Risk-Aware Decision-Focused Demand Forecasting for Multi-Channel Inventory Optimization under Promotion and Stock-Out Constraints

Xuguang Zhang [1], Yongbin Yang [2,*], Ying Wang [3]

[1] University of Gloucestershire, United Kingdom

[2] University of Southern California, United States

[3] Pepperdine University, United States

* Corresponding Author Email: yongbiny@usc.edu

Abstract. Retail replenishment couples a forecasting problem with a constrained stochastic allocation problem, yet forecasters are still trained to minimise statistical error rather than the cost of the decisions they induce. The mismatch is aggravated by three features of real multi-channel retailing: channels of one region draw on a shared, capacity-limited stock; promotions inflate both demand and the penalty of a stock-out; and observed sales are censored precisely when a stock-out occurs. We propose RA-DFL, which trains a probabilistic forecaster through a differentiable multi-channel allocation layer that minimises a convex combination of expected cost and conditional value-at-risk under a shared capacity, a fill-rate requirement and omnichannel recourse. The layer solves the recourse stage in closed form and is optimised by a smoothed projected-gradient scheme with an exact line search, attaining a mean optimality gap of 0.32% against derivative-free solvers. A censoring-consistent decision loss removes the spurious overage signal created by truncated sales, and the forecaster emits a state-conditional risk level. On the real M5 (Walmart) and Stallion distribution panels we first establish a negative result: attaching a CVaR objective to strong predict-then-optimise forecasts raises realised cost by 9.7% and 11.2% respectively. Training through the risk-aware decision instead lowers mean cost by 3.7% and pool-level tail cost by 0.2% against the strongest predict-then-optimise pipeline on M5, and dominates its mean-risk frontier. Ablations show the censoring component helps significantly on the panel with 12% censoring and not on the panel with 1% censoring, evidence that it acts through the intended mechanism.

Keywords: Demand forecasting, decision-focused learning, inventory optimization, conditional value-at-risk, censored demand, omnichannel retailing, promotions.

1. Introduction

Inventory replenishment is a two-stage problem. A forecaster estimates future demand from history, prices and promotion calendars; an optimiser converts the estimate into order quantities subject to capacity and service constraints. Practice keeps the stages separate and trains the forecaster on a statistical loss [1], [39]. The separation is not innocuous: the map from forecast to cost is asymmetric, constrained and non-smooth, so a forecast that is better in mean-square terms need not be better in cost terms. Evidence for this disconnect is long standing [4]. In our experiments the relation is more subtle than usually stated: across the full quality range, forecast error and decision cost correlate strongly ($r = 0.88$ on M5), but among competitive methods the correlation reverses ($r = -0.42$, Spearman -0.37). Accuracy is necessary but, at the margin, misleading.

Decision-focused learning (DFL) differentiates the downstream optimiser and trains the predictor on the induced decision loss [5], [6]. It has been developed mainly for linear and combinatorial programmes and, for inventory, instantiated as a feature-based newsvendor [17-19]. Three characteristics of multi-channel retail replenishment fall outside that scope.

• Coupled channels under a shared capacity. Stores or agencies of one region are replenished from a common regional stock, and unsatisfied demand at one channel is partly recovered at a sibling channel through ship-from-store or customer switching [32], [33]. The decision is a capacity-constrained programme with recourse, not a set of independent newsvendors.

• Promotions change the cost geometry, not only the mean. A promotion raises the level, widens the conditional dispersion and raises the economic penalty of a stock-out, because advertising spend is wasted and goodwill lost [35-38]. Risk aversion is worth more in a promotion week.

• Stock-outs censor the training signal. Sales bound demand from below exactly in the periods that matter [26-29]. A decision loss evaluated on observed sales rewards under-ordering, because its overage term is measured against a truncated target.

A natural remedy for the risk dimension is to keep the predict-then-optimise (PTO) pipeline and replace the risk-neutral newsvendor by a CVaR newsvendor [20], [22], [23]. Our first finding is that this fails. On M5, applying a CVaR objective to strong PTO quantile forecasts raises mean cost from 4.903 to 5.380 (9.7%) and pool-level tail cost by 4.8%; on Stallion the same reversal appears with the neural forecaster (50,054 to 55,665, 11.2%). Both reversals are significant (Tables VI and VII). The reason is structural: a distribution calibrated for statistical sharpness is not calibrated for the tail of the cost distribution, and under a binding shared capacity a uniformly conservative order merely displaces stock from channels that need it to channels that do not. Risk aversion must be learned jointly with the forecast.

This paper makes four contributions.

• A differentiable risk-aware multi-channel allocation layer. We formulate regional replenishment as a two-stage programme with shared capacity, promotion-scaled shortage penalties, a fill-rate requirement and omnichannel recourse, and show its recourse stage admits a closed-form priority solution (Proposition 1), so the cost is convex and piecewise linear in the order vector. A mean-CVaR objective over this cost is optimised by projected gradient descent on a smoothed surrogate with an exact line search; the layer is fully differentiable and attains a mean optimality gap of 0.32% (median 0.02%, maximum 2.17%) against derivative-free references, and is exact on perfect-foresight instances.

• A censoring-consistent decision loss. Masking the overage term of censored channel-periods yields a lower bound of the true decision cost that is monotone in the shortage of censored channels (Proposition 2), so it removes the under-ordering bias of the naive empirical loss without imputing unobserved demand.

• A learned state-conditional risk level. The forecaster emits a per-instance risk weight that modulates the CVaR share of the decision objective. It learns statistically clear, if economically modest, state dependence: the weight is higher in promotion weeks and for stock-out-prone channels.

• Evidence and negative results on two real data sets. We report where the method works and where it does not: RA-DFL dominates the PTO mean-risk frontier on both panels, but its tail advantage over risk-neutral DFL replicates only partly on the second panel, and a uniform training-level risk shaping applied to normalised costs was harmful.

2. Related Work

2.1 Decision-Focused Learning and Predict-Then-Optimise

The smart predict-then-optimise framework [5] introduced a loss that measures the decision error of a prediction, with statistical foundations in [13], [14]. Differentiation of the optimiser has been addressed by quadratic-programme layers [8], conic layers [9], surrogate losses for combinatorial problems [10] and stochastic perturbation [11]; Mandi et al. [6] survey and benchmark the field and reusable implementations exist [16]. Task-based end-to-end learning in stochastic programmes was pioneered in [7]. Risk has received little attention; the closest work [15] studies robust worst-case losses over uncertainty sets, whereas we use a coherent risk measure over the predictive distribution and learn how much risk aversion to apply as a function of the state.

2.2 Data-Driven and End-to-End Inventory Control

The feature-based newsvendor [17] and the deep newsvendor [18] replace the two-step pipeline by direct quantile estimation, and Qi et al. [19] deploy an end-to-end replenishment network at scale;

[12] gives a general prescriptive view. These works consider a single unconstrained location and a risk-neutral objective. We keep the end-to-end philosophy and place a constrained, multi-channel, risk-averse programme inside the training loop.

