Demand Estimation with Unobserved Stockout Substitution: Exact Likelihood Maximization

Chenhao Wang
School of Economics and Management, Tongji University, Shanghai, China, chenhaowang@tongji.edu.cn

Hansheng Jiang
Rotman School of Management, University of Toronto, Toronto, Ontario, Canada, hansheng.jiang@rotman.utoronto.ca

Ningyuan Chen
Department of Management, University of Toronto Mississauga; Rotman School of Management, University of Toronto, Ontario, Canada, ningyuan.chen@utoronto.ca

When a product is sold out, customers who prefer it may buy a substitute or leave without purchasing. Efficient demand estimation in this setting is challenging when inventory is observed only at replenishments, because the stockout times, and hence the products available to each customer, are unobserved. Estimating demand efficiently under stockout substitution has been a challenging problem since Anupindi et al. (1998), and it has become more pressing as the scale of such problem instances grows thanks to innovations in retailing such as unmanned stores. We provide a complete solution to this problem for the multinomial logit (MNL) model, the most widely used model of customer choice. Using a continuous-time representation of the purchase process developed in Zhou et al. (2026), we show that the exact likelihood and its gradient are one-dimensional integrals, no matter how many products, customers, or stockouts there are. Both can be evaluated at a cost linear in the numbers of products and periods. Our method estimates demand for 2,000 products from 2,000 selling periods in under ten minutes.

Key words: demand estimation; stockout substitution; multinomial logit; aggregate sales; maximum likelihood estimation; Poisson arrivals

1. Introduction

Consider a retailer who periodically replenishes an array of products in a vending machine. The inventory of each product is observed only at replenishments, immediately before and after each one. Between two replenishments, which we call a *selling period*, customers arrive sequentially and choose from the products that are still available; a customer whose preferred product has sold out may choose a substitute or leave without purchasing. The retailer observes none of these choices and, in particular, does not know when each product sold out. The question is whether the retailer can estimate the demand for each product from the initial inventories and the number of units of each product sold in each period, which we call the terminal sales. The same data structure (illustrated in Figure 1 and formally introduced in Section 2) arises in traditional and unmanned retail stores, where inventory is typically counted only at replenishments.

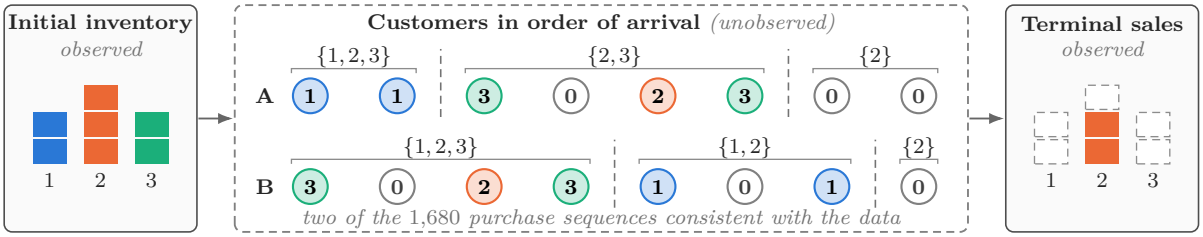

Figure 1 One example selling period

Notes. The initial inventory is $(2, 3, 2)$, and eight customers arrive. Sequences A and B both produce the terminal sales $(2, 1, 2)$, in which products 1 and 3 are sold out. Each circle shows a customer's choice, with 0 denoting no purchase. The brackets indicate the products available to each customer.

Stockouts cannot be ignored: after a product sells out, its sales stop and part of its demand shifts to the remaining products. Accounting for this shift is difficult because stockouts occur at unobserved times. The terminal sales reveal which products sold out, but not when, and therefore not which products were available to each customer. The same terminal sales can result from many purchase sequences, i.e., orders in which customers bought products or left, and their number grows exponentially with the number of customers. In the small example of Figure 1, with three products and eight customers, 1,680 purchase sequences are consistent with the observed initial inventory and terminal inventory/sales. These sequences do not share the same likelihoods, because a product that sells out early changes the choices of all later customers. For instance, product 1 sells out after the second customer in sequence A but only after the seventh in sequence B, so customers 3 to 7 can choose product 1 in sequence B but not in sequence A. The likelihood of the observed data is therefore a sum over all the purchase sequences consistent with the initial inventory and terminal sales.

