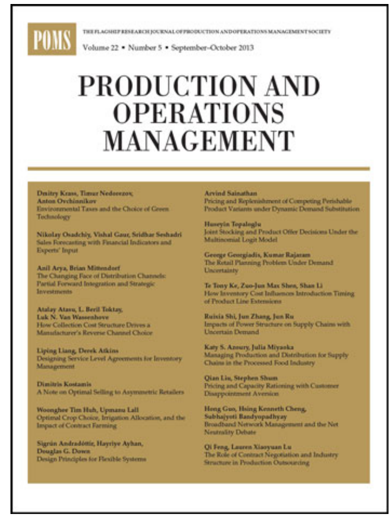
Accepted Manuscript

Title: Dynamic Pricing and Information Disclosure for Fresh Produce: An Artificial Intelligence Approach



Authors: Cenying Yang, Yihao Feng, Andrew Whinston

DOI: https://doi.org/doi:10.1111/poms.13525
Reference: POMS 13525
To appear in: Production and Operations Management

Please cite this article as: Yang Cenying.,et.al., Dynamic Pricing and Information Disclosure for Fresh Produce: An Artificial Intelligence Approach.
Production and Operations Management (2021), https://doi.org/doi:10.1111/poms.13525

Cenying Yang (corresponding author)
Department of Information Systems, College of Business, City University of Hong Kong, Hong Kong SAR, China
cenyyang@um.cityu.edu.hk
+1 512-736-8838
Personal zoom ID: 238 942 8994
Note: I am currently quarantined in a hotel in Hong Kong. You can reach out to me through my US number during my quarantine time by July 8, 2021. You can always reach out to me via zoom or email.

Yihao Feng
Department of Computer Science, College of Natural Sciences, University of Texas at Austin, Austin, Texas, US
yihao@cs.utexas.edu
+1 603-306-3311

Andrew Whinston
Department of Information Systems, Risks, and Operations Management, McCombs Business School of Business, University of Texas at Austin, Austin, Texas, US
abwhins@gmail.com
+1 512-436-8224

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/POMS.13525

Dynamic Pricing and Information Disclosure for Fresh Produce: An Artificial Intelligence Approach

Abstract

Failing to sell fresh produce before expiration not only hurts the bottom line of grocery retailers, but also leads to food waste. This work combines dynamic pricing and information disclosure to help retailers to effectively sell fresh produce and promote sustainability. We focus on a quality-based pricing strategy and whether retailers should disclose information on food quality to customers. We consider a model where a monopolistic retailer sells fresh produce to customers who have different perceptions about food quality within a given time period. We employ a deep reinforcement learning algorithm to derive the optimal pricing and information strategies. Our simulation results show that a quality-based pricing strategy yields lower prices than a pricing strategy that does not consider quality. Lower prices drive demand, thus improving profits and reducing food waste. Additionally, we show that, when an information strategy is allowed, the prices in a quality-based pricing strategy stay the same or even increase during the selling season. This is because information disclosure helps align customers' biased perceptions on food quality with the actual levels. We show that a combination of quality-based pricing and information disclosure further improves profits and reduces food waste when a large portion of customers consider quality to be lower than actual levels. To implement our ideas, we propose a cloud-based automated system that integrates sensor data, artificial intelligence, and customer communications. Our results have profound implications for the food industry on managing fresh produce.

Keywords: Dynamic pricing, information disclosure, reinforcement learning, sustainability

History: Received: March 2020; Accepted: June 2021 by Subodha Kumar, after 3 revisions.

1. Introduction

Perishable fresh produce comprises 53.8% of grocery chain revenues, totaling $267.8 billion in 2015 (Sanders 2020); yet, in the United States, 10%, or 43 billion pounds, of the food in grocery stores is thrown away every year.[1] This is equivalent to a $47 billion loss for retailers (Buzby and Hyman 2012). Thus, it is a big challenge for grocery retailers to sell fresh produce before it expires. Failing to do so not only hurts retailers' bottom line, but also leads to food waste. One key component of this challenge is to have a better understanding of food quality, which is measured in terms of remaining days,[2] *over time*. The rise of the Internet of Things (IoT) has made this possible. Sensor data on temperature and humidity and images of products can empower retailers to monitor remaining days in real time. For instance, Walmart has been leveraging IoT to determine the freshness and remaining days of food on a granular level.[3] However, given that retailers have the ability to monitor food quality (i.e., remaining days) in real time, how can they leverage this information on food quality to make better decisions to sell fresh produce more effectively?

When it comes to fresh produce, food safety is a top concern,[4] and how customers perceive products directly affects their purchase decisions. Customers can form their *subjective* perceptions about the quality of fresh produce through physical examination. As pointed out by Grunert (2005), customers' subjective perceptions about fresh produce usually diverge from the true level of quality. Empirical evidence also supports this perception bias. Previous work has found that packaging (Bou-Mitri et al. 2021), and country of origin (Elliott and Cameron 1994) can affect customers' perception of quality, thus affecting their purchase decisions. Then, if information on food quality is accessible to grocery retailers in real time, information asymmetry leaves it up to the retailer whether to disclose information on food quality when some customers could perceive it to be lower than the actual level, and others higher. Previous studies have explored information disclosure on product quality when customers are uncertain about product quality in a monopolistic market (Grossman 1981, Zhu et al. 2021), in a competitive market (Guo and Zhao 2009, Kuksov and Lin 2010), and across distribution channels (Cao et al. 2020, Guan and Chen 2017). However,

[1]Quest: Food waste statistics, the reality of food waste.
[2]Hereafter, we use the terms *food quality* and *remaining days* interchangeably.
[3]Supermarket News: Walmart introduces Eden, its high-tech fresh-food initiative.
[4]Food Dive: Food safety top concern for consumers.

they focus on one-time disclosures. In our case, during the selling season, retailers have multiple chances to disclose information. This temporal dimension adds a new trade-off where what retailers disclose today will alter customers' perceptions of food quality in the future and, thus, their purchase decisions in the future as well (Ely 2017). Therefore, *when* to disclose information during the selling season matters, since information disclosure changes customers' biased perceptions about the quality of fresh produce over time and, thus, their purchase decisions.

Pricing is another tool that can be used to influence customer demand. Dynamic pricing has long been used to address the pricing of perishable products (for reviews, see, e.g., Bitran and Caldentey 2003, Chen and Chen 2015, Shen and Su 2007). Previous works focus on perishable products whose quality does not change over time. Thus, they consider demand in a certain period that depends only on prices (Gallego and Van Ryzin 1994, Golrezaei et al. 2020). However, fresh produce is unique because it deteriorates over time. Since food quality is a major consideration in customers' purchase decisions, the demand for fresh produce depends on its deteriorating quality as well. Without taking changing quality into account, retailers might not be able to correctly tailor prices to demand. Thus, it is important to derive a quality-based pricing strategy in the case of fresh produce. Additionally, little work has addressed a combined strategy of *dynamic* pricing and information disclosure over time. When information disclosure is allowed, what would be the optimal dynamic pricing strategy?

Therefore, given that retailers are able to monitor food quality in real time, our paper describes a first attempt to explore a quality-based pricing strategy and information disclosure to help grocery retailers effectively sell fresh produce and promote sustainability.[5] We are interested in the following research question: *how can the monitoring of food quality (remaining days) help a grocery retailer better manage perishable fresh produce?* More specifically, *when should a retailer disclose information on food quality to customers? How should a retailer adjust prices over time, especially when information disclosure is allowed?* We study a problem where a monopolistic retailer sells single fresh produce to heterogeneous customers. We consider heterogeneity in two dimensions: (1) customers have different perceptions about the quality of the product over time and (2) customers

[5]We consider sustainability in terms of reducing food waste.

have different valuations of the product over time. The goal of the retailer is to determine optimal pricing and information strategies to effectively sell fresh produce in terms of achieving high profits and reducing food waste.

We formulate the problem as a Markov decision process (MDP) where the environment is the market demand. In practice, market demand is likely unknown or hard to quantify. Traditional decision making solutions, such as dynamic programming (DP), make assumptions about market demand to simplify the problem (e.g., Bitran and Mondschein 1997, Gallego and Van Ryzin 1994). In a response to better resemble practical situations, more recent works have relaxed assumptions about the demand and proposed optimal heuristics solutions when retailers have limited demand information (Cohen et al. 2018, Ke et al. 2019, Miao and Chao 2021). Despite the effort, these approaches still rely on some assumptions about the demand. Advances in artificial intelligence (AI), such as deep reinforcement learning (DRL), open another route to approach the MDP problem. Through continuous trial-and-error interactions with an unknown demand, DRL-based AI agents learn by themselves to achieve optimal strategies *without* imposing assumptions on the demand (Sutton and Barto 2018). Thus, DRL can adapt well to complex, evolving real-world environments, and be generalized to solve a diversity of problems. DRL has tremendous potential to revolutionize operations practice. Scholars have adopted DRL to solve challenging real-world problems, such as large-scale fleet management (e.g., Qin et al. 2020). Thus, these practical considerations motivate us to adopt DRL for pricing and information disclosure as well. Specifically, we employ proximal policy optimization (PPO) (Schulman et al. 2017). The retailer's goal is to maximize total discounted profits during a predefined selling window.

