

Coordinating Replenishment and Dynamic Pricing for Perishable Omnichannel Retail: A Fixed-Point Dynamic-Programming–Reinforcement-Learning Framework

Ayad Hendalianpour

sanwood84@gmail.com

University of Tehran

Mehdi Abbasi

Islamic Azad University of Shiraz

Research Article

Keywords: Perishable products, Omnichannel, Dynamic pricing, Inventory management, Dynamic Programming-Reinforcement Learning

Posted Date: October 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11265948/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Coordinating Replenishment and Dynamic Pricing for Perishable Omnichannel Retail: A Fixed-Point Dynamic-Programming–Reinforcement-Learning Framework

Abstract

For omnichannel retailers, managing perishable products with limited shelf lives requires tightly coordinated decisions that balance profitability, waste and service quality under uncertain, time-sensitive demand. This study presents an integrated optimisation framework that unifies long-term replenishment planning with short-term adaptive pricing for fresh products, decoupling the two decisions by timescale: dynamic programming devises lot-sizing and replenishment policies in advance, modelling freshness decay, delivery constraints and waste costs, while reinforcement learning updates channel-specific prices in real time in response to stochastic demand and inventory ageing. We further treat the freshness guarantee offered to online customers as a decision variable rather than a fixed parameter. Using a transparently calibrated simulation of a fresh-produce retailer with a four-day shelf life and a three-day replenishment cycle over a one-year horizon, we establish two results. First, whether perishability binds at all is governed by the ratio of the replenishment cycle to the shelf life: under daily delivery waste is 0.04% and no policy has anything to optimise, whereas a three-day cycle raises it to 6.6%. Second, a stronger freshness guarantee raises online demand but strands ageing stock, lowering profit while raising waste, so the optimal promise is the operational minimum. Benchmarking the coordinated framework in the binding regime against a one-pass sequential variant, a standalone dynamic program, a standalone reinforcement-learning agent, and a two-stage stochastic program shows that a purely learning-based agent is decisively the weakest, and that closing the replenishment-pricing feedback loop delivers its gain in waste control rather than profit.

Keywords: Perishable products; Omnichannel; Dynamic pricing; Inventory management; Dynamic Programming-Reinforcement Learning.

1. Introduction

Retail has moved toward omnichannel selling, in which consumers buy through several channels at once, including Buy-Online-Pick-up-from-Store (BOPS), Buy-Online-Home-Delivery (BOHD), and Buy-in-Store-Pick-from-Store (BSPS) channels, as identified in studies by Hendalianpour et al., (2020); M. Liu et al., (2024); Zhao et al., (2023). Although this has improved the overall experience for consumers, it has also posed a significant degree of complexity, especially for food items with limited shelf lives. For managing such products, decisions regarding replenishment and pricing must be well coordinated to maximize profitability, minimize wastage, and address diverse consumer demands for food freshness, as identified in studies by Amorim et al., (2019); Pilati et al., (2024). In this regard, this study aims to address this gap by developing a hybrid approach that incorporates Dynamic

Programming (DP) for strategic decisions regarding lot sizing and Reinforcement Learning (RL) for pricing decisions.

The rationale behind the hybridization of the approaches stems from the inherent limitations of the individual methodologies. On the one hand, exact DP faces the so-called “curse of dimensionality” in the context of the omnichannel environment, as the state space should capture not only the total inventory but also its age distribution across different channels, making the computation intractable (Taudes and Rudloff 2012; Zhao, Chen, and Tang 2025). On the other hand, pure RL, despite its ability to cope with large-dimensional state spaces, may be myopic and focus only on the short-term reward, ignoring the long-term inventory-related consequences, including inventory aging and waste (Ganesan, Sundararaj, and Srinivas 2021; Wang et al. 2025). Thus, the proposed framework combines the forward-looking and structured order replenishment policy of DP with the ability of RL to learn and react to real-time changes in the environment, as the former optimizes the order quantities with consideration of the aforementioned inventory-related factors, while the latter adjusts the prices based on the real-time inventory levels and consumer behavior (Keskin, Li, and Song 2022; Kopalle et al. 2023; Syed et al. 2024).

This study considers the joint optimization of lot-sizing and dynamic pricing for perishable products in an omnichannel retailing context and highlights the heterogeneity of consumers in terms of product freshness. It highlights the gap in literature on inventory management and dynamic pricing and the need for strategies to improve efficiency, reduce waste and enhance profitability. A major challenge is the diversity of consumer preferences for product qualities and shelf life, and operational constraints such as delivery time and out of stock. This work suggests a data-driven framework for the integration of long-term inventory planning based on DP and real-time pricing adaptation based on RL. The DP module determines the lot sizes in a way to minimize the lot sizes subject to freshness and demand elasticity constraints. The RL module determines the prices under stochastic demands and inventory levels, and it takes into account delivery costs and product aging among different retail channels. Accordingly, the study addresses the following research questions:

- How does product freshness affect customer demand on various channels?
- What are the optimal lot sizes and replenishment policies when there are multiple purchasing and delivery options?
- How do dynamic pricing strategies contribute to waste reduction and profitability amidst heterogeneous consumer preferences?
- How do operational factors such as waste penalties, procurement costs, and freshness guarantees affect overall system performance?

Having established the scope and the questions addressed, the key innovations of the study are as follows, the first being the paper's central finding and the remainder the methodology through which it is obtained and tested:

- A scope condition for perishable inventory systems: whether perishability binds at all, and therefore whether adaptive replenishment and pricing can improve anything, is governed by the ratio of the replenishment cycle to the usable shelf life rather than by the sophistication of the policy. This reframes the delivery schedule as the prior determinant of performance and shows that near-zero waste reported under short replenishment cycles is a property of the delivery calendar rather than of the algorithm.
- A unified framework that decouples strategic replenishment (DP) from adaptive pricing (RL) and coordinates them through a joint fixed point, solving each decision at its natural timescale; its demonstrated benefit over a one-pass sequential procedure is the stabilisation of waste rather than an increase in profit (Section 5.6).
- A hybrid approach that combines RL for adaptive pricing with DP for strategic inventory planning; benchmarking shows that a standalone learner asked to acquire both decisions is decisively the weakest of the methods tested, which is the empirical case for imposing DP structure on replenishment rather than learning it.
- Real operational factors such as demand elasticity, freshness decay, inventory aging, and delivery constraints are considered.
- Freshness is modelled as an endogenous, demand-shaping state variable that affects demand elasticity and consumer channel choice, rather than merely a cost of ageing; this enables explicit modelling of heterogeneous consumer behaviour across online and in-store channels and channel-specific, freshness-aware pricing strategies.
- Actionable managerial insights that help retail managers by providing specific guidance on how to optimize inventory allocation, pricing, and waste reduction across multiple channels.

The remainder of the paper is organized as follows: Section 2 reviews the literature and preliminaries. Section 3 states the problem, while Section 4 outlines the solution approach. Section 5 presents numerical results, while Section 6 discusses managerial implications. Section 7 concludes the study.

2. Literature Review and Preliminaries

2.1. Literature Review

Perishable retail operations research problems can be divided along two major axes of coordination for perishable products: the level of joint versus separate solution of the inventory and pricing models, and the use of optimization versus learning. While significant progress has been made along each of these axes individually, the integration of these methods within the framework of freshness-sensitive omnichannel retail remains in its infancy.

The first category of research is concerned mainly with the inventory management of perishable items. De Moor et al. (2022) improve deep RL for inventory management using reward shaping, showing the improved stability of policies in stochastic environments. Bu et al. (2023) prove the asymptotic robustness of base-stock controls under different lifetimes and demands. Pilati et al. (2024) extend reorder-level controls to include environmental sustainability, while van Donselaar & Broekmeulen, (2025) introduce operational estimators to minimize waste and maximize shelf availability. These papers advance the theory of replenishment; however, pricing is usually modeled as an external parameter, and demand is not a strategic response to coordinated operational decisions.

The second wave of research focuses on dynamic pricing for perishable goods. Song et al. (2021) study the interplay of channel control and pricing, structurally modeling the optimal policies. Feng et al. (2022) incorporate product age and display into demand models. Yang et al. (2022) investigate AI-based pricing with information disclosure approaches. For omnichannel retailing, Zhao et al. (2023) emphasize the importance of reference freshness in dual-channel models, and Liu et al. (2024) investigate BOPS models and market reach. Syed et al. (2024) and Mamoudan et al. (2024) take data analytics and machine learning approaches to inform pricing. Although these papers contribute to our knowledge of price sensitivity and consumer perception, the inventory is frequently treated as a constraint instead of a decision variable to be made jointly, hindering intertemporal coordination.

The third category of studies explicitly couples inventory and pricing decisions through DP or parametric optimization. Keskin et al. (2022) develop a data-driven joint approach under dynamic retail environments. Zhou et al. (2023) and Özbilge et al. (2024) investigate joint pricing approaches incorporating the effects of aging products and donation options. Cachon & Swinney (2009) offer key insights into inventory commitment and price markdowns in the context of strategic consumer behavior, and Dey et al. (2024) extend the analysis of coordination to multi-tier perishable supply chains. These models show that coordination improves system performance, but they remain intractable at scale; the usual remedies are to discretize state spaces or imposing structural assumptions. The traditional DP approach provides important structural insights into such models but becomes computationally infeasible as the state-action space is enlarged by the simultaneous consideration of product freshness decay, continuous pricing, and cross-channel substitution.

Recent developments bring RL into play to overcome these scalability issues. Yavuz & Kaya (2024) show that deep RL can cope with coupled inventory-pricing dynamics in freshness-sensitive settings, and Kavooosi et al. (2025) extend this to continuous action spaces with practical perishability constraints. Nevertheless, the majority of current learning-based solutions are limited to single-channel retail networks and assume perishability as an exogenous decay rate instead of a demand-shaping state variable. The current literature, arranged along four structural aspects, is shown in Table 1. Overall, the impression from the table is that literature generally advances along one or two, but typically no more than two, of the structural aspects at a time. In particular, there is no study that simultaneously considers

omnichannel substitution, freshness effects on demand, joint optimization of inventory and price, and adaptive RL.

Table 1. Structural Positioning of Prior Research and the Present Study

Study	Multi-channel Structure	Endogenous Freshness Dynamics	Joint Inventory–Pricing	Adaptive Learning
De Moor et al. (2022)	No	No	Inventory only	Yes
Bu et al. (2023)	No	No	Inventory only	No
Pilati et al. (2024)	No	Partial	Inventory-focused	No
van Donselaar & Broekmeulen, (2025)	No	No	Inventory-focused	No
Song et al. (2021)	Partial	No	Pricing-focused	No
Feng et al. (2022)	No	Partial	Pricing + lot-sizing	No
Yang et al. (2022)	No	No	Pricing-focused	AI-based
Zhao et al. (2023)	Yes	Partial	Pricing-focused	No
Liu et al. (2024)	Yes	No	Channel strategy	No
Syed et al. (2024)	Partial	No	Pricing-focused	Data-driven
Mamoudan et al. (2024)	No	No	Pricing-focused	ML-based
Keskin et al. (2022)	No	No	Yes	Data-driven
Zhou et al. (2023)	No	Partial	Yes	No
Özbilge et al. (2024)	No	Partial	Yes	No
Cachon & Swinney, (2009)	No	No	Yes	No
Dey et al. (2024)	No	No	Yes	No
Yavuz & Kaya, (2024)	No	Partial	Yes	Deep RL
Kavoosi et al. (2025)	No	Partial	Yes	Deep RL
This Study	Yes	Yes (demand-linked)	Yes	Adaptive RL (continuous)

In this paper, we review the literature on inventory and pricing models, with an emphasis on perishable products in an omnichannel setting. It reveals a major gap in previous research, which has generally focused on only two of four structural dimensions relevant to this issue. It is argued that the effective management of perishables requires an overall approach that takes into account all the dimensions simultaneously, with particular focus on the relationship between freshness, inventory and price over time.

In this study, we propose a novel hybrid framework of DP and RL to handle the complexity of freshness and omnichannel dynamics. This two-pronged approach separates long term strategic planning from short term operational control, creating an alignment that allows for effective management of perishables. The first major contribution of the research is to conceptualize freshness as an endogenous state variable that impacts demand elasticity and consumer channel choice. This is in contrast to traditional models that have focused primarily on the costs associated with product ageing such as disposal or salvage costs.