This estimation problem dates back at least to Anupindi et al. (1998) and remains an active area of research (see Section 1.1). Existing likelihood-based methods enumerate, sample, or approximate the purchase sequences, and each approach breaks down as the numbers of products and stockouts grow. Expectation–maximization (EM) methods average over the possible times and orders of stockouts within each period, which is tractable only when few products have stockouts: Anupindi et al. (1998) carry out the EM calculations for up to two stockouts per period, and Conlon and Mortimer (2013) report that direct calculation is feasible for up to three. Bayesian data augmentation (Musalem et al. 2010, Wan et al. 2018) samples the purchase sequences; in our experiments, it takes several minutes even with ten products. Shao and Kleywegt (2020) group the purchase sequences by the order in which products sell out and sample these orders. Their sampling and approximation method reduces the computational cost, but introduces simulation error or approximation bias that can affect parameter estimates, and controlling these errors requires additional computation. This tradeoff between computational cost and estimation accuracy is compounded because the likelihood must be evaluated for every selling period at every iteration of the optimization. To our knowledge, the exact

likelihood has not been computed efficiently at a realistic scale, even under the multinomial logit (MNL) model, the most widely used model of customer choice. The gap has widened as data have grown: assortments now often contain hundreds or thousands of products.

In this paper, we close this gap for the MNL model. We build on a recent result of Zhou et al. (2026), who represent the purchase sequence under MNL by a continuous-time system of independent Poisson processes and use it to study inventory allocation under dynamic substitution. We use this representation for a different purpose: to compute the probability or likelihood of an entire vector of terminal sales. Our finding is surprising: the exact likelihood of each selling period is a *one-dimensional* integral, no matter how many products, customers, or stockouts there are. The gradient of the log-likelihood is also a one-dimensional integral, and both the likelihood and its gradient can be computed efficiently through the integral expression. These results amount to a complete solution to the problem: the purchase sequences are integrated out *analytically* rather than enumerated, sampled, or approximated, and the cost of evaluating the likelihood and its gradient is linear in both the number of products and the number of selling periods. In the numerical experiments, our method estimates demand for 2,000 products from 2,000 selling periods within ten minutes on a single GPU, whereas no other benchmarks complete much smaller instances within the time limit.

1.1. Literature Review

A stream of literature studies demand estimation with stockouts, typically in setups that differ from ours. We compare these studies with ours below. Anupindi et al. (1998) and Conlon and Mortimer (2013) both estimate demand from periodic inventory records, as we do. Anupindi et al. (1998) study Poisson demand with a particular substitution pattern, and Conlon and Mortimer (2013) consider more flexible choice models, including nested and random-coefficients logit. Both use EM algorithms to handle the unobserved stockout times. However, the expectation step becomes complicated as the number of stockouts in a period grows as illustrated in Figure 1, and both studies carry out the exact calculations only when few products sell out in a period.

Musalem et al. (2010) and Wan et al. (2018) tackle the problem by Bayesian data augmentation, sampling the unobserved purchase sequences that are consistent with the data. Their setups differ slightly: Musalem et al. (2010) allow heterogeneous customer preferences, and Wan et al. (2018) consider nested logit models with substitution across both products and stores. Both treat the total number of customers in each period, including those who leave without purchasing, as known, obtained from total transactions or estimated externally (see Section 2 for more details). However, sampling the purchase sequences is computationally intensive, and in our experiments it is practical only for the smallest instances (Section 6).

Shao and Kleywegt (2020) is closest to ours, because they also attempt to formulate the likelihood exactly and maximize it directly. They jointly estimate a Poisson arrival rate and the parameters of a

general discrete-choice model from initial inventories and terminal sales, without observing individual choices, no-purchases, or stockout times. They group choice sequences by their stockout orders, i.e., the order in which products sell out. This allows them to simplify the calculation of the likelihood significantly. However, the likelihood still requires enumerating or sampling stockout orders, which remains prohibitive for large instances in our experiments. Our computation of the likelihood avoids enumerating or sampling high-dimensional latent spaces altogether, and this is our most important contribution to the literature. Shao and Kleywegt (2020) also leave identification open, suggesting that because stockouts create more variety in choice sets, the number of potential customers may be identifiable even when the initial assortment is fixed; we prove this under MNL (Section 5). Table 1 summarizes the studies above.