Our simulation results demonstrate the importance of a quality-based pricing strategy and information disclosure. A quality-based pricing strategy enables retailers to improve profits and reduce food waste compared to a pricing strategy that does not take quality into account. We find that prices generally decrease over time. This result is consistent with prior literature on dynamic pricing when customers' valuation does not increase over time (e.g., Bitran and Mondschein 1997, Gallego and Van Ryzin 1994). We also find that, when food quality is accounted for, the retailer tends to set lower prices over time, and these low prices drive demand and thus improve effectiveness.

Improvement arises because the ability to monitor the quality of fresh produce can provide the retailer *implicit* information about how demand responds to prices, given the current quality. This implicit information enables the retailer to better tailor prices to the demand, which depends on both the price and quality. This intuition is in line with operations practice where demand learning is the key for dynamic pricing (Chen and Chen 2015). Additionally, we show that when information disclosure is allowed, the retailer is able to further improve profits and reduce food waste when a large portion of customers have perceptions of low quality. The retailer has greater incentives to disclose information at an early stage when customers' perception biases are high. Under such an information disclosure strategy, customers' biased perceptions about quality gradually approach the true level. Zhu et al. (2021) seem to support our finding. They show that the seller needs to disclose information when there is large quality perception variability among customers. More interestingly, such alignment between customers' perceptions and the true quality enables the retailer to maintain prices at high levels during the selling season. Such a trade-off between dynamic pricing and information disclosure resembles that of Sanders (2020), who empirically finds that waste bans can disincentivize a store from adopting dynamic pricing, leading to reduced food waste.

Contribution. We contribute to the literature in several important ways. Our work enriches our understanding of information disclosure when customers have heterogeneous biased perceptions about product quality. A large body of literature has identified factors that affect sellers' disclosure decision (e.g., Liu et al. 2019, Zhu et al. 2021). We extend the literature by looking at multiple decisions on information disclosure over time, since we are considering deteriorating fresh produce. This temporal dimension adds a new trade-off in terms of the disclosures of today affecting customers' beliefs and thus their purchase behavior tomorrow. Additionally, we contribute to the vast literature on dynamic pricing (for reviews, see Bitran and Caldentey 2003, Chen and Chen 2015, Shen and Su 2007) as the first to incorporate changing product quality and combine dynamic pricing with dynamic information disclosure. By doing so, we also contribute to a burgeoning literature on grocery retailing and food waste (Belavina 2021, Belavina et al. 2017, Sanders 2020) by showing that a combination of dynamic pricing and information disclosure helps reduce food waste at retail stores. Finally, our work describes a response to a call by Caro et al. (2020) for more research on the impact of IoT on store operations. We extend a rapidly growing literature on the value of IoT by showcasing the value of IoT-based monitoring capabilities in managing perishable products more effectively. Methodologically, we are the first to apply DRL to dynamic pricing and information disclosure. Traditional methods require assumptions on market demand to derive optimal strategies (e.g., Bitran and Mondschein 1997, Gallego and Van Ryzin 1994). Despite recent efforts in relaxing assumptions and proposing problem-specific solutions (Cohen et al. 2018, Ke et al. 2019, Miao and Chao 2021), approaches with greater adaptability and generalizability are needed. DRL is one such approach. Our work is among the few to describe an application of DRL in operations management. ***Managerial implications***. Managerially speaking, our findings suggest that quality-based dynamic pricing is critical for the effective sale of fresh produce. It benefits the retailer, customers, and society in terms of improved profits, increased consumer surplus, and reduced food waste. Additionally, the incorporation of information disclosure enables the retailer to maintain prices at high levels during the selling season. A more stable pricing scheme could be preferred by those retailers who have concerns about dynamic pricing. The incorporation of information disclosure also leads to a higher price per unit sold, which decreases consumer surplus. Our results also suggest that a combination of quality-based dynamic pricing and information disclosure brings retailers the greatest benefits in markets where customers are very risk averse in terms of quality, customers' perception biases are high, and the demand rate is low. Lastly, to implement our ideas in practice, we propose a cloud-based automated system that integrates IoT, AI, and customer communications (see Figure 1 and Online Appendix A for a detailed discussion). This system receives sensor data and/or images of produce to estimate the remaining days of fresh produce. Given the remaining days and inventory level, the AI module employs a DRL algorithm to derive optimal pricing and information disclosure decisions. The information is conveyed to customers through a communication module. We further conduct a cost analysis to shed light on the incentives of the retailer to adopt this automated system. In this paper, we focus on the AI module and explore optimal pricing and information strategies given IoT-based monitoring capabilities.

The remainder of the paper is organized as follows. Section 2 reviews the literature. Section 3 formulates the model, followed by the methodology in Section 4. Section 5 presents the results, and

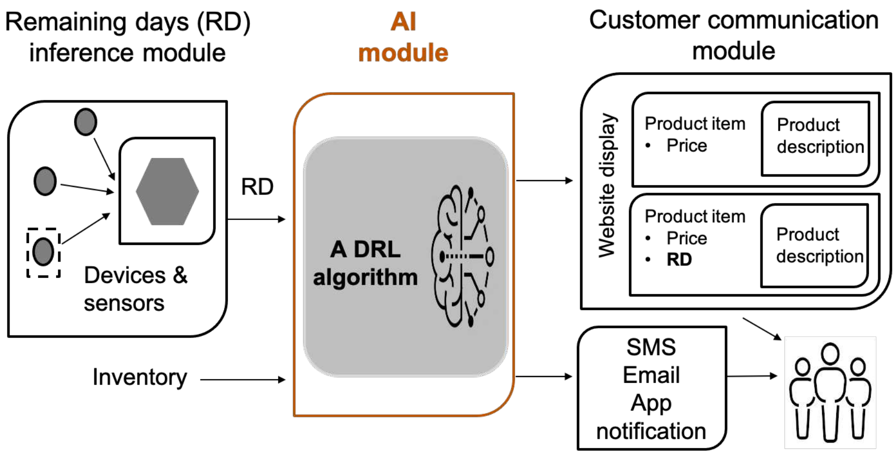


Figure 1: Design of the cloud-based automated system

Section 6 checks their robustness. We conclude the paper in the final section.

2. Literature review

2.1 Product information disclosure

Our work is closely related to the literature on product information disclosure. The central issue is the conditions under which firms would have incentives to disclose product information when there is information asymmetry between firms and customers. The general unraveling theory states that a monopolist will voluntarily disclose private information on product quality if disclosure and quality verification are costless (Grossman 1981, Milgrom 1981). Subsequent works investigate the factors that can lead to partial or no information disclosure. When information disclosure incurs a cost, the disclosure will happen only if the quality is above a threshold (Jovanovic 1982). Guo and Zhao (2009) show that firms in competitive markets reveal less information than a monopoly firm. The possible reason could be reduced expected benefits from quality disclosure or fear of free riding by competitors (Jin 2005). Kuksov and Lin (2010) show that, while a high-quality firm always has incentives to disclose information, the decision of information disclosure by a low-quality firm depends on the intensity of competition and the extent of free riding on information. The recent rise of online search/reviews offers easy access to customers to learn about the quality of products. Liu et al. (2019) allow customers to conduct research to resolve valuation uncertainties by incurring a search cost. The authors show that information disclosure depends on customers' uncertainties in their valuations of the product and the magnitude of customers' search costs. Zhu et al. (2021) take into account customers learning from peers who have used the product before. They show that firms with high- or low-quality goods prefer not disclosing quality, while a medium-quality firm might disclose its quality level. This is because customers' quality perception variability is high when the quality is medium. Information disclosure helps reduce such variability. The authors also show that disclosure strategy impacts the optimal pricing decision in a nonmonotonic way. Other works have explored optimal information disclosure strategies among different players in a distribution channel (e.g., Cao et al. 2020, Guan and Chen 2017, Sun and Tyagi 2020).

The main distinguishing element of our research compared to previous works is the incorporation of changing product quality. This incorporation leads to a problem of *dynamic* information disclosure, that is, the firm has multiple chances to decide whether to disclose information during the selling season. This temporal dimension adds a new trade-off in terms of the seller's incentive to disclose information, since what a retailer discloses today will change customers' perceptions about product quality and their purchase decisions in the future.

2.2 Dynamic pricing of perishable products

Additionally, our work relates to a vast literature on dynamic pricing. Since the seminal work by Gallego and Van Ryzin (1994), many studies have generalized the original monopoly single-product Poisson arrival demand model with myopic customers. Zhao and Zhang (2000) extend the problem to nonhomogeneous Poisson arrival demand. Li and Huh (2011) consider a problem of multiple differentiated products in a competitive market. They find that, although competition drives up the total market share and drives down prices, the market share and revenue of a particular product can increase or decrease due to competition. Hu et al. (2019) build a product return model by augmenting the classic monopolist's dynamic pricing framework. They find that ignoring returns leads to overpricing and can cause significant revenue loss. Golrezaei et al. (2020) considers strategic customers who are heterogeneous with respect to their initial valuations and the rates at which their initial valuations decrease. The authors show that delayed allocation and dynamic pricing can be effective screening tools for maximizing firm profit and improving social welfare.

