Int. J. Production Economics 292 (2026) 109824

Contents lists available at ScienceDirect

International Journal of Production Economics

journal homepage: www.elsevier.com/locate/ijpe

Dynamic expiration date-based discounting of fresh food products☆

Rene Haijema [a],*, Lisan Duijvestijn [a], Renzo Akkerman [a], Frans Cruijssen [b]

[a] *Operations Research and Logistics, Wageningen University, The Netherlands*
[b] *Tilburg School of Economics and Management, Tilburg University, The Netherlands*

ARTICLE INFO

Keywords:
Retail
Markdown
Pricing
Food waste
Consumer behavior
Dynamic programming

ABSTRACT

To reduce food waste, many supermarkets discount food products that are close to their expiration date. In practice, this is done either by discount labels put on the product or by electronic shelf labels (or digital price tags) showing the price per expiration date. Digital price tags allow to easily change the price of products and to apply different discount rates to items with different expiration dates. An important question to practitioners is when and how much discount to offer. In this study, we use Stochastic Dynamic Programming (SDP) to derive optimal expiration-date-based discounting policies for a profit-maximizing retailer who sells a product with m periods (e.g., days) of shelf life. We compare various discounting strategies, such as static last-day discounting, optimal dynamic last-day, and last-two-days discounting, against the no-discounting strategy.

The model allows products of different expiration dates to be in stock simultaneously, as replenishment happens every period. In the last-day discounting policies, two selling prices co-exist: the regular price and the discounted price. When applying a last-two-days discounting policy, three selling prices co-exist. Demand and product withdrawal depend on both price and product age (freshness). We consider different customer picking behavior, and divide customers into First-Expiry-First-Out (FEFO) and Last-Expiry-First-Out (LEFO) consumers (i.e, customers that pick the oldest items first and customers that take the freshest items available). For LEFO customers, we also consider that a fraction of these customers will pick discounted old items (depending on the size of discount). Finally, extra demand is attracted as long as discounted products are available.

Optimal policies are derived by SDP and evaluated by simulation to generate insights into the impact of discounting on profits, sales, fill rates, and waste. Various key factors, such as shelf life, customer picking behavior, and discount sensitivity are analyzed in detail. The results show that the last-two-days discounting policy performs well. Averaged over all experiments, this policy demonstrates a 3.8% increase in profits compared to no-discounting, and a waste reduction from 5.6% to 3.6%. Smaller, but still significant improvements are shown over simpler discounting policies.

1. Introduction

Perishable food products – such as dairy, meat, fruit, vegetables, and bakery products – are crucial for the profitability and store image of supermarkets (Ferguson and Ketzenberg, 2006; Tsiros and Heilman, 2005). In the United Kingdom, these categories account for 52.4% of its unit sales (Dudlicek et al., 2019) and greatly contribute to supermarket profitability (Tsiros and Heilman, 2005). Additionally, perishable products can be used to create a competitive advantage, as they play a significant role in attracting consumers to a particular supermarket over others (Ferguson and Ketzenberg, 2006). Product availability is thus key for many supermarkets.

As a result of the focus on product availability, a certain amount of fresh food expires before it is sold and consumed. At the supermarket level, in 2020, around 1.7% of the food went to waste (Vollebregt, 2020). This number is much higher for certain product categories: 7.7% for bread and pastry; 2.9% for fresh meat and fish; 2.7% for potatoes, vegetables, and fruit; and 1.4% for dairy, eggs, and chilled convenience products (Vollebregt, 2020, 2023). This not only has a negative impact on sustainability but also on the supermarkets' profits (Scholz et al., 2015). Their profits can increase by 15% when the waste of perishable goods is reduced (Tsiros and Heilman, 2005).

Currently, supermarkets are already practicing various strategies to minimize food waste, such as offering discounts for products that are near their expiration date, selling food boxes with near-to-expiry products (e.g., Too Good To Go), and donating surplus food to food banks. These measures help to prevent the food from going to waste. In

☆ This article is part of a Special issue entitled: 'ISIR 2024' published in International Journal of Production Economics.

* Corresponding author.
E-mail address: rene.haijema@wur.nl (R. Haijema).

https://doi.org/10.1016/j.ijpe.2025.109824
Received 20 December 2024; Received in revised form 3 September 2025; Accepted 10 October 2025
Available online 19 October 2025

Fig. 1. Example of a digital price tag with two different expiration-based discounts.
Source: Taken from S&K Solutions GmbH (2025).

this study, we focus on the impact of expiration-date-based discounting. That is, the price of products depends on their expiration date. Prices can be adjusted by putting a discount label (sticker) on the product, or by using electronic shelf labels or digital price tags, like in Fig. 1 (taken from S&K Solutions GmbH (2025)).

The advent of new technology such as electronic shelf labels and digital price tags, see e.g. Herbon et al. (2014), place discounting at the center of attention of retailers as the new tools enable for smarter and more targeted discounting policies. For a recent report, Akkas (2024) interviewed 19 retailers globally to understand their practices in managing food surplus. It turned out that many retailers implement discounts on the same day or a day before expiration. For longer products with a longer shelf life, several retailers initiate discounts even earlier (e.g., 3–4 days prior to expiration).

Furthermore, a group of innovative retailers adopts a two-step discounting approach, such as starting discounting early at say 30% discount and when time pass increasing it to 60%–70%, if the product still remains unsold. Two retailers even used a three-step discounting strategy. Although smart discounting has the potential to increase sales and reduce waste, there is currently no consensus about the optimal strategy. As Akkas (2024) states, no retailer has fully resolved the discounting challenge. Even the most advanced retailers continue to experiment with various strategies to balance profit optimization and waste minimization. This state of continuous improvement in the discounting processes renders it timely to investigate optimal strategies for this challenge.

Next to the retailer, also consumers play a crucial role in generating food waste in supermarkets. Harcar and Karakaya (2005) found that many consumers inspect expiration dates when purchasing perishable products. They are aware of the expiry date and adjust their purchasing behavior accordingly. Stenmarck et al. (2011) showed that some consumers tend to choose products with the longest best-before dates, even if they intend to consume them the same day. This results in increased waste levels, if older items remain unsold. This type of consumer picks items from the shelf in the order Last-Expiry-First-Out (LEFO). Therefore, in this paper, we call them LEFO consumers, as opposed to FEFO consumers who pick in order of First-Expiry-First-Out (FEFO), see Ostermeier et al. (2021) and Brandimarte and Gioia (2022). In well-organized shops oldest products are at the front row on a shelf. FEFO consumers buy the oldest products, either for sustainability reasons, or because they grab an arbitrary item from the front of the display (where retailers tend to position the oldest items) . For a more in-depth discussion on customer behavior and sensitivity to product freshness, we refer to Herbon (2014) and Chang and Su (2022).

Discounting the oldest items in stock makes them more attractive. The higher the discount, the more LEFO consumers are expected to buy a discounted item instead of a fresher item. Although it has been found that not all LEFO consumers are price-sensitive. Aschemann-Witzel (2018) and Hansen et al. (2024). As discounted items will also be picked by FEFO consumers, discounting can be detrimental to profit. This loss in profit is partly compensated for by a decrease in waste and by an increase in total sales as discounted items generate extra sales. Whether to discount or not, and how much discount to offer are delicate decisions that should anticipate many aspects, such as consumer behavior, the number of items in stock, and their ages. As these numbers vary, dynamic discounting policies, i.e. strategies where the price of a product is adjusted dynamically over time based on factors such as the number of items left, the selling season, or the product expiration date, are more suitable than static or fixed discounting strategies.

According to a recent literature review (Riesenegger et al., 2023), many studies on dynamic pricing assume that one finds at most two different expiration dates. Consequently, most optimization studies are limited to two co-existing selling prices. In practice, consumers may find more than two different expiration dates on the shelf, as fresh products are often replenished every day. Using 2D codes and digital price tags, retailers can then also use more than two expiration-date-based prices. At the start of every period, the prices, or discount rates, can be adjusted dynamically based on the actual number of items in stock and their ages.

Only a few studies so far focus on setting with more than two expiration dates on display, and more research is needed to reduce food waste and maintain profitability. Determining optimal prices, or discount rates, is rather complex in settings where customers may choose from items with three or more different expiration dates.

In this paper, we derive optimal dynamic expiration-date-based discounting policies for perishable inventory systems with possibly more than two co-existing expiration dates. We study the relevant but understudied setting of a non-monopolistic retailer that has items with more than two different expiration dates simultaneously on sale. Furthermore, we consider that a significant fraction of the consumers pick in FEFO order (i.e., they buy the oldest items in stock even when these are not discounted) and that LEFO customers might be convinced to buy older items when discounts are offered.

We formulate the dynamic optimization problem as a Markov decision process (MDP) that considers an order lead time of one period. We solve the problem by stochastic dynamic programming for a fixed shelf life of $m \in \{3, 4, 5\}$ periods. Without loss of generality, we assume a period to be a day, but one may also choose to set it to half a day, two days, or a week. Next to optimizing two discount rates in a dynamic last-two-days discounting policy, we also determine optimal dynamic discounting policies with a single discount rate.

This paper is further organized as follows. First, in Section 2 an overview of the current literature on pricing and discounting of perishable products is provided with emphasis on multiple age classes. In Section 3 we formulate an MDP model. Sections 4 and 5 present a numerical study for comparing different discounting policies for a broad set of experiments that vary in maximal shelf life, mean demand, and consumer responses to freshness and discounts. Section 6 closes the paper with a discussion and conclusion.

2. Literature

Discounting is a well known approach to reduce food waste, especially for perishable products with a relatively short shelf life (Nijs et al., 2001). As the consumers' willingness to pay declines with the age of the product, offering early discounts is effective in food waste reduction (Tsiros and Heilman, 2005). When the discount applies to all items with the same expiration date, this is referred to as expiration-date-based discounting.

Expiration-date-based discounting belongs to the broader stream of literature on dynamic pricing. Dynamic pricing is a form of revenue management that is applied in various settings, ranging from the airline industry to e-commerce and food retailing. In this section, we discuss relevant literature on (dynamic) pricing and discounting of perishable food products with a short shelf life, such as packed fresh food.

Table 1
Overview of relevant literature on multi-period pricing models.

Paper	Number of co-existing		Product	Multiple	Method[b]
	Ages	Prices	picking	restocks[a]	
Single age class					
Chatwin (2000)	1	1	–	No	SDP
Liu et al. (2008)	1	1	–	No	SDP, Sim
Zhang et al. (2015)	1	1	–	No	PMP, Sim
Rabbani et al. (2016)	1	1	–	No	MA, Sim
Li and Teng (2018)	1	1	–	No	MA, Sim
Duan and Liu (2019)	1	1	–	No	PMP, Sim
Two age classes					
Chew et al. (2014)	2	2	DA	Yes	SDP
Fan et al. (2020)	2	2	U	No	SDP
Scholz and Kulko (2022)	2	2	U	No	Sim
Sanders (2024)	2	2	F	No	SDP, Sim
Three or more ages, ≤2 prices					
Chua et al. (2017)	$m \in \{2, 4\}$	2	LP	Yes	SDP
Kaya and Ghahroodi (2018)	$m \in \{1, 2, 3\}$	1	L, LP	Yes	SDP
Buisman et al. (2019)	$m \in \{5, 8, 10\}$	2	F, L, LP	Yes	Sim
Three or more ages, >2 prices					
Chung and Li (2014)	$m \in \{7, 11, 15\}$	$\leq m$	LP	Yes	Sim
Adenso-Díaz et al. (2017)	$m = 10$	$\leq m$	DA	No	Sim
Yavuz and Kaya (2024)	$m \in \{2, 3, 4, 5\}$	$\leq m$	DA	Yes	SDP, RL
This paper	$\mathbf{m \in \{3, 4, 5\}}$	**2–3**	**F, L, LP**	**Yes**	**SDP, Sim**

[a] F = FEFO = Oldest first; L = LEFO = Youngest first; LP = LEFO customers give priority to lowest priced items; U = by a utility function depending on price and/or age; DA = demand is for specific age class (or younger).
[b] SDP = Stochastic Dynamic Programming; Sim = Simulation; MA = Mathematical Analysis; PMP = Pontryagin's Maximum Principle; RL = Reinforcement Learning.

The literature on modeling and optimization of expiration-date-based discounting can be categorized based on the number of age classes that are modeled and the number of selling prices that co-exist, i.e., the number of different prices and expiration dates one may find on the shelf simultaneously. Table 1 gives an overview of the most relevant papers, which we will briefly discuss in the remainder of this section.

