

Designing Surprise Bags for Surplus Foods

Fan Zhou

Shenzhen Finance Institute, School of Management and Economics, The Chinese University of Hong Kong, Shenzhen, zhoufan@cuhk.edu.cn

Hansheng Jiang

Rotman School of Management, University of Toronto, hansheng.jiang@rotman.utoronto.ca

Andrea Li

Too Good To Go, andreali@toogoodtogo.ca

Joline Uichanco

Tandon School of Engineering, New York University, ju441@nyu.edu

Abstract. *Problem definition:* Surplus-food marketplaces, such as Too Good To Go, enable retailers to monetize leftover inventory and reduce food waste by selling discounted “surprise bags.” Participating stores, however, face complex operational challenges. They must jointly set a reservation price and the number of bags made available before daily demand and end-of-day surplus are realized, while customers’ *ex-post* experiences affect future demand through the platform’s rating system. Moreover, although evenly distributing surplus across bags is intuitive and operationally convenient, it remains unclear whether bag-value differentiation can improve consumer experience and profitability. *Methodology/results:* We study these questions through an infinite-horizon two-stage stochastic dynamic program. In each period, the store first sets a reservation price per bag and the number of bags available and then, after observing accepted reservations and surplus, allocates surplus food across surprise bags, potentially supplementing from regular inventory. We show that this infinite-dimensional allocation problem can be reduced to a scalar choice of average bag value. Specifically, the optimal allocation is characterized by the upper concave envelope of the consumer utility function and requires at most two bag types. Building on this structural reduction, we develop linear value-function and deterministic fluid approximations that provide explicit performance guarantees and transparent implementation rules. The linear value-function approximation captures most of the dynamic-program profit improvement and produces the lowest average modeled total waste across the main grid, while the deterministic fluid approximation yields transparent steady-state comparative statics. *Managerial implications:* Price and the reservation cap are not interchangeable: price shifts latent requests and customer net utility, whereas the cap truncates high-request realizations before surplus is known. This division of control provides a basis for platform decision-support tools that balance retailer profit with food recovery.

Key words: food surprise bag, food waste, reputation, platforms, dynamic programming

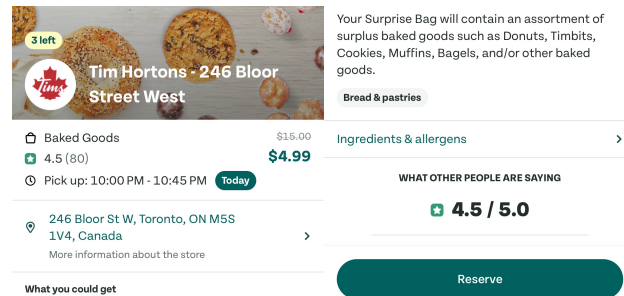
1. Introduction

Over one billion tons of food are wasted globally each year, contributing to roughly 10% of human-caused greenhouse gas emissions and generating economic losses exceeding one trillion U.S. dollars (United Nations 2024, WWF 2024). A large share of this waste is generated downstream, at grocery stores and restaurants, where demand uncertainty and perishability jointly result in unsold items near the end of the selling day. For example, in 2023 U.S. food retailers generated 4.45 million tons of surplus food, while foodservice operators generated 12.7 million tons (ReFED 2025). These figures underscore both the environmental stakes and the operational opportunity: improved mechanisms for reallocating surplus can simultaneously reduce waste and create value for firms and consumers.

One popular mechanism is the surplus-food marketplace platforms that connect consumers with retailers’ unsold food at discounted prices. Prominent examples include Too Good To Go (TGTG), Flashfood, treat-

sure, Phenix, and Tabete (Toeniskoetter 2022). Since its launch in 2016, TGTG has become a global leader in this space, connecting over 170,000 stores and restaurants with more than 100 million users across 18 countries and reporting over 350 million meals saved (TGTG 2024). It operates as a two-sided marketplace: retailers list their surplus inventory as *surprise bags* through the platform, while consumers browse, reserve, and pay for them via the TGTG mobile app. Each surprise bag is a discounted bundle of a store's end-of-day surplus, whose exact contents are unknown at the time of reservation and revealed only at pickup. Figure 1 is a typical listing on the app, showing that the only information known to the customer is the store (Tim Hortons), the bag type (baked goods), the reservation price (\$4.99), the promised value of the contents (\$15.00), the number of bags left (3), the store rating on the platform (4.5 stars), and the pickup time.

Figure 1 Too Good To Go user interface screenshot for a cafe (upper and lower portions shown left to right).



TGTG's model creates value across the supply chain: food retail stores gain a new revenue channel for food that would otherwise be discarded, consumers enjoy deeply discounted food, and society benefits from the reduction in food waste. Our 2024 semi-structured interviews with two store operators likewise suggest that cannibalization of regular sales is limited (interview notes are available from the authors). Because surprise bags are opaque, residual-inventory products with uncertain contents and limited availability, the app is more likely to attract casual shoppers, whereas regular customers who value a reliable channel for specific items are better served through regular sales. Practitioner accounts likewise report limited concerns about cannibalization (see, e.g., Schneider 2023, Too Good To Go 2025a,b). This is also consistent with qualitative evidence that selling leftovers through a separate channel or platform can help salvage surplus without materially affecting the original sales channel (Riesenegger and Hübner 2022).

1.1. Operational Challenges and Research Questions

Surprise bags pose an operational challenge for retailers. The core difficulty is one of commitment under two-sided uncertainty: a store must set both the reservation price and the number of bags available (the reservation cap) before knowing either how many customers will request a bag or how much surplus will remain at closing, and must allocate the surplus to the reserved bags once the day's sales are complete. These decisions directly shape customer experience: the price paid and the promised value set natural benchmarks

for value, and customers evaluate the bag contents against those benchmarks when deciding to rate the store. Since TGTG’s platform aggregates these ratings and makes them highly visible to prospective buyers (see [Figure 1](#)), today’s reservation price, reservation cap, and the surplus allocation directly influence tomorrow’s demand (via ratings), creating an intertemporal tension between short-run profit and long-run reputation building. The operators we interviewed described static rules of thumb, including pricing near the platform benchmark and allocating similar value across bags. Neither reported explicitly incorporating the consequences of these decisions for future ratings, despite ratings being highly salient to consumers and strongly influencing future sales. Specifically, the store’s decisions have the following tradeoffs:

- i) *Reservation control under reputation.* Reservations are jointly shaped by the store’s rating, its bag price, and its reservation cap. A lower price attracts more requests and improves net customer utility, but may reduce margin. A larger reservation cap allows the store to serve more customers, but increases the risk that realized surplus must be spread too thinly or supplemented from regular inventory at cost. Bag price and the reservation cap are therefore distinct but coordinated controls: price affects both who wants to buy and how buyers evaluate the transaction, whereas the reservation cap determines how many requests the store accepts.

Because TGTG users are often casual or new discovery shoppers at a given store, the store’s rating on the platform serves as the primary signal of quality available to prospective buyers. When quality is unobservable before purchase, reputation is theoretically predicted to be the key mechanism through which consumers assess quality and make purchase decisions ([Shapiro 1983](#)). Empirically, seller ratings have been shown to significantly affect buyer behavior on platform markets ([Nosko and Tadelis 2015](#)). Consistent with this, users in the Reddit community [r/toogoodtogo](#) report applying explicit rating thresholds when deciding whether to reserve, for instance, “below 4.5 makes me pause, below 4.3 is an absolute no” ([r/toogoodtogo 2024](#)), making ratings a first-order driver of demand on TGTG.

- ii) *Supplementation decisions when surplus is insufficient.* If realized surplus is insufficient to deliver satisfactory bags, the store can supplement from regular inventory to avoid disappointing customers. Such supplementation decisions improve consumer experience and protect future demand, but they incur costs that may exceed the margins from surprise-bag sales.
- iii) *Allocation of food across bags.* After surplus is realized and potentially supplemented, the store must decide how to distribute food across all reserved bags. Both operators described even splitting as their usual practice, so that each bag has roughly the same monetary value. While intuitive and easy to implement, equal allocation ignores how customers evaluate their bag relative to what they paid and what was promised, and therefore can be suboptimal. Consider a store that promises a bag value of \$30 and charges a reservation price of \$10, with two reserved bags and surplus totaling \$120 in original retail value, but due to perishability has a *quality-adjusted* value of \$40. For illustration, suppose customers derive positive utility only if quality-adjusted value reaches the benchmark of \$30. (The

example is deliberately stark; the model in the paper allows customer valuation to be nonlinear and reference-dependent.) Equal splitting gives each customer quality-adjusted value \$20, yielding zero utility for both customers. A differentiated allocation produces quality-adjusted values \$10 and \$30, yielding positive utility for the second customer and making differentiation strictly preferable to equal splitting. This raises the question of whether a simple *differentiated* allocation can improve customer experience while remaining practical to execute.

These operational challenges give rise to three research questions: i) *Reservation control*: How should a store jointly set its reservation price and the reservation cap as a function of its current rating, balancing current margin, demand uncertainty, fulfillment costs, and future reputation? ii) *Supplementation*: Given realized requests and surplus food value, how should the store choose the supplemented food value, weighing the costs of supplementing against the reputational benefit of meeting customer expectations? iii) *Allocation*: Given the total surplus and supplemented food value, how should the store distribute the value across reserved bags to maximize overall customer satisfaction?

1.2. Our Contributions

We develop an infinite-horizon stochastic dynamic program that captures the interaction among pricing, the reservation cap, bag design, and reputation. In each period, a store faces a two-stage decision problem: it first sets a reservation price and reservation cap as a function of its current rating (Stage I), which determine latent customer requests and accepted reservations; then, after accepted reservations and end-of-day surplus are realized (Stage II), the store chooses the total value delivered to customers (potentially supplementing from regular inventory) and allocates that value across bags. Consumer satisfaction feeds into next period's rating through an exponential-smoothing update rule, generating an intertemporal trade-off between short-run profit and long-run reputation building. Our main findings are as follows.

A tractable characterization and operational guideline for within-day allocation. Conditional on a target average bag value ℓ , the optimal allocation achieves an average customer valuation equal to $\hat{v}(\ell)$, where $\hat{v}(\cdot)$ is the *upper concave envelope* of the consumer valuation function $v(\cdot)$, and can be implemented with at most two bag types. This reduces the Stage-II allocation from a distributional choice to a scalar choice of the average-value ℓ . If $\hat{v}(\ell) > v(\ell)$, a two-type design strictly increases expected satisfaction; if v is concave on the feasible support, equal splitting is optimal.

Dynamic reservation control and bag value under evolving ratings. Embedding the within-day optimal allocation result into the intertemporal problem yields a two-stage stochastic dynamic program in which the store's rating evolves endogenously with customer satisfaction. Price shifts latent requests and customer net utility, whereas the reservation cap truncates only high-request realizations before surplus is known. Under a linear approximation of the value function, it is optimal to open slots (by increasing the reservation cap) until the optimal price equals the expected marginal scarcity cost of another accepted

reservation. In the optimal deterministic steady state, by contrast, a strictly binding cap is dominated by a small price increase. This contrast identifies stochastic demand spikes and a binding price constraint as the settings in which the reservation cap has a distinct role. Conditional on accepted reservations and realized surplus, the optimal bag-design structure remains unchanged.

Practical approximation policies and steady-state insights. We develop two approximation policies with provable performance bounds. The linear value function approximation (LVA) maps the long-run value of reputation into an explicit within-period incentive weight, yielding a simple “use surplus first; supplement only if it pays” rule and a low-dimensional joint price-cap problem. The deterministic fluid approximation replaces within-period uncertainty with mean demand and mean surplus, enabling steady-state comparative statics through a responsiveness index that combines store patience and rating-system responsiveness.

Numerical evidence on profitability and waste reduction. We evaluate the policies through numerical experiments calibrated with a prospect-theory utility function. In the numerical study, the envelope-based static policy raises discounted per-period profit by 5.5% and reduces modeled total waste by 27.1%; LVA raises profit by 28.2% and delivers the lowest waste, 66.7% below Baseline; and the full dynamic program raises profit by 34.0% while reducing waste by 64.0%. These comparisons separate two layers of operational value: envelope allocation can improve on uniform allocation, whereas joint dynamic control of the price, the reservation cap, and the target allocation value can add further profit and waste reductions. They also reveal a profit-waste frontier: the most profitable policy is not always the lowest-waste policy. When the store also chooses the promised value, DP and LVA both lower the promised value, price, and average bag value and raise the reservation cap relative to the fixed-promise benchmark, but only DP produces a statistically distinguishable reduction in total waste.

Although our model takes retailer profit as the objective, the analysis also informs platform-level decision support because platforms benefit from store participation and consumer satisfaction. The results suggest that store-facing tools should coordinate reservation price, reservation caps, and bag design, while recognizing that rating-system responsiveness changes the strength of long-run quality incentives.

2. Literature Review

In recent years, the agricultural and food industries have attracted substantial scholarly attention, particularly regarding sustainability and efficiency challenges in food supply chains (e.g., [Murali et al. 2019](#), [Boyabatli et al. 2021](#), [Dong et al. 2022](#), [Adebola et al. 2022](#), [Agrawal and Zhang 2026](#)). Within operations management, food waste has become a central concern. For example, [Akkaş and Gaur \(2022\)](#) outline a comprehensive research agenda to address this issue.

A large literature examines the drivers of food waste and potential mitigation strategies. One prominent source is waste from food that remains nutritionally viable but lacks aesthetic appeal. This stream of research studies the marketing of “ugly” produce and shows that selling cosmetically imperfect food

can improve both profitability and waste reduction (Han et al. 2023, Kazaz et al. 2024). Relatedly, Crama et al. (2023) show that stringent cosmetic standards can generate substantial food loss. Another source is “multiple-dates-led expiration waste,” which occurs when products with different freshness levels are displayed together, increasing the likelihood that near-expiration items go unsold. Prior work has examined this issue from both empirical and analytical perspectives, focusing on display and inventory strategies that reduce such waste (Jain et al. 2023, Atan et al. 2025). Spoilage during transportation is another important contributor, motivating research on shipping policies that reduce waste before products reach retailers (see, e.g., Akkaş and Honhon 2022, 2023). Donating surplus food to food banks is another established mitigation strategy with economic benefits. Lowrey et al. (2023) show that food donations not only reduce waste but also increase food prices and store profits. Related operational work studies how to improve food bank performance through better promotion design (Ahire and Pekgün 2018), staffing (Ata et al. 2019), and distributional fairness (Alptekinoglu and Benade 2024). Emerging business models and technologies also contribute to more sustainable and waste-conscious food retail. For example, research on online grocery platforms shows that subscription delivery fees encourage smaller, more frequent orders, reducing household food waste and increasing retailer revenues (Belavina et al. 2017). Other studies examine promotional mechanisms such as buy-one-get-one-free offers and their effect on store waste (Wu and Honhon 2023). Related research compares centralized and decentralized food-safety audits (Dong et al. 2022). Technological innovations, including blockchain (Keskin et al. 2024) and artificial intelligence (Nu et al. 2024), have likewise gained traction as tools to improve transparency, monitor spoilage, and optimize supply chain operations.

Our work, motivated by Too Good To Go, is closely related to Yang and Yu (2025), who study the interaction between sellers’ production decisions and consumer purchases in surplus-food clearance sales. These sales occur after regular hours and target a specific subset of customers. They show that surprise-bag clearance sales can increase store profits and reduce in-store food waste, but may also raise consumer-side food waste when customers receive more food than they need. Recent empirical evidence from Dong et al. (2025) further shows that pricing matters on Too Good To Go: rigid discount policies can reduce seller participation and limit the platform’s ability to match supply and demand.

We study store decision-making on the platform from a different angle. Specifically, we model endogenous reputation dynamics, shaped by past customer satisfaction and influencing future demand, and develop a dynamic framework capturing the trade-off between current payoffs and long-run reputation building. We allow for vertically differentiated surprise bags within a day and introduce a generalized consumer utility function to characterize the optimal (potentially uneven) allocation of leftover food across bags.

Beyond food waste, this paper contributes to the literature on probabilistic or opaque goods, whose attributes consumers cannot fully observe before purchase (see, e.g., Fay and Xie 2008, Jerath et al. 2010, Zhang et al. 2015, Elmachtoub and Hamilton 2021, Chen et al. 2021, Fay and Gheibi 2024). Our study

lies at the intersection of opaque selling and reputation management. In this literature, consumers' learning about the quality of probabilistic goods is often described as "anecdotal reasoning" (Huang and Yu 2014). Building on this idea, Huang and Yin (2021) study a two-period model of probabilistic selling with vertically differentiated goods. They compare opaque and transparent selling and show that optimal first-period opaque selling assigns a higher probability to high-quality goods. Our work also relates to the use of opaque selling to clear leftover inventory. In addition to Yang and Yu (2025), Ren and Huang (2022) study this problem and similarly find that, relative to transparent selling, opaque selling of leftover inventory increases profits and supports higher prices for high-quality products.

Finally, this paper contributes to the literature on reputation management. Bar-Isaac and Tadelis (2008) provide a broad overview of this area. Within rating systems, prior studies examine how pricing shapes reputation, both theoretically (e.g., Carnehl et al. 2024, Shin et al. 2023) and empirically (e.g., Cabral and Hortacsu 2010, Luca and Reshef 2021). Our paper focuses on the interaction between reputation dynamics and the store's dynamic quality choice, which remains underexplored in the operations management literature. To the best of our knowledge, this paper is among the very few studies examining opaque selling as a tool to reduce food waste, and the first to study the optimal dynamic design of probabilistic goods for food-waste reduction under evolving reputation.

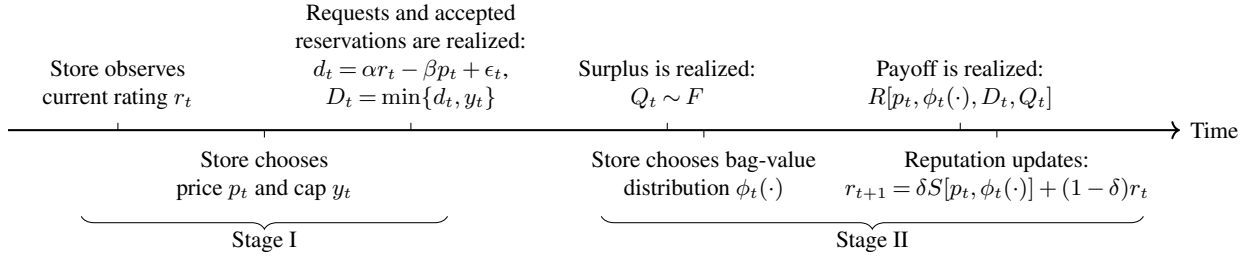
3. Model Setup

We study a profit-maximizing food retailer (hereafter, the *store*) that sells its surplus through a surplus-food platform by offering surprise bags. Time is discrete, with each period $t = 0, 1, 2, \dots$ corresponding to an active reservation-and-collection cycle (i.e., typically one day in which the store lists bags), in which the store first posts a price and a reservation cap, accepts reservations up to that cap, and later fulfills those reservations with end-of-cycle surplus. Store nonparticipation is outside the modeled active cycle.

Each period t unfolds in two stages:

- (1) **Stage I:** The store observes its current platform rating r_t and jointly chooses a reservation price p_t and a reservation cap y_t . Customers generate latent requests (or demand) d_t , and of those the store accepts $D_t = \min\{d_t, y_t\}$ reservations. Note that in this paper we use "demand" and "requests" interchangeably to refer to d_t .
- (2) **Stage II:** After the end-of-day surplus Q_t is observed, the store decides how to allocate food across the D_t accepted reservations and whether to supplement from regular inventory. Each bag is characterized by its *food value* x , a quality-adjusted summary of the quantity, freshness, and mix of items packed into the bag, expressed in retail-value-equivalent units. The allocation decision is represented through a bag-value distribution $\phi_t(\cdot)$, where $\phi_t(x)$ denotes the proportion of bags assigned food value x .

This two-stage decision process is modeled after the platforms operational structure. In practice, customers require lead time to discover available bags on the app and travel to the store. Therefore, stores must

Figure 2 Sequence of events in period t with two stages

announce prices and post reservation slots hours before closing (Stage I). By contrast, actual surplus is realized and observed only at the end of the day (Stage II).

The store's *state variable* is its current platform rating r_t , which summarizes all information relevant to future demand. Customers collect their bags at the end of each cycle, and their average satisfaction (determined by the bag contents relative to the price paid and the promised value) feeds into a rating update that sets r_{t+1} , influencing future demand. Since today's pricing, reservation-cap, and allocation decisions shape tomorrow's reputation and demand, the store's problem has an inherently dynamic structure, which we model as an infinite-horizon discounted dynamic program.

The sequence of events within one reservation-and-collection cycle is summarized in Figure 2. The rest of this section formalizes each component in turn, beginning with Stage II, which introduces the core primitives of food value and bag design, before turning to consumer utility and reputation dynamics, and finally Stage I.

3.1. Stage II: Surplus, Bag Design, and Supplementation

After reservations have been confirmed and regular sales have concluded, the store enters Stage II with D_t reserved bags to fulfill, where D_t is the realized number of accepted reservations from Stage I, formally defined in Section 3.3. Moreover, the store observes the total value of surplus that it can use to fulfill the reserved bags. Surplus food (or leftover food, used interchangeably) varies not only in *quantity* but also in attributes such as freshness, cosmetic condition, and product mix. Let Q_t denote its total quality-adjusted food value in period t , expressed in retail-value-equivalent units, where $Q_t \sim F$, F is time-stationary, and $Q_t \geq 0$ almost surely.

We treat the regular-channel's starting inventory and daily demand as exogenous to the store's decisions on the platform. Their net outcome is summarized by the end-of-cycle surplus Q_t . Thus the store's platform problem only needs to track Q_t , while interactions with the regular channel enter through the surplus distribution F and the supplementation cost c . We assume $\{Q_t\}$ is i.i.d. across active listing cycles.

Bag design. Let x denote an individual bag's quality-adjusted food value, measured in dollars. Unlike the sum of the items' original retail values, x reflects the bag's quantity, freshness, condition, and product mix. Under perfect divisibility, let $\mathcal{X} = [\underline{x}, \bar{x}]$ be the feasible range of x . The food allocation decision is

represented by a bag-value distribution $\phi_t(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^+$, where $\phi_t(x)$ is the proportion of bags in period t allocated food value x , normalized so that

$$\int_{\mathcal{X}} \phi_t(x) dx = 1.$$

We write ϕ_t as a density for notational convenience. The results also allow ϕ_t to be supported on finitely many points (e.g., a two-point distribution corresponds to offering two types of bags).