See Bitran and Caldentey (2003), Chen and Chen (2015) and Shen and Su (2007) for a comprehensive review. To position our work with respect to existing work, we introduce specific demand complexity that is caused by customers' biased perceptions about the changing product quality. Fresh produce is unique in terms of deteriorating quality. Customers' perceptions about the deterioration can differ from the true case (Grunert 2005). Since the demand for fresh produce depends on both price and customers' perceived quality, it is critical to incorporate this bias inherent in customers' perceptions into the pricing problem. In line with Golrezaei et al. (2020), we model customers' heterogeneous perceptions about quality as different rates at which fresh produce deteriorates. However, we distinguish ourselves by considering a joint problem of dynamic pricing and information disclosure when customers have biased perceptions about food quality.

In terms of methodology, most previous works make assumptions on the demand, whereas, in reality, demand information is often limited. This practical consideration has spurred research in active demand learning (e.g., Miao and Chao 2021) and robustness optimization (e.g., Cohen et al. 2018). However, the methods still rely on some assumptions about the demand, leading to problem-dependent heuristics. DRL provides an opportunity to solve a diversity of problems without making assumptions about the demand. Though DRL has been proven successful in games (e.g., Silver et al. 2016), applications of DRL in operations practice remain scarce. Our work describes a first attempt to apply DRL in dynamic pricing and information disclosure.

2.3 Value of IoT

Finally, our work complements a rapidly growing stream of literature on IoT. IoT refers to the network of things, where things are connected via smart sensors to the Internet (Li et al. 2015). This real-time connection of physical and digital systems has been changing the way companies manage operations, products, and services (Caro et al. 2020, Kumar et al. 2018, Olsen and Tomlin 2020). The application of IoT in operations management can be traced back to the introduction of radio-frequency identification, which has proven powerful in tracking and monitoring inventory through the supply chain (for a review, see Dutta et al. 2007). The recent surge in ubiquitous sensors has further allowed for much richer information regarding the state of the surrounding environment, which has revived research in the food industry and sustainable operations. For instance, enabled by IoT, fertilizers and pesticides are applied only where they are needed, and when they are needed. Sensors can measure moisture, nitrogen, and other chemical contents in the soil and monitor animal and insect conditions to provide guidelines for farmers (Dong 2021).

At final retail stores, research on the value of IoT on store operations remains scarce. Caro et al. (2020) call for further work to show the impact of IoT data on store operations. Atasu et al. (2020) also point out that supply chain traceability, large-scale sensor-based measurements, and the availability of other data will allow operations management researchers to explore novel research questions about environmental and social sustainability. Our work describes a response to both calls in the context of grocery retailing. We investigate how businesses can leverage IoT-based monitoring capabilities on food quality to sell products more effectively and thus promote sustainability through two operations weapons: dynamic pricing and information disclosure.

3. Model

Consider a grocery retailer who needs to sell fresh produce to heterogeneous customers within time T. The retailer tries to adjust the price and information strategies to effectively sell products. At the beginning of the planning horizon, the retailer has C units in inventory and replenishment is not allowed. The unit cost of inventory is q. We assume that food quality deteriorates according to an exponential function (Bai and Kendall 2008) $\theta_t = Qe^{-\gamma t} + \epsilon_t$, where γ represents the true deterioration rate and ϵ_t is noise that is normally distributed according to $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. We check the robustness of our results by using a linear form of deteriorating quality.

Customers are aware of the exponential nature of quality deterioration, but they have different perceptions of the deterioration rate, denoted as $\hat{\gamma}$. Thus customers have different perceptions of the quality of fresh produce $\hat{\theta}_t = Qe^{-\hat{\gamma} t} + \epsilon_t$. We consider three groups of customers: (1) customers who perceive food as deteriorating more quickly than it actually does and thus perceive its quality as low, (2) customers who perceive food as deteriorating at the same rate as it actually does and thus perceive its quality as being the same as its true state, and (3) customers who perceive food as deteriorating more slowly than it actually does and thus perceive its quality as being high. We consider three cases in terms of the portion of customers with low perceived quality (see Table 1). In case 1, we consider the same portions of customers with low, the same, and high perceived quality. In case 2, customers with low perceived quality comprise 50%, and the other two groups

Table 1: Case settings

	Case 1	Case 2	Case 3
Customers with low perceived quality	33.3%	50%	75%
Customers with same perceived quality	33.3%	25%	12.5%
Customers with high perceived quality	33.3%	25%	12.5%

25% each. Finally, in case 3, customers with low perceived quality comprise 75%, and the other two groups 12.5% each. We are especially interested in the case with a large portion of customers with low perceived quality. Because food safety is a big concern, risk-averse customers are likely to perceive the quality as lower than it actually is, especially for produce that previously suffered a high-profile food accident.[6] In addition, customers have cosmetic standards for fresh produce. They consider ugly food, such as blemished tomatoes, low quality and refrain from buying them. Cases 2 and 3 will therefore arise for some types of fresh produce, such as produce that has been involved in a high-profile accident and ugly produce. We check the robustness of our results by considering customers with a high perception of quality as being in the majority.

Additionally, customers are heterogeneous in terms of their reservation price and reservation quality. Thus, customers hold different valuations of the product over time. For instance, a customer with high income is more likely to have a high reservation price and high reservation quality, caring more about quality than price. A customer with low income is more likely to have a low reservation price and low reservation quality, caring more about price than quality. We incorporate this heterogeneity by assuming Weibull distributions $F(\cdot)$ and $M(\cdot)$ for customers' reservation price and reservation quality, respectively. We assume that customers arrive according to a Poisson distribution and are myopic. They make purchases immediately if the listed price is below their reservation price and their perceived quality is above the reservation quality. The probability that a customer will make a purchase is given by $(1 - F(p_t))M(\hat{\theta}_t)$.

We formulate the problem as an MDP that is characterized by $\{S, A, R, T\}$, where S is the state space, A is the action space, R is the reward signal, and T is the transition probability that determines the next state, given the current states and actions, denoted by $T(s'|s, a)$. We consider

[6] For instance, an outbreak of salmonella Newport infections that is linked to red onions has spread to 47 states since July, 2020.

two environment states: the inventory level and the food quality (i.e., remaining days). At time t, the retailer observes the true state of the food quality, θ_t, and the inventory level, c_t. The retailer then has two actions to take: setting the price p_t and determining whether to *truthfully* disclose information on food quality, denoted as m_t. The evolution of the inventory level and remaining days, that is, $T(s'|s, a)$, is governed by the demand in each period. We assume the demand is unknown to the retailer or is hard to quantify, which is likely to be the case in practice. However, the retailer can influence the demand with a pricing strategy and an information strategy. Finally, we consider the total discounted profit during the selling season T as the reward signal in our model. The discount rate is denoted as α. We consider three models to evaluate the effectiveness of incorporating quality into a pricing strategy and incorporating information disclosure.

3.1 Benchmark model: Pricing without quality

Extensive work has derived an optimal pricing strategy for perishable products when product quality is not taken into account (e.g., Bitran and Mondschein 1997, Gallego and Van Ryzin 1994). Since fresh produce deteriorates over time and food quality is a major factor in customers' purchase decisions, it is important to incorporate food quality into the pricing strategy. To demonstrate the importance of such incorporation, we use the model that does not consider food quality as the
一个状态 库存水平，一个动作 定价
benchmark, where we have one state, the inventory level, and one action, pricing. At time t, a random number of customers arrive at the store. They independently decide whether to buy one unit of the product, based on their reservation price and reservation quality. If the inventory, c_t, is greater than the number of purchases, n_t, the inventory is reduced by n_t; otherwise, the inventory becomes zero and $n_t - c_t$ of the demand is lost. At the end of the selling season, any leftover inventory is thrown away, with no salvage value.

3.2 Model 1: Quality-based pricing

两个状态：库存和食品质量 一个动作：价格

Model 1 grants the retailer access to information on food quality. We have two states, the inventory level and food quality, and one action, pricing. The comparison between the benchmark model and Model 1 sheds light on the importance of accounting for food quality when deriving an optimal pricing strategy.

3.3 Model 2: Quality-based pricing with information disclosure

两个状态：库存和食品质量，一个动作：价格

Table 2: Model settings

	Benchmark	Model 1	Model 2
State	c_t	c_t, θ_t	c_t, θ_t
Action	p_t	p_t	p_t, m_t
Reward	$\sum_t \alpha^t \min(n_t, c_t)p_t$ $-Cq$	$\sum_t \alpha^t \min(n_t, c_t)p_t$ $-Cq$	$\sum_t \alpha^t \min(n_t, c_t)p_t$ $-Cq$

When an information strategy is incorporated, we have two states, inventory level and food quality, and two actions. One action is pricing, and the other is information disclosure. Food quality deteriorates over time. The retailer observes the true quality θ_t at time t and decides whether to truthfully disclose the quality to customers. We consider m_t as a continuous variable that ranges between zero and one, that is, $m_t \in [0, 1]$. We thus obtain the probability that a retailer will disclose the information on food quality to customers at time t.

