Overview of the **main criticisms DRO researchers raise against predict-then-optimize approaches**, grouped by conceptual level.

## What Bootstrapping Time Series Actually Does

Bootstrapping a time series (e.g., via **block bootstrap**, **stationary bootstrap**, or **model-based residual bootstrap**) aims to generate **resampled trajectories** that reflect the uncertainty and dependence structure of the data.

So, instead of a single estimated distribution P^, you get many simulated versions P^1​,P^2​,…,P^B​, which approximate sampling variability.

This can be used to:

* Estimate **forecast uncertainty**,
* Compute **confidence or prediction intervals**,
* Evaluate **decision robustness** across resampled series.

### **1. Objective Misalignment**

**Criticism:**  
PtO optimizes the prediction model for statistical accuracy (e.g., minimizing MSE or cross-entropy), not for decision quality under uncertainty.

**Explanation:**  
Even small errors in the predicted mean or quantiles can lead to poor downstream decisions if the cost function is highly nonlinear or asymmetric.  
In contrast, DRO directly incorporates the decision objective and uncertainty into a single optimization problem (In recent years, distributionally robust optimization (DRO) has emerged as a promising data-driven approach to tackle these challenges. DRO aims to find a minimax robust optimal decision that minimizes the expected loss under the most adverse distribution within a predefined set of relevant distributions, known as an ambiguity set.)

Min\_x sup\_ P∈P ​E\_P ​[c(x,ξ)] 

where P is an ambiguity set around the empirical distribution.

**In short:** PtO can be statistically optimal but decision-suboptimal, because it separates prediction and optimization.

|  |  |
| --- | --- |
| **1. Objective misalignment** (i.e. prediction error minimized ≠ decision/regret error) | • Elmachtoub, A. N., & Grigas, P. (2022). *Smart “Predict, then Optimize”*. *Management Science*, 68(1), 9-26. [Emergent Mind+1](https://www.emergentmind.com/papers/1710.08005?utm_source=chatgpt.com)  • *Generalization Bounds in the Predict-then-Optimize Framework*. O. El Balghiti, A. Elmachtoub, P. Grigas, Ambuj Tewari. NeurIPS 2019. [papers.nips.cc](https://papers.nips.cc/paper/9585-generalization-bounds-in-the-predict-then-optimize-framework?utm_source=chatgpt.com)  • *Risk Bounds and Calibration for a Smart Predict-then-Optimize Method*. Heyuan Liu & Paul Grigas, 2021 (arXiv / math.OC) [arXiv](https://arxiv.org/abs/2108.08887?utm_source=chatgpt.com) |

**Key refs & quotes**

* Elmachtoub & Grigas — *Smart “Predict, then Optimize”* (arXiv / Management Science).  
  Quote: “machine learning tools are intended to minimize prediction error and do not account for how the predictions will be used in the downstream optimization problem.” ([arXiv](https://arxiv.org/abs/1710.08005?utm_source=chatgpt.com" \o "Smart \"Predict, then Optimize\"))  
  *Annotation:* This paper introduces the SPO loss precisely because standard predictive losses are not aligned with decision loss; it formalizes the misalignment and proposes a decision-aware surrogate (SPO+). ([INFORMS Pubs Online](https://pubsonline.informs.org/doi/10.1287/mnsc.2020.3922?utm_source=chatgpt.com))
* El Balghiti, Elmachtoub, Grigas, & Tewari — *Generalization Bounds in the Predict-then-Optimize Framework* (NeurIPS 2019).  
  Quote: “we provide an assortment of generalization bounds for the SPO loss function.” ([papers.nips.cc](https://papers.nips.cc/paper/9585-generalization-bounds-in-the-predict-then-optimize-framework?utm_source=chatgpt.com))  
  *Annotation:* Provides theoretical analysis showing how predictive loss vs. decision loss generalize differently — further evidence that optimizing prediction error need not minimize decision regret. ([ambujtewari.com](https://www.ambujtewari.com/research/balghiti19generalization.pdf?utm_source=chatgpt.com))

### **Bootstrap and Objective Misalignment**

✖ **Does not address:** The fact that PtO still optimizes the wrong loss.