Secondly, the proposed framework innovatively integrates lot sizing and dynamic continuous pricing in an omnichannel model, where the DP segment determines optimal ordering amounts, long-run inventory strategies and considers factors like delivery lead times and cost asymmetries between online and offline inventory turnover rates. Simultaneously, the RL component deals with dynamic pricing based on the changing demand and inventory condition, coupling pricing to replenishment through the coordination loop rather than setting the two in isolation. Also, the framework provides a

road to scalability that preserves the economic integrity of the model. This work combines structured economic insights with adaptive flexibility, unlike earlier high-dimensional DP solutions that often run into practical solving difficulties or black-box learning frameworks lacking economic clarity.

Finally, the model is well aware of the operational features associated with the various channels, adapted to reflect FIFO inventory management and assortment differences. In summary, the contributions of this study significantly improve four structural elements, i.e., omnichannel retailing dynamics, endogenous freshness modeling, interactivity of pricing and inventory, and adaptive learning capability, and thus advance the current understanding in literature regarding perishable product management in omnichannel environments.

2.2. Preliminaries

A. Dynamic Pricing

With the enhancement of technology, dynamic pricing has evolved from 'haggling' prices with customers. It allows a firm to dynamically change its prices through algorithms and data analytics to increase its revenue while offering and serving the customers with personalized shopping experience (Hendalianpour 2020; Vogt et al. 2025). It is defined as variations in price derived by four main factors—periods or time, places or locations, product configurations, and consumer segmentation—besides general market demand drivers (Liu, Hendalianpour, and Hamzehlou 2021). It has shifted from a static price, which remains the same in all circumstances, to dynamic pricing, where the price constantly changes. This is an example of online retailers charging different prices based on changing customer behavior or in-store algorithm methods modifying prices based on customer identification (Kopalle et al. 2023). This paper addresses the first two: period and place. Temporal price variation occurs while holding location, product, and consumer characteristics constant. One of the major theoretical concepts is reference price formation, an adaptive expectation process based on past price experiences (Blattberg and Neslin 1990; Greenleaf 1995; Kopalle and Winer 1996). Consumer demand responds to the difference between the reference price and the observed price—the greater the perceived gain, the higher the demand, and vice versa. This consideration is consistent with Prospect Theory, describing the way in which buyers make decisions about purchases against their subjective perception of gains and losses. Firms can use such causal mechanisms to their advantage by timing price increases and temporary price reductions aimed at maximizing revenue, especially among promotion-oriented buyers (Mishra, Kushwaha, and Gupta 2024; Kopalle and Lindsey-Mullikin 2003; Lee, Shin, and Moon 2024; Lin, Gui, and Lu 2024). Understanding temporal dynamics provides a strong theoretical basis for modeling the behavior of consumers over time and developing appropriate pricing strategies.

B. Dynamic Programming & Reinforcement Learning

DP offers a structured approach to sequential decision making under uncertainty, owing to which one can model system states and value functions precisely; this also allows inventory managers to see

the effect of current decisions on future costs (Bellman 1966; Porteus 2002). Nevertheless, the theoretical strength of DP leads to computational infeasibility in high dimensional omnichannel settings, particularly for perishable items in which product age and spoilage rates also become a factor. This, sometimes referred to as the “curse of dimensionality,” restricts the applicability of exact DP solutions (Baghalian, Rezapour, and Farahani 2013; Porteus 1990).

Recent studies have overcome some of these limitations by combining Deep Reinforcement Learning (DRL) with DP, which enables the approximation of value functions and the derivation of near-optimal policies. Large state spaces, nonlinearities in interactions across freshness, pricing, and inventory, and demand heterogeneity across channels can be modeled using neural networks in DRL (Dehaybe, Catanzaro, and Chevalier 2024; Lee, Shin, and Moon 2024). Furthermore, DRL allows for end-to-end optimization of ordering, allocation, and pricing policies, enabling reduction of waste and enhancement of margin (Gijbsbrechts et al. 2022).

DRL has been reported to reduce inventory waste, improve demand-forecast accuracy and allocate resources better across channels (Boute et al. 2022; Lee, Shin, and Moon 2024). Such results confirm that the combination of DP with DRL forms a practical framework for making sequential decisions in complex, real-world environments and enables near-optimal policy extraction where traditional DP would be infeasible due to computational reasons.

Still, several challenges remain. For instance, policy stability, large-scale empirical validation, and modeling of sophisticated customer behaviors in omnichannel settings remain open issues. Future research is needed on robust neural network design, validation under realistic operational constraints, and capturing dynamic interactions between freshness, pricing, and inventory in multichannel systems (De Moor, Gijbsbrechts, and Boute 2022; Gijbsbrechts et al. 2022; Yan et al. 2022). Filling these gaps will improve the practical value of DRL-based inventory management and further develop the theoretical underpinnings of sequential decision-making in perishable and multichannel supply chains.

C. Perishable Products

Perishable products are those goods whose quality, functionality, or value deteriorate over time until they become spoiled, depleted, or obsolete. According to Amorim et al., (2019), one of the core concepts in managing such products is the shelf life, which is considered to be the period where the product maintains appropriate conditions for sale or consumption. In highly perishable products like fresh fruits, vegetables, or blood, organoleptic and physical properties degrade continuously, making shelf life stochastic because it depends on product characteristics, storage conditions, and origin. On the other hand, products such as canned foods and beverages have a deterministic shelf life because expiration dates involve limit periods beyond which they cannot be used anymore (Solari et al. 2025).

While prior literature recognizes such operational hurdles, an integrated theoretical framework of product characteristics, supply chain complexity, and sustainability considerations is still lacking. For instance, Buisman et al., (2019) state that the main challenge lies in their short shelf life, while

Amorim et al., (2019) advocate for the need to consider variability resulting from storage and origin with adaptive management strategies. Sensitivity to environmental conditions, together with limited shelf life, may render coordination among all actors along the supply chain highly indispensable because both profitability and consumer satisfaction are concerned (Anggraini, F. Yu, and Wibowo 2025; Labin et al. 2022). Three dimensions are vital in the design of sustainable supply chains for perishable products.

- **Problem Characteristics:** There are certain intrinsic complexities related to perishable supply chains: restricted storage capacity, stringent regulatory requirements, and time-sensitive inventory management. These constraints put additional pressure on finding integrated planning approaches which optimize product flow while minimizing spoilage.
- **Risk and Uncertainty:** Variability in supply and demand, fluctuating prices, and dependence on transport and monitoring technologies contribute to increasing uncertainty. Traditional inventory models do not capture this dynamic. Thus, robust optimization with advanced forecasting is crucial to minimize disruptions and ensure operational stability (Anggraini, F. Yu, and Wibowo 2025).
- **Modeling and Solution Strategies:** Advanced computational techniques, including machine learning, RL, and hybrid predictive models, have the potential to enhance operational efficiency, decision-making, and resilience. Integrating large-scale data allows for real-time monitoring and adaptive management to address stochasticity in product degradation and demand patterns.

The body of literature that applies integration of DP and RL to perishable-product supply chains remains limited, and the existing studies address only isolated components of the broader problem. Mamoudan et al., (2024) examined food-product pricing using a hybrid machine-learning and coordinated approach, demonstrating that independent procurement decisions lead to higher retail profits compared to coordinated strategies. However, their model abstracts away from the core complexities of perishability, specifically inventory ageing, shelf-life dynamics, waste generation, and freshness-based demand variability. As a result, while their findings contribute to pricing theory, they do not provide insights into how Interdependent Decisions influence replenishment decisions or inventory evolution over time. Similarly, Dey et al., (2024) developed digital-technology-enabled coordinated pricing strategies for a three-tier perishable supply chain and evaluated multiple power structures, including centralized, decentralized, and revenue-sharing contracts. Their analysis incorporates important parameters such as quality deterioration, cost optimization coefficients, and technology adoption costs. Yet the study focuses primarily on inter-tier pricing structures in a traditional supply chain setting and does not consider omnichannel retail environments, where both online and in-store channels interact, share inventory, and serve consumers with widely different freshness expectations. Moreover, the absence of dynamic pricing mechanisms, real-time decision-making, and

reinforcement-learning-based adaptation limits the practical applicability of their model in fast-paced retail contexts where demand uncertainty and freshness decay occur simultaneously.

The studies indicate a wide gap in research for the simultaneous consideration of strategic replenishment and lot sizing, omnichannel dynamic pricing, and the operational problems of perishable goods in supply chains. The lack of integrated models neglects the critical interactions of pricing, freshness, delivery speed and inventory management. This paper develops an integrated framework to jointly determine replenishment and RL based dynamic pricing decisions for a freshness-sensitive omnichannel environment. The results provide actionable insights for retailers to improve profitability, service quality and reduce waste.

3. Problem Statement and Mathematical Modeling

The research focuses on the optimal lot sizing, inventory management, and dynamic pricing of perishable products in the context of omnichannel retail, considering the different consumer behaviors, different modes of delivery, and the freshness of the products. In the context of the trends of BOPS, BOHD, and BSPS, the trade-offs between inventory, waste reduction, and profit maximization are critical. The critical issue in the decision process of the retailers relates to the sensitivity of the demand of the customers with respect to the freshness of the products, as the online customers are highly sensitive to the shelf life of the products, and the customers in the stores demonstrate heterogeneous behaviors, impacting the decision process. In the context of dynamic pricing, the focus relates to the reduction of waste, especially with respect to the products that are near expiration and are offered at discounts for sale. On the other hand, the wrong pricing strategies may result in losses, impacting on the profitability of the inventory.

The complexity of the retail operational environment is also influenced by the time sensitivity of the delivery process and the constraints of the last mile of the delivery process, which affects consumer satisfaction and the likelihood of purchases. In addition, inventory shortages lead to loss of sales, which can subsequently lead to revenue loss and consumer dissatisfaction. As the online platform is managed through the warehouse management system technology based on the FIFO principle, the inventory sort process is found to be inconsistent. Hence, an urgent need exists to formulate a strategic model to optimize the allocation of inventory to maximize the sale while minimizing the wastage. Most existing studies do not combine the strategic interactions between retail stakeholders with the way perishable value decays over time. The challenges that were encountered by the study can be summarized as follows:

- **Channel-Specific Dynamics**
 - Online consumers (BOHD/BOPS) request fresher products; they also seek shorter delivery times, while in-store consumers of BSPS show mixed freshness preferences, selecting oldest/newest items.

- Conflicting inventory allocation policies: Online channels use FIFO (via advanced WMS) in order to guarantee freshness, but in-store channels face randomness in terms of stock rotation.
- Last-mile delivery costs and time sensitivity for BOHD compete with BOPS's operational efficiency.
- **Perishability & Waste**
 - Products whose shelf-life $=1$ are sold at discounted prices in-store; unsold items become waste, incurring disposal costs.
 - Dynamic pricing has to consider the freshness decay over time while balancing the demand across channels.
- **Strategic Interactions**
 - Consumers channel choice depends on the price, freshness, and delivery time; hence, the pricing and inventory strategies are interdependent.
 - Retailers have to anticipate competitor actions, such as pricing wars, and consumer behavior, such as sensitivity to discounts, within a dynamic optimization framework.
- **Operational Constraints**
 - Inventory requires optimal lot-sizing decisions to avoid shortages or overstocking.
 - Shortages incur lost-sale penalties while overstocking risks waste, especially for short shelf-life products.