Customers hold different perceptions of how fresh produce deteriorates over time. We consider three groups of customers: (1) those who perceive the quality as low perceive a high deterioration rate, denoted as $\hat{\gamma}_1$; (2) those who perceive the quality as the same as the true state perceive a true deterioration rate, denoted as $\hat{\gamma}_2$, where $\hat{\gamma}_2 = \gamma$; and (3) those who perceive quality as high perceive a low deterioration rate, denoted as $\hat{\gamma}_3$. We assume that customers are aware of the retailer's truthful revelations. Thus, having received information disclosed by the retailer, customers observe the true state of the food quality θ_t and update their perceptions about the deterioration rate according to Bayes rule:

$$p(\hat{\gamma}_i|\theta_t) = \frac{p(\theta_t|\hat{\gamma}_i)p(\hat{\gamma}_i)}{p(\theta_t)}, \tag{1}$$

where i represents the group of customers. Once customers update their belief on the deterioration rate $\hat{\gamma}_i$, they will obtain their perceived quality of fresh produce $\hat{\theta}_t$ based on the exponential function, taking noise into account. We assume that customers' perceived deterioration rate $\hat{\gamma}_i$ has a prior distribution $p_0(\hat{\gamma}_i) \sim N(\mu_{i0}, \sigma_{i0}^2)$. The settings of our models are summarized in Table 2.

4. Methodology: A DRL algorithm

To solve the MDP problem, we employ DRL for two major reasons. First, DRL can be adapted to real-world complex, evolving demand by learning directly from the data, without making assumptions about the demand. Second, DRL can handle continuous and high-dimensional product features and action spaces by leveraging neural networks to approximate complex dynamics. In our case, the food quality (i.e., remaining days), pricing, and information disclosure are continuous. Thus, DP, a method that is widely used in previous literature (e.g., Bitran and Mondschein 1997, Gallego and Van Ryzin 1994), is not appropriate in our application. Specifically, we adopt proximal policy optimization (PPO) as our AI agent to make pricing and information disclosure decisions over time. The AI agent observes the current inventory level and remaining days (i.e., food quality) and decides what price to set and whether to disclose information on remaining days. Then, the agent observes the feedback (i.e., whether profits improve or decrease) to evaluate the actions. The agent will learn from millions of such interactions with demand, to continuously improve pricing and information disclosure strategies. The retailer's goal is to obtain optimal strategies to maximize the total discount profit during the selling season.

Formally, our AI agent starts with a stochastic policy π that is parameterized with ω, denoted as π_ω. A stochastic policy is defined as the probability distribution of actions, given current states, that is, $\pi_\omega(a_t = a|s_t = s)$. In our case, the policy gives the probability distribution of pricing and information disclosure, given the current inventory level and remaining days (i.e., food quality). Following the policy π_ω, the AI agent observes a sequence of states, actions, and rewards, called a trajectory, denoted by $\tau = (s_0, a_0, s_1, r_1, a_1 \ldots)$. A trajectory is jointly determined by the policy π_ω and the transition probability of the environment $T(s_{t+1}|s_t, a_t)$. The agent's goal is to learn an optimal policy that maximizes the discounted total profit during the selling season T:

$$J(\omega) = \mathbb{E}_{\rho_\omega(\tau)}[G_t], \tag{2}$$

where G_t is the discounted *future* reward, starting from the current period t. Finding an optimal policy that maximizes the discounted *total* reward is equivalent to finding an optimal policy that maximizes the discounted *future* reward at each time point. The expectation is taken with respect to the distribution of trajectories ρ, which is jointly induced by policy and environment dynamics, that is, $\rho_\omega(\tau) = d(s_0)\pi(a_0|s_0)T(s_1|s_0, a_0)\cdots$, where $d(s_0)$ is the distribution of the initial state s_0.

We employ gradient descent and, in each iteration, update the parameters by

$$\omega_{t+1} = \omega_t + \beta_{step} \nabla J(\omega_t), \tag{3}$$

where β_{step} is the step size. According to the policy gradient theorem (Sutton and Barto 2018), the gradient of $J(\omega)$ can be written as

$$\nabla_\omega J(\omega) = \mathbb{E}_\pi [G_t \sum_t \nabla_\omega \log \pi(a_t|s_t)]. \tag{4}$$

The derivative of the objective function $J(\omega)$ no longer requires knowledge of the distribution of states or the demand dynamics $T(s_{t+1}|s_t, a_t)$. This algorithm is also known as a REINFORCE algorithm (Williams 1992). It describes one way in which DRL can learn an optimal policy without knowledge of or assumptions on the underlying demand. However, REINFORCE algorithms suffer from several problems. First, due to the expectation in equation (4), it is usually not possible to compute the gradient of the objective function $J(\omega)$ exactly. The REINFORCE algorithm employs stochastic gradient estimators to approximate the gradient of the expected return based on a batch of sampled trajectories. However, sampling leads to high gradient estimation variance, making the training unstable and slowing down the convergence. Additionally, REINFORCE algorithms require a large number of samples to perform a gradient update, making the training unstable as well. Therefore, we employ PPO to alleviate those issues (Schulman et al. 2017), as follows:[7]

- **Value function as a baseline.** A widely used approach to reducing the gradient estimation variance while keeping unbiased estimation is to subtract a baseline value from the return G_t. PPO uses the state value function $V(s)$ as such a baseline. A state value function represents the future discounted reward when policy π starts with state s. We denote $A_t = G_t - V(s_t)$ as the advantage function and can rewrite equation (2) as $J(\omega) = \mathbb{E}_{\rho_\omega(\tau)}[A_t]$.

- **Trust region.** PPO introduces trust regions in the objective function (Schulman et al. 2015). A trust region constrains the extent to which a policy can be updated in each step. This

[7] Even with those techniques, we do acknowledge that in practice, not all companies can generate a large number of samples to implement RL. Currently, only tech giants, like Amazon, have the ability to implement RL in practice.

approach avoids large changes in parameters and makes training process more stable and less sensitive to hyperparameter tuning.

- **Importance sampling and clipped function.** Importance sampling is widely used to update policy by sampling data from old policy. This method improves sampling efficiency, since sampling from the new policy every time is computationally expensive. The ratio of importance sampling, $\frac{\pi_\omega(a_t|s_t)}{\pi_{\omega_{\text{old}}}(a_t|s_t)}$, measures how different the policy is from the old one. A larger importance ratio will make the training process unstable. Thus, to avoid high gradient estimate variance, PPO further clips the importance sampling ratio within a proper range.

Two issues remain when implementing PPO. First, which optimal policy π should the AI agent learn? Second, how should the AI agent learn the value function $V(s)$? We address those issues by leveraging deep neural networks to approximate both the policy and value functions. Deep neural networks have shown power in learning complicated nonlinear relations (e.g., Silver et al. 2016). Many real-world problems, such as ours, involve nonlinear rather than linear relations between the input and output. Thus, when implementing PPO, we employ two feedforward neural networks to learn the policy function and the value function, respectively, as follows:

- **Policy architecture.** We assume a Gaussian distribution for the policy. The mean of the distribution is approximated through a two-layer-deep feedforward neural network to learn the policy, that is, a mapping from states to a distribution of actions (see Figure 2). Going from one layer to another, a set of units compute the weighted sum of their inputs from the previous layers and pass the result to the next layers through a *nonlinear function*. The final output is the mean of a multivariate normal distribution whose dimension is the same as that of the action.

- **Value function (advantage function).** We parameterize the value function with ϕ. We implement a two-layer–deep feedforward neural network to learn the value function V_ϕ (i.e., a mapping from states to a value). The architecture is the same as for the policy, except that the output is a numeric scalar value. To calculate the advantage value, we further obtain the future discounted total reward G_t through Monte Carlo sampling.