2.1. Single age class

Typically, in models with only one age class, one (implicitly) assumes that when new items arrive, the remaining old stock is disposed of in a secondary market, or wasted. The key questions are (i) when to adjust the price, and (ii) when to order new items (and thus to dispose of the remaining items). Most studies assume a profit-maximizing monopolistic retailer who can freely set the price. In most studies, the demand function depends on the price and the age of the products in stock. As all items in stock are of the same age, one does not need to make assumptions about the picking order. The mathematical analysis is relatively simple by focusing on a single replenishment cycle. The problem can be solved by a finite horizon Stochastic Dynamic Program if the replenishment cycle is fixed. Alternatively, one may optimize the cycle length by disposing of any remaining stock at the end of the replenishment cycle. Next we discuss the relevant papers in this category.

Liu et al. (2008) study a single perishable product for which the demand is decreasing linearly in price and exponentially in product age. They maximize the expected profit over a sales cycle by dynamically adjusting the price and the order quantity. Similarly to the approach that we will follow in out paper, Stochastic Dynamic Programming (SDP) is used to derive analytical results and the numerical results are obtained by simulation.

Zhang et al. (2015) consider an initial inventory of a perishable product that can be sold in a sales cycle of T periods. The product does not decay in the first t_d periods, but thereafter it decays at a constant rate θ. Inventory levels go down not only because of consumer demand but also as a result of product disposal. Consumer demand depends on the price as well as the quantity of product in stock, but not on the age of the products. Optimal control theory is used to maximize profit by setting the sales cycle T, and (dynamically) controlling the sales price in continuous time. Rabbani et al. (2016) also consider a deterioration-free period of length t_d, after which deterioration results in disposal and lost sales. Contrary to Zhang et al. (2015), the demand does depend on product age, price, and the change in price. Demand is deterministic, but deterioration is modeled by Weibull distributions.

Li and Teng (2018) integrate the demand models in the above two papers: demand is a function of price, reference price, age, and stock level. Both the demand and the deterioration are deterministic. The product has a fixed maximal shelf life of m periods, and product freshness at age t, is set to $f(t) = \frac{m-t}{m}$. The demand decreases linearly over time: at time $t = m$ the demand will be zero. Structural results are derived for the optimal pricing strategy and illustrated by two numerical examples.

Duan and Liu (2019) study a finite-horizon problem of selling an initial inventory of some perishable products until it is sold out. No production and replenishment occur during the horizon. The demand function is a linear function of price, reference price, and product age. The reference price is determined by the weighted average historical price. Using Pontryagin's maximum principle, optimal dynamic prices are determined.

Typically in the above studies, the profit is optimized by a monopolist by setting a single price and assuming a single product without age differences. As the price can be set freely and the demand volume depends on this price, it is necessary to simultaneously optimize the pricing and the ordering decisions.

2.2. Two age classes

Models with two age classes distinguish between 'old' and 'new' items in stock. In most studies, authors assume one single review period or a shelf life covering two review periods. SDP can be applied to determine the optimal prices of the old and new products. Below, we discuss the most relevant papers with two age classes in chronological order.

Chew et al. (2014) do not consider a food product but rather a generic product of which different versions are brought to the market over time. The analogy is that products of the same version are of the same age class in our context. Rather than having a single total demand function, they model the demand for each specific age class that depends on the price level for that class and that of neighboring classes. Demand transfers from one age class to another are proportional to the price difference. In case of a stock out of one age class, demand for that age class is lost and not transferred to other age classes. In their study, the authors obtain optimal pricing decisions through SDP for products with two age classes, and for more age classes, they propose a heuristic based on the optimal solution to a single-period problem.

Chua et al. (2017) present three models on products with a shelf life of two periods. All consumers are assumed to pick in LEFO order, unless a discounted product is available. Depending on the size of the discount x_t at time t, a fraction $a(x_t)$ of the consumers picks the discounted products. In one of the models, extra demand is attracted as long as discounted products are available. The number of extra customers depends on the discount rate and on how many base-demand customers purchase a discounted product.

Fan et al. (2020) consider a monopolistic retailer who sells fresh food products with at most two different expiration dates in stock simultaneously. By means of SDP, an optimal pricing and ordering strategy is determined. The customer's decision whether to buy, and if so, from which age class, is stochastic and follows a utility model, which is linear in product freshness and price. The SDP is solved with a finite horizon of two periods, during which only at the start a replenishment order is set which arrives instantaneously.

Scholz and Kulko (2022) consider a retailer that simultaneously offers fresh strawberries and strawberries that are three days old. The number of items demanded from an age class depends on their price as well as the price of products in the other age class. Product withdrawal and purchasing probabilities are derived from a utility function, which is fitted to willingness-to-pay data derived from online experiments. Dynamic pricing policies are evaluated using Monte Carlo Simulation.

Recently, Sanders (2024) applies a structural econometric model to study the extent to which dynamic pricing is a more effective way to reduce food waste than a 'waste ban'. Therefore, counterfactual simulations are executed, using data from a US grocery chain. The sales transaction data for artisanal bread shows that demand depends on the time of the day and the day of the week. The optimal pricing problem is modeled as solved by a finite horizon SDP with a four-dimensional state: the day of the week, the time of the day, and the number of products in stock of the two vintages. An order is placed on at the first day, and stochastic demand is fulfilled in FEFO order. Sanders concludes that dynamic pricing could result in higher profits and lower waste.

2.3. Three or more co-existing age classes

Only a few papers allow for more than two co-existing age classes ($m > 2$), which give rise to an m-dimensional inventory state denoting the number of products in stock of each age. This category can be further divided in studies that optimize at most 2 prices, and studies that allow for three or more prices, as we will consider.

Chua et al. (2017) present three models for products with a shelf life of two periods, and one model with a longer shelf life of $m > 2$ periods. The latter model studies for each of the m age classes, the decision whether to set a fixed discount or not. In addition, one decides on an optimal order quantity. All consumers select in LEFO order, starting with the discounted products. The demand volume is insensitive to price and the ages of the products in stock. Numerical solutions are presented only for a shelf life of up to 3 periods. For $m = 4$ age classes, the authors suggest an approximation based on a two-class model.

Kaya and Ghahroodi (2018) present a model for $m \in \{2, 3\}$. They study the simultaneous optimization of order quantity and pricing. Demand is Poisson distributed with a mean that depends on the price and the age of the products. Consumers pick products in LEFO order, and optimal ordering and pricing policies are computed by SDP. When new products are ordered, the old products are removed from stock and sold in a secondary market. The key question is thus when to order a new batch. Because the old products are removed, all products in stock are of the same age class, and the SDP state remains relatively low-dimensional. In addition, they consider the case of two- or three-period shelf life (and thus two or three co-existing age classes).

Buisman et al. (2019) study the impact of optimal discounting of products with a dynamic shelf life and products with a maximum shelf life of 5 to 10 periods. Simulation is applied to mimic product quality decay in a supply chain consisting of a retailer distribution center, and a set of retail outlets. Product quality deteriorates exponentially due to the growth of microbes. Consumer demand is Poisson distributed with weekday-dependent means that do not depend on product age or discount level. Consumer demand is split into FEFO and LEFO demand. LEFO demand is met before meeting FEFO demand. For various scenarios, optimal safety factors and last-day discount levels are computed by enumeration.

None of the studies discussed so far considers more than two co-existing prices, which is a key part of our study. Only three papers consider multiple co-existing selling prices. Chung and Li (2014) model what they call 'need-driven' demand using simulation. Customers require products to have a minimum remaining shelf life, which is a stochastic number sampled for each consumer. In case multiple items in stock fulfill this criterion, consumers select first based on price (lowest first), and if prices are identical, they choose the product with the longest remaining shelf life. Hence, all consumers are LEFO consumers who pick an older item if the price is lower regardless of the size of the discount. The discount rate does also not influence the demand volume. Four pricing policies are defined ranging from no discounting and single discount levels to $m-1$ discount levels. Price levels are set by a fixed logic that distributes the aggregate discount over multiple age classes. Replenishment happens every period by a base stock policy. The authors use simulation to compare the profit and waste under the four pricing policies for different values of the base stock level. Their results show that the pricing policy does not affect which base stock level yields the highest profit.

Adenso-Díaz et al. (2017) study how the depletion of a given initial stock is influenced by dynamic pricing. In their work, they simulate the inventory dynamics over a short time horizon spanning the maximum shelf life m during which no replenishment happens. Demand is modeled by m deterministic demand functions that depend on the price and the product's age. Whether dynamic pricing results in an increase in revenues is shown to depend on the age profile of the products in stock.

In general, the pricing problem for settings with more than two co-existing age classes and prices can still be formulated as a Markov Decision Process (MDP), but hardly no paper applies an exact method like SDP due to the exponential growth in the number of states as the shelf life increases. Yavuz and Kaya (2024) apply SDP only for products with a shelf life of two periods. For products with a maximum shelf life of more than two periods, they find an approximate solution by reinforcement learning. RL is no exact method and the authors then also report an optimality gap of 5% (for the cases with a shelf life of 2 periods). Although reinforcement learning can be effective in optimizing the pricing decisions (even in combination with the ordering decision), the time needed for training can still be very large (a few hours). The decisions space studied by Yavuz and Kaya (2024) is therefore limited to simple parameterized ordering policies like (R, Q).

2.4. Product withdrawal

Many studies on dynamic pricing of food products assume that consumers are rational and select the freshest item if prices are equal. This studies ignore that, in practice, many consumers pick an older product while a fresher item is available at the same price. Hansen et al. (2024) conclude, based on field experiments, that almost half of the consumers choose an older item while an equally priced fresher one is available. Three possible explanations are: (i) consumers are uninformed because of choice frictions (i.e., executing a rational search for the item with the highest utility is too time-consuming), (ii) consumers simply ignore expiration dates, or (iii) consumers choose to select to oldest items to prevent food waste. Uninformed consumers typically take an arbitrary item from the front of the shelf, which is the oldest item as many retailers position the oldest items at the front. Effectively, these consumers pick items in FEFO order even when these are not discounted. The existence of FEFO consumers should be included dynamic pricing models as it strongly affects the inventory dynamics.

2.5. Number of re-stocks and optimization method

Expiration-date-based pricing problem are often studied by SDP, Simulation, mathematical analysis (MA) (calculus), or control theory (Pontryagins's Maximum Principle, PMP). MA and PMP is primarily applied in settings with a single co-existing age class, and a single replenishment at the start of the horizon. In these studies one often determines the optimal length of a replenishment cycle given a demand function that depends on stock age (freshness) and price. Simulation can be applied to settings with any number of co-existing age classes and prices. Simulation is no optimization, as it takes input on which pricing strategies to evaluate and to compare.

SDP can be used to find an optimal pricing strategy. In settings of a monopolistic retailer, one needs to align the prices with the ordering quantity, as the price strongly affects the demand. We focus on a setting of a non-monopolistic retailer, e.g. supermarket, who considers the regular selling price and the order quantity to be given. His focus is on the operational in-store decision on how to set discount rates for items that are about to expire. The impact of discounts on demand is then limited, as it applies only as long as discounted products are available, and discounts are not published but only visible to in-store customers. In infinite horizon SDP, multiple replenishment moments are modeled. In finite horizon SDP, one often limits to a single replenishment cycle.

SDP methods are less applicable when m gets large due to the large number of possible values of the state vector. Also, the number of action values needs to be limited to maintain tractable models. RL is a flexible methods to approximately solve a MDP that allows for multiple restocks, and more than 2 co-existing age classes, and prices. Training time of a RL model maybe very time consuming but RL scales well.

2.6. Research gap and contribution

New technologies allow retailers to differentiate prices for items of different expiration dates. However, determining optimal prices is an understudied research area, especially for the context of a non-monopolistic retailer that sells to both FEFO and LEFO customers. For this practical context, we optimize two discount rates by SDP and we uniquely derive optimal discounting rates for products with a shelf life of up to 5 periods (e.g. days). As replenishment happens every period, one could thus have items in stock of 5 different expiration dates. Our results can serve as new benchmarks for future research that apply approximation methods, like reinforcement learning.

The results provide managerial insights in the added value of applying two discount rates over using just one. In previous studies that consider two (or more) discount rates, the discount rates are not subject to optimization but instead are fixed (e.g. by some formula that depends on the product age). This paper thus leverages insights in the value of differentiated discounting. It gives insights on how to set discount rates that both reduce food waste and maximize profit. These insights are particularly valuable, as many retailers still struggle to determine how deep their discounts should be, and whether to offer earlier but smaller discounts.

In the next section, we will describe our modeling approach in detail.