Bootstrapping may improve uncertainty estimation, but unless you incorporate the **decision objective into the resampling-based training**, you’re still separating prediction and optimization.  
You would need decision-focused learning or stochastic programming for that.

### **2. Lack of Robustness to Distributional Shift**

**Criticism:**  
PtO assumes that the predictive model generalizes perfectly — i.e., that the distribution seen in training is the same as in deployment.

**Explanation:**  
When there is a distributional shift (e.g., new demand patterns, new customers, changing prices), PtO decisions can degrade sharply.  
DRO, on the other hand, explicitly guards against worst-case shifts within a predefined ambiguity set, often defined via Wasserstein distance or f-divergences.

**In short:** PtO is brittle under data mismatch; DRO is robust by design.

|  |  |
| --- | --- |
| **2. Lack of robustness to distributional shift** | • *Minimax Regret Optimization for Robust Machine Learning under Distribution Shift*, Alekh Agarwal & Tong Zhang, COLT / NeurIPS -type setting (PMLR) 2022. [Proceedings of Machine Learning Research](https://proceedings.mlr.press/v178/agarwal22b.html?utm_source=chatgpt.com)  • Also works on DRO in ML (e.g. group DRO, domain adaptation) implicitly argue that PtO that uses only training distribution fails under shift.  • Review “Predict-and-Optimize Techniques for Data-Driven Optimization Problems: A Review”, Neural Processing Letters 2025, which discusses shortcomings of decoupled PO vs methods that account for uncertainty in deployment. [SpringerLink](https://link.springer.com/article/10.1007/s11063-025-11746-w?utm_source=chatgpt.com) |

**Key refs & quotes**

* Mohajerin Esfahani & Kuhn — *Data-driven Distributionally Robust Optimization using the Wasserstein metric* (Math. OR / 2018).  
  Quote: “we seek decisions that perform best in view of the worst-case distribution within this Wasserstein ball.” ([SpringerLink](https://link.springer.com/article/10.1007/s10107-017-1172-1?utm_source=chatgpt.com))  
  *Annotation:* Formal DRO construction showing how decision rules can explicitly guard against distributional shift — directly contrasts PtO’s reliance on a single estimated distribution. ([arXiv](https://arxiv.org/abs/1505.05116?utm_source=chatgpt.com" \o "Data-driven Distributionally Robust Optimization Using the ...))
* (Recent) *Decision-Focused Evaluation of Worst-Case Distribution Shift* — discussion and examples showing how PtO can fail under adversarial shifts.  
  Quote: “Task-aware/decision-aware methods can be tested by worst-case shifts in distribution.” ([arXiv](https://arxiv.org/html/2407.03557v1?utm_source=chatgpt.com" \o "Decision-Focused Evaluation of Worst-Case Distribution Shift))  
  *Annotation:* Recent work framing distributional shift as a decision-facing worst-case problem, reinforcing the DRO critique of PtO under deployment shifts. ([arXiv](https://arxiv.org/html/2407.03557v1?utm_source=chatgpt.com" \o "Decision-Focused Evaluation of Worst-Case Distribution Shift))

### **Bootstrap and Distributional Shift (Nonstationarity)**

✖ **Does not address:** Changes in the data-generating process.

Bootstrapping assumes that the observed series is representative of the future (stationary or at least weakly dependent).  
If your system undergoes **structural breaks**, **regime changes**, or **policy shifts**, bootstrap resamples just replicate past patterns — they don’t anticipate future distributional shifts.

→ DRO, in contrast, explicitly models adversarial or worst-case deviations, so it’s more robust to this.