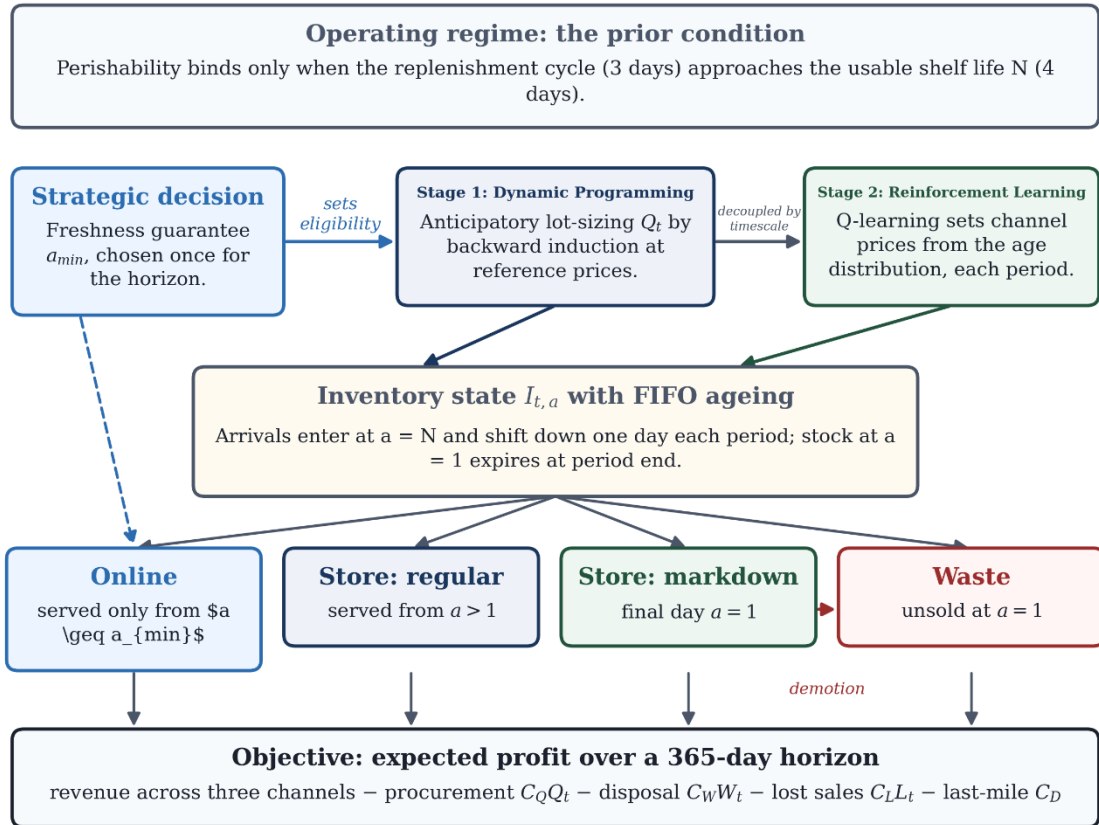


Fig. 1: Research Framework, Single-Objective Optimization with Penalty Terms

The challenges and complexities outlined above provide the premise for this study, which intends to create an integrated optimization model for the determination of lot sizes, the inventory replenishment strategy, and the pricing mechanism for the sale of perishable goods in an omnichannel retail setting. The research framework for the study is outlined in Fig. 1. The factors considered for the purpose of this study are the elasticity of demand, different levels of freshness, sensitivity in pricing, different delivery times, and waste minimization strategies. The hybrid model, using the combination of the DP and RL techniques, offers a structured framework for retailers that could guide strategic decisions for the data-driven purpose of improving efficiency, reducing waste, and increasing profitability in the sale of perishables in omnichannel retailing. The intention here is to create an optimized model that could determine the optimal lot sizes, inventory, and pricing mechanism for the sale of perishables in omnichannel retailing. Specifically, the model is required to:

- Include strategic decision-making between stakeholders: retailers and consumers.
- Optimize channel-specific policies that ensure optimum BOPS vs. BOHD fulfilment under freshness decay and delivery time constraints.
- Minimize waste and shortage penalties while maximizing profitability.

The study shows retailers how to coordinate omnichannel operations and price dynamically against perishability. Integration of DP and RL will address strategic competition and consumer behavior with a novel framework for the management of perishables in a complex retail ecosystem.

3.1. Assumptions

- The initial inventory of each product in the first period is zero. Deliveries do not arrive every day: the retailer is replenished two to three times per week, modelled as a three-day replenishment cycle, against a usable shelf life of $N = 4$ days. This ratio of cycle to shelf life determines whether stock ages before it sells, and Section 5.2 shows whether perishability binds at all.
- Demand is sensitive to product freshness (i.e., expiry date or shelf-life)
- Online consumers are more sensitive to the product's freshness.
- To meet online consumer expectations (for fresh products), retailers can promise min. shelf-life.
- The store's consumer expectations are mixed (some pick the freshest available, some pick items with min. shelf-life, some pick randomly, and some pick the oldest one). The store may try to sort them randomly, but it cannot do it constantly, and different products with different shelf-lives are being mixed.
- The online demand is equipped with full WMS technology so it can follow FIFO.
- Items with shelf-life = 1 (i.e., expiring today) are sold 70% of the original price in store. If the store cannot sell them, they will be considered waste.

- The shortage could occur as a lost sale (not in back order), which incurs a penalty cost.
- Adjust prices based on inventory levels, demand fluctuations, and product freshness to maximize profit or minimize waste.
- Shelf-life data visibility is considered in item-level data, where the inventory system knows how many of each item, with the exact remaining shelf life, is on hand.

3.2. Modeling

Sets

- T : Period (*in this study is a “day”*) $t \in T$
- A : Shelf-life age $a \in A$
- C : Sales channels {online (BOHD/BOPS), store (BSPS)} $c \in C$

Parameters

- C_Q : Cost per unit to procure and replenish inventory (includes transportation, handling).
- C_w : Disposal cost per unit for unsold expired items (environmental fees, labor).
- C_L : Opportunity cost per unit for unmet demand (lost revenue + customer dissatisfaction)
- C_D : Last-mile delivery cost per unit for BOHD orders.
- $X\%$: Discount Rate, Markdown applied to in-store items expiring today ($a = 1$).
- α_{online} : Base demand intercept
- β_{online} : Price elasticity
- γ_{online} : Freshness sensitivity
- α_{store} : Store regular demand intercept
- β_{store} : Store regular price elasticity
- α_{disc} : Store discounted demand intercept
- β_{disc} : Store discounted price elasticity
- Freshness guarantee: minimum remaining shelf-life (in days) promised for online
- a_{min} : orders. DECISION VARIABLE, a_{min} in $\{1, \dots, N\}$, chosen once for the planning horizon (see Section 5.3); it is not a fixed parameter
- f_{BOHD} : Fraction of online sales via BOHD
- N : Maximum shelf-life (days)

Decision Variables

- Q_t : Units ordered and received at the start of period t .
- $p_{online,t}$: Price per unit for online orders in period t .
- $p_{store,t}$: Price per unit for in-store regular sales in period t .

$I_{t,a}$	:	Units of inventory with remaining shelf-life a at the start of t .
$S_{online,t}^a$	:	Units sold online from inventory of age a in t .
$S_{store,t}^a$	:	Units sold in-store from inventory $a > 1$
$S_{store,t}^{disc}$	:	Units sold in-store from inventory $a = 1$.
W_t	:	An unsold unit of $a = 1$ discarded at the end of t .
$L_{online,t}$	:	Online lost sales in t
$L_{store,t}^{reg}$	:	Store regular lost sales in t
$L_{store,t}^{disc}$	:	Store discounted lost sales in t

Demand Functions

$$D_{online,t} = \exp(\alpha_{online} - \beta_{online} \cdot p_{online,t} + \gamma_{online} \cdot \log a_{min}) \quad (1) \quad \text{Online Demand}$$

$$D_{store,t}^{reg} = \exp(\alpha_{store} - \beta_{store} \cdot p_{store,t}) \quad (2a) \quad \begin{array}{l} \text{In-store demand} \\ \text{Regular Sales } a > 1 \end{array}$$

$$D_{store,t}^{disc} = \exp(\alpha_{disc} - \beta_{disc} \cdot X\%p_{store,t}) \quad (2b) \quad \begin{array}{l} \text{In-store demand} \\ \text{Discounted Sales } (a = 1) \end{array}$$

Objective Function

$$\begin{aligned} \max \sum_{t=1}^T & \left[p_{online,t} \cdot \sum_{a>1, a \geq a_{min}} S_{online,t}^a + p_{store,t} \sum_{a>1, a \geq a_{min}} S_{store,t}^a \right. \\ & + (X \cdot p_{store,t}) \cdot S_{store,t}^{disc} \\ & - \left(C_Q Q_t + C_W W_t + C_L (L_{online,t} + L_{store,t}^{reg} + L_{store,t}^{disc}) \right) \\ & \left. + C_D f_{BOHD,t} + \sum_{a>1, a \geq a_{min}} S_{online,t}^a \right] \end{aligned} \quad (3) \quad \begin{array}{l} S_{BOHD,t} : \\ \text{BOHD-} \\ \text{specific sales} \\ \text{(subset of } S_{online,t}) \end{array}$$

Constraints

$$I_{t+1,N} = Q_t \quad (\text{New Inventory}) \quad (4a) \quad \text{Inventory Balance}$$

$$I_{t+1,a} = I_{t,a+1} - S_{online,t}^{a+1} - S_{store,t}^{a+1} \quad \forall t \in T, a > 1 \quad (4b)$$

$$\begin{array}{l} \text{Only use inventory with } a \geq a_{min} \\ S_{online,t}^a \leq I_{t,a} \quad \forall t \in T, a \geq a_{min} \end{array} \quad (5a) \quad \begin{array}{l} \text{Online Sales Fulfillment} \\ \text{(FIFO)} \end{array}$$

Total online sales cannot exceed demand

$$\sum_{a>1, a \geq a_{\min}}^T S_{\text{online},t}^a$$

$$\leq \exp(\alpha_{\text{online}} - \beta_{\text{online}} \cdot p_{\text{online},t} + \gamma_{\text{online}} \cdot \log a_{\min}) \quad \forall t \in T$$

(5b)

Cumulative online sales up to age a cannot exceed demand, ensuring older stock is prioritized.

$$L_{\text{online},t} \geq \alpha_{\text{online}} - \beta_{\text{online}} \cdot p_{\text{online},t} + \gamma_{\text{online}} \cdot a_{\min}$$

$$- \sum_{a>1, a \geq a_{\min}} S_{\text{online},t}^a$$

(5c) Lost sales

$$S_{\text{store},t}^a \leq I_{t,a} \quad \forall t \in T, a \geq 1$$

(6a)

$$\sum_{a>1, a \geq a_{\min}}^T S_{\text{store},t}^a$$

$$\leq \exp(\alpha_{\text{store}} - \beta_{\text{store}} \cdot p_{\text{store},t}) \quad \forall t \in T$$

(6b)

Store Regular Sales $a > 1$

$$L_{\text{store},t}^{\text{reg}} \geq (\alpha_{\text{store}} - \beta_{\text{store}} \cdot p_{\text{store},t}) - \sum_{a>1, a \geq a_{\min}} S_{\text{store},t}^a$$

(6c)

$$S_{\text{store},t}^{\text{disc}} \leq I_{t,1} \quad \forall t \in T$$

(7a)

$$S_{\text{store},t}^{\text{disc}} \leq \exp[\alpha_{\text{disc}} - \beta_{\text{disc}}(X \cdot p_{\text{store},t})] \quad \forall t \in T$$

(7b)

Store Discounted Sales $a = 1$

$$L_{\text{store},t}^{\text{disc}} \geq (\alpha_{\text{disc}} - \beta_{\text{disc}}(X \cdot p_{\text{store},t})) - S_{\text{store},t}^{\text{disc}}$$

(7c)

$$W_t = \max(I_{t,1} - S_{\text{store},t}^{\text{disc}}, 0)$$

(7)

Waste Calculation

$$L_t = \max(D_{\text{online},t} - S_{\text{online},t}, 0)$$

(8)

Lost Sales Penalty

$$+ \max(D_{\text{store},t}^{\text{reg}} - S_{\text{store},t}^{\text{reg}}, 0)$$

$$Q_t, I_{t,a}, S_{\text{online},t}^a, S_{\text{store},t}^a, S_{\text{store},t}^{\text{disc}}, W_t \geq 0$$

(9)

$\forall t \in T, a \in A$

$$L_{\text{online},t}, L_{\text{store},t}^{\text{reg}}, L_{\text{store},t}^{\text{disc}} \geq 0$$

(10)

$\forall t \in T$

$$p_{\text{online},t} \in [p_{\min}, p_{\max}] \text{ and } p_{\text{store},t} \in [p_{\min}, p_{\max}]$$

(11)

$\forall t \in T$

$$I_{1,a} = 0$$

(12)

$\forall a \in A$ (zero initial inventory)

4. Solution Approach

The hybrid methodology that combines the techniques of DP and RL has been used to tackle the unique challenges that are posed by the management of perishable goods in the context of omnichannel retailing, in which strategic planning and adaptive decisions are both critical. The former offers a deterministic and rigorous methodology for optimizing decisions regarding the lot sizing problem, which ensures that the inventory replenishment decisions are made in such a manner that there are minimal instances of both waste and loss due to inadequate inventory levels. The backward induction methodology used in the context of the DP model ensures that the optimal order quantities Q_t are determined for the different stages while considering the future states as well. The determination of the order quantities becomes all the more relevant in the context of managing perishable goods, as these

have a finite life and the decisions regarding the order quantities are made in a manner that the overall cost, including the cost of waste and loss, is optimized for the entire planning horizon. However, the application of the DP methodology faces significant challenges in the context of the complexities that are posed by the uncertain demand, price sensitivity, and competitive pricing strategies, which are stochastic in nature.

