



From Demand Signals to Retail Decisions in Business Intelligence Research

Mustafa Musa^{1,*} Raden Nur Rachman Dzakiyullah²

¹ Center of Research and Innovation Management, Universiti Teknikal Malaysia Melaka, Malaysia

² Department of Information System, Universitas Alma Ata, Yogyakarta, Indonesia

Emails: mustafmusa@utem.edu.my · nurrachman.dzakiyullah@almaata.ac.id

Received: August 04, 2026 Revised: September 03, 2026 Accepted: October 05, 2026 ★ Corresponding author

ABSTRACT

Retail analytics has moved beyond improving point forecasts, yet evidence about the value of a forecast for pricing, replenishment, fulfillment, and managerial adoption remains dispersed. This structured review compares 36 journal studies published from 2023 through 2025, identified within a search window ending in January 2026, across four connected areas: demand forecasting (12), pricing and commercial levers (7), inventory and fulfillment (8), and decision systems (9). Each study is examined for its information setting, analytic approach, decision connection, and evidence boundary. The synthesis identifies four recurring discontinuities: sales data can conceal unmet demand; forecast scores can misrank policies; a policy that works in a model may be infeasible in a distributed retail organization; and staff may override recommendations whose value has not been made legible. A decision-centered evaluation structure links data availability, uncertainty estimation, feasible action, adoption, and realized economic outcomes. The paper develops a research agenda around observable-demand corrections, decision-calibrated uncertainty, joint pricing and fulfillment, transportable evaluations, and human-in-the-loop deployment. The review is deliberately bounded rather than a claim of exhaustive coverage or pooled effect sizes. It offers business intelligence researchers a common basis for studying whether analytical improvements survive the operational decisions they are meant to inform.

Keywords: Business intelligence ▪ Retail analytics ▪ Demand forecasting ▪ Prescriptive analytics ▪ Inventory ▪ Decision support

1. INTRODUCTION

A retailer may improve a forecast and still order too much, display the wrong price, or route an online order from the wrong store. These outcomes do not contradict a smaller forecast error. They expose a gap between predicting a quantity and making a constrained decision with that prediction. Business intelligence is useful to a firm when information changes an action and the action improves a defensible business outcome. The sequence is less direct than many model comparisons imply: transactions record sales rather than all demand; promotional responses change over time; inventory

and delivery commitments constrain what can be offered; and managers may reject a recommended markdown.

Recent work makes this an especially useful point at which to take stock. Retail forecasting studies now address censored sales, sparse products, social signals, cross-product relationships, and large probabilistic catalogs [1–4]. Pricing research links historical demand, finite stock, perishability, and the behavior of staff using a decision support system [5–7]. Omnichannel operations studies couple store stock with online allocation and shipping promises [8–10]. Yet these streams often measure success at different places in the decision chain.

A forecast score, a simulated profit comparison, a field adoption rate, and an executive interview are informative, but none can substitute for the others.

This paper asks: *How has recent retail business intelligence research connected demand information to an operational choice, and where does that connection remain untested?* We examine 36 distinct journal studies in four linked streams. The contribution is an evidence map, not an estimate of a common treatment effect. It distinguishes what each stream observes and optimizes; develops a compact analytic structure for comparing studies; and turns recurrent gaps into testable questions. Retail is the focal market because the same sales and inventory signals feed forecasting, price, assortment, replenishment, and fulfillment. Adjacent restaurant and multiretailer distribution studies are included when they clarify a transferable decision problem.

2. REVIEW METHOD

2.1 Scope and selection

We used a structured, purposive review of peer-reviewed journal articles with final issue dates from January 2023 through January 2026. Twelve related topic searches were run in OpenAlex, including retail demand forecasting, pricing and inventory, omnichannel fulfillment, probabilistic forecasting, and decision-focused learning. Each query filtered OpenAlex publication dates to 2023–2025, returned at most the first 100 records by relevance, and was matched to a defined set of 29 journals in forecasting, operations, information systems, marketing, and business analytics. Deduplication by DOI produced 147 candidate records. Targeted publisher and DOI searches added four relevant records, for 151 unique candidates. Title and available abstract screening retained studies addressing a business demand signal, a retail or closely related operating setting, or a specified decision downstream of prediction. The final 36 were selected to cover the connected decision chain, not as a random or exhaustive sample of the 151. The exact queries, journal filter, and screening rules are recorded in the accompanying search protocol.

Exact titles, authors, journal information, and DOIs were checked against Crossref records and, where online and final issue years differed, against publisher issue information. The 36 retained records and coding fields are supplied with the paper. All included studies have final journal issue years from 2023 through 2025; publication years in the text and reference list follow the final issue year.

Selection favors papers that allow comparison across the prediction-to-decision chain. The relevance cap and journal filter mean that this review cannot identify every study published in the period; the category counts are properties of the selected corpus, not publication trends. We did not calculate pooled effect sizes: forecast errors, profit objectives, service measures, and behavioral outcomes differ in population, unit, and counterfactual. Editorial reviews and sources after January 2026 are outside the analytic corpus. Article abstracts and publisher descriptions support the coding; where a publication does not establish actual deployment, the table describes a model or evaluation rather than a realized business gain.

2.2 Comparison structure

The unit of analysis is the study. For each record we distinguish (i) the information setting; (ii) the analytic mechanism; (iii) the decision that could use its output; and (iv) the strongest evidence boundary visible in the study description. These columns appear in Tables 1–4. A forecast can be decision-relevant without estimating a policy's causal gain, while an optimization model can specify a policy without observing its adoption. The boundary column prevents the two claims from being collapsed.

For a retail decision made at time t , let X_t denote information available before action, D_t latent customer demand, A_t an implementable action, and Y_t a subsequent business outcome. Observed sales may satisfy $S_t = \min(D_t, I_t)$ when stock I_t is binding. An analytic system constructs a predictive distribution $\hat{F}_t(D | X_t)$ and recommends

$$A_t^* \in \arg \max_{a \in \mathcal{A}_t} \mathbb{E}_{\hat{F}_t}[u(a, D_t) | X_t], \quad (1)$$

where $\mathcal{A}_t$ includes inventory, labor, delivery, and policy constraints. Ex post business value requires comparison with a credible alternative action, not only with another prediction:

$$V(\pi, \pi_0) = \mathbb{E}[Y_t(\pi(X_t)) - Y_t(\pi_0(X_t))]. \quad (2)$$

The potential outcomes in Equation 2 are generally unobservable together. A simulation reports value under its assumed demand and cost model; a randomized or well-identified field design can make a stronger claim about deployment. This distinction guides the synthesis.