3. Modeling expiration-date-based discounting

The key question we aim to answer is: What is the impact of different discounting policies on the profit, waste, and fill rate of the retailer? In particular, we are interested in the difference between last-day and last-two-days discounting, as well as how deep the discount(s) should be (i.e., the size of the discount). These questions are answered by formulating and solving the problem as a Markov decision process (MDP) that optimizes the discounting levels of the last two age classes. By restricting the action space, the model is also used to derive an optimal dynamic last-day discounting policy. By further restricting the action space to a single discount value, the models boils down to a Markov chain to evaluate a fixed-discount policy and the no-discounting policy.

A description of the problem, including a graphical representation of the decisions over time and the notation we use, is given in Section 3.1. In Section 3.2, the mathematical model is presented in detail. When formulating the model, we use general terminology for time periods, which in practice would be days, as most fresh products can be ordered every day in a retail context. This also implies that modeling of shelf life in days, and that discounting decisions are made once per day. Other resolutions can also be dealt with in our model (for instance, by assuming that a period is half a day), but in the numerical analysis presented in Section 5, we work with days.

3.1. Problem formulation and notations

We consider a profit-maximizing grocery retailer (e.g., supermarket) that sells a perishable food product with a fixed shelf life of m periods. Items whose expiration date has passed can no longer be sold and are disposed of at a unit disposal cost (w). New items are ordered at the start of every period (i.e., the review period $R = 1$) using a base stock policy and delivered at the end of the period (i.e., lead time $L = 1$). The (constant) base stock level B is a strategic decision aiming at a sufficiently high product availability such that consumers will not switch to a competitor. We model a single non-monopolistic retailer and competing retailers are not modeled. An item is purchased at cost c, and sold at its regular unit selling price r if no discounts applies. Note, as the retailer is not a monopolist, the regular selling price r is an exogenous parameter.

At the start of the period, the retailer can decide on two discount levels: x_1 for items that expire at the end of the period (last-day discounting), and x_2 for items with a remaining shelf life of two days, which will expire the next period (next-to-last day discounting).

Regular demand d is uncertain and modeled by a stationary discrete probability distribution $p(d)$ with mean μ and standard deviation σ. In our model, we split this demand (using stochastic rounding) between customer groups with different picking behavior and we generate additional demand in case there are discounted products to also incorporate possible price-sensitive customers.

Regarding picking behavior, some consumers prefer items with the longest expiration date over items with a shorter expiration date: these consumers pick items from the shelf in order of Last-Expired-First-Out (LEFO). Other consumers are insensitive to this and pick the items presented in front of the shelf, in the order First-Expired-First-Out (FEFO). Whenever discounted products are available, some price-sensitive LEFO consumers might pick a discounted product instead of picking the last-expired items first. Thus, we need to distinguish between three types of consumers in our modeling: LEFO consumers who are price sensitive and pick discounted products whenever available, LEFO consumers who are not price sensitive, and FEFO consumers.

We add a fourth consumer type to model the possible extra demand that is generated by the price discount. This extra demand depends on the size of the discount and is limited by the number of discounted items. In this way, our model still considers the regular (exogenous) demand for the product at full price, but also considers price sensitivity

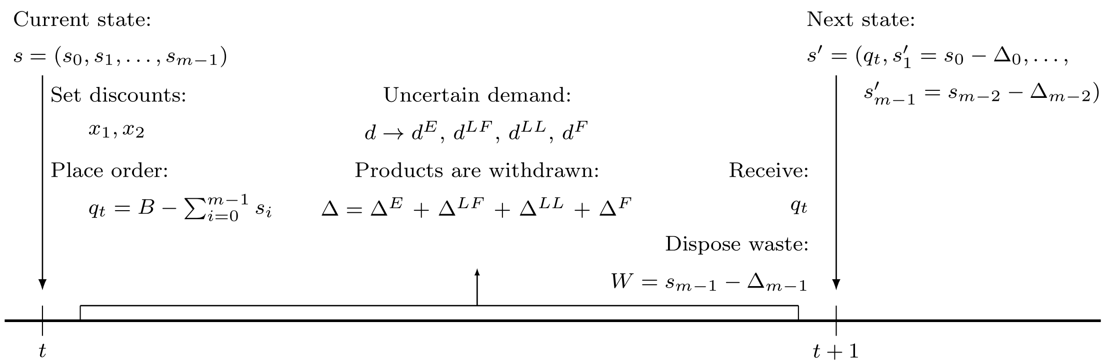

Fig. 2. Model diagram with states, actions, and events.

related to the discounted products. The possible extra demand is scaled with the regular demand for the product, but may also originate from consumers who intended to buy another (substitute) product. As discounts are made available at the start of the period, the extra demand is happening at the start of the period while discounted items are still available. For easy notation, the extra demand is modeled and met before the regular demand for d items is met.

3.2. MDP model

An MDP model is defined by the state space, action space, the events and state transitions, the rewards, and Bellman's optimality equation. These concepts will be defined in the remainder of this section. In Fig. 2, a timeline with the states, decisions, and events is given to illustrate the main structure of the model.

State and state space

At the start of every period, the retailer inspects the inventory of each of the m age classes. Typically, such information may be available in enterprise resource planning systems. The number of items in stock of age i is denoted by s_i. Thus, s_0 is the number of 'new' products in stock, and s_{m-1} is the number of the oldest items in stock, which will expire at the end of the period. The vector $s = (s_0, s_1, \ldots, s_{m-1})$ represents the state of the system at the start of a period. Let $\mathcal{S}$ denote the state space, which consists of all possible values of vector s. As we study a context in which discounted products are displayed on the same shelf as the not-discounted products, the state is represented as a single stock vector. Thereby we implicitly assume that these items are offered on the same shelf and are to some extent substitutes for customers (e.g. the LEFO customers that might be tempted to become FEFO customers if those items are discounted). In a different context with discounted products sold at a different locations, one may model the inventories separately by two vectors, and in that case one could also add a demand stream for customers that are solely interested in discounted products.

Action and action space

At the start of a period, the discounting decisions $(x_1, x_2) \in \mathcal{X}$ are made, where $\mathcal{X}$ is the action space. We restrict the action space by setting $x_1 \geq x_2$, i.e., fresher items do not get a higher discount than the oldest items in stock. Furthermore, a replenishment order is placed of size:

$$q_t = B - \sum_{i=0}^{m-1} s_i. \tag{1}$$

These q_t items will be delivered at the end of the period to acknowledge a lead time of one period. Note that the replenishment order follows a typical base stock policy, which is not subject to optimization. This keeps the action space limited, but it is also a common approach in a context in which replenishment decisions are driven by a target service level.

Stochastic state transitions and rewards

The state transitions and rewards are stochastic due to the uncertainty in the demand. The regular demand is $d \in \mathcal{D}$ with probability $P(d)$. $\mathcal{D}$ is a discrete finite set of possible demand outcomes: $\mathcal{D} = \{0, 1, 2, \ldots, D\}$. The availability of discounted items of age i generates extra demand d_i^E:

$$d_i^E = \lceil \delta \cdot x_i \cdot d \rfloor,$$

where $\lceil y \rfloor$ is a shortcut notation for the stochastic rounding of y:

$$\lceil y \rfloor = \begin{cases} \lceil y \rceil & \text{with probability } y - \lfloor y \rfloor \\ \lfloor y \rfloor & \text{else.} \end{cases}$$

Parameter δ indicates the price elasticity of demand. We assume that the extra demand occurs at the start of the day (while discounted items are still available), and that it is met before the regular demand. The number of items picked to meet the extra demand, Δ_i^E, cannot exceed s_i:

$$\Delta_i^E = \min\{d_i^E, s_i\}$$

The regular demand level d is split into three demands, one for each of the three types of customers that we introduced earlier: demand by FEFO customers (d^F), demand by price-sensitive LEFO customers (d^{LF}), and the other LEFO customers (d^{LL}).

The FEFO demand is for $d^F = \lceil f \cdot d \rfloor$ items, and is obtained by stochastic rounding of $f \cdot d$. Here, f is the percentage of FEFO customers. The LEFO demand, $d - d^F$, is split into price-sensitive customers d^{LF} who are willing to pick a discounted product, and the rest. As we model two discount levels, x_1 for items of age $m-1$ and x_2 for items of age $m-2$, we define d_{m-i}^{LF} for the number of LEFO consumers that is willing to pick a product of age $m-i$:

$$d_{m-i}^{LF} = \begin{cases} \lceil \gamma \cdot x_i \cdot (d - d^F) \rfloor & \text{if } i \in \{1, 2\} \\ 0 & \text{otherwise.} \end{cases}$$

Here, γ is used to model the sensitivity of the customer to the discount level. Again, the quantity picked cannot exceed the number items left of age $i \in \{0, 1, \ldots, m-1\}$:

$$\Delta_{m-i}^{LF} = \min\{d_i^{LF}, s_i - \Delta_i^E\}.$$

The remaining LEFO demand is then $d^{LL} = d - d^F - d_{m-1}^{LF} - d_{m-2}^{LF}$, leading to the following quantity picked:

$$\Delta_i^{LL} = \begin{cases} \min\{\ d^{LL},\ s_i\ \} & \text{if } i = 0 \ \&\ m > 2 \\ \min\{\ d^{LL} - \sum_{j=0}^{i-1} \Delta_j^{LL},\ s_i - \Delta_i^E - \Delta_i^{LF}\ \} & \text{else.} \end{cases}$$

In the model, we thus sequence the demand as follows: the extra demand d^E, the price-sensitive LEFO consumer (d^{LF}), the remaining LEFO consumers, and finally the FEFO consumers. The total number of items picked from age i is:

$$\Delta_i = \Delta_i^E + \Delta_i^{LF} + \Delta_i^{LL} + \Delta_i^F. \tag{2}$$

The state transition from state s in period t to state s' at period $t+1$, when demands in period t are $(d^E, d^{LF}, d^{LL}, d^F)$, is

$$s'_i = \begin{cases} q & \text{if } i = 0 \\ s_{i-1} - \Delta_{i-1} & \text{else} \end{cases} \quad (3)$$

and the resulting waste in that period is

$$W = s_{m-1} - \Delta_{m-1}. \quad (4)$$

Rewards

The reward in period t is the profit that is earned, which depends on the reference price (r), purchasing price (c), disposal cost (w), the discounts (x_1, x_2), the order quantity q, and the number of sold items Δ_i for each age i. The reward when demand is $(d^E, d^{LF}, d^{LL}, d^F)$ is:

$$R(s, x \mid d^E, d^{LF}, d^{LL}, d^F) = r \cdot \sum_{i=0}^{m-1} \Delta_i - x_1 \cdot r \cdot \Delta_{m-1} - x_2 \cdot r \cdot \Delta_{m-2} - c \cdot q - w \cdot W \quad (5)$$

The (unconditional) expected rewards $r(s, x)$ and state transition probabilities $p(s, s' \mid x)$ can be derived by iterating over all possible values of the regular demand $d \in \mathcal{D}$ and enumerating all possible related demand vectors $(d^E, d^{LF}, d^{LL}, d^F)$.

Bellman's optimality equation and stochastic dynamic programming

Let $V_n(s)$ be the expected reward over n periods when starting in state s and optimal (state-dependent) actions (x_1, x_2) are taken in each period. By this definition $V_0(s) = 0$, for all states $s \in \mathcal{S}$, and V_n can be defined recursively:

$$\forall s \in S: \quad V_n(s) = \max_{x \in X(s)} \left\{ r(s, x) + \sum_{s' \in S} p(s, s' \mid x) \cdot V_{n-1}(s') \right\} \quad (6)$$

An optimal policy π^* is set by

$$\forall s \in S: \quad \pi^*(s) = \lim_{n \to \infty} \arg\max_{x \in X(s)} \left\{ r(s, x) + \sum_{s' \in S} p(s, s' \mid x) \cdot V_n(s') \right\} \quad (7)$$

The resulting infinite horizon MDP is solved by Stochastic Dynamic Programming (using value iteration), see Puterman (2014). The difference $g_n = V_n - V_{n-1}$ converges to a constant vector $\lim_{n\to\infty} g_n$, with each element equal to the optimal expected profit per period g. The iterative procedure of evaluating Eq. (6) for increasing values of n stops as soon as $\max g_n - \min g_n$ gets smaller than a pre-specified small value ϵ. For more details, we refer to Puterman (2014).

4. Design of experiments

In this section, we present the design of 17 experiments by which we numerically compare the performance of four discounting policies. The result are presented in Section 5.