	Choice model	Number of customers	Latent stockout history
Anupindi et al. (1998)	Poisson demand with one-stage substitution	Estimated	EM; up to two stockouts per period
Musalem et al. (2010)	Random-coefficients logit	Given	Sampled by MCMC
Conlon and Mortimer (2013)	Nested and random-coefficients logit	Given	EM; up to three stockouts per period
Wan et al. (2018)	Nested logit; substitution across stores	Given	Sampled by MCMC
Shao and Kleywegt (2020)	General discrete choice	Estimated	Enumerated or sampled
This paper	MNL	Given or estimated	Integrated out analytically in the exact likelihood function

Table 1 Closely related studies

When individual transaction times, and hence stockout times, are observed or can be inferred, the problem becomes much more tractable, even though no-purchases and potential customer arrivals remain unobserved. Letham et al. (2016) exploit this information in a Bayesian model with nonhomogeneous Poisson arrivals and heterogeneous preferences. They construct a transaction-time likelihood from purchase rates and note that the MNL no-purchase parameter is not identifiable when the arrival function is unknown. Chen et al. (2026) develop a related model with covariate-dependent arrivals and stockout-based substitution and derive an analytic transaction-time likelihood. Karabatı et al. (2009) group point-of-sale intervals by availability scenario and infer substitution from differences in purchase rates. Zheng et al. (2026) estimate base demand from observations unlikely to contain substitution and recover substitution probabilities by comparing demand across stockout conditions.

Even when all choices and stockout times are observed, Liu et al. (2026) show that maximum likelihood estimates of MNL choice probabilities are biased by stockout-induced endogeneity, and they remove the bias by dropping the purchases that trigger stockouts.

A broader literature jointly estimates customer arrivals and preferences when no-purchases are unobserved but the offered assortment is recorded for each period or transaction. Under MNL, the arrival rate and the no-purchase probability are separated either through a common arrival rate (Talluri and van Ryzin 2004, Newman et al. 2014) or by combining period-specific arrival rates with external market-share information (Vulcano et al. 2012, Abdallah and Vulcano 2021). Related methods accommodate rank-based, Markov chain, nested logit, and mixture-of-logit choice models (van Ryzin and Vulcano 2017, Şimşek and Topaloglu 2018, Cho et al. 2023, Wang and Xu 2025), incorporate competitor information (Talluri and Tekin 2025), guard against errors in market-share information (Cho et al. 2024), or develop integer-programming and moment-based estimators (Subramanian and Harsha 2021, Li et al. 2025). In all of these studies, the assortment faced by each customer is known. In ours, it changes within the period at unobserved times, and we show that these unobserved changes are nevertheless enough to separate a common arrival rate from the MNL preferences.

A separate line of literature studies demand estimation in inventory control with censoring. Classical work treats a single product (Nahmias 1994, Agrawal and Smith 1996), and later work exploits the timing of stockouts (Jain et al. 2015) or corrects for inaccurate inventory records (Mersereau 2015); Steeneck et al. (2022) estimate lost sales for substitutable products when on-shelf availability is uncertain, using external market-share and availability information. Our setting differs in that inventory records are accurate and no external information is used, but the times at which products sell out, and hence the choice sets of later customers, are unobserved.

Finally, assortment and inventory decisions under stockout substitution have attracted considerable attention recently. Several studies estimate substitution from retail sales and then optimize assortments or inventories, typically treating availability as observed or constant within a period (Kök and Fisher 2007, Fisher and Vaidyanathan 2014, Akchen and Caro 2026); Lee et al. (2016) do so for new and used textbooks from season sales with unobserved stockout times. With known demand parameters, a large literature optimizes assortments and inventories under stockout-based substitution (Smith and Agrawal 2000, Mahajan and van Ryzin 2001, Honhon et al. 2010), often through tractable approximations (Hopp and Xu 2008, Liang et al. 2021, Aouad and Segev 2023, Zhang et al. 2025), and related work learns demand while making these decisions (Chen and Chao 2020, Liang et al. 2026). The continuous-time representations of MNL substitution developed for such problems (Zhou et al. 2026, Zhou and Zhang 2026) are the starting point of our analysis.