Supplementation. Our model allows the store to deliver a total food value, $D_t \cdot \int_{\mathcal{X}} x \phi_t(x) dx$, that exceeds the total surplus Q_t . Where permitted by the platform, the store can *supplement* the food surplus with items prepared or stocked through its normal operations and diverted from the regular sales channel. We model the marginal *shadow cost* of supplying one additional unit of retail-value-equivalent beyond realized surplus as $c \geq 0$, giving rise to the total supplementation cost

$$C[\phi_t(\cdot), D_t, Q_t] = c \cdot \left(D_t \cdot \int_{\mathcal{X}} x \phi_t(x) dx - Q_t \right)^+. \quad (1)$$

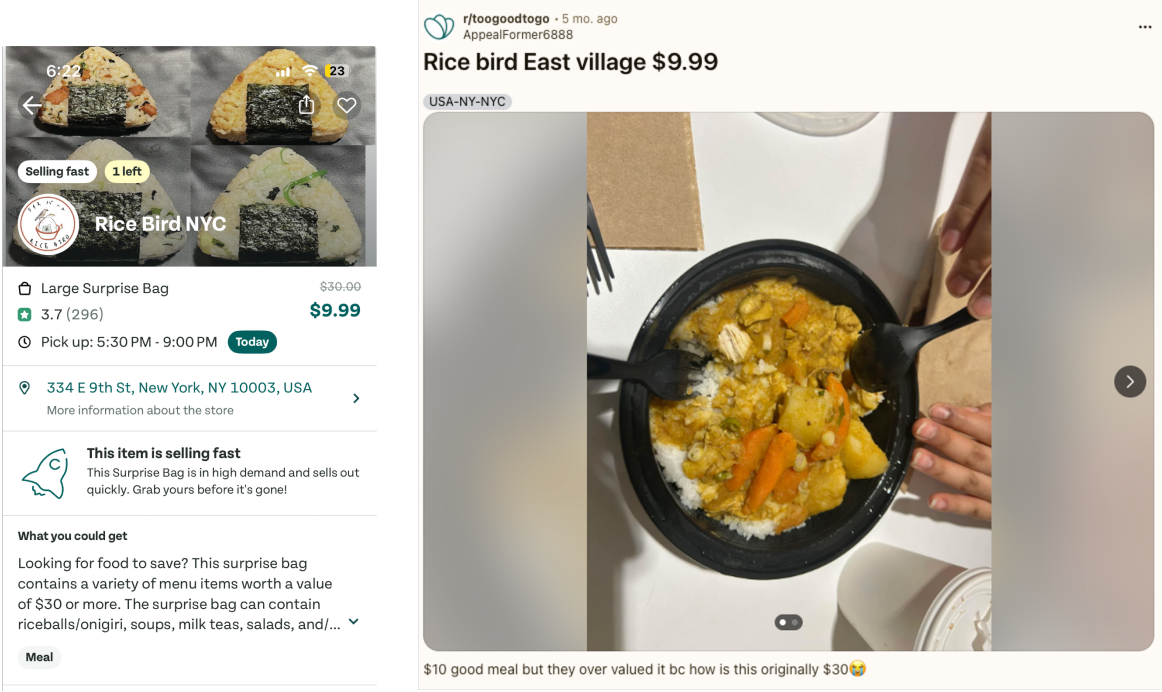
REMARK 1 (SUPPLEMENTATION COST c). Because food value is expressed in retail-value-equivalent units, c is a dimensionless shadow cost *per unit of food value* supplied beyond realized surplus. In the baseline calibration we focus on $c \in [0, 1]$, meaning that one unit of added food value costs at most one unit in shadow cost. Operationally, c captures the incremental handling cost and forgone profit from diverting food from the regular sales channel when realized surplus Q_t falls short of the intended total delivered value.

The size of c depends on the diverted food's alternative use. It is lower when the store draws from prepared or stocked items with little expected regular-channel value, and higher when it diverts retail-ready inventory likely to sell at full price. In the latter case, the shadow cost also includes possible stockout risk and replenishment frictions.

The linear form is appropriate when marginal opportunity cost is approximately constant. A convex cost of the total shortfall still depends on the bag-value distribution only through its mean, so the fixed-mean envelope and its at-most-two-bag-type implementation remain valid; the supplementation and price-cap rules change, as discussed in [Section 7](#). The dynamic monotonicity result below additionally assumes $\bar{p} \geq c\bar{x}$.

3.2. Customer Utility and Reputation Dynamics

Surprise bags are distributed via a blind, first-come, first-served mechanism, so stores do not customize bag contents or assignments based on consumer profiles. After customers collect their bags, they rate their experience on the platform. TGTG displays to prospective shoppers an aggregate store rating over a certain aggregation window, for example, claiming that “the score is based on ratings from the past two months”.

Figure 3 A Reddit post about a Too Good To Go surprise bag sold at \$10 with a promised value of \$30.

(a) A store listing

(b) A consumer complains on Reddit after perceiving the value as less than \$30

Consumer utility. Consumers' valuation of receiving a bag with food value x is given by a continuous, non-decreasing function $v(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$. We interpret $v(x) = \mathbb{E}_\theta[v_\theta(x)]$ as the population-average valuation, where θ indexes heterogeneous consumers. Because bags are randomly assigned and the store cannot condition bag value on θ , expected satisfaction depends on the allocation through this average function. We treat the served-customer mix, and hence v , as stable over time. If types were observable at packing or consumers could select among bag types, the allocation problem would instead need to condition on type. This restriction is consistent with the **TGTG Store Terms** (Section 8.2), under which consumer data are not shared with stores without prior consent or a regulatory requirement.

Customers evaluate realized quality-adjusted food value x relative to the promised value in the listing. **Figure 3** shows a \$30 original retail value against a \$9.99 reservation price and a customer perceiving the bag's value as below \$30. We allow such perceived shortfalls: lower x reduces customer valuation and, through ratings, future demand. This reference dependence motivates the nonlinear, potentially non-concave $v(\cdot)$ in Example 1, consistent with prospect theory (Kahneman and Tversky 1979).

EXAMPLE 1. Drawing on prospect theory (Kahneman and Tversky 1979), let $u(z)$ denote consumers' perceived value of a gain or loss z relative to a reference point. A widely used specification (Tversky and Kahneman 1992) is

$$u(z) = z^\eta \cdot \mathbf{1}_{z \geq 0} - \lambda(-z)^\eta \cdot \mathbf{1}_{z < 0},$$

where $\eta \in (0, 1)$ captures diminishing sensitivity to gains/losses and $\lambda > 1$ captures loss aversion, the tendency to weigh losses more heavily than equivalent gains. Let h denote the original retail value promised in the listing (\$30 in Figure 3). Because x is quality-adjusted, it may be below h even when the bag meets its promised original retail value. In the context of surprise bags, this promised value anchors the customer's evaluation of the realized bundle. We model this by setting

$$v(x) = u(x - h). \quad (2)$$

Note that $v(\cdot)$ does not depend on the price paid p_t . This is consistent with the standard treatment in the pricing literature, where consumer valuation is defined independently of price, and net utility is the difference between valuation and price paid.

Reputation dynamics. Given a food allocation decision $\phi_t(\cdot)$ and price p_t , consumer satisfaction (S) is measured by population-average net utility, defined as the average perceived value (V) of delivered bags minus the price. Formally,

$$V[\phi_t(\cdot)] := \int_{\mathcal{X}} v(x) \phi_t(x) dx, \quad \text{and} \quad S[p_t, \phi_t(\cdot)] := V[\phi_t(\cdot)] - p_t. \quad (3)$$

Since consumer satisfaction $S[p_t, \phi_t(\cdot)]$ captures the average gap between perceived bag value and price paid, it is the natural input to the rating update. We model this through an exponential-smoothing update:

$$r_{t+1} = \delta S[p_t, \phi_t(\cdot)] + (1 - \delta)r_t, \quad (4)$$

where $\delta \in (0, 1)$ is the weight assigned to the current periods average customer satisfaction. A smaller δ corresponds to a longer effective memory in the displayed rating; for example, a score described as based on ratings from the past two months corresponds to a slower-updating rating signal. This reduced-form reputation model is widely used to represent evolving consumer beliefs and online ratings in both theoretical and empirical studies (see, e.g., Gaur and Park 2007, Yang et al. 2019, Cho et al. 2024, Carnehl et al. 2024).

REMARK 2 (AVERAGE SATISFACTION AND SELECTIVE RATINGS). Our baseline reputation update in (4) assumes that all consumers are equally likely to submit a rating and receive equal weight in the aggregate satisfaction measure. In practice, rating submission may be selective, with dissatisfied consumers who receive lower-value bags more likely to leave ratings or complaints. We capture this “unhappy consumers speak louder” effect through a selective-rating model and examine its implications numerically in Section 6.3.

3.3. Stage I: Price, Reservation Cap, and Demand

At the beginning of period t , the store observes its current platform rating r_t , the aggregate score displayed to prospective customers on the app, and chooses a reservation price $p_t \in [0, \bar{p}]$ per bag and a reservation cap $y_t \in [\underline{y}, \bar{y}]$, where $0 < \underline{y} < \bar{y}$. The cap y_t is the maximum number of bags made available during the

reservation window. Condition $\underline{y} > 0$ restricts attention to active listings; $\bar{y}$ is the maximum feasible listing quantity, and $\bar{p}$ is the platform price limit. The analytical rating state lies in the compact invariant interval $\mathcal{R} := [\underline{r}, \bar{r}]$, where $\underline{r} := v(\underline{x}) - \bar{p}$ and $\bar{r} := v(\bar{x})$; invariance follows because the update in (4) is a convex combination of the current rating and one-period satisfaction.

Consumers arrive during the reservation window to reserve bags. Let

$$d_t = d(r_t, p_t, \epsilon_t) := \alpha r_t - \beta p_t + \epsilon_t \quad (5)$$

denote latent reservation demand, that is, the number of visiting customers who would intend to reserve a bag assuming infinite availability. Here, $\alpha > 0$ and $\beta > 0$ capture sensitivity to rating and price, respectively, and $\epsilon_t \sim G$ is a mean-zero shock with $\mathbb{E}_G[\epsilon] \equiv 0$. The shocks ϵ_t and Q_t are independent within each cycle and have continuous distributions. We focus on interior optima. The relevant expected reduced objectives are differentiable on the interior, and differentiation may be interchanged with expectation. We also assume that the upper support of G is sufficiently wide for any interior reservation cap to bind with positive probability over the active state–price range. The number of *accepted* reservations is thus

$$D_t = D(r_t, p_t, y_t, \epsilon_t) := \min\{d(r_t, p_t, \epsilon_t), y_t\}. \quad (6)$$

All accepted reservations are fulfilled; store-initiated cancellation after acceptance is outside the modeled action set. We restrict attention to state–action pairs with positive latent demand over the relevant price menu and shock support. Formally, over $\mathcal{R}$ and the support of ϵ , we assume $\alpha r - \beta p + \epsilon > 0$ for all $r \in \mathcal{R}$, $p \in [0, \bar{p}]$. For analytical tractability, we treat d_t , y_t , and D_t as continuous, which can be interpreted as a fluid approximation for large-scale reservation volumes. For an exact individual-choice foundation, suppose N potential customers have i.i.d. participation costs $\nu \sim \text{Unif}[0, \bar{\nu}]$. Customers request iff $ar - bp - \nu \geq 0$, where $a, b, \bar{\nu} > 0$. For $0 < ar - bp < \bar{\nu}$, expected latent requests are $N \Pr(\nu \leq ar - bp) = (Na/\bar{\nu})r - (Nb/\bar{\nu})p$, giving $\alpha = Na/\bar{\nu}$ and $\beta = Nb/\bar{\nu}$; ϵ_t captures aggregate shocks.

REMARK 3 (DISTINCT ROLES OF PRICE AND THE RESERVATION CAP). Price and the reservation cap are imperfect substitutes for controlling accepted volume, but they have different consequences. Raising price reduces latent demand and also lowers the net utility of customers who purchase. Lowering the reservation cap truncates high-demand realizations without directly changing the price paid. The two decisions must therefore be coordinated: price manages margin, demand, and customer net utility, while the cap manages exposure to surplus scarcity.

3.4. The Store’s Optimization Problem

The store’s one-period payoff is revenue from surprise bags net of supplementation cost:

$$R_t = R[p_t, \phi_t(\cdot), D_t, Q_t] = p_t \cdot D_t - C[\phi_t(\cdot), D_t, Q_t]. \quad (7)$$

Any surplus not distributed through surprise bags becomes in-store waste, expressed as $W_t := (Q_t - D_t \cdot \int_{\mathcal{X}} x \phi_t(x) dx)^+$. Although our model does not explicitly track household waste, distributing the same surplus across more bags reduces the amount of food per household, which may increase the likelihood of timely consumption. We revisit this implication numerically in [Section 6](#).

We impose the following maintained conditions, used to establish that the optimal value function is nondecreasing in reputation:

ASSUMPTION 1. $(1 - \delta)\beta \geq \delta\alpha$, and $\bar{p} \geq c\bar{x}$.

The first condition ensures that the price increase used in [Proposition 1](#) does not lower the next-period rating. The second ensures that any additional accepted volume when the price reaches $\bar{p}$ is weakly profitable even at the maximum bag value. Together, these sufficient conditions imply that a higher rating cannot reduce the store's value.

A stationary policy π consists of a price rule $p^\pi : r \mapsto [0, \bar{p}]$ and a cap rule $y^\pi : r \mapsto [\underline{y}, \bar{y}]$ in Stage I, and a food allocation rule in Stage-II that corresponds to a valid bag-value distribution $\phi^\pi(\cdot | r, p, D, Q)$ on $\mathcal{X}$. Let Π denote the set of all such admissible stationary policies. The store chooses the policy π that maximizes its expected discounted profit:

$$\max_{\pi \in \Pi} \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t \cdot R[p_t^\pi, \phi_t^\pi(\cdot), D(r_t, p_t^\pi, y_t^\pi, \epsilon_t), Q_t] \right] \quad (8a)$$

$$\text{subject to } \int_{\mathcal{X}} \phi_t^\pi(x) dx = 1, \quad (8b)$$

$$r_{t+1} = \delta S[p_t^\pi, \phi_t^\pi(\cdot)] + (1 - \delta) r_t, \quad (8c)$$

$$\epsilon_t \sim G, Q_t \sim F, \quad (8d)$$

where $\gamma \in (0, 1)$ is the discount factor. Constraint (8b) ensures that ϕ_t^π is a valid distribution, (8c) imposes the reputation dynamics in (4), and (8d) specifies the randomness in consumer demand and surplus supply.

REMARK 4 (PROFIT OBJECTIVE AND WASTE REDUCTION). Although the objective (8a) is the store's profit from the platform, it is economically connected to the goal of reducing food waste. The platform's ability to sustain and scale food recovery depends on stores' participation incentives, so store profitability affects platform viability and the long-run volume of surplus diverted from disposal. At the same time, with an endogenous reservation cap, higher profit need not come from selling more bags and need not mechanically reduce in-store waste. The profit-waste alignment is therefore a quantitative question rather than a modeling assumption. In [Section 6](#), we numerically quantify how our proposed policies affect food waste reduction at both the store and consumer levels.

4. Structure and Properties of Optimal Allocation

The stochastic dynamic program in [Section 3](#) presents two interrelated challenges. First, each period involves a *two-stage* decision: the store posts a price and reservation cap before demand and surplus are

realized, and then assembles bags after observing accepted reservations and surplus. Second, the Stage-II bag-value distribution $\phi(\cdot)$ is an *infinite-dimensional* decision variable, and the bilinear interaction between accepted demand and average bag value in the supplementation cost function eq. (1) makes the joint objective generally non-concave.

We address these challenges through a sequence of structural reductions. Section 4.1 shows that the infinite-dimensional Stage-II decision (i.e., the bag-value distribution) can be reduced to a one-dimensional choice of the *average* bag value ℓ . Section 4.2 then characterizes the *conditional optimal distribution* for a given ℓ through a concavification argument, showing that the store needs at most two bag types to maximize average consumer satisfaction. Together, these results yield the reduced Bellman equation (18), which serves as the foundation for the approximation policies developed in Section 5.

4.1. Stage II Action Space Reduction

Let $J^*(r)$ denote the optimal expected discounted profit starting from reputation r . By the Bellman optimality condition, $J^*(\cdot)$ satisfies

$$J^*(r) = \max_{\substack{p \in [0, \bar{p}] \\ y \in [\underline{y}, \bar{y}]}} \mathbb{E}_\epsilon \left[p \cdot D(r, p, y, \epsilon) + \mathbb{E}_Q \left[\max_{\phi(\cdot)} K(r, p, y, \epsilon, Q, \phi(\cdot)) \right] \right], \quad (9)$$

where the within-period Stage II contribution K is defined as

$$K(r, p, y, \epsilon, Q, \phi(\cdot)) := -C[\phi(\cdot), D(r, p, y, \epsilon), Q] + \gamma J^*(\delta S[p, \phi(\cdot)] + (1 - \delta)r), \quad (10)$$

The two-stage decision structure is reflected directly in the Bellman equation. In Stage I, the store chooses price p and reservation cap y before observing latent demand. In Stage II, after observing both accepted reservations D and surplus Q , the store chooses the bag-value distribution $\phi(\cdot)$. The reservation-cap decision changes the distribution of realized accepted demand but does not alter the conditional Stage-II allocation problem. Once D and Q have been observed, y has no additional role.

We first turn our attention to the Stage II problem of choosing ϕ to optimize eq. (10). The key observation is that the supplementation cost $C[\phi(\cdot), D, Q] = c \cdot (D \cdot \ell(\phi) - Q)^+$ depends only on $\phi(\cdot)$ through its mean:

$$\ell(\phi) := \int_{\mathcal{X}} x \phi(x) dx \in \mathcal{X}.$$

Therefore, any two distributions with the same mean generate identical supplemental costs. Among such cost-equivalent distributions, the only difference in eq. (10) lies in the continuation value. This motivates asking: among all distributions with mean ℓ , which one maximizes the continuation value?

Answering this question requires understanding two properties: how the bag-value distribution ϕ affects next period's reputation r' , and how continuation value $J^*(r')$ depends on r' . The first is straightforward: a higher average bag valuation $V[\phi]$ raises satisfaction $S[p, \phi] = V[\phi] - p$, which raises next period's rating

through the update $r' = \delta S + (1 - \delta)r$. The second is less obvious and is established by the following proposition.

PROPOSITION 1. *Under **Assumption 1**, $J^*(r)$ is nondecreasing in r .*

A higher rating increases latent demand. At the higher rating, the store can retain the optimal cap and Stage-II allocation from the lower rating and raise the associated price by α/β times the rating increase, up to $\bar{p}$, without reducing accepted reservations. **Assumption 1** ensures that this choice yields weakly higher current payoff and continuation value, so a higher rating cannot reduce the store's value.

With J^* nondecreasing in r , the continuation value in **eq. (10)** is weakly increasing in $S[p, \phi] = V[\phi] - p$, so among all bag-value distributions with the same mean, the store weakly prefers and may select one that maximizes $V[\phi]$. This is formalized in the following corollary.

COROLLARY 1. *Fix any Stage-II state and realization (r, p, D, Q) . There exists an optimal Stage-II decision $\phi_t^*(\cdot)$ such that*

$$V[\phi_t^*(\cdot)] \geq V[\phi'_t(\cdot)], \quad \text{for any } \phi'_t(\cdot) \text{ such that } \int_{\mathcal{X}} x \phi'_t(x) dx = \int_{\mathcal{X}} x \phi_t^*(x) dx. \quad (11)$$

Corollary 1 reduces the Stage-II infinite-dimensional problem to a two-step procedure. First, choose a target average bag value $\ell \in \mathcal{X}$. Second, given ℓ , solve for the distribution $\phi^*(\cdot | \ell)$ that maximizes consumer satisfaction among all distributions with the same mean ℓ :

$$\max_{\phi(\cdot): \mathcal{X} \rightarrow \mathbb{R}_+} V[\phi] = \int_{\mathcal{X}} v(x) \phi(x) dx \quad (12a)$$

$$\text{subject to } \int_{\mathcal{X}} \phi(x) dx = 1, \quad \text{and} \quad \int_{\mathcal{X}} x \phi(x) dx = \ell. \quad (12b)$$

We refer to the optimal solution $\phi^*(\cdot | \ell)$ as the *conditional optimal food distribution*. This distribution depends only on the primitive utility function $v(\cdot)$ and the target mean ℓ , and not on the within-period information (r, p, D, Q) .

Substituting $\phi^*(\cdot | \ell)$ into **(9)**, the Bellman equation reduces to

$$J^*(r) = \max_{\substack{p \in [0, \bar{p}] \\ y \in [\underline{y}, \bar{y}]}} \mathbb{E}_{\epsilon} \left[p \cdot D(r, p, y, \epsilon) + \mathbb{E}_Q \left[\max_{\ell \in \mathcal{X}} K(r, p, y, \epsilon, Q, \ell) \right] \right], \quad (13)$$

where the Stage II search is now over $\ell \in \mathcal{X}$ to optimize the following Stage II objective:

$$K(r, p, y, \epsilon, Q, \ell) := -c \cdot (\ell D(r, p, y, \epsilon) - Q)^+ + \gamma J^*[\delta (V[\phi^*(\cdot | \ell)] - p) + (1 - \delta)r]. \quad (14)$$

Here, $\phi^*(\cdot | \ell)$ is the optimal solution to a semi-infinite linear program **(12)**: the decision variable $\phi(\cdot)$ is an infinite-dimensional object (a probability density), but the constraints are finite (a normalization constraint and a mean constraint). This may appear to offer little computational advantage, since the decision variable

remains infinite-dimensional. However, as we will establish next in [Section 4.2](#), its optimal value $V[\phi^*(\cdot | \ell)]$ can be specified analytically as a function of ℓ alone, without requiring $\phi^*(\cdot | \ell)$ to be computed explicitly. This allows the Bellman equation to be written entirely in terms of ℓ , reducing the Stage II search to a one-dimensional optimization over $\ell \in \mathcal{X}$.

4.2. Conditional Optimal Food Distribution

We now characterize the conditional optimizer $\phi^*(\cdot | \ell)$ of the semi-infinite linear problem (12). A standard duality argument implies that the optimal value $V[\phi^*(\cdot | \ell)]$ equals the *upper concave envelope* of $v(\cdot)$ evaluated at ℓ , and that $\phi^*(\cdot | \ell)$ has probability mass on at most two points. This “one- or two-point” structure is consistent with classic moment-problem results (e.g., [Scarf et al. 1957](#)) and can also be derived using the concavification approach common in information design (e.g., [Kamenica and Gentzkow 2011](#)).

PROPOSITION 2. *The optimal solution to Problem (12), $\phi^*(\cdot | \ell)$, can be chosen to be supported on at most two points in $\mathcal{X}$ and achieves*

$$V[\phi^*(\cdot | \ell)] = \hat{v}(\ell), \quad (15)$$

where $\hat{v}(\cdot)$ is the upper concave envelope of the consumer utility function $v(\cdot)$.