### **3. Ignorance of Uncertainty Quantification**

**Criticism:**  
PtO typically produces point forecasts and ignores uncertainty in the predictions.

**Explanation:**  
Decisions like inventory stocking or portfolio allocation require understanding not just expected outcomes, but also variance and tail risk.  
DRO treats the entire distribution of the uncertain parameters as the decision input, not just a point estimate.

**In short:** PtO discards uncertainty; DRO formalizes it.

|  |  |
| --- | --- |
| **3. Ignorance of uncertainty quantification** | • Many works under DRO and decision-focused learning point out that PtO gives point estimates only; e.g. *Generalization Bounds in the Predict-then-Optimize Framework* again, for understanding how the loss of decision errors depends on true parameter uncertainty. [papers.nips.cc](https://papers.nips.cc/paper/9585-generalization-bounds-in-the-predict-then-optimize-framework?utm_source=chatgpt.com)  • Also, the *Smart Predict, then Optimize* original SPO paper discusses “misspecification” to show how models optimized for pure prediction error can perform poorly in decision‐making. [arXiv+1](https://arxiv.org/abs/1710.08005?utm_source=chatgpt.com) |

**Key refs & quotes**

* Elmachtoub & Grigas — *Smart “Predict, then Optimize”*.  
  Quote: “SPO loss function which measures the decision error induced by a prediction.” ([arXiv](https://arxiv.org/abs/1710.08005?utm_source=chatgpt.com" \o "Smart \"Predict, then Optimize\"))  
  *Annotation:* By defining a decision loss, SPO highlights the need to go beyond point forecasts — i.e., to measure how prediction uncertainty maps to decision regret. ([optimization-online.org](https://optimization-online.org/wp-content/uploads/2018/12/6398.pdf?utm_source=chatgpt.com))
* Liu & Grigas — *Risk Bounds and Calibration for a Smart Predict-then-Optimize Method* (2021).  
  Quote: “we develop risk bounds and uniform calibration results for the SPO+ loss.” ([arXiv](https://arxiv.org/abs/2108.08887?utm_source=chatgpt.com" \o "Risk Bounds and Calibration for a Smart Predict-then-Optimize Method))  
  *Annotation:* This paper extends SPO by deriving statistical risk and calibration results that connect surrogate optimization to the underlying decision risk — a step toward quantifying uncertainty’s effect on decisions. ([arXiv](https://arxiv.org/pdf/2108.08887?utm_source=chatgpt.com" \o "Risk Bounds and Calibration for a Smart Predict-then- ...))

### **Bootstrap and Uncertainty Quantification**

✔ **Addresses:** PtO’s ignorance of uncertainty.

Bootstrapping gives an empirical distribution over forecasts or model parameters.  
You can then make distribution-aware decisions, e.g.:

min⁡x1B∑b=1Bc(x,ξ(b)),\min\_x \frac{1}{B} \sum\_{b=1}^B c(x, \xi^{(b)}),xmin​B1​b=1∑B​c(x,ξ(b)),

or minimize a quantile or CVaR over bootstrapped outcomes.

→ This explicitly accounts for **forecast uncertainty**, a step toward the DRO philosophy.

### **4. Overconfidence and Lack of Performance Guarantees**

**Criticism:**  
PtO provides no guarantees on out-of-sample performance or regret.

**Explanation:**  
Even if a predictive model achieves low test error, there is no formal bound on how suboptimal the induced decision is when the true distribution deviates from the estimated one.  
DRO provides worst-case guarantees and out-of-sample performance bounds — the main selling point of the framework.