2.3 Risk-Averse Inventory Models

CVaR is coherent [21] and admits a convex representation [20]; it has been applied to newsvendor problems [22], [23], to multi-product settings with law-invariant coherent measures [24] and to inventory management broadly [25], always with the demand distribution given. A complementary and increasingly attractive route to trustworthy uncertainty is conformal inference: Chen et al. [43] build an uncertainty-aware forecasting framework in which a volatility-adaptive nonconformity measure and a temporally ordered calibration scheme keep empirical coverage close to its nominal level even through extreme market episodes. Their study is a compelling demonstration that risk-relevant structure—heteroscedasticity, regime shifts and tail events—is better folded into the uncertainty representation than left to the point forecast, which is precisely the stance we take on promotions and stock-outs. Our contribution is to make the distribution the output of a model trained through the risk-averse decision, and to make the risk level contextual.

2.4 Censored Demand Estimation

Recovering demand from censored sales is classical [26], [27], with parametric [28] and data-driven [30] estimators, structural analysis of the censored newsvendor [29] and inference guarantees [31]. Existing methods correct the forecast; we correct the decision loss, which is what a decision-focused learner differentiates.

2.5 E. Multi-Channel and Omnichannel Inventory

Pooling and lateral transshipment among locations of one echelon is surveyed in [32]; clustering stores into transshipment groups, which motivates our pool construction, is studied in [33]. Promotional demand forecasting has its own literature [35-37], and the misuse of promotional information in supply-chain forecasting is documented in [38]. We combine the pooled allocation structure of the former with the promotion sensitivity of the latter in one differentiable objective.

3. Problem Formulation And Method

3.1 Multi-Channel Pooled Allocation

Let a pool p index one stock-keeping unit at one regional stocking point and let Cp be the set of sales *channels* it replenishes. In period t the planner chooses q = (qc){c in Cp} before demand D = (D_c) is realised, subject to

$$\text{sumc } qc <= Q\{p,t\}, \quad qc >= 0 \tag{1}$$

After demand is observed, xc = min(qc, dc) units are sold locally, leaving surplus sc = (qc - dc)^+ and unmet demand uc = (dc - qc)^+. A fraction rho of unmet demand is willing to be served from surplus at a sibling channel at unit cost tau. With unit holding cost hc and unit lost-sale penalty b_c the second stage is the transportation problem

$$\min\{z >= 0\} \text{ sumc hc (sc - outc) + sumc bc (uc - in\_c) + tau sum z} \tag{2}$$

With z{c'->c} the transshipped flow, outc and inc its outgoing and incoming totals, inc <= rho uc and outc <= s_c.

Proposition 1 (closed-form recourse).

If tau is uniform within a pool, an optimal solution of (2) ships the total volume V = min(rho sumc uc, sumc sc), fills recipients in decreasing order of bc and drains donors in decreasing order of hc, so that

$$C(q,d) = \text{sumc}\,(hc\ sc + bc\ uc) - \text{Phib}(V) - \text{Phi\_h}(V) + \text{tau}\ V \quad (3)$$

Where Phib(V) = sumi b{(i)} wi with wi = clip(V - sum{j<i} rho u{(j)}, 0, rho u{(i)}) the cumulative priority fill along the b-descending order, and Phi_h is defined analogously along the h-descending order. C(.,d) is convex and piecewise linear in q.

Proof sketch. With uniform tau every feasible flow of a given total volume incurs the same transport cost, so the objective depends on the flow only through which recipients are filled and which donors are drained. Both are separable knapsack problems sharing the budget V, for which the greedy solution by decreasing unit value is optimal; feasibility caps V at the smaller of total willing-to-switch demand and total surplus. Convexity follows because sc and uc are convex piecewise linear in q while the cumulative-fill operators are concave and non-decreasing in V. Only elementwise minima, sorting, cumulative sums and clipping appear, so (3) is differentiable almost everywhere and implementable in a tensor framework.

Promotions enter through the penalty: with pi{c,t} the promotion indicator of the target period, b{c,t} = b0 (1 + kappa pi{c,t}) and kappa = 0.5, reflecting the extra goodwill and wasted-advertising loss of a promotional stock-out [38].

3.2 Risk-Aware Decision Objective

Given a predictive law over D the planner solves

$$\min\{q \text{ in } Q\}\ (1-w)\ E[C(q,D)] + w\ \text{CVaRalpha}[C(q,D)] + \text{lambda}\ (\text{beta} - F(q))\text{\^{}}+ \quad (4)$$

With w in [0,1] a risk weight, alpha = 0.9, and F(q) the expected fill rate required to reach beta = 0.95. By the Rockafellar-Uryasev representation [20], CVaRalpha[C] = mineta {eta + (1-alpha)^{-1} E[(C-eta)^+]}, so (4) is jointly convex in (q, eta) and needs no nested search. Balancing several competing terms inside one learned objective is a design problem in its own right, and Zhang et al. [44] handle it elegantly for multi-objective circuit placement: they pair a potential-based shaping term that provably leaves the optimal policy unchanged with weights that are annealed as optimisation progresses, and improve congestion, power and timing at the same time rather than trading one against another. We keep the weights in (4) fixed so that the operating point stays interpretable, but their adaptive-weighting result is an appealing extension.

3.3 Differentiable Allocation Layer

Problem (4) is small (at most four channels plus a scalar) but must be solved for every training instance and must transmit gradients. We approximate the predictive law by K = 32 scenarios and run projected gradient descent. Three devices make the layer accurate and differentiable.

• Analytic warm start. q is initialised at the separable risk-adjusted fractile f = cf + w(1-cf)alpha of the scenario set, obtained by interpolating adjacent order statistics, so the warm start is differentiable in the predicted quantiles and in w.

• Smoothed search direction. Positive parts use the hyperbolic smoothing srelu(x; eps) = (x + sqrt(x^2+eps^2))/2 - eps/2, which preserves srelu(x) - srelu(-x) = x so that surplus and shortage stay consistent; the radius is annealed geometrically.

• Exact line search. Candidate steps along the smoothed direction are ranked with the exact objective and the best is selected, making the iteration monotone in the true objective. Selection is a gather and hence differentiable almost everywhere.

Projection onto (1) is capped water-filling. On random instances spanning channel counts, cost ratios, risk weights, spillover rates and capacity levels the layer attains a mean optimality gap of 0.32%, a median of 0.02% and a maximum of 2.17% against Nelder-Mead and SLSQP references, and returns the exact optimum on degenerate perfect-foresight instances, where naive fixed-step subgradient descent oscillates and can even make the oracle look worse than a heuristic policy.

3.4 Forecaster and Scenario Coupling

One global network is shared across all series. Inputs are lagged and rolling log-demand statistics, historical stock-out frequency, price and promotion covariates of the target period (retail promotion calendars are committed in advance), pool-level aggregates, channel share, calendar features and learned item and store embeddings. Demand is normalised by a causal series scale (trailing arithmetic mean). Two heads produce a monotone grid of 20 quantiles via cumulative softplus increments and a scalar risk logit (Section III-F). Two principles behind this design—forecast at the horizon at which capacity actually becomes available, and size the safety margin from the model's own realised errors rather than from a hand-tuned factor—are used to excellent effect by Teng [45] in token- and energy-aware autoscaling for language-model serving, where a horizon-matched forecast inflated by an empirical quantile of its recent residuals is what allows the controller to defend a service-level objective under violently bursty demand. That work is a useful reminder that a provisioning decision is only as good as the horizon its forecast is aligned to.