3. RELATED WORK

3.1 Forecasting and demand information

Table 1 maps twelve studies. They do not all solve the same statistical problem. Ye et al. [11] address seasonal retail series with limited observations, while Wellens et al. [12] compare simpler tree-based methods using explanatory variables and connect forecasting to inventory experiments. The latter paper's retailer data cover 4,523 products; its business contribution lies partly in showing how an accessible model can compete when practical features are available. Long et al. [4] instead confront the scale and intermittency of a large e-commerce catalog. A forecasting architecture suited to a small seasonal series cannot be presumed to transfer to hundreds of thousands of sparse product histories.

Several studies question the target itself. When a product is unavailable, zero sales need not mean zero demand. Sousa et al. [1] explicitly model censoring for new fashion products; Adam et al. [13] consider zero- and low-sales products in long-tail demand estimation; and Bekci et al. [14] study learning for a warehouse and multiple stores under censored demand. These problems differ in data-generating process, but all warn against treating observed transactions as a complete record of consumer willingness to buy. Additional information is not automatically useful: Fu and Fisher [2] investigate fashion-related social media signals, whereas Lei et al. [15] transfer category-level information to product predictions and Liu et al. [3] model product relationships with a graph. Their common business question is whether a new sig-

nal has incremental value after accounting for the operating decision and its information cost.

Uncertainty matters at the margin. Wang et al. [16] combine probabilistic forecasts of intermittent demand and explicitly consider inventory utility, while Long et al. [4] develop scalable probabilistic retail forecasts. Store-SKU graph forecasting is tied to replenishment planning in Ma [17]. The setting of Chae et al. [18], a restaurant chain, permits comparison of data sources and learning models across demand horizons. Spatial dependence can add information at aggregate levels: Zhu et al. [19] use an attention-based ConvLSTM to incorporate sales from neighboring retailers and compare the resulting forecasts with benchmarks. This is evidence about prediction, not realized gains from a decision policy. Forecast quality also depends on the human who creates or revises the plan. Fahimnia et al. [20] investigate service-level anchoring under promotions and perishability, separating a demand belief from a desired service target. Taken together, these papers make a case for evaluating a forecast on the marginal decision it changes, with training information fixed at the actual decision time.

3.2 Pricing, promotions, and commercial levers

The seven pricing studies in Table 2 stretch from analytical policy design to observed managerial use. Joint pricing and inventory are modeled for an omnichannel retailer by Liu et al. [21]; Syed et al. [5] examine dynamic pricing for perishable goods through multiple retail cases and the organizational process of data-driven digital transformation. The finite-stock constraint in Li and Zheng [22] and censored offline records in Tang et al. [7] put different limits on price learning. In the first, a merchant trades off exploration and remaining stock while using external information; in the second, historical prices and inventory determine which sales observations are informative. Neither setting permits an analyst to treat prices as randomized inputs merely because they are recorded in a database.

The decision environment also affects implementation. Chehrizi et al. [23] explain rigidity in perishable grocery prices through inventory information frictions and present proof-of-concept evidence from two industry pilots. Caro and Sáez de Tejada Cuenca [6] examine managers' adherence to a clearance markdown decision support system at Zara; recommendations are only consequential if users accept or productively revise them. Oh et al. [24] study return policy, price, and inventory as interacting behavioral levers. These studies expose a missing bridge in many policy comparisons: a predicted revenue-maximizing price is not a realized price, and price changes can alter both demand observations and subsequent inventory. Regional forecasting in Zhu et al. [19] also belongs in this bridge: better aggregate predictions may inform inventory coordination across locations, but the study evaluates forecast performance rather than realized policy outcomes.

3.3 Inventory, assortment, and fulfillment

Table 3 presents eight studies whose operational choice is more explicit. Ship-from-store couples the replenishment decision for physical stores to online order allocation [8, 9]. Arslan et al. [10] add standard shipping contracts and their

delivery commitments; Sun et al. [25] formulate robust inventory policies under demand uncertainty. Kolyaei et al. [26] use reinforcement learning for replenishment and fulfillment over a capacitated multicycle network. These contributions model distinct feasible sets: a policy judged on the same forecast but a different promise, capacity, or stock-sharing rule can have a different economic ranking.

Central information is not always available centrally. In Baron et al. [27], stores hold local information and may accept or reject fulfillment requests; a central optimizer cannot assume complete compliance. Deng et al. [28] couple warehouse placement and demand matching across zones, while Li et al. [29] show why the assortment stocked at an e-commerce warehouse depends on fulfillment costs as well as product popularity. Comparisons confined to fill rate can miss order splitting or spillover cost. Across this group, the principal contrast is between solving a decision under stipulated demand and measuring how the policy behaves when demand, staff response, and delivery conditions shift.

3.4 Learning systems and organizational use

The nine studies in Table 4 span methods and adoption. Bekci et al. [14] address learning when lost sales censor demand; Kaynov et al. [30] examine deep reinforcement learning across one-warehouse multiretailer inventory variants; and Temizöz et al. [31] introduce deep controlled learning against established inventory heuristics. The latter comparison matters because a complex learning system needs an operationally credible baseline, not merely an untuned generic algorithm. Stratigakos et al. [32] combine probabilistic forecasts with a downstream decision objective, while Jia et al. [33] tie demand scenarios to online inventory routing. Both narrow the gap between a distributional score and a business choice, although their value remains conditional on the decision model and information process used in evaluation.

Other papers address the pieces a mathematical policy can leave out. Brau et al. [34] draw on interviews with 25 executives across the retail supply chain and show the continuing role of human judgment; Robertson et al. [35] use 23 interviews with retail executives and AI vendors to examine organizational change in the retail value chain. Brau et al. [36] discuss how judgment and predictive analytics are integrated in demand planning. The interviews and conceptual integration do not estimate an inventory policy's financial return; they explain why algorithm performance cannot be interpreted as adoption. Demand estimation for zero- and low-sales products [13] is a complementary information problem. A sensible program must test measurement, optimization, and organization together.