4.1. Parameter settings

The experiments relate to the context of a supermarket selling a fresh product with a short shelf life of 3–5 days, such as packed fresh-cut vegetables, meat products, and ready-to-eat meals. Replenishment orders are placed daily by a base stock policy with base stock level $B = \mu \cdot (R+L) + z \cdot \sigma \cdot \sqrt{R+L}$, where z is the safety factor. Note that the base stock level B does not account for any additional demand induced by the discounting strategy, as doing so would undermine the purpose of discounting – to sell surplus inventory – and could lead to a self-reinforcing cycle of continued discounting. To test, for the robustness of the discounting policies, we evaluate the discounts rates also for lower and higher values of B, see Section 5.6.

As demand is for a (small) discrete number of items (and depending on the number of customers visiting the store), in many similar studies demand is assumed to be Poisson distributed. Hence $\sigma = \sqrt{\mu}$. This corresponds to a reasonable degree of demand uncertainty, especially for lower levels of demand. The Poisson distribution excludes negative values and resembles a right-skewed distribution (especially, for lower values of μ), which often fits well with retail demands.

Table 2
Fixed values of parameters.

Parameter	Notation	Value
Review period (days)	R	1
Lead time (days)	L	1
Reference price (euro)	r	2.50
Purchase price (euro)	c	1.75
Disposal cost (euro)	w	0.10

Table 3
Experimental values of parameters.

Parameter	Notation	Experimental values
Shelf life (days)	m	3, **4**, 5
Mean consumer demand (days)	μ	2, **4**, 6
Fraction FEFO consumer	f	0, 0.25, **0.5**, 0.75, 1.0
Discount sensitivity	γ	0, **1.0**, 1.75, 2.5
Price elasticity of demand	δ	0, 0.3, **0.55**, 0.8
Safety factor	z	1.0, **1.5**, 2.0

We set the economic parameters to the values reported in Table 2. The (regular) profit margin is fixed at 30% of the regular selling price. The waste costs relate to the cost of disposing of an item, which may include handling costs as well as the economic costs of emptying a waste container.

The study comprises a base case and sixteen deviations from it by sequentially varying the shelf life (m), the mean demand (μ), the price elasticity of demand (δ), the fraction of FEFO consumers (f), the discount sensitivity of LEFO consumers γ, and the safety factor z. These values are deterministic in our model, but reflect the context of the specific retailer or retail store studied (e.g., stores in different kind of neighborhoods would have different price sensitivities with regards to discounts and different fractions of LEFO and FEFO customers.)

Table 3 summarizes the value of the experimental factors, with the base case values highlighted in bold.

The shelf life parameter plays a crucial role in inventory management and directly influences waste and shortages. In the context of this study, the product shelf life is set to 3, 4, or 5 days. The choice of a 4-day shelf life for the base case is supported by van Donselaar et al. (2006), wherein 4 days is identified as the median for the shelf life of days fresh products in a single supermarket. According to the same study, the average weekly sales of fresh products in supermarkets is 33.8 products, which implies an average of 4.8 products sold per day. Therefore, in this study we consider three mean demand levels $\mu = \{2, 4, 6\}$. For the base case, $\mu = 4$ is used. In all experiments demand is modeled by a Poisson distribution with mean μ.

The actual fraction of FEFO consumers (f) depends on the type of product and differs per store. Therefore, this study will explore a wide range of values: $f \in \{0, 0.25, 0.5, 0.75, 1.0\}$. In the base case, we assume an equal division of LEFO and FEFO consumers: $f = 0.5$. The wide range of value of f is supported by literature. Studies that apply mathematical analysis, assume either pure LEFO $f = 0$ or pure FEFO $f = 1$. In contrast many numerical studies, consider a mix of FEFO and LEFO. Santos et al. (2022) suggests $f = 0.2$, whereas Hübner et al. (2024) estimates that, if items on the shelves are sorted by expiration date with oldest items displayed at the front, f is in the range 0.65 to 0.75. Other studies assume $f = 0.45$ (Tromp et al., 2012) and $f = 0.6$ (Buisman et al., 2019).

The discount sensitivity of LEFO customers (γ) refers to the degree to which customers react to price reductions by transitioning from LEFO to FEFO behavior. A discount of $x\%$ makes $\gamma \cdot x\%$ of the LEFO consumers willing to pick in FEFO order. In line with Buisman et al. (2019), for the base case we set $\gamma = 1$. When $\gamma = 0$, no LEFO consumers will prefer a discounted item over a fresher item, while with $\gamma = 2.5$, a 40% discount will trigger all LEFO customers to buy a discounted product (whenever available).

The price elasticity of demand (δ) models the extra demand generated by price discounts. In the base case, we set $\delta = 0.55$, which implies that for every one percent reduction in price, the demand increases by 0.55%. A study conducted by Andreyeva et al. (2010) establishes the price elasticity of food within the range of 0.3 to 0.8, indicating an inherent inelasticity in the demand for perishable products, meaning that price variations only have a moderate impact on demand levels. Additionally, to show the effects of perfect inelasticity, $\delta = 0$ is added to the experiments. This implies that demand is not affected by price changes. To study these effects, we set $\delta \in \{0, 0.3, 0.55, 0.8\}$.

The safety factor (z) determines the height of the base stock level $B = \mu \cdot (R+L) + z \cdot \sqrt{\mu \cdot (R+L)}$. A high value of B is needed to buffer for demand fluctuations, and thus to ensures a high on-shelf-availability of products. For the base case we set $z = 1.5$, then about 94% of the demand can be met from stock. With a higher safety factor, the supermarket buffers for high realizations of demand volumes to ensure a high fill rate. A more high-end supermarket has a higher safety factor ($z = 2.0$), compared to a low-end supermarket ($z = 1.0$). As demand is Poisson distributed and a base stock policy is followed with target stock level $B = \mu \cdot (R+L) + z \cdot \sqrt{\mu \cdot (R+L)}$, we expect the fill rate to be between 89% and 98%. The value of the base stock level in the base case is thus 12, and varies from 11 (if $z = 1$) to 14 (if $z = 2$).

4.2. Four discounting policies

For each experiment, we compare the performance of four discounting policies. We apply a grid of discount levels ranging from 0% to 40% with a grid size of 5%: $\mathcal{G} = \{0, 0.05, 0.1, \dots, 0.4\}$. To optimize and evaluate the expected profit for each discounting policy, we limit the action space $\mathcal{X}$ as follows:

- D2 = dynamic last-two-days discounting: $\mathcal{X} = \{(x_1, x_2) \mid x_1 \geq x_2 \wedge x_1, x_2 \in \mathcal{G}\}$
- D2S = dynamic last-two-days discounting at same rate: $\mathcal{X} = \{(x_1, x_2) \mid x_1 = x_2 \wedge x_1, x_2 \in \mathcal{G}\}$
- D1 = dynamic last-day discounting: $\mathcal{X} = \{(x_1, 0) \mid x_1 \in \mathcal{G}\}$ as $x_2 \equiv 0$,
- NO = no-discounting policy: $\mathcal{X} = \{(0, 0)\}$.
- FO = fixed optimal last-day discount policy: $\mathcal{X} = \{(x_1^*, 0)\}$. The optimal (profit-maximizing) static discount level $x_1^* \in \mathcal{G}$, is determined by evaluating the expected profit for each $x_1 \in \mathcal{G}$.

4.3. Computational accuracy and running times

To numerically solve the MDP models by value iteration, we truncate the discrete demand distribution at the smallest level of D for which:

$$\sum_{d=D}^{\infty} \frac{\mu^d}{d!} \cdot e^{-\mu} \leq 0.001.$$

Hence, the finite discrete demand distribution $P(d)$ is

$$P(d) = \begin{cases} \frac{\mu^d}{d!} & \text{if } d \in \{0, 1, \dots, D-1\} \\ 1 - \sum_{d=0}^{D-1} P(d) & \text{if } d = D \\ 0 & \text{otherwise.} \end{cases}$$

For $\mu = 2, 4$, and 6 this implies D is set to 8, 12, and 15 respectively.

The MDP model is coded in Python, and solved using MDPtoolbox (Chades et al., 2014). For all policies we stop value iteration by setting $\epsilon = 0.001$, hence the expected profit is accurate up to 0.001. The running time depends on the number of states, the number of actions, and the number of demand levels that may occur per day.

To evaluate the waste and fill rate of all discounting policies that are derived through value iteration on the described MDP, in the next section, these policies are simulated. Although the truncation of the demand distribution is primarily needed for solving the MDP, we use the same truncated distribution in the simulation. The simulation lasts 100,000 periods, which results in an accuracy similar to MDP: average profit levels are at most 0.001 off from the nearly-exact values reported by value iteration.

Table 4
Average profit, sales, fill rate, and waste for each discounting policy.

	NO	FO	D1	D2S	D2
Profit (€)	2.50	2.54	2.56	2.57	2.58
# sold	3.88	3.94	3.95	3.96	3.97
% fill rate	97.27%	97.25%	97.25%	97.14%	97.15%
% waste	5.61%	4.23%	3.98%	3.74%	3.60%

5. Results

In this section we present a sensitivity study based on the results for the design of experiments. In addition, we discuss the robustness of the discounting policies when applied to setting with higher a longer review period and/or lead time, and when the base stock level deviates.

5.1. Overview: average profit, sales, fill rate, and waste

Table 4 reports the performance of the four discounting policies averaged over all 17 experiments. Under the no-discounting policy (NO), sales are on average 3.88 per period, which implies a fill rate of $\frac{3.88}{4} \cdot 100\% = 97\%$, and an average waste of 5.61%. The fill rate is evaluated in relation to the regular demand and excludes the extra demand, as the extra demand comes from non-planned purchases by consumers. FO reduces the waste to 4.23%, whereas the dynamic policies D1 and D2 reduce the waste to 3.98% and 3.6%, respectively. With waste at 3.74%, policy D2S performs in between policies D1 and D2, as might be expected. For situations in which two co-existing discounting would be difficult to implement, this means that the D2S strategy would lead to about half the benefit of moving from last-day discounting to last-two-days discounting (assuming optimal discounting percentages).

In general, discounting results in more items sold, but the fill rate remains virtually the same. Apparently, the extra sold items are primarily realized by the extra demand, which in turn explains a great part of the profit increase.

5.2. Sensitivity analysis: profit and waste per experiment

Table 5 shows the relative waste and profit for each experiment. The profits of FO, D1, D2S, and D2 are the percentage increases over the no-discounting policy (NO). In the top row, we read the averages: FO yields a 2.07% higher profit, and D1 and D2 improve by 2.88% and 3.83%, respectively. For D2S, the average improvement is 3.46%. Besides the increase in average profit, we observe a significant reduction in waste: NO results in 5.6% of the items turning into waste, whereas waste is only 3.6% for D2.

The largest increase in profit and reduction in waste is achieved in settings where the average waste is high, and/or where consumers are very price sensitive, like in experiment 2, 3, 7, 9, 10, 13, and 16. In Experiments 2 and 3 waste tends to be high in case of no discounting due to a low fraction of consumers accepting the oldest products ($f = 0$ or 0.25). In experiment 7 the waste under no discounting is high due to a high safety stock ($z = 2$). In experiments 9 and 10, the profit increase and reduction in waste is high as LEFO consumers are more sensitive to discounts ($\gamma = 1.75$, or 2.5). Similarly, in experiment 13, the profit is higher due to a higher price elasticity ($\delta = 0.8$), which generates more demand. In Experiment 16 the shelf life is shorter ($m = 3$), causing much more waste if no discounts would be offered.

Table 5
Overview of daily profit and waste for each experiment and discounting policy.