Additionally, the proposed approach has a strong link to the operation of an omnichannel environment. Indeed, the online channels (BOHD/BOPS) demand strict fulfillment rates, which in turn require the application of FIFO principles to ensure the sale of the oldest inventory first. On the contrary, the in-store channels (BSPS) benefit from the RL component, which facilitates the adjustment of prices to expedite the sale of products close to their expiration date. By differentiating the decision on lot size (optimized by DP) from the decision on the pricing strategy (learned by RL), the proposed model efficiently balances the trade-offs between the two conflicting objectives of minimizing waste by efficiently managing the inventory and maximizing profit by exploiting the price elasticity of demand. To ensure that the two decisions are genuinely coordinated rather than solved in a single one-directional pass, the replenishment target and the pricing policy are solved to a joint fixed point: the DP re-optimises the order-up-to level under the prices the RL agent actually learns, and the two are iterated until the order target no longer moves (Algorithm 1, Stage 3). The framework is benchmarked in Section 5.6 against a one-pass sequential variant, a standalone DP with static list pricing, a standalone RL agent that learns both ordering and pricing, and a two-stage stochastic program, all within the binding regime and under common random numbers. The model is evaluated in Section 5. The Hybrid DP–RL Model for Perishable Goods is proposed below in the parametric form.

Algorithm 1. Hybrid DP–RL procedure for Problem Statement

Inputs: planning horizon T ; shelf-life L ; channel demand functions $D_p(p, \text{age})$; discount factor δ ;
procurement cost c_o , waste penalty π_m , lost-sales penalty π_l ; price grid P .
RL hyperparameters: learning rate α , exploration rate ϵ , episodes N .
Output: order policy $Q^*(x)$ and pricing policy $\pi^*(x)$.

Stage 1 (Offline DP, lot-sizing under expected price):
for each terminal state x at $t = T$: $V_t(x) \leftarrow 0$
for $t = T-1$ down to 0:
for each inventory-age state $x \in \mathcal{X}$:
for each feasible order quantity $q \in \{0, \dots, Q_{\max}\}$:
Compute expected one-period profit $g(x, q)$ using mean demand and current price proxy $\bar{p}$
Compute expected next state $E[x' | x, q]$ under FIFO aging and demand realisation
 $Q_t(x, q) \leftarrow g(x, q) + \delta \cdot E[V_{t+1}(x')]$
 $V_t(x) \leftarrow \max_o Q_t(x, q)$; $Q^*(x) \leftarrow \operatorname{argmax}_o Q_t(x, q)$

Stage 2 (Online RL, channel-specific dynamic pricing given Q^*):
Initialize Q-table $Q_{ri}(x, p) \leftarrow 0$ for all $(x, p) \in \mathcal{X} \times \mathcal{P}$
for episode = 1 to N :
 $x \leftarrow$ initial inventory state
for $t = 0$ to $T-1$:
Order $q \leftarrow Q^*(x)$ (replenishment from DP policy)

Choose price $p \in \mathcal{P}$ via ε -greedy on $Q_{ri}(x, \cdot)$
Sample stochastic demand $d \sim D(p, age)$
Allocate sales by FIFO; compute realized profit $r = revenue - c_o q - \pi_m \cdot waste - \pi_l \cdot$
lost
 $x' \leftarrow$ next state after aging and demand realisation
 $Q_{ri}(x, p) \leftarrow Q_{ri}(x, p) + \alpha[r + \delta \cdot \max_{p'} Q_{ri}(x', p') - Q_{ri}(x, p)]$
 $x \leftarrow x'$
Decay ε
 $\pi^*(x) \leftarrow \arg\max_p Q_{ri}(x, p)$
Return (Q^*, π^*) .

A) Demand model (deterministic mean functions used in DP expectation; stochastic draws used in simulation/RL):

I) Online mean demand:

$$d_{online}(p_{online,t}) = \alpha_{online} \cdot \exp\left(-\beta_{online} \cdot p_{online,t} + \gamma_{online} \frac{a_{min}}{N}\right)$$

II) Store-regular mean demand:

$$d_{store}(p_{store,t}) = \alpha_{store} \cdot \exp(-\beta_{store} \cdot p_{store,t})$$

III) Store-discount means demand (price indexed to discounted store price $X \cdot p_{store}$)

$$d_{disc}(p_{store,t}) = \alpha_{disc} \cdot \exp(-\beta_{disc} \cdot X \cdot p_{store,t})$$

IV) Discount factor for intertemporal value: $\delta \in (0,1]$

B) State, Decisions, and Inventory Dynamics

I) State

- $I_{t,a}$: on-hand inventory at start of period t with age a ($1 =$ will expire at end of t ; $N =$ freshest).
- Vector $I_{t,a} = (I_t^1, \dots, I_t^N)$

II) Decisions

- Lot sizing (DP): order quantity $Q_t \in \mathbb{Q} \subset \mathbb{Z} \geq 0$
- Pricing (RL): $p_{online,t}, p_{store,t} \in [p_{min}, p_{max}]$ chosen from discretized action sets.

III) Within-period sequence

- Ageing and replenishment (start of period t): $\tilde{I}_t^1 = I_t^2, \tilde{I}_t^2 = I_t^3, \dots, \tilde{I}_t^{N-1} = I_t^N, \tilde{I}_t^N = Q_t$
Denote $\tilde{I}_t = (\tilde{I}_t^1, \dots, \tilde{I}_t^N)$
- Demands (means) at prices $(p_{online,t}, p_{store,t})$

$$D_{online,t} \sim \text{Stoch} \left(d_{online}(p_{online,t}) \right), D_{store,t} \sim \text{Stoch} \left(d_{store}(p_{store,t}) \right), D_{disc,t} \sim \text{Stoch} \left(d_{disc}(p_{store,t}) \right)$$

DP's expectation uses the means $d(0)$; simulation/RL draw from positive distributions (e.g., log-normal) centered at these means.

C) Fulfilment rules (FIFO by age buckets):

- online: allocated from ages $a = a_{min}, a_{min} + 1, \dots, N$
- Store-regular: allocate from ages $a = 2, 3, \dots, N$
- Store-discount: allocated from age $a = 1$ only.

I) Define sales vectors $S_{online,t}^a, S_{store,t}^a \geq 0$ and scalar $S_{disc,t}^a \geq 0$ such that:

$$\sum_{a=a_{min}}^N S_{online,t}^a \leq D_{online,t}, \sum_{a=2}^N S_{store,t}^a \leq D_{store,t}, S_{disc,t}^a \leq D_{disc,t}$$

and they respect availability (FIFO) from $\tilde{I}_t$.

II) Lost sales:

$$L_{online,t} = D_{online,t} - \sum_{a \geq a_{min}}^N S_{online,t}^a, L_{store,t}^{reg} = D_{store,t} - \sum_{a \geq 2}^N S_{store,t}^a, L_{store,t}^{disc} = D_{disc,t} - S_{disc,t}^a$$

III) Expiration (waste) at period end: $W_t = \tilde{I}_t^1 - S_t^{disc}$
(all remaining age 1 after selling is waste).

IV) Next state I_{t+1} (leftovers after sales, already age-shifted by construction): $I_{t+1}^1 = 0, I_{t+1}^a = \tilde{I}_t^a - S_{online,t}^a - S_{store,t}^a$ (for $a = 2, \dots, N$) with $S_{disc,t}^a$ taken from $a = 1$.

D) Per-period Profit (given $Q_t, p_{online,t}, p_{store,t}$):

I) Revenue

$$R_t = p_{online,t} \sum_{a \geq a_{min}}^N S_{online,t}^a + p_{store,t} \sum_{a \geq 2}^N S_{store,t}^a + X p_{store,t} S_{disc,t}$$

II) Cost

$$C_t = C_Q \cdot Q_t + C_w \cdot W_t + C_L (L_{online,t}, L_{store,t}^{reg}, L_{store,t}^{disc}) + C_{DFBOHD} \left(\sum_{a \geq a_{min}} S_{online,t}^a \right)$$

III) Profit

$$\Pi_t = R_t - C_t$$

E) DP for Lot Sizing (with Expected Demand)

State space discretization for DP (inventory level $\times$ “oldest age present”):

I) Inventory grid: $x \in \{0, \Delta, 2\Delta, \dots, \bar{I}\}$ where $\Delta_{inventory_step}$ and $\bar{I}_{max_inventory}$

II) Age index $a \in \{1, \dots, N\}$

Bellman recursion for value function $V_t(x, a)$ with prices treated as exogenous (or using an average price proxy $\bar{p}$):

$$V_t(x, a) = \max_{Q_t \in \mathcal{Q}} \left\{ \underbrace{\mathbb{E}[\Pi_t | x, a, Q, \bar{p}]}_{\text{expecte one period profit}} + \delta V_{t+1}(x', a') \right\}, V_{T+1}(0) = 0$$

where the transition $(x, a, Q) \mapsto (x', a')$ is induced by the ageing, FIFO allocation, and post-sales leftovers mapped back to the discretized next-period state. The DP policy is:

$$Q_t^*(x, a) \in \arg \max_{Q_t \in \mathcal{Q}} \{0\}$$

F) RL for Pricing (Q-Learning, Discrete Actions)

I) Define a pricing state s from the real inventory vector I_t by:

$$s(I_t) = \left(\left\lceil \left(\sum_a I_{t,a} \right) / \Delta \right\rceil - 1 \right) N + \text{age_state}(I_t)$$

where $\text{age_state}(I_t) = \min(\max\{a: I_{t,a} > 0\}, N)$, and clamp to the valid range.

Action spaces (discrete price grids of size A):

$$A_{\text{online}} = \{p_{\min} + \frac{k}{A}(p_{\max} - p_{\min})\}_{k=1}^A$$

$$A_{\text{store}} = \{p_{\min} + \frac{k}{A}(p_{\max} - p_{\min})\}_{k=1}^A$$

Given the DP lot size Q_t^* at the mapped DP state, the per-period reward to pricing is the realized profit Π_t from Section E under stochastic demands.

III) Q-updates (learning rate η , discount δ):

$$Q_{\text{online}}(s, a) \leftarrow Q_{\text{online}}(s, a) + \eta \left[\Pi_t + \delta \max_{a'} Q_{\text{online}}(s', a') - Q_{\text{online}}(s, a) \right]$$

$$Q_{\text{store}}(s, a) \leftarrow Q_{\text{store}}(s, a) + \eta \left[\Pi_t + \delta \max_{a'} Q_{\text{store}}(s', a') - Q_{\text{store}}(s, a) \right]$$

Exploration: ϵ -greedy on each channel's action set.

Stage 3 (Coordination, joint fixed point). Stages 1 and 2 are embedded in an outer loop so that replenishment and pricing are solved jointly rather than in a single pass. The DP order-up-to target of Stage 1 is computed under an expected price pair; the RL agent of Stage 2 then learns channel prices given that target; the expected price pair is updated to the demand-weighted prices the learned policy actually charges; and Stages 1-2 are repeated until the order-up-to target no longer changes. Because the optimal target falls as prices rise, this feedback lowers the target to match the prices the agent learns, which is the coordination a one-pass DP-then-RL procedure lacks.

5. Numerical Results and Discussion

This section reports and interprets the performance of the framework. Section 5.1 states the calibration and protocol. Section 5.2 establishes the condition under which perishability binds at all, which proves to be the governing result of the study. Section 5.3 reports the freshness-guarantee decision. Section 5.4 tests robustness across the cost and behavioural parameters, and Section 5.5 reports the precision of the estimator and the claims it will not support.

5.1. Calibration and experimental protocol

The instance is a transparently calibrated simulation of a fresh-produce retailer rather than an estimate from transaction records. This bounds the claims: the results describe the structure of the trade-offs a retailer of this shape faces, not a forecast of its accounts.

The usable shelf life of strawberries at retail is four days ($N = 4$). Deliveries arrive two to three times per week, modelled as a three-day replenishment cycle. Procurement cost is 8.0 per unit, disposal 2.0, the lost-sale penalty 3.0 and last-mile fulfilment 1.5 per online unit, with 60% of online volume delivered to the home. Units on their final day ($a = 1$) are marked down to $X = 70\%$ of the store price. Demand shocks are multiplicative lognormal with unit mean and standard deviation 0.30. The planning horizon is 365 days.

Two protocol choices are load bearing rather than cosmetic, for reasons Section 5.5 makes concrete. Evaluation uses common random numbers, so every policy meets an identical demand path within a replication and all comparisons are paired. And each reported figure averages over three to five independent training runs of the Q-learning agent rather than resting on one. Only the two demand intercepts are calibrated, and they are set so that sales volume matches the retailer’s operating point. Everything else in Table 2 is an outcome of the model, not a target of the fit, which is what makes the table a test rather than a tautology.

Table 2. Calibration against the operating point ($a_{min} = 2$, five training seeds, 25 replications).