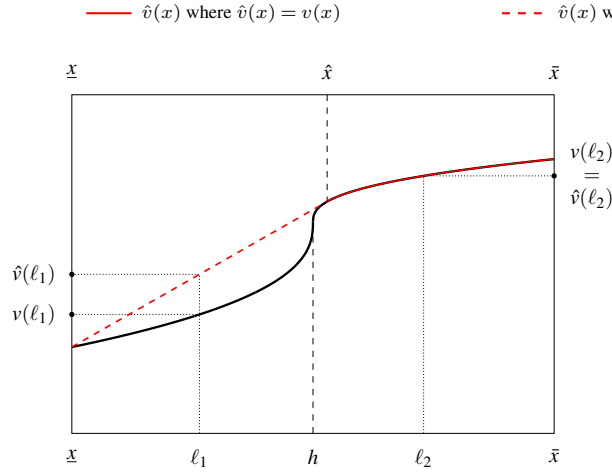
[Proposition 2](#) delivers a direct operational message: the optimal allocation requires offering at most two bag types. A single type corresponds to evenly allocating food across all bags so that each bag has the same value. A two-type allocation introduces differentiation, with some bags containing a higher food value than others. Which case arises depends on the relationship between $v(\ell)$ and $\hat{v}(\ell)$: if $v(\ell) = \hat{v}(\ell)$, then equal splitting is optimal and every bag receives value ℓ ; if $\hat{v}(\ell) > v(\ell)$, then the store should offer two bag types. The specific bag values that constitute the optimal two-type design depend on the shape of $v(\cdot)$, which in turn depends on how customers anchor on the promised value h , as the following example illustrates.

EXAMPLE 2. Recall the prospect-theory valuation function $v(x)$ introduced in [example 1](#). [Figure 4](#) plots $v(x)$ (black curve) and its upper concave envelope $\hat{v}(x)$ (red curve). As expected from the prospect-theory specification, $v(x)$ is nonlinear and non-concave in a neighborhood of the promised value h : it is concave above h , reflecting diminishing sensitivity to gains, and convex below h , reflecting diminishing sensitivity to losses. The envelope equals $v(x)$ at $x = \underline{x}$ and for all $x \geq \hat{x}$ (solid red) for some threshold $\hat{x}$, while replacing the non-concave portion on $(\underline{x}, \hat{x})$ with a linear segment (dashed red).

Depending on whether the target mean ℓ falls above or below $\hat{x}$, the optimal allocation that maximizes customer satisfaction takes one of two forms:

- a) **High target mean** ($\ell \geq \hat{x}$). Since $\hat{v}(\ell) = v(\ell)$, offering one bag type is optimal: $\phi^*(\cdot | \ell) = \Delta_\ell$, where Δ_x denotes unit mass (Dirac distribution) at x .
- b) **Low target mean** ($\underline{x} < \ell < \hat{x}$). Since $\hat{v}(\ell) > v(\ell)$, offering two bag types is optimal, with mixing weights chosen to deliver mean ℓ :

$$\phi^*(\cdot | \ell) = \frac{\hat{x} - \ell}{\hat{x} - \underline{x}} \Delta_{\underline{x}} + \frac{\ell - \underline{x}}{\hat{x} - \underline{x}} \Delta_{\hat{x}}. \quad (16)$$

Figure 4 Customer valuation function $v(x)$ in **example 1** and its upper concave envelope $\hat{v}(x)$.

Note. $\hat{v}(x) = v(x)$ for $x \in \{\underline{x}\} \cup [\hat{x}, \bar{x}]$, and $\hat{v}(x) > v(x)$ for $x \in (\underline{x}, \hat{x})$.

This strategy yields a strictly higher average perceived value than a single-type allocation:

$$\frac{\hat{x} - \ell}{\hat{x} - \underline{x}} v(\underline{x}) + \frac{\ell - \underline{x}}{\hat{x} - \underline{x}} v(\hat{x}) = \hat{v}(\ell) > v(\ell). \quad (17)$$

The gap $\hat{v}(\ell) - v(\ell)$ therefore measures the value created by bag differentiation.

At the boundary $\ell = \underline{x}$, the only feasible allocation is $\Delta_{\underline{x}}$ and $\hat{v}(\underline{x}) = v(\underline{x})$.

The benefit of a two-type design comes from diminishing sensitivity to losses, which makes valuation convex below the promised benchmark h . Loss aversion, captured by $\lambda > 1$, determines the relative weight of losses and gains. Equal splitting gives every customer food value ℓ and places all customers in the loss region when $\ell < h$. A two-type design instead combines the lower support point $\underline{x}$ with a higher point $\hat{x} \geq h$. With the mixing weights in (17), expected customer valuation $\hat{v}(\ell)$ exceeds that of equal splitting at food value ℓ .

COROLLARY 2. *If $v(\cdot)$ is concave on $\mathcal{X}$, then a degenerate distribution $\phi^*(\cdot | \ell) = \Delta_{\ell}$ is optimal.*

Corollary 2 identifies the boundary of the differentiation result. Differentiation creates value only where $v(\cdot)$ is nonconcave on the feasible support. If utility is concave on that support, equal splitting is optimal.

4.3. Reformulated Bellman Equation and Optimal Policy

The results of **Section 4.1** and **Section 4.2** together reduce the store's infinite-dimensional problem to two Stage-I scalar decisions, price and the reservation cap, followed by one Stage-II scalar decision, the target average bag value. Therefore, a policy π reduces to three scalar rules: a pricing rule $r \mapsto p^\pi(r)$, a reservation-cap rule $r \mapsto y^\pi(r)$, and a target average bag-value rule $(r, p, D, Q) \mapsto \ell^\pi(r, p, D, Q)$. The bag-value distribution rule $\phi^\pi(\cdot)$ is fully recovered from $\ell^\pi(\cdot)$: given the realized target $\ell = \ell^\pi(r, p, D, Q)$, if $v(\ell) = \hat{v}(\ell)$, all bags receive value ℓ ; if $v(\ell) < \hat{v}(\ell)$, the store uses the two contact points spanning the affine

segment of $\hat{v}$ that contains ℓ , with weights chosen to attain mean ℓ . Equation (16) gives the corresponding weights for the prospect-theory specification in Example 2.

For any admissible policy $\pi \in \Pi$, let $J^\pi(r)$ denote the expected discounted value-to-go achieved under policy π starting from reputation r . The optimal value function is $J^*(r) := \sup_{\pi \in \Pi} J^\pi(r)$. Substituting $V[\phi^*(\cdot | \ell)] = \hat{v}(\ell)$ from Proposition 2 into (13), we can write $J^*(r)$ as

$$J^*(r) = \max_{\substack{p \in [0, \bar{p}] \\ y \in [\underline{y}, \bar{y}]}} \mathbb{E}_\epsilon \left\{ p \cdot D(r, p, y, \epsilon) + \mathbb{E}_Q \left[\max_{\ell \in \mathcal{X}} \gamma J^*(\delta(\hat{v}(\ell) - p) + (1 - \delta)r) - c \cdot (\ell D(r, p, y, \epsilon) - Q)^+ \right] \right\}. \quad (18)$$

The optimal policy, denoted by $\pi^* = (p^*(\cdot), y^*(\cdot), \ell^*(\cdot))$, is the solution to (18). Under the baseline fixed-promise specification, the consumer utility function $v(\cdot)$ is determined by pre-listing information—the product category and the promised value (e.g., “baked goods worth \$30”)—so $v(\cdot)$ and its upper concave envelope $\hat{v}(\cdot)$ can be computed offline.

The Stage-II objective in (18) is generally *non-concave*, which makes global comparative statics for the joint Stage-I policy difficult to establish. The reservation cap also changes the role of price: with an endogenous cap, a high-reputation store can limit accepted demand through y while using price to manage margin and customer satisfaction. Thus the optimal price need not be maximal at high ratings. When the reservation cap is locally slack, the model reduces to a local price-adjustment problem.

Food allocation response to within-period uncertainty. The remaining Stage-II decision is how to set the average bag value ℓ^* in response to accepted reservations D and surplus Q .

PROPOSITION 3. *Let $\ell^*(r, p, D, Q)$ denote the maximal element of the Stage-II argmax set. Given r and p , $\ell^*(r, p, D, Q)$ is (i) weakly decreasing in D and (ii) weakly increasing in realized surplus Q .*

Furthermore, if the optimal value function $J^*(r)$ is concave, then $\ell^*(r, p, D, Q)$ is also (iii) weakly increasing in price p and (iv) weakly decreasing in rating r .

In Stage II, the store chooses ℓ by trading off reputational gains, through $\hat{v}(\ell)$, against supplementation cost when $\ell D > Q$. Results (i) and (ii) characterize how the store responds to within-period uncertainty. When accepted reservations D are unexpectedly high, the store must spread the surplus across more bags, making it costlier to maintain high bag values; the optimal response is to decrease the average bag value. Conversely, when realized surplus Q is unexpectedly high, the store can increase the average bag value without incurring additional supplementation cost. The reservation cap and the target average bag value therefore address the same mismatch at different times. Holding (r, p, ϵ, Q) fixed, tightening a binding cap lowers D and, by part (i), weakly raises the target average bag value; it can thereby reduce the need for two-type differentiation. Conditional on the realized number of accepted reservations, the bag-value response itself is unchanged.

Results (iii) and (iv) are conditional on concavity of $J^*(r)$; results (i) and (ii) do not require it. A higher price lowers net consumer satisfaction, so the store optimally compensates by raising the average bag value to protect its rating. A higher current rating provides more reputational slack, allowing the store to offer a lower average bag value.

5. Approximations

Although the action-space reduction in Section 4 collapses the second-stage decision to a scalar, the Bellman equation (18) remains a two-stage stochastic control problem with multiple continuous decisions and nested expectations. Solving it exactly requires substantial computation, which limits its usefulness as a scalable decision-support tool. Motivated by implementability and interpretability, we develop two approximation policies that preserve the key economics of reputation building while remaining computationally efficient. The first, a linear value function approximation (LVA), yields a closed-form bag-value rule and a low-dimensional joint reservation-control problem. The second, a deterministic fluid approximation (DFA), collapses the two-stage decision into a single stage where the target average bag value depends on the current rating only. Both approximations admit explicit performance guarantees relative to the optimal DP.

5.1. Linear Value Function Approximation (LVA)

Let $\mathcal{T}$ denote the Bellman optimality operator on the right-hand side of (18). The LVA replaces the optimal future value $J^*(\cdot)$ with an affine function $\tilde{J}(\cdot)$: $J^*(r) \approx \tilde{J}(r) := \tilde{J}_0 + kr$, where k is a parameter approximating the marginal value of reputation and $\tilde{J}_0$ is a constant. The characterization below assumes the nondegenerate case $k > 0$; when $k = 0$, the LVA policy is defined directly by the one-step and Stage-II argmaxes below. The key simplification from the affine approximation is that the discounted future value $\gamma\tilde{J}(r')$ becomes linear in the next period's rating r' . By the reputation update (4), $r' = \delta S + (1 - \delta)r$ where $S := \hat{v}(\ell) - p$ is the current consumer satisfaction. Substituting $\tilde{J}(r') = \tilde{J}_0 + kr'$ and collecting terms, the continuation value reduces to a linear function of S : a unit increase in satisfaction raises the continuation value by $\gamma\delta k$. This motivates defining

$$\omega := \gamma\delta k, \quad (19)$$

which is positive when $k > 0$. We call ω the *marginal value of satisfaction*: it converts a unit of customer utility into discounted future value. A larger ω means that current satisfaction has greater continuation value because reputation is more valuable, the store is more patient, or ratings respond more quickly.

Replacing the continuation value J^* in (18) with $\tilde{J}$ gives the following one-step look-ahead objective:

$$(\mathcal{T}\tilde{J})(r) = \gamma\tilde{J}_0 + \gamma k(1 - \delta)r + \max_{\substack{p \in [0, \bar{p}] \\ y \in [\underline{y}, \bar{y}]}} \mathbb{E}_{\epsilon, Q} \left[p \cdot D(r, p, y, \epsilon) - \omega p + \max_{\ell \in \mathcal{X}} \left\{ \omega \hat{v}(\ell) - c(\ell D(r, p, y, \epsilon) - Q)^+ \right\} \right]. \quad (20)$$

where $D(r, p, y, \epsilon) = \min\{d(r, p, \epsilon), y\}$. The LVA policy $\tilde{\pi} = \{\tilde{p}(\cdot), \tilde{y}(\cdot), \tilde{\ell}(\cdot)\}$ selects the maximizing actions on the right-hand side of (20), where $\tilde{p} : r \mapsto [0, \bar{p}]$, $\tilde{y} : r \mapsto [\underline{y}, \bar{y}]$, and $\tilde{\ell} : (r, p, D, Q) \mapsto \mathcal{X}$. This

one-step objective preserves a closed-form Stage-II rule and a bounded Stage-I search over price and the reservation cap.

LVA Policy Characterization. The store never deliberately leaves surplus unused while simultaneously supplementing from regular inventory. Thus, if supplementation occurs, the chosen average bag value must satisfy $\ell \geq Q/D$, and the relevant tradeoff is between the reputational value of a higher target, $\omega \hat{v}(\ell)$, and the supplementation cost $cD\ell$. Define the LVA target bag value

$$\ell^\circ(D) := \arg \max_{\ell \in \mathcal{X}} \{\omega \hat{v}(\ell) - cD\ell\}. \quad (21)$$

The largest optimizer is selected if the maximizer is not unique. When it is interior and $\hat{v}'$ is locally strictly decreasing, it satisfies $\hat{v}'(\ell^\circ(D)) = cD/\omega$, equivalently $\ell^\circ(D) = (\hat{v}')^{-1}(cD/\omega)$. Here, $\ell^\circ(D)$ is the reputation-adjusted service target: it is higher when satisfaction is more valuable and lower when more reservations must be served.

For the Stage-I decision, let $m(D, Q)$ denote the marginal scarcity cost created by one additional accepted reservation:

$$m(D, Q) = \begin{cases} c\ell^\circ(D), & Q \leq D\ell^\circ(D), \\ \omega \hat{v}'(\frac{Q}{D}) \frac{Q}{D^2}, & D\ell^\circ(D) < Q < D\bar{x}, \\ 0, & Q \geq D\bar{x}. \end{cases} \quad (22)$$

At a kink of $\hat{v}$, use the relevant one-sided slope. The term $m(D, Q)$ links the *ex-ante* reservation decision to *ex-post* surplus scarcity. In the first regime, one more accepted reservation requires supplementation, so the marginal cost is $c\ell^\circ(D)$. In the second regime, the store uses surplus only, and the marginal cost is reputational dilution from spreading a fixed surplus pool across more customers. In the third regime, surplus is abundant enough that an additional reservation creates no scarcity cost.

Write $[z]_{\mathcal{X}} := \min\{\bar{x}, \max\{\underline{x}, z\}\}$ for truncation onto $\mathcal{X}$. We formalize the optimal LVA policy below.

PROPOSITION 4. *Under the LVA policy, the store uses surplus before supplementing and chooses the Stage-II average bag value*

$$\tilde{\ell}(D, Q) = \left[\max \left\{ \frac{Q}{D}, \ell^\circ(D) \right\} \right]_{\mathcal{X}}. \quad (23)$$

At an interior Stage-I optimum, the reservation price equals the expected marginal scarcity cost:

$$\tilde{p}(r) = \mathbb{E}_Q[m(\tilde{y}(r), Q)]. \quad (24)$$

The Stage-II rule is directly implementable. The store first uses realized surplus. If surplus per accepted reservation is below the reputation-adjusted target $\ell^\circ(D)$, it supplements only up to that target. The final mean value is then implemented with the one- or two-type envelope rule in [Proposition 2](#).

The Stage-I rule shows how price and the reservation cap share the role of demand control. At an interior optimum, (24) says that the store opens additional slots until the revenue from one more accepted reservation

equals its expected scarcity cost. For comparison, when the cap is fixed at a nonbinding upper bound, the price condition becomes

$$\tilde{p} = \frac{\alpha r}{2\beta} - \frac{\omega}{2\beta} + \frac{1}{2} \mathbb{E}_{\epsilon, Q} [m(D(r, \tilde{p}, \tilde{y}, \epsilon), Q)]. \quad (25)$$

The three terms are the myopic monopoly price, a reputation discount, and a scarcity markup. Price manages margin and customer net utility, while the cap limits exposure to high-demand scarcity states.

Figure 5 Decision flow of the LVA heuristic. The joint reservation-control rule $(\tilde{p}(r), \tilde{y}(r))$ and the target average bag value $\ell^\circ(D)$ are precomputed offline and stored as lookup tables. The flow applies to fulfilled periods with $D > 0$; no Stage-II allocation is made in a no-sale period.

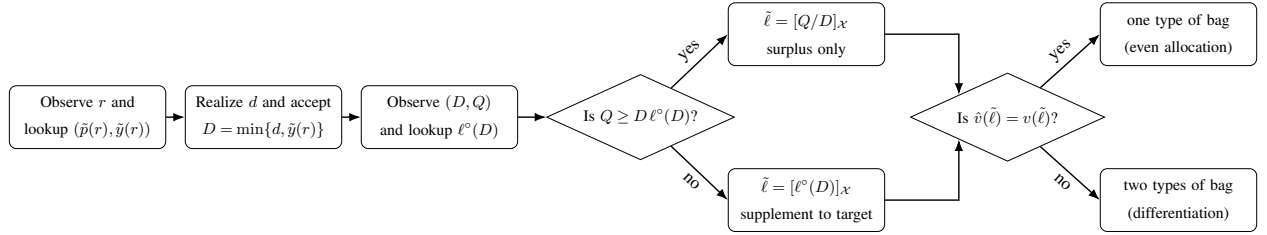


Figure 5 summarizes the decision flow of the LVA policy. The joint Stage-I rule $(\tilde{p}(r), \tilde{y}(r))$ and the target $\ell^\circ(D)$ can be precomputed offline and stored as lookup tables; for discrete bag counts, the platform can enumerate feasible caps and solve a scalar price problem for each cap. At runtime, the manager looks up the price and cap, compares Q/D with $\ell^\circ(D)$ after (D, Q) is observed, and then recovers the final bag-value distribution from $\tilde{\ell}$ using the procedure in Section 4.3.

Appendix A presents a square-root example, $v(x) = \theta\sqrt{x}$, that yields explicit closed-form expressions for the Stage-II target and marginal scarcity cost. The joint Stage-I pair $(\tilde{p}(r), \tilde{y}(r))$ is then obtained through the bounded search in (20).

Performance Guarantee. The following proposition quantifies the worst-case optimality loss of the LVA policy.

PROPOSITION 5. *There exist LVA parameters $\tilde{J}_0$ and $k \geq 0$ such that, for all $r \in [\underline{r}, \bar{r}]$,*

$$J^*(r) - J^{\tilde{\pi}}(r) \leq \frac{2\gamma\alpha(\bar{p} + c\bar{x})(\bar{r} - \underline{r})}{3(1 - \gamma)[1 - \gamma(1 - \delta)]}.$$

Proposition 5 bounds the LVA loss by a term proportional to the feasible rating-state range, the sensitivity of latent requests to reputation, and the one-period payoff scale. The bound is smaller when the store places less weight on future profits, when demand is less sensitive to reputation, and when ratings update faster, which reduces the persistence of approximation errors in the reputation state.

REMARK 5 (ESTIMATING LVA PARAMETER k). The performance guarantee in Proposition 5 is an *existence* result: it asserts that there exist coefficients $(\tilde{J}_0, k)$ for which the LVA policy attains a controlled worst-case optimality gap. The LVA decisions depend only on the *slope* k (the intercept $\tilde{J}_0$ cancels from the

argmax). For implementation, LVA can therefore be viewed as a one-parameter family of policies indexed by k , which summarizes how aggressively the store trades current profit for future reputation.

In a TGTG setting, a retailer can tune k offline using lightweight simulation or historical bootstrapping. For a grid of candidate k values, one can (i) compute the induced policy via the closed-form Stage-II rule (23) and a bounded search for $(\tilde{p}(r), \tilde{y}(r))$, and (ii) evaluate long-run profit under empirical samples of (ϵ, Q) . This procedure avoids solving the full dynamic program and yields an interpretable parameter k (the marginal value of reputation) that can be adapted to store- and category-specific environments.

5.2. Deterministic Fluid Approximation

The LVA approach retains the two-stage structure, allowing a store to adjust its bag design in response to accepted reservations and end-of-day surplus. In practice, however, some stores may find it operationally simpler to commit to the bag design at the start of each day, before demand and surplus are observed. This motivates the *deterministic fluid approximation* (DFA), a standard technique in stochastic control (see, e.g., Najafi et al. 2024, Chan et al. 2024), which replaces stochastic processes with their mean behavior. In our setting, DFA yields a policy that depends only on the current rating r , which is easy to compute and implement while preserving the key tradeoff between current revenue and reputation building.

The DFA model is a deterministic system where latent requests and surplus are equal to their mean values

$$\bar{d}(r, p) := \mathbb{E}[d_t] = \alpha r - \beta p, \quad \bar{D}(r, p, y) := \min\{\bar{d}(r, p), y\}, \quad \bar{Q} := \mathbb{E}_F[Q].$$

Let $\hat{J}(r)$ denote the optimal value function under the deterministic system. It satisfies the Bellman equation:

$$\hat{J}(r) = \max_{\substack{p \in [0, \bar{p}], y \in [\underline{y}, \bar{y}] \\ \ell \in \mathcal{X}}} \left\{ p \bar{D}(r, p, y) - c[\ell \bar{D}(r, p, y) - \bar{Q}]^+ + \gamma \hat{J}((1 - \delta)r + \delta(\hat{v}(\ell) - p)) \right\}. \quad (26)$$

The DFA policy $\hat{\pi} = \{\hat{p}(\cdot), \hat{y}(\cdot), \hat{\ell}(\cdot)\}$ is the optimal solution to (26), where $\hat{p} : r \mapsto [0, \bar{p}]$, $\hat{y} : r \mapsto [\underline{y}, \bar{y}]$, and $\hat{\ell} : r \mapsto \mathcal{X}$. Since all three rules are functions of the current rating r , they can be precomputed over a grid of rating values and stored as a lookup table, making real-time implementation straightforward.

Performance Guarantee. The following proposition bounds the optimality loss of the DFA policy.

PROPOSITION 6. *If ϵ and Q have finite second moments, then for every admissible initial rating r ,*

$$J^*(r) - J^{\hat{\pi}}(r) \leq \frac{2}{1 - \gamma} \left[(\bar{p} + c\bar{x})\sqrt{\text{Var}(\epsilon)} + c\sqrt{\text{Var}(Q)} \right].$$

Proposition 6 is tightest when future losses are discounted heavily, demand and surplus shocks are small, and supplementation is inexpensive. Because the cap truncates demand, the bound contains the revenue term $\sqrt{\text{Var}(\epsilon)}: \mathbb{E}[\min\{\alpha r - \beta p + \epsilon, y\}]$ need not equal $\min\{\alpha r - \beta p, y\}$. The term involving c captures the cost of ignoring realized surplus fluctuations.

5.3. Steady State under the Deterministic System

DFA enables a steady-state analysis of how the store's long-run pricing, reservation-cap, and bag-design decisions depend on the operational environment and platform design. Specifically, we show that, in a deterministic system, the store's rating and decisions may converge to a *steady state* $(\hat{r}, \hat{p}, \hat{y}, \hat{\ell})$. We characterize an *interior active-supplementation* steady state in which accepted demand is non-zero, $\bar{D}(\hat{r}, \hat{p}, \hat{y}) > 0$, and supplementation is active, $\hat{\ell}\bar{D}(\hat{r}, \hat{p}, \hat{y}) > \bar{Q}$. Appendix C.4 gives conditions under which the steady-state necessary conditions admit a unique interior solution and discusses the associated boundary candidates.