Experiment		Profit	% Profit gain over NO				Waste				
		NO	FO	D1	D2S	D2	NO	FO	D1	D2S	D2
Average		€2.50	2.07%	2.88%	3.46%	3.83%	5.6%	4.2%	4.0%	3.7%	3.6%
1	Base case	€2.59	0.13%	1.14%	1.28%	1.60%	4.4%	3.9%	3.3%	3.2%	3.0%
2	$f = 0$	€1.74	19.70%	21.02%	26.21%	27.79%	13.8%	6.9%	6.8%	5.1%	4.8%
3	$f = 0.25$	€2.15	7.86%	8.52%	9.87%	10.54%	9.6%	5.0%	5.0%	4.4%	4.1%
4	$f = 0.75$	€2.82	0.00%	0.06%	0.02%	0.07%	1.4%	1.4%	1.3%	1.4%	1.3%
5	$f = 1$	€2.88	0.00%	0.00%	0.00%	0.00%	0.6%	0.6%	0.6%	0.6%	0.6%
6	$z = 1$	€2.66	0.00%	0.50%	0.50%	0.67%	2.8%	2.8%	2.2%	2.2%	2.1%
7	$z = 2$	€2.26	1.95%	3.87%	5.04%	5.78%	8.9%	6.2%	5.9%	5.5%	5.2%
8	$\gamma = 0$	€2.59	0.00%	0.45%	0.42%	0.73%	4.4%	4.4%	4.0%	3.9%	3.7%
9	$\gamma = 1.75$	€2.59	0.83%	1.91%	2.08%	2.42%	4.4%	2.7%	2.8%	2.7%	2.6%
10	$\gamma = 2.5$	€2.59	1.72%	2.65%	2.87%	3.21%	4.4%	2.4%	2.5%	2.3%	2.2%
11	$\delta = 0$	€2.59	0.00%	0.01%	0.00%	0.01%	4.4%	4.4%	4.4%	4.4%	4.4%
12	$\delta = 0.3$	€2.59	0.00%	0.24%	0.30%	0.43%	4.4%	4.4%	4.1%	4.0%	3.9%
13	$\delta = 0.8$	€2.59	1.69%	2.57%	2.89%	3.32%	4.4%	2.3%	2.6%	2.4%	2.3%
14	$\mu = 2$	€0.90	0.84%	2.01%	2.92%	3.06%	13.2%	11.9%	11.6%	11.2%	11.1%
15	$\mu = 6$	€4.17	0.15%	0.92%	0.90%	1.24%	2.1%	1.8%	1.4%	1.3%	1.2%
16	$m = 3$	€2.08	0.37%	2.69%	2.93%	3.79%	10.3%	9.2%	8.0%	8.0%	7.5%
17	$m = 5$	€2.80	0.00%	0.37%	0.51%	0.51%	1.6%	1.6%	1.2%	1.2%	1.1%

Experiment 6 and 7 reveal that a higher profit level can be achieved by lowering the base stock level, but it results in a (much) lower product availability. By enumerating a range of values for z, one could determine a base stock level B that maximizes the profit. However, in practice, consumers who get disappointed too often will choose to not return and to continue their grocery shopping at a competitor. Therefore, in this study the safety stock is a fixed strategic decision aiming at a sufficiently high product availability. If it would be open for optimization, one may consider to add the order quantity as an extra dimension to the action space of the MDP model. Note that the impact of discounting on the demand is small and only applies as long as discounted items are available.

The biggest profit gain (27.79%) is achieved by D2 when all consumers pick the freshest items first ($f = 0$). In that setting, D2 reduces the average waste from 13.8% to 4.8%. Comparing D2S and D2, we see that the additional flexibility in using two different discount percentages provides a significant profit gain and waste reduction. Including D1 in this comparison, we also see that in some cases, D2S performs worse than D1, implying that in those cases, discounting the items from the two last age classes leads to too many items being discounted. This is especially clear in experiments 4, 8, and 15, i.e., when there are many FEFO customers, when there is no discount sensitivity, and when there is higher demand. In those cases, it is more likely that older products are sold without discounting and also discounting the items in the next-to-last age class would not be beneficial.

Note that for $m = 3$ FO hardly improves over NO, whereas D1, D2S, and D2 yield an increases in profit between 2.69% and 3.79%. We observe in 7 out of 17 experiments (namely experiments 4, 5, 6, 8, 11, 12, and 17) that FO does not improve NO in terms of profit or waste. This means that the optimal fixed discount level x_1^* is 0%. In other words, applying a discount would result in a lower profit. In most of these experiments, dynamic discounting has a slightly positive effect on the average profit and waste.

Most notably, experiment 8 shows the results when LEFO consumers are not responsive to discounts at all ($\gamma = 0$). Then the discount is installed only to attract some extra demand, which results in an increase in the average profit by 0.45% (D1) and by 0.73% (D2). As mentioned above, in this case, D2S would then lead to more items being discounted than necessary to achieve this effect.

When $\delta = 0$, the price elasticity is zero, and thus the extra demand is zero, and any effect of discounting is then explained by LEFO consumers who accept an older discounted item over new items. Also, in the case of a longer shelf life ($m = 5$), FO has no positive effect on profit and waste, whereas dynamic discounting has a positive impact on both profit and waste.

Discounting (by FO, D1, D2S, or D2) has only a minimal effect on profit and waste when many consumers pick the oldest first ($f = 0.75$, $f = 1$), or when the price elasticity is zero ($\delta = 0$). When $f = 0.75$ and $f = 1$, many customers accept the oldest items even when they are not discounted. Providing a discount to FEFO consumers is pointless, as it results in a loss of profit. This loss in profit margin is not compensated for by extra sales or the very few, if any, LEFO customers who opt for the discounted products.

New information technology in the form of digital price tags allows for cost-effectively setting multiple expiration-date-based prices. It is therefore interesting to analyze in which cases last-two-days discounting (D2) is much better than last-day discounting (FO or D1). This question is in many research papers unanswered, as their focus is on optimizing one price or at most two prices. Fig. 3 visualizes the profit gain of FO, D1, and D2 compared to NO discounting (we did not include D2S in this figure as it would mostly lie between D1 and D2 following the results in Table 5). It clearly demonstrates that last-two-days discounting (D2) yields consistently more profit than last-day discounting (FO and D1). The difference is higher in settings with a higher price elasticity δ, or when waste is relatively high, i.e., when more consumers pick in LEFO order, when the safety stock is high, and when shelf life is short.

5.3. Impact of discounting on sales

he profit increase from D2 is largely attributable to reduced waste and higher sales, as long as discounted items remain available. Fig. 4 shows for FO, D1, and D2 the increase in sales compared to no discounting (NO). Clearly, D2 generates more sales than D1 and FO. The extra sales are in most experiments around 2%. Much higher figures relate to experiments in which high discounting rates are optimal: i.e., in cases of a high average waste (i.e., low f, high z, low M) or when the price elasticity δ is high.

Extra sales do not always imply a high profit. For example, the extra sales generated by FO is in some experiments ($\gamma = 1.75$ or 2.5, $\delta = 0.8$) more than that of D1, while levels of D1 are, by definition, greater than or equal to those of FO.

5.4. Size of discounts and frequencies

To inspect the average discount rate set by the three policies FO, D1, and D2, we calculate the weighted average last-day discount (x_1) per item in stock. Table 6 reports the results.

Averaged over all experiments, we observe that the average discount rate of FO is 8.5%. For D1, this is 9.4%. For D2, the last-day

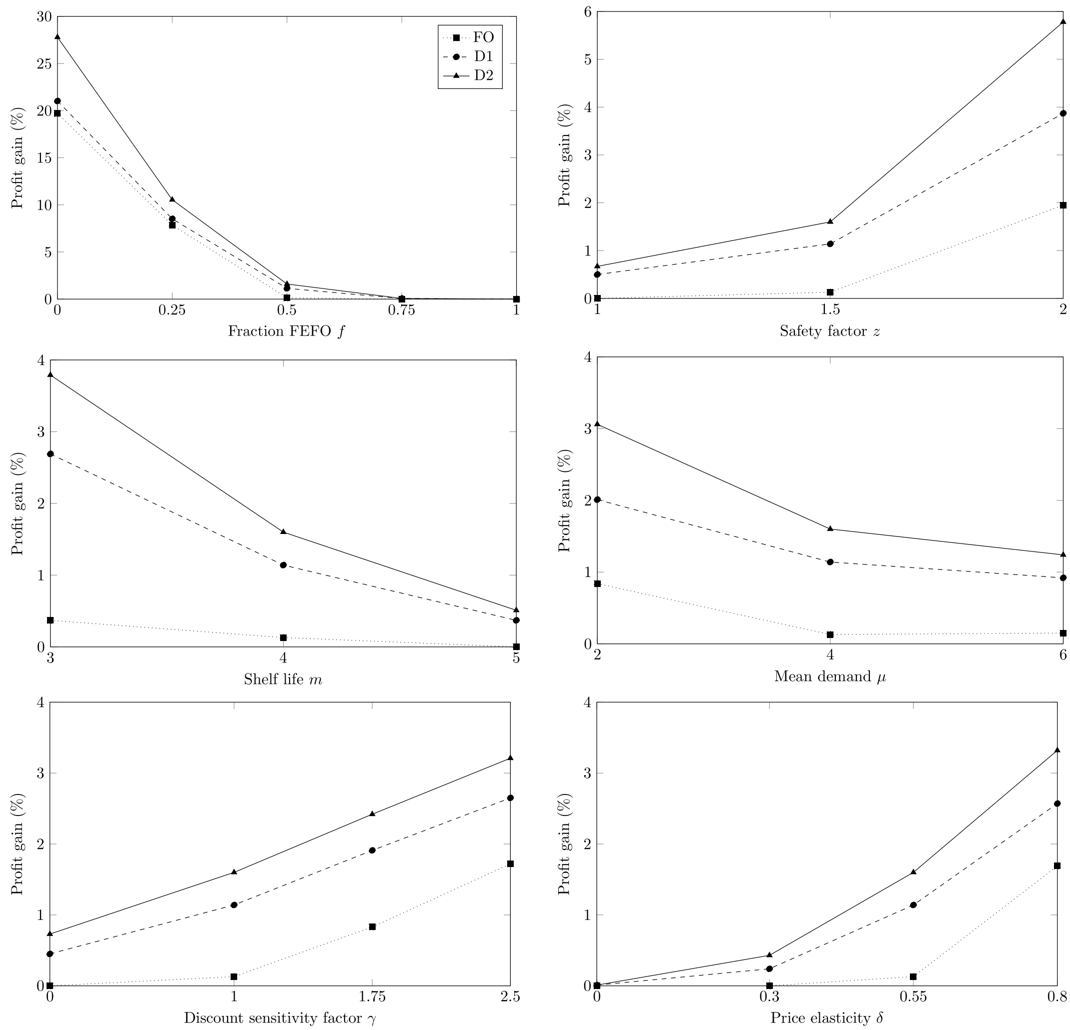

Fig. 3. Effect of parameters f, z, m, μ, γ, and δ on profit gain over no discounting.

discount rate (x_1) falls in between that of FO and D1. The next-to-last day discount rate (x_2) of D2 is significantly lower. When comparing with the results in Table 5, we observe higher discounts in experiments with high waste. The high discount rates of FO (25% or more) apply to experiments with no or few FEFO consumers ($f = 0$, or 0.25). A medium discount rate (15%–20%) is observed in four experiments, where waste figures are high as a result of a high safety stock, or a high price sensitivity (i.e., high value of δ and γ). Overall, the average discount rates of all policies are higher.

Next, we zoom in on how many products in the base case receive a last-day discount and how often each discount level is selected. From Fig. 5 we read that under D1, 15% of the items ordered are still unsold on the last day before expiration. About 9% get a discount of 5% or more, and 6.5% get a discount of 10% or more. For D2, fewer items end up at the last age class (13.6% vs. 15%), and fewer products get a discount of 5% or more. Remarkably, the same fraction of products get a discount of 10% or more.

Fig. 6 shows for the items that are still unsold on the day of expiration, how often each discounting percentage applies. 38%–40% of these oldest items are not discounted. For items that are discounted, the discount rate is most often in the range 5%–15%.

These results trigger us to investigate in which states what discounting rates apply, and whether the dynamic policies reveal a clear structure. We focus on the optimal policy D1 for the base case. The optimal discounting policy turns out to be rather complicated and not easy to capture in a simple rule, so a heuristic dynamic policy, although outside the scope of this paper, can be interesting to explore.

In Fig. 7, we present a heat map of policy D1 for the base case. The table shows for each discount level x_1, the weighted average number

Table 6
Overview of the average discount rate for each experiment and discounting policy.