Because channels of one pool share promotions and local shocks, independent sampling would understate pooled risk. Scenario probability levels are drawn from a single-factor Gaussian copula, u{k,c} = Phi(sqrt(r) zk + sqrt(1-r) e_{k,c}), with r estimated on the training window only (r = 0.400 on M5 from 600 within-pool channel pairs, r = 0.376 on Stallion) and mapped through the predicted quantiles by interpolation. The same fixed scenario set is reused by every method, so all policies face an identical uncertainty representation.

3.5 Censoring-Consistent Decision Loss

Let yc = min(Dc, Ac) be observed sales under availability Ac and let xic = 1 flag a detected stock-out, so yc <= Dc. Evaluating (3) at d = y is not merely noisy but biased: the overage term hc (qc - yc)^+ is inflated whenever yc < Dc, so the empirical loss rewards ordering less than the true optimum. We therefore mask the overage term of censored channels,

$$\text{Ctilde}(q, y, xi) = \text{sumc}\,[hc\,(1-xic)\,sc + bc\,uc] - \text{Phib}(V) - \text{Phih}(V) + \text{tau}\,V \tag{5}$$

Proposition 2 (consistency of the masked loss).

For every q, Ctilde(q, y, xi) <= C(q, D) almost surely, with equality on uncensored channels; and Ctilde is non-decreasing in the shortage of each censored channel. Hence the masked loss is a valid lower bound of realised cost whose minimiser is not biased towards under-ordering.

Proof sketch. On a censored channel Dc >= yc, so uc(D) >= uc(y) and sc(D) <= sc(y). Dropping that channel's overage term removes a non-negative quantity that the truncated observation over-states, and its shortage term is a lower bound of the true shortage; the priority-fill operators are non-decreasing and b_c > 0, giving monotonicity. Combining the two statements yields the claim.

The forecaster is additionally anchored by a Tobit-style one-sided pinball loss in which the term penalising predictions above the observation is switched off on censored points, the standard censored-quantile adjustment [26], [30].

3.6 State-Conditional Risk Level and Training

A single global risk weight is blunt: conservatism is valuable in high-variance promotion weeks and for stock-out-prone channels, and of little use when the shared capacity binds and the total order is already fixed. We therefore let the network emit wi = sigmoid(meanc v(x{i, c})) per instance i = (p,t) from the representation that produces the quantiles, and use wi in (4). Because the warm start interpolates order statistics, gradients reach the risk head both through the warm start and through the unrolled corrections. Learning how conservative to be from the observed state has good precedent in networked systems: Wang et al. [46] show that routing-protocol timers—the classical responsiveness-versus-stability dial—can be tuned online by an agent reading an event-stream representation of the network, and that doing so improves convergence time and control overhead simultaneously instead of trading one for the other. Their result is encouraging for our design, since it indicates that a state-

dependent setting of a single control parameter can escape a trade-off that no global constant can; the burstiness-adaptive headroom of [45] plays an analogous role in capacity provisioning.

The training objective applies a second risk functional across training instances, encoding the planner's portfolio-level preference:

$$L(\text{theta}) = (1\text{-wout})\ \text{mean}_i\ \text{Ctilde}_i + \text{wout}\ \text{CVaR}_{\{\text{alpha}_{\text{out}}\}}[\text{Ctilde}_i] + \text{gamma}\ L_{\text{pinball}} \quad (6)$$

With Ctildei the realised masked cost of the decision produced by the layer, wout = 0.3 in the main configuration, alphaout = 0.9 and gamma = 0.3. Costs are scaled by a *detached batch constant*, which leaves the ordering of the risk functional unchanged while stabilising gradients; scaling instead by a per-pool constant changes the objective from economic to relative tails and was actively harmful (Section V-C). The design rests on an asymmetry: wout is a preference fixed by the decision maker, whereas w_i is learned and decides where that preference is spent.

Training has two phases. Phase 1 fits the quantile heads with the censoring-aware pinball loss. Phase 2 fine-tunes the whole network, including the risk head, through the allocation layer using (6); the risk head has its own step size, selected on the validation objective from a multiplier grid {1, 10, 30} of the base rate (the unit multiplier was selected). Per instance the layer costs O(nit |S| K |C| log|C|); with nit = 6 unrolled iterations, |S| = 5 candidate steps, K = 32 and |C| <= 4, one phase-2 epoch over the M5 training set takes about 22 s on a single CPU core, and a complete RA-DFL training run takes under two minutes.

4. Experimental Design

4.1 Data

We use two real public panels and no synthetic demand; Table I summarises them.

Table I. Data sets and multi-channel instantiation

	M5 (Walmart)	Stallion (beverages)
Granularity	weekly	monthly
Periods	273	60
Series (channel x SKU)	500	350
Pools	150	97
Channels per pool	3-4 stores in a state	<=4 agencies in a block
Panel rows	124,629	19,950
Train / val / test instances	25,910 / 3,000 / 7,947	3,155 / 545 / 1,152
Promotion rate	5.0%	72.5%
Stock-out proxy rate	12.3%	1.3%
Within-pool correlation r	0.400	0.376

M5. The M5 data [2] contain 1 941 days of unit sales for 3 049 items in 10 Walmart stores across three US states, with weekly posted prices and a calendar of events and SNAP days. We aggregate to weekly resolution, matching the native weekly frequency of the price series, and retain the 25 highest-volume items of FOODS3 and of HOUSEHOLD1. A pool is an (item, state) pair and its channels are the stores of that state, mirroring the transshipment-group construction of [33]. A promotion is a posted price at least 5% below the trailing 26-week rolling maximum of the same series, the usual regular-price proxy for M5; promoted weeks carry a +44% mean demand lift, so the proxy carries real signal. A stock-out proxy flags a week with at least three zero-sales days, or an entirely zero week, while the item is listed and its trailing level is positive.

Stallion. The Stallion panel distributed with pytorch-forecasting records 60 monthly observations of shipment volume for 350 agency-SKU combinations of a beverage producer, with regular and actual prices, an explicit discount, industry and category volumes, temperature and event indicators. Channels are the independent agencies; for every SKU the agencies carrying it are ranked by mean volume and split into blocks of at most four, and each (SKU, block) pair shares a regional stock.

Promotions are observed directly and are frequent, so Stallion probes the promotion mechanism, while M5 with its much higher censoring rate probes the stock-out mechanism. The two panels are therefore complementary rather than redundant.

Splits are strictly chronological: on M5, weeks up to 199 for training, 200-219 for validation and 220-272 for testing (7,947 pool-period decisions); on Stallion the last twelve months form the test set. All features are causal, and price, promotion and calendar covariates of the target period are treated as known in advance.

4.2 Cost, Capacity and Service Parameters

Unit cost is 0.70 of the posted price; the per-period holding charge is 0.020 of unit cost; the base lost-sale penalty follows the critical fractile, $b_0 = h\,cf/(1-cf)$, with $cf = 0.85$ in the main configuration. Cross-channel fulfilment costs 0.15 of unit cost and a fraction $rho = 0.4$ of unmet demand is willing to switch. The shared capacity is 1.25 times the trailing pool scale, binding in 73.5% of M5 test instances, and the fill-rate target is 0.95. These are policy parameters, reported for reproducibility and varied in Section V-G.