Responsiveness Index. A key quantity governing the steady state is the *responsiveness index* κ :

$$\kappa := \sum_{t=1}^{\infty} \gamma^t \delta (1 - \delta)^{t-1} = \frac{\gamma \delta}{1 - \gamma(1 - \delta)}. \quad (27)$$

The responsiveness index is the discounted cumulative effect of one unit of current satisfaction on future ratings. It increases when the store is more forward-looking or when ratings place more weight on recent outcomes.

In the interior deterministic regime considered below, a strictly demand-binding reservation cap is dominated by a small price increase; hence we can select a nonbinding cap and characterize the steady state through $(\hat{r}, \hat{p}, \hat{\ell})$.

PROPOSITION 7. *Under the deterministic system, an optimal interior active-supplementation steady state $(\hat{r}, \hat{p}, \hat{y}, \hat{\ell})$ at which $\hat{v}$ is differentiable and $\bar{d}(\hat{r}, \hat{p}) < \bar{y}$ satisfies*

$$\hat{r} = \hat{v}(\hat{\ell}) - \hat{p}, \quad \hat{v}'(\hat{\ell}) = c \left(1 + \frac{\beta}{\alpha \kappa} \right) \quad \text{and} \quad \frac{\bar{d}(\hat{r}, \hat{p})}{\hat{p} - c\hat{\ell}} = \beta + \alpha \kappa, \quad (28)$$

where the steady-state price satisfies $\hat{p} > c\hat{\ell}$ and can be written as

$$\hat{p} = \frac{\alpha \hat{v}(\hat{\ell}) + c\hat{\ell}(\beta + \alpha \kappa)}{\alpha(1 + \kappa) + 2\beta}. \quad (29)$$

Moreover, any feasible nonbinding cap $\hat{y} \geq \bar{d}(\hat{r}, \hat{p})$ is payoff-equivalent, and when $\bar{d}(\hat{r}, \hat{p}) \geq \underline{y}$ we select $\hat{y} = \bar{d}(\hat{r}, \hat{p})$ as the minimal-cap normalization.

With deterministic demand and an interior price, price can rise without changing accepted volume, so a binding cap is dominated. Under stochastic demand, the same increase suppresses ordinary-state reservations, whereas the cap clips only the high-demand tail. The reservation cap is thus primarily an uncertainty-control instrument, a role captured by LVA and DP but absent from the deterministic steady state.

The first condition in (28) is a *fixed-point condition*: in steady state, the store rating equals the net utility delivered to consumers.

The second is a marginal optimality condition characterizing $\hat{\ell}$: at the optimal steady-state bag value, the marginal reputational benefit $\kappa \hat{v}'(\hat{\ell})$ equals the effective marginal cost of supplementation $c(\kappa + \beta/\alpha)$. The

factor $\kappa + \beta/\alpha$ reflects that higher bag value not only increases supplementation cost directly but also raises future demand through reputation, which in equilibrium puts upward pressure on price and further increases the cost of serving each demand.

The third is a marginal optimality condition characterizing $\hat{p}$ under the normalized nonbinding cap. A unit price increase generates an immediate marginal profit of $\bar{d}(\hat{r}, \hat{p}) - \beta(\hat{p} - c\hat{\ell})$. At the same time, a higher price reduces net consumer satisfaction, which depresses the store's rating and reduces future demand. The total discounted future demand loss from a unit price increase is $\alpha\kappa(\hat{p} - c\hat{\ell})$. At the optimal steady-state price $\hat{p}$, these two marginal effects exactly offset: the immediate marginal profit gain equals the long-run reputational cost.

If $\hat{p} = \bar{p}$, a strictly demand-binding cap may remain because price cannot be raised further. The listing bounds may also determine which nonbinding cap is selected. These are boundary regimes rather than the interior steady state characterized in [Proposition 7](#).

COROLLARY 3. *Under the conditions of Proposition 7, suppose additionally that the bag-value equation in (28) has a unique interior solution and that $\hat{v}$ is twice differentiable there with $\hat{v}''(\hat{\ell}) < 0$. Along parameter changes that preserve the interior active-supplementation, nonbinding-cap regime:*

- (i) *The steady-state bag value $\hat{\ell}$ is unique and strictly increasing in the responsiveness index κ and in α , and strictly decreasing in β and in c .*
- (ii) *The steady-state reputation $\hat{r}$ is strictly increasing in κ and therefore increasing in γ and δ .*

[Corollary 3](#) clarifies how the store's long-run strategy is shaped by the operating environment and platform design. A larger responsiveness index κ , driven by either a higher weight on future profits or a faster-updating rating system, raises both the steady-state bag value $\hat{\ell}$ and the steady-state rating $\hat{r}$. A higher rating elasticity α has the same qualitative effect because quality investment converts more strongly into future demand, whereas a higher price elasticity β or supplementation cost c lowers the target bag value because preserving quality is less profitable or more expensive. From a platform-design perspective, shortening the rating window raises κ and strengthens incentives to maintain bag quality.

In the numerical experiments, we examine how these comparative statics from the deterministic system carry over to the stochastic baseline model.

6. Numerical Study

This section compares Baseline, Envelope, LVA, and DP and then investigates whether the analytical insights from the model survive changes in utility curvature, selective ratings, and store control of the promised value h . [Appendix D](#) provides the implementation details.

Numerical design. We use the loss-averse valuation (Tversky and Kahneman 1992)

$$v(\ell) = \begin{cases} v_0 + \lambda_+(\ell - \ell_{\text{ref}})^\eta & \text{if } \ell \geq \ell_{\text{ref}}, \\ v_0 - \lambda_-(\ell_{\text{ref}} - \ell)^\eta & \text{if } \ell < \ell_{\text{ref}}, \end{cases} \quad (30)$$

with $\ell_{\text{ref}} = h_0 = 3.5$, $\eta = 0.6$, $(\lambda_+, \lambda_-) = (1.5, 3.0)$, and $v_0 = 3.0$. Its nonconcavity makes two-type allocation potentially valuable.

We enumerate parameter combinations over a grid that combines $\alpha, \beta \in \{1.8, 1.9, 2.0, 2.1, 2.2\}$ and $c \in \{0.75, 0.80, 0.85, 0.90, 0.95\}$ with five truncated-normal demand-shock and scaled-beta surplus environments, yielding 625 operating environments, each of which defines an independent simulation experiment. Each policy uses $T = 2,000$, ten initial ratings on $[3, 5]$, and eight common-random-number replications; long-run operating measures discard the first 200 periods. The DP solver, numerical LVA search, and simulator all use $D_t = \min\{(\alpha r_t - \beta p_t + \epsilon_t)^+, y_t\}$; when $D_t = 0$, Stage II is skipped and $r_{t+1} = r_t$. The DP solver and simulator use the same projection of candidate ratings onto $[0, 5]$. The numerical positive part therefore relaxes the analytical positive-demand support condition: the dynamic analytical results are proved for strictly positive latent demand, whereas numerical DP additionally accommodates no-sale states. Common random numbers support paired policy comparisons, the ten initial ratings cover different starting states, and bootstrapping environment identifiers summarizes variation across the 625 operating environments rather than a population of stores.

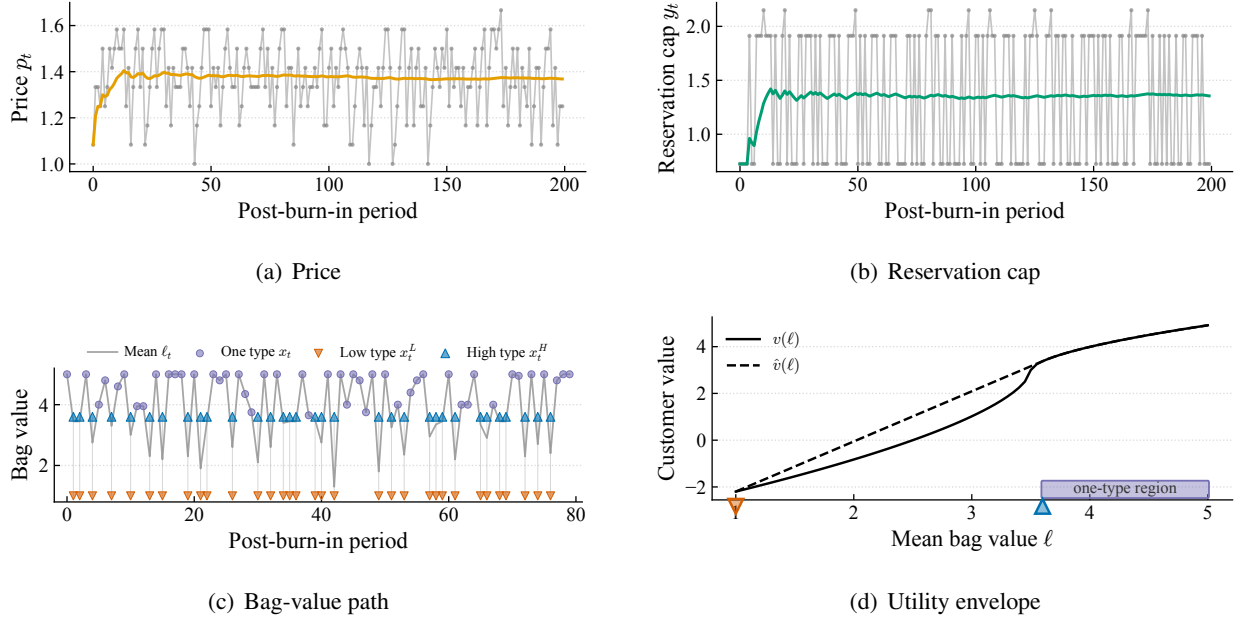
Four Policies for Comparison. When $D_t > 0$, Baseline fixes $p_t = h_0/3$, following the platform’s approximate one-third pricing convention (Too Good To Go 2026), and $y_t = 5 = \bar{y}$, the least restrictive feasible cap. Baseline assigns uniform bags at mean $[Q_t/D_t]_{\mathcal{X}}$. Envelope changes only the bag allocation to the optimal one- or two-type design. LVA adapts price, cap, and the Stage-II target using an affine continuation value; DP solves the full discretized program. Because the LVA slope is fitted to the DP value function, this oracle-calibrated LVA isolates affine-approximation loss rather than deployment-calibration error.

6.1. Trajectories of Price, Cap, and Bag Design under Optimal DP

We illustrate DP at the central case $\alpha = \beta = 2$ and $c = 0.95$, with ϵ a truncated normal shock of scale 0.5 and Q a scaled Beta(2, 2) draw on $[2, 5]$. Figure 6(a)–(c) displays post-burn-in sample paths of price, the reservation cap, and bag value, together with the utility envelope in panel (d).

Figure 6 shows that, after the simulation’s initial burn-in phase, price is relatively stable, the pre-set cap clips high-request realizations, and Stage-II bag design adapts to accepted reservations and realized surplus. The cap binds in 40.6% of DP periods and 80.6% of LVA periods across the grid, showing that upper-tail reservation control is active in the stochastic policy paths. By contrast, the fixed benchmark cap $y = 5$ does not bind under Baseline or Envelope in the same post-burn-in simulations. Two types appear only on a strict envelope segment. The plotted x is quality-adjusted food value.

Figure 6 Optimal price p_t , reservation cap y_t , bag-value path, and utility envelope under stochastic Q and ϵ . Panels (a)–(b) show 200 post-burn-in periods and panel (c) shows 80. Gray lines in (a)–(b) are sample-path decisions, and colored lines are running averages; in (c), gray is the target mean and orange/blue mark the low/high supports.



6.2. Profit and Waste Comparison Across Policies

We compare normalized discounted average profit, based on $\Pi_t = p_t D_t - c(D_t \ell_t - Q_t)^+$, and a modeled total-waste proxy. The latter combines surplus left in store with at-risk excess delivered above household need:

$$W_t^{\text{total}} = (Q_t - D_t \ell_t)^+ + \xi D_t \int_{\bar{\ell}_{\text{need}}}^{\infty} (x - \bar{\ell}_{\text{need}}) \phi_t(x) dx.$$

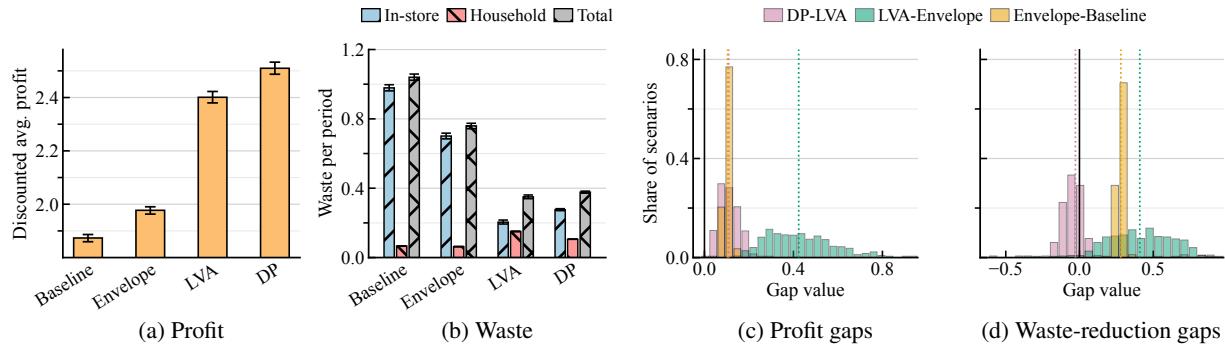
We set the at-risk weight $\xi = 0.3$ and household-need benchmark $\bar{\ell}_{\text{need}} = 1.1h_0$. Figure 7 reports policy levels and matched policy gaps across the 625 environments.

Conditional differentiation has meaningful value, but joint dynamic control matters more. LVA captures 82.9% of DP's profit improvement over Baseline; DP's remaining mean advantage is 0.109, or 4.3% of its profit. Because the dynamic policies also change the fulfillment target, this increment is not attributable to price and the cap alone. DP maximizes profit, whereas LVA produces slightly less modeled waste and shifts its location between store and households. Profit optimization therefore need not minimize waste.

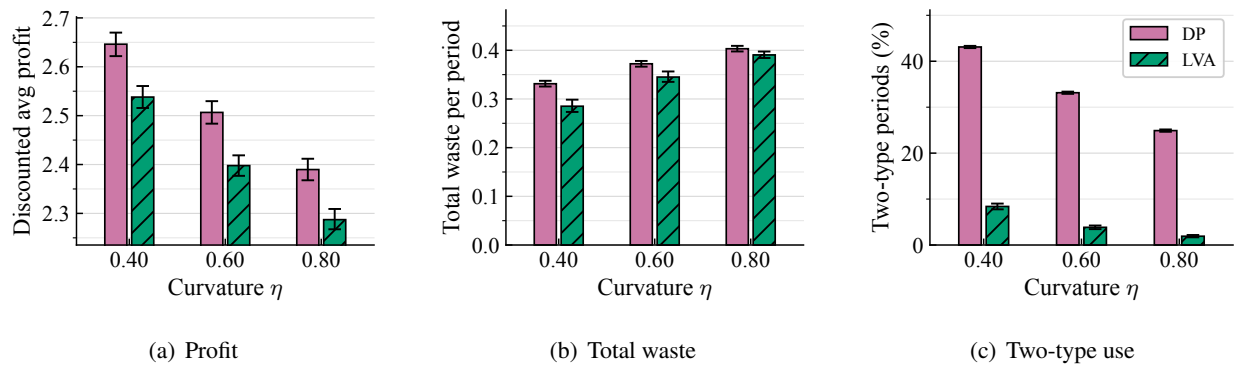
6.3. Extensions: Utility Curvature, Selective Ratings, and Endogenous Promised Value

We next examine three extensions: alternative utility curvature, selective ratings, and store control of promised value. Each extension retains the policy comparison relevant to the change.

Utility curvature. As shown in (30), η captures diminishing sensitivity to gains and losses relative to the promised value, while loss aversion remains fixed. A lower η means that a given change in bag value

Figure 7 Policy levels and paired scenario-level gaps.

Notes: Panels (a)–(b) report means and 95% scenario-bootstrap intervals. Panels (c)–(d) report common-random-number paired-gap shares; positive favors the first-listed policy and, in (d), greater waste reduction. Colored dotted lines mark the corresponding mean gaps; the black vertical line marks zero.

Figure 8 Utility curvature, two-type use, and policy outcomes on the main 625-environment grid.

Notes: Bars report DP and LVA means with 95% scenario-level bootstrap confidence intervals over 625 operating environments. Panels (a)–(c) show discounted average profit, total waste, and two-type-use% among fulfilled post-burn-in periods, respectively.

has a larger effect on customer valuation when the bag value is close to the promised value than when it is farther away. A two-type allocation can therefore generate a larger increase in average customer valuation by moving some bags across the promised value while holding the average bag value fixed. As η rises toward one, customer valuation becomes closer to linear, so this gain over equal splitting becomes smaller. Consistent with this mechanism, Figure 8 shows that, as η increases from 0.40 to 0.80, DP's two-type use falls from 43.1% to 24.9%, and LVA's from 8.4% to 1.9%.

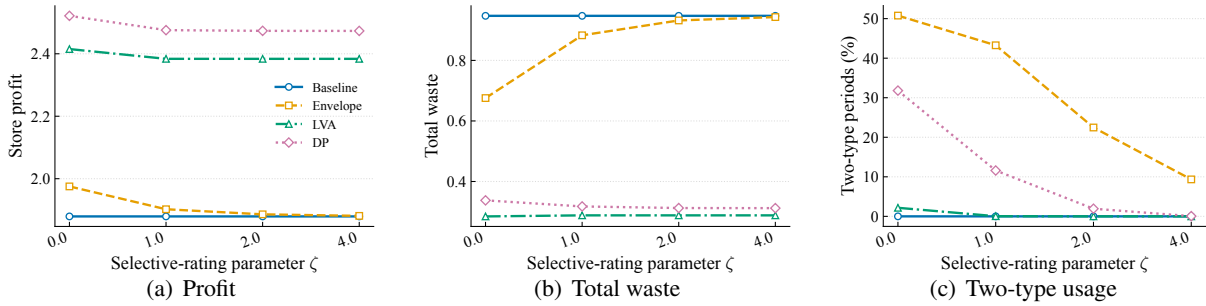
DP's paired profit advantage over LVA remains 0.102–0.109 (0.109 at the baseline $\eta = 0.60$), with positive bootstrap intervals throughout. Thus, as diminishing sensitivity weakens, both policies use equal splitting more often, although they continue to use two bag types in selected periods. Total waste nevertheless rises and LVA's waste advantage shrinks in the tested grid. Hence, more frequent use of equal splitting does not imply lower total waste.

Selective ratings. A low-value bag may make a consumer more likely to leave a rating. We model the relative rating propensity as $\psi_\zeta(x) = \exp\left(-\zeta \frac{x-\underline{x}}{\bar{x}-\underline{x}}\right)$, where $\zeta \in \{0, 1, 2, 4\}$ and $\zeta = 0$ corresponds to the baseline, and a larger ζ places greater weight on ratings from lower-value bags. The corresponding rating-weighted distribution is

$$\phi_\zeta^\psi(x) = \frac{\phi(x)\psi_\zeta(x)}{\int_{\mathcal{X}} \phi(z)\psi_\zeta(z)dz}, \quad (31)$$

and satisfaction is evaluated under this rating-weighted distribution. For each target mean, an optimal allocation can still be chosen with at most two bag types. We use this property to recompute all four policies;

Figure 9 Selective-rating extension as rating propensity shifts toward low-value bags. Lines report means over five matched stochastic environments at $\alpha = \beta = 2$ and $c = 0.95$



As shown in Figure 9, placing greater rating weight on lower-value bags sharply reduces the share of periods using two bag types. Envelope and DP use two bag types much less often as ζ increases, while LVA's already infrequent use disappears; Baseline always uses equal splitting. Selective ratings give the lower-value bags in a two-type allocation a disproportionate effect on future reputation, thereby reducing the benefit of differentiation. Profit falls under Envelope, LVA, and DP, most visibly under Envelope because differentiation is its only departure from Baseline; the declines under LVA and DP are modest because these policies can also adjust price, the reservation cap, and the target average bag value. Total waste does not move uniformly: it rises under Envelope toward the Baseline level, changes little under LVA, and falls modestly under DP. Thus selective ratings consistently discourage two-type allocation, but their waste effect depends on how the policy adjusts the store's other decisions.

Endogenous promised value h . The baseline model fixes the promised value at $h_t = h_0 = 3.5$. Motivated by Too Good To Go's approximate convention that the reservation price is one third of the bag's original retail value (Too Good To Go 2026), we allow the store to choose h_t jointly with price and the reservation cap in Stage I from $\mathcal{H} = \{3.0, 3.5, 4.0, 4.5, 5.0\}$, subject to $2p_t \leq h_t \leq 4p_t$. A chosen promise becomes the valuation reference and also shifts latent requests:

$$d(r, p, h, \epsilon) = \alpha r - \beta p + \rho(h - h_0) + \epsilon, \quad D(r, p, y, h, \epsilon) = \min\{d(r, p, h, \epsilon)^+, y\}.$$

We hold $\mathcal{X}$ fixed across h and recompute the valuation function and envelope for each promise. Here ρ measures the increase in latent requests per unit of promise; we use $\rho \in \{0.50, 0.55, 0.60\}$, below the baseline rating and price coefficients $\alpha = \beta = 2$. **Table 1** focuses on LVA and DP; Baseline and Envelope retain fixed h_0 to preserve their benchmark definitions. The comparison uses 100 independently generated stochastic environments at $c = 0.95$ and matched random-number streams, averaging the three ρ values within each environment before inference. All 300 endogenous-promise DP value functions are nondecreasing on the rating grid, validating the envelope reduction; Appendix D reports the diagnostic.

Table 1 Operating behavior of LVA and DP under fixed and endogenous promised values.

<i>Panel A: Promise and price</i>					
Policy	Promise rule	Mean promise	Within-path SD(h_t)	Mean price	Within-path SD(p_t)
LVA	Fixed h_0	3.500	0.000	1.742	0.026
		[3.500, 3.500]	[0.000, 0.000]	[1.741, 1.744]	[0.022, 0.030]
LVA	Endogenous	3.105	0.337	1.476	0.110
		[3.093, 3.118]	[0.310, 0.363]	[1.444, 1.503]	[0.090, 0.131]
DP	Fixed h_0	3.500	0.000	1.391	0.145
		[3.500, 3.500]	[0.000, 0.000]	[1.376, 1.406]	[0.142, 0.149]
DP	Endogenous	3.172	0.458	1.338	0.161
		[3.149, 3.198]	[0.416, 0.502]	[1.327, 1.348]	[0.155, 0.167]
<i>Panel B: Cap and fulfillment</i>					
Policy	Promise rule	Mean cap	Within-path SD(y_t)	Mean bag value	Two-type share (%)
LVA	Fixed h_0	0.783	0.378	4.561	2.4
		[0.772, 0.793]	[0.345, 0.412]	[4.548, 4.573]	[2.1, 2.7]
LVA	Endogenous	0.994	0.923	4.336	5.4
		[0.979, 1.008]	[0.901, 0.946]	[4.322, 4.351]	[5.2, 5.6]
DP	Fixed h_0	1.356	0.614	3.904	34.8
		[1.332, 1.381]	[0.594, 0.635]	[3.892, 3.915]	[34.3, 35.4]
DP	Endogenous	1.483	0.781	3.715	27.4
		[1.439, 1.526]	[0.752, 0.812]	[3.697, 3.734]	[26.0, 28.8]

Notes: Cells report means with 95% scenario-bootstrap confidence intervals in brackets; the three ρ values are averaged within each matched environment.