Experiment		FO	D1	D2 x_1	D2 x_2
Average		8.5%	9.4%	8.9%	2.8%
1	Base case	5%	8.6%	8.7%	2.1%
2	$f = 0$	30%	28.4%	23.1%	12.3%
3	$f = 0.25$	25%	23.5%	19.4%	7.0%
4	$f = 0.75$	0%	1.4%	1.5%	0.1%
5	$f = 1$	0%	0.1%	0.1%	0.0%
6	$z = 1$	0%	5.7%	5.8%	1.2%
7	$z = 2$	15%	14.3%	13.9%	4.8%
8	$\gamma = 0$	0%	4.7%	6.6%	1.6%
9	$\gamma = 1.75$	15%	9.8%	9.2%	2.2%
10	$\gamma = 2.5$	15%	10.0%	9.5%	2.3%
11	$\delta = 0$	0%	0.3%	0.3%	0.0%
12	$\delta = 0.3$	0%	3.0%	3.7%	0.8%
13	$\delta = 0.8$	20%	14.1%	13.8%	4.0%
14	$\mu = 2$	10%	10.3%	9.9%	3.7%
15	$\mu = 6$	5%	8.6%	10.1%	2.2%
16	$m = 3$	5%	9.4%	9.2%	2.7%
17	$m = 5$	0%	6.9%	7.1%	1.5%

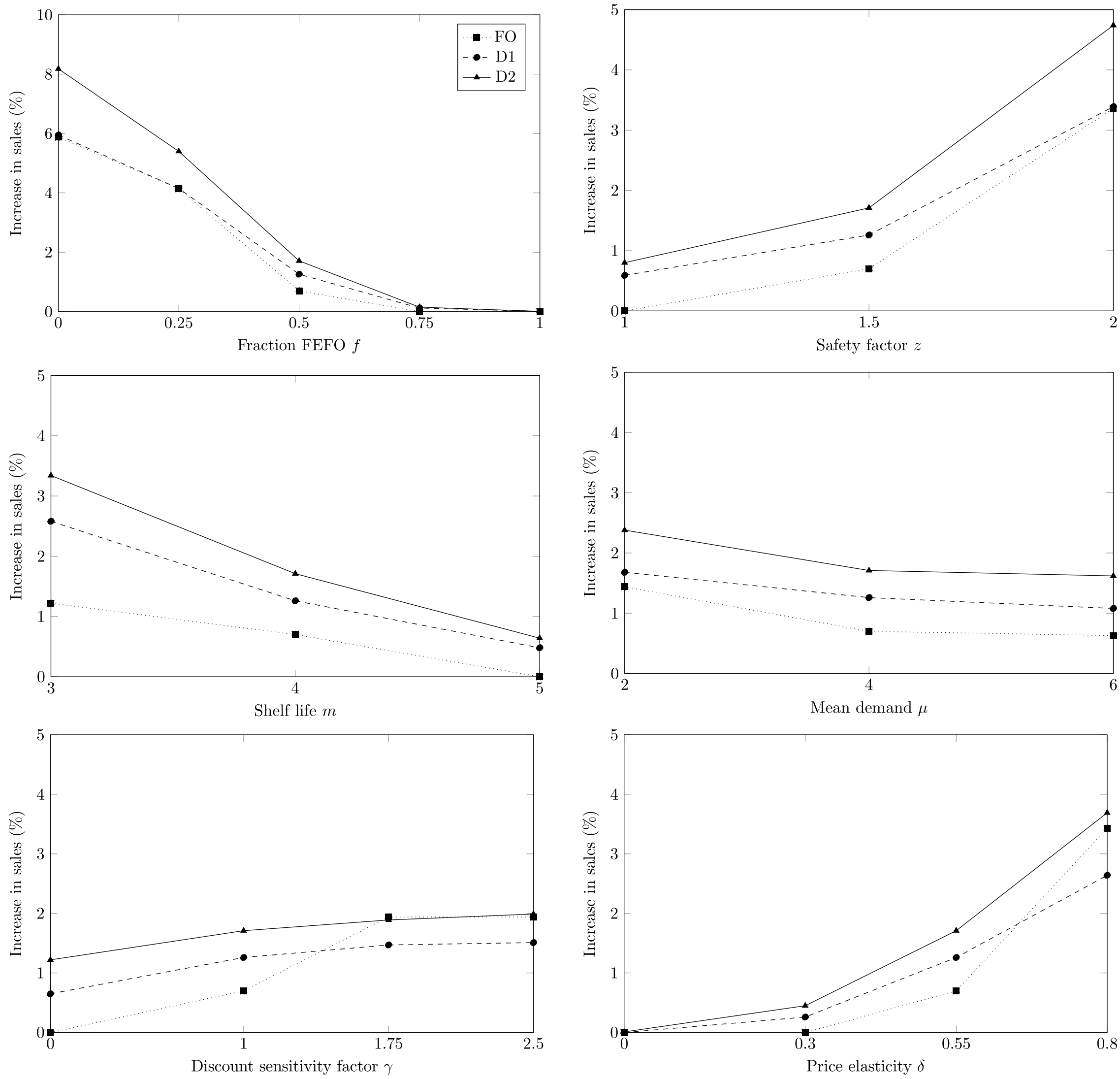

Fig. 4. Effect of parameters f, z, m, μ, γ, and δ on extra sales.

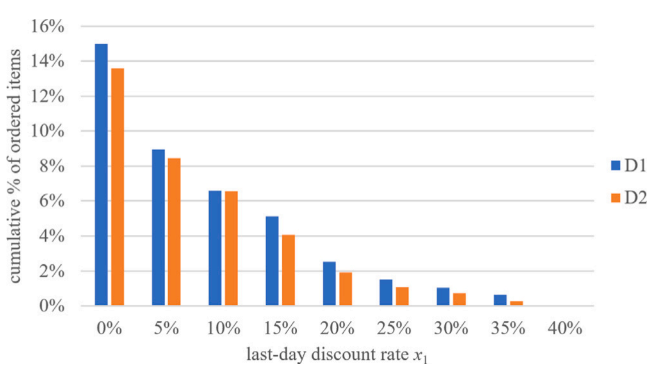

Fig. 5. Fraction of ordered items that are discounted on the last day at rate x_1 or higher (base case).

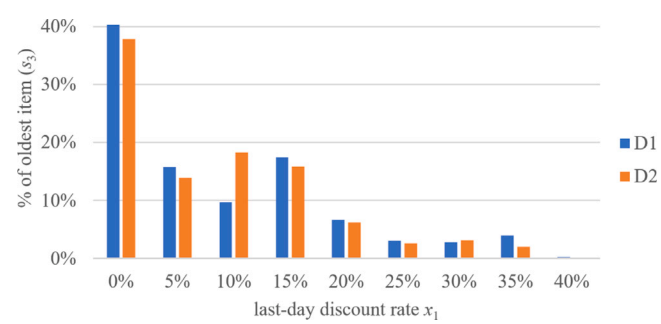

Fig. 6. Fraction of oldest items that are discounted on the last day at rate x_1 (base case).

of product in stock of age $i \in \{0, \ldots, m-1\}$:

$$\forall x \in \mathcal{X}: \quad \bar{s}_i(x_1) = \frac{\sum_{s \in \mathcal{S}(x_1)} g(s) \cdot s_i}{\sum_{s \in \mathcal{S}(x_1)} g(s)},$$

where $\mathcal{S}(x_1) = \{s \in \mathcal{S} \; : \; s_3 > 0 \;\wedge\; \pi^*(s) = x_1\}$, and $g(s)$ denotes the number of days state s occurs in a simulation of 100,000 periods. Note that states with $s_3 = 0$ are excluded, as in that case there are no items to discount. This corresponds to 68,115 out of 100,000 simulated days. The last column $h(x_1)$ indicates how often the discount rate x_1 is selected when products of age 3 are available ($s_3 > 0$).

Fig. 7 highlights that a low discount rate relates to a relatively large number of fresh products in stock. And, typically, high discount rates apply when relatively many products in stock are two days old (i,e, when $\bar{s_2}(x_1)$ is high). The highest discount rate is rarely set: at only

x_1	$\bar{s}_0(x_1)$	$\bar{s}_1(x_1)$	$\bar{s}_2(x_1)$	$\bar{s}_3(x_1)$	total	$h(x_1)$
0%	4.0	1.3	0.9	1.5	7.7	50.7%
5%	3.4	1.4	1.0	2.8	8.5	10.9%
10%	2.7	1.8	2.0	3.0	9.4	6.2%
15%	2.8	1.3	2.2	2.8	9.1	11.8%
20%	2.9	2.0	2.2	2.4	9.5	5.3%
25%	3.2	2.5	2.4	1.5	9.7	3.8%
30%	2.6	2.3	3.2	1.1	9.2	5.0%
35%	2.1	2.2	4.3	1.3	9.9	5.9%
40%	1.2	0.6	6.2	1.6	9.6	0.3%

Fig. 7. Weighted average number of products in stock in each age class and discount level x_1 for policy D1 (base case).

Table 7
Structure of D1: optimal discount x_1 for a selection of states (base case).

s_0	s_1	s_2	s_3	x_1
4	4	2	0	0%
4	3	2	1	20%
4	2	2	2	0%
4	1	2	3	10%
4	0	2	4	15%

Table 8
Structure of D1: optimal discount x_1 for a selection of states (base case).

s_0	s_1	s_2	s_3	x_1
0	5	3	2	5%
1	4	3	2	15%
2	3	3	2	20%
3	2	3	2	15%
4	1	3	2	15%
5	0	3	2	10%

0.3% of the days where $s_3 > 0$. Most often, the discount rates are 0%, 5%, or 15%.

Predicting the optimal discount level at individual states appears to be difficult, as discount levels vary significantly. For example, in Table 7 the optimal discount levels of D1 are presented for a subset of five states. In all states, 10 items are in stock, of which four are new (zero days old), and two are two days old. The optimal discount levels vary from 0% to 20%, depending on the division of the other four items over s_1 and s_3. Zero discount applies when 0 or 2 items are old, when 1 item is old 20% discount is optimal, and when 3 or 4 items are old, the optimal discount level is 10% and 15%, respectively.

In addition, the distinction between old stock and new stock does not explain the variation in discount levels. In Table 8 the optimal discount levels of D1 are presented for six states. In all states, 10 items are in stock, of which three are two days old and two are three days old. The optimal discount levels vary from 5% to 20%, depending on the division of the other five items over s_0 and s_1. A low discount level applies when no item is present in one of the age classes. Surprisingly, if all five other items are new, the discount is higher than in the case they are one day old. This is not due to the stopping criterion of VI, as we get the same optimal discount levels when VI is executed at a higher level of precision ($\epsilon = 0.00001$). Hence, in the long run, it must be (slightly) better to discount in the latter state at 10% than at a lower value. We conclude that the optimal policy is complex and difficult to capture in simple rules.

We explored predicting the discount levels of the base case, using a regression and classification study using weights $g(s)$ to predict x_1 based on (s_0, s_1, s_2, s_3). Simple models like weighted linear regression or weighted logistic regression did not fit well, nor did they after some feature engineering. Apparently, the variation between states that yield the same discount rate appears to be large. Tree-based predictions work reasonably well but result in deep trees with a large number of nodes. Hence, determining a heuristic for setting discount rates is a relevant open research question.

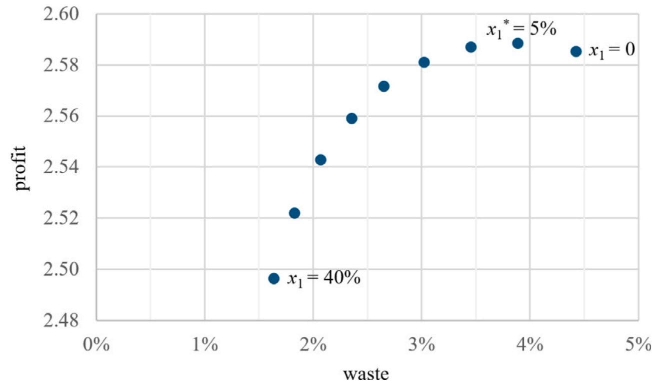

Fig. 8. Effect of sub-optimal fixed discounting on profit and waste (base case).

5.5. Sub-optimal static discounting

Many retailers that apply a fixed discount rate often install much higher discount percentages than the optimal levels of FO that we find in this study. Fig. 8 demonstrates for the base case the optimization of FO and the impact of sub-optimal fixed last-day discount levels on the profit and waste. For the base case, the profit is maximal at a fixed discount level of 5%. No discounting ($x_1 = 0$) results in a slightly lower profit and higher waste (4.4% versus 3.9%). Higher discount rates result in lower profit and less waste. Note that the marginal effect of discounting on waste gets smaller, whereas the negative impact on profit gets larger.

Often, we observe in (Dutch) supermarkets fixed discount rates of 35%. At 35% last-day discounting, the expected profit is about 2.5% lower than the optimal profit level (2.522 versus 2.588), whereas waste is reduced from 3.9% to 1.8%.

For the base case, a retailer who does not want to waste more than 2.5% should accept a profit that is 1% below optimal when he sticks to applying a fixed last-day discounting policy. Alternatively, he can choose to apply a more advanced discounting policy or adjust the ordering policy (e.g., lowering the safety stock).

5.6. Robustness to deviations in base stock level, review period and lead time

A practical question is whether the ranking of the policies will change when suboptimal discount rates apply. Specifically, we examine the policy ranking when the base stock level, review period, and/or lead time deviate from the value assumed in optimizing FO, D1, D2S, and D2.

Deviation from the base stock level B

To assess the robustness of the policy ranking to deviations from the base stock level, we simulate the policies with base stock level B_{sim} instead of B, which was used in the optimization of the discounting rates. We limit the analysis to the base case in which $B = 12$. The optimal discount rates for $B = 12$ are simulated with B_{sim} set to 11, 12, and 13. In Table 9, we report the profit gain of the policies compared to no discounting (NO) and the relative waste. We observe the ranking is the same but the difference in profit gain is larger for a larger value of B_{sim}, as in that case more items risk to turn into waste and thus more items get discounted.