**In short:** PtO decisions lack robustness certificates; DRO provides them.

|  |  |
| --- | --- |
| **4. Overconfidence and lack of performance guarantees / regret bounds** | • *Risk Bounds and Calibration for a Smart Predict-then-Optimize Method* (Liu & Grigas 2021) gives theoretical bounds transferring excess risk in the SPO+ surrogate to decision loss. [arXiv](https://arxiv.org/abs/2108.08887?utm_source=chatgpt.com)  • *Generalization Bounds in the Predict-then-Optimize Framework*, El Balghiti et al. (2019) gives generalization bounds for the decision loss. [papers.nips.cc](https://papers.nips.cc/paper/9585-generalization-bounds-in-the-predict-then-optimize-framework?utm_source=chatgpt.com) |

**Key refs & quotes**

* Liu & Grigas — *Risk Bounds and Calibration for a Smart Predict-then-Optimize Method* (2021).  
  Quote: “we show that the empirical minimizer of the SPO+ loss achieves low excess true risk with high probability.” ([arXiv](https://arxiv.org/abs/2108.08887?utm_source=chatgpt.com" \o "Risk Bounds and Calibration for a Smart Predict-then-Optimize Method))  
  *Annotation:* Gives formal excess-risk bounds translating surrogate loss minimization into decision guarantees — addressing the DRO community’s demand for performance certificates. ([papers.neurips.cc](https://papers.neurips.cc/paper/2021/file/b943325cc7b7422d2871b345bf9b067f-Paper.pdf?utm_source=chatgpt.com))
* El Balghiti et al. — *Generalization Bounds in the Predict-then-Optimize Framework*.  
  Quote: “derive bounds based on the Natarajan dimension.” ([papers.nips.cc](https://papers.nips.cc/paper/9585-generalization-bounds-in-the-predict-then-optimize-framework?utm_source=chatgpt.com))  
  *Annotation:* Supplies generalization theory for decision loss, which is necessary to provide any claim that PtO solutions won’t catastrophically fail out of sample. ([ambujtewari.com](https://www.ambujtewari.com/research/balghiti19generalization.pdf?utm_source=chatgpt.com))

### **Bootstrap and Partial Robustness to Sampling Variability**

✔ **Partially addresses:** Lack of robustness to sampling noise.

By simulating data perturbations consistent with your estimated process, bootstrapping can reveal **how stable your decision is to sample variation** — roughly analogous to a data-driven ambiguity set centered around your empirical distribution.

→ In this sense, bootstrap resamples can approximate the **inner supremum over distributions** in DRO, though informally.

### **Bootstrap and Some Protection Against Overconfidence**

✔ **Addresses:** Overconfidence and lack of performance guarantees (to a degree).

Bootstrap distributions allow you to estimate **out-of-sample regret or risk bounds empirically**, e.g., via percentile or bias-corrected intervals for expected cost.

→ It does not yield formal worst-case guarantees, but it gives **empirical performance envelopes**.

### **5. Inefficiency in Data Usage**

**Criticism:**  
PtO uses data inefficiently because it ignores how data affects the optimal decision boundary.

**Explanation:**  
In many practical problems, only parts of the predictive distribution affect the optimal decision (e.g., the quantile relevant for a newsvendor problem).  
DRO and related end-to-end approaches (decision-focused learning) exploit this by weighting data according to decision sensitivity.

**In short:** PtO trains on the wrong loss; DRO (or end-to-end learning) uses data in a decision-aware way.

|  |  |
| --- | --- |
| **5. Data inefficiency / only parts of the predictive distribution matter** | • The SPO framework (Elmachtoub & Grigas) shows that even simpler models (linear) trained with a decision-aware loss (SPO+) can outperform more flexible models trained for pure predictive accuracy, especially under misspecification. This implies more efficient use of data for decision performance. [arXiv+1](https://arxiv.org/abs/1710.08005?utm_source=chatgpt.com)  • *A Note on Task-Aware Loss via Reweighing Prediction Loss by Decision-Regret*, Connor Lawless & Angela Zhou, 2022. They propose reweighting prediction error by decision regret to improve efficiency. [arXiv](https://arxiv.org/abs/2211.05116?utm_source=chatgpt.com) |