As shown in **Table 1**, both policies choose a lower average promised value and price, a higher reservation cap, and a lower average bag value. Relative to fixed h_0 , they therefore make a larger number of lower-value bags available. The paired profit gains are 0.256 for DP and 0.148 for LVA. DP's total-waste change is -0.041 (95% interval $[-0.051, -0.032]$), whereas LVA's is 0.012 ($[-0.005, 0.031]$). Endogenizing h thus changes how the store sets the number of bags available together with the promised and delivered value per bag; the different waste effects show that this change can alter the profit-waste ranking. Appendix D shows that Corollary 3's comparative statics in κ also survive stochastic shocks.

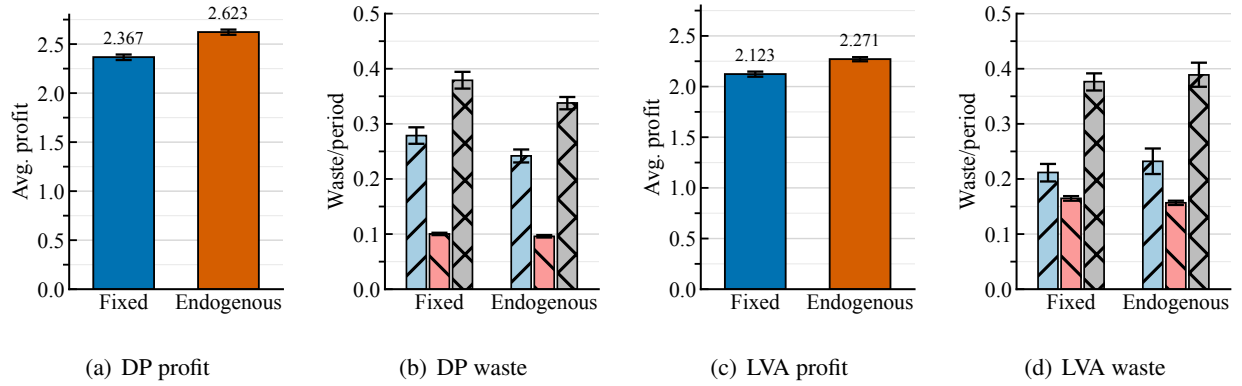


Figure 10 Outcome effects of endogenous promised value under DP and LVA.

Notes: Bars report means and 95% scenario-bootstrap confidence intervals after averaging the three ρ values within each matched environment. Waste bars report in-store, household, and total waste. LVA and DP use their own endogenous-promise rules.

7. Conclusion

Surplus-food retailers must sell opaque, perishable inventory under demand and surplus uncertainty while managing a visible platform reputation. We model their joint *ex-ante* price and reservation-cap decisions and *ex-post* bag design in an infinite-horizon dynamic program.

A key result is that an optimal *ex-post* bag design can always be implemented with at most two bag types. We show that the infinite-dimensional bag-design problem reduces to choosing the scalar target ℓ . When the upper concave envelope coincides with the utility function at ℓ , a single bag type (equal splitting) is optimal; when the envelope lies strictly above the utility function, an optimal implementation uses two bag types and yields a higher average customer valuation than equal splitting at the same mean.

Embedding the envelope reduction in the reputation problem yields the implementable LVA policy and the steady-state DFA approximation. In the main four-policy grid, DP remains on the profit frontier and LVA on the lower-waste frontier. These rankings persist across the tested utility curvatures, although two-type use declines as customer response becomes more linear, making differentiation most relevant on selected fulfillment days rather than as a routine rule. Selective ratings further reduce the value of differentiation by making low-value bags more salient in the rating signal. When promised value is also endogenous, both policies make more lower-value bags available, but only DP yields a statistically distinguishable reduction in total waste, so the fixed-promise waste ranking need not persist.

Operationally, price manages margin, latent requests, and customer net utility; the reservation cap limits high-demand commitments before surplus is known; and Stage-II bag design absorbs the remaining reservation-surplus mismatch. When promised value is flexible, these decisions determine whether the store makes fewer higher-value bags or more lower-value bags available. A binding cap is unnecessary for predictable interior demand but has a distinct role under demand uncertainty or when price is constrained. Because the retailer-optimal cap need not maximize food recovery, platform tools should report both the

profit and recovery consequences of recommended reservation caps. Rating-system design also matters because the representativeness of the rating signal determines how costly differentiated low-value bags become.

Limitations. First, the reservation cap is a single *ex-ante* listing decision. Stores may instead revise listings as sales and expected surplus evolve, so displayed bags remaining can reflect both reservations and later quantity changes. Our results apply when listings are set before demand and surplus are realized.

Second, the rating update gives constant weight to each period's average satisfaction. A volume-weighted historical average would require rating volume as a second state. Exponential smoothing is most appropriate when ratings per period are stable or recent ratings receive greater weight; otherwise decisions may depend on both the rating and rating volume.

Third, the monotonicity argument in **Proposition 1** uses both conditions in **Assumption 1**: if $(1 - \delta)\beta < \delta\alpha$, global monotonicity can fail; if $\bar{p} < c\bar{x}$, additional accepted volume need not be profitable. If either condition fails, **Proposition 2**'s fixed-mean envelope remains a customer-satisfaction benchmark, but **Corollary 1** need not identify a dynamically optimal allocation.

Fourth, equation (1) assumes constant marginal supplementation cost. Convex cost from congestion, overtime, or disruption would reduce supplementation in large-shortfall states and favor tighter *ex-ante* reservation control. The fixed-mean envelope and its at-most-two-bag-type implementation remain valid because cost still depends on the bag-value distribution only through its mean. The monotonicity result extends if its profitability bound uses the largest feasible marginal cost; otherwise monotonicity and the LVA and steady-state rules can change.

References

- Adebola O, Arora P, Zhang C (2022) Farm equipment sharing in emerging economies. Available at SSRN 4190725 .
- Agrawal V, Zhang C (2026) Sustainable sourcing of agricultural products: Fixed vs. flexible premiums. *Management Science* URL <http://dx.doi.org/10.1287/mnsc.2025.01093>, articles in Advance.
- Ahire SL, Pekgün P (2018) Harvest hope food bank optimizes its promotional strategy to raise donations using integer programming. *Interfaces* 48(4):291–306.
- Akkaş A, Gaur V (2022) OM forumreducing food waste: An operations management research agenda. *Manufacturing & Service Operations Management* 24(3):1261–1275.
- Akkaş A, Honhon D (2022) Shipment policies for products with fixed shelf lives: Impact on profits and waste. *Manufacturing & Service Operations Management* 24(3):1611–1629.
- Akkaş A, Honhon D (2023) Determining maximum shipping age requirements for shelf life and food waste management. *Production and Operations Management* 32(7):2173–2188, URL <http://dx.doi.org/10.1111/poms.13963>.
- Alptekinoglu A, Benade G (2024) Achieving Rawlsian justice in food rescue. Available at SSRN 4992842 .
- Ata B, Lee D, Sönmez E (2019) Dynamic volunteer staffing in multicrop gleaning operations. *Operations Research* 67(2):295–314.
- Atan Z, Honhon D, Pan XA (2025) Displaying and discounting perishables: Impact on retail profits and waste. *Management Science* 72(5):4106–4123, URL <http://dx.doi.org/10.1287/mnsc.2023.00316>.
- Bar-Isaac H, Tadelis S (2008) Seller reputation. *Foundations and Trends® in Microeconomics* 4(4):273–351.
- Belavina E, Girotra K, Kabra A (2017) Online grocery retail: Revenue models and environmental impact. *Management Science* 63(6):1781–1799.
- Boyabatli O, Shao L, Zhou YH (2021) Integrated optimization of farmland cultivation and fertilizer application: Implications for farm management and food security. Working paper.
- Cabral L, Hortacsu A (2010) The dynamics of seller reputation: Evidence from eBay. *The Journal of industrial economics* 58(1):54–78.
- Carnehl C, Stenzel A, Schmidt P (2024) Pricing for the stars: Dynamic pricing in the presence of rating systems. *Management Science* 70(3):1755–1772.
- Chan TC, Huang SY, Sarhangian V (2024) Dynamic control of service systems with returns: Application to design of postdischarge hospital readmission prevention programs. *Operations Research* 73(4):2242–2263, URL <http://dx.doi.org/10.1287/opre.2022.0066>.
- Chen N, Elmachtoub AN, Hamilton ML, Lei X (2021) Loot box pricing and design. *Management Science* 67(8):4809–4825.
- Cho S, Pekgün P, Janakiraman R, Wang J (2024) The competitive effects of online reviews on hotel demand. *Journal of Marketing* 88(2):40–60.

- Crama P, Zhou YH, Wang M (2023) Impact of cosmetic standards on food loss. Available at SSRN 4452800 .
- Dong AY, Siddiq A, Zhang J (2025) Empirical market design in surplus food marketplaces. Available at SSRN 5648812 .
- Dong L, Rashkova I, Shi D (2022) Food safety audits in developing economies: Decentralization vs. centralization. *Manufacturing & Service Operations Management* 24(6):2863–2881.
- Elmachtoub AN, Hamilton ML (2021) The power of opaque products in pricing. *Management Science* 67(8):4686–4702.
- Fay S, Gheibi S (2024) The effect of probabilistic selling on channel dynamics in supply chains. *Manufacturing & Service Operations Management* 26(2):632–645.
- Fay S, Xie J (2008) Probabilistic goods: A creative way of selling products and services. *Marketing Science* 27(4):674–690.
- Gaur V, Park YH (2007) Asymmetric consumer learning and inventory competition. *Management Science* 53(2):227–240.
- Han Z, Hu B, Dawande M (2023) Seeing beauty in ugly produce: A food waste perspective. Available at SSRN 4349504 .
- Huang T, Yin Z (2021) Dynamic probabilistic selling when customers have boundedly rational expectations. *Manufacturing & Service Operations Management* 23(6):1597–1615.
- Huang T, Yu Y (2014) Sell probabilistic goods? A behavioral explanation for opaque selling. *Marketing Science* 33(5):743–759.
- Jain N, Kabra A, Karamshetty V (2023) Until later is preferred over sooner: Multiplicity in product expiration dates and food waste in retail stores. Available at SSRN 4317868 .
- Jerath K, Netessine S, Veeraraghavan SK (2010) Revenue management with strategic customers: Last-minute selling and opaque selling. *Management Science* 56(3):430–448.
- Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.
- Kamenica E, Gentzkow M (2011) Bayesian persuasion. *American Economic Review* 101(6):2590–2615.
- Kazaz B, Xu F, Yu H (2024) Retailing strategies of imperfect produce and the battle against food waste. Available at SSRN 4951154 .
- Keskin NB, Li C, Song JSJ (2024) The blockchain newsvendor: Value of freshness transparency and smart contracts. *Management Science* 71(8):6666–6682, URL <http://dx.doi.org/10.1287/mnsc.2021.02949>.
- Lowrey JD, Richards TJ, Hamilton SF (2023) Food donations, retail operations, and retail pricing. *Manufacturing & Service Operations Management* 25(2):792–810.
- Luca M, Reshef O (2021) The effect of price on firm reputation. *Management Science* 67(7):4408–4419.
- Murali K, Lim MK, Petrucci NC (2019) The effects of ecolabels and environmental regulation on green product development. *Manufacturing & Service Operations Management* 21(3):519–535.
- Najafi S, Duenyas I, Jasin S, Uichanco J (2024) Multiproduct dynamic pricing with limited inventories under a cascade click model. *Manufacturing & Service Operations Management* 26(2):554–572.
- Nosko C, Tadelis S (2015) The limits of reputation in platform markets: An empirical analysis and field experiment. Working Paper 20830, National Bureau of Economic Research.
- Nu Y, Belavina E, Girotra K (2024) Using artificial intelligence to reduce food waste. Available at SSRN 4826777 .
- ReFED (2025) Five years since COVID-19 upended our food system. URL <https://refed.org/articles/five-years-since-covid-19-upended-our-food-system/>, march 20. Accessed: 2026-09-22.
- Ren H, Huang T (2022) Opaque selling and inventory management in vertically differentiated markets. *Manufacturing & Service Operations Management* 24(5):2543–2557.
- Riesenegger L, Hübner A (2022) Reducing food waste at retail stores—an explorative study. *Sustainability* 14(5):2494, URL <http://dx.doi.org/10.3390/su14052494>.
- r/toogoodtogo (2024) What rating is too low? Reddit, URL https://www.reddit.com/r/toogoodtogo/comments/1fi4vio/what_rating_is_too_low/, accessed: 2026-03-01.
- Scarf HE, Arrow K, Karlin S (1957) *A min-max solution of an inventory problem* (Rand Corporation Santa Monica).
- Schneider M (2023) Does “too good to go” do more harm to your business than it does good? URL https://www.linkedin.com/posts/mausch_does-too-good-to-go-do-more-harm-to-your-activity-7130464942930972672-AttC/, LinkedIn post.
- Shapiro C (1983) Premiums for high quality products as returns to reputations. *The Quarterly Journal of Economics* 98(4):659–680.
- Shin D, Vaccari S, Zeevi A (2023) Dynamic pricing with online reviews. *Management Science* 69(2):824–845.
- TGTG (2024) World’s largest online marketplace for surplus food, Too Good To Go launches down under to help fight food waste. URL <https://www.toogoodtogo.com/en-au/press/melbourne-launch>, august 28. Accessed: 2026-09-22.
- Toeniskoetter C (2022) A new app aims to tackle food waste by linking eateries with surplus food to consumers. *The New York Times* URL <https://www.nytimes.com/2022/09/20/climate/food-waste-app.html>, accessed: 2024-09-18.
- Too Good To Go (2025a) Case study: Kismet Bagels reduces waste & grows customer base. URL <https://www.toogoodtogo.com/en-us/blog/kismet-bagels>, company blog post.
- Too Good To Go (2025b) How Pemberton Farms cut waste without adding work. URL <https://www.toogoodtogo.com/en-us/blog/pemberton-farms-grocery>, company blog post.
- Too Good To Go (2026) How does the too good to go app work? URL <https://www.toogoodtogo.com/en-us/how-does-the-app-work>, accessed: 2026-07-25.
- Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty* 5:297–323.
- United Nations (2024) Food waste index report 2024. URL <https://wedocs.unep.org/20.500.11822/45230>.
- Wu Q, Honhon D (2023) Don’t waste that free lettuce! Impact of BOGOF promotions on retail profit and food waste. *Production and Operations Management* 32(2):501–523.
- WWF (2024) Food waste. URL <https://www.worldwildlife.org/initiatives/food-waste>, accessed: 2024-09-17.
- Yang L, Yu M (2025) Too good to go: Combating food waste with surprise clearance. *Management Science* 72(3):2307–2325, URL <http://dx.doi.org/10.1287/mnsc.2023.03001>.
- Yang M, Zheng Z, Moorkerjee V (2019) Prescribing response strategies to manage customer opinions: A stochastic differential equation approach. *Information Systems Research* 30(2):351–374.
- Zhang Z, Joseph K, Subramaniam R (2015) Probabilistic selling in quality-differentiated markets. *Management Science* 61(8):1959–1977.

Appendix A: An Example Value Function for LVA

Suppose the concave envelope takes the form $\hat{v}(\ell) = \theta\sqrt{\ell}$ with $\theta > 0$. Then $\hat{v}'(\ell) = \theta/(2\sqrt{\ell})$, and the interior target in (21) is $\ell^\circ(D) = \left(\frac{\omega\theta}{2cD}\right)^2$. Substituting into (23) yields the explicit Stage-II rule

$$\tilde{\ell}(D, Q) = \left[\max \left\{ \frac{Q}{D}, \left(\frac{\omega\theta}{2cD} \right)^2 \right\} \right]_x, \quad D = \min\{\alpha r - \beta p + \epsilon, y\}.$$

In this square-root case, both the supplementation target and the marginal scarcity cost admit transparent power laws. First, $\ell^\circ(D) \propto D^{-2}$: when accepted reservations increase, the LVA target value per bag drops rapidly, reflecting diminishing marginal satisfaction from additional value. Second, for $D > 0$, the marginal scarcity cost in (22) reduces to

$$m(D, Q) = \begin{cases} \frac{\omega^2\theta^2}{4cD^2}, & Q \leq D\ell^\circ(D), \\ \frac{\omega\theta}{2} \cdot \frac{\sqrt{Q}}{D^{3/2}}, & D\ell^\circ(D) < Q < D\bar{x}, \\ 0, & Q \geq D\bar{x}. \end{cases}$$

The first regime ($Q \leq D\ell^\circ(D)$) corresponds to topping up to the target; the marginal scarcity cost is then exactly the per-bag top-up cost $c\ell^\circ(D)$, which decreases in D because the target itself falls with demand. The second regime corresponds to using surplus only; here $m(D, Q)$ captures the reputational loss from diluting a fixed surplus pool across more bags, and its $D^{-3/2}$ dependence shows that the dilution penalty becomes much smaller when demand is large. The square-root specification provides closed-form Stage-II targets and scarcity costs. The joint Stage-I pair $(\tilde{p}, \tilde{y})$ is then obtained through the bounded search in (20).

Appendix B: Proofs

B.1. Proof of Proposition 1

Let $J_k(r)$ denote the optimal k -period discounted value function with terminal value 0. Then $J_0 \equiv 0$ and, for $k \geq 0$,

$$J_{k+1}(r) = \max_{\substack{p \in [0, \bar{p}] \\ y \in [\underline{y}, \bar{y}]}} \mathbb{E}_\epsilon \left[pD(r, p, y, \epsilon) + \mathbb{E}_Q \left[\max_{\phi(\cdot)} \left\{ -C[\phi(\cdot), D(r, p, y, \epsilon), Q] + \gamma J_k(\delta(V[\phi] - p) + (1 - \delta)r) \right\} \right] \right].$$

We first show by induction that each J_k is weakly increasing in r .

Fix $r' > r$ and $k \geq 0$. Let $(p_k^*(r), y_k^*(r))$ be an optimal Stage-I pair for $J_{k+1}(r)$. For each (ϵ, Q) , let $\phi_k^*(r, \epsilon, Q)$ be an optimizer of the inner Stage-II problem at $(r, p_k^*(r), y_k^*(r), \epsilon, Q)$, and let

$$\ell^*(\epsilon, Q) := \int_{\mathcal{X}} x \phi_k^*(r, \epsilon, Q)(x) dx$$

denote its mean. At state r' , consider the feasible price

$$p' := \min \left\{ \bar{p}, p_k^*(r) + \frac{\alpha}{\beta}(r' - r) \right\}$$

and the same reservation cap $y_k^*(r)$. After observing the same realization (ϵ, Q) , use the same Stage-II allocation $\phi_k^*(r, \epsilon, Q)$. Define $d := \alpha r - \beta p_k^*(r) + \epsilon$, $d' := \alpha r' - \beta p' + \epsilon$, $D := \min\{d, y_k^*(r)\}$, and $D' := \min\{d', y_k^*(r)\}$. By construction, $d' \geq d$ and therefore $D' \geq D$ for every ϵ . Moreover, by the active-cycle support restriction in Section 3.3, $d > 0$ under the policies considered. For the continuation value, the next-period ratings are

$$r_{\text{next}} = (1 - \delta)r + \delta(V[\phi_k^*] - p_k^*(r)) \quad \text{and} \quad r'_{\text{next}} = (1 - \delta)r' + \delta(V[\phi_k^*] - p'),$$

so

$$r'_{\text{next}} - r_{\text{next}} = (1 - \delta)(r' - r) - \delta(p' - p_k^*(r)) \geq \frac{(1 - \delta)\beta - \delta\alpha}{\beta}(r' - r) \geq 0,$$

where the last inequality uses Assumption 1. By the induction hypothesis, J_k is weakly increasing, so $\gamma J_k(r'_{\text{next}}) \geq \gamma J_k(r_{\text{next}})$. For the one-period payoff, write

$$\Delta\Pi := p'D' - p_k^*(r)D - c[(\ell^*D' - Q)^+ - (\ell^*D - Q)^+].$$

Because $x \mapsto x^+$ is 1-Lipschitz and $D' \geq D$, we have $(\ell^*D' - Q)^+ - (\ell^*D - Q)^+ \leq \ell^*(D' - D)$, hence $\Delta\Pi \geq (p' - c\ell^*)D' - (p_k^*(r) - c\ell^*)D$. If $p' = p_k^*(r) + \frac{\alpha}{\beta}(r' - r) \leq \bar{p}$, then $D' = D$, so $\Delta\Pi \geq (p' - p_k^*(r))D \geq 0$. If instead $p' = \bar{p} < p_k^*(r) + \frac{\alpha}{\beta}(r' - r)$, then Assumption 1 and $\ell^* \leq \bar{x}$ imply $\bar{p} - c\ell^* \geq 0$, and therefore

$$\Delta\Pi \geq (\bar{p} - c\ell^*)D' - (p_k^*(r) - c\ell^*)D \geq (\bar{p} - c\ell^*)D - (p_k^*(r) - c\ell^*)D = (\bar{p} - p_k^*(r))D \geq 0.$$

Thus, for every realization (ϵ, Q) , the constructed feasible decision at state r' yields a weakly larger current payoff and a weakly larger continuation value than the optimal decision at state r . Taking expectations gives $J_{k+1}(r') \geq J_{k+1}(r)$. Since $J_0 \equiv 0$, every J_k is weakly increasing. It remains to pass to the infinite horizon. Fix an initial state r and any admissible policy π . Let $R_t^\pi := p_t^\pi D_t^\pi - C[\phi_t^\pi(\cdot), D_t^\pi, Q_t]$ denote the period- t payoff, and let $\ell_t^\pi := \int_{\mathcal{X}} x \phi_t^\pi(x) dx \in \mathcal{X}$ be the corresponding mean bag value. Since $V[\phi_t^\pi] \leq v(\bar{x})$ and $p_t^\pi \geq 0$,

$$r_{t+1}^\pi = (1 - \delta)r_t^\pi + \delta(V[\phi_t^\pi] - p_t^\pi) \leq (1 - \delta)r_t^\pi + \delta v(\bar{x}),$$

which implies $r_t^\pi \leq \bar{r}(r) := \max\{r, v(\bar{x})\}$ for all t . Under the active-cycle demand restriction,

$$|R_t^\pi| \leq p_t^\pi D_t^\pi + c(\ell_t^\pi D_t^\pi - Q_t)^+ \leq (\bar{p} + c\bar{x})D_t^\pi \leq (\bar{p} + c\bar{x})\bar{y}.$$

Hence $M := (\bar{p} + c\bar{x})\bar{y} < \infty$ satisfies $\mathbb{E}|R_t^\pi| \leq M$ for all t and all admissible π . Therefore,

$$\sup_{\pi} \mathbb{E} \left[\sum_{t=k}^{\infty} \gamma^t |R_t^\pi| \middle| r_0 = r \right] \leq \frac{\gamma^k}{1 - \gamma} M \xrightarrow{k \rightarrow \infty} 0.$$

So the k -period value functions converge pointwise to the infinite-horizon value function: $J_k(r) \rightarrow J^*(r)$. Since each J_k is weakly increasing, the limit $J^*(r)$ is also weakly increasing. $\square$

B.2. Proof of Corollary 1

Proof. Fix (r, p, D, Q) and let $\bar{\phi}(\cdot)$ be any optimal Stage-II distribution. Let

$$\bar{\ell} := \int_{\mathcal{X}} x \bar{\phi}(x) dx$$

be its mean bag value. Among all distributions on $\mathcal{X}$ with mean $\bar{\ell}$, choose one that maximizes $V[\phi]$; denote it by $\phi^*(\cdot)$. Such a distribution exists because the feasible set is tight on the compact interval $\mathcal{X}$ and the objective is continuous.