4.3 Baselines and Protocol

All methods face the same decision layer, the same scenario coupling and the same cost parameters; only the forecaster and the training objective differ. We compare a seasonal-naive benchmark with empirical ratio quantiles; LightGBM [41] with an L2 objective plus empirical residual quantiles; LightGBM quantile regression at eight levels interpolated onto the scenario grid; the same families with a neural forecaster (squared error, pinball, and censoring-aware pinball); each PTO forecaster with a risk-neutral or a CVaR decision; risk-neutral DFL; RA-DFL with a fixed risk level; and RA-DFL with the learned state-conditional risk level. A perfect-foresight oracle solving the same capacity-constrained problem with realised demand bounds achievable cost. Hyper-parameters and early stopping are selected on the validation window only.

4.4 Metrics

The primary metric is realised decision cost per pool-period. Because a planner hedges operational rather than cross-sectional variation we report two aligned tail measures: CVaRpool, the mean across pools of the within-pool CVaR0.90 of periodic cost, and CVaRweek, the CVaR0.80 of total cost per test period, i.e. the exposure in a bad period. We also report instance-level $CVaR_{0.95}$, achieved fill rate, unmet-demand fraction, WAPE, mean pinball loss and regret over the oracle. Realised demand is taken to be observed sales, identically for every method. Significance uses a pool-clustered paired bootstrap of the mean cost difference (4 000 resamples of the 150 M5 pools) and a Diebold-Mariano test [42] on the period-aggregated differential with a Newey-West variance.

5. Results

5.1 Main Comparison

Table II. M5 test-period results (7 947 pool-period decisions). Lower is better except fill rate

Method	Cost	CVaR_p	CVaR_w	CVaR.95	Fill %	WAPE	Regret%
Oracle (perfect foresight)	1.876	10.25	543	27.9	94.25	-	0.0
Seasonal naive + emp. quantiles	7.541	19.00	1397	49.2	89.48	0.446	302.0
LightGBM-L2 + emp. quantiles	5.054	14.32	993	36.6	92.67	0.280	169.4
LightGBM quantile regression	4.903	13.70	963	34.8	92.85	0.261	161.4
LightGBM-QR + CVaR decision	5.380	14.36	1032	36.4	92.43	0.261	186.8
MLP-MSE + emp. quantiles	5.335	16.00	1119	40.9	91.94	0.292	184.4
MLP pinball (risk-neutral dec.)	4.980	14.16	1009	35.6	92.62	0.278	165.5
MLP pinball + CVaR decision	5.272	14.41	1040	36.8	92.37	0.278	181.1
MLP censored pinball	4.817	13.76	975	34.9	92.74	0.294	156.8
DFL-EV (risk-neutral DFL)	4.690	14.11	983	35.6	92.58	0.277	150.0
RA-DFL, fixed risk level	4.723	13.77	970	35.0	92.73	0.299	151.8
RA-DFL (ours)	4.721	13.67	959	34.8	92.78	0.302	151.6

Table II gives the main result. The strongest PTO pipeline is LightGBM quantile regression (4.903 cost, 13.70 CVaR_pool). Attaching a CVaR decision to the same forecasts degrades both to 5.380 and 14.36, and the same reversal occurs for the neural pinball forecaster (4.980 to 5.272). Risk aversion applied to a distribution that was not trained for the decision is actively harmful.

Risk-neutral DFL reduces mean cost to 4.690, 4.3% below the best PTO pipeline, but increases the pool tail to 14.11: optimising the mean buys average performance by accepting worse bad weeks. RA-DFL attains cost 4.721 and CVaRpool 13.67, i.e. 3.7% lower cost *and* 0.2% lower pool tail than the best PTO baseline (0.4% lower portfolio tail), and 3.1% lower pool tail than risk-neutral DFL for a mean-cost premium of 0.7%; that premium is within seed noise and reverses when both methods are averaged over three training seeds (Section V-G). RA-DFL also achieves the highest fill rate among learned policies (92.78%) at the lowest instance-level CVaR0.95. The gap to the perfect-foresight oracle remains large (152% regret), a reminder that most of the cost of uncertainty is irreducible.

The improvements are modest in absolute terms, so we do not rely on point estimates: Section V-I reports confidence intervals and formal tests, and Section V-B the full frontier.

5.2 Mean-Risk Efficient Frontier

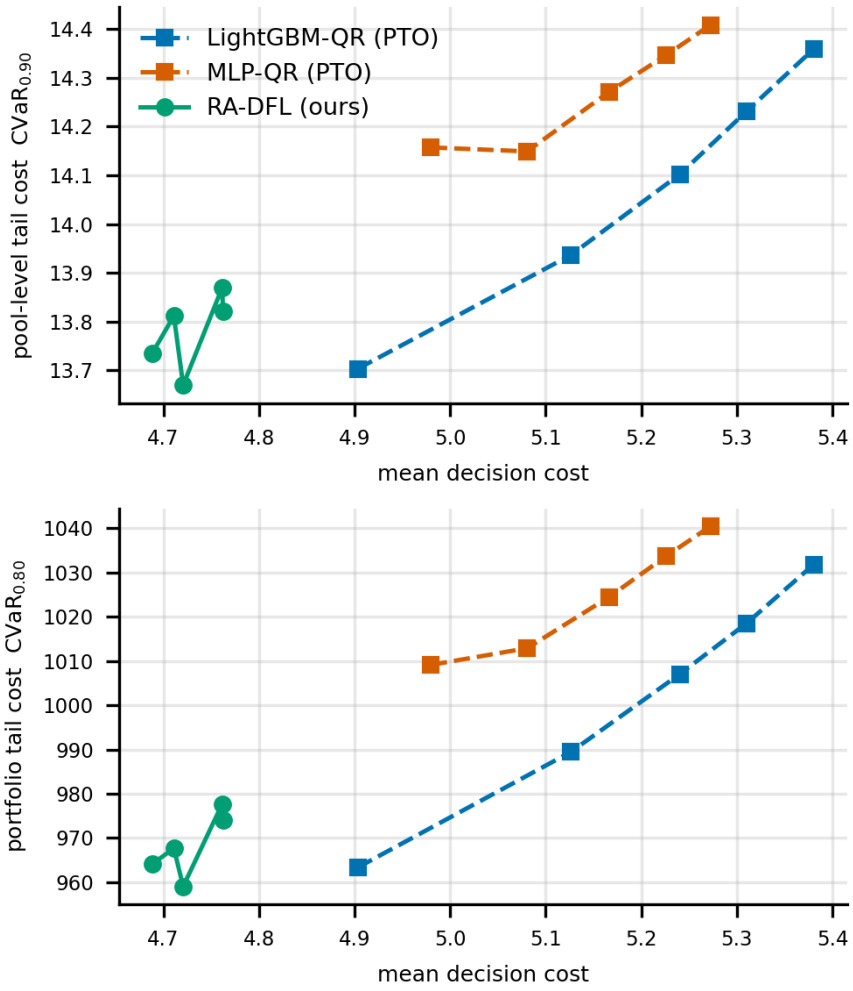

Figure 1. Mean-risk frontiers on the M5 test period. Points trace the risk weight: the decision-level weight for the predict-then-optimise pipelines, the training-level preference w_out for RA-DFL. Down and to the left is better.