**Key refs & quotes**

* Elmachtoub & Grigas — *Smart “Predict, then Optimize”*.  
  Quote: “linear models trained using SPO+ loss tend to dominate random forest algorithms” (in misspecified settings). ([arXiv](https://arxiv.org/abs/1710.08005?utm_source=chatgpt.com" \o "Smart \"Predict, then Optimize\"))  
  *Annotation:* Shows that decision-aware training can extract the data components that matter for the downstream optimization, giving better decision performance with simpler models (i.e., more efficient use of data). ([optimization-online.org](https://optimization-online.org/wp-content/uploads/2018/12/6398.pdf?utm_source=chatgpt.com))
* Lawless & Zhou — *A Note on Task-Aware Loss via Reweighing Prediction Loss by Decision-Regret* (2022).  
  Quote: “we propose a decision-aware version of predict-then-optimize” using reweighting by decision regret. ([arXiv](https://arxiv.org/abs/2211.05116?utm_source=chatgpt.com" \o "A Note on Task-Aware Loss via Reweighing Prediction Loss by Decision-Regret))  
  *Annotation:* Proposes practical reweighting to focus the predictor on decision-relevant errors — a direct response to critiques about data inefficiency in vanilla PtO. ([arXiv](https://arxiv.org/pdf/2211.05116?utm_source=chatgpt.com" \o "A Note on Task-Aware Loss via Reweighing Prediction ...))

### **6. Poor Handling of Rare but Costly Events**

**Criticism:**  
PtO often underestimates the importance of rare, high-impact events.

**Explanation:**  
Because predictive losses (like MSE) focus on average performance, rare events that strongly affect decision costs (e.g., stockouts, system failures) may be ignored.  
DRO explicitly accounts for these by penalizing distributions that deviate in such tails.

**In short:** PtO underweights tails; DRO emphasizes them.

|  |  |
| --- | --- |
| **6. Poor handling of rare but costly events / tail risks** | • While not always explicitly in PtO vs DRO literature, many DRO papers are motivated by wanting to guard against worst-case distributions—including rare events. For example, the DRO literature (various) around CVaR, worst‐case f-divergence, etc.  • Also, *Predict-and-Optimize Techniques for Data-Driven Optimization Problems: A Review* mentions how small prediction errors in tails can lead to large decision losses. [SpringerLink](https://link.springer.com/article/10.1007/s11063-025-11746-w?utm_source=chatgpt.com) |

**Key refs & quotes**

* Rockafellar & Uryasev — *Conditional Value-at-Risk for General Loss Distributions* (Math. Programming / 2000/2002 and subsequent work).  
  Quote: “Conditional Value-at-Risk” (classic formulation for tail-risk minimization). ([sites.math.washington.edu](https://sites.math.washington.edu/~rtr/papers/rtr187-CVaR2.pdf?utm_source=chatgpt.com))  
  *Annotation:* CVaR is the canonical way to incorporate tail risk directly into the objective; DRO similarly targets worst-case tail behavior, while PtO (MSE) tends to ignore tails. ([sites.math.washington.edu](https://sites.math.washington.edu/~rtr/papers/rtr187-CVaR2.pdf?utm_source=chatgpt.com))
* Esfahani & Kuhn — *Wasserstein DRO* (2018).  
  Quote: “ambiguity sets contain all distributions that are close to a nominal distribution with respect to the prescribed probability metric.” ([SpringerLink](https://link.springer.com/article/10.1007/s10107-017-1172-1?utm_source=chatgpt.com))  
  *Annotation:* By constructing ambiguity sets, Wasserstein-DRO can explicitly inflate the mass in adverse tails — a capability PtO lacks unless tail models are explicitly built. ([arXiv](https://arxiv.org/abs/1505.05116?utm_source=chatgpt.com" \o "Data-driven Distributionally Robust Optimization Using the ...))

### **Bootstrap and Tail Event Emphasis**

✖ **Partially addresses:** Rare but costly events.

Bootstrapping reweights existing data, so rare events that didn’t appear much in the historical record will still be underrepresented.  
You’d need tail modeling (e.g., EVT or stress testing) or adversarial perturbations (as in DRO) to fully capture tail risk.