4. COMPARATIVE ANALYSIS

4.1 What is observed and what is optimized

The four streams differ along two axes: the quality of the recorded target and the proximity of the outcome to an implemented action. Table 5 compares their typical inferential limits. Forecast models often evaluate held-out demand or sales; policies usually optimize a modeled margin or cost; human-centered work observes adherence or describes organizational routines. None is intrinsically superior. Their

Table 1. Demand forecasting studies in the reviewed corpus ($n = 12$)

Study	Analytic approach	Data or information setting	Decision connection	Evidence boundary
Ye et al. (2024)	Fourier grey seasonal model	Seasonal retail series	Demand forecast	Forecast comparison; small-sample setting
Wellens et al. (2024)	Simplified tree learners	Belgian retailer; 4,523 products	Forecast to inventory	Retail data and inventory experiments
Sousa et al. (2025)	Two-stage censored-demand model	New fashion products	Demand forecast	Stockout censoring addressed
Fu and Fisher (2023)	Social-data prediction	Fashion styles and social signals	Demand forecast	Retail application; short life cycles
Wang et al. (2024)	Probabilistic forecast pooling	Intermittent demand series	Forecast to inventory	Accuracy and inventory utility compared
Lei et al. (2025)	Transfer learning across levels	Category and product sales	Demand forecast	Pooling of information across hierarchy
Liu et al. (2025)	Graph neural network	Product relations and sales	Demand forecast	Cross-category structure modeled
Ma (2024)	Graph recurrent network	Store-SKU attributes	Replenishment plan	Forecast and store-level ordering linked
Long et al. (2025)	Boosted-tree probabilistic model	Large e-commerce catalog	Demand forecast	Scalability and intermittency assessed
Chae et al. (2024)	ML and deep-learning comparison	Large restaurant chain	Demand forecast	External data and horizons evaluated
Fahimnia et al. (2025)	Behavioral forecasting study	Promotion and perishability scenarios	Human demand plan	Service-level anchoring examined
Zhu et al. (2025)	Spatial-attention ConvLSTM	Sales across neighboring retail regions	Aggregate demand forecast	Real-data benchmarks; model evaluation

The final column states a comparison boundary, not a common effect size. SKU = stock-keeping unit.

Table 2. Pricing and commercial-decision studies ($n = 7$)

Study	Analytic approach	Data or information setting	Decision connection	Evidence boundary
Liu et al. (2024)	Joint dynamic optimization	Omnichannel uncertain demand	Price and inventory	Model-based policy evaluation
Syed et al. (2024)	Multi-case study of dynamic pricing	Perishable-goods retail cases	Pricing process and inventory	Case evidence; not a policy trial
Caro and Sáez de Tejada Cuenca (2023)	DSS adherence analysis	Zara markdown decisions	Manager price override	Field behavior, not algorithm alone
Chehrizi et al. (2025)	Information-friction model	Perishable grocery; two pilots	Price adjustment	Inventory visibility and rigidity
Tang et al. (2025)	Offline causal pricing	Censored historical sales	Feature-based price	Logging and censoring constraints
Li and Zheng (2024)	Contextual online pricing	External covariates; finite stock	Dynamic price	Inventory-constrained learning model
Oh et al. (2024)	Behavioral multi-lever study	Retail return-policy decisions	Returns, price, stock	Joint behavioral decision setting

“Model-based” does not imply a live randomized deployment. DSS = decision support system.

Table 3. Inventory, assortment, and fulfillment studies ($n = 8$)

Study	Analytic approach	Data or information setting	Decision connection	Evidence boundary
Bansal et al. (2024)	Integrated stochastic model	Ship-from-store network	Replenish and allocate	Coupled channel decisions
Goedhart et al. (2023)	Joint network policy	Store inventory and online orders	Replenish and fulfill	Myopic allocation benchmark
Arslan et al. (2025)	Shipping-contract optimization	Omnichannel delivery promises	Fulfill and replenish	Contract and inventory coupled
Sun et al. (2025)	Data-driven robust policy	Omnichannel demand uncertainty	Inventory policy	Uncertainty set drives robustness
Kolyaei et al. (2025)	Reinforcement learning	Capacitated multicycle network	Replenish and fulfill	Policy evaluated in model setting
Baron (2024)	Decentralized optimization	Stores with private local information	Allocate online orders	Acceptance behavior modeled
Deng et al. (2024)	Policies and heuristics	Multi-warehouse online retailer	Stock and match demand	Capacity and zone demand coupled
Li et al. (2024)	Assortment optimization	E-commerce warehouse catalog	Stock assortment	Fulfillment cost vs. fill rate

The column labels indicate each paper’s specified decision, without asserting comparable implementation gains.

evidentiary claims address different links in Equations 1 and 2. A technically excellent forecast can still have little value when a stock constraint binds, and a simulated policy gain can disappear when a store rejects an order or a manager overrides a price.

A simple newsvendor illustration makes the metric problem explicit. With unit underage cost c_u and overage cost c_o , the cost-minimizing order is the τ quantile of demand, where

$$\tau = \frac{c_u}{c_u + c_o}, \quad q^* = F_D^{-1}(\tau). \quad (3)$$

A model can reduce mean absolute error while worsening calibration near τ ; it may then produce a worse order. Conversely, a model with modest point accuracy may provide a better upper-tail estimate. The issue becomes sharper under stockouts because the training variable S_t is censored, not D_t . The combination of probabilistic forecasts and inventory metrics in Wang et al. [16], censoring corrections in Sousa et al. [1], and downstream pooling in Stratigakos et al. [32] all address parts of this chain. They do not imply that one forecasting technique is uniformly best across margins, service targets, and products.

Table 4. Learning and decision-system studies ($n = 9$)

Study	Analytic approach	Data or information setting	Decision connection	Evidence boundary
Bekci et al. (2023)	Learning with censoring	One-warehouse multistore network	Allocate inventory	Unknown demand and lost sales
Kaynov et al. (2024)	Deep reinforcement learning	One-warehouse multiretailer system	Reorder and allocate	Policy compared across variants
Temizöz et al. (2025)	Deep controlled learning	Lost-sales inventory systems	Inventory control	Heuristic benchmark emphasized
Brau et al. (2024)	Qualitative grounded theory	25 retail-supply-chain executives	Managerial judgment	Interview evidence and organizational context
Robertson et al. (2025)	Expert-interview framework	23 retail executives and AI vendors	AI implementation	Organizational change, not cost trial
Adam et al. (2024)	Two-stage demand estimator	Long-tail sales with zero products	Demand estimation	Selection bias in sparse catalogs
Stratigakos et al. (2025)	Decision-focused pooling	Multiple probabilistic forecasts	Downstream decision	Forecast score vs. realized objective
Jia et al. (2025)	Scenario predict-optimize	One-warehouse multiretailer routing	Inventory and route	Correlated/lumpy demand scenarios
Brau et al. (2023)	Human-analytics integration	Digital supply-chain planning	Demand planning	Judgment and prediction interface

The interview counts come from the respective published abstracts.