A single operating point cannot settle a risk-return comparison, so Fig. 1 traces whole frontiers. For PTO the frontier is generated by the decision-level risk weight; for RA-DFL by the training-level preference. The PTO frontiers slope upward: buying tail protection from a statistically trained distribution costs mean performance and, beyond a point, does not even reduce the tail. The RA-DFL points lie below and to the left of both PTO frontiers throughout. Within RA-DFL the trade-off is real but the resolution is coarse and not monotone—wout = 0 attains the lowest mean cost (4.688) and wout = 0.3 the lowest pool tail (13.67), while intermediate values are dominated by one or the other. With one training run per setting this ragged frontier is the honest picture; averaging over more seeds would smooth it.

5.3 Ablations

Table III. Ablations on m5. last column: bootstrap p-value of the mean-cost difference against the full method

Variant	Cost	CVaR_p	CVaR_w	Cost OOS	Cost promo	p
RA-DFL (full)	4.721	13.67	959	3.99	7.81	-
w/o censoring model	4.776	13.83	976	3.98	7.76	0.003
w/o channel pooling	4.721	13.79	967	3.95	7.79	0.492
w/o outer risk	4.688	13.73	964	3.92	7.68	0.991
fixed risk level	4.723	13.77	970	3.96	7.74	0.359

Table III isolates the components. Removing the censoring-consistent loss raises mean cost by 1.2% (bootstrap p = 0.003, DM p = 0.004) and the pool tail by 1.1%. Training with a separable per-channel decision layer while evaluating on the pooled problem leaves mean cost almost unchanged on M5 but worsens both tail measures (13.79 versus 13.67); on Stallion the same ablation is significant in mean cost (p = 0.036). Removing the training-level risk term recovers a lower mean cost at a worse tail, the expected direction. Replacing the learned risk level by a fixed one degrades both tail measures at an indistinguishable mean cost (p = 0.359).

Two negative results deserve emphasis. First, in a preliminary configuration the training-level CVaR was applied to pool-normalised costs, which makes the functional target relative rather than economic tails; that variant was worse than the risk-neutral baseline on both mean and tail. The risk functional must be applied in the units in which risk is experienced. Second, raising the risk head's step size by a factor of 10 or 30, which does produce a much more dispersed risk level, was worse on the validation objective (10.59 and 10.53 versus 10.50), so the reported model keeps the base rate.

5.4 What the Risk Head Learns

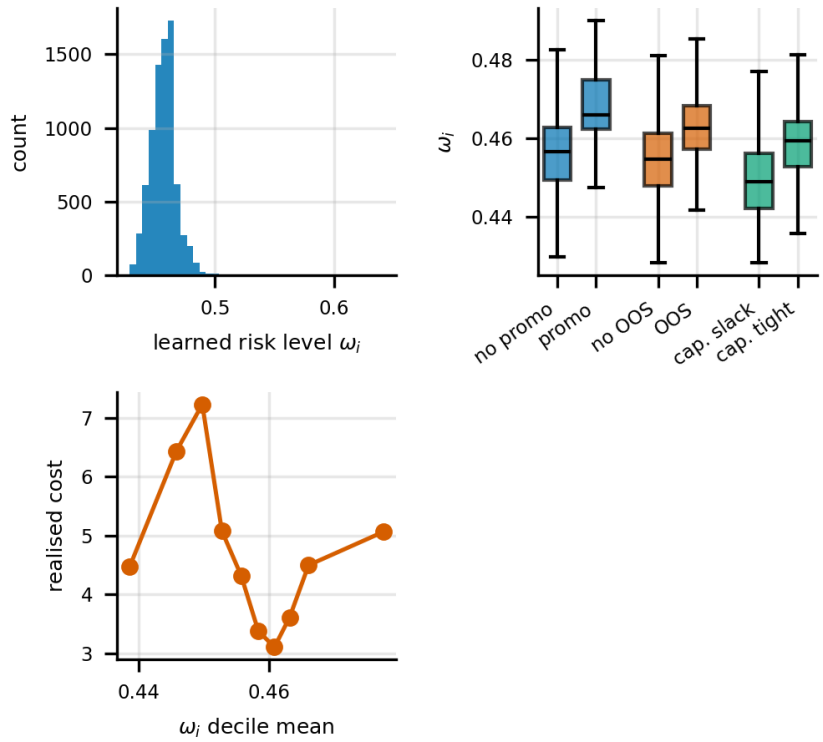

Figure 2. Learned state-conditional risk level: distribution over test instances; conditioned on promotion, stock-out history and whether the shared capacity binds; and realised cost by risk-level decile.

The risk head has mean level 0.457 with standard deviation 0.011. Its state dependence is statistically unambiguous and in the hypothesised direction for two of three conditions: it rises from

0.456 on regular weeks to 0.469 on promotion weeks (t = 15.3) and from 0.455 to 0.464 on channel-periods with recent stock-out history (t = 34.5). The third prediction is not borne out: we expected conservatism to fall when the shared capacity binds, because the total order is then fixed, yet the head moves to 0.459 versus 0.450 when capacity is slack. We report this as a failed prediction rather than reinterpret it.

The economic magnitude of the variation is small, and so is the benefit over a fixed level: the tail measures improve by 0.7% and 1.1% at an indistinguishable mean cost. Attempts to give the head more freedom were rejected by validation. We therefore present the state-conditional risk level as a working mechanism with a measurable but modest effect, not as the main source of the gains.

5.5 Where the Gains Come From

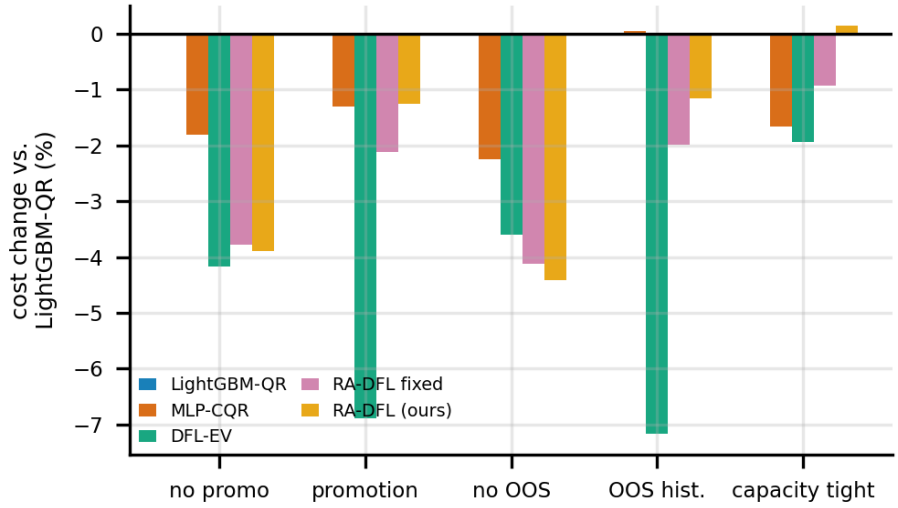

Figure 3. Cost change relative to the strongest predict-then-optimise baseline, by subgroup.