### **7. Conceptual Disconnect Between Prediction and Decision (v.1)**

**Criticism:**  
PtO treats forecasting and optimization as two independent pipelines, while in many problems (e.g., operations, energy, logistics), prediction and decision are tightly coupled.

**In short:** PtO is modular but disconnected; DRO is integrated and coherent.

|  |  |
| --- | --- |
| **7. Conceptual disconnect between prediction and decision / two-stage vs joint** | • Again, *Smart “Predict, then Optimize”* (Elmachtoub & Grigas 2017) presents SPO to directly tie prediction and decision. [arXiv+1](https://arxiv.org/abs/1710.08005?utm_source=chatgpt.com)  • *Decision Trees for Decision-Making under the Predict-then-Optimize Framework*, Elmachtoub, Liang & McNellis (2020) PMLR. They build decision trees whose splits are chosen based on decision loss rather than just predictive error. [Proceedings of Machine Learning Research+1](https://proceedings.mlr.press/v119/elmachtoub20a?utm_source=chatgpt.com) |

**Key refs & quotes**

Elmachtoub & Grigas — *Smart “Predict, then Optimize”*.  
Quote: “we propose a new and very general framework, called Smart ‘Predict, then Optimize’ (SPO), which directly leverages the optimization problem.” ([optimization-online.org](https://optimization-online.org/wp-content/uploads/2018/12/6398.pdf?utm_source=chatgpt.com))  
*Annotation:* The SPO program is an archetypal “join the loop” response to the PtO disconnect: train with decision loss so prediction and decision objectives are aligned. ([INFORMS Pubs Online](https://pubsonline.informs.org/doi/10.1287/mnsc.2020.3922?utm_source=chatgpt.com))

* TeamCore / tutorial notes on decision-focused learning (differentiating through optimization).  
  Quote: “Differentiating through optimization problems is non-trivial ... creating an end-to-end pipeline requires a differentiable 'surrogate' optimization problem.” ([teamcore.seas.harvard.edu](https://teamcore.seas.harvard.edu/learning-loss-functions-predict-then-optimize/?utm_source=chatgpt.com))  
  *Annotation:* Technical challenges of end-to-end methods (discrete decisions, non-differentiability) explain why PtO remains popular despite its disconnect — and motivate surrogate / reweighting approaches. ([teamcore.seas.harvard.edu](https://teamcore.seas.harvard.edu/learning-loss-functions-predict-then-optimize/?utm_source=chatgpt.com))

### **Summary Table**

| **Aspect** | **Predict-then-Optimize** | **Distributionally Robust Optimization** |
| --- | --- | --- |
| Objective | Predict accurately | Decide robustly |
| Uncertainty | Often ignored or reduced to a point | Explicitly modeled as a set |
| Sensitivity to shift | High | Low |
| Guarantees | None | Worst-case bounds |
| Tail events | Underweighted | Emphasized |
| Philosophy | Sequential (two-stage) | Unified (joint) |

**Short synthesis**

* **SPO and its followups** (Elmachtoub & Grigas; El Balghiti et al.; Liu & Grigas) are the central decision-aware literature that explicitly articulates the objective-misalignment critique and propose both surrogates and risk bounds. ([arXiv](https://arxiv.org/abs/1710.08005?utm_source=chatgpt.com" \o "Smart \"Predict, then Optimize\"))
* **DRO literature** (Esfahani & Kuhn; classical robust optimization texts) provides the theoretical and practical frameworks for guarding against distributional shift and tail/worst-case events — the principal alternatives critics advocate. ([SpringerLink](https://link.springer.com/article/10.1007/s10107-017-1172-1?utm_source=chatgpt.com))
* **Practical task-aware methods** (Lawless & Zhou; team tutorials) propose easy-to-use baselines and explain engineering barriers to end-to-end adoption, bridging theory and practice. ([arXiv](https://arxiv.org/pdf/2211.05116?utm_source=chatgpt.com" \o "A Note on Task-Aware Loss via Reweighing Prediction ...))