Table 5. Cross-stream comparison of evidence and managerial interpretation

Stream	Typical observed signal	Principal decision lever	Informative test	Claim that requires further evidence
Demand forecasting	Sales, exogenous covariates, product relationships	Order quantity or planning input	Later-period forecast and inventory sensitivity	Improvement in realized profit
Pricing	Historical prices, inventory, sales, promotions	Markdown or dynamic price	Policy comparison under feasible price set	Causal revenue lift from adoption
Fulfillment	Store and warehouse stock, online orders	Source, ship, replenish, assortment	Network cost and service under capacity	Transfer to different contracts and store behavior
Decision systems	Censored histories, expert judgment, overrides	Algorithmic recommendation and use	Strong heuristic baseline plus adherence	Sustained economic value after rollout

4.2 Information timing and counterfactual validity

A forecast must use only information known when an order, price, or allocation was set. A later stockout label, completed promotion, revised inventory count, or observed competitor response can leak future information into a retrospective test. The need is clearest for online price learning [7, 22] and dynamic fulfillment [9, 26]. Evaluation should reconstruct the historical information set $\mathcal{S}_t$, then compare policies on a common decision calendar. A time split alone is insufficient if same-period features were not available at the intended decision time.

Counterfactual evaluation is harder than predictive validation. Historical actions were chosen by a prior policy, so records are selected; price and promotional exposure are particularly nonrandom. Estimating Equation 2 from logs requires overlap between candidate actions and logged actions, a credible demand-response model, or a field experiment. Stockouts induce an additional missing-outcome problem. Offline pricing under censoring [7] and the estimate of long-tail demand [13] make these limitations visible. A paper that reports a scenario-model cost difference should identify the assumed demand distribution and cost parameters; a paper that claims actual commercial gain should show the comparison design that supports it.

4.3 Uncertainty, constraint, and organization

Uncertainty is often passed through a pipeline as if every stage were separable. But inventory changes the observed training target; prices change demand; and fulfillment changes

both availability and customer lead time. Omnichannel policy models [8, 10, 25] bring constraints into the optimization, while collaborative demand forecasts and store-level replenishment [3, 17] alter the information entering it. The resulting systems are coupled. Comparing model classes on one prediction score while holding the feasible set implicit can misstate the manager's actual choice.

Implementation adds a second coupling. A store's private information [27] and a markdown manager's willingness to follow a DSS [6] affect the feasible and chosen actions. Interview research [34, 35] gives context for these frictions; it is an input to system design rather than a direct estimate of algorithmic value. The practical unit of analysis is therefore an *implemented policy with a defined override rule*, not a model saved to a dashboard. This interpretation helps reconcile why advanced methods, simpler tree models [12], and human forecasts may coexist in one firm.

5. METHODOLOGICAL ASSESSMENT

5.1 Four kinds of evidence

The word “performance” can denote four different findings in business intelligence. *Predictive validity* means that a model estimates a target more usefully on later observations than a stated comparator. *Decision validity* means that its predictions induce a better feasible action under an explicit operating objective. *Implementation validity* means that people and systems actually execute the action when needed. *Causal business impact* means that the executed policy changes an outcome relative to a credible counterfactual. Each is valu-

able, but none follows automatically from the previous one. A seasonal forecasting comparison [11] primarily addresses the first claim. Inventory experiments with the tree-based retail forecasts of Wellens et al. [12] connect the first two under assumed costs. The markdown adherence study of Caro and Sáez de Tejada Cuenca [6] observes a piece of implementation that a model-only paper cannot recover. Shipping-contract optimization [10] clarifies the feasible decision but cannot, without deployment evidence, establish actual rollout profit. Executive interviews [34, 35] explain organizational mechanisms, not a policy's causal return. Table 6 supplies appraisal questions that respect those differences. The risk column identifies an inference hazard; it is not a score assigned to a particular paper.

5.2 Forecast error and decision regret

For an action a and demand d , let $L(a, d)$ be the operating loss. The action a_F minimizes its expectation under the true demand distribution, whereas $a_{\hat{F}}$ uses a forecast. Decision regret is

$$R(\hat{F}; F) = \mathbb{E}_F[L(a_{\hat{F}}, D)] - \mathbb{E}_F[L(a_F, D)] \geq 0. \quad (4)$$

The true distribution is unknown in ordinary practice. A study approximates regret through held-out outcomes, a specified simulator, or experimental assignment. Under the newsvendor loss, the economically relevant feature of a forecast is its behavior near the critical quantile in Equation 3. Under a shipping promise, route capacity and the distribution of delivery times also matter. A useful evaluation therefore reports both a proper predictive score and the resulting decision cost, while labeling the latter as simulated or observed.

Product heterogeneity can reverse model rankings. A method may reduce volume-weighted error on stable staples while degrading the upper tail for items whose shortage is expensive. Its average forecast score improves, yet shortage cost rises. The category pooling, sparse e-commerce, and intermittent-demand papers [4, 15, 16] motivate this design concern; it is not a claim that any named method failed. Researchers should show uncertainty and cost differences by intermittency, lead time, promotion state, and item value. How those strata are weighted is a managerial judgment, not a neutral statistical choice.

When price or assortment changes, historical sales no longer identify demand under the proposed action without additional assumptions. Price affects demand; an assortment changes what can be chosen; both alter the future training data. Offline learning from censored prices [7] and dynamic pricing under external information [22] face related but different restrictions. A comparison must state its support for alternative actions and whether demand is assumed invariant to the policy. If that assumption is only plausible for small changes, the tested action set should reflect that limit.

5.3 Validation at the actual decision time

Every input has a creation time, arrival time, revision policy, and action deadline. A forecast intended for an order placed on Monday may use a promotion planned on Friday but cannot use that promotion's realized sales. An inventory position can be revised for returns after a routing decision.

A delivery promise can change within minutes. For each variable X_{jt} , researchers should document whether $X_{jt} \in \mathcal{I}_t$, the information available to the decision maker when action A_t was chosen. A generic random train-test split does not establish this condition.

Holdout construction should follow decision cadence. A later-quarter replenishment evaluation needs intact ordering cycles; random row splitting may put closely related store-SKU histories into both sets. Cross-store tests address transfer to new locations, and cross-product tests address cold starts; neither substitutes for a later-period test. Seasonal series, fashion launches, and sparse catalogs have different failure points [1, 11, 13]. Their item counts, histories, and metric aggregation should be disclosed separately.

A shared simulator can compare algorithms under controlled demand and cost assumptions. Its substitution rules, capacities, lead times, perishability, and policy baselines should be explicit, with common demand scenarios for competitors. A shadow deployment can then reveal whether required state variables arrive on time and whether actions fit the live workflow. A controlled rollout is the appropriate next step for a realized impact claim. Each stage supports a useful but bounded inference. Treating the simulator as a field experiment obscures where the evidence ends.