2. Data and Model

A retailer sells J products, indexed by $j \in [J] = \{1, \ldots, J\}$, across N selling periods, indexed by $n = 1, \ldots, N$. In period n, the retailer starts with the observed inventory vector $\boldsymbol{y}_n = (y_{1,n}, \ldots, y_{J,n}) \in \mathbb{Z}_+^J$, where $\mathbb{Z}_+ = \{0, 1, \ldots\}$. Customers then arrive sequentially and choose among the products that remain in stock or leave without a purchase (choosing the *outside option*). To describe the dynamics of the inventory process, let $\boldsymbol{I}_{t,n} = (I_{1,t,n}, \ldots, I_{J,t,n})$ denote the inventory immediately after customer t has chosen, with $\boldsymbol{I}_{0,n} = \boldsymbol{y}_n$. Customer t sees the available set $S_{t,n} = \{j \in [J] : I_{j,t-1,n} > 0\}$. Product j has attraction $\theta_j = e^{v_j} > 0$, and the outside option, product 0, is normalized to have attraction $\theta_0 = 1$, or equivalently $v_0 = 0$. Observing $S_{t,n}$, the choice of customer t in period n, $C_{t,n} \in \{0, 1, \ldots, J\}$, satisfies the classic MNL model:

$$\mathbb{P}(C_{t,n} = j \mid S_{t,n}) = \begin{cases} \dfrac{\theta_j}{1 + \sum_{k \in S_{t,n}} \theta_k}, & j \in S_{t,n}, \\ \dfrac{1}{1 + \sum_{k \in S_{t,n}} \theta_k}, & j = 0, \\ 0, & j \in [J] \setminus S_{t,n}. \end{cases} \tag{1}$$

Inventory evolves according to

$$I_{j,t,n} = I_{j,t-1,n} - \mathbb{I}_{\{C_{t,n} = j\}}, \qquad j \in [J], \tag{2}$$

where $\mathbb{I}_{\{\cdot\}}$ denotes the indicator function. Thus each arrival either leaves inventory unchanged or consumes one unit of an available product. If customer t buys the last unit of product j, then j is removed from $S_{t+1,n}$ and all subsequent available sets. At the end of the period, the retailer observes terminal sales $\boldsymbol{x}_n = (x_{1,n}, \ldots, x_{J,n})$, a realization of the random terminal-sales vector $\boldsymbol{X}_n$.

We consider two data settings. In the *known-arrival setting*, the observed data are $\mathcal{D}_N^{\mathrm{K}} = \{(\boldsymbol{y}_n, \boldsymbol{x}_n, T_n)\}_{n=1}^N$, and the unknown parameter is $\boldsymbol{\theta}$, or equivalently $\boldsymbol{v} = \log \boldsymbol{\theta}$. In the *Poisson-arrival setting*, the observed data are $\mathcal{D}_N^{\mathrm{P}} = \{(\boldsymbol{y}_n, \boldsymbol{x}_n)\}_{n=1}^N$, while $\widetilde{T}_n \sim \mathrm{Poi}(\lambda)$ independently across periods (we use $\widetilde{T}_n$ to distinguish from T_n because $\widetilde{T}_n$ are unobserved latent variables). Here $\lambda > 0$ is the expected number of arrivals per selling period. Representing each period by the interval $[0, 1]$, we refer to λ as the arrival rate. In addition to $\boldsymbol{v}$, this rate is unknown and to be estimated. Figure 2 summarizes the inventory process and the information structure.