## ⚖️ Summary

| **Criticism** | **Can Bootstrapping Help?** | **Comment** |
| --- | --- | --- |
| Objective misalignment | ❌ | Still separates prediction from decision |
| Distributional shift | ❌ | Bootstraps assume stationarity |
| Uncertainty quantification | ✅ | Provides empirical forecast distributions |
| Lack of robustness guarantees | ⚠️ | Gives empirical intervals, not formal bounds |
| Data inefficiency | ⚠️ | Only marginal improvement |
| Rare events | ⚠️ | Limited by data availability |
| Predict–decision disconnect | ❌ | Still two-stage |

## 🧭 Bottom Line

* **Bootstrapping helps with epistemic uncertainty** (uncertainty that arises from a lack of knowledge, finite-sample variability)  
  → useful for empirical robustness analysis and quantile decisions.
* **It does not help with structural uncertainty** (distributional shift, model misspecification)  
  → which DRO explicitly guards against.
* You can view **bootstrap-based stochastic optimization** as a Monte Carlo approximation to the DRO inner supremum when your ambiguity set is around the empirical distribution.

**Philosophical view of misalignment.**  
A recurrent criticism of the predict-then-optimize (PtO) paradigm is the so-called objective misalignment between the statistical loss used to train predictive models and the operational cost that ultimately matters in decision making. From a pragmatic standpoint, however, blaming PtO for misalignment can be interpreted as wishing that the world’s data-generating process conformed more closely to the decision maker’s objective function. In this sense, the problem may lie less in the decoupling between prediction and optimization than in the brittleness or misspecification of the downstream model itself. The decision-focused learning literature responds by redefining the training criterion to reflect the operational consequences of predictive errors, whereas the robust optimization perspective instead accepts the world’s variability as given and seeks solutions that perform well across plausible distributions. The tension between these views captures a fundamental philosophical divide: whether misalignment reflects a modeling deficiency that learning can correct, or an intrinsic property of a world that cannot—and perhaps should not—be coerced into alignment.

**Philosophical view of misalignment.**  
A recurrent criticism of the predict-then-optimize (PtO) paradigm is the so-called objective misalignment between the statistical loss minimized during prediction and the operational cost ultimately optimized downstream (Elmachtoub and Grigas 2022; Bengio et al. 2020). From a classical operations research perspective, however, attributing performance degradation to such misalignment can be seen as implicitly assuming that the world’s data-generating process should conform to the decision maker’s cost function. In this sense, the underlying problem may lie less in the decoupling of prediction and optimization than in the brittleness or misspecification of the downstream model itself (Bertsimas and Kallus 2020). The decision-focused learning literature addresses the issue by redefining the training objective to reflect the operational consequences of predictive errors (Donti, Amos, and Kolter 2017; Mandi et al. 2020), whereas the robust optimization and distributionally robust optimization traditions accept uncertainty as an intrinsic property of the environment and seek prescriptions that perform reliably across plausible distributions (Ben-Tal et al. 2009; Delage and Ye 2010). The tension between these positions reflects a deeper philosophical divide: whether misalignment represents a modeling deficiency that learning can correct, or a structural feature of an uncertain world that cannot—and perhaps should not—be forced into alignment.

### ✅ **References (suggested list for the paper)**

* Ben-Tal, A., El Ghaoui, L., & Nemirovski, A. (2009). Robust Optimization. Princeton University Press.
* Bertsimas, D., & Kallus, N. (2020). “From predictive to prescriptive analytics.” Management Science, 66(3), 1025–1044.
* Delage, E., & Ye, Y. (2010). “Distributionally robust optimization under moment uncertainty with application to data-driven problems.” Operations Research, 58(3), 595–612.
* Donti, P., Amos, B., & Kolter, J. Z. (2017). “Task-based end-to-end model learning in stochastic optimization.” NeurIPS.
* Elmachtoub, A. N., & Grigas, P. (2022). “Smart ‘predict, then optimize.’” Operations Research, 70(5), 2952–2970.
* Bengio, Y., Lodi, A., & Prouvost, A. (2020). “Machine learning for combinatorial optimization: a methodological tour d’horizon.” European Journal of Operational Research, 290(2), 405–421.
* Mandi, J., Bucarey, V., et al. (2020). “Smart predict-and-optimize for hard combinatorial problems.” AAAI Conference on Artificial Intelligence.