5.4 Counterfactuals and operating cost

Historical actions were selected by a prior pricing, ordering, or fulfillment policy. Logged observations are consequently uneven across possible actions. A retrospective comparison that imputes profit for an unobserved price needs credible overlap or a model of demand response; stockouts add missing demand. The causal pricing approach of Tang et al. [7] makes the challenge explicit. For policies affecting multiple stores, interference is possible: sending an order from one store changes the stock available to local and neighboring customers. A store-level randomized trial must account for that network relationship rather than assuming independent units.

Aggregate value can also conceal where costs move. An online order allocated to a store may shorten delivery while consuming stock reserved for walk-in demand [8, 9]. A warehouse assortment can lower shipping cost but change fill rate and choice [29]. The outcome vector should include net contribution after fulfillment and spoilage, service, stockouts, customer promises, and material workload. A study may reasonably examine only a subset; it should name omitted channels before recommending a policy.

Operational compliance is another part of the counterfactual. In a centralized optimization, every store executes a chosen action; in Baron et al. [27], stores possess local information and can reject requests. A markdown manager may depart from the system for a reason absent from its data [6]. Reporting recommendation quality, adoption, and outcomes conditional on adoption separately is informative, but conditioning on adopters can select favorable cases. Where possible, assignment to recommendation access gives an intention-to-treat estimate, followed by analysis of overrides and their recorded reasons. This helps distinguish useful local correction from avoidable reluctance to use the system.

Net value should include the cost of acquiring external signals,

Table 6. Appraisal questions for retail decision-intelligence claims

Claim examined	Minimum question	Principal failure mode	Illustrative studies
Target validity	Does the label measure demand or censored sales?	Availability and returns distort the target	[1, 13, 14]
Temporal validity	Was each covariate known before the action?	Later information leaks backward	[12, 19]
Predictive validity	Are strong baselines and uncertainty tested later?	Mean error hides sparse and tail failures	[4, 16]
Decision validity	Are actions feasible under common costs?	Service targets reverse rankings	[8, 25, 32]
Policy evaluation	Are logs informative about alternative actions?	Selection and poor overlap bias value	[7, 22]
Implementation	Are recommendations and overrides observed?	Suggested actions never become actual actions	[6, 27, 34]
Transfer	Does value survive a new regime or store?	One catalog or network defines the result	[4, 26, 30]
Business impact	Is the outcome comparison counterfactual?	Simulated value becomes “realized” gain	[10, 35]

integrating data, maintaining the model, training staff, and resolving exceptions. A signal that arrives after an ordering deadline has zero value for that order regardless of its predictive content. In Equation 5, C_z makes the information cost explicit. The same principle applies to a policy: a percentage gain in forecast accuracy does not establish a viable service without the baseline, decision horizon, implementation cost, and uncertainty around net benefits.

6. DISCUSSION

6.1 Comparisons that are legitimate across streams

The 36 studies invite comparison, but the choice of denominator matters. Predictive studies can be compared by whether they evaluate later outcomes, address censoring, quantify uncertainty, and relate their target to a decision. Optimization papers can be compared by the decisions and constraints they encode, the strength of their baseline, and the sensitivity of their results to demand and cost assumptions. Behavioral and organizational studies can be compared by what they actually observe about recommendation use. Comparing a forecast error from one paper with a modeled profit increase from another is not meaningful; the target, policy, and evidence differ.

A narrower comparison can still yield insight. Ye et al. [11] consider limited seasonal data, Lei et al. [15] share information across a product hierarchy, and Long et al. [4] handle a very large sparse catalog. All three address how much history a business has for an item, but their solutions occupy different corners of a data-regime map. The researcher deciding among them should first diagnose series length, cross-item similarity, intermittency, and computation budget. A comprehensive benchmark would cross those four dimensions while holding the horizon, product mix, and business loss fixed. It should permit simpler methods to win where data are thin, rather than assuming a complex architecture has universal value.

A similar comparison links Sousa et al. [1], Bekci et al. [14], and Adam et al. [13]: they each resist a naive reading of zero or low sales. Yet new-product censoring, multistore lost sales, and long-tail choice selection require different missing-data

assumptions. A common label called “demand correction” would hide those distinctions. A future study could compare partial-identification bounds, imputation, and structural models on the same availability logs, with a sensitivity range for consumers who switch to substitutes. The main reporting unit would be the resulting order decision, not a reconstructed demand series presumed to be ground truth.

The pricing and fulfillment literature displays a second division. Liu et al. [21] jointly determine price and inventory under omnichannel uncertainty; Bansal et al. [8] join replenishment with online allocation; Arslan et al. [10] account for delivery commitments; and Li et al. [29] choose a warehouse assortment under fulfillment costs. Each paper integrates a different pair of levers. A comprehensive retail policy would need several at once, but adding every lever to one model can make identification and implementation opaque. A disciplined sequence is to add one decision interaction at a time and report the marginal value of that coupling. For example, fix the replenishment policy when testing a price-allocation interaction; then allow replenishment to adapt and quantify the additional gain. This design identifies which coupling matters, rather than crediting an entire system for an unexplained difference.

Table 7 organizes twelve study pairs or small groups by what can be learned from their difference. The final column is a prospective comparison, not a reported head-to-head result. It specifies the common information and decision conditions needed before algorithms or policies can be ranked. Such contrasts are useful to research teams designing a study with a retail partner: they show which business assumption should be held constant and which should be varied.

6.2 An empirical sequence for researchers

A research partnership can implement the agenda as a sequence of increasingly strong tests. First, establish the decision ledger. Each row should represent a choice available at a known time: a store-SKU order, a category markdown, a fulfillment request, or a delivery promise. Record the feasible actions and constraints as they were then, not as revised later. Join data by event time, retain stock and offer exposure,

Table 7. Twelve targeted contrasts and the experiments they motivate

Studies	Substantive contrast	Common condition for a fair comparison	Outcome to report
[11, 12]	Seasonal small samples vs. feature-rich trees	Same retail series, forecast origin, and available covariates	Horizon error and order cost
[4, 12]	Practical simplicity vs. catalog scale	Same product strata and compute budget	Accuracy, latency, and maintenance
[1, 13]	Stockout censoring vs. long-tail selection	Availability and offer-exposure logs	Demand bounds and chosen quantity
[2, 3]	External social signals vs. product relationships	Same history, action deadline, and acquisition costs	Incremental net information value
[16, 32]	Probabilistic pooling vs. decision-focused pooling	Common distributions, loss, and inventory service target	Proper score and policy regret
[8, 17]	Store-level replenishment vs. online allocation	Same store stock and demand calendar	Shortage and fulfillment cost
[20, 36]	Service anchoring vs. planner integration	Separated demand belief and required service	Forecast revision and action quality
[7, 21]	Joint price-stock choice vs. logged-price censoring	Common price support and demand response	Policy value and identification range
[6, 34]	Observed markdown adherence vs. executive judgment	Recorded exposure, discretion, and override reason	Compliance and override benefit
[5, 23]	Inventory visibility vs. perishable pricing	Same shelf-life and inventory information	Markdown timing and waste
[9, 27]	Central allocation vs. private store information	Same network and store acceptance behavior	Feasible service and coordination cost
[30, 31]	Generic deep RL vs. controlled learning	Same simulator, heuristic baseline, and tuning budget	Cost, violation, and robustness

and mark any variable whose publication or system-arrival time follows the choice. This step is less glamorous than model development, but errors here invalidate an otherwise sophisticated comparison.