The Stage-II objective is

$$W(\phi(\cdot); r, p, D, Q) = -C[\phi(\cdot), D, Q] + \gamma J^*(\delta(V[\phi(\cdot)] - p) + (1 - \delta)r).$$

Because ϕ^* and $\bar{\phi}$ have the same mean, their supplementation costs are identical: $C[\phi^*(\cdot), D, Q] = C[\bar{\phi}(\cdot), D, Q]$. By construction, $V[\phi^*(\cdot)] \geq V[\bar{\phi}(\cdot)]$. The common price term $-p$ appears in both rating arguments, so Proposition 1 implies

$$J^*(\delta(V[\phi^*(\cdot)] - p) + (1 - \delta)r) \geq J^*(\delta(V[\bar{\phi}(\cdot)] - p) + (1 - \delta)r).$$

Thus $W(\phi^*(\cdot); r, p, D, Q) \geq W(\bar{\phi}(\cdot); r, p, D, Q)$. Since $\bar{\phi}$ was optimal, ϕ^* is also optimal and maximizes $V[\phi]$ among distributions with the same mean. $\square$

B.3. Proof of Proposition 2 and Corollary 2

Proof of Proposition 2. Interpret ϕ in (12) as a probability measure on $[\underline{x}, \bar{x}]$. When ϕ admits a density with respect to Lebesgue measure, we use the notation $\phi(x) dx$ as shorthand. Fix $\ell \in [\underline{x}, \bar{x}]$, and let ϕ be any feasible distribution satisfying $\int_{\mathcal{X}} x \phi(dx) = \ell$.

Because $\hat{v}$ is the upper concave envelope of v , we have $\hat{v}(x) \geq v(x) \forall x \in [\underline{x}, \bar{x}]$. Hence

$$V[\phi] = \int_{\mathcal{X}} v(x) \phi(dx) \leq \int_{\mathcal{X}} \hat{v}(x) \phi(dx).$$

Since $\hat{v}$ is concave and ϕ has mean ℓ , Jensen's inequality implies

$$\int_{\mathcal{X}} \hat{v}(x) \phi(dx) \leq \hat{v}\left(\int_{\mathcal{X}} x \phi(dx)\right) = \hat{v}(\ell).$$

Therefore every feasible ϕ satisfies $V[\phi] \leq \hat{v}(\ell)$. So the optimal value of (12) is at most $\hat{v}(\ell)$. It remains to construct a feasible distribution attaining $\hat{v}(\ell)$. Let $\mathcal{O} := \{x \in [\underline{x}, \bar{x}] : v(x) = \hat{v}(x)\}$. If $\ell \in \mathcal{O}$, then the degenerate distribution $\phi^*(\cdot) = \Delta_{\ell}$ is feasible, has mean ℓ , and satisfies $V[\phi^*] = v(\ell) = \hat{v}(\ell)$. Hence Δ_{ℓ} is optimal. Now suppose $\ell \notin \mathcal{O}$, equivalently $\hat{v}(\ell) > v(\ell)$. Since v is continuous, the set $\{x \in [\underline{x}, \bar{x}] : \hat{v}(x) > v(x)\}$ is open, so ℓ belongs to some open interval $(x^-, x^+) \subseteq [\underline{x}, \bar{x}]$. By the standard geometry of the smallest concave majorant on an interval, $\hat{v}$ is affine on $[x^-, x^+]$ and satisfies $\hat{v}(x^-) = v(x^-)$ and $\hat{v}(x^+) = v(x^+)$. Define $a := \frac{x^+ - \ell}{x^+ - x^-}$ and $1 - a := \frac{\ell - x^-}{x^+ - x^-}$, and let $\phi^*(\cdot) := a \Delta_{x^-} + (1 - a) \Delta_{x^+}$. Then ϕ^* is feasible and has mean

$$\int_{\mathcal{X}} x \phi^*(dx) = ax^- + (1 - a)x^+ = \ell.$$

Using the contact conditions and the affinity of $\hat{v}$ on $[x^-, x^+]$, we obtain

$$V[\phi^*] = av(x^-) + (1 - a)v(x^+) = a\hat{v}(x^-) + (1 - a)\hat{v}(x^+) = \hat{v}(\ell).$$

Thus ϕ^* is optimal. We conclude that the optimal value of (12) is $\hat{v}(\ell)$ and that there exists an optimal solution supported on at most two points. This selected optimizer is the $\phi^*(\cdot | \ell)$ stated in the proposition. Because v is nondecreasing, its upper concave envelope $\hat{v}$ is also nondecreasing. $\square$

Proof of Corollary 2. If $v(\cdot)$ is concave on $\mathcal{X}$, then its upper concave envelope coincides with v itself: $\hat{v}(\ell) = v(\ell)$ for all $\ell \in \mathcal{X}$. Therefore $\ell \in \mathcal{O}$ for every $\ell \in \mathcal{X}$, and the degenerate distribution $\phi^*(\cdot | \ell) = \Delta_{\ell}$ is optimal by the first case of the proof of Proposition 2. $\square$

B.4. Proof of Proposition 3

Proof. Fix (r, p, D, Q) and define the objective

$$\Psi(\ell; r, p, D, Q) := \gamma J^*(\delta(\hat{v}(\ell) - p) + (1 - \delta)r) - c(\ell D - Q)^+, \quad \ell \in \mathcal{X}.$$

Let $\mathcal{L}^*(r, p, D, Q) := \arg \max_{\ell \in \mathcal{X}} \Psi(\ell; r, p, D, Q)$ denote the maximizer set. When $\mathcal{L}^*$ is not a singleton, we use the maximal selector $\bar{\ell}^*(r, p, D, Q) := \sup \mathcal{L}^*(r, p, D, Q)$. The maximizer set is nonempty because $\mathcal{X}$ is compact and the objective is upper semicontinuous in ℓ ; continuity of J^* follows from Lemma C.1 below.

Part 1 (monotonicity in Q). Fix (r, p, D) and define $g(\ell, Q) := -c(\ell D - Q)^+$. For any $Q_2 \geq Q_1$, consider

$$\Delta(\ell) := g(\ell, Q_2) - g(\ell, Q_1) = -c(\ell D - Q_2)^+ + c(\ell D - Q_1)^+.$$

By the active-cycle assumption, $D > 0$. Using the identity

$$(\ell D - Q)^+ = \begin{cases} 0, & \ell \leq Q/D, \\ \ell D - Q, & \ell > Q/D, \end{cases}$$

we obtain

$$\Delta(\ell) = \begin{cases} 0, & \ell \leq Q_1/D, \\ c(\ell D - Q_1), & Q_1/D < \ell \leq Q_2/D, \\ c(Q_2 - Q_1), & \ell > Q_2/D, \end{cases}$$

which is nondecreasing in ℓ . Hence $g(\ell, Q)$ has increasing differences in (ℓ, Q) . Because the remaining term $\gamma J^*(\delta(\hat{v}(\ell) - p) + (1 - \delta)r)$ does not depend on Q , $\Psi(\ell; r, p, D, Q)$ also has increasing differences in (ℓ, Q) . By Topkis' Monotonicity Theorem, the maximal selector $\bar{\ell}^*(r, p, D, Q)$ is nondecreasing in Q . This proves (ii).

Part 2 (monotonicity in D). Fix (r, p, Q) and view $g(\ell, D) := -c(\ell D - Q)^+$ for $D > 0$. For $D_2 \geq D_1 > 0$, define

$$\tilde{\Delta}(\ell) := g(\ell, D_2) - g(\ell, D_1) = -c(\ell D_2 - Q)^+ + c(\ell D_1 - Q)^+.$$

For any fixed $\ell \in \mathcal{X}$, the mapping $D \mapsto c(\ell D - Q)^+$ is convex and nondecreasing, and its right-derivative equals $c\ell \mathbf{1}\{\ell D \geq Q\}$. Consequently, for $D_2 \geq D_1$,

$$c(\ell D_2 - Q)^+ - c(\ell D_1 - Q)^+ = c \int_{D_1}^{D_2} \ell \mathbf{1}\{\ell s \geq Q\} ds$$

is nondecreasing in ℓ , which implies $\tilde{\Delta}(\ell)$ is nonincreasing in ℓ .

Therefore $g(\ell, D)$ has decreasing differences in (ℓ, D) on $D > 0$. Since the continuation-value term is independent of D , Ψ has decreasing differences in (ℓ, D) . Applying Topkis' Monotonicity Theorem yields that the maximal selector $\bar{\ell}^*(r, p, D, Q)$ is nonincreasing in D for $D > 0$. This proves (i).

Part 3 (monotonicity in p and r). Fix (D, Q) and note that

$$\Psi(\ell; r, p, D, Q) = \gamma J^*(\delta(\hat{v}(\ell) - p) + (1 - \delta)r) - c(\ell D - Q)^+.$$

Assume now that J^* is concave, as stated in Proposition 3. Because $\hat{v}$ is nondecreasing, for any $p_2 \geq p_1$ we can write

$$\Psi(\ell; r, p_2, D, Q) - \Psi(\ell; r, p_1, D, Q) = \gamma \left(J^*(x(\ell) - \delta(p_2 - p_1)) - J^*(x(\ell)) \right),$$

where $x(\ell) := \delta(\hat{v}(\ell) - p_1) + (1 - \delta)r$ is nondecreasing in ℓ . For any fixed $\Delta \geq 0$, concavity of J^* implies that the map $z \mapsto J^*(z - \Delta) - J^*(z)$ is nondecreasing. Hence the display above is nondecreasing in ℓ , so Ψ has increasing differences in (ℓ, p) . Similarly, for any $r_2 \geq r_1$,

$$\Psi(\ell; r_2, p, D, Q) - \Psi(\ell; r_1, p, D, Q) = \gamma \left(J^*(y(\ell) + (1 - \delta)(r_2 - r_1)) - J^*(y(\ell)) \right),$$

where $y(\ell) := \delta(\hat{v}(\ell) - p) + (1 - \delta)r_1$ is nondecreasing in ℓ . For any fixed $\Delta \geq 0$, concavity of J^* implies that the map $z \mapsto J^*(z + \Delta) - J^*(z)$ is nonincreasing. Therefore the display above is nonincreasing in ℓ , so Ψ has decreasing differences in (ℓ, r) .

The cost term does not depend on p or r . Applying Topkis' Monotonicity Theorem, the maximal selector $\bar{\ell}^*$ can be chosen to be nondecreasing in p and nonincreasing in r . This proves (iii) and (iv). $\square$

Appendix C: Proofs for Section 5

All proofs use the active-cycle support condition stated in Section 3.3; equivalently, one may restrict the price set to state-dependent admissible prices that keep latent demand positive.

C.1. Proof of Proposition 4

Proof. Fix a rating r . The proof proceeds in four steps.

Step 1. Substitute the affine value approximation. Let $\tilde{J}(r) = \tilde{J}_0 + kr$ and define $\omega := \gamma\delta k$. By assumption, $\omega > 0$; recall also that $\hat{v}$ is concave and increasing. Evaluating the Bellman operator at $\tilde{J}$ and collecting terms gives

$$(\mathcal{T}\tilde{J})(r) = \gamma\tilde{J}_0 + \gamma k(1 - \delta)r + \max_{\substack{0 \leq p \leq \bar{p} \\ y \leq y \leq \bar{y}}} \mathbb{E}_{\epsilon, Q} \left[pD(r, p, y, \epsilon) - \omega p + \max_{\ell \in \mathcal{X}} \{ \omega \hat{v}(\ell) - c(\ell D(r, p, y, \epsilon) - Q)^+ \} \right].$$

Since $\gamma\tilde{J}_0 + \gamma k(1 - \delta)r$ does not depend on the controls, the control problem decomposes into a Stage-II choice of ℓ given (r, p, y, ϵ, Q) and a Stage-I choice of (p, y) . Define the maximized Stage-II value

$$\Phi(D, Q) := \max_{\ell \in \mathcal{X}} \{ \omega \hat{v}(\ell) - c(\ell D - Q)^+ \}.$$

Substituting the maximized Stage-II value back into the LVA objective gives the Stage-I problem

$$(\tilde{p}(r), \tilde{y}(r)) \in \arg \max_{\substack{0 \leq p \leq \bar{p} \\ y \leq y \leq \bar{y}}} \mathbb{E}_{\epsilon, Q} \left[pD(r, p, y, \epsilon) - \omega p + \Phi(D(r, p, y, \epsilon), Q) \right]. \quad (\text{C.1})$$

Step 2. Stage-II rule. Fix (r, p, y, ϵ, Q) and write $D := D(r, p, y, \epsilon)$. By the active-cycle assumption, $D > 0$. Define

$$\Psi(\ell; D, Q) := \omega \hat{v}(\ell) - c(\ell D - Q)^+, \quad \ell \in \mathcal{X}.$$

If $Q/D \geq \bar{x}$, then every feasible $\ell \in \mathcal{X}$ satisfies $\ell \leq Q/D$, so $(\ell D - Q)^+ = 0$ for all feasible ℓ and $\Psi(\ell; D, Q) = \omega \hat{v}(\ell)$. Because $\hat{v}$ is increasing, the maximizer is $\bar{\ell} = \bar{x}$, which agrees with (23).

Hence suppose $Q/D < \bar{x}$. On the no-supplement region $\{\ell \in \mathcal{X} : \ell \leq Q/D\}$, we have $(\ell D - Q)^+ = 0$, so $\Psi(\ell; D, Q) = \omega \hat{v}(\ell)$, which is increasing in ℓ . Thus, whenever this region is nonempty, its best feasible point is $\ell = Q/D$. On the supplementation region $\{\ell \in \mathcal{X} : \ell \geq Q/D\}$, we have

$$\Psi(\ell; D, Q) = \omega \hat{v}(\ell) - cD\ell + cQ.$$

The function $g_D(\ell) := \omega \hat{v}(\ell) - cD\ell$ is concave. Let $\ell^\circ(D)$ be its maximal optimizer over $\mathcal{X}$, as defined immediately before Proposition 4. Hence $g_D(\ell)$ is weakly increasing up to $\ell^\circ(D)$ and weakly decreasing thereafter, so its maximal optimizer over the restricted feasible interval

$$\mathcal{X} \cap [Q/D, \infty) = [\max\{\underline{x}, Q/D\}, \bar{x}]$$

is the projection of $\ell^\circ(D)$ onto that interval, namely

$$\left[\max \left\{ \frac{Q}{D}, \ell^\circ(D) \right\} \right]_{\mathcal{X}}.$$

When $Q/D \in \mathcal{X}$, the no-supplement and supplementation objectives coincide at $\ell = Q/D$, so this restricted-region maximizer weakly dominates the best no-supplement point. When $Q/D < \underline{x}$, the no-supplement region is empty. Therefore,

$$\tilde{\ell}(D, Q) = \left[\max \left\{ \frac{Q}{D}, \ell^\circ(D) \right\} \right]_{\mathcal{X}},$$

proving (23).

Step 3. Marginal scarcity cost. At points where Φ is differentiable in D , define the marginal scarcity cost as $m(D, Q) := -\partial_D \Phi(D, Q)$. The three branches follow from the envelope theorem. If $Q < D\ell^\circ(D)$, then the Stage-II optimizer is $\ell = \ell^\circ(D)$ and

$$\Phi(D, Q) = \omega \hat{v}(\ell^\circ(D)) - c(D\ell^\circ(D) - Q),$$

so $\Phi_D(D, Q) = -c\ell^\circ(D)$. If $D\ell^\circ(D) < Q < D\bar{x}$, then the optimizer is $\ell = Q/D$ and $\Phi(D, Q) = \omega \hat{v}(Q/D)$, so

$$\Phi_D(D, Q) = -\omega \hat{v}'\left(\frac{Q}{D}\right) \frac{Q}{D^2}.$$

If $Q > D\bar{x}$, then the optimizer is $\ell = \bar{x}$ and $\Phi(D, Q) = \omega \hat{v}(\bar{x})$, hence $\Phi_D(D, Q) = 0$. Therefore, the marginal scarcity cost is exactly (22).

Step 4. Derive the Stage-I FOCs. Under the maintained regularity in Section 3.3,

$$\frac{\partial D(r, p, y, \epsilon)}{\partial p} = -\beta \mathbf{1}_{\{d(r, p, \epsilon) < y\}}, \quad \frac{\partial D(r, p, y, \epsilon)}{\partial y} = \mathbf{1}_{\{d(r, p, \epsilon) > y\}}.$$

Using $\Phi_D(D, Q) = -m(D, Q)$ in (C.1), the first-order condition with respect to price is

$$0 = \mathbb{E}_{\epsilon, Q} [D(r, p, y, \epsilon) - \omega - \beta(p - m(D(r, p, y, \epsilon), Q)) \mathbf{1}_{\{d(r, p, \epsilon) < y\}}].$$

On the event $\{d(r, p, \epsilon) < y\}$, $D(r, p, y, \epsilon) = d(r, p, \epsilon)$. Evaluating at $(p, y) = (\tilde{p}, \tilde{y})$ gives the price first-order condition. The first-order condition with respect to the cap is

$$0 = \mathbb{E}_{\epsilon, Q} [(p - m(D(r, p, y, \epsilon), Q)) \mathbf{1}_{\{d(r, p, \epsilon) > y\}}].$$

On the event $\{d(r, p, \epsilon) > y\}$, $D(r, p, y, \epsilon) = y$. Evaluating at $(p, y) = (\tilde{p}, \tilde{y})$ gives the cap first-order condition.

At $(p, y) = (\tilde{p}, \tilde{y})$, the cap first-order condition factors into

$$\Pr(d(r, \tilde{p}, \epsilon) > \tilde{y}) (\tilde{p} - \mathbb{E}_Q[m(\tilde{y}, Q)]) = 0,$$

and division by the positive leading probability gives $\tilde{p} = \mathbb{E}_Q[m(\tilde{y}, Q)]$. For comparison, when the cap is fixed at a nonbinding upper bound, $D(r, \tilde{p}, \tilde{y}, \epsilon) = d(r, \tilde{p}, \epsilon)$ and $\mathbf{1}_{\{d(r, \tilde{p}, \epsilon) < \tilde{y}\}} = 1$. The price condition becomes

$$\alpha r - \beta \tilde{p} - \omega - \beta \tilde{p} + \beta \mathbb{E}_{\epsilon, Q}[m(D(r, \tilde{p}, \tilde{y}, \epsilon), Q)] = 0,$$

which rearranges to (25). □

C.2. Proof of Proposition 5

Proof. Let $W := \bar{r} - \underline{x}$, using the invariant analytical rating interval defined in Section 3.3. If $W = 0$, choose $\tilde{J} = J^*$ on the singleton state space and the result is immediate. Hence suppose $W > 0$. First we show that the optimal value function is Lipschitz continuous in the rating state on this interval.

LEMMA C.1. Assume the feasible sets satisfy $p \in [0, \bar{p}]$, $y \in [\underline{y}, \bar{y}]$, and $\ell \in \mathcal{X}$. Then the optimal infinite-horizon value function J^* is Lipschitz continuous in r with constant

$$L := \frac{\alpha(\bar{p} + c\bar{x})}{1 - \gamma(1 - \delta)}. \quad (\text{C.2})$$

In particular, for any $r, r' \in [\underline{r}, \bar{r}]$, there is $|J^*(r') - J^*(r)| \leq L|r' - r|$.

Proof of Lemma C.1. We use value iteration and induction. Define the finite-horizon value functions $J_0^*(r) \equiv 0$ and, for $k \geq 0$,

$$J_{k+1}^*(r) = \max_{\substack{0 \leq p \leq \bar{p} \\ \underline{y} \leq y \leq \bar{y}}} \mathbb{E}_\epsilon \left[pD(r, p, y, \epsilon) + \mathbb{E}_Q \max_{\ell \in \mathcal{X}} \left\{ \gamma J_k^*(\delta(\hat{v}(\ell) - p) + (1 - \delta)r) - c(\ell D(r, p, y, \epsilon) - Q)^+ \right\} \right].$$

By the contraction property of the Bellman operator, $J_k^*(r) \rightarrow J^*(r)$ as $k \rightarrow \infty$ for each r . We claim that for all $k \geq 0$ and all $r, r' \in [\underline{r}, \bar{r}]$,

$$|J_k^*(r') - J_k^*(r)| \leq L|r' - r|. \quad (\text{C.3})$$

The base case $k = 0$ is immediate since J_0^* is constant.

Fix $k \geq 0$ and assume (C.3) holds for J_k^* . Let $r' > r$ and denote $\Delta r := r' - r > 0$. Let $(p_k^*(r'), y_k^*(r'))$ be an optimal Stage-I pair attaining $J_{k+1}^*(r')$. For each realization of (ϵ, Q) , let $\ell_k^*(r', p_k^*(r'), y_k^*(r'), \epsilon, Q)$ be a measurable maximizer of the inner problem at state r' and pair $(p_k^*(r'), y_k^*(r'))$. For notational brevity, write $\ell_k^*(r', \epsilon, Q) := \ell_k^*(r', p_k^*(r'), y_k^*(r'), \epsilon, Q)$.