Quantity	Operating point	Calibrated model	Status
Units sold per day	38.8	37.4	calibration target
Online: store split	approximately even	5,942: 6,672	calibration target
Mean inventory (units)	52	47.9	outcome
Waste (% of units sold)	8–12% (fresh-produce norm)	7.2%	outcome
Service level	—	84.9%	outcome

Two outcomes deserve emphasis. Mean inventory of 47.9 units against a reported 52 was not fitted, and neither was the waste rate of 7.2%, which nonetheless lands inside the band reported for fresh produce. That an optimised policy sits at the lower edge of that band is itself informative: it suggests the observed industry range reflects operating practice rather than a physical floor, and that roughly a third of it is recoverable by policy alone. Section 5.2 explains why a model of this kind can report near-zero waste as an artefact rather than an achievement.

5.2. The condition under which perishability binds

A unit is discarded only if it survives $N - 1$ days unsold. Whether that ever happens is governed by the ratio of the replenishment cycle to the shelf life, and this ratio, rather than the sophistication of the policy, decides whether a perishable-inventory model has anything to optimise at all.

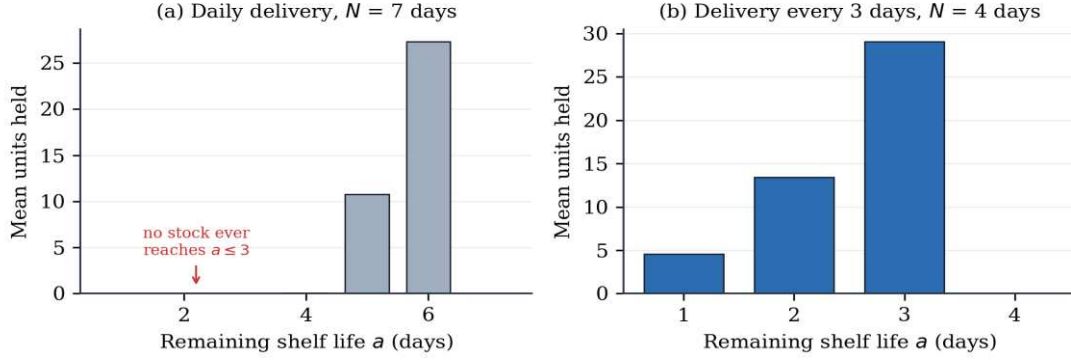


Fig. 2. Mean inventory by remaining shelf life. Under daily delivery against a seven-day life (a) stock never ages; under a three-day cycle against a four-day life (b) it does.

Fig. 2 shows why. In panel (a), under daily replenishment against a seven-day life, stock concentrates at $a = 6$ and no unit ever reaches $a \leq 3$. The age distribution is effectively degenerate: inventory turns over in under a day, expiry is a tail event, and the eligibility constraint $a \geq a_{min}$ is slack for any guarantee the retailer might plausibly offer. A model in this regime will report near-zero waste no matter how the policy is designed and will therefore attribute to the policy an outcome that is a property of the calendar. Panel (b) shows the same system under the schedule the retailer actually operates. Stock now spans the full age range, and expiry is a live risk the policy must manage. The distinction matters beyond this instance, because it identifies a scope condition for an entire literature. Studies of perishable inventory that assume continuous or daily replenishment against a multi-day shelf life are, without saying so, studying a system in which perishability is inert. Their reported waste figures are not evidence about the policy.

Table 3. Performance across replenishment cycles ($N = 4$, $a_{min} = 2$, three training seeds).

Replenishment cycle	Profit	s.d. (seeds)	Waste (units)	Waste (%)	Service (%)	Mean inventory
Every 1 day	80,336	213	4	0.04	81.7	38.1
Every 2 days	74,164	1,151	605	4.49	85.8	56.8
Every 3 days	66,575	1,821	911	6.58	84.8	47.2
Every 4 days	31,812	2,437	774	6.00	69.3	47.8

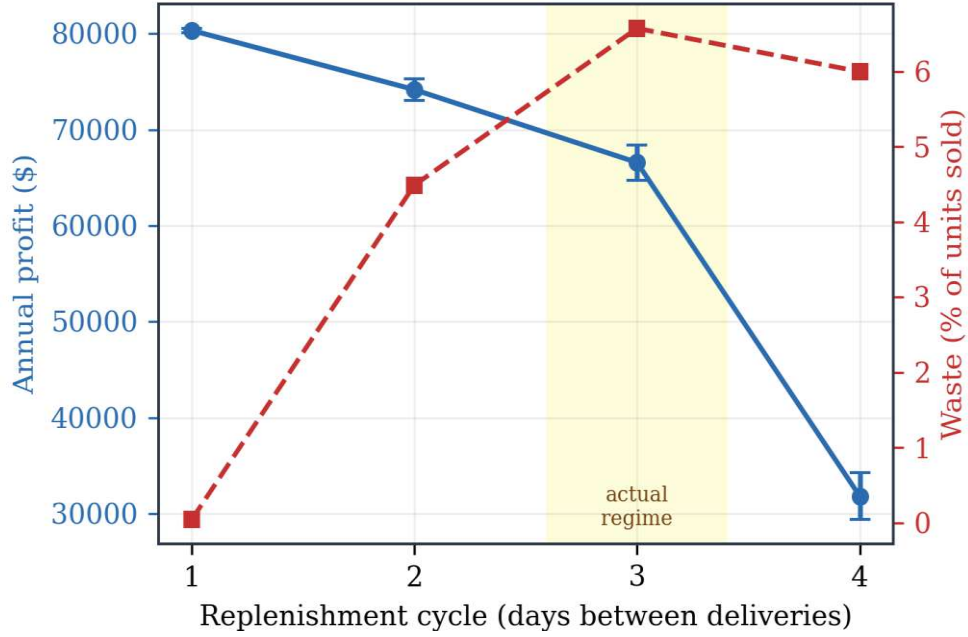


Fig. 3. Profit and waste against the replenishment cycle. (The shaded band marks the regime the retailer actually operates)

Table 3 and Fig. 3 quantify the transition. Waste climbs from 0.04% under daily delivery to 6.58% at a three-day cycle, a factor of roughly one hundred and sixty, while profit falls monotonically as inventory is held longer and financed for longer. The collapse at a four-day cycle, where the cycle equals shelf life, has a different cause from the rest of the curve and repays attention: waste actually falls slightly, from 6.58% to 6.00%, while service drops from 84.8% to 69.3% and profit halves. The retailer cannot cover the tail of a cycle as long as the life of the product without ordering quantities that expire before they can be sold, so the optimiser retreats, protects margin, and simply forgoes the demand. Once the cycle reaches the shelf life, the binding constraint switches from waste to availability, and the business ceases to be viable in its current form rather than merely becoming wasteful. This same ratio also governs whether the coordination studied in Section 5.6 has anything to act on: the order-up-to target responds to price only where perishability binds and demand is price-elastic, so the benchmark comparison there is deliberately run at the price-elastic operating point rather than the base calibration used above, in which the target is insensitive to price and the methods coincide.

5.3. The freshness guarantee as a decision

Treating the guarantee a_{min} as a decision rather than a fixed parameter yields Table 4.

Table 4. Performance against the freshness guarantee ($\gamma_{online} = 0.3$, five training seeds, 25 replications).

a_{min} (days)	Profit	s.d. (seeds)	Waste (units)	Waste (%)	Service (%)	Online units
------------------	--------	--------------	---------------	-----------	-------------	--------------

1	68,056	1,960	318	2.6	83.9	5,401
2	65,848	1,408	985	7.2	84.9	5,942
3	47,352	5,072	1,612	12.1	76.0	4,480
4	22,144	6,999	2,571	21.6	68.8	2,746

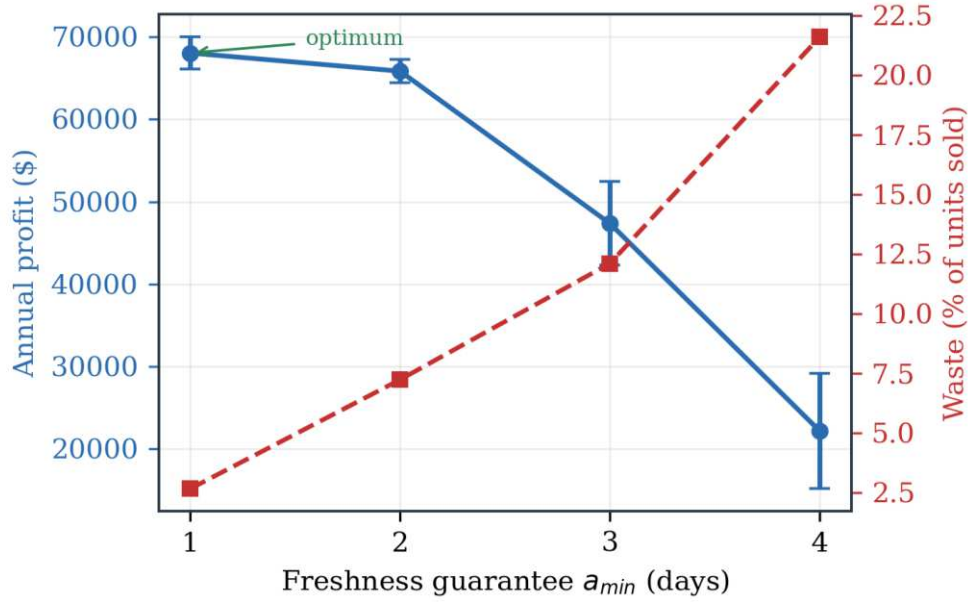


Fig. 4. Profit and waste against the freshness guarantee. Error bars give the standard deviation across training seeds.

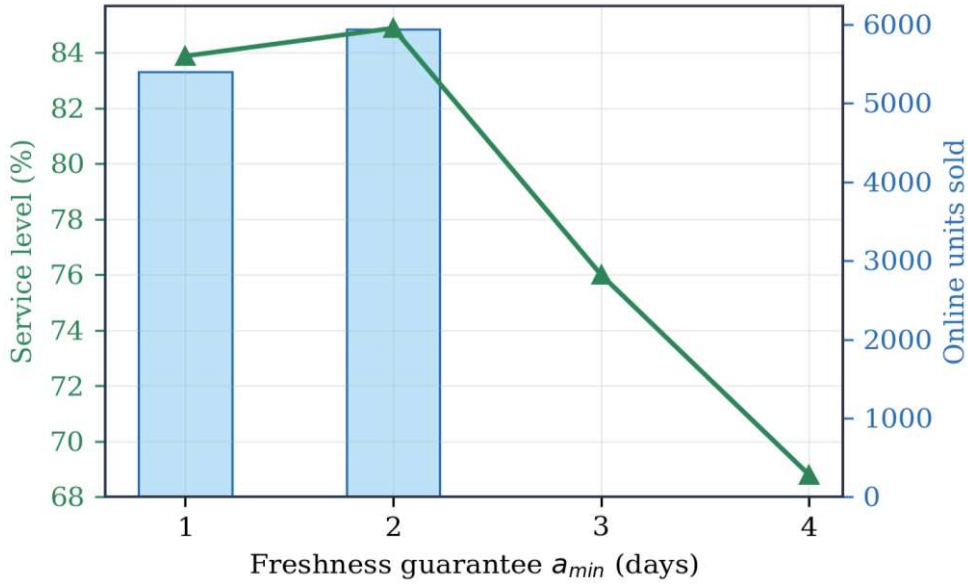


Fig. 5. Service level (line) and online volume (bars) against the guarantee. Beyond two days, eligible stock becomes scarce enough to cost availability.

The demand side behaves exactly as the specification requires: online volume rises from 5,401 to 5,942 units when the promise moves from one day to two, a gain of 10%, and the guarantee is doing real commercial work. The supply side, however, more than cancels it. Every unit below the threshold is excluded from the online channel and demoted to the store, then to markdown, then to disposal, and

waste triples from 318 to 985 units. Profits fall by 3.2%. The comparison is paired across seeds, and the sign is stable on every one of the five.

The interpretive point is not simply that profit declines but that the two curves in Fig. 4 diverge rather than trade off. A conventional operations trade-off would show one metric improving as the other worsens, leaving the manager a genuine choice about where to sit. Here profit and waste deteriorate together as the promise tightens, so there is no frontier to negotiate and no defensible reason to move up the axis. The guarantee is not a lever with two settings and a preference between them; over this range it is simply a cost.

Fig. 5 explains the mechanism at the far end. Service holds at roughly 84% for one- and two-day promises, then falls to 76.0% and 68.8%. At a three-day guarantee against a four-day life, only stock in its first day of shelf presence is eligible for online sale, so the online channel is competing for a sliver of each delivery. The guarantee stops rationing freshness and starts rationing availability, and the demand it attracted can no longer be served.