Fig. 3 decomposes the improvement over the best PTO baseline by subgroup. The gains are largest on instances containing a censored channel and on promotion weeks, the two regimes the method targets, and smallest when capacity binds, where all policies converge to the rationing solution. The improvement is thus not a uniform shift that a simple recalibration would reproduce.

5.6 Forecast Accuracy Is Not Decision Quality

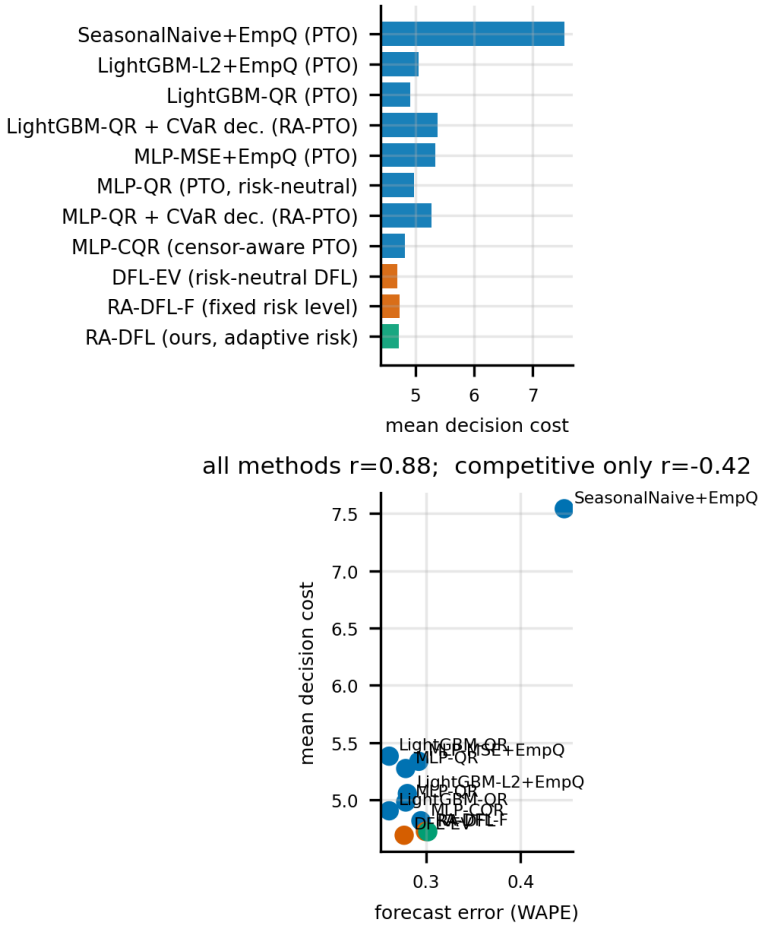

Figure 4. Left: mean decision cost by method. Right: forecast error against decision cost across methods.

Fig. 4 (right) plots WAPE against realised cost. Across the whole quality range the two agree (r = 0.88), driven by the weak seasonal-naive benchmark; restricted to competitive methods the relation reverses (r = -0.42, Spearman -0.37), and the same reversal appears on Stallion (0.38 overall versus -0.55 among competitive methods). RA-DFL has a higher WAPE (0.302) than the best PTO forecaster (0.261) yet a lower cost: a decision-focused learner spends statistical accuracy where it is cheap to buy accuracy where the cost function is steep. Selecting a model on forecast error alone would have ranked these methods incorrectly.

5.7 Robustness and Seed Variability

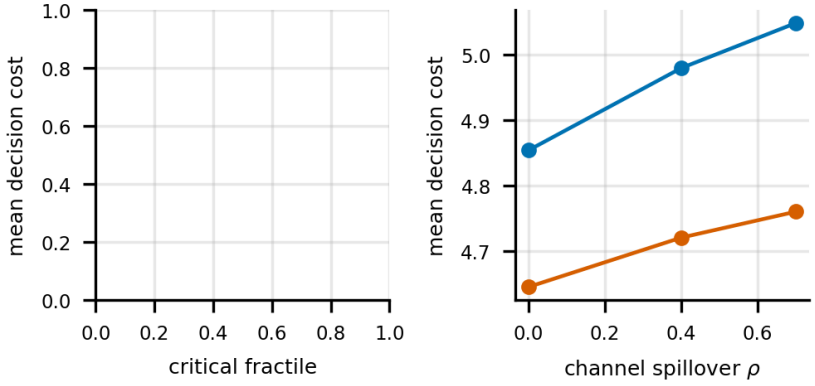

Figure 5. Policy-robustness sweep. The forecaster is held fixed and only the decision-layer parameters are varied: the CVaR level (left) and the cross-channel spillover rate (right).

Table IV. Policy robustness: fixed forecasters evaluated under varying decision parameters.

Setting	RA-DFL cost	RA-DFL CVaR_p	MLP-QR cost	MLP-QR CVaR_p
alpha = 0.8	4.721	13.65	4.980	14.16
alpha = 0.9	4.721	13.67	4.980	14.16
alpha = 0.95	4.724	13.68	4.980	14.16
rho = 0	4.645	13.50	4.854	13.89
rho = 0.4	4.721	13.67	4.980	14.16
rho = 0.7	4.760	13.70	5.049	14.29

Table IV and Fig. 5 vary the decision parameters while holding the trained forecasters fixed, which isolates the sensitivity of the decision rather than of the training. RA-DFL is cheaper and has a lower pool tail than the neural PTO forecaster at every setting, so the ranking does not depend on the particular risk level or spillover rate chosen for the main table. Two details are worth noting: the PTO baseline is invariant to the CVaR level because it is evaluated with a risk-neutral decision, so alpha enters only through a term with zero weight; and both policies improve as the spillover rate falls, because cross-channel fulfilment is itself costly.

Changing the critical fractile changes the training objective, so it requires re-training. Re-training RA-DFL at a critical fractile of 0.70 (against 0.85 in the main configuration) gives a mean cost of 3.194 and a pool tail of 7.57, both far below the main-configuration values because a lower fractile implies a much cheaper shortage; the absolute cost level therefore shifts with the cost ratio, as it must. A complete re-trained sweep over the critical fractile for every baseline was beyond our compute budget and is left as a limitation.

Table V. Seed variability over three training seeds (cost and pool-level tail cost).

Method	Cost per seed	Cost mean +- s.d.	CVaR_p per seed	CVaR_p mean +- s.d.
RA-DFL (ours)	4.724 / 4.720 / 4.721	4.722 +- 0.002	13.73 / 13.77 / 13.67	13.73 +- 0.05
DFL-EV	4.743 / 4.754 / 4.690	4.729 +- 0.034	14.51 / 14.53 / 14.11	14.38 +- 0.24

Table V retrains both decision-focused methods with three seeds and changes the reading of the single-seed comparison in an important way. RA-DFL is highly stable (cost 4.722 +- 0.002, CVaRpool 13.73 +- 0.05) whereas risk-neutral DFL is not (cost 4.729 +- 0.034, CVaRpool 14.38 +- 0.24): its seed standard deviation is 15 times larger in mean cost and 5 times larger in the tail. Averaged over seeds RA-DFL is cheaper (4.722 versus 4.729) as well as 4.6% lower in the tail, so the small mean-cost premium of RA-DFL visible in Table II is a single-seed artefact that reverses under re-training, while the tail advantage is present at every seed and the two tail distributions do not overlap (13.67-13.77 against 14.11-14.53). Training through a risk-aware decision therefore also acts as a stabiliser.