Since $J_{k+1}^*(r)$ is the supremum over all admissible controls, it is at least the value obtained by using the same controls $(p_k^*(r'), y_k^*(r'), \ell_k^*(r', \epsilon, Q))$ at state r . Define

$$D'_\epsilon := D(r', p_k^*(r'), y_k^*(r'), \epsilon), \quad D_\epsilon := D(r, p_k^*(r'), y_k^*(r'), \epsilon).$$

Therefore, defining

$$x_{r'} := \delta(\hat{v}(\ell_k^*(r', \epsilon, Q)) - p_k^*(r')) + (1 - \delta)r' \quad \text{and} \quad x_r := \delta(\hat{v}(\ell_k^*(r', \epsilon, Q)) - p_k^*(r')) + (1 - \delta)r,$$

we obtain

$$\begin{aligned} J_{k+1}^*(r') - J_{k+1}^*(r) &\leq \mathbb{E}_\epsilon \left[p_k^*(r')D'_\epsilon - p_k^*(r')D_\epsilon \right. \\ &\quad \left. + \mathbb{E}_Q \left\{ \gamma(J_k^*(x_{r'}) - J_k^*(x_r)) - c \left[(\ell_k^*(r', \epsilon, Q)D'_\epsilon - Q)^+ - (\ell_k^*(r', \epsilon, Q)D_\epsilon - Q)^+ \right] \right\} \right]. \end{aligned}$$

We bound the three differences inside the expectations. First, since $D(r, p, y, \epsilon) = \min\{\alpha r - \beta p + \epsilon, y\}$ and $z \mapsto \min\{z, y\}$ is nondecreasing and 1-Lipschitz,

$$0 \leq D(r', p, y, \epsilon) - D(r, p, y, \epsilon) \leq \alpha \Delta r.$$

Thus

$$p_k^*(r') \left(D(r', p_k^*(r'), y_k^*(r'), \epsilon) - D(r, p_k^*(r'), y_k^*(r'), \epsilon) \right) \leq \alpha p_k^*(r') \Delta r.$$

Second, the next-period rating differs only through the term $(1 - \delta)r$, so $x_{r'} - x_r = (1 - \delta)\Delta r$. By the induction hypothesis (C.3),

$$\gamma(J_k^*(x_{r'}) - J_k^*(x_r)) \leq \gamma L(1 - \delta)\Delta r.$$

Finally, since the positive-part operator is 1-Lipschitz,

$$\begin{aligned} & c(\ell_k^*(r', \epsilon, Q)D(r, p_k^*(r'), y_k^*(r'), \epsilon) - Q)^+ - c(\ell_k^*(r', \epsilon, Q)D(r', p_k^*(r'), y_k^*(r'), \epsilon) - Q)^+ \\ & \leq c \left| \ell_k^*(r', \epsilon, Q) \left(D(r, p_k^*(r'), y_k^*(r'), \epsilon) - D(r', p_k^*(r'), y_k^*(r'), \epsilon) \right) \right| \\ & \leq \alpha c \ell_k^*(r', \epsilon, Q) \Delta r \leq \alpha c \bar{x} \Delta r, \end{aligned}$$

where the last inequality uses $\ell_k^*(r', \epsilon, Q) \in \mathcal{X}$. Combining the bounds and taking expectations yields

$$\begin{aligned} J_{k+1}^*(r') - J_{k+1}^*(r) & \leq \alpha p_k^*(r') \Delta r + \gamma L(1 - \delta) \Delta r + \alpha c \bar{x} \Delta r \\ & \leq (\alpha \bar{p} + \gamma L(1 - \delta) + \alpha c \bar{x}) \Delta r = L \Delta r, \end{aligned}$$

where the last equality uses (C.2). This proves (C.3) for $r' > r$; the case $r' < r$ follows by symmetry. Letting $k \rightarrow \infty$ and using $J_k^*(r) \rightarrow J^*(r)$ for each r , we obtain $|J^*(r') - J^*(r)| \leq L|r' - r|$, which completes the proof. $\square$

With this Lipschitz bound in hand, the rest of the proof has two ingredients: we first construct an affine surrogate $\tilde{J}$ with small uniform error on $[\underline{r}, \bar{r}]$, and then translate that approximation error into a loss bound for the greedy policy induced by $\tilde{J}$.

Because $p \in [0, \bar{p}]$ and $\ell \in \mathcal{X}$, the one-period satisfaction $\hat{v}(\ell) - p$ lies in $[\underline{r}, \bar{r}]$. Therefore, whenever $r \in [\underline{r}, \bar{r}]$, the next rating

$$(1 - \delta)r + \delta(\hat{v}(\ell) - p)$$

is a convex combination of two points in $[\underline{r}, \bar{r}]$ and remains in the same interval. Hence $[\underline{r}, \bar{r}]$ is invariant under all admissible controls.

Choose $r_L := \underline{r} + \frac{W}{6}$ and $r_U := \bar{r} - \frac{W}{6}$, and let $\tilde{J}(r) = \tilde{J}_0 + kr$ be the chord through $(r_L, J^*(r_L))$ and $(r_U, J^*(r_U))$, where

$$k := \frac{J^*(r_U) - J^*(r_L)}{r_U - r_L} \quad \text{and} \quad \tilde{J}_0 := J^*(r_L) - kr_L.$$

This choice leaves endpoint strips of width $W/6$ and a middle interval of length $2W/3$, which balances the two approximation bounds below. By Proposition 1 and Lemma C.1,

$$0 \leq k = \frac{J^*(r_U) - J^*(r_L)}{r_U - r_L} \leq L.$$

Let $\tilde{\pi} = (\tilde{p}, \tilde{y}, \tilde{\ell})$ denote the greedy policy induced by $\tilde{J}$, and define $\varepsilon_J := \max_{r \in [\underline{r}, \bar{r}]} |J^*(r) - \tilde{J}(r)|$.

Step 1 (bounding ε_J). For any $r \in [r_L, r_U]$, write

$$r = (1 - s)r_L + sr_U, \quad \text{where} \quad s := \frac{r - r_L}{r_U - r_L} \in [0, 1].$$

Because $\tilde{J}$ is the chord through $(r_L, J^*(r_L))$ and $(r_U, J^*(r_U))$, there is $\tilde{J}(r) = (1 - s)J^*(r_L) + sJ^*(r_U)$. Therefore,

$$\begin{aligned} |J^*(r) - \tilde{J}(r)| & = |(1 - s)(J^*(r) - J^*(r_L)) - s(J^*(r_U) - J^*(r))| \\ & \leq (1 - s)|J^*(r) - J^*(r_L)| + s|J^*(r_U) - J^*(r)| \\ & \leq (1 - s)L(r - r_L) + sL(r_U - r) \\ & = 2Ls(1 - s)(r_U - r_L) \leq \frac{L}{2}(r_U - r_L) = \frac{LW}{3}. \end{aligned}$$

If $r \notin [r_L, r_U]$, let a denote the nearest endpoint of $[r_L, r_U]$ (so $a = r_L$ for $r \leq r_L$ and $a = r_U$ for $r \geq r_U$). Since $\tilde{J}(a) = J^*(a)$,

$$|J^*(r) - \tilde{J}(r)| \leq |J^*(r) - J^*(a)| + |\tilde{J}(r) - \tilde{J}(a)| \leq (L + |k|) |r - a| \leq 2L|r - a|.$$

Each outer strip has width $W/6$, so $|r - a| \leq W/6$ and therefore again $|J^*(r) - \tilde{J}(r)| \leq LW/3$. Hence $\varepsilon_J \leq \frac{LW}{3}$.

Step 2 (policy loss bound in terms of ε_J). We next show that

$$J^*(r) - J^{\tilde{\pi}}(r) \leq \frac{2\gamma}{1-\gamma} \varepsilon_J \text{ for all } r \in [\underline{r}, \bar{r}].$$

For any stationary policy $\pi = (p, y, \ell)$ and any bounded continuation value function $J(\cdot)$ on $[\underline{r}, \bar{r}]$, write $D_\pi(r, \epsilon) := D(r, p(r), y(r), \epsilon)$ and define the Bellman operator

$$(\mathcal{T}^\pi J)(r) := \mathbb{E}_\epsilon \left[p(r) D_\pi(r, \epsilon) + \mathbb{E}_Q \left\{ \gamma J(r_\pi^+(r, \epsilon, Q)) - c(\ell(r, p(r), D_\pi(r, \epsilon), Q)) D_\pi(r, \epsilon) - Q \right\}^+ \right],$$

where

$$r_\pi^+(r, \epsilon, Q) := \delta(\hat{v}(\ell(r, p(r), D_\pi(r, \epsilon), Q)) - p(r)) + (1 - \delta)r.$$

Throughout this step, we work on the feasible rating interval $[\underline{r}, \bar{r}]$, so $r_\pi^+(r, \epsilon, Q) \in [\underline{r}, \bar{r}]$ whenever $r \in [\underline{r}, \bar{r}]$.

By definition, the value under policy π is the fixed point of $\mathcal{T}^\pi$, i.e., $J^\pi = \mathcal{T}^\pi J^\pi$. Let π^* be an optimal stationary policy, so $J^* = J^{\pi^*} = \mathcal{T}^{\pi^*} J^*$, and define $M := \sup_{x \in [\underline{r}, \bar{r}]} (J^*(x) - J^{\tilde{\pi}}(x)) \geq 0$.

For any two bounded functions J_1, J_2 and any $r \in [\underline{r}, \bar{r}]$,

$$|(\mathcal{T}^\pi J_1)(r) - (\mathcal{T}^\pi J_2)(r)| \leq \gamma \sup_{x \in [\underline{r}, \bar{r}]} |J_1(x) - J_2(x)|.$$

Fix $r \in [\underline{r}, \bar{r}]$. Because $\tilde{\pi}$ is greedy with respect to $\tilde{J}$, the optimal policy π^* is a feasible comparison, so $(\mathcal{T}^{\tilde{\pi}} \tilde{J})(r) \geq (\mathcal{T}^{\pi^*} \tilde{J})(r)$. Therefore,

$$J^*(r) - J^{\tilde{\pi}}(r) = (\mathcal{T}^{\pi^*} J^*)(r) - (\mathcal{T}^{\tilde{\pi}} J^{\tilde{\pi}})(r) \leq (\mathcal{T}^{\pi^*} J^*)(r) - (\mathcal{T}^{\pi^*} \tilde{J})(r) + (\mathcal{T}^{\tilde{\pi}} \tilde{J})(r) - (\mathcal{T}^{\tilde{\pi}} J^{\tilde{\pi}})(r).$$

The first term is at most $\gamma \varepsilon_J$ by the contraction bound. For the second term, by definition of $\mathcal{T}^{\tilde{\pi}}$,

$$(\mathcal{T}^{\tilde{\pi}} \tilde{J})(r) - (\mathcal{T}^{\tilde{\pi}} J^{\tilde{\pi}})(r) = \gamma \mathbb{E} \left[\tilde{J}(r_\pi^+(r, \epsilon, Q)) - J^{\tilde{\pi}}(r_\pi^+(r, \epsilon, Q)) \right] \leq \gamma(\varepsilon_J + M),$$

because, for every $x \in [\underline{r}, \bar{r}]$,

$$\tilde{J}(x) - J^{\tilde{\pi}}(x) \leq |\tilde{J}(x) - J^*(x)| + (J^*(x) - J^{\tilde{\pi}}(x)) \leq \varepsilon_J + M.$$

Hence

$$J^*(r) - J^{\tilde{\pi}}(r) \leq 2\gamma \varepsilon_J + \gamma M \quad \forall r \in [\underline{r}, \bar{r}].$$

Taking the supremum over r on the left-hand side gives $M \leq 2\gamma \varepsilon_J + \gamma M$, hence $M \leq \frac{2\gamma}{1-\gamma} \varepsilon_J$. Finally, since $J^*(r) - J^{\tilde{\pi}}(r) \leq M$ for every r ,

$$J^*(r) - J^{\tilde{\pi}}(r) \leq \frac{2\gamma}{1-\gamma} \varepsilon_J \leq \frac{2\gamma}{1-\gamma} \cdot \frac{LW}{3} = \frac{2\gamma \alpha(\bar{p} + c\bar{x})}{3(1-\gamma)[1-\gamma(1-\delta)]} [\bar{p} + \hat{v}(\bar{x}) - \hat{v}(\underline{x})], \quad \forall r \in [\underline{r}, \bar{r}].$$

□

C.3. Proof of Proposition 6

Proof. Define the optimal Bellman operator $\mathcal{T}$ by

$$\mathcal{T}J(r) := \max_{\substack{p \in [0, \bar{p}] \\ y \in [\underline{y}, \bar{y}]}} \mathbb{E}_{\epsilon, Q} \left\{ pD(r, p, y, \epsilon) + \max_{\ell \in \mathcal{X}} [\gamma J(r^+(r, p, \ell)) - c(\ell D(r, p, y, \epsilon) - Q)^+] \right\},$$

where $r^+(r, p, \ell) := \delta(\hat{v}(\ell) - p) + (1 - \delta)r$, $\bar{d}(r, p) := \alpha r - \beta p$, $\bar{D}(r, p, y) := \min\{\bar{d}(r, p), y\}$, and $\bar{Q} := \mathbb{E}[Q]$. Define the fluid Bellman operator $\hat{\mathcal{T}}$ by

$$\hat{\mathcal{T}}J(r) := \max_{\substack{p \in [0, \bar{p}], y \in [\underline{y}, \bar{y}] \\ \ell \in \mathcal{X}}} \left\{ p\bar{D}(r, p, y) - c(\ell \bar{D}(r, p, y) - \bar{Q})^+ + \gamma J(r^+(r, p, \ell)) \right\}.$$

For the DFA policy $\hat{\pi} = \{\hat{p}(\cdot), \hat{y}(\cdot), \hat{\ell}(\cdot)\}$, define the true-system and fluid policy-evaluation operators by

$$(\mathcal{T}^{\hat{\pi}}J)(r) := \mathbb{E}_{\epsilon, Q} \left[\hat{p}(r)D(r, \hat{p}(r), \hat{y}(r), \epsilon) - c(\hat{\ell}(r)D(r, \hat{p}(r), \hat{y}(r), \epsilon) - Q)^+ + \gamma J(r^+(r, \hat{p}(r), \hat{\ell}(r))) \right],$$

and

$$(\hat{\mathcal{T}}^{\hat{\pi}}J)(r) := \hat{p}(r)\bar{D}(r, \hat{p}(r), \hat{y}(r)) - c(\hat{\ell}(r)\bar{D}(r, \hat{p}(r), \hat{y}(r)) - \bar{Q})^+ + \gamma J(r^+(r, \hat{p}(r), \hat{\ell}(r))).$$

Fix an admissible rating $r, p \in [0, \bar{p}]$, $y \in [\underline{y}, \bar{y}]$, and any function J for which the displayed quantities are finite. For each realization (ϵ, Q) , define

$$\Phi_J(r, p, y, \epsilon, Q) := \max_{\ell \in \mathcal{X}} \left\{ \gamma J(r^+(r, p, \ell)) - c[\ell D(r, p, y, \epsilon) - Q]^+ \right\}.$$

Using $|x^+ - y^+| \leq |x - y|$ and $|\max_x f(x) - \max_x g(x)| \leq \max_x |f(x) - g(x)|$, we obtain

$$\begin{aligned} |\Phi_J(r, p, y, \epsilon, Q) - \Phi_J(r, p, y, 0, \bar{Q})| &\leq \max_{\ell \in \mathcal{X}} \left| c(\ell D(r, p, y, \epsilon) - Q)^+ - c(\ell \bar{D}(r, p, y) - \bar{Q})^+ \right| \\ &\leq c \cdot \max_{\ell \in \mathcal{X}} |\ell(D(r, p, y, \epsilon) - \bar{D}(r, p, y)) - (Q - \bar{Q})| \\ &\leq c\bar{x}|\epsilon| + c|Q - \bar{Q}|. \end{aligned}$$

The last inequality uses

$$|D(r, p, y, \epsilon) - \bar{D}(r, p, y)| = |\min\{\bar{d}(r, p) + \epsilon, y\} - \min\{\bar{d}(r, p), y\}| \leq |\epsilon|.$$

Taking expectation over (ϵ, Q) yields

$$\left| \mathbb{E}_{\epsilon, Q} \Phi_J(r, p, y, \epsilon, Q) - \Phi_J(r, p, y, 0, \bar{Q}) \right| \leq c\bar{x} \mathbb{E}|\epsilon| + c \mathbb{E}|Q - \bar{Q}|. \quad (\text{C.4})$$

The revenue term satisfies $|pD(r, p, y, \epsilon) - p\bar{D}(r, p, y)| \leq \bar{p}|\epsilon|$. Combining this bound with (C.4) and $|\max_x f(x) - \max_x g(x)| \leq \max_x |f(x) - g(x)|$ gives the one-step perturbation bound

$$|\mathcal{T}J(r) - \hat{\mathcal{T}}J(r)| \leq \eta := (\bar{p} + c\bar{x})\mathbb{E}|\epsilon| + c\mathbb{E}|Q - \bar{Q}|, \quad \text{for all admissible } r. \quad (\text{C.5})$$

By the same argument,

$$|\mathcal{T}^{\hat{\pi}}J(r) - \hat{\mathcal{T}}^{\hat{\pi}}J(r)| \leq \eta, \quad \text{for all admissible } r. \quad (\text{C.6})$$

We now work with finite-horizon value functions. Define

$$J_0^*(r) = \hat{J}_0(r) = 0, \quad J_{n+1}^* := \mathcal{T}J_n^*, \quad \hat{J}_{n+1} := \hat{\mathcal{T}}\hat{J}_n,$$

and similarly

$$J_0^{\hat{\pi}}(r) = \hat{J}_0^{\hat{\pi}}(r) = 0, \quad J_{n+1}^{\hat{\pi}} := \mathcal{T}^{\hat{\pi}} J_n^{\hat{\pi}}, \quad \hat{J}_{n+1}^{\hat{\pi}} := \hat{\mathcal{T}}^{\hat{\pi}} \hat{J}_n^{\hat{\pi}}.$$

Let $M_n := \sup_{r \text{ admissible}} |J_n^*(r) - \hat{J}_n(r)|$. We claim that $M_n \leq \eta \sum_{t=0}^{n-1} \gamma^t = \eta \frac{1-\gamma^n}{1-\gamma}$. This is true for $n = 0$. If it holds for n , then for any admissible r ,

$$|J_{n+1}^*(r) - \hat{J}_{n+1}(r)| = |\mathcal{T} J_n^*(r) - \hat{\mathcal{T}} \hat{J}_n(r)| \leq |\mathcal{T} J_n^*(r) - \hat{\mathcal{T}} J_n^*(r)| + |\hat{\mathcal{T}} J_n^*(r) - \hat{\mathcal{T}} \hat{J}_n(r)| \leq \eta + \gamma M_n.$$

Therefore

$$M_{n+1} \leq \eta + \gamma M_n \leq \eta \sum_{t=0}^n \gamma^t.$$

Hence

$$|J_n^*(r) - \hat{J}_n(r)| \leq \eta \frac{1-\gamma^n}{1-\gamma}, \quad \text{for all admissible } r. \quad (\text{C.7})$$

Exactly the same induction, now using (C.6), yields

$$|J_n^{\hat{\pi}}(r) - \hat{J}_n^{\hat{\pi}}(r)| \leq \eta \frac{1-\gamma^n}{1-\gamma}, \quad \text{for all admissible } r. \quad (\text{C.8})$$

It remains to pass to the infinite-horizon limit. For any feasible policy π and any admissible initial state r , the reputation process satisfies $r_{t+1} \leq (1-\delta)r_t + \delta\hat{v}(\bar{x})$, hence $r_t \leq \max\{r, \hat{v}(\bar{x})\}$ for all $t \geq 0$. Since prices and bag values are bounded and accepted demand is positive and bounded above by $\bar{y}$ under the active-cycle assumption, the one-period reward under any feasible policy satisfies

$$|R_t^\pi| \leq (\bar{p} + c\bar{x})D_t^\pi + c|Q_t| \leq (\bar{p} + c\bar{x})\bar{y} + c|Q_t|.$$

The right-hand side has finite expectation, uniformly in t . Therefore discounted tails vanish, and the finite-horizon values converge pointwise: $J_n^*(r) \rightarrow J^*(r)$, $\hat{J}_n(r) \rightarrow \hat{J}(r)$, $J_n^{\hat{\pi}}(r) \rightarrow J^{\hat{\pi}}(r)$, and $\hat{J}_n^{\hat{\pi}}(r) \rightarrow \hat{J}^{\hat{\pi}}(r)$.

Letting $n \rightarrow \infty$ in (C.7) and (C.8) gives $|J^*(r) - \hat{J}(r)| \leq \frac{\eta}{1-\gamma}$ and $|J^{\hat{\pi}}(r) - \hat{J}^{\hat{\pi}}(r)| \leq \frac{\eta}{1-\gamma}$ for all admissible r . Since $\hat{\pi}$ is optimal for the fluid problem, $\hat{J}^{\hat{\pi}} = \hat{J}$. Therefore, for any admissible r ,

$$0 \leq J^*(r) - J^{\hat{\pi}}(r) \leq |J^*(r) - \hat{J}(r)| + |J^{\hat{\pi}}(r) - \hat{J}(r)| \leq \frac{2\eta}{1-\gamma}.$$

Finally, by Cauchy–Schwarz,

$$\mathbb{E}|\epsilon| \leq \sqrt{\mathbb{E}[\epsilon^2]} = \sqrt{\text{Var}(\epsilon)}, \quad \mathbb{E}|Q - \bar{Q}| \leq \sqrt{\mathbb{E}[(Q - \bar{Q})^2]} = \sqrt{\text{Var}(Q)}.$$

Hence $\eta \leq (\bar{p} + c\bar{x})\sqrt{\text{Var}(\epsilon)} + c\sqrt{\text{Var}(Q)}$. Substituting this into the previous inequality completes the proof. $\square$

C.4. Proof of Proposition 7

Proof. We first show that the reservation cap need not bind at the interior deterministic steady state characterized in the proposition. Suppose, to the contrary, that $\hat{y} < \bar{d}(\hat{r}, \hat{p})$ and $0 < \hat{p} < \bar{p}$. Let $g = \bar{d}(\hat{r}, \hat{p}) - \hat{y} > 0$. Increase the price by a small $\eta > 0$, holding $\hat{y}$ and $\hat{\ell}$ fixed, where $\eta < \frac{g}{\alpha+\beta}$. Under this perturbation, the deterministic rating path remains between $\hat{r}$ and $\hat{r} - \eta$. Hence latent demand remains above $\hat{y}$, so accepted demand remains equal to $\hat{y}$. Supplementation cost and delivered bag value are unchanged, while per-period revenue increases by $\eta\hat{y}$, contradicting optimality. Therefore an interior deterministic steady state must satisfy $\hat{y} \geq \bar{d}(\hat{r}, \hat{p})$. We can therefore select any feasible

nonbinding cap and proceed with accepted demand $D = \bar{d}(r, p)$. Recall the deterministic model with latent requests $\bar{d}(r, p) = \alpha r - \beta p$, accepted demand $\bar{D}(r, p, y) = \min\{\bar{d}(r, p), y\}$, and reputation update

$$r_{t+1} = (1 - \delta)r_t + \delta(\hat{v}(\ell_t) - p_t), \quad \delta \in (0, 1). \quad (\text{C.9})$$

For the local perturbation argument, select an equivalent cap with slack, $\hat{y} > \bar{d}(\hat{r}, \hat{p})$; after deriving the first-order conditions, we return to the minimal-cap normalization $\hat{y} = \bar{d}(\hat{r}, \hat{p})$. In a neighborhood of this slack-cap representation, accepted demand equals latent requests. Let the local one-period payoff be

$$\Pi(p, \ell, r) := p \bar{d}(r, p) - c(\ell \bar{d}(r, p) - \bar{Q})^+. \quad (\text{C.10})$$

We first derive the steady-state identities in (28)–(29). We then add a short existence/uniqueness/convergence argument under an explicit interiority condition, and finally summarize the non-interior regimes.

