Inventory management of vertically differentiated perishable products with stock-out based substitution

Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-10 (2022) 2683–2688

Inventory management of vertically differentiated perishable products with stock-out based substitution

Daniele Giovanni Gioia[1] * **Leonardo Kanashiro Felizardo**[1] ** **Paolo Brandimarte**[1] ***

* *e-mail: daniele.gioia@polito.it, ORCID: 0000-0001-8979-4174*
** *e-mail: leonardo.kanashiro@polito.it, ORCID: 0000-0002-2871-860X*
*** *e-mail: paolo.brandimarte@polito.it, ORCID: 0000-0002-6533-3055*
[1] *Department of Mathematical Sciences DISMA, Politecnico di Torino, 10129 Corso Duca degli Abruzzi 24 Turin (TO) Italy.*

Abstract: The need for optimal inventory control strategies for perishable items is of the utmost importance to reduce the large share of food products that expire before consumption and to achieve responsible food stocking policies. Our study allows for a multi-item setting with substitution between similar goods, deterministic deterioration, delivery lead times and seasonality. Namely, we model demand by a linear discrete choice model to represent a vertical differentiation between products. The verticality assumption is further applied in a novel way within product categories. Specifically, the same product typology is vertically decomposed according to the age of the single stock-keeping unit in a quality-based manner. We compare two different policies to select the daily size of the orders for each product. On the one hand, we apply one of the most classical approaches in inventory management, relying on the Order-Up-To policy, modified to deal with the seasonality. On the other hand, we operate a state-of-the-art actor-critic technique: Soft Actor-Critic (SAC). Although similar in terms of performance, the two policies show diverse replenishment patterns, handling products differently.

Keywords: Multi-item inventory systems, Perishable products, Inventory control, Reinforcement learning, Discrete choice models

1. INTRODUCTION AND PAPER POSITIONING

The management of the inventories is a well-known problem in production economics, where many of the available studies allow the substitution of different products when there is no deterioration (Shin et al., 2015). Although the practical benefits of a control strategy on substitutable products are renowned (Fisher and Raman, 2010), when dealing with perishability, the literature is quite scarce (Janssen et al., 2016), even if the response of the consumers to shelf out-of-stocks of perishable products is actually more favourable (Van Woensel et al., 2007). The literature on inventory systems for perishable products is commonly labelled according to the classification proposed by (Goyal and Giri, 2001) and then adopted by (Janssen et al., 2016). Specifically, they discriminate between *Obsolescence* and *Deterioration*. In brief, when there is obsolescence, the item may lose value due to technological or market competition reasons, causing a drastic reduction in the price of the good. Conversely, deterioration refers to the loss of value of the item due to ageing and many other possible reasons related to the internal characteristics of the product itself. We include this work in this last family, precisely among products with a maximum deterministic lifetime (e.g., vegetables and blood), thus further excluding circumstances where the value decays due to other reasons (e.g., gasoline vaporization), usually modelled by a stochastic lifetime. Our work assumes a multi-item setting with substitution between similar goods, entailing inventory and assortment planning problems. They can be characterized (Shin et al., 2015) by the substitution mechanism, the actor in charge of the substitution, and the substitution direction (one-way/two-way). When the perishable products are blood platelets (Haijema et al., 2005), it may be natural to model a *supplier-driven* substitution that exploits the compatibility of different blood groups. On the contrary, we presume a supermarket oriented application, where the substitution is *consumer-driven* and happens due to a stock-out of the first consumer's choice, thence the consumer decides on his own when to substitute a product. One possible way to model the stock-out substitution employs an exogenous pre-defined share of the consumers (Hendrix et al., 2019; Buismann et al., 2020) wishing to replace their preference when facing a stock-out, enforcing an implicit simplification of the phenomena. A more flexible approach relies on the *Discrete Choice Methods* that generate endogenous utility-based substitution rates (Transchel, 2017; Transchel et al., 2021) at the cost of increased model complexity.