In the rest of the paper, when we discuss a particular period and there is no confusion, we suppress the index n and use $(\boldsymbol{y}, \boldsymbol{x})$ to denote the observed inventories and sales, with $0 \le x_j \le y_j$ for every product. Let $\mathcal{A} = \{j : y_j > 0\}$ be the initially available products, $\mathcal{B} = \{j \in \mathcal{A} : x_j = y_j\}$ the products that sell out by the end of the period, and $\mathcal{C} = \mathcal{A} \setminus \mathcal{B}$ the products that remain in stock. Adding the subscript n denotes these sets for period n.

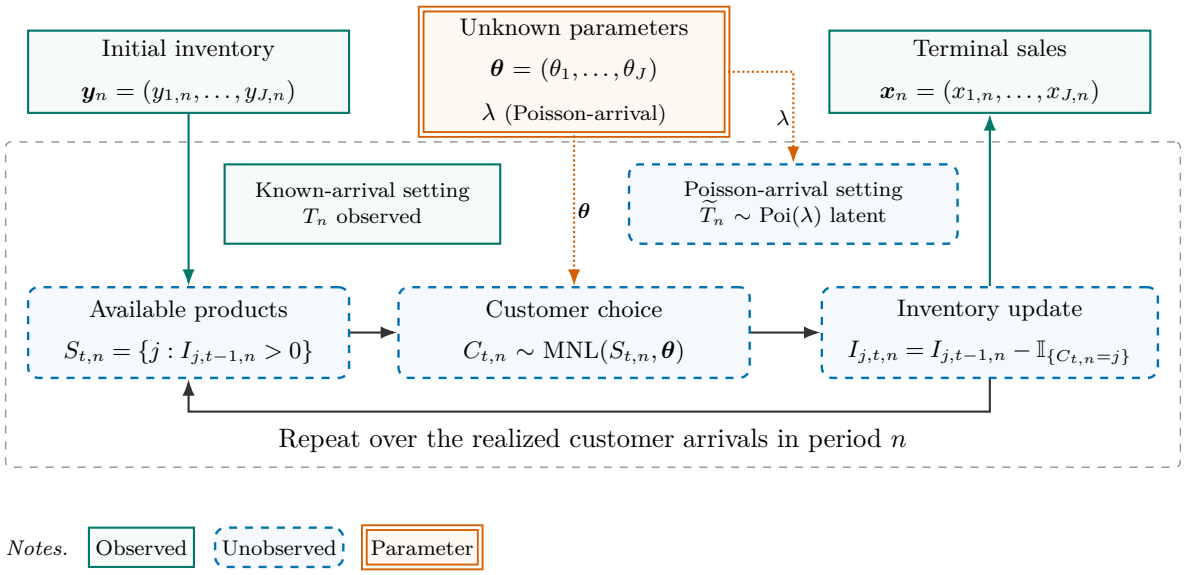

Figure 2 Data-generating process for one selling period

3. Continuous-Time Representation and Exact Likelihood

A natural choice to estimate the parameters in Section 2 is a likelihood-based method. Consider the known-arrival setting for simplicity. For a trajectory of purchases $\boldsymbol{c} = (c_1, \ldots, c_T)$, including no-purchase outcomes, let $\boldsymbol{C} = (C_1, \ldots, C_T)$ and let $S_t(\boldsymbol{c})$ be the available set obtained by applying (2). Its *complete-data likelihood* is

$$\mathbb{P}_{\boldsymbol{v}}(\boldsymbol{C} = \boldsymbol{c} \mid \boldsymbol{y}, T) = \prod_{t=1}^{T} \frac{e^{v_{c_t}}}{1 + \sum_{k \in S_t(\boldsymbol{c})} e^{v_k}}.$$

However, the seller does not observe the choice trajectory, which must be consistent with the terminal sales:

$$\mathcal{P}(\boldsymbol{x}, T) = \left\{ \boldsymbol{c} \in \{0, 1, \ldots, J\}^T : \sum_{t=1}^{T} \mathbb{I}_{\{c_t = j\}} = x_j \text{ for every } j \in [J] \right\}.$$