Second, choose the business comparator. A retailer may already use base-stock rules, a category planner's forecast, a markdown schedule, or a routing heuristic. The value of an advanced method is relative to that practice. The comparator must be tuned using the same historical window and constraints as the proposed policy. Compare simple statistical models, tree ensembles, and structured methods where appropriate; the favorable inventory baseline in Temizöz et al. [31] illustrates the importance of a serious heuristic. If a model requires more compute or delayed external data, those resources belong in the comparison.

Third, diagnose observability before estimation. A stockout means that actual demand is partly hidden. If shelf availability is recorded with error, evaluate a range of plausible corrections. For pricing, tabulate the observed action support by product, store, and time; a method cannot credibly estimate what would happen at a price never offered to comparable customers without additional structure. For fulfillment, distinguish a request sent to a store from its acceptance and final delivery. These checks define which policy changes are evaluable with existing logs and which require prospective experimentation.

Fourth, pre-specify business loss and decision units. A shortage penalty should reflect the intended service consequence, not be tuned afterward to favor a method. The same holds for spoilage, shipping, return handling, and labor. Report a distribution of results across plausible cost ratios because managers can legitimately disagree about trade-offs. The objective should be calculated per order, item, or customer exposure and then aggregated with transparent weights. Otherwise a gain on high-volume items can conceal concentrated losses on a minority of high-value or vulnerable customer segments.

Fifth, perform staged validation. Out-of-time testing checks drift in information and demand. Cross-store or cross-region testing checks transfer. Scenario analysis checks policy behavior under disrupted supply and exceptional promotions. Shadow evaluation measures whether proposed ac-

tions are available, timely, and intelligible. A randomized rollout, if feasible, estimates realized value under adoption and spillovers. The steps need not all appear in one article; explicit reporting of the completed stage lets later teams build on it without exaggerating certainty.

Sixth, report the failure cases. An average profit gain can co-exist with occasional severe service failures, infeasible stock movements, or highly unstable prices. Present distributions, worst-case groups, and constraint violations. An automated recommender should expose whether a recommendation is sensitive to one uncertain cost or demand parameter. Managers can then set a review threshold around high-stakes exceptions instead of applying a fixed confidence score to every case. The goal is a decision process that remains defensible when forecasts are wrong.

6.3 Research infrastructure and reproducibility

Public retail datasets make comparison possible, but transaction logs rarely include the whole decision context. They may omit shelf availability, private promotions, labor limits, returns, contracted delivery costs, and the recommendations staff saw. A benchmark built only on sales can support forecasting research while remaining inadequate for a causal inventory policy test. A useful shared dataset should publish a data dictionary with event timestamps, action alternatives, exposure, observed outcomes, and missingness rules. If sensitive fields cannot be released, a secure evaluation service can run submitted policies on held-out events and publish a predefined metric vector without revealing customer records.

Reproducibility also requires fixing the sequence in which an analyst sees data. For each model update, record the training cutoff, feature generation time, hyperparameter selection window, and deployment version. An online policy changes the data it will later learn from; thus its logs should include action probabilities or the rule by which actions were selected. Without these, offline evaluation of a new policy can be unidentified even when the sales table is large. For price experiments, randomized variation should be designed within legal and commercial constraints; for store allocation, cluster assignment may reduce spillovers. These are study design issues, not merely software documentation.

The field should also be cautious about benchmark saturation.

A public leaderboard can encourage repeated model fitting to the same evaluation periods and optimizing a score unrelated to operating cost. A rolling-origin test with locked later periods and multiple cost structures would be more informative. Alongside accuracy, report calibration, computational expense, data latency, decision regret under documented assumptions, and a policy's feasibility rate. The evidence from Wellens et al. [12] on practical models and Long et al. [4] on large-catalog scale suggests that a reproducible result should include both statistical and operational resources.

Finally, interdisciplinary replication would strengthen the link between operations and organizational research. A field system can retain recommendation logs and override reasons while forecasting researchers test calibration and operations researchers test policy robustness. Human factors researchers can examine whether managers understand uncertainty displays and whether exceptions convey useful local knowledge. A shared protocol avoids the false choice between trusting an algorithm and dismissing it; it records when a human correction helps, when it harms, and what information caused the difference. That is a measurable business intelligence question.

7. RESEARCH CHALLENGES

7.1 Recovering demand from transactions

Sales reflect availability, display, assortment, price, and channel. Latent demand is obscured by missed purchases during stockouts [1, 14] and unobserved low-probability alternatives in a long-tail catalog [13]. Promotion exposure and perishability also shape demand plans [20], while regional sales correlations may improve aggregate forecasts [19]. A common evaluation dataset should preserve availability histories, lost-sale proxies, returns, promotion exposure, and regional identifiers, with timestamps at the decision level. These fields are commercially sensitive and often missing. When they are absent, researchers should report a bound or sensitivity analysis rather than relabel sales as unconstrained demand.

7.2 Aligning validation with business loss

Accuracy metrics aggregate across items and periods even though stockout, markdown, and fulfillment costs are asymmetric. Equation 3 shows that the relevant tail depends on the cost ratio. Comparative work should report a proper predic-

tive score, calibration at the decision quantile, and the same policy's cost or service on a later horizon. A policy objective can itself be misspecified: a carrier contract, spoilage rate, or store labor constraint may change. Sensitivity to those parameters is as important as sensitivity to a learning algorithm's hyperparameters.

7.3 Transporting evidence across operating regimes

The studies cover seasonal series, fashion life cycles, restaurant horizons, sparse e-commerce catalogs, perishable groceries, and multistore distribution [2, 4, 11, 18, 23, 30]. Their time scales, assortment breadth, and service promises differ. A comparison that averages scores over a single dataset cannot establish external validity. Researchers should disclose the regime in which a policy is identified, test on later periods and distinct store or product groups, and specify what must be re-estimated after a promotion, supply disruption, or channel shift. Repeated tuning on a public benchmark can also overstate transfer to a private retail network.