5.4. Sensitivity analysis

The results above are conditional on a particular cost structure. To establish which conclusions are robust and which are artefacts of the calibration, each parameter was varied one at a time around the base case, holding the rest fixed, with every cell averaged over three training seeds. Table 5 reports the outcomes and Fig. 6 ranks the parameters by the profit swing they induce.

Table 5. One-at-a-time sensitivity analysis ($a_{min} = 2$, three training seeds, 25 replications).

Parameter	Value	Profit	Δ vs base	Waste (%)	Service (%)
Base case	—	66,648	—	6.6	84.8
Procurement cost	6	95,023	+42.6%	6.0	83.8
Procurement cost	10	33,727	-49.4%	10.3	82.7
Disposal cost	0.5	68,698	+3.1%	9.0	85.5
Disposal cost	5	66,507	-0.2%	4.0	82.9
Lost-sale penalty	1	73,417	+10.2%	5.1	80.8
Lost-sale penalty	6	60,314	-9.5%	11.1	86.5
Markdown multiplier	0.5	62,731	-5.9%	5.1	76.7
Markdown multiplier	0.85	68,841	+3.3%	7.5	88.0
Demand volatility	0.2	62,363	-6.4%	9.4	83.4
Demand volatility	0.5	60,646	-9.0%	10.7	82.0
Freshness elasticity	0.15	65,047	-2.4%	6.5	84.6
Freshness elasticity	0.6	73,993	+11.0%	6.3	85.7

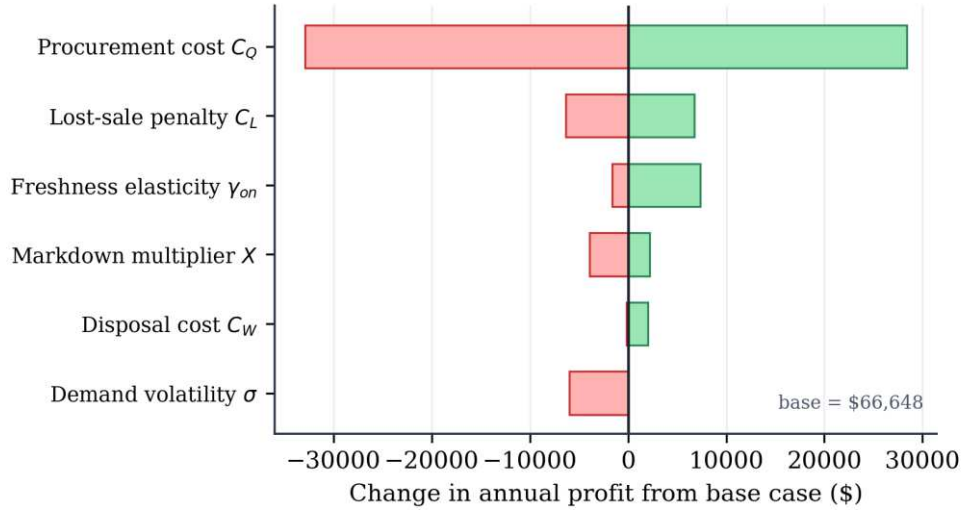


Fig. 6. Tornado diagram. Parameters ranked by the profit swing induced across their tested range.

Fig. 6 delivers an uncomfortable result for the framework itself. Procurement cost dominates everything else by an order of magnitude: moving C_Q from 6 to 10, a 67% increase, moves annual profit from 95,023 to 33,727, a swing of 61,296 against a base of 66,648. No other parameter comes close. For a retailer, the honest reading is that sourcing terms are worth more than any refinement of the pricing policy, and that a procurement negotiation of a few percent dominates the entire optimisation problem this paper solves. The framework is worth deploying, but it is worth deploying second.

The disposal cost yields the most useful managerial finding in the study, and it is one that only a sensitivity analysis could surface. Raising C_W from 0.5 to 5, a tenfold increase, changes profit by only -3.2%, but it cuts waste from 9.0% to 4.0%, a reduction of 55%. Waste is therefore cheap to eliminate in this system. A retailer under regulatory or reputational pressure to cut food waste can do so by raising the internal shadow price of disposal and will surrender only a few percent of profit for more than half the waste. This is a rare case where a sustainability objective and a commercial one are nearly separable, and it is invisible without varying the parameter.

The lost-sale penalty shows the opposite: a genuine expensive trade-off. Raising C_L from 1 to 6 lifts service from 80.8% to 86.5%, but profit falls by 18% and waste more than doubles, from 5.1% to 11.1%. Pursuing availability in a perishable category means ordering into uncertainty, and the stock that covers the demand that does not arrive becomes waste. Service targets set without reference to shelf life are therefore expensive twice over, in margin and in disposal, and a retailer that treats its service level as a fixed corporate constant is buying both.

Markdown depth runs against the intuition that clearance protects perishables. A shallow markdown ($X = 0.85$) outperforms a deep one ($X = 0.5$) by 9.7%. Deep discounting does reduce waste, from 7.5% to 5.1%, but it surrenders more margin than the disposal it avoids is worth. One caveat belongs with this result rather than in a footnote: the service metric counts unmet markdown demand as a lost sale, so the apparent service falling under deep discounting is partly definitional, because a

deep discount attracts bargain demand that only expiring stock can serve. The profit ranking does not depend on that convention; the service ranking does.

Freshness elasticity behaves as the specification predicts. Raising γ_{online} from 0.15 to 0.6 lifts profit at a two-day guarantee from 65,047 to 73,993, since the same promise buys more demand. This is consistent with the guarantee becoming worth offering where customers value it strongly enough, but it does not establish where the switch occurs: locating that threshold requires the full guarantee grid to be re-run at each elasticity with seed averaging, which is left to future work rather than asserted here.

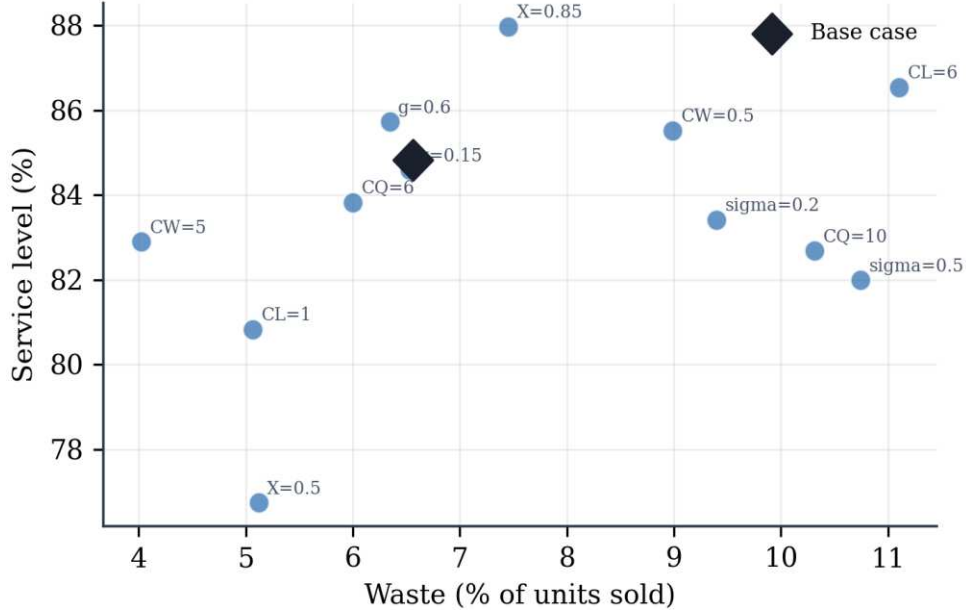


Fig. 7. Waste and service outcomes across all sensitivity settings. No parameter reaches the low-waste, high-service corner.

Fig. 7 plots every setting in waste–service space, and the empty region is the informative part. No configuration reaches the low-waste, high-service corner. Settings that buy service (a high lost-sale penalty, a shallow markdown) carry more waste; settings that suppress waste (a high disposal cost, a deep markdown) surrender service. The cloud traces a frontier that the parameters move along rather than outward, which is the clearest statement available here of what the policy can and cannot do: within a given replenishment regime, the framework chooses a position on the frontier, but it does not shift it. Only the delivery cycle does that, and Table 3 shows the cost of moving it.

One result does not fit and is reported rather than smoothed. Demand volatility is non-monotone: profit at $\sigma = 0.20$ (62,363) is lower than at the base $\sigma = 0.30$ (66,648), before falling again at $\sigma = 0.50$ (60,646). The base-case advantage is roughly 4,285, around 2.5 training standard deviations, so it sits only marginally outside the noise documented in Section 5.5. We do not claim that moderate volatility is beneficial; the more probable explanation is that the reference-price approximation inside the DP is better matched to the base calibration than to the low-variance one. Resolving it would require more training seeds than were run.

5.5. Precision and what it rules out

Repeated training of the Q-learning agent on an identical instance returns materially different policies. Across five training seeds at $a_{min} = 1$, annual profit spans roughly 4,901 with a standard deviation of 1,960. At $a_{min} = 3$ and $a_{min} = 4$ the seed standard deviations rise to 5,072 and 6,999: a scarcer eligible pool makes the pricing problem harder to learn, so the variance of the estimator is itself a function of the decision being evaluated, and comparisons across the grid are heteroscedastic. Two consequences follow and are reported rather than suppressed. First, any comparison resting on a single training run is unreliable at this margin. In preliminary work the ranking of the one-day and two-day guarantees reversed under a different seed, and only paired seed averaging resolved it; a single-run study would have reported the opposite headline with equal confidence. Second, differences below roughly two thousand cannot be separated from training variance, and the framework should not be credited with margins beneath that floor. The gaps in Tables 3 and 4 survive because they are large, not because the estimator is sharp. We note this as a general caution for the reinforcement-learning-in-operations literature, where single-run results remain common and where the seed variance is rarely reported at all.

5.6. Benchmarking against standalone and stochastic methods

The associate-editor concern that a decoupled DP-then-RL procedure may not constitute genuine coordination is addressed directly here. We compare five methods in the binding regime (three-day cycle, four-day life) under common random numbers with paired evaluation and seed averaging: the coordinated DP-RL of Section 4 solved to a joint fixed point; a one-pass sequential DP-then-RL in which the DP orders under a fixed mid-range price proxy and the RL agent then prices given that fixed order policy, with no feedback; a standalone DP paired with a single static list price; a standalone reinforcement-learning agent that must learn both ordering and pricing over the joint action space; and a two-stage stochastic program that fixes an order-up-to level and a static price pair against sampled demand scenarios with the period allocation as recourse. All results are averaged over eight training seeds with twenty-five paired replications each; the calibration is the price-elastic operating point at which the order-up-to target responds to price, which is the regime in which coordination is testable.

Table 6. Benchmark comparison in the binding regime (eight training seeds, 25 paired replications).

Method	Profit	s.d. (seeds)	Waste (%)	Service (%)	Mean inv.
Coordinated DP-RL (fixed point)	87,962	5,004	3.7	79.0	35.1
Sequential DP-then-RL (one pass)	92,189	5,598	10.3	84.6	52.5
Standalone DP + static list price	86,123	--	1.4	70.7	40.3
Standalone RL (orders + prices)	49,658	12,996	0.4	55.0	24.0
Two-stage stochastic program	87,901	--	2.9	72.4	40.8

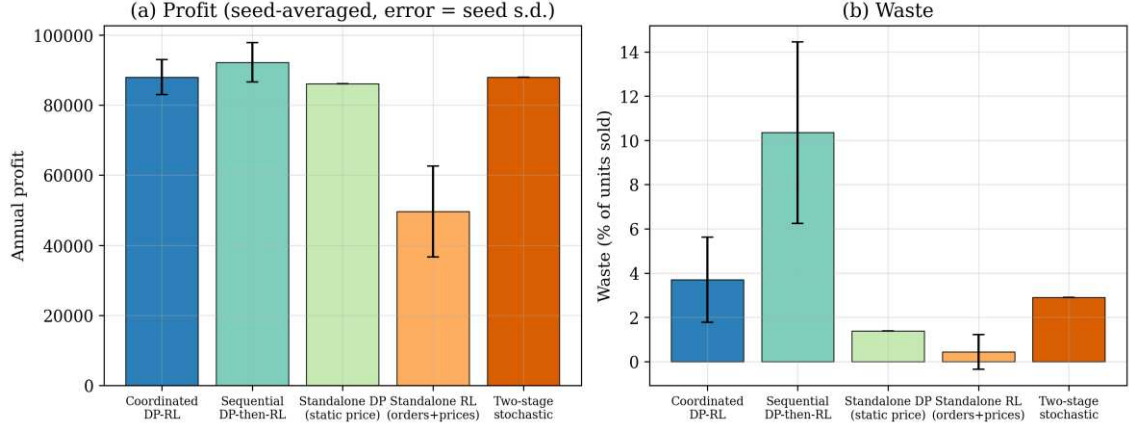


Fig. 8. Seed-averaged profit (error bars: seed standard deviation) and waste for the five methods in the binding regime.