Because $x_j \le y_j$, every trajectory in $\mathcal{P}(\boldsymbol{x}, T)$ is consistent with the initial inventory and the terminal sales: before any purchase of j, fewer than $x_j \le y_j$ units of j have been sold. The number of such trajectories can be exponentially large in T, as illustrated in Figure 1. The *observed-data* likelihood aggregates the complete-data likelihood over possible choice trajectories:

$$L(\boldsymbol{v} \mid \boldsymbol{y}, \boldsymbol{x}, T) = \sum_{\boldsymbol{c} \in \mathcal{P}(\boldsymbol{x}, T)} \mathbb{P}_{\boldsymbol{v}}(\boldsymbol{C} = \boldsymbol{c} \mid \boldsymbol{y}, T). \tag{3}$$

Note that we treat $\boldsymbol{y}$ and T as given and $\boldsymbol{x}$ as the random sample being generated when expressing the likelihood in this way. Directly enumerating $\mathcal{P}(\boldsymbol{x}, T)$ can therefore be computationally prohibitive, particularly when likelihood evaluation is repeated during optimization. Applying EM does not by itself remove this difficulty: its expectation step must still average over latent histories conditional on

terminal sales. This latent-history calculation is a central challenge in Musalem et al. (2010), Conlon and Mortimer (2013), Shao and Kleywegt (2020), and Steeneck et al. (2022).

We next represent the process in a different way to provide a tractable characterization of the likelihood. We use the *continuous-time representation* developed for stockout-based substitution in Zhou et al. (2026). The key idea is to embed the discrete-time events in Section 2, such as $C_{t,n}$ and $I_{j,t,n}$, in a continuous-time system with *Poisson processes*. More precisely, consider $J+1$ independent Poisson processes $(\Pi_j(\tau))_{\tau\geq 0}$, $j=0,\dots,J$, with rates θ_j and $\theta_0=1$. We use τ for continuous time and retain t for the customer index. For $j\geq 1$, only the first y_j counts can represent sales; process 0 records no-purchase outcomes and remains uncapped. Define the capped processes by

$$N_0(\tau)=\Pi_0(\tau), \qquad N_j(\tau)=\min\{y_j,\Pi_j(\tau)\}, \qquad j\in[J]. \tag{4}$$

Note that the capped processes are still independent. This differs from the discrete-time model, where products share the same sequence of customers. To connect the two systems, define $N_{\text{tot}}(\tau) \coloneqq N_0(\tau)+\sum_{j=1}^{J} N_j(\tau)$ as the total number of customer choices by (continuous) time τ, including no purchases. For each integer $t\geq 1$, define two quantities

$$\sigma_t=\inf\{\tau\geq 0: N_{\text{tot}}(\tau)\geq t\}, \qquad D_t=j \iff N_j(\sigma_t)-N_j(\sigma_t^-)=1, \qquad j=0,\dots,J,$$

where $N_j(\sigma_t^-)=\lim_{\tau\uparrow\sigma_t} N_j(\tau)$. Here, σ_t is the first time at which $N_{\text{tot}}(\tau)$ reaches t, while D_t records which process contributes the t-th choice. In this definition, $D_1, D_2, \dots$ in the continuous-time system represents sequence of choices and can be mapped to the customer choices $C_1, C_2, \dots$ in the discrete-time system. Zhou et al. (2026, Section 4.1) give the following equivalence in distribution:

LEMMA 1 **(Continuous-time representation).** *For every fixed $t\in\mathbb{Z}_+$,*

$$(C_1,\dots,C_t)\overset{d}{=}(D_1,\dots,D_t), \tag{5}$$

where recall that C_t and D_t are the t-th purchased product (including the outside option) in the discrete-time and continuous-time systems, respectively.

We provide a brief introduction to the intuition behind Lemma 1 while leaving the details to Zhou et al. (2026). The result hinges on the Independence of Irrelevant Alternatives (IIA) property of the MNL choice probability, which keeps the choice-probability ratios among available alternatives unchanged as other products sell out. Under the exponential interarrival times of independent Poisson processes, these choice probabilities become competing clocks, with the memoryless property preserving the same choice dynamics after each event. These properties (IIA and memoryless) guarantee that although the inventory processes of different products in the discrete-time system seem to be more intertwined, they are essentially "independent" under the continuous-time view.