**Practical implications.**  
The philosophical divide between decision-focused and robust approaches has concrete methodological implications. In data-rich environments with stable and well-specified cost structures, decision-focused learning can yield substantial gains by tailoring predictions to the aspects of uncertainty that most affect downstream performance. Conversely, in settings characterized by structural change, model misspecification, or limited data, robust or distributionally robust formulations provide a more prudent hedge against the fragility of decision-aware learning (Bertsimas et al. 2019; Esfahani and Kuhn 2018). In practice, the choice between the two should reflect not only the availability of predictive accuracy but also the decision maker’s tolerance for model risk and the degree of confidence in the operational objective itself. As such, the trade-off is less about algorithmic preference than about epistemic stance—whether one assumes that the decision model is essentially correct and should guide learning, or that it is approximate and therefore must be protected from overcommitment to any single predictive representation.

### Additional references to complement the above:

* Bertsimas, D., Gupta, V., & Kallus, N. (2019). “Data-driven robust optimization.” Mathematical Programming, 167(2), 235–292.
* Esfahani, P. M., & Kuhn, D. (2018). “Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations.” Mathematical Programming, 171(1–2), 115–166.

**Bootstrapped SAA as implicit distributional robustness.**  
An alternative to explicit distributionally robust formulations is to propagate predictive uncertainty through bootstrapped scenario generation and solve the resulting problem via Sample Average Approximation (SAA). In this approach, resampling the historical data produces an ensemble of empirical distributions whose variability reflects sampling error in the estimated model. Solving the optimization problem over the aggregated bootstrap scenarios thus yields a decision that performs well across a spectrum of plausible distributions, effectively conferring a form of implicit robustness. Although such bootstrapped SAA schemes lack the formal ambiguity sets and worst-case guarantees of distributionally robust optimization, they approximate the same underlying intent—hedging against estimation error and model misspecification by broadening the support of the empirical distribution. In practice, this simulation-based approach can be computationally simpler and more transparent, while achieving much of the practical protection that motivates explicit DRO formulations (cf. Efron and Tibshirani 1994; Shapiro, Dentcheva, and Ruszczyński 2021).

**References**

* Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap.* Chapman & Hall/CRC.
* Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2021). *Lectures on Stochastic Programming: Modeling and Theory* (3rd ed.). SIAM.

**Application to time-series forecasting.**  
In time-series settings, bootstrapped SAA acquires additional interpretive force. Resampling the residuals or blocks of observations generates alternative realizations of the underlying stochastic process, thereby capturing both estimation error in model parameters and uncertainty about temporal dependence. Each bootstrapped forecast trajectory can be viewed as a plausible future consistent with the observed dynamics, and optimization across their ensemble yields policies that are less sensitive to transient fluctuations or structural breaks. In effect, the bootstrap extends the empirical support of the predictive distribution along the temporal dimension, producing a data-driven approximation of the ambiguity sets commonly used in distributionally robust optimization. This alignment between resampling-based uncertainty propagation and robust optimization principles makes the approach particularly attractive when analytical characterization of forecast uncertainty is intractable or when the decision environment evolves faster than formal DRO calibration would allow (see Politis and Romano 1994; Paparoditis and Politis 2001).

**Additional references**

* Politis, D. N., & Romano, J. P. (1994). “The stationary bootstrap.” *Journal of the American Statistical Association*, 89(428), 1303–1313.
* Paparoditis, E., & Politis, D. N. (2001). “Tapered block bootstrap.” *Biometrika*, 88(4), 1105–1119.