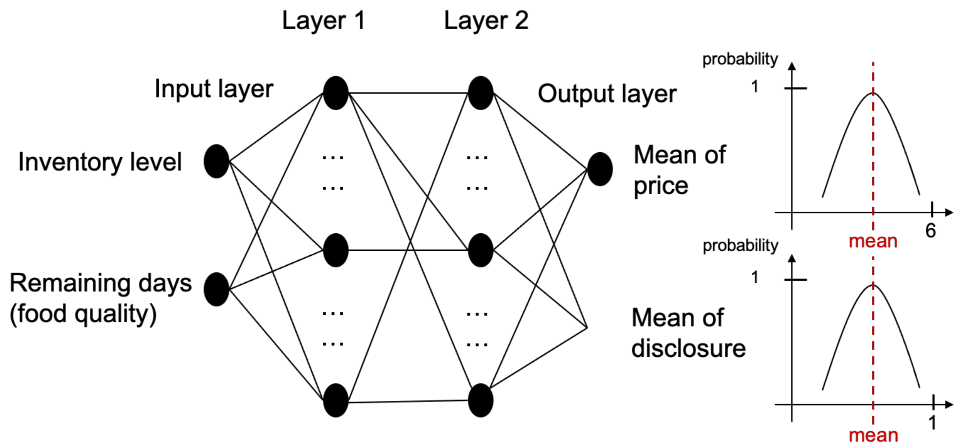


Figure 2: Policy architecture: neural network

The framework of our PPO implementation is illustrated in Figure 3. Given the current state s_t, the AI agent decides the action a_t to take according to neural network policy π_ω. After the AI agent takes those actions, demand will be realized and the agent will observe a profit and the next state s_{t+1}. Finally, the AI agent will update the policy in the direction consistent with the advantage function, which is approximated by another neural network. We train the parameters of the neural networks to maximize the expected future discounted reward by using backpropagation.[8] We then employ stochastic gradient descent to update the policy gradient. The gradient descent updates the gradient based on the full data set, whereas stochastic gradient descent performs updates based on a small batch of the original data. Thus, stochastic gradient descent is much faster to train and more computationally efficient. Online Appendix B provides a more detailed description of how we implement PPO in our case.

5. Simulations and results

We limit the selling period to $T = 12$ days (approximately two weeks), and the price ranges from zero to six.[9] We start with $C = 500$ units of inventory, and the unit cost of inventory is $q = 3$. Customers arrive according to a Poisson distribution, with an average intensity rate of 70. We set the true deterioration rate γ at 0.1. We consider three groups of customers: (1) for customers who

[8] Backpropagation is a procedure for computing the gradient of a neural network parameter with respect to an objective function. It is a practical application of the chain rule for derivatives.

[9] According to the US Department of Agriculture, the retail prices of most fresh produce fall in this range.

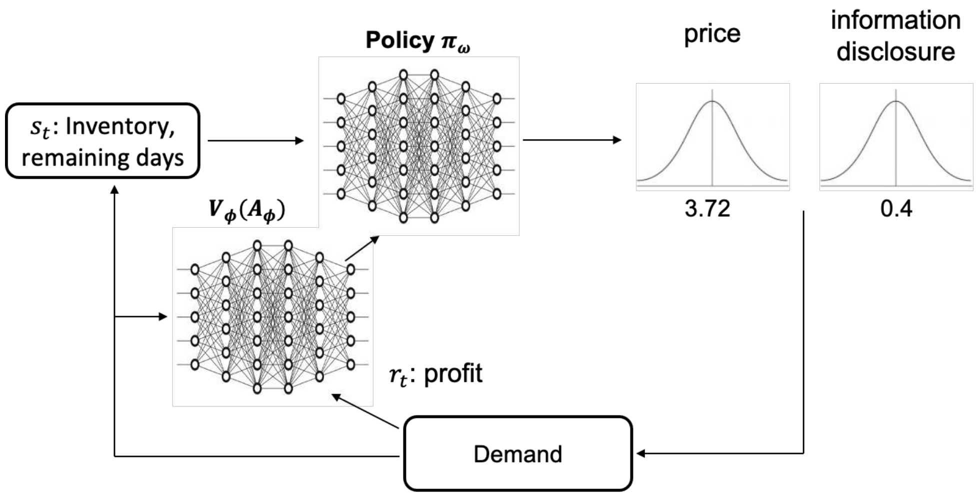


Figure 3: Implementation of PPO

perceive the quality as lower than the actual level, we set their deterioration rate $\hat{\gamma}_1 = 0.15$; (2) for customers who perceive the quality as the same as the actual level, we set their deterioration rate $\hat{\gamma}_2 = 0.1$; and (3) for customers who perceive the quality as higher than the actual level, we set their deterioration rate $\hat{\gamma}_3 = 0.05$. After observing the true quality disclosed by the retailer, each group of customers updates their beliefs on the deterioration rate $\hat{\gamma}_i$, based on the Bayesian rule. We employ *maximum a posteriori* estimate (Hastie et al. 2005) to update customers' perceived deterioration rate $\hat{\gamma}_i$ iteratively (See Online Appendix C for a detailed description).

Further, we consider three cases in terms of different portions of customers with different perceptions of quality. In case 1, we have equal-sized groups of customers with low, the same, and high perceived quality. In case 2, 50% of the customers have low perceived quality, with the other two groups comprising 25% each. In case 3, 75% of the customers have low perceived quality, with the other two groups comprising 12.5% each. See Online Appendix D for more details on the specifications of our simulated environments. We train each model for around 500 iterations, and each model achieves convergence after around 100 iterations. To save space, we show only the convergence of Model 2 in Figure 4. The convergence of the benchmark and that of Model 1 exhibit similar patterns. We then run the simulation for each trained model 30,000 times and take the average of the discounted total profit, the leftover inventory at the end of the selling period, prices, information strategies, and total demand.

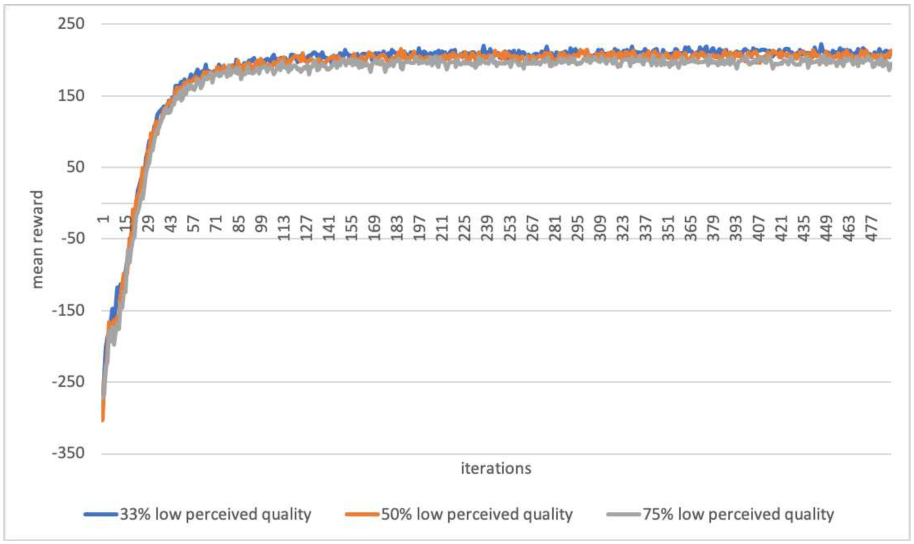


Figure 4: Convergence of Model 2

5.1 Benchmark versus Model 1

A major contribution of our work is the explicit consideration of food quality when deriving pricing and information strategies in a stochastic environment. Our results underscore the importance of a quality-based pricing strategy when the quality is changing over time. If we compare the benchmark model and Model 1 in the three different cases, a quality-based pricing strategy produces higher profits and less leftover inventory than a pricing strategy that does not take changing quality into account (see Tables 3 and 4). Specifically, compared to the benchmark, the profits in Model 1 increase by 7.6%, 7.5%, and 12% in cases 1 to 3, respectively, and the leftover inventory decreases by 74.1%, 61.4%, and 64.9%. The differences are statistically significant.

Table 3: Profit comparison

	Benchmark	Model 1	Model 2	Model 1 - Benchmark	Model 2 - Model 1
Case 1: 33% with low perceived quality	386.36	415.83	422.16	29.47***	6.33*
Case 2: 50% with low perceived quality	350.79	377.32	411.81	26.53***	34.49***
Case 3: 75% with low perceived quality	279.76	313.33	398.97	33.57***	85.64***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Leftover inventory comparison

	Benchmark	Model 1	Model 2	Model 1 - Benchmark	Model 2 - Model 1
Case 1: 33% with low perceived quality	26.3	6.8	7.9	-19.5***	1.1*
Case 2: 50% with low perceived quality	31.7	12.2	6.6	-19.5***	-5.6***
Case 3: 75% with low perceived quality	43.3	15.2	7.9	-28.1***	-7.3***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In line with prior literature (e.g., Bitran and Mondschein 1997, Gallego and Van Ryzin 1994), prices generally decrease over time (see Figures 5 to 7). Additionally, the prices in Model 1 are lower than those in the benchmark model. This result suggests that, if food quality is not taken into account, the retailer tends to set prices too high. High prices discourage purchases, and much inventory is thus left over at the end. The lower prices in Model 1 make up for customers' consideration of food quality and encourage more purchases. Thus, the demand in Model 1 is higher than in the benchmark model in cases 1 to 3 (see Table 5). Consequently, the retailer is able to sell more product and achieve higher profits. The ability to better adjust prices to demand in Model 1 is due to the capability of monitoring food quality in real time. This capability allows the retailer to obtain *implicit* information about how demand responds to prices, given current quality. This implicit information helps the retailer better tailor prices to demand. The intuition echoes previous literature, which states that demand learning is critical for dynamic pricing when information about the demand is limited (Chen and Chen 2015).