7.4 Evaluating adoption and accountability

Model recommendations compete with local knowledge and incentives. In markdowns [6], an override can be an informed correction or an avoidable bias; without recording reasons, the two cannot be separated. Interviews document organizational challenges [34, 35], but future field designs should record recommendation exposure, acceptance, override reason, stock and price at the time of action, and downstream outcome. A business intelligence system should show the expected consequence and uncertainty of each recommendation and offer a route for legitimate human correction. Accountability requires deciding who can override a policy and how the override is audited.

8. FUTURE RESEARCH DIRECTIONS

Table 8 states seven research directions as designs rather than slogans. They target concrete uncertainties in the reviewed evidence and indicate what would count as progress. Several can be pursued with privacy-preserving collaborations between firms and researchers; others require controlled simulation before any live test. Neither a larger model nor a larger data table alone resolves the identification problem.

Table 8. Decision-centered research agenda derived from the 36-study review

Research question	Comparison design	Minimum reporting set	Studies
What changes when stockout-censored sales are corrected?	Same store-SKU history, fixed policy and horizon; vary demand reconstruction	Availability, censoring rule, calibration, lost-sales and cost sensitivity	[1, 13, 14]
When does better probability calibration improve orders?	Compare models at common cost ratios and lead times	Tail calibration, proper score, realized/simulated service and cost	[4, 16, 32]
Do social and relational signals pay for acquisition?	Feature ablation at predecision timestamp across cold- and warm-start groups	Incremental value, feature latency, missingness and data cost	[2, 3, 15]
Can price and fulfillment be jointly evaluated?	Common demand-response environment with capacity and delivery promises	Price support, spillovers, fulfillment costs, profit uncertainty	[8, 10, 21]
How robust are policies under regime change?	Train on earlier regime, evaluate after shift and across stores	Re-estimation rule, drift, constraint violation, net cost	[19, 25, 26]
When should a manager override the system?	Record recommendations and reasons; randomize explanation or review protocol where feasible	Adoption, override benefit, heterogeneity and accountability	[6, 34, 35]
Which gains persist after operational rollout?	Stage simulation, shadow deployment and controlled field comparison	Baseline policy, compliance, spillover, service, net margin	[9, 12, 33]

A useful first design compares two models in a fixed decision setting. Researchers can pre-register a loss function and estimate the uncertainty of a value difference on sequential holdouts. The baseline should be a competent operational heuristic, as the inventory learning comparison of Temizöz et al. [31] emphasizes, and its tuning budget should be documented. Next, evaluate the same models under several under-age/overage ratios and service constraints. A rank reversal is a substantive result because it tells a manager when predictive superiority ceases to matter.

A second design measures the incremental business value of information. For a candidate signal Z_t observed before action, define

$$\Delta_Z = V(\pi_{X,Z}, \pi_0) - V(\pi_X, \pi_0) - C_Z, \quad (5)$$

where C_Z is its acquisition, integration, and governance cost. Social signals, product graphs, and category pooling [2, 3, 15] motivate the test. Comparing the same policy class with and without Z_t isolates the signal more cleanly than comparing two unrelated algorithms. The design should also report missing-data rates and latency; a signal that arrives after the order deadline has no operational value for that order.

A third design brings decision rights into evaluation. Shadow recommendations can be displayed without changing actions, to record discrepancies and reasons. A staged trial can then randomize recommendation access, explanations, or review capacity across comparable units while accounting for spillovers between stores. This is more demanding than a retrospective dashboard analysis, but it connects the behavioral findings of Caro and Sáez de Tejada Cuenca [6] with the organizational mechanisms in Brau et al. [34] and the operational policies in Bansal et al. [8]. Where randomization is impossible, researchers should be explicit about policy logging, overlap, and sensitivity to confounding.

Finally, research should examine joint rather than sequential choices. For example, a price affects demand, demand affects stock and delivery times, and the fulfillment source affects future store availability. A tractable experimental environment could couple one promotional category, a small store network, and an explicit shipping contract, then publish its logs and constraints. The aim is not to replace detailed field studies with a universal benchmark. It is to make failure modes reproducible: leakage, censoring, tail miscalibration, capacity violation, and staff override can each be switched on and measured.

9. LIMITATIONS

This is a bounded review of 36 selected journal studies, not an exhaustive census or a formal meta-analysis. The four clusters simplify work that crosses categories; for example, a store-level forecasting paper also informs replenishment, and a pricing paper also estimates demand. Coding relies on publication descriptions and is interpretive. The evidence-boundary labels describe the kind of support visible in the source, not a formal risk-of-bias score. Exact bibliographic records and the study-by-study coding accompany the source package, allowing readers to revise classifications. The cut-off at January 2026 excludes subsequent research, and the agenda should be updated as evaluations with live operational outcomes appear.

10. CONCLUSION

Retail business intelligence is strongest when it treats a forecast as an input to a constrained, adopted, and evaluated decision. Across 36 recent studies, the field has produced credible advances in recovering demand signals, representing uncertainty, learning prices, allocating inventory, and explaining organizational use. The studies also reveal a persistent gap between a successful model comparison and demonstrated business value. The most useful next papers will specify the historical information set, correct or bound censored demand, compare a policy with a realistic alternative, report decision-level costs and uncertainty, and observe whether managers actually implement the recommendation. Those steps make business data analytics accountable to the commercial decisions it was built to improve.

REFERENCES

- [1] Sousa, M. S., Loureiro, A. L. D., & Miguéis, V. L. Predicting demand for new products in fashion retailing using censored data. *Expert Systems with Applications*, 259, Article 125313, 2025.
- [2] Fu, Y., & Fisher, M. The value of social media data in fashion forecasting. *Manufacturing & Service Operations Management*, 25(3), 1136–1154, 2023.
- [3] Liu, J., Wang, G., Zhao, H., Lu, M., Huang, L., & Chen, G. Beyond complements and substitutes: A graph neural network approach for collaborative retail sales forecasting. *Information Systems Research*, 36(4), 1993–2016, 2025.
- [4] Long, X., Bui, Q., Oktavian, G., Schmidt, D. F., Bergmeir, C., Godahewa, R., Lee, S. P., Zhao, K., & Condylis, P. Scalable probabilistic forecasting in retail with gradient boosted trees: A practitioner's approach. *International Journal of Production Economics*, 279, Article 109449, 2025.
- [5] Syed, T. A., Aslam, H., Bhatti, Z. A., Mehmood, F., & Pahuja, A. Dynamic pricing for perishable goods: A data-driven digital transformation approach. *International Journal of Production Economics*, 277, Article 109405, 2024.
- [6] Caro, F., & Sáez de Tejada Cuenca, A. Believing in analytics: Managers' adherence to price recommendations from a DSS. *Manufacturing & Service Operations Management*, 25(2), 524–542, 2023.
- [7] Tang, J., Qi, Z., Fang, E., & Shi, C. Offline feature-based pricing under censored demand: A causal inference approach. *Manufacturing & Service Operations Management*, 27(2), 535–553, 2025.
- [8] Bansal, V., Bisi, A., Roy, D., & Venkateshan, P. Integrated inventory replenishment and online demand allocation decisions for an omnichannel retailer with ship-from-store strategy. *European Journal of Operational Research*, 316(3), 1085–1100, 2024.
- [9] Goedhart, J., Haijema, R., Akkerman, R., & de Leeuw, S. Replenishment and fulfillment decisions for stores in an omnichannel retail network. *European Journal of Operational Research*, 311(3), 1009–1022, 2023.
- [10] Arslan, B., Schrottenboer, A. H., & Atan, Z. Optimal fulfillment and replenishment for omnichannel retailers with standard shipping contracts. *European Journal of Operational Research*, 323(2), 642–656, 2025.