Among many discrete choice methods, one way to decide which one fits the problem best is to distinguish between *horizontally* and *vertically* differentiated products. The former concerns idiosyncratic preferences that are not purely based on a quality-price ranking (Transchel et al.,

10.1016/j.ifacol.2022.10.115

2021), whereas the latter considers cases where the choice would be driven only by the quality if the price of the products were the same. Our work looks at grocery applications, where private labels and famous brands typically coexist, thus leading to a vertical differentiation of very similar products that we model by means of a *Linear discrete choice method*, widely employed to tackle verticality between goods. However, notice that we do not seek an optimal pricing as (Transchel, 2017) and we consider the price of the products as exogenously determined.

The model complexity usually constrains the proposed solution methods. Analytical approaches on substitution (Transchel et al., 2021) struggle to manage multiple time periods inventories, especially when a positive lead time and a fixed shelf life is associated with the products. To deal with sequential decisions in multiple time periods, a more general and flexible alternative is dynamic programming (Hendrix et al., 2019), but exact solutions provided by this strategy may suffer from the curse of dimensionality when value and policy functions are explicitly represented by lookup tables (Brandimarte, 2021; Powell, 2021), requiring the use of approximated methods. The classical heuristic approach applied to the discrete-time inventory problems with periodic-reviews considers *Order-Up-To* (OuT) rules. Several examples in the perishable literature handle both scenarios with no substitution (Haijema and Minner, 2019) and multi-item ones (Buismann et al., 2020), where products are subject to replacement. Those procedures allow a larger model flexibility and solve the problem by a direct policy function approximation.

In the following work, we present and test our novel designed multi-item perishable environment that, overcoming the various limitations of the above-mentioned approaches, allows for fixed lead times, shelf lives, seasonality and vertical substitution between products all at once. We execute the well-known OuT policy as benchmark to find an optimal replenishment policy, further applying a state-of-the-art reinforcement learning technique, Soft Actor-Critic (SAC) (Haarnoja et al., 2018). This latter flexible methodology approximates both value and policy functions using artificial neural networks and we eventually show its greater stability due to a smoother policy shape.

2. THE MODEL

The model we propose considers a simulation of a perishable items retailer that has to decide about the order quantity of J vertically differentiated products. The characteristics of each product category j are:

- LT_j: The time necessary to receive the goods after placing the order (expressed in days), identified as Lead-Time.
- SL_j: The residual life of the product at the delivery time, expressed as Shelf-Life.
- The selling price p_j and the cost c_j per unit.
- The perceived quality q_j of the item, summarized by a single value.

Each simulation step takes into account a day, following the structure provided in Fig 1. Specifically, at the beginning of day t, we deliver the awaited orders and the inventory is updated accordingly. We then sample the number of consumers employing a Poisson(λ_t) distribution that contemplates a weekly seasonality scaling factor in its parameter

$$\lambda_t = \eta_k \lambda, \qquad k = t_{\mathsf{mod7}}.$$

Table 1. Weekly seasonality pattern employed.

Weekday	Mon	Tue	Wed	Thr	Fri	Sat	Sun
k	0	1	2	3	4	5	6
Seasonality Factor (η_k)	0.68	0.76	0.76	0.76	0.99	1.52	1.52

The adopted seasonality factors pattern is shown in Table 1, but it represents just one of the infinite possibilities. A further assumption associate to each consumer no more than one stock-keeping unit, thus there is a one-to-one proportion between the product demand and the number of consumers.

Once we know the number of consumers, we iterate for each of them by making use of a linear discrete choice model, either to select which product to buy among the available ones or to purchase nothing. When there are no more consumers, the retailer discards all the expired items, and observes the state of the system (S_t). The adopted policy suggests the action x_t (the new order quantity) at the end of the day and the request is placed consistently, returning the daily profit C_t.