Two findings are robust across seeds. First, the standalone reinforcement-learning agent is decisively the weakest of the five, earning roughly 50,000 against 86,000-92,000 for every structured method and carrying by far the largest seed variance (standard deviation near 13,000). An unstructured learner required to discover ordering and pricing simultaneously explores the joint action space poorly and does not reliably reach a good policy. This is the central empirical case for imposing dynamic-programming structure on the replenishment decision rather than learning it, and it holds on every seed.

Second, closing the replenishment-pricing feedback loop delivers its benefit in waste control rather than in profit. The coordinated fixed point lowers the order-up-to target once it internalises the RL agent learns, cutting waste from 10.3% to 3.7% on average and, more tellingly, removing the occasional waste blow-ups the sequential variant suffers (up to nearly 20% on an adverse seed). On profit, however, the paired difference between the coordinated and sequential methods is -4,227 with a standard deviation of 7,396 over eight seeds: the sign is not stable, and the gap lies within training variance, so we do not claim a profit improvement from coordination. The coordinated method also runs at a lower mean inventory (35 against 52 units), consistent with its tighter replenishment.

The structured methods cluster within roughly ten percent of one another, and several of the pairwise gaps are inside the seed variance documented in Section 5.5. This is the honest reading the estimator supports: the framework's value is in the ordering of methods it establishes, structure over unstructured learning, and feedback for waste stabilisation, rather than in a single headline profit margin, which at this precision could not be separated from training noise.

6. Managerial Insights

6.1. The freshness guarantee is a decision, not a constant

The promise a retailer makes about remaining shelf life sits on both sides of its own operation, and that is what makes it a decision rather than a marketing parameter. On the demand side it lifts online volume, since online demand scales as a_{min} raised to the power γ_{online} . On the supply side it restricts

which units may serve that demand, because an online order can be filled only from stock with at least a_{min} days remaining. Stock below the threshold is not lost, but it is demoted: to the store at full price while $a > 1$, to markdown at $X.p_{store}$ on its final day, and to disposal if it fails to clear.

The guarantee is therefore a demand lever with a supply cost, and the two do not balance by construction. In this calibration they do not balance at all. Moving from a one-day to a two-day promise buys 10% more online volume and costs 3.2% of profit, while tripling waste. The intuition that a stronger freshness promise is straightforwardly a competitive good does not survive contact with the inventory arithmetic.

6.2. Freshness guarantees and waste move together

A retailer under pressure to reduce food waste might expect a stronger freshness promise to help, on the reasoning that it forces faster turnover. The model says the reverse. Waste rises monotonically with the guarantee, from 2.6% at $a_{min} = 1$ to 21.6% at $a_{min} = 4$, because every unit excluded from the online channel is pushed toward markdown and disposal. A freshness guarantee is a promise about what the customer receives; it says nothing about what the retailer discards, and tightening it moves discards the wrong way. Where waste reduction is the objective, the effective levers are the replenishment cycle and the depth of markdown, not the promise.

6.3. The replenishment cycle is the binding constraint

The comparison in Fig. 2 and Table 3 carries the most general lesson in this study. Under daily delivery against a seven-day life, no pricing or replenishment policy has anything to work with: no unit ages, and waste is 0.04% whatever the optimiser does. Under a three-day cycle against a four-day life, the same policy space suddenly matters. The prior question, before a retailer invests in adaptive pricing for perishables, is whether its replenishment cadence and shelf life leave any perishability to manage. Where they do not, sophisticated machinery earns nothing; where they do, it earns a great deal. This ordering (diagnose the regime, then choose the policy) reverses how such projects are usually scoped.

6.4. Boundary conditions

Four limits bound the claims above. The guarantee is optimised as a single horizon-wide commitment, whereas a retailer able to vary it by season or product line faces a larger problem. The results are conditional on γ_{online} , the elasticity of online demand with respect to the promise, which is the parameter they are most sensitive to and the one least likely to be known in practice; a retailer whose customers value the guarantee far more strongly than assumed here would face a different optimum. The estimates carry the training variance documented in Section 5.5. And the instance is calibrated to an operating point rather than estimated from transaction data, so the magnitudes should be read as the structure of a trade-off and not as a forecast.

6.5. Reproducibility statement

The complete implementation, comprising the DP recursion, the Q-learning agent, the outer loop over the freshness guarantee, the seed-averaging protocol and the common-random-numbers scheme, is provided as supplementary material and will be deposited in a public repository on acceptance. All parameters are stated in Section 5.1. No proprietary data are required to reproduce any figure or table in this paper.

7. Conclusion

This study set out to coordinate replenishment and dynamic pricing for perishable products in omnichannel retail through a hybrid framework that decouples the two decisions by timescale, assigning anticipatory lot-sizing to DP and adaptive channel pricing to RL. In the course of specifying and testing that framework we were led to a prior question that the perishable-inventory literature has largely left implicit, and the answer to it reorganises what the rest of the paper can claim.

The principal contribution is that finding. Whether perishability binds at all is governed by the ratio of the replenishment cycle to the usable shelf life, not by the design of the policy. Under daily delivery against a seven-day life, inventory turns over in under a day, no unit ever approaches expiry, and the system reports waste of 0.04% regardless of how it is controlled. Under the three-day cycle the retailer actually operates against a four-day life, waste rises to 6.58% and the same policy space begins to matter. This is a scope condition rather than a parameter choice. It identifies when the machinery in this paper, and in substantial adjacent literature, is capable of earning anything, and it implies that near-zero waste reported under short replenishment cycles is a property of the calendar being credited to the algorithm.

The second contribution is to treat the freshness guarantee offered to online customers as a decision variable rather than a fixed parameter. Doing so exposes a lever that sits on both sides of the retailer’s own operation: it lifts online demand while restricting the stock eligible to serve it. In this calibration the supply cost dominates. A two-day promise buys 10% more online volume, triples waste and reduces profit, and the optimum is the operational minimum. The wider point is structural rather than numerical. Profit and waste worsen together as the promise tightens, so the guarantee is not a frontier the retailer negotiates but a cost it should minimise unless customers value freshness far more strongly than assumed here.

Theoretically, these results argue for a change in how models of perishable systems are specified and reported. The replenishment cadence and the usable shelf life are not incidental parameters to be fixed at convenient values; jointly they determine whether the phenomenon under study occurs in the model at all, yet neither is routinely reported. A model can be internally correct, solved to optimality and validated against real operating characteristics while describing a regime in which its subject matter is inert. The remedy is cheap: report the cycle-to-shelf-life ratio and the resulting age distribution, as in Fig. 2, before reporting any waste result. The study also contributes to methodological caution. Training

a tabular agent on an identical instance across seeds produced profit differences large enough to reverse the ranking of adjacent policies, and single-run reporting remains common in this literature. Seed averaging with paired evaluation is not a refinement here; it is the difference between a result and its opposite.

For practice, four implications follow, and they are ordered by the size of the effect rather than by the interest of the method. First, sourcing dominates. The sensitivity analysis puts procurement cost an order of magnitude ahead of every other parameter, so a retailer should exhaust its supply terms before investing in adaptive pricing. Second, the delivery schedule is the real decision. It determines whether a perishability problem exists, and no pricing policy will substitute for it. Third, waste is cheap to cut. Raising the internal shadow price of disposal tenfold removes more than half the waste for a few percent of profit, which makes an environmental target achievable at a cost most retailers would accept, and this separability is invisible without varying the parameter. Fourth, service targets set without reference to shelf life are expensive twice over, in margin and in disposal, because the stock that covers demand which does not arrive becomes the waste.

The study has clear limits. Validation rests on a transparently calibrated simulation rather than estimated demand, so the magnitudes describe the structure of a trade-off and not a forecast; the intercepts were fitted to volume alone, and mean inventory and waste emerged as outcomes, but that is corroboration rather than estimation. The results are conditional on the freshness elasticity, the parameter they are most sensitive to and the one least likely to be known in practice. The guarantee is optimised as a single horizon-wide commitment. The demand specification is separable across channels, so the model represents three demand streams competing for shared inventory rather than consumers substituting between channels on price and freshness, which is a weaker structure than the omnichannel framing suggests. And the estimator carries the training variance documented in Section 5.5.

Those limits set the agenda. The most valuable next step is to replace the separable demand system with a discrete-choice specification in which channel shares respond to relative price and freshness, which would make substitution endogenous and allow the elasticity to be estimated rather than assumed. A second is to locate the elasticity threshold at which a stronger guarantee begins to pay, which requires the full guarantee grid to be re-run at each elasticity under seed averaging. We benchmark the framework within the binding regime against a one-pass sequential variant, a standalone DP with static pricing, a standalone reinforcement-learning agent, and a two-stage stochastic program (Section 5.6). Two findings are robust across training seeds: a standalone learner that must discover ordering and pricing jointly is decisively the weakest of the five, which is the central case for imposing dynamic-programming structure; and closing the feedback loop between replenishment and pricing yields its demonstrable benefit in waste control rather than in profit, roughly halving waste relative to the sequential variant and removing its occasional blow-ups, while the profit difference between the two lies within training variance. Extending the comparison to a deep reinforcement-learning agent remains worthwhile. Beyond these, the framework extends naturally to multiple products with

substitution, to guarantees that vary by season or product line, and to competition between retailers whose freshness promises are strategic rather than parametric. Each of these is worth doing. None of them, on the evidence here, will matter as much to a retailer as the delivery schedule and the price it pays for the fruit.

Conflict of Interest

The authors declare that they have no conflict of interest.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

- Amorim, Pedro, Bernardo Almada-Lobo, Pedro Campelo, and Fábio Neves-Moreira. 2019. "Consistent Vehicle Routing Problem with Service Level Agreements: A Case Study in the Pharmaceutical Distribution Sector." *Article in European Journal of Operational Research*, ahead of print, 2019. <https://doi.org/10.1016/j.ejor.2018.07.030>.
- Anggraini, Violita, Vincent F. Yu, and Budhi Sholeh Wibowo. 2025. "Multi-Objective Optimization in Sustainable Supply Chains: A Review of Methods and Applications for Perishable Products." *Proceedings of the International Conference on Industrial Engineering and Operations Management* (Michigan, USA), February 18, 2025. <https://doi.org/10.46254/an15.20250458>.
- Baghalian, Atefeh, Shabnam Rezapour, and Reza Zanjirani Farahani. 2013. "Robust Supply Chain Network Design with Service Level against Disruptions and Demand Uncertainties: A Real-Life Case." *European Journal of Operational Research* 227 (1): 199–215. <https://doi.org/10.1016/j.ejor.2012.12.017>.
- Bellman, Richard. 1966. "Dynamic Programming." *Science* 153 (3731): 34–37. <https://doi.org/10.1126/science.153.3731.34>.
- Blattberg, RC, and SA Neslin. 1990. *Sales Promotion: Concepts, Methods and Strategies, 1990*. In *Englewood Cliffs, NJ: Prentice Hall*. 1990. https://www.researchgate.net/profile/Richard_Briesch/publication/256241591_Sales_Promotions/links/0046352a116791d6ae000000/Sales-Promotions.pdf.
- Boute, Robert N., Joren Gijsbrechts, Willem van Jaarsveld, and Nathalie Vanvuchelen. 2022. "Deep Reinforcement Learning for Inventory Control: A Roadmap." In *European Journal of Operational Research*, 298:401–412. no. 2. North-Holland, April 2022. <https://doi.org/10.1016/j.ejor.2021.07.016>.
- Bu, Jinzhi, Xiting Gong, and Xiuli Chao. 2023. "Asymptotic Optimality of Base-Stock Policies for Perishable Inventory Systems." *Management Science* 69 (2): 846–864. <https://doi.org/10.1287/mnsc.2022.4400>.
- Buisman, M. E., R. Haijema, and J. M. Bloemhof-Ruwaard. 2019. "Discounting and Dynamic Shelf Life to Reduce Fresh Food Waste at Retailers." *International Journal of Production Economics* 209 (March 2019): 274–284. <https://doi.org/10.1016/j.ijpe.2017.07.016>.
- Cachon, Gérard P., and Robert Swinney. 2009. "Purchasing, Pricing, and Quick Response in the Presence of Strategic Consumers." *Management Science* 55 (3): 497–511. <https://doi.org/10.1287/mnsc.1080.0948>.
- Dehaybe, Henri, Daniele Catanzaro, and Philippe Chevalier. 2024. "Deep Reinforcement Learning for Inventory Optimization with Non-Stationary Uncertain Demand." *European Journal of Operational Research* 314 (2): 433–445. <https://doi.org/10.1016/j.ejor.2023.10.007>.
- Dey, Sushil Kumar, Kaustav Kundu, and Prasun Das. 2024. "Digital Technology Based Game-Theoretic Pricing Strategies in

