V_D$ for $\rho_\tau \in \{0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95\}$.

Metrics

- Feasibility: fraction of instances satisfying $V_{RS} \leq \tau$.
- Relative optimality: V_{RS}/V_D for feasible instances.

Findings

- DRS becomes increasingly infeasible as ρ_τ increases.
- ConfRS remains feasible for most instances until the most stringent targets.
- Among feasible instances, ConfRS attains a higher mean V_{RS}/V_D with lower dispersion.



Real data: perishable e-grocery inventory management

Data

- 50,000 store-SKU pairs over 97 days, spanning 898 stores in 18 cities.
- Contexts include product-location identifiers, transactions, stockouts, prices, promotions, calendar variables, and weather.
- Treat each store-SKU trajectory as a cross-sectional instance within the common holdout window.

Does the **deep-learning demand predictor** provide a strong enough anchor for the downstream robust inventory model?

Train and select the best deep-learning demand predictor

- Train three deep-learning demand predictors: TFT, DLinear, and SSA.
- Evaluate on a held-out test set under metrics WAPE, WPE, and MAE.
- Compare performance across all store-SKU pairs, high-volume SKUs, and low-volume SKUs.

	Overall			High-volume			Low-volume		
Method	WAPE	WPE	MAE	WAPE	WPE	MAE	WAPE	WPE	MAE
TFT	**29.2%**	5.4%	**0.363**	**22.8%**	**2.1%**	**0.644**	**38.1%**	10.0%	**0.267**
DLinear	30.6%	6.6%	0.381	23.5%	2.7%	0.666	40.6%	12.0%	0.283
SSA	31.1%	-1.2%	0.387	26.0%	-4.0%	0.739	38.5%	2.7%	0.267

Predict, calibrate, then solve the robust inventory model

Multiperiod robust perishable inventory management problem

$$\begin{aligned} \min_{u_0,\dots,u_{T-1}} \quad & \max_{\boldsymbol{w}\in\mathcal{U}_\alpha} \sum_{k=0}^{T-1} \big\{ cu_k + R(x_{k+1}) \big\} \\ \text{s.t.} \quad & x_{k+1} = x_k + u_k - w_k, \qquad k = 0,\dots,T-1, \\ & R(x) = \max\{hx, -px\}. \end{aligned}$$

c, h, and p denote procurement, holding, and shortage costs.

Predict, calibrate, then solve the robust inventory model

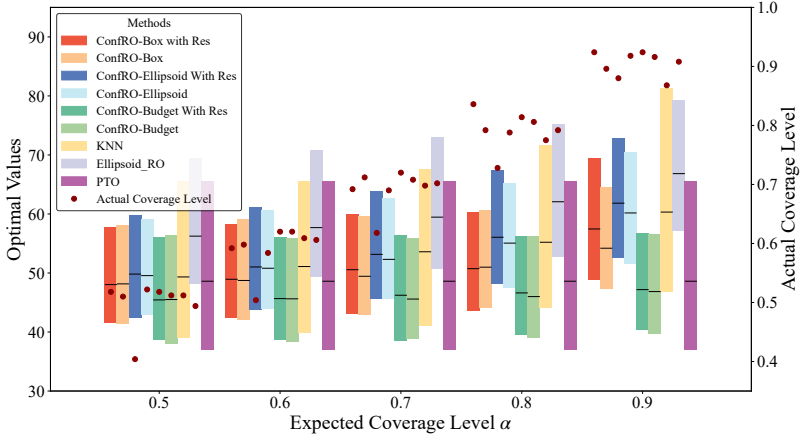
Multiperiod robust perishable inventory management problem

$$\min_{u_0,\dots,u_{T-1}} \quad \max_{\boldsymbol{w}\in\mathcal{U}_\alpha} \sum_{k=0}^{T-1} \{cu_k + R(x_{k+1})\}$$
$$\text{s.t.} \quad x_{k+1} = x_k + u_k - w_k, \qquad k = 0,\dots,T-1,$$
$$R(x) = \max\{hx, -px\}.$$

c, h, and p denote procurement, holding, and shortage costs.

- Forecast store-SKU demand with TFT.
- Calibrate $\mathcal{U}_\alpha$ using Box, Ellipsoid, or Budget scores, with or without residual scaling.
- Compare $\alpha \in [0.5, 0.9]$ with KNN/KMeans RO, global RO, and PTO; stress-test demand with independent $\mathrm{Unif}[0.8, 1.4]$ multipliers.

Budget ConfRO reaches the low-cost reliability frontier



Under demand perturbations, Budget ConfRO attains lower mean cost and lower cost variability than PTO while controlling reliability.

Practical insights from the numerical study

Prioritize forecast quality before tuning robustness

Strong context-specific forecasts shrink calibrated residual scores and reduce the conservatism passed to the downstream robust model.

Favor parsimonious geometries that control aggregate deviations

Budget scores balance interpretability and tractability, avoiding coordinate-wise Box conservatism and covariance-estimation instability in Ellipsoid sets.

Treat residual scaling as validation-sensitive

Adaptive scaling can help under data heteroscedasticity, but a secondary residual model may introduce estimation noise that offsets its benefit.

Takeaways

1. Conformal score as a robustness scale

A conformal score can serve as the missing unit that translates black-box predictions into downstream robustness.

2. Unified robust framework for prediction-driven decision-making

The same score-calibrated interface supports reliability guarantees, acceptable-target selection, and fragility analysis around fixed predictors.

3. ConfRO–ConfRS equivalence and interpretation

Under regularity conditions, ConfRO and ConfRS are two parameterizations of one robustness frontier, and fragility interprets the marginal cost of robustness.

Thank you!

Lingjie Zhao, Hansheng Jiang, Wei Qi.
Conformal Robustness in Prediction-Driven Decision-Making, SSRN 5338354.