2.1 The discrete choice method

When we process a consumer, we assume a discrete choice method to calculate a utility function that allow for heterogeneous agents with a individual quantitative evaluation of the price-quality ratio of the available products. The discrete choice linear model computes, for each consumer n and product j

$$U_{nj} = \theta_n q_j - p_j, \tag{1}$$

where θ_n is a stochastic representation of the consumer's valuation, sampled by a beta distribution. The trade-off between the price p_j and the quality q_j is evaluated with regard to each product j in a linear way. The purchase happens when the utility of at least one item is greater than zero, by selecting the one with the highest positive value. Analytically, the consumer n picks:

$$\arg\max \{U_{nj}, j = 0, \ldots, J\},$$

where we introduce a null utility U_{n0} for the no purchase option, interpreted as an additional dummy product with no price and zero quality.

We modify the model to deal with perishability by increasing the initial number of products J to

$$\bar{J} = \sum_{j=1}^{J} \mathsf{SL}_j, \tag{2}$$

simulating the quality reduction due to the ageing process by an additional vertical differentiation of the items **within** their product category, according to their *Residual Life*. Practically speaking, we disaggregate the product categories by assigning each item to a group j based on the residual life (i.e., generalizing the quality definition), thus expanding the initial set of choices J to $\bar{J}$ ones from the consumer's point of view. However, the retailer will continue to select the orders quantity only for the J original products.

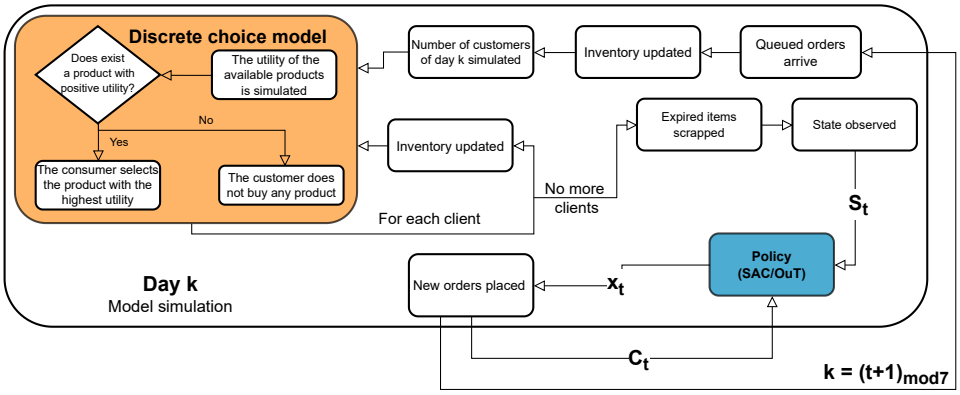


Fig. 1. Framework of the simulation and interaction with the policy.

Table 2. Product A and B characteristics.

	Price (p)	Quality (q)	Cost (c)	SL	LT
A	6	24	3	4	3
B	4	20	2	2	2

Closed-form optimal assortment and inventory replenishment results are computable in a single-period scenario, when no shelf-life and lead-time are present. In facts, these additions make the problem complex to investigate analytically and open up the way to approximated approach. Several works (Pan and Honhon, 2012; Transchel, 2017; Transchel et al., 2021), furnish a deep understanding of the behaviour of the model when employed in non-perishable retail settings, analyzing the optimal assortment, inventory and pricing. One of the principal traits of the linear approach analyzed in these works is the quick identification of structural properties that characterize the optimal assortment of products. Furthermore, dominated products (i.e., items that the consumers would never buy if others are available) can be analytically recognized. These results come from the quality, the price and the margin of the available options. However, in our setting, the lead-time length and a higher shelf-life may guide the policy of the inventory as well. Unfortunately, those potential advantages are not easily measurable.