At a steady state, $(\hat{r}, \hat{p}, \hat{y}, \hat{\ell})$ satisfies $\hat{r} = (1 - \delta)\hat{r} + \delta(\hat{v}(\hat{\ell}) - \hat{p})$, which rearranges to the first equality in (28): $\hat{r} = \hat{v}(\hat{\ell}) - \hat{p}$. Let $\hat{J}$ denote the DFA value function in (26). Since $T(r, p, \ell) := (1 - \delta)r + \delta(\hat{v}(\ell) - p)$ satisfies $T(\hat{r}, \hat{p}, \hat{\ell}) = \hat{r}$, the Bellman equation at $\hat{r}$ implies

$$\hat{J}(\hat{r}) = \Pi(\hat{p}, \hat{\ell}, \hat{r}) + \gamma \hat{J}(\hat{r}) \quad \Rightarrow \quad \hat{J}(\hat{r}) = \Pi(\hat{p}, \hat{\ell}, \hat{r}) / (1 - \gamma). \quad (\text{C.11})$$

The proof uses two-period perturbations that return reputation to its steady-state level at $t = 2$, so they have no effect on later payoffs.

Next, fix $\ell_t \equiv \hat{\ell}$ and perturb price in periods $t = 0, 1$ so that $r_2 = \hat{r}$. For ε small, consider the feasible policy

$$(p_0, \ell_0) = (\hat{p} + \varepsilon, \hat{\ell}), \quad (p_1, \ell_1) = (\hat{p} - (1 - \delta)\varepsilon, \hat{\ell}), \quad (p_t, \ell_t) = (\hat{p}, \hat{\ell}), \quad t \geq 2.$$

Using (28) and (C.9), the induced reputation path is

$$r_0 = \hat{r}, \quad r_1 = \hat{r} - \delta\varepsilon, \quad r_2 = \hat{r}, \quad r_t = \hat{r} \quad (t > 2).$$

Let $V_p(\varepsilon)$ denote the value from $\hat{r}$ under this perturbed policy. Because $r_t = \hat{r}$ for all $t \geq 2$,

$$V_p(\varepsilon) = \Pi(\hat{p} + \varepsilon, \hat{\ell}, \hat{r}) + \gamma \Pi(\hat{p} - (1 - \delta)\varepsilon, \hat{\ell}, \hat{r} - \delta\varepsilon) + \gamma^2 \hat{J}(\hat{r}).$$

Optimality of $(\hat{p}, \hat{\ell})$ at $\hat{r}$ implies $\hat{J}(\hat{r}) \geq V_p(\varepsilon)$. Using (C.11) to eliminate $\hat{J}(\hat{r})$ gives, for all sufficiently small ε ,

$$0 \geq \left[\Pi(\hat{p} + \varepsilon, \hat{\ell}, \hat{r}) - \Pi(\hat{p}, \hat{\ell}, \hat{r}) \right] + \gamma \left[\Pi(\hat{p} - (1 - \delta)\varepsilon, \hat{\ell}, \hat{r} - \delta\varepsilon) - \Pi(\hat{p}, \hat{\ell}, \hat{r}) \right]. \quad (\text{C.12})$$

Define $F_p(\varepsilon)$ as the right-hand side of (C.12). Then $F_p(0) = 0$ and, by interiority of $\hat{p}$, we have $F_p(\varepsilon) \leq 0$ for ε of both signs in a neighborhood of 0. Because the active-supplementation condition gives $\hat{\ell} \bar{d}(\hat{r}, \hat{p}) > \bar{Q}$, the positive-part term is locally linear around $(\hat{p}, \hat{\ell}, \hat{r})$ and Π is differentiable there. Hence $F'_p(0) = 0$, which yields

$$0 = \partial_p \Pi(\hat{p}, \hat{\ell}, \hat{r}) + \gamma \left(-(1 - \delta) \partial_p \Pi(\hat{p}, \hat{\ell}, \hat{r}) - \delta \partial_r \Pi(\hat{p}, \hat{\ell}, \hat{r}) \right).$$

Therefore

$$(1 - \gamma(1 - \delta)) \partial_p \Pi(\hat{p}, \hat{\ell}, \hat{r}) = \gamma \delta \partial_r \Pi(\hat{p}, \hat{\ell}, \hat{r}). \quad (\text{C.13})$$

Next, let

$$\Delta v(\varepsilon) := \hat{v}(\hat{\ell} + \varepsilon) - \hat{v}(\hat{\ell})$$

and consider the two-period perturbation

$$\begin{aligned}(p_0, \ell_0) &= (\hat{p}, \hat{\ell} + \varepsilon), \\ (p_1, \ell_1) &= (\hat{p} + (1 - \delta)\Delta v(\varepsilon), \hat{\ell}), \\ (p_t, \ell_t) &= (\hat{p}, \hat{\ell}), \quad t \geq 2.\end{aligned}$$

Using (28) and (C.9), the induced reputation path is

$$r_0 = \hat{r}, \quad r_1 = \hat{r} + \delta\Delta v(\varepsilon), \quad r_t = \hat{r} \quad (t \geq 2).$$

Optimality, interiority, and differentiability of $\hat{v}$ at $\hat{\ell}$ therefore give

$$0 = \partial_{\ell}\Pi(\hat{p}, \hat{\ell}, \hat{r}) + \gamma\hat{v}'(\hat{\ell}) \left[(1 - \delta)\partial_p\Pi(\hat{p}, \hat{\ell}, \hat{r}) + \delta\partial_r\Pi(\hat{p}, \hat{\ell}, \hat{r}) \right].$$

Equation (C.13) implies that the bracketed term multiplied by γ equals $\partial_p\Pi(\hat{p}, \hat{\ell}, \hat{r})$. Hence

$$\partial_{\ell}\Pi(\hat{p}, \hat{\ell}, \hat{r}) + \hat{v}'(\hat{\ell})\partial_p\Pi(\hat{p}, \hat{\ell}, \hat{r}) = 0. \quad (\text{C.14})$$

Because $\hat{\ell}\bar{d}(\hat{r}, \hat{p}) > \bar{Q}$, locally

$$\Pi(p, \ell, r) = (p - c\ell)(\alpha r - \beta p) + c\bar{Q}.$$

Hence $\partial_p\Pi = \alpha r - 2\beta p + \beta c\ell$, $\partial_r\Pi = \alpha(p - c\ell)$, and $\partial_{\ell}\Pi = -c(\alpha r - \beta p)$.

Substituting these into (C.13) and (C.14) gives

$$(1 - \gamma(1 - \delta))(\alpha\hat{r} - 2\beta\hat{p} + \beta c\hat{\ell}) = \gamma\delta\alpha(\hat{p} - c\hat{\ell}), \quad (\text{C.15})$$

$$\hat{v}'(\hat{\ell})(\alpha\hat{r} - 2\beta\hat{p} + \beta c\hat{\ell}) = c(\alpha\hat{r} - \beta\hat{p}). \quad (\text{C.16})$$

Using $\kappa = \gamma\delta/[1 - \gamma(1 - \delta)]$, (C.15) can be written as

$$\alpha\hat{r} - 2\beta\hat{p} + \beta c\hat{\ell} = \alpha\kappa(\hat{p} - c\hat{\ell}) \iff \bar{d}(\hat{r}, \hat{p}) = (\beta + \alpha\kappa)(\hat{p} - c\hat{\ell}).$$

This is the third equality in (28). Because demand is strictly positive in the interior active regime, it also implies $\hat{p} > c\hat{\ell}$.

Substituting this identity into (C.16) gives $\hat{v}'(\hat{\ell})\alpha\kappa(\hat{p} - c\hat{\ell}) = c(\beta + \alpha\kappa)(\hat{p} - c\hat{\ell})$, hence

$$\hat{v}'(\hat{\ell}) = c \left(1 + \frac{\beta}{\alpha\kappa} \right),$$

which is the second equality in (28). Finally, combining $\hat{r} = \hat{v}(\hat{\ell}) - \hat{p}$ with $\bar{d}(\hat{r}, \hat{p}) = (\beta + \alpha\kappa)(\hat{p} - c\hat{\ell})$ yields (29). The perturbation argument at the start of the proof shows that any interior deterministic steady state can be represented with a nonbinding cap. Hence any feasible cap $\hat{y} \geq \bar{d}(\hat{r}, \hat{p})$ is payoff-equivalent in this deterministic steady state. When $\bar{d}(\hat{r}, \hat{p}) \geq \underline{y}$, selecting $\hat{y} = \bar{d}(\hat{r}, \hat{p})$ gives the minimal-cap normalization.

To characterize solutions of the necessary conditions, assume additionally that $\hat{v}$ is differentiable on $\mathcal{X}$ and that $\hat{v}'$ is continuous and strictly decreasing. Define $\chi := c \left(1 + \frac{\beta}{\alpha\kappa} \right)$. If

$$\hat{v}'(\bar{x}) < \chi < \hat{v}'(\underline{x}), \quad (\text{C.17})$$

where the endpoint derivatives are understood one-sidedly, then by continuity and strict monotonicity of $\hat{v}'$ there exists a unique $\hat{\ell} \in (\underline{x}, \bar{x})$ such that $\hat{v}'(\hat{\ell}) = \chi$. Define $\hat{p}$ by (29) and $\hat{r} = \hat{v}(\hat{\ell}) - \hat{p}$. If, in addition, $0 < \hat{p} < \bar{p}$, $0 <$

$\bar{d}(\hat{r}, \hat{p}) < \bar{y}$, and $\hat{\ell} \bar{d}(\hat{r}, \hat{p}) > \bar{Q}$, then $(\hat{r}, \hat{p}, \hat{y}, \hat{\ell})$, with any feasible nonbinding cap $\hat{y} \geq \bar{d}(\hat{r}, \hat{p})$, is the unique interior candidate satisfying (28)–(29) up to the payoff-equivalent cap normalization. These conditions establish uniqueness among interior solutions to the steady-state necessary conditions; they do not by themselves provide a global Bellman verification.

Moreover, once this pair $(\hat{p}, \hat{\ell})$ is fixed, the induced deterministic rating dynamics are

$$r_{t+1} = (1 - \delta)r_t + \delta(\hat{v}(\hat{\ell}) - \hat{p}) = (1 - \delta)r_t + \delta\hat{r}.$$

Hence, for any initial rating r_0 , $r_t = \hat{r} + (1 - \delta)^t(r_0 - \hat{r}) \rightarrow \hat{r}$. Therefore the deterministic trajectory generated by repeatedly applying the steady-state pair $(\hat{p}, \hat{\ell})$ and any equivalent nonbinding cap converges globally to $(\hat{r}, \hat{p}, \hat{y}, \hat{\ell})$.

Finally, when (C.17) fails, the necessary conditions instead identify a boundary candidate in the bag-value dimension: if $\chi \geq \hat{v}'(\underline{x})$, then the quality choice is pushed to the lower boundary $\hat{\ell} = \underline{x}$; if $\chi \leq \hat{v}'(\bar{x})$, then it is pushed to the upper boundary $\hat{\ell} = \bar{x}$. Likewise, if the price reconstructed from (29) violates $0 < p < \bar{p}$, the associated candidate lies at $p = 0$ or $p = \bar{p}$. $\square$

C.5. Proof of Corollary 3

Proof. The additional conditions in Corollary 3 permit local implicit differentiation at the unique interior solution.

(i) **Uniqueness and comparative statics of $\hat{\ell}$.** From Proposition 7, $\hat{v}'(\hat{\ell}) = \chi(\alpha, \beta, \kappa, c)$, where $\chi(\alpha, \beta, \kappa, c) := c(1 + \frac{\beta}{\alpha\kappa})$. The additional condition in Corollary 3 gives a unique interior solution, and $\hat{v}''(\hat{\ell}) < 0$ makes it locally regular.

For comparative statics, implicitly differentiate $\hat{v}'(\hat{\ell}) = \chi(\alpha, \beta, \kappa, c)$ with respect to $\theta \in \{\kappa, \alpha, \beta, c\}$:

$$\hat{v}''(\hat{\ell}) \frac{\partial \hat{\ell}}{\partial \theta} = \frac{\partial \chi}{\partial \theta}.$$

Since $\hat{v}''(\hat{\ell}) < 0$, the sign of $\partial \hat{\ell} / \partial \theta$ is the opposite of the sign of $\partial \chi / \partial \theta$. Direct calculation yields

$$\frac{\partial \chi}{\partial \kappa} = -\frac{c\beta}{\alpha\kappa^2} < 0, \quad \frac{\partial \chi}{\partial \alpha} = -\frac{c\beta}{\alpha^2\kappa} < 0, \quad \frac{\partial \chi}{\partial \beta} = \frac{c}{\alpha\kappa} > 0, \quad \frac{\partial \chi}{\partial c} = 1 + \frac{\beta}{\alpha\kappa} > 0,$$

which implies that $\hat{\ell}$ is strictly increasing in κ and α , and strictly decreasing in β and c .

(ii) **Monotonicity of $\hat{r}$ in κ .** Using (29) and $\hat{r} = \hat{v}(\hat{\ell}) - \hat{p}$, define

$$\hat{r}(\ell, \kappa) := \frac{(\alpha\kappa + 2\beta)\hat{v}(\ell) - c(\beta + \alpha\kappa)\ell}{\alpha(1 + \kappa) + 2\beta}.$$

Along the steady state, where $\hat{\ell}$ varies with κ , $\frac{d\hat{r}}{d\kappa} = \frac{\partial \hat{r}}{\partial \kappa} + \frac{\partial \hat{r}}{\partial \ell} \cdot \frac{d\hat{\ell}}{d\kappa}$. From part (i), $d\hat{\ell}/d\kappa > 0$. A direct derivative computation gives

$$\frac{\partial \hat{r}}{\partial \kappa} = \frac{\alpha(\alpha\hat{v}(\ell) - c(\alpha + \beta)\ell)}{(\alpha(1 + \kappa) + 2\beta)^2}.$$

Evaluating at $\ell = \hat{\ell}$ and using (29), we have $\alpha\hat{v}(\hat{\ell}) - c(\alpha + \beta)\hat{\ell} = (\alpha(1 + \kappa) + 2\beta)(\hat{p} - c\hat{\ell})$. Therefore, $\bar{d}(\hat{r}, \hat{p}) = \hat{D} > 0$ implies $\hat{p} > c\hat{\ell}$, hence $\partial \hat{r} / \partial \kappa > 0$ at $\ell = \hat{\ell}$. Moreover,

$$\frac{\partial \hat{r}}{\partial \ell} = \frac{(\alpha\kappa + 2\beta)\hat{v}'(\ell) - c(\beta + \alpha\kappa)}{\alpha(1 + \kappa) + 2\beta},$$

and substituting $\hat{v}'(\hat{\ell}) = c(1 + \beta/(\alpha\kappa))$ from (28) yields

$$\frac{\partial \hat{r}}{\partial \ell} \Big|_{\ell=\hat{\ell}} = \frac{2c\beta(1 + \beta/(\alpha\kappa))}{\alpha(1 + \kappa) + 2\beta} > 0.$$

Therefore both terms in $d\hat{r}/d\kappa$ are positive, and $\hat{r}$ is strictly increasing in κ .

(iii) **Implications for γ and δ .** Since $\kappa = \gamma\delta/[1 - \gamma(1 - \delta)]$, we have

$$\frac{\partial \kappa}{\partial \gamma} = \frac{\delta}{(1 - \gamma(1 - \delta))^2} > 0 \quad \text{and} \quad \frac{\partial \kappa}{\partial \delta} = \frac{\gamma(1 - \gamma)}{(1 - \gamma(1 - \delta))^2} > 0.$$

Thus the monotonicity of $\hat{\ell}$ and $\hat{r}$ in γ and δ follows from parts (i)–(ii). $\square$

Appendix D: Numerical Details

Numerical grids and stochastic environments. The main parameter grid is $\alpha, \beta \in \{1.8, 1.9, 2.0, 2.1, 2.2\}$ and $c \in \{0.75, 0.80, 0.85, 0.90, 0.95\}$, with $\eta = 0.60$, $\gamma = 0.90$, and $\delta = 0.10$. The action ranges are $\mathcal{X} = [1, 5]$, $p \in [0, 5]$, and $y \in [0.25, 5]$, with $h_0 = 3.5$. The dynamic program is solved on 61 rating grid points on $[0, 5]$, 61 price grid points, 41 reservation-cap grid points, and 81 bag-value grid points. Thus the baseline DP search is over (r, p, y, ℓ) , and the extension with endogenous promised value adds h . Modified policy iteration uses tolerance 10^{-5} , at most 260 iterations, 7-point equiprobable quantile discretizations for (ϵ, Q) , and linear interpolation on the rating grid. The DP solver and simulator project candidate next-period ratings to $[0, 5]$; this safeguard did not bind over the saved fixed- and endogenous-promise policy grids. Each policy is evaluated from ten evenly spaced initial ratings on $[3, 5]$, with $T = 2,000$, a 200-period burn-in for long-run operating measures, and eight common-random-number replications. After averaging replications and initial states within each environment, scenario-bootstrap intervals resample environment identifiers 2,000 times.

Reported profit is the normalized discounted average

$$\bar{\Pi}^\gamma = \frac{\sum_{t=0}^{T-1} \gamma^t \Pi_t}{\sum_{t=0}^{T-1} \gamma^t}, \quad \Pi_t = p_t D_t - c(D_t \ell_t - Q_t)^+.$$

The DP solver, numerical LVA search, and simulator all use $D_t = \min\{(\alpha r_t - \beta p_t + \epsilon_t)^+, y_t\}$. When $D_t = 0$, no Stage-II allocation is made and the rating remains unchanged; the period contributes zero profit and household waste, and Q_t units of in-store waste. These numerical rules allow no-sale states outside the positive-demand region maintained in the analytical model.

For each deterministic grid point in the main policy comparison, we draw five stochastic environments. The demand shock follows a mean-zero truncated normal distribution with support $[-3\sigma_\epsilon, 3\sigma_\epsilon]$, where $\sigma_\epsilon \sim U[0.35, 0.65]$. End-of-cycle surplus follows a scaled beta distribution on $[Q_{\min}, Q_{\max}]$, where $Q_{\min} \sim U[1.5, 2.5]$, $Q_{\max} \sim U[4.5, 5.5]$, and the beta shape parameters are drawn from $U[1.5, 4.0]$. The same quantile discretization is used when solving the DP and when evaluating approximate policies.

Policy and extension details. The oracle-calibrated LVA estimates the affine slope parameter k by least-squares fitting to the DP value function on the rating grid. This design isolates approximation-form error and does not test how k would be learned out of sample. Outside the extension with endogenous promised value, when delivered bag value exceeds household regular needs $\bar{\ell}_{\text{need}} = 1.1 \ell_{\text{ref}}$, the excess amount is wasted with probability weight $\xi = 0.3$. That extension instead keeps this need benchmark fixed at $1.1h_0$, as specified below.

For the selective-rating extension, we use five stochastic environments at $\alpha = \beta = 2$ and $c = 0.95$. For each target mean on the 81-point bag-value grid, we enumerate every feasible one- and two-point support and select the distribution that maximizes the rating-weighted satisfaction ratio in (31); we call this the rating-weighted optimal allocation. When $\zeta = 0$, this procedure recovers the ordinary envelope allocation.

The extension with endogenous promised value keeps $(\lambda_+, \lambda_-) = (1.5, 3.0)$ and compares the fixed benchmark $h_t = h_0 = 3.5$ with endogenous choices from $\mathcal{H} = \{3.0, 3.5, \dots, 5.0\}$. Both regimes obey $2p_t \leq h_t \leq 4p_t$ and use the centered demand adjustment $\rho(h_t - h_0)$. Within each matched environment, the quality-adjusted-value support $\mathcal{X} = [1, 5]$, the surplus distribution, and c are held fixed across h ; the customer-valuation function and its envelope are recomputed for each promise. Baseline and Envelope retain their fixed-promise definitions; DP searches jointly over (p, y, h) at every rating-grid point, while LVA searches over the same actions under its separately fitted affine continuation-value approximation. We use $\rho \in \{0.50, 0.55, 0.60\}$ in 100 independently generated stochastic environments at $\alpha = \beta = 2$ and $c = 0.95$, with matched random-number streams across policies. For each endogenous policy, scenario-level metrics are averaged equally across the three ρ values before resampling the 100 scenario identifiers; the fixed result is retained once. Household regular needs remain fixed at $\bar{\ell}_{\text{need}} = 1.1h_0$ rather than moving mechanically with the chosen promise.

Because the price–promise band does not ensure the sufficient profitability condition in Assumption 1, we checked the computed DP value function directly. We classified it as nondecreasing when every adjacent difference on the 61-point rating grid was at least -10^{-10} ; all 300 solves (100 environments at each of the three ρ values) passed, with a maximum downward change of zero. The largest final sup-norm change between value iterations was 9.96×10^{-6} , below the solver tolerance 10^{-5} . Thus the envelope reduction used by DP remains valid in the endogenous-promise extension.

Responsiveness-index comparative statics. To extend the DFA comparative statics in Corollary 3 to the stochastic model, we vary $\kappa = \gamma\delta/(1 - \gamma(1 - \delta))$ through two channels:

- (a) *Discount-factor channel*: $\gamma \in [0.50, 0.90]$ with $\delta = 0.10$.
- (b) *Rating-update channel*: $\delta \in [0.03, 0.10]$ with $\gamma = 0.90$.

All other parameters stay at baseline ($\alpha = 2$, $\beta = 2$, $c = 0.95$). Each channel uses an 8-point equispaced grid, with the baseline value at the upper bound, yielding κ up to 0.47. All tested points satisfy $(1 - \delta)\beta \geq \delta\alpha$ and converge. At each grid point, we solve the DP with the numerical settings above and simulate using the protocol in Section 6. Post-burn-in averages of ℓ_t , r_t , and p_t serve as stochastic analogues of ℓ^* , r^* , and p^* .

Figure D.1 Comparative statics for the responsiveness index κ , varied through γ and δ .

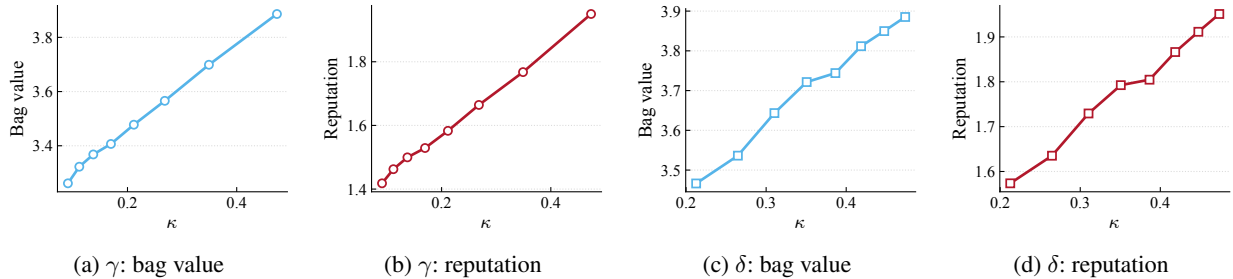


Figure D.1 shows that increasing κ through either γ or δ raises long-run bag value and reputation, matching the deterministic force in Corollary 3 and supporting κ as a useful summary statistic for reputation incentives in the stochastic model.