- a Three-Tier Perishable Food Supply Chain.” *Annals of Operations Research*, May 7, 2024, 1–27. <https://doi.org/10.1007/s10479-024-06021-5>.
- Donselaar, Karel van, and Rob Broekmeulen. 2025. “Easy-to-Use Estimators for Waste, on Shelf Availability and Number of Orders in a Periodic Review Inventory System with Perishable Items.” *International Journal of Production Economics* 284 (June 2025): 109608. <https://doi.org/10.1016/j.ijpe.2025.109608>.
- Feng, Lin, Wan Chih Wang, Jinn Tsair Teng, and Leopoldo Eduardo Cárdenas-Barrón. 2022. “Pricing and Lot-Sizing Decision for Fresh Goods When Demand Depends on Unit Price, Displaying Stocks and Product Age under Generalized Payments.” *European Journal of Operational Research* 296 (3): 940–952. <https://doi.org/10.1016/j.ejor.2021.04.023>.
- Ganesan, Viswanath Kumar, David Sundararaj, and Ananda Padmanaba Srinivas. 2021. “Adaptive Inventory Replenishment for Dynamic Supply Chains with Uncertain Market Demand.” In *Lecture Notes in Mechanical Engineering*, 325–335. Springer Science and Business Media Deutschland GmbH, 2021. https://doi.org/10.1007/978-981-15-5689-0_28.
- Gijsbrechts, Joren, Robert N. Boute, Jan A. Van Mieghem, and Dennis J. Zhang. 2022. “Can Deep Reinforcement Learning Improve Inventory Management? Performance on Lost Sales, Dual-Sourcing, and Multi-Echelon Problems.” *Manufacturing and Service Operations Management* 24 (3): 1349–1368. <https://doi.org/10.1287/msom.2021.1064>.
- Greenleaf, Eric A. 1995. “The Impact of Reference Price Effects on the Profitability of Price Promotions.” *Marketing Science* 14 (1): 82–104. <https://doi.org/10.1287/mksc.14.1.82>.
- Hendalianpour, Ayad. 2020. “Optimal Lot-Size and Price of Perishable Goods: A Novel Game-Theoretic Model Using Double Interval Grey Numbers.” *Computers & Industrial Engineering*, September 2, 2020, 106780. <https://doi.org/10.1016/j.cie.2020.106780>.
- Hendalianpour, Ayad, Mahnaz Fakhrabadi, Mohamad Sadegh Sangari, and Jafar Razmi. 2020. “A Combined Benders Decomposition and Lagrangian Relaxation Algorithm for Optimizing a Multi-Product, Multi-Level Omni-Channel Distribution System.” *Scientia Iranica*, ahead of print, 2020. <https://doi.org/10.24200/sci.2020.53644.3349>.
- Kavoosi, Alireza, Reza Tavakkoli-Moghaddam, Hedieh Sajedi, Nazanin Tajik, and Keivan Tafakkori. 2025. “Dynamic Pricing and Inventory Control of Perishable Products by a Deep Reinforcement Learning Algorithm.” *Expert Systems with Applications* 291 (October 2025): 128570. <https://doi.org/10.1016/j.eswa.2025.128570>.
- Keskin, N. Bora, Yuexing Li, and Jing-Sheng Song. 2022. “Data-Driven Dynamic Pricing and Ordering with Perishable Inventory in a Changing Environment.” *Management Science* 68 (3): 1938–1958. <https://doi.org/10.1287/mnsc.2021.4011>.
- Kopalle, Praveen K., and Joan Lindsey-Mullikin. 2003. “The Impact of External Reference Price on Consumer Price Expectations.” *Journal of Retailing* 79 (4): 225–236. <https://doi.org/10.1016/j.jretai.2003.09.002>.
- Kopalle, Praveen K., Koen Pauwels, Laxminarayana Yashaswy Akella, and Manish Gangwar. 2023. “Dynamic Pricing: Definition, Implications for Managers, and Future Research Directions.” *Journal of Retailing* 99 (4): 580–593. <https://doi.org/10.1016/j.jretai.2023.11.003>.
- Kopalle, Praveen K., and Russell S. Winer. 1996. “A Dynamic Model of Reference Price and Expected Quality.” *Marketing Letters* 7 (1): 41–52. <https://doi.org/10.1007/BF00557310>.
- Lagin, Madelen, Johan Håkansson, Carin Nordström, Roger G. Nyberg, and Christina Öberg. 2022. “Last-Mile Logistics of Perishable Products: A Review of Effectiveness and Efficiency Measures Used in Empirical Research.” *International Journal of Retail and Distribution Management* 50 (13): 116–139. <https://doi.org/10.1108/IJRDM-02-2021-0080>.
- Lee, Junhyeok, Youngchul Shin, and Ilkyeong Moon. 2024. “A Hybrid Deep Reinforcement Learning Approach for a Proactive Transshipment of Fresh Food in the Online–Offline Channel System.” *Transportation Research Part E: Logistics and Transportation Review* 187 (July 2024): 103576. <https://doi.org/10.1016/j.tre.2024.103576>.
- Lin, Xi, Luyi Gui, and Yixin Lu. 2024. “Managing Sales via Livestream Commerce: Implications of Price Negotiation and Consumer Price Search.” *Production and Operations Management*, ahead of print, February 8, 2024. <https://doi.org/10.1177/10591478231224930>.
- Liu, Minjian, Shaofu Du, Tengfei Nie, and Yangguang Zhu. 2024. “Pricing, Channel Design, and Platform Operations with

- Buy-Online-and-Pick-up-in-Store in Online Food Delivery Market.” *International Journal of Production Economics* 274 (August 2024): 109293. <https://doi.org/10.1016/j.ijpe.2024.109293>.
- Liu, Peide, Ayad Hendalianpour, and Mohammad Hamzehlou. 2021. “Pricing Model of Two-Echelon Supply Chain for Substitutable Products Based on Double-Interval Grey-Numbers.” *Journal of Intelligent & Fuzzy Systems* Preprint (Preprint): 1–23. <https://doi.org/10.3233/JIFS-201206>.
- Mamoudan, Mobina Mousapour, Zahra Mohammadnazari, Ali Ostadi, and Ali Esfahbodi. 2024. “Food Products Pricing Theory with Application of Machine Learning and Game Theory Approach.” *International Journal of Production Research* 62 (15): 5489–5509. <https://doi.org/10.1080/00207543.2022.2128921>.
- Mishra, Manish, Rohit Kushwaha, and Nimit Gupta. 2024. “Impact of Sales Promotion on Consumer Buying Behavior in the Apparel Industry.” *Cogent Business and Management* 11 (1): 2310552. <https://doi.org/10.1080/23311975.2024.2310552>.
- Moor, Bram J. De, Joren Gijsbrechts, and Robert N. Boute. 2022. “Reward Shaping to Improve the Performance of Deep Reinforcement Learning in Perishable Inventory Management.” *European Journal of Operational Research* 301 (2): 535–545. <https://doi.org/10.1016/j.ejor.2021.10.045>.
- Özbilge, Armağan, Elkafi Hassini, and Mahmut Parlar. 2024. “Optimal Pricing and Donation Policy for Fresh Goods.” *European Journal of Operational Research* 312 (1): 198–210. <https://doi.org/10.1016/j.ejor.2023.06.020>.
- Pilati, Francesco, Marco Giacomelli, and Matteo Brunelli. 2024. “Environmentally Sustainable Inventory Control for Perishable Products: A Bi-Objective Reorder-Level Policy.” *International Journal of Production Economics* 274 (August 2024): 109309. <https://doi.org/10.1016/j.ijpe.2024.109309>.
- Porteus, Evan. 2002. *Foundations of Stochastic Inventory Theory*. In *Foundations of Stochastic Inventory Theory*. 2002. <https://doi.org/10.1515/9781503619883>.
- Porteus, Evan L. 1990. “Stochastic Inventory Theory.” *Stochastic Models* 2 (1990): 605–652. <https://www.sciencedirect.com/science/article/pii/S0927050705801768>.
- Solari, Federico, Natalya Lysova, Andrea Volpi, Roberto Montanari, and Eleonora Bottani. 2025. “Simulation of Reordering Policies for Perishable Products under Service Level Constraints: Sensitivity Analysis Based on Design of Experiments and Response Surface Methodology.” *International Journal of Systems Science: Operations & Logistics* 12 (1). <https://doi.org/10.1080/23302674.2025.2485201>.
- Song, Boqian, Michael Z.F. Li, and Weifen Zhuang. 2021. “Dynamic Channel Control and Pricing of a Single Perishable Product on Multiple Distribution Channels.” *European Journal of Operational Research* 288 (2): 539–551. <https://doi.org/10.1016/j.ejor.2020.06.004>.
- Syed, Tahir Abbas, Haris Aslam, Zeeshan Ahmed Bhatti, Fahad Mehmood, and Aseem Pahuja. 2024. “Dynamic Pricing for Perishable Goods: A Data-Driven Digital Transformation Approach.” *International Journal of Production Economics* 277 (November 2024): 109405. <https://doi.org/10.1016/j.ijpe.2024.109405>.
- Taudes, Alfred, and Christian Rudloff. 2012. “Integrating Inventory Control and a Price Change in the Presence of Reference Price Effects: A Two-Period Model.” *Mathematical Methods of Operations Research* 75 (1): 29–65. <https://doi.org/10.1007/s00186-011-0374-1>.
- Vogt, Thomas, Ben Lowery, Anna-Lena Sachs, and Ulrich W. Thonemann. 2025. “Inventory Control and Picking Behavior: The Roles of Sustainability Messages and Price Discounts.” *European Journal of Operational Research*, ahead of print, October 1, 2025. <https://doi.org/10.1016/j.ejor.2025.09.028>.
- Wang, Guanghui, Yumin Li, Clark X. Cheng, Ruyi Li, Cynthia Xin Ding, Yue Zhang, and Yan Liu. 2025. “Integrated Optimization of Equipment Degradation Modeling and Spare Parts Inventory for Predictive Maintenance in Power Systems.” *Sustainable Energy Technologies and Assessments* 83 (November 2025): 104626. <https://doi.org/10.1016/j.seta.2025.104626>.
- Yan, Yimo, Andy H.F. Chow, Chin Pang Ho, Yong-Hong Kuo, Qihao Wu, and Chengshuo Ying. 2022. “Reinforcement Learning for Logistics and Supply Chain Management: Methodologies, State of the Art, and Future Opportunities.”

- Transportation Research Part E: Logistics and Transportation Review* 162 (June 2022): 102712. <https://doi.org/10.1016/j.tre.2022.102712>.
- Yang, Cenyang, Yihao Feng, and Andrew Whinston. 2022. "Dynamic Pricing and Information Disclosure for Fresh Produce: An Artificial Intelligence Approach." *Production and Operations Management* 31 (1): 155–171. <https://doi.org/10.1111/poms.13525>.
- Yavuz, Tuğçe, and Onur Kaya. 2024. "Deep Reinforcement Learning Algorithms for Dynamic Pricing and Inventory Management of Perishable Products." *Applied Soft Computing* 163 (September 2024): 111864. <https://doi.org/10.1016/j.asoc.2024.111864>.
- Zhao, Xiuji, Piao Chen, and Loon Ching Tang. 2025. "Condition-Based Maintenance via Markov Decision Processes: A Review." *Frontiers of Engineering Management* 12 (2): 330–342. <https://doi.org/10.1007/s42524-024-4130-7>.
- Zhao, Yan, Yanhui Li, Qi Yao, and Xu Guan. 2023. "Dual-Channel Retailing Strategy vs. Omni-Channel Buy-Online-and-Pick-up-in-Store Behaviors with Reference Freshness Effect." *International Journal of Production Economics* 263 (September 2023): 108967. <https://doi.org/10.1016/j.ijpe.2023.108967>.
- Zhou, Haijie, Kebin Chen, and Shengbin Wang. 2023. "Two-Period Pricing and Inventory Decisions of Perishable Products with Partial Lost Sales." *European Journal of Operational Research* 310 (2): 611–626. <https://doi.org/10.1016/j.ejor.2023.03.010>.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Highlights.docx](#)