5.2 Model 1 versus Model 2

The incorporation of information disclosure helps the retailer achieve higher profits and have less leftover inventory when a large portion of the customers have low perceived quality (see Tables 3 and 4). When equal amounts of customers have perceptions of low, the same, and high quality, respectively, Models 1 and 2 yield roughly the same revenue and leftover inventory levels. The differences are not statistically significant.

Table 5: Total demand comparison

	Benchmark	Model 1	Model 2	Model 1 - Benchmark	Model 2 - Model 1
Case 1: 33% with low perceived quality	475.9	496.2	494.4	20.3***	-1.8*
Case 2: 50% with low perceived quality	469.3	489.0	496.4	19.7***	7.4***
Case 3: 75% with low perceived quality	457.1	486.3	495.0	29.2***	8.8***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

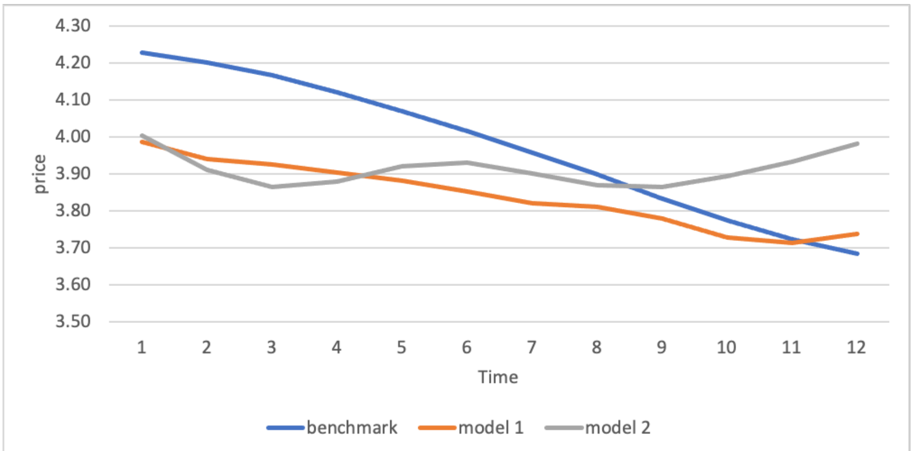


Figure 5: Optimal pricing strategy in case 1

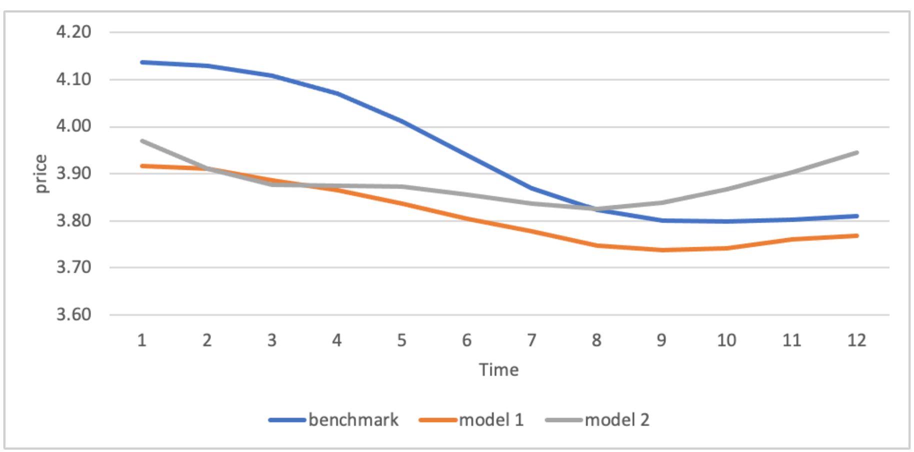


Figure 6: Optimal pricing strategy in case 2

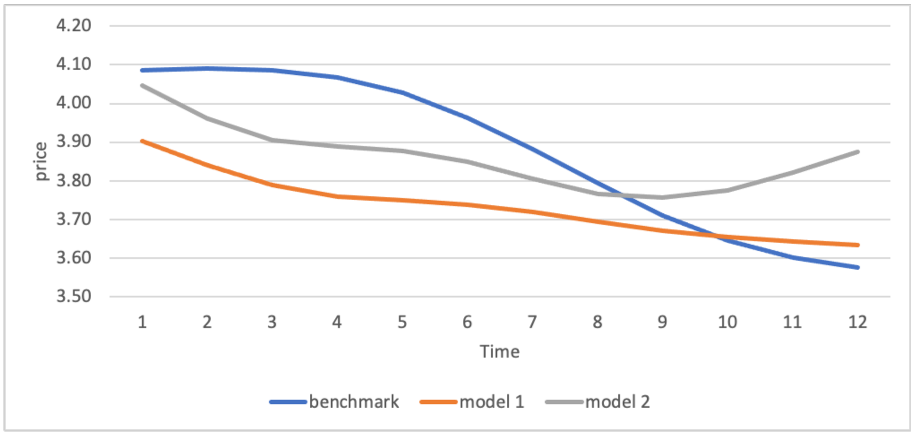


Figure 7: Optimal pricing strategy in case 3

However, Model 2 increases profits by 9.1% and 27.3% and reduces leftover inventory by 46.0% and 48.2% in cases 2 and 3, respectively, compared to Model 1. The superiority is due to the retailer's ability to align customers' biased perceptions about food quality with the true level through information disclosure. When customers' perception biases are high at the beginning of the selling season, it is optimal for the retailer to disclose quality information (see Figure 8). This result echoes that of Zhu et al. (2021), who show that retailers will disclose product quality information when customers' quality perception variability is high. We further show that, when the portion of customers with low perceived quality increases, the retailer has greater incentives to disclose information over time. The retailer stops disclosing information earlier in case 1 than in case 2, whereas, in case 3, the retailer has incentives to disclose information throughout the selling season. Thus, we show that the retailer does not have incentives to disclose quality information all the time. Our results provide a plausible explanation for the incomplete disclosure in many markets. In particular, our results imply that some biases between the customers' perceptions about the quality and the true quality can be beneficial, since the retailer can leverage this bias to extract value from those customers with the true perceived quality and high perceived quality. Consequently, the retailer has incentives to withhold quality information. However, when the dominant population perceives quality as low, it is better to disclose information throughout the selling season, to make sure this majority of customers knows the true state of the quality of the produce.

More interestingly, when information disclosure is allowed, prices stay the same or even increase during the selling season. In addition, the prices in Model 2 are generally higher than those in Model 1, especially in the later stages (see Figures 5 to 7). This is because, after the retailer discloses product information in the early stages, customers' perceptions about food quality become increasingly aligned with the true level in later stages. The alignment not only drives up demand (see Table 5), but also enables the retailer to maintain prices at a high level. The larger the portion of customers with low perceived quality, the more important it is to incorporate quality information disclosure, and the greater the improvement from the use of a combination of quality-based pricing and information disclosure.

5.3 Roles of quality-based dynamic pricing and information disclosure

Our results show that IoT-based monitoring capability opens up new opportunities for dynamic pricing and information disclosure, but the two operations tools play different roles. A quality-based dynamic pricing facilitates demand learning and thus enables the retailer to better tailor prices to demand. Information disclosure aligns customers' biased perception about the quality with the true state and thus affects the demand. When and how to employ the two tools largely depend on the portion of customers who have biased perceptions about the quality. When a large portion of them consider quality to be low and information disclosure is not allowed, the retailer can only use prices to influence demand. Thus, the retailer ends up pushing the prices lower and lower to get customers to buy the product. This leads to a decreasing average price per unit sold from case 1 to case 2 to case 3 in the benchmark and Model 1 (Table 6).

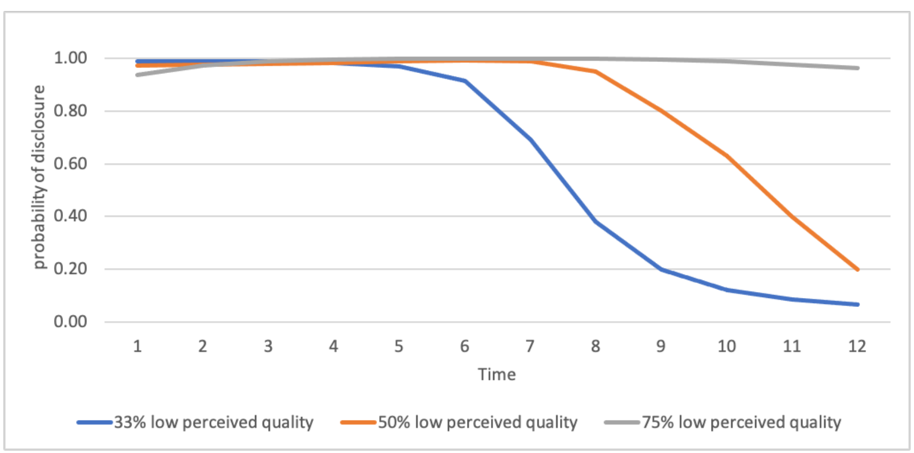


Figure 8: Optimal information disclosure

In Model 2, however, the average price remains roughly the same across the three cases. Thus, when a large portion of the customers have low perceived quality, quality-based pricing alone is not sufficient to effectively sell fresh produce. Our results speak to the importance of incorporating information disclosure in markets where customers are very risk averse about product quality, such as markets for ugly produce and produce that has been involved in a high-profile accident.