2.2 Case study: model for two perishable products

To better explain how the consumer choice model is modified to deal with our perishable products oriented system, let us consider two vertically differentiated items, A and B, featuring the price and quality reported in Table 2. Assuming a beta(2,3) distribution for θ in Eq (1) and, initially, no quality depreciation due to the ageing, we can graphically represent their utility in Fig 2a by means of the two lines $\theta q_A - p_A$ in red and $\theta q_B - p_B$ in blue. The share of the captured demand by each product is visually suggested by the area of the distribution where the lines are dominant. Furthermore, we can identify:

- $\theta_1 = \frac{p_B}{q_B} = \frac{1}{5}$, where the consumers have sufficient utility to purchase something,
- $\theta_2 = \frac{p_A - p_B}{q_A - q_B} = \frac{1}{2}$, where the highest utility switch from B to A.

Following this setting, when A is preferred, but we have a stock-out, consumers have enough utility on B to substitute the item. Whereas, if the stock-out concerns B, the substitution happens only if $\theta > \frac{p_A}{q_A} = \frac{1}{4}$.

Ageing process When the stored goods get old, it is reasonable to assume a quality reduction. Pointing out the *Residual Life* of an item by a superscript q_j^{RL} , we expect:

$$q_j^{\mathsf{SL}} \geq \cdots \geq q_j^0, \quad \forall j.$$

This effect leads to the *Last-In-First-Out* (LIFO) consumers' behaviour illustrated in Fig 2b, where we only examine product B, applying an age decomposition such that we divide the original utility-line with respect to its age. Since we consider the price as an exogenous variable, the dashed-line $\theta q_B^1 - p_B$ shall always lie under $\theta q_B^2 - p_B$ in blue, intersecting in $\theta = 0$, thence all the consumers will prefer the newest item in blue until it will eventually run out, then they will pick the older one. Although we consider an exogenous price, the retailer may apply a discounted price to modify the LIFO behaviour into a (*First-In-First-Out*) FIFO/LIFO mixture. This strategy is way different from a dynamic pricing approach because it fixes a pre-determined discount on expiring products, moving upward the line of the age-decoupled items. However, a future challenge might be to optimize the pricing of expiring products to maximize the revenue. In Fig 2c, we examine both A and B, but we assume that only B is affected by ageing, applying a discounted price adjustment such that:

$$q_B^2 = 20 \rightarrow q_B^1 = 18, \quad p_B^2 = 4 \rightarrow p_B^1 = 3.3. \tag{3}$$

The selected parameters halve the preferences of B, causing half of the people to act FIFO and the rest LIFO. The estimation of the perceived quality factor is not easy, but the model we presented shows great flexibility to build personalized FIFO/LIFO mixtures in a simulation environment. Furthermore, the applied strategy takes naturally into account the ageing process thanks to the vertical differentiation assumption of the method and can be tuned by decomposing according to age batches.

Details of the employed model In the remainder of this work, we analyze two scenarios concerning products A and B of Table 2. We first consider the entire shelf-life of products A and B without any discount strategy, we decompose A as we did with B in Fig 2b, transforming the two lines of Fig 2a in $\mathsf{SL}_B + \mathsf{SL}_A$ ones, as illustrated in Fig 2d. Nevertheless, since we have no historical data to train the simulation model, it is not obvious how set

the quality depreciation when items get older. To provide a reliable approximation of the system dynamics when the price is fixed, we thus consider the linear model on products A and B of Fig 2a to select a category, then we assume that consumers follow a LIFO behaviour, purchasing the newest item available per chosen article kind. What changes from an exact decomposed simulation is a further implicit assumption. Specifically, we say that any consumer who favours A will not switch to B simply because of the freshness. This generates an approximation depicted in green in Fig 2d, where the *age-based substitution* would have happened, depending on the relative quality depreciation of the different articles. Additionally, notice that the positive utility of aged goods crosses the purchase line of Fig 2d with a slightly higher value. Even so, assuming a small daily depreciation and a shelf-life of a few days, we consider the effect negligible. The assumptions we made are easy to remove when precise data are available and, thanks to discounted prices, more complex substitution patterns are achievable. Nonetheless, if the approximation zone is dense of intersections, the required precision to exactly model the crossing points where the freshness substitution happens may boils down to an approximation anyway.