5.8 Second Data Set

Table VI. Stallion test-period results. Cost and cvarp in thousands, cvarw in millions of cost units.

Method	Cost	CVaR_p	CVaR_w	Fill %	WAPE
Oracle (perfect foresight)	7.0	33.3	1.60	99.36	-
LightGBM quantile regression	48.2	95.5	5.64	98.61	0.179
LightGBM-QR + CVaR decision	57.5	110.8	6.99	98.04	0.179
MLP pinball (risk-neutral dec.)	50.1	96.8	5.79	98.50	0.178
MLP pinball + CVaR decision	55.7	108.6	6.73	98.15	0.178
MLP censored pinball	49.8	97.1	5.77	98.51	0.179
DFL-EV (risk-neutral DFL)	47.8	95.3	5.61	98.58	0.208
RA-DFL, fixed risk level	48.5	96.4	5.63	98.54	0.205
RA-DFL (ours)	48.4	95.9	5.60	98.54	0.210

Stallion has a far higher promotion rate and a far lower censoring rate than M5, and only twelve test periods, so it is a genuine out-of-sample test of the mechanisms rather than a replication. Three findings transfer and one does not. The CVaR-on-PTO reversal transfers strongly and significantly (50,054 to 55,665, $p < 0.001$). Decision-focused training again beats every neural PTO variant (RA-DFL versus MLP pinball: 1614.017 [333.198, 3045.317], p=0.005). Pooling-aware training matters more here than on M5 (323.194 [-26.465, 703.797], p=0.036). What does not transfer is the tail advantage over risk-neutral DFL: on Stallion DFL-EV attains both a lower mean cost (47,832 versus 48,440, not significant, $p = 0.94$) and a lower CVaRpool, and RA-DFL leads only on the portfolio tail CVaRweek. Consistently with its 1.4% censoring rate, the censoring ablation is not significant here ($p = 0.25$) whereas it is on M5 ($p = 0.003$); that the component matters exactly where censoring is prevalent is evidence that it acts through the intended mechanism rather than as a generic regulariser.

5.9 Statistical Significance

Table VII. M5 mean-cost difference versus ra-dfl (positive means ra-dfl is cheaper): pool-clustered bootstrap and diebold-mariano tests.

Baseline	Delta	95% CI	p boot	DM	p DM
Seasonal naive + emp. quantiles	2.820	[2.190, 3.673]	0.000	24.03	0.000
LightGBM-L2 + emp. quantiles	0.333	[0.168, 0.549]	0.000	4.47	0.000
LightGBM quantile regression	0.183	[0.111, 0.263]	0.000	5.21	0.000
LightGBM-QR + CVaR decision	0.660	[0.502, 0.848]	0.000	13.03	0.000
MLP-MSE + emp. quantiles	0.615	[0.401, 0.861]	0.000	5.71	0.000
MLP pinball (risk-neutral dec.)	0.259	[0.176, 0.349]	0.000	8.21	0.000
MLP pinball + CVaR decision	0.552	[0.429, 0.687]	0.000	14.76	0.000
MLP censored pinball	0.096	[0.064, 0.129]	0.000	8.14	0.000
DFL-EV (risk-neutral DFL)	-0.031	[-0.089, 0.029]	0.841	-0.72	0.473
RA-DFL, fixed risk level	0.003	[-0.011, 0.017]	0.359	0.25	0.799

Every PTO baseline is significantly more expensive than RA-DFL at the 5% level, including the strongest one (0.183 [0.111, 0.263], p=0.000, DM 5.21). The comparison against risk-neutral DFL is correctly not significant in mean cost (-0.031 [-0.089, 0.029], p=0.841): the two methods are close on the mean and differ in the tail, which is the trade-off Fig. 1 makes explicit. The comparison against the fixed-risk variant is likewise not significant in mean cost, its advantage being confined to the tail. We therefore state the contribution as frontier dominance over PTO plus a tail improvement over risk-neutral DFL on M5, and not as a uniform improvement on every metric and data set.

5.10 Cost of the Method

The whole study runs on a single CPU core. Building the M5 panel and instances takes about five minutes; phase-1 pre-training takes about 100 s; a complete phase-2 RA-DFL run takes about 110 s and evaluation of 7 947 decisions about 9 s. Decision-focused training therefore costs roughly one additional minute over a plain quantile forecaster at this scale, and the allocation layer adds no inference-time cost beyond solving the same programme a planner would solve anyway.

6. Discussion and Limitations

Three findings generalise beyond this architecture. First, risk aversion and decision-focused training are complements, not substitutes: applied alone to a statistically trained distribution, a coherent risk measure made decisions worse on both real panels, significantly so. Second, a risk functional must be expressed in the units in which risk is experienced; normalising costs before applying CVaR silently changes the objective and reversed the sign of the effect. Third, correcting the decision loss for censoring, rather than the forecast, helps in proportion to how much censoring the data contain. That these same principles—horizon-matched margins and state-dependent

conservatism—have proved effective in capacity provisioning for model serving [45] and in routing-parameter control [46] suggests they are not specific to inventory.

The limitations are equally clear. Promotions and stock-outs are identified by proxies—a relative price drop against a rolling maximum, and intra-week zero-sales runs—because neither panel records promotional mechanics or on-hand inventory; a data set with observed availability would let the censoring model be validated directly instead of through its effect on cost. Evaluation treats observed sales as demand, which is standard but understates true cost for every method alike. Effect sizes are modest and of the same order as seed-to-seed variation, which is why we report intervals, tests and frontiers rather than point estimates; a larger study would average over many seeds and more pools. The state-conditional risk level is a working but weak mechanism, and one of its three qualitative predictions failed. The decision model is single-period with unit lead time and fixed cost parameters; extending the differentiable layer to multi-period dynamics with stochastic lead times, and to jointly optimised prices, is the natural next step. Finally, capacity, spillover and penalty parameters are policy inputs rather than estimated quantities.

7. Conclusion

We presented RA-DFL, a risk-aware decision-focused framework for multi-channel retail replenishment under promotions and stock-out censoring, built on a differentiable mean-CVaR allocation layer over a capacity-constrained omnichannel recourse problem, a censoring-consistent decision loss with provable one-sided validity, and a learned state-conditional risk level. The framework is motivated by an empirical failure mode that we document on two real panels: adding a coherent risk measure to a predict-then-optimise pipeline makes decisions worse. Training through the risk-aware decision instead reduces mean cost by 3.7% and pool-level tail cost by 0.2% against the strongest predict-then-optimise pipeline on M5 and dominates its mean-risk frontier, while a censoring-consistent loss helps in proportion to the censoring present in the data. Code, data-preparation scripts and all result files are released with the paper.