Additionally, when information disclosure is allowed, we show that the optimal prices remain roughly the same during the selling season. This finding indicates that a fixed pricing strategy might be sufficient in combination with information disclosure. Similar results are found by Sanders (2020), who shows that waste bans discourage the retailer from adopting dynamic pricing and that a combination of waste bans and fixed pricing strategy helps reduce food waste. Despite recent managerial push for dynamic pricing enabled by new technology (Xu et al. 2019), dynamic pricing has not hit the store level in any significant way in the United States.[10] Our results indicate one possibility of keeping prices relatively stable and achieving high performance by incorporating information disclosure.

Finally, quality-based dynamic pricing and information disclosure improve the retailer's profits, but they impose different welfare implications for customers. In line with Stamatopoulos et al. (2019), we consider that a lower average price per unit sold[11] implies a higher rate of consumer surplus. We show that Model 1 has a lower average price per unit sold than the benchmark model across the three cases (Table 6). The results indicate that quality-based dynamic pricing increases consumer surplus. The increase can be attributed to the fact that the retailer will lower prices to account for the deteriorating quality after having obtained implicit information about the demand, which depends on both the price and customers' perceived quality. We also reveal that the average price per unit sold is higher in Model 1 than in Model 2 (Table 6). This is because information disclosure drives demand through perception alignment. After customers' biased perceptions are aligned with the true state, the retailer is able to set higher prices, thus decreasing consumer surplus.

[10]RetailWire: Has dynamic pricing hit a rut?

[11]The average price per unit sold is calculated as total rewards divided by total demand.

Table 6: Average price per unit sold

	Benchmark	Model 1	Model 2
Case 1: 33% with low perceived quality	3.964	3.861	3.888
Case 2: 50% with low perceived quality	3.963	3.839	3.851
Case 3: 75% with low perceived quality	3.894	3.729	3.836

5.4 Cost analysis

Till now, we have focused on the benefit of IoT-based monitoring capabilities. The benefits depend on the adoption of IoT-based monitoring capability and AI methods.[12] In practice, however, deploying and maintaining IoT and AI require additional costs (see Online Appendix A for a more detailed discussion). Such costs are likely to prevent retailer adoption. To evaluate the issue, we increase the unit cost in Models 1 and 2 and compare their performance with that of the benchmark. We consider cases where the unit cost in Models 1 and 2 is 3%, 6%, and 10% higher, respectively, than in the benchmark model. We calculate the difference in profits (leftover inventory) between Model 1 (Model 2) and the benchmark. A negative difference in profit (leftover inventory) indicates lower profit (leftover inventory) in Model 1 (Model 2) than in the benchmark model. Our results show that, when the unit cost incurred by the adoption of IoT and AI increases by 10%, the benchmark model yields higher profits than Models 1 and 2 across all three cases (Table 7). When the cost increase is moderate (i.e., 3% and 6%), the superiority of Model 2 still holds in case 3, when the majority of customers have low perceived quality. In terms of leftover inventory, a higher unit cost does not affect the role of quality-based dynamic pricing and information disclosure in reducing leftover inventory. Both Models 1 and 2 produce much less food waste than the benchmark. The differences are negative and remain roughly the same, no matter how much the unit cost increases for each case (Table 8). Our results suggest that negative environmental externality is not internalized by a profit-maximizing retailer. Though the retailer can improve profits by not adopting IoT and AI when the cost of adoption is high, it is socially undesirable. Thus, the social planner must incentivize retailers to adopt IoT to monitor food quality and employ AI methods to make business

[12] See the automated system we propose in the beginning (Figure 1).

decisions on the quality.

Table 7: Profit comparison when the unit cost increases in Models 1 and 2

	Model 1 - Benchmark			Model 2 - Benchmark		
	3% increase	6% increase	10% increase	3% increase	6% increase	10% increase
Case 1	−21.16	−70.64	−121.02	−14.43	−67.06	−119.10
Case 2	−25.14	−72.90	−122.69	**10.43**	−36.76	−86.29
Case 3	−16.29	−67.12	−114.76	**71.13**	**20.09**	−27.86

Table 8: Leftover inventory comparison when the unit cost increases in Models 1 and 2

	Model 1 - Benchmark			Model 2 - Benchmark		
	3% increase	6% increase	10% increase	3% increase	6% increase	10% increase
Case 1	−19.19	−18.95	−18.57	−19.22	−19.43	−14.85
Case 2	−21.21	−23.00	−20.69	−23.03	−24.45	−24.10
Case 3	−30.56	−29.45	−28.81	−31.61	−32.23	−36.11

6. Robustness checks

We first test the robustness of our results with respect to the portion of customers with low perceived quality. We examine the cases in which customers with *high* perceived quality account for the majority of the customer base. The superiority of a quality-based pricing strategy over a pricing strategy without quality still holds. However, when the majority of customers perceive quality as high, the retailer has little incentive to disclose information. However, the incorporation of information disclosure does not hurt performance, as indicated by the insignificant difference between Models 1 and 2 in profit and leftover inventory (See Section E.1 in Online Appendix E).

Additionally, we change the parameters of the true deterioration rate γ and customers' perceived deterioration rate $\hat{\gamma}$, since these parameters directly affect customers' behavior. We use a different bias between the true deterioration rate and customers' biased perceptions about deterioration rate. Keeping the true deterioration rate γ at 0.1, we set the perceived deterioration rate $\hat{\gamma}$ of customers with low perceived quality at 0.12, and that of customers with high perceived quality at 0.08. The main results remain qualitatively the same (see Section E.2 in Online Appendix E). Comparing the results with greater bias between the true deterioration rate and customers' perceived deterioration rate, we find that the retailer has less incentive to disclose quality information. This finding lends support to the motivation behind information disclosure. Additionally, we find that the retailer is able to set higher prices when the bias is smaller. This finding lends further support to the trade-off between pricing and information disclosure. Greater alignment between customers' perceptions and the true remaining days (i.e., food quality) enables the retailer to set higher prices. We also change the true deterioration rate γ to 0.2 and set the perceived deterioration rates $\hat{\gamma}$ of customers with low and high perception quality, respectively, at 0.25 and 0.15. The main results remain qualitatively the same (see Section E.2 in Online Appendix E).

Furthermore, we change the parameter of demand rate, which is the customers' arrival rate of Poisson distribution. Again, the main results remain qualitatively the same (see Section E.3 in Online Appendix E). Our results also indicate that when demand rate is high, the benefit that information disclosure could bring diminishes. This is intuitive as the retailer does not need to align customers' biased perceptions in order to drive the demand up. Finally, we change the assumption of the exponential deterioration of food quality. We use a linear form of food deterioration and assume a normal distribution for customers' prior belief on the deterioration rate. Again, our main results remain qualitatively the same (see Section E.4 in Online Appendix E).

In addition to testing the robustness of our model, we also test the robustness of our method with respect to key parameters in PPO algorithm. We conduct sensitivity analysis with respect to hidden size of neural networks, step size, batch size, and clip parameter. Our results show that PPO is quite robust to those parameters. Total rewards and leftover inventory remain similar across different settings (see section E.5 in Online Appendix E).

7. Conclusions and discussions

Since fresh produce is an important part of grocery retailing, the design of a system to effectively sell fresh produce and thus promote sustainability is of great significance. The current work is among the first to explore optimal quality-based pricing and information disclosure dynamically to achieve the goal. Our findings yield many managerial implications.

First, we find that a quality-based pricing strategy is a win–win–win strategy. It helps increase profits for the retailer, reduces food waste for the society, and increases consumer surplus through charging a lower average price. The improvement comes from the retailer's ability to observe implicit information about how demand responds to prices, given quality. By doing so, we extend the literature on the value of IoT (Caro et al. 2020, Kumar et al. 2018, Olsen and Tomlin 2020) by showcasing the value of IoT-enabled monitoring capabilities. We also contribute to the literature on dynamic pricing by incorporating changing product quality and considering customers' heterogeneity with respect to their biased perceptions about the rate at which fresh produce deteriorates.

Additionally, we show that information disclosure can further improve profits and reduce the remaining inventory when a large portion of customers have low perceived quality. The improvement comes from the alignment of customers' biased perceptions about the deterioration rate of food quality with the true deterioration rate. Such alignment enables the retailer to maintain prices at high levels during the selling season. Thus, we uncover a similar role of information disclosure as the waste bans shown by Sanders (2020). Our results also suggest one possibility of keeping prices stable and achieving high performance by incorporating information disclosure. Third, we show that a combination of a quality-based dynamic pricing strategy and information disclosure strategy generates the highest profits and lowest food waste in markets where customers are very risk averse in terms of quality, customers' perceptions about the quality are high, and the demand rate is low.