The second scenario will employ the discount policy of Eq (3), tested by maintaining the approximate LIFO policy on product A, whereas applying the values in Eq (3) on B, forcing a FIFO behaviour on half of the consumers who prefer B.

3. DYNAMICS OF THE POLICIES

Seeking an optimal replenishment policy for the two items produces a sequential optimization problem that can be tackled by numberless strategies (Powell, 2021). Regardless of the method employed, we firstly define the state, the actions and the reward of the problem.

3.1 State, action and reward functions

State variable The state variable of the system is observed at the end of each business day, after the expired items are scrapped. To define the state variable, we introduce the quantities related to the inventory and to the awaited orders as:

- $O^{\mathsf{d}}_{t,j}$: Ordered products of category $j \in \{\mathrm{A}, \mathrm{B}\}$ awaited in $d \in \{0, \dots, \mathsf{LT}-1\}$ days at time t. [1]
- $I^{d}_{t,j}$: Observed inventory for products $j \in \{\mathrm{A}, \mathrm{B}\}$ with a residual life of $d \in \{1, \dots, \mathsf{SL}\}$ days at time t.
- Week-day k, as defined in Table 1.

The observed state variable at time t for product j will be:

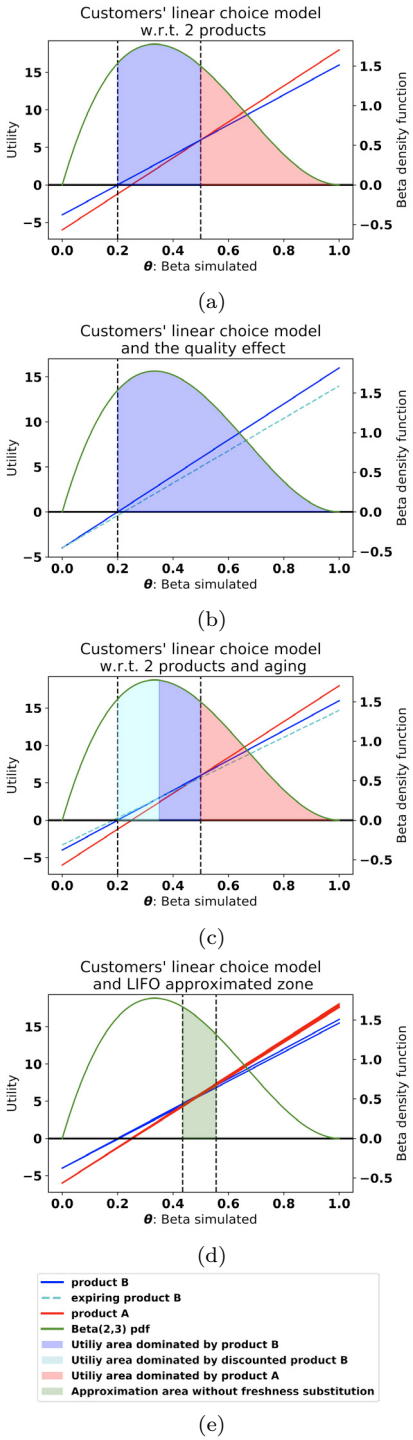
$$S_{t,j} = [O^{\mathsf{LT}-1}_{t,j}, \dots, O^{0}_{t,j} | I^{\mathsf{SL}}_{t,j}, \dots, I^{1}_{t,j}], \tag{4}$$

hence, the complete state variable in our scenarios will be

$$S_t = [S_{t,A} | S_{t,B} | k].$$

Notice that the dimension of the state space will be $\sum_j((\mathsf{SL}_j - 1) + \mathsf{LT}_j) + 1$, that is because when the state is observed, the expired items with 0 residual life are scrapped yet and the maximum lead time is reduced through the day to $\mathsf{LT} - 1$.