- [11] Ye, L., Xie, N., Boylan, J. E., & Shang, Z. Forecasting seasonal demand for retail: A Fourier time-varying grey model. *International Journal of Forecasting*, 40(4), 1467–1485, 2024.
- [12] Wellens, A. P., Boute, R. N., & Udenio, M. Simplifying tree-based methods for retail sales forecasting with explanatory variables. *European Journal of Operational Research*, 314(2), 523–539, 2024.
- [13] Adam, H., He, P., & Zheng, F. Machine learning for demand estimation in long tail markets. *Management Science*, 70(8), 5040–5065, 2024.
- [14] Bekci, R. Y., Gümüő, M., & Miao, S. Inventory control and learning for one-warehouse multistore system with censored demand. *Operations Research*, 71(6), 2092–2110, 2023.
- [15] Lei, D., Qi, Y., Liu, S., Geng, D., Zhang, J., Hu, H., & Shen, Z. J. M. Pooling and boosting for demand prediction in retail: A transfer learning approach. *Manufacturing & Service Operations Management*, 27(6), 1779–1794, 2025.
- [16] Wang, S., Kang, Y., & Petropoulos, F. Combining probabilistic forecasts of intermittent demand. *European Journal of Operational Research*, 315(3), 1038–1048, 2024.
- [17] Ma, S. Retail store-SKU level replenishment planning with attribute-space graph recurrent neural networks. *Expert Systems with Applications*, 249, Article 123727, 2024.
- [18] Chae, B. K., Sheu, C., & Park, E. O. The value of data, machine learning, and deep learning in restaurant demand forecasting: Insights and lessons learned from a large restaurant chain. *Decision Support Systems*, 184, Article 114291, 2024.
- [19] Zhu, B., Sun, Z., vanden Broucke, S., Lan, K., & Xiao, D. Aggregate sales forecasting based on spatial correlation in retail. *Journal of Theoretical and Applied Electronic Commerce Research*, 20(4), Article 334, 2025.
- [20] Fahimnia, B., Tan, T., & Tahirov, N. Service-level anchoring in demand forecasting: The moderating impact of retail promotions and product perishability. *International Journal of Forecasting*, 41(2), 554–570, 2025.
- [21] Liu, S., Wang, J., Wang, R., Zhang, Y., Song, Y., & Xing, L. Data-driven dynamic pricing and inventory management of an omni-channel retailer in an uncertain demand environment. *Expert Systems with Applications*, 244, Article 122948, 2024.
- [22] Li, X., & Zheng, Z. Dynamic pricing with external information and inventory constraint. *Management Science*, 70(9), 5985–6001, 2024.
- [23] Chehrizi, N., Sanders, R. E., & Stamatopoulos, I. Inventory information frictions explain price rigidity in perishable groceries. *Marketing Science*, 44(2), 411–436, 2025.
- [24] Oh, H. K., Abdulla, H., & Oliva, R. Behavioral multi-lever decision-making: A study of consumer return policy, price, and inventory decisions. *Journal of Operations Management*, 70(1), 137–156, 2024.
- [25] Sun, Y., Qiu, R., Shao, S., Sun, M., & Fan, Z. P. Omnichannel strategies and data-driven robust inventory policies with demand uncertainties. *Computers & Operations Research*, 173, Article 106830, 2025.
- [26] Kolyaei, M., Zhang, L., & Blom, M. Inventory replenishment and fulfilment decisions for an omnichannel retailer: A reinforcement learning-based method. *International Journal of Production Research*, 63(24), 9571–9592, 2025.
- [27] Baron, O., Cire, A. A., & Savaser, S. K. Decentralized online order fulfillment in omni-channel retailers. *Production and Operations Management*, 33(8), 1719–1738, 2024.
- [28] Deng, Q., Li, X., Lim, Y. F., & Liu, F. Optimal policies and heuristics to match supply with demand for online retailing. *Manufacturing & Service Operations Management*, 26(5), 1925–1944, 2024.
- [29] Li, X., Lin, H., & Liu, F. Should only popular products be stocked? Warehouse assortment selection for e-commerce companies. *Manufacturing & Service Operations Management*, 26(4), 1372–1386, 2024.
- [30] Kaynov, I., van Knippenberg, M., Menkovski, V., van Breemen, A., & van Jaarsveld, W. Deep reinforcement learning for one-warehouse multi-retailer inventory management. *International Journal of Production Economics*, 267, Article 109088, 2024.
- [31] Temizöz, T., Imdahl, C., Dijkman, R., Lamghari-Idrissi, D., & van Jaarsveld, W. Deep controlled learning for inventory control. *European Journal of Operational Research*, 324(1), 104–117, 2025.
- [32] Stratigakos, A., Pineda, S., & Morales, J. M. Decision-focused linear pooling for probabilistic forecast combination. *International Journal of Forecasting*, 41(3), 1112–1125, 2025.
- [33] Jia, M., Schrottenboer, A. H., & Chen, F. Scenario predict-then-optimize for data-driven online inventory routing. *Transportation Science*, 59(5), 1032–1056, 2025.
- [34] Brau, R. I., Sanders, N. R., Aloysius, J., & Williams, D. Utilizing people, analytics, and AI for decision making in the digitalized retail supply chain. *Journal of Business Logistics*, 45(1), Article e12355, 2024.
- [35] Robertson, J., Botha, E., Oosthuizen, K., & Montecchi, M. Managing change when integrating artificial intelligence (AI) into the retail value chain: The AI implementation compass. *Journal of Business Research*, 189, Article 115198, 2025.
- [36] Brau, R., Aloysius, J., & Siemsen, E. Demand planning for the digital supply chain: How to integrate human judgment and predictive analytics. *Journal of Operations Management*, 69(6), 965–982, 2023.