The profit gain can be higher when discount rates are optimized for $B = B_{sim}$, as shown in Experiment 6 and 7. In Experiment 6 and 7 in

Table 9
Robustness of policy ranking in deviations in base stock level.

B_{sim}	Profit	% Profit gain over NO				Waste				
	NO	FO	D1	D2S	D2	NO	FO	D1	D2S	D2
$B-1=11$	€2.66	0.00%	0.46%	0.46%	0.57%	2.8%	2.8%	2.2%	2.2%	2.1%
$B=12$ (base case)	€2.58	0.09%	1.10%	1.21%	1.58%	4.4%	3.9%	3.3%	3.2%	3.0%
$B+1=13$	€2.44	0.85%	2.13%	2.53%	3.11%	6.5%	4.4%	4.6%	4.4%	4.2%

Table 10
Robustness of policy ranking in deviations in review period and lead time. Discount rates are optimized for the base case $R = L = 1$ (For FO the rate is thus set to 5%.)

R	L	Profit	% Profit gain over NO				Waste				
		NO	FO	D1	D2S	D2	NO	FO	D1	D2S	D2
1	1	€2.58	0.13%	1.14%	1.28%	1.6%	4.4%	3.9%	3.3%	3.2%	3.0%
1	2	€2.36	0.56%	1.99%	2.60%	3.0%	7.1%	6.3%	5.3%	5.1%	4.9%
2	1	€1.97	0.33%	1.65%	1.69%	1.63%	11.5%	10.8%	10.2%	10.1%	10.1%
2	2	€1.78	0.68%	1.82%	1.75%	1.76%	13.7%	12.9%	12.4%	12.4%	12.4%

Table 5, we have investigated optimal discount rates for $B = 11$, and 14 (by setting $z = 1$ respectively 2). This shows the same policy ranking for different values of the base stock level.

Larger review period R and lead time L

The review period is the number of periods between two successive decision moments (for discounting and ordering). We distinct between the review period for the discounting decision, which is 1 in all experiments, and the order review period R which we will set to $R = 1$ or $R = 2$ in the analysis below. The lead time L is the number of periods between the ordering and the receipt of new items. For many Dutch supermarkets holds $R = L = 1$. However, in case of smaller or more remote stores, the review period or lead time could be two rather than one. A lead time of two periods applies when more time is needed for the production and distribution planning and operations.

So far, the discount rates of FO, D1, D2S, and D2 are optimized assuming the review period R and the lead time L are both one day. We assess the robustness of the discount rates by applying them to a setting with $R = 2$, and/or $L = 2$. The review period R relates to the ordering decision, and not to the discounting decision (which is taken every day). Therefore, in the simulation model we assume discount rates are still adjusted at the start of every day, but ordering happens every R days and deliveries arrive L days later. The inventory and ordering dynamics are adjusted accordingly. That is, orders are placed only on days t for which hold $t \mod R = 0$ (where $\mod$ is the modulo operator that returns the remainder after integer division). The base stock level B_{sim} is set to $\mu \cdot (R + L) + z \cdot \sigma \cdot \sqrt{R + L}$, and rounded to the nearest integer. In the base case ($R = L = 1$) this yields $B_{sim} = 12$. When R or L is two (but not both), B_{sim} is set to 17. And for $R = L = 2$ holds $B_{sim} = 22$.

Table 10 shows the impact of applying the optimal discount rates for $R = L = 1$ to settings where R and L are 1 or 2. When R or L (or both) are two than the profit decreases compared to the base case with $R = L = 1$, as the inventory system is less responsive to demand fluctuations. The ranking of the policies is hardly affected. When $R = 1$ and $L = 2$ the profit gain over NO is larger as the safety stock and waste is larger.

When the review period is two ($R = 2$), orders are placed only every other day (if $t \mod R = 0$). The average order size is thus about twice as large than in case of $R = 1$. The replenishment is thus less smooth and less responsive to demand fluctuations, which results in a higher waste percentage and a lower average profit. As the shelf life in the base case is four days, at most two different expiration dates are found on the shelf. Hence D2S and D2 hardly improve over D1. As the discount rates are optimal for $R = L = 1$, and likely sub-optimal for other values, it may happen that D1 is slightly better than D2S and D2, which we observe when $R = L = 2$. As the differences are very small, we conclude the policy ranking is robust to changes in R, and L.

6. Discussion and conclusion

6.1. Scientific contribution

In this paper, we have formulated the expiration-date-based discounting problem of a non-monopolistic, profit-maximizing retailer as a Markov decision problem (MDP), and solved it by stochastic dynamic programming (SDP). The resulting optimal policies are simulated to also evaluate their impact on product waste and fill rate.

Previous literature has been limited in terms of the number of co-existing age classes and prices, as well as in how consumer picking behavior is included. In many studies, it is assumed that consumers behave rational, but in practice a large part of the consumers pick old products from the shelf even when they are not discounted. In this paper, we do include these so-called FEFO consumers next to LEFO consumers, which is important as it significantly influences the optimal discounting policy. Combined with our focus on more than one discount rate, the results provide insights in the value of differentiated discounting, and how this can be used to both maximize profits and reduce food waste.

We optimize two discount rates by SDP, and we uniquely derive optimal discounting rates for products with a shelf life of up to five periods (e.g. days). As replenishment happens every period, one could have items in stock of five different expiration dates. Our results can serve as new benchmarks for future research that applies approximation methods, like reinforcement learning.

6.2. Practical implications

One of the main questions driving this research was whether it is useful to start discounting in an earlier stage, using the last-two-days discounting policy. Our results do confirm its usefulness: Dynamic last-two-days discounting (D2) results in all 17 experiments in the highest profit and lowest waste level. The structure of the optimal discounting policy is unfortunately complex and difficult to capture in simple rules to use in practice. The numerical results generated some insights and guidelines for practitioners to further develop and experiment with their discounting policies. Furthermore, the optimal results obtained in this paper, are an important benchmark for heuristics and approximation methods such as a machine learning methods, like reinforcement learning (RL).

In practice, many retailers apply a fixed discount percentage of 20% to 40% to items that are about to expire (e.g. at the last day). Remarkably, in our results, the average optimal discount percentage of all policies is much lower than the ones we see being used in practice. The optimal discount level of D2 is on average 8.9%, but it varies between 0 and 40% depending on the number of items in stock and their ages. Often, no discount is provided even when old products are in stock. Under D2, 8.5% of all items ordered get eventually discounted, whereas under the optimal fixed discounting policy (FO) more than 15% of the items get discounted (as discounts are installed every day to all items that are about to expire that day). Dynamic policies with fixed discount rates can also be studied with the MDP model after making some adjustments (e.g. to optimize when during the day to start applying a fixed discount), as we will discuss in the next subsection on future research.

Interviews with supply chain managers of a few large retailers revealed that, in practice, little is known about what discount level is best. As a result, these practitioners also wonder whether the discounting percentages they use may be too high. High fixed discount levels are effective in reducing food waste, but could negatively impact profit. The effect on profit is less visible to store managers than the impact on food waste.

This study sheds a light on potential savings in waste and profit by a numerical comparison of different policies. On average the profit level of D2 is 3.8% higher than with no discounting (NO), 1.7% higher than the optimal fixed discount policy (FO), and 1% higher than the optimal dynamic last-day discount policy (D1). Applying identical dynamic discount rates to items with the last two expiration dates (as in policy D2S), yields a slightly lower profit gain over NO of 3.5%. The average waste is reduced from 5.6% in case of no discounting to 3.6% by last two days discounting. Last-day dynamic discounting (D1) yield an average waste of 4.0%. Lower levels of waste can be achieved by increasing the waste disposal cost.

Note that the profit gains and the savings in waste strongly depend on consumers preference for fresh items and sensitivity to discounts. We model the willingness of LEFO consumer to buy discounted items instead of the freshest items. This fits well to the case of displaying discounted items on the same shelf next to the fresher items. When discounted items are displayed in a separate section in the store, fewer LEFO consumers become aware of their availability. This can be captured by the model by parameter γ, which controls the fraction of LEFO consumers that respond to a discount. We recommend to carefully estimate consumers picking behavior and their sensitivity to discounts and expiration dates.

Although superior in both waste reduction and increasing profit, the dynamic last-two-days discounting policy (D2) is more difficult to implement without the technological possibilities (i.e., digital price tags) described in the introduction. In case stickers have to be put manually on the product, discounting different items with different discount percentages is labor-intensive and potentially prone to errors. Policy D2S is then easier as it puts the same discount label on items with different expiration dates. To reduce re-labeling every day, one may limit the number of discount levels to choose from to only a few. This would possibly simplify the process, but would also reduce the average profit gain.

In situations in which waste is relatively low, e.g. when older products are sold anyway (due to a higher share of FEFO customers, or the customer base has a low discount sensitivity), it could be better to only discount the items on the last day. Or in case of high labor cost, or high IT investment cost, to not discount at all.

6.3. Further research

This research shows that the effect of discounting policies strongly depends on how consumer trade off product shelf life (as a proxy to product quality) and price. An important stream of future research is to better understand how expiration dates and discounts affect consumer behavior: e.g., regarding the product withdrawal, the price elasticity, and the discount sensitivity of LEFO consumers. Consumers' responses to discounts may strongly depend on where discounted items are located. Whether discounted items should be stored in a separate section or on the same shelf remains an open research question. On the one hand, displaying discounted items at a separate central location makes discounted items more visible to consumers in search for discounted items and thus potentially generating more demand. Furthermore, a central location may be more efficient to maintain, e.g., to adjust discount labels manually, and to remove expired items. On the other hand, it may encourage consumers buying less items at the full price by first shopping at the section of discounted items. When discounted items are less visible to LEFO consumers, fewer LEFO consumers may be triggered to change their undesired picking behavior. Another drawback of separate storage is the risk that consumers encounter stock-outs of full-priced products while discounted items remain available in the separate section. Hence more research is needed to determine when to display on the same shelf and when to store discounted items separate from the full-priced items.

Another line of future research may focus on other discounting policies. The MDP framework can fit other discounting policies than those considered in this paper. For example, in practice one may also apply a dynamic, fixed-rate discounting policy, which determines every day (based on the inventory state) whether a discount at a fixed percentage should be installed or not. This type of dynamic On–Off policy, can be optimized by limiting the action space to only two options: a fixed discount percentage or no discount (0%). Further, one may consider to optimize the timing of a discount (during the day) rather applying discounts from the start of a day. Therefore a day can be split into sub-period at which the discounting decision is reviewed. The state of the MDP model is then expanded by an index indicating the time slot during the day, and the MDP would then become periodic. Alternatively, the model can be adjusted to optimize the number of items to discount rather than the discounting percentage.

In our research we fixed the ordering policy to a base stock policy with a constant base stock level. This reflects practical situations where safety stock is a strategic decision aimed at ensuring sufficiently high product availability. When including the order quantity in a profit maximization model, stock-outs may occur too frequently, leading to disappointed consumers. Nevertheless, future research could focus on optimizing both the order quantity and the discount rates. This could be addressed in the MDP model by adding a penalty for missed demand to the objective function, and/or by imposing a (stock-dependent) minimum order quantity in the action space to limit stock-outs.

In this paper, we assumed the order review period and lead time are both one day. An interesting direction for future research is optimizing the discount rates for settings where the review period and lead time differ from one, as motivated in Section 5.6. The MDP model can be adjusted, but its solution process becomes more elaborate. When the review period is set to two or more periods, while keeping $L = 1$, then the number of periods elapsed since the last order should be added to the state description, causing the MDP to become periodic. Similarly, a smaller lead time of say half a day, can be handled by splitting a period into two sub-periods and adding an index for the sub-period to the state space, A larger lead time of two or more, while keeping $R = 1$, requires extending the state with the outstanding order(s). Then the state space may become too large to determine an optimal solution using an (exact) dynamic-programming-based methods, like value iteration. In such cases, reinforcement learning (RL) approaches could be used to find approximate solutions.

As retailers like supermarkets sell many perishable products that are to some extend substitutes to each other, future research could investigate the interaction between discounting, product substitution, and waste. When one product is on discount the demand for another product may drop, which may cause more waste of that product. Thus discounting may reduce waste of the discounted product, but may result in more waste (and discounting) of other products. Depending on consumers' picking behavior, their price elasticity, their sensitivity to discounts, and their willingness to substitute, it may even be better to not discount at all or even to donate nearly expired products to charity organization. The impact of discounting in a multi-product setting is understudied. It can be modeled as an MDP, but its optimal solution is intractable due to the explosion in the number of states. Again, a RL approach could be applied to determine dynamic discounting strategies.