[1] For the sake of readability we dropped the j dependence on LT and SL.



(a) (b) (c) (d) (e)

Fig. 2. Linear discrete choice model of (a) two products; (b) one aging product where the price is not discounted according to the quality; (c) two products where the cheapest one is discounted when expiring; (d) 2 product with 4-fold and 2-fold decomposed lines and the age-based substitution approximation zone. The beta distribution is provided to show the area of the requests per product. Figure (e) shows the legend of the plots.

Action space The decision variable is the daily order quantity for each product category. When the order is placed, it is subject to the entire lead time, thus, if the products are either A or B, the time t action can be

identified as

$$x_t = [O^{\mathsf{LT_A}}_{t,A}, O^{\mathsf{LT_B}}_{t,B}]. \quad (5)$$

It follows that the policy π we want to design will be

$$x_t = X^{\pi}(S_t). \quad (6)$$

Reward function We deal with a sequential optimization problem that considers an offline learning approach, where the single reward is myopic due to the gap between the moment when the items are purchased and the one where they are actually sold. In fact, the seasonality effect introduced in Table 1 produces a quantity gap between what is sold on a day and what it is bought to be prepared for the next days. On average this leads to negative reward on days before weekends and substantially favourable ones during them. When the action is selected, the retailer pays $\sum_j c_j O^{\mathsf{LT}}_{t,j}$ and earns from the sold items of the day $\sum_j p_j L_{t,j}(S_{t,j})$, where $L_{t,j}(S_{t,j})$ indicates the sold items of category j at time t, subject to the available inventory constraint and governed by the demand uncertainty. [2] The consequent reward will be

$$C(x_t, L_t) = \sum_j \left(p_j L_{t,j}(S_{t,j}) - c_j O^{\mathsf{LT}}_{t,j}\right). \quad (7)$$

Domain of the variables The domain of the variables depends on how large we set the maximum quantity per day. Assuming a Poisson distribution, we set as upper bound to the number of daily consumers $4\max_k(\lambda_k)$. Roughly speaking, we add 3 standard deviations on the expected value of the distribution on its higher weekly factor. We further multiply this value by the theoretical market share of the products, analytically computed by the cumulative density function F of the beta distribution and the linear utility method. Therefore,

$$\mathsf{UB}_j = 4 \max_k \lambda_k \cdot (F(\theta_{j+1}) - F(\theta_j)) \quad \forall j \in \{\text{A,B}\},$$

where θ_j indicates the switch-preference point. The details of this calculation are available in (Pan and Honhon, 2012). The lower bound will instead be $\mathsf{LB}_A = \mathsf{LB}_B = 0$. These two bounds define the action space as:

$$\mathsf{LB}_j \leq x_{t,j} \leq \mathsf{UB}_j \quad \forall t, \forall j \in \{\text{A,B}\}.$$

The code of the simulation environment is available at: `https://github.com/DanieleGioia/PerishableDCM`.

3.2 Policies

Order-Up-To policy Our benchmark follows the simple and popular *Order-Up-To* policy, extensively studied in (Haijema and Minner, 2019) when employed on perishable products, adapted to the seasonality as

$$X^{\mathsf{OuT}}_j(S_t) = \left(z_{j,k} - (\sum_d O^d_{t,j} + I^d_{t,j})\right) \quad \forall j \in \{\text{A,B}\},$$

where we order by comparing a week-day dependent trained variable $z_{j,k}$ to the current inventory and to the queue of orders, maximizing the following objective function

$$\sum_t C(x_t, L_t) = \sum_t C(X^{\mathsf{OuT}}_A(S_t), X^{\mathsf{OuT}}_B(S_t), L_t),$$

evaluated over the whole training horizon for each $z_{j,k}$. The consequent dimension of the global optimization problems is 7 times the number of products. We employ the `PySOT` surrogate optimization environment of (Eriksson et al., 2019), implementing a *Stochastic Radial Basis Function* optimization strategy from (Regis and Shoemaker, 2007).