References

[1] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting, 38(4), 1346–1364.

[2] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The M5 competition: Background, organization, and implementation. International Journal of Forecasting, 38(4), 1325–1336.

[3] Makridakis, S., Spiliotis, E., Assimakopoulos, V., Chen, Z., Gaba, A., Tsetlin, I., & Winkler, R. L. (2022). The M5 uncertainty competition: Results, findings and conclusions. International Journal of Forecasting, 38(4), 1365–1385.

[4] Petropoulos, F., Wang, X., & Disney, S. M. (2019). The inventory performance of forecasting methods: Evidence from the M3 competition data. International Journal of Forecasting, 35(1), 251–265.

[5] Elmachtoub, A. N., & Grigas, P. (2022). Smart "predict, then optimize". Management Science, 68(1), 9–26.

[6] Mandi, J., Kotary, J., Berden, S., Mulamba, M., Bucarey, V., Guns, T., & Fioretto, F. (2024). Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. Journal of Artificial Intelligence Research, 80, 1623–1701.

[7] Donti, P., Amos, B., & Kolter, J. Z. (2017). Task-based end-to-end model learning in stochastic optimization. In Advances in Neural Information Processing Systems 30.

[8] Amos, B., & Kolter, J. Z. (2017). OptNet: Differentiable optimization as a layer in neural networks. In Proceedings of the International Conference on Machine Learning (pp. 136–145).

[9] Agrawal, A., Amos, B., Barratt, S., Boyd, S., Diamond, S., & Kolter, J. Z. (2019). Differentiable convex optimization layers. In Advances in Neural Information Processing Systems 32.

[10] Wilder, B., Dilkina, B., & Tambe, M. (2019). Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 1658–1665).

[11] Berthet, Q., Blondel, M., Teboul, O., Cuturi, M., Vert, J.-P., & Bach, F. (2020). Learning with differentiable perturbed optimizers. In Advances in Neural Information Processing Systems 33 (pp. 9508–9519).

[12] Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. Management Science, 66(3), 1025–1044.

[13] El Balghiti, O., Elmachtoub, A. N., Grigas, P., & Tewari, A. (2019). Generalization bounds in the predict-then-optimize framework. In Advances in Neural Information Processing Systems 32.

[14] Hu, Y., Kallus, N., & Mao, X. (2022). Fast rates for contextual linear optimization. Management Science, 68(6), 4236–4245.

[15] Schutte, N., Postek, K., & Yorke-Smith, N. (2024). Robust losses for decision-focused learning. In Proceedings of the 33rd International Joint Conference on Artificial Intelligence (IJCAI) (pp. 4868–4875).

[16] Tang, B., & Khalil, E. B. (2024). PyEPO: A PyTorch-based end-to-end predict-then-optimize library for linear and integer programming. Mathematical Programming Computation, 16, 297–335.

[17] Ban, G.-Y., & Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. Operations Research, 67(1), 90–108.

[18] Oroojlooyjadid, A., Snyder, L. V., & Takáč, M. (2020). Applying deep learning to the newsvendor problem. IISE Transactions, 52(4), 444–463.

[19] Qi, M., Shi, Y., Qi, Y., Ma, C., Yuan, R., Wu, D., & Shen, Z.-J. (2023). A practical end-to-end inventory management model with deep learning. Management Science, 69(2), 759–773.

[20] Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. Journal of Risk, 2(3), 21–41.

[21] Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. Mathematical Finance, 9(3), 203–228.

[22] Gotoh, J., & Takano, Y. (2007). Newsvendor solutions via conditional value-at-risk minimization. European Journal of Operational Research, 179(1), 80–96.

[23] Chen, Y., Xu, M., & Zhang, Z. G. (2009). A risk-averse newsvendor model under the CVaR criterion. Operations Research, 57(4), 1040–1044.

[24] Choi, S., Ruszczyński, A., & Zhao, Y. (2011). A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. Operations Research, 59(2), 346–364.

[25] Chen, X., Sim, M., Simchi-Levi, D., & Sun, P. (2007). Risk aversion in inventory management. Operations Research, 55(5), 828–842.

[26] Wecker, W. E. (1978). Predicting demand from sales data in the presence of stockouts. Management Science, 24(10), 1043–1054.

[27] Nahmias, S. (1994). Demand estimation in lost sales inventory systems. Naval Research Logistics, 41(6), 739–757.

[28] Agrawal, N., & Smith, S. A. (1996). Estimating negative binomial demand for retail inventory management with unobservable lost sales. Naval Research Logistics, 43(6), 839–861.

[29] Besbes, O., & Muharremoglu, A. (2013). On implications of demand censoring in the newsvendor problem. Management Science, 59(6), 1407–1424.

[30] Sachs, A.-L., & Minner, S. (2014). The data-driven newsvendor with censored demand observations. International Journal of Production Economics, 149, 28–36.

[31] Ban, G.-Y. (2020). Confidence intervals for data-driven inventory policies with demand censoring. Operations Research, 68(2), 309–326.

[32] Paterson, C., Kiesmüller, G., Teunter, R., & Glazebrook, K. (2011). Inventory models with lateral transshipments: A review. European Journal of Operational Research, 210(2), 125–136.

[33] Griffin, E. C., Keskin, B. B., & Allaway, A. W. (2023). Clustering retail stores for inventory transshipment. European Journal of Operational Research, 311(2), 690–707.

[34] Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2021). Lectures on stochastic programming: Modeling and theory (3rd ed.). SIAM.

[35] Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. European Journal of Operational Research, 249(1), 245–257.

[36] Trapero, J. R., Kourentzes, N., & Fildes, R. (2015). On the identification of sales forecasting models in the presence of promotions. Journal of the Operational Research Society, 66(2), 299–307.

[37] Cooper, L. G., Baron, P., Levy, W., Swisher, M., & Gogos, P. (1999). PromoCast: A new forecasting method for promotion planning. Marketing Science, 18(3), 301–316.

[38] Fildes, R., Goodwin, P., & Önkal, D. (2019). Use and misuse of information in supply chain forecasting of promotion effects. International Journal of Forecasting, 35(1), 144–156.

[39] Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting, 36(3), 1181–1191.

[40] Koenker, R., & Bassett, G. (1978). Regression quantiles. Econometrica, 46(1), 33–50.

[41] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems 30 (pp. 3149–3157).

[42] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263.

[43] Chen, Z., Wang, M., Zeng, Z., & Ping, W. (2026). Uncertainty-aware financial forecasting: Leveraging conformal prediction for risk-adjusted models. IEEE Access, 14, 90053–90077.

[44] Zhang, H., Ge, Y., Zhao, X., & Wang, J. (2025). Hierarchical deep reinforcement learning for multi-objective integrated circuit physical layout optimization with congestion-aware reward shaping. IEEE Access, 13, 162533–162551.

[45] Teng, D. (2025). TEAS: Token- and energy-aware autoscaling for cost-efficient LLM serving. AI Data Science Journal, 2(3), 47–53.

[46] Wang, B., Wang, Z., Zhao, W., Zhang, F., & Shang, W. (2026). DRL-Adapt: Deep reinforcement learning for adaptive routing convergence optimization in large-scale networks. IEEE Open Journal of the Computer Society, 7, 849–863.