Finally, our work is among the few to describe an application of DRL in operations management. Our results show the promise of applying an automated system (Figure 1) to better manage fresh produce in practice. The automated system depends on IoT-enabled monitoring capability and AImethods (i.e., DRL). Despite its promise, the implementation of such an automated system requires additional costs which are likely to hinder a profit-maximizing retailer from adopting the system, thus creating more food waste. Thus, a social planner should provide incentives for retailers to leverage IoT and AI methods.

Our work presents several avenues for future research. First, future work could look into the problem of joint decisions on pricing strategy and inventory management. Without modeling product quality as deteriorating over time, past research has laid a strong foundation with both theoretical and empirical results (for a review, see Elmaghraby and Keskinocak 2003). Taking the decaying quality of products into account could lead to new insights. Second, future work could extend our model to consider strategic customers. Strategic behavior adds extra complexity to analyzing both pricing and information strategies. However, it is worth noting that strategic behavior is more likely to occur among price-sensitive customers, while quality-sensitive customers care more about the quality of fresh produce and are thus less likely to wait for the price drop. Finally, while we consider a monopolistic market, it would be interesting to explore optimal pricing and information strategies in a competitive market.

Acknowledgment

We extend our thanks to Amazon's Grocery Supply Chain group for constructive comments and suggestions regarding the design of the cloud-based automated system.

References

Atasu, A., C. J. Corbett, X. Huang, L. B. Toktay. 2020. Sustainable operations management through the perspective of manufacturing & service operations management. *Manufacturing & Service Operations Management*, **22**(1), 146-157.

Bai, R., G. Kendall. 2008. A model for fresh produce shelf-space allocation and inventory management with freshness-condition-dependent demand. *INFORMS Journal on Computing*, **20**(1), 78-85.

Belavina, E. 2021. Grocery store density and food waste. *Manufacturing & Service Operations Management*, **23**(1), 1-18.

Belavina, E., K. Girotra, A. Kabra. 2017. Online grocery retail: Revenue models and environmental impact. *Management Science*, **63**(6), 1781-1799.

Bitran, G., R. Caldentey. 2003. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, **5**(3), 203-229.

Bitran, G., S. Mondschein. 1997. Periodic pricing of seasonal products in retailing. *Management Science*, **43**(1), 64-79.

Bou-Mitri, C., M. Abdessater, H. Zgheib, Z. Akiki. 2021. Food packaging design and consumer perception of the product quality, safety, healthiness and preference. *Nutrition & Food Science*, **51**(1), 71-86

Buzby, J. C., J. Hyman. 2012. Total and per capita value of food loss in the United States. *Food Policy*, **37**(5), 561-570.

Cao, H., X. Guan, T. Fan, L. Zhou. 2020. The acquisition of quality information in a supply chain with voluntary vs. mandatory disclosure. *Production and Operations Management*, **29**(3), 595-616.

Caro, F., A. G. Kök, V. Martínez-de-Albéniz. 2020. The future of retail operations. *Manufacturing & Service Operations Management*, **22**(1), 47-58.

Chen, M., Z. L. Chen. 2015. Recent developments in dynamic pricing research: Multiple products, competition, and limited demand information. *Production and Operations Management*, **24**(5), 704-731.

Cohen, M. C., R. Lobel, G. Perakis. 2018. Dynamic pricing through data sampling. *Production and Operations Management*, **27**(6), 1074-1088.

Dong, L. 2021. Toward resilient agriculture value chains: challenges and opportunities. *Production and Operations Management*, **30**(3): 666-675.

Dutta, A., H. L. Lee, S. Whang. 2007. RFID and operations management: technology, value, and incentives. *Production and Operations Management*, **16**(5), 646-655.

Elliott, G. R., R. C. Cameron. 1994. Consumer Perception of Product Quality and the Country-of-Origin Effect1. *Journal of international Marketing*, **2**(2), 49-62.

Elmaghraby, W., P. Keskinocak. 2003. Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science*, **49**(10), 1287-1309.

Ely, J. C. 2017. Beeps. *American Economic Review*, **107**(1), 31-53.

Gallego, G., G. Van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, **40**(8), 999-1020.

Golrezaei, N., H. Nazerzadeh, R. Randhawa. 2020. Dynamic pricing for heterogeneous time-sensitive customers. *Manufacturing & Service Operations Management*, **22**(3), 562-581.

Grossman, S. J. 1981. The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics*, **24**(3), 461-483.

Grunert, K. G. 2005. Food quality and safety: Consumer perception and demand. *European Review of Agricultural Economics*, **32**(3), 369-391.

Guan X., Y. J. Chen. 2017. The interplay between information acquisition and quality disclosure. *Production and Operations Management*, **26**(3), 389-408.

Guo L., Y. Zhao. 2009. Voluntary quality disclosure and market interaction. *Marketing Science*, **28**(3), 488-501.

Hastie, T., R. Tibshirani, J. Friedman, J. Franklin. 2005. The elements of statistical learning: Data mining, inference and prediction. *Mathematical Intelligencer*, **27**(2), 83-85.

Hu, X., Z. Wan, N. N. Murthy. 2019. Dynamic pricing of limited inventories with product returns. *Manufacturing & Service Operations Management*, **21**(3), 501-518.

Jin, G. Z. 2005. Competition and disclosure incentives: An empirical study of HMOs. *RAND Journal of Economics*, April (1), 93-112.

Jovanovic, B. 1982. Truthful disclosure of information. *Bell Journal of Economics*, **13**(1), 36-44.

Ke, J., D. Zhang, H. Zheng. 2019. An Approximate Dynamic Programming Approach to Dynamic Pricing for Network Revenue Management. *Production and Operations Management*, **28**(11), 2719-2737.

Kuksov, D., Y. Lin. 2010. Information provision in a vertically differentiated competitive marketplace. *Marketing Science*, **29**(1), 122-138.

Kumar, S., V. Mookerjee, A. Shubham. 2018. Research in operations management and information systems interface. *Production and Operations Management*, **27**(11), 1893-1905.

Li, H., W. T Huh. 2011. Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications. *Manufacturing & Service Operations Management*, **13**(4), 549-563.

Li, S., L. Da Xu, S. Zhao. 2015. The internet of things: a survey. *Information Systems Frontiers*, **17**(2), 243-259.

Liu, Y., W. L. Cooper, Z. Wang. 2019. Information provision and pricing in the presence of consumer search costs. *Production and Operations Management*, **28**(7), 1603-1620.

Miao, S., X. Chao. 2021. Dynamic joint assortment and pricing optimization with demand learning. *Manufacturing & Service Operations Management*, **23**(2), 525-545.

Milgrom, P. R. 1981. Good news and bad news: Representation theorems and applications. *Bell Journal of Economics*, **12**(2), 380-391.

Olsen, T. L., B. Tomlin. 2020. Industry 4.0: Opportunities and challenges for operations management.

Manufacturing & Service Operations Management, **22**(1), 113-122.

Qin, Z., X. Tang, Y. Jiao, F. Zhang, Z. Xu, H. Zhu, J. Ye. 2020. Ride-hailing order dispatching at DiDi via reinforcement learning. *INFORMS Journal on Applied Analytics*, **50**(5), 272-286.

Sanders, R. E. 2020. Dynamic pricing and organic waste bans: A study of grocery retailers' incentives to reduce food waste. Available at SSRN 2994426.

Schulman, J., S. Levine, P. Abbeel, M. Jordan, P. Moritz. 2015. Trust region policy optimization. *Proceedings of the 32nd International Conference on Machine Learning*, **37**, 1889-1897.

Schulman, J., F. Wolski, P. Dhariwal, A. Radford, O. Klimov. 2017. Proximal policy optimization algorithms. *arXiv*, 1707.06347v2.

Shen, Z. J. M., X. Su. 2007. Customer behavior modeling in revenue management and auctions: A review and new research opportunities. *Production and Operations Management*, **16**(6), 713-728.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, **529**(7587), 484.

Stamatopoulos, I., N. Chehrazi, A. Bassamboo. 2019. Welfare implications of inventory-driven dynamic pricing. *Management Science*, **65**(12), 5741-5765.

Sun, M., R. K. Tyagi. 2020. Product fit uncertainty and information provision in a distribution channel. *Production and Operations Management*, **29**(10), 2381-2402.

Sutton, R. S., A. G. Barto. 2018. *Reinforcement Learning: An Introduction.* MIT Press, Cambridge, MA.

Williams, R. J. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, **8**(3-4), 229-256.

Xu, J., P. S. Fader, S. Veeraraghavan. 2019. Designing and evaluating dynamic pricing policies for major league baseball tickets. *Manufacturing & Service Operations Management*, **21**(1), 121-138.

Zhao, W., Y. Zheng. 2000. Optimal dynamic pricing for perishable assets with nonhomogeneous demand. *Management Science*, **46**(3), 375-388.

Zhu, H., Y. Yu, S. Ray. 2021. Quality disclosure strategy under customer learning opportunities. *Production and Operations Management*, **30**(4): 1136-1153.