[2] The subscript j ranges from 1 to $\bar{J}$, thus considering any discount policy on c_j.

Table 3. Test results on 5 different seeds.

	Policy	Avg profit: mean±std	Avg waste: mean±std
No discount strategies	OuT	585.43±12.31	3.03±1.64
	SAC	600.59±2.14	3.17±0.82
Discount strategy on B	OuT	584.48±17.36	2.35±0.36
	SAC	600.52±1.00	2.51±0.62

Soft Actor-Critic policy The Soft Actor-Critic technique (Haarnoja et al., 2018) uses artificial neural networks to approximate both a value function that provides the value of each action acting greedily and a direct policy. The term "soft" comes from the entropy approach employed in updating the value function and the Q-function. SAC also maximizes the expected entropy of the policy and the objective function, in order to explore the set of possible actions better. The policy networks encourage exploration, which increases the probability to escape from local minima. In this work we employ a multilayer perceptron (MLP) architecture of the policy and the Q networks.

4. EXPERIMENTAL VERIFICATION

We consider two metrics to evaluate the policies performance: the average profit per day and the average number of scrapped items per day, selected for their managerial and environmental value. We first learn the policies by a simulation approach, stopped when a sufficient tolerance is reached for both the methods. The metrics are successively computed in an out-of-sample test of 200 weeks, repeating the experiments with five different seeds. The parameters assumed for the products are those presented in Table 2, with respect to which we investigate the discount strategy effect introduced in Sec. 2 and the performance of SAC and OuT. In Table 3 we observe that:

- There is a decrease in the waste quantity when discount strategies are applied, whereas the profit remains stable.
- The SAC policy outperforms OuT in terms of average profit per day and maintains a lower variability between simulations.

This latter point raises questions about the shape of the two policies. We investigate this issue by inspecting the size of the orders per product in the first month of the test. For the sake of brevity, we report only the discounted scenario in Fig 3, where two entirely different strategies arise. On the one hand, SAC maintains a constant quantity of the A product in the queue, exploiting the B product only to fulfill the weekend peaks. So, in short, it uses the higher shelf life of A, hedging the stock-out risk. On the other hand, OuT follows an opposite strategy, where A is ordered in large quantities right on time for the weekends, while B is ordered without peaks, thus exploiting the shorter lead time and suffering from its shelf life. Under both approaches, doing the math of the linear discrete choice model for the selected beta distribution, we notice that the retailer forces the substitution of B with A. The original

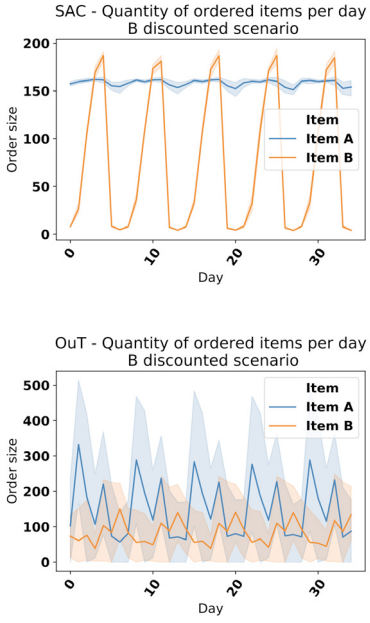


Fig. 3. Size of the orders on the first month of the out-of-sample test: OuT and SAC policies. Confidence intervals (.95) are provided w.r.t. 5 different simulations.

demand for B is approximately 1.5 times the A one, but the vendor prefers A due to a higher margin and a higher shelf life.