6.4. Conclusion

To conclude, the MDP framework provides a powerful approach to designing expiration-date-based discounting policies. We demonstrated when and how much profit can be gained and how much waste can be reduced by implementing such policies. In our analysis we model FEFO and LEFO consumers, and allow for more than two co-existing age classes, and two discount levels.

The tractability of exact dynamic programming methods, as well as the complexity of resulting dynamic policies, motivates future research into heuristic approaches, such as reinforcement learning. Although these methods do not guarantee an optimal solution, they can potentially provide good solutions reasonably fast, facilitating implementation, and experimentation in practice. The optimal solutions derived from our MDP approach could serve as useful benchmarks for these machine learning methods. The use of machine learning approaches provide opportunities to extend the model to multiple products, other discounting policies, a dynamic ordering policy, and a larger review periods and lead times.

Despite its limitations and opportunities for further research, this paper provides valuable insights for supermarkets aiming to optimize their profit and waste levels. The provided model can be tailored to specific product and store specifications. It demonstrates the added value of dynamic discounting over static approaches, as well as the benefits of applying two discount levels simultaneously instead of just one. Furthermore, it contributes to the ongoing discussion and future research on determining optimal discount levels in practice.

CRediT authorship contribution statement

Rene Haijema: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Formal analysis, Conceptualization. **Lisan Duijvestijn:** Software, Methodology, Formal analysis, Conceptualization. **Renzo Akkerman:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Frans Cruijssen:** Writing – review & editing, Validation, Funding acquisition.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 program under grant agreement No 101036388.

Data availability

No data was used for the research described in the article.

References

Adenso-Díaz, B., Lozano, S., Palacio, A., 2017. Effects of dynamic pricing of perishable products on revenue and waste. Appl. Math. Model. 45, 148–164. http://dx.doi.org/10.1016/j.apm.2016.12.024.

Akkas, A., 2024. Food Surplus Management in Retailing: A Global Perspective. Isenberg School of Management, University of Massachusetts Amherst.

Andreyeva, T., Long, M.W., Brownell, K.D., 2010. The impact of food prices on consumption: A systematic review of research on the price elasticity of demand for food. Am. J. Public Health 100 (2), 216–222. http://dx.doi.org/10.2105/AJPH.2008.151415.

Aschemann-Witzel, J., 2018. Consumer perception and preference for suboptimal food under the emerging practice of expiration date based pricing in supermarkets. Food Qual. Pref. 63, 119–128. http://dx.doi.org/10.1016/j.foodqual.2017.08.007, https://www.sciencedirect.com/science/article/pii/S0950329317301908.

Brandimarte, P., Gioia, D.G., 2022. Base-stock inventory control rules for perishable items: Impact of problem features and robustness to consumer behavior misspecification. SSRN Electron. J. http://dx.doi.org/10.2139/ssrn.4084025, https://www.ssrn.com/abstract=4084025.

Buisman, M.E., Haijema, R., Bloemhof-Ruwaard, J.M., 2019. Discounting and dynamic shelf life to reduce fresh food waste at retailers. Int. J. Prod. Econ. 209, 274–284. http://dx.doi.org/10.1016/j.ijpe.2017.07.016.

Chades, I., Chapron, G., Cros, M.-J., Garcia, F., Sabbadin, R., 2014. MDPtoolbox: A multi-platform toolbox to solve Markov decision processes. Ecography 37 (9), 916–920.

Chang, H.-H., Su, J.-W., 2022. Sustainable consumption in Taiwan retailing: The impact of product features and price promotion on purchase behaviors toward expiring products. Food Qual. Pref. 96, 104452. http://dx.doi.org/10.1016/j.foodqual.2021.104452, https://www.sciencedirect.com/science/article/pii/S0950329321003347.

Chatwin, R.E., 2000. Optimal dynamic pricing of perishable products with stochastic demand and a finite set of prices. European J. Oper. Res. 125 (1), 149–174. http://dx.doi.org/10.1016/S0377-2217(99)00211-8.

Chew, E.P., Lee, C., Liu, R., Hong, K.S., Zhang, A., 2014. Optimal dynamic pricing and ordering decisions for perishable products. In: International Journal of Production Economics. vol. 157, Elsevier B.V., pp. 39–48. http://dx.doi.org/10.1016/j.ijpe.2013.12.022, (1).

Chua, G.A., Mokhlesi, R., Sainathan, A., 2017. Optimal discounting and replenishment policies for perishable products. Int. J. Prod. Econ. 186, 8–20. http://dx.doi.org/10.1016/j.ijpe.2017.01.016.

Chung, J., Li, D., 2014. A simulation of the impacts of dynamic price management for perishable foods on retailer performance in the presence of need-driven purchasing consumers. J. Oper. Res. Soc. 65 (8), 1177–1188. http://dx.doi.org/10.1057/jors.2013.63.

Duan, Y., Liu, J., 2019. Optimal dynamic pricing for perishable foods with quality and quantity deteriorating simultaneously under reference price effects. Int. J. Syst. Sci.: Oper. Logist. 6 (4), 346–355. http://dx.doi.org/10.1080/23302674.2018.1465618.

Dudlicek, J., Goldschmidt, B., Kleckler, A., Martin, K., 2019. Personal business: Putting the shopper first is the best strategy for success. Progress. Groc. 22–33, https://issuu.com/ensembleiq/docs/pg-0719?fr=sZTFiOTcxMjU.

Fan, T., Xu, C., Tao, F., 2020. Dynamic pricing and replenishment policy for fresh produce. Comput. Ind. Eng. 139, 106127. http://dx.doi.org/10.1016/j.cie.2019.106127.

Ferguson, M., Ketzenberg, M.E., 2006. Information sharing to improve retail product freshness of perishables. Prod. Oper. Manage. 15 (1), 57–73. http://dx.doi.org/10.1111/j.1937-5956.2006.tb00003.x.

Hansen, K., Misra, K., Sanders, R.E., 2024. Uninformed choices in perishables. Mark. Sci. 43 (4), 751–777.

Harcar, T., Karakaya, F., 2005. A cross-cultural exploration of attitudes toward product expiration dates. Psychol. Mark. 22 (4), 353–371. http://dx.doi.org/10.1002/mar.20063.

Herbon, A., 2014. Dynamic pricing vs. acquiring information on consumers' heterogeneous sensitivity to product freshness. Int. J. Prod. Res. 52 (3), 918–933. http://dx.doi.org/10.1080/00207543.2013.843800.

Herbon, A., Levner, E., Cheng, T., 2014. Perishable inventory management with dynamic pricing using time–temperature indicators linked to automatic detecting devices. Int. J. Prod. Econ. 147, 605–613. http://dx.doi.org/10.1016/j.ijpe.2013.07.021, https://www.sciencedirect.com/science/article/pii/S0925527313003332.

Hübner, A., Ostermeier, M., Schäfer, F., Winkler, T., 2024. Seeking the freshest? Preventing customer picking for expiration dates with retail operations. [Available at SSRN]. http://dx.doi.org/10.2139/ssrn.4731442.

Kaya, O., Ghahroodi, S.R., 2018. Inventory control and pricing for perishable products under age and price dependent stochastic demand. Math. Methods Oper. Res. 88 (1), 1–35. http://dx.doi.org/10.1007/s00186-017-0626-9.

Li, R., Teng, J.T., 2018. Pricing and lot-sizing decisions for perishable goods when demand depends on selling price, reference price, product freshness, and displayed stocks. European J. Oper. Res. 270 (3), 1099–1108. http://dx.doi.org/10.1016/j.ejor.2018.04.029.

Liu, X., Tang, o., Huang, P., 2008. Dynamic pricing and ordering decision for the perishable food of the supermarket using RFID technology. Asia Pac. J. Mark. Logist. 20 (1), 7–22. http://dx.doi.org/10.1108/13555850810844841.

Nijs, V.R., Dekimpe, M.G., Steenkamp, J.B.E., Hanssens, D.M., 2001. The category-demand effects of price promotions. Mark. Sci. 20 (1), 1–22. http://dx.doi.org/10.1287/mksc.20.1.1.10197.

Ostermeier, M., Düsterhöft, T., Hübner, A., 2021. A model and solution approach for store-wide shelf space allocation. Omega 102, 102425. http://dx.doi.org/10.1016/j.omega.2021.102425, https://www.sciencedirect.com/science/article/pii/S0305048321000347.

Puterman, M.L., 2014. Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons.

Rabbani, M., Pourmohammad Zia, N., Rafiei, H., 2016. Joint optimal dynamic pricing and replenishment policies for items with simultaneous quality and physical quantity deterioration. Appl. Math. Comput. 287–288, 149–160. http://dx.doi.org/10.1016/j.amc.2016.04.016.

Riesenegger, L., Santos, M.J., Ostermeier, M., Martins, S., Amorim, P., Hübner, A., 2023. Minimizing food waste in grocery store operations: Literature review and research agenda. Sustain. Anal. Model. 3, 100023.

Sanders, R.E., 2024. Dynamic pricing and organic waste bans: A study of grocery retailers' incentives to reduce food waste. Mark. Sci. 43 (2), 289–316. http://dx.doi.org/10.1287/mksc.2020.0214.

Santos, M.J., Martins, S., Amorim, P., Almada-Lobo, B., 2022. On the impact of adjusting the minimum life on receipt (MLOR) criterion in food supply chains. Omega (United Kingdom) 112, 102691. http://dx.doi.org/10.1016/j.omega.2022.102691.

Scholz, K., Eriksson, M., Strid, I., 2015. Carbon footprint of supermarket food waste. Resour. Conserv. Recycl. 94, 56–65. http://dx.doi.org/10.1016/j.resconrec.2014.11.016.

Scholz, M., Kulko, R.D., 2022. Dynamic pricing of perishable food as a sustainable business model. Br. Food J. 124 (5), 1609–1621. http://dx.doi.org/10.1108/BFJ-03-2021-0294.

S&K Solutions GmbH, 2025. "Too good for the garbage can"- less waste in food retailing with the help of electronic price tags. https://www.e-shelf-labels.com/news/news-about-labels-and-signage/too-good-for-the-garbage-can-less-waste-in-food-retailing-with-the-help-of-electronic-price-tags.html.

Stenmarck, Å., Werge, M., Hanssen, O.J., Silvennoinen, K., Katajajuuri, J.-M., 2011. Initiatives on Prevention of Food Waste in the Retail and Wholesale Trades. Nordic Council of Ministers, http://dx.doi.org/10.6027/tn2011-548, http://urn.kb.se/resolve?urn=urn:nbn:se:norden:org:diva-1655.

Tromp, S.O., Rijgersberg, H., Pereira Da Silva, F., Bartels, P., 2012. Retail benefits of dynamic expiry dates - simulating opportunity losses due to product loss, discount policy and out of stock. Int. J. Prod. Econ. 139 (1), 14–21. http://dx.doi.org/10.1016/j.ijpe.2011.04.029.

Tsiros, M., Heilman, C.M., 2005. The effect of expiration dates and perceived risk on purchasing behavior in grocery store perishable categories. J. Mark. 69 (2), 114–129. http://dx.doi.org/10.1509/jmkg.69.2.114.60762.

van Donselaar, K., van Woensel, T., Broekmeulen, R., Fransoo, J., 2006. Inventory control of perishables in supermarkets. Int. J. Prod. Econ. 104 (2), 462–472. http://dx.doi.org/10.1016/j.ijpe.2004.10.019, https://www.sciencedirect.com/science/article/pii/S0925527305000381.

Vollebregt, H.M., 2020. Dutch supermarkets provide insights into food waste. https://www.wur.nl/en/research-results/research-institutes/food-biobased-research/show-fbr/dutch-supermarkets-provide-insights-into-food-waste-f00dwa5.htm. (Accessed 02 October 2024).

Vollebregt, M., 2023. Monitor voedselverspilling: Wat is aan voedsel verloren en verspild in de nederlandse retail in 2023. Wageningen Food & Biobased Research, Rapport 2585, https://doi.org/10.18174/672427.

Yavuz, T., Kaya, O., 2024. Deep reinforcement learning algorithms for dynamic pricing and inventory management of perishable products. Appl. Soft Comput. 163, 111864.

Zhang, J., Wang, Y., Lu, L., Tang, W., 2015. Optimal dynamic pricing and replenishment cycle for non-instantaneous deterioration items with inventory-level-dependent demand. Int. J. Prod. Econ. 170, 136–145. http://dx.doi.org/10.1016/j.ijpe.2015.09.016.