5. CONCLUSIONS

We apply a linear discrete choice model to an inventory management problem in a multi-item setting where stock-out substitution occurs, allowing for fixed deterioration, lead times, and seasonality. We investigate two different policies to optimize the replenishment strategy of the inventory and, although the performances are similar, the size of the orders is altogether different. Specifically, the Soft Actor-Critic method achieves a more stable policy with better average daily profit and similar waste levels. Further investigation is required for reinforcement learning approaches to be used as production solution, improving the methods by exploiting the problem structure, but either way, the results are promising. Future works will further explore the value of the lead time and the shelf life on the retailer's choices and the combined effect of these features with the other economics. Moreover, experiments with more than two products will be researched as well.

ACKNOWLEDGEMENTS

This research was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES - Coordination for the Improvement of Higher Education Personnel, Finance Code 001, grant 88882.333380/2019-01), Brazil

REFERENCES

Brandimarte, P. (2021). *From Shortest Paths to Reinforcement Learning: A MATLAB-based Tutorial on Dynamic Programming*. Springer.

Buismann, M., Haijema, R., and Hendrix, E. (2020). Retailer replenishment policies with one-way consumer-based substitution to increase profit and reduce food waste. *Logistics Research*, 13.

Eriksson, D., Bindel, D., and Shoemaker, C.A. (2019). pysot and poap: An event-driven asynchronous framework for surrogate optimization. *arXiv preprint arXiv:1908.00420*.

Fisher, M. and Raman, A. (2010). *The new science of retailing: how analytics are transforming the supply chain and improving performance*. Harvard Business Review Press.

Goyal, S. and Giri, B. (2001). Recent trends in modeling of deteriorating inventory. *European Journal of Operational Research*, 134(1), 1–16.

Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, 1861–1870. PMLR.

Haijema, R. and Minner, S. (2019). Improved ordering of perishables: The value of stock-age information. *International Journal of Production Economics*, 209, 316–324.

Haijema, R., van der Wal, J., and van Dijk, N.M. (2005). Blood platelet production: a multi-type perishable inventory problem. In H. Fleuren, D. den Hertog, and P. Kort (eds.), *Operations Research Proceedings 2004*, 84–92. Springer Berlin Heidelberg, Berlin, Heidelberg.

Hendrix, E., Ortega Lopez, G., Haijema, R., Buisman, M., and García Fernandez, I. (2019). On computing optimal policies in perishable inventory control using value iteration. *Computational and Mathematical Methods*, 1.

Janssen, L., Claus, T., and Sauer, J. (2016). Literature review of deteriorating inventory models by key topics from 2012 to 2015. *International Journal of Production Economics*, 182, 86–112.

Pan, X.A. and Honhon, D. (2012). Assortment planning for vertically differentiated products. *Production and Operations Management*, 21(2), 253–275.

Powell, W.B. (2021). *Reinforcement Learning and Stochastic Optimization: A Unified Framework for Sequential Decisions*. Wiley.

Regis, R.G. and Shoemaker, C.A. (2007). A stochastic radial basis function method for the global optimization of expensive functions. *INFORMS Journal on Computing*, 19(4), 497–509.

Shin, H., Park, S., Lee, E., and Benton, W. (2015). A classification of the literature on the planning of substitutable products. *European Journal of Operational Research*, 246(3), 686–699.

Transchel, S. (2017). Inventory management under price-based and stockout-based substitution. *European Journal of Operational Research*, 262(3), 996–1008.

Transchel, S., Buisman, M.E., and Haijema, R. (2021). Joint assortment and inventory optimization for vertically differentiated products under consumer-driven substitution. *European Journal of Operational Research*.

Van Woensel, T., van Donselaar, K., Broekmeulen, R., and Fransoo, J. (2007). Consumer responses to shelf out-of-stocks of perishable products. *International Journal of Physical Distribution & Logistics Management*, 37, 704–718.