One implication of Lemma 1 is that, if we plug in $t = T_n$, the number of customers in period n, then $(N_1(\sigma_{T_n}), \ldots, N_J(\sigma_{T_n}))$ have the same joint distribution as the terminal-sales vector $\boldsymbol{X}_n$, conditional on $(\boldsymbol{y}_n, T_n)$. We note that the underlying Poisson processes are independent as stochastic processes, so $N_j(\tau)$s are independent at any fixed τ. However, $(N_1(\sigma_{T_n}), \ldots, N_J(\sigma_{T_n}))$ are dependent because the common random time σ_{T_n} is determined by the total arrivals reaching T_n.

Using Lemma 1, we next analyze the likelihood of terminal sales $(x_{1,n}, \ldots, x_{J,n})$, or equivalently the joint distribution of $(N_1(\sigma_{T_n}), \ldots, N_J(\sigma_{T_n}))$, through the continuous-time system. To do that, we first study the joint probability at a fixed continuous time, without conditioning on a stopping time. Given the terminal sales $\boldsymbol{x}$, one can recover $\mathcal{B}$ and $\mathcal{C}$, the subsets of products that are sold out and in stock at the end of the period, respectively. For total no-purchase m and a fixed time $\tau > 0$, by independence of the Poisson processes, we have

$$G_m(\tau; \boldsymbol{\theta}) \coloneqq \mathbb{P}(N_0(\tau) = m, N_1(\tau) = x_1, \ldots, N_J(\tau) = x_J) = \pi_m(\tau) \prod_{j \in \mathcal{C}} \pi_{x_j}(\theta_j \tau) \prod_{j \in \mathcal{B}} \overline{F}_{y_j}(\theta_j \tau). \tag{6}$$

Here, $\pi_k(\mu) \coloneqq \mathbb{P}(\mathrm{Poi}(\mu) = k) = e^{-\mu}\mu^k/k!$ is the probability mass function (PMF) of the Poisson distribution, and $\overline{F}_k(\mu) \coloneqq \mathbb{P}\{\mathrm{Poi}(\mu) \geq k\} = \sum_{l=k}^{\infty} e^{-\mu}\mu^l/l!$ is the tail probability. The joint distribution reflects products in $\mathcal{B}$ and $\mathcal{C}$ differently: for $j \in \mathcal{C}$, $N_j(\tau) = x_j < y_j$ is equivalent to $\Pi_j(\tau) = x_j$; for $j \in \mathcal{B}$, $N_j(\tau) = y_j$ is equivalent to $\Pi_j(\tau) \geq y_j$. The number of no purchases $N_0(\tau)$ remains uncapped.

The expression (6) gives the probability of the state $(m, \boldsymbol{x})$ at a fixed continuous time τ. To relate it to the likelihood of the discrete-time system, we need to integrate out τ. More precisely, we are interested in the probability of the event $(N_0(\tau), N_1(\tau), \ldots, N_J(\tau)) = (m, x_1, \ldots, x_J)$ for some $\tau > 0$. Once the continuous-time system ever reaches the state $(m, x_1, \ldots, x_J)$, in the coupled discrete-time system, what we observe is the terminal sales being $\boldsymbol{x} = (x_1, x_2, \ldots, x_J)$ given $T = m + x_1 + \cdots + x_J$ customers and initial inventory $\boldsymbol{y}$. Vice versa, if the discrete-time system observes the terminal sales observation $\boldsymbol{x}$ under T and $\boldsymbol{y}$, then the continuous-time system $(N_0(\tau), N_1(\tau), \ldots, N_J(\tau))$ ends in $(m, x_1, \ldots, x_J)$.

Lemma 2 **(Terminal-sales probability representation)**. *The probability of* $(N_0, N_1, \ldots, N_J)$ *hitting* $(m, x_1, \ldots, x_J)$ *is*

$$\mathbb{P}((N_0(\tau), N_1(\tau), \ldots, N_J(\tau)) = (m, x_1, \ldots, x_J) \text{ for some } \tau > 0) = (1 + \sum_{j \in \mathcal{C}} \theta_j) \int_0^{\infty} G_m(\tau; \boldsymbol{\theta})\, d\tau. \tag{7}$$

To understand (7), it is helpful to view $(N_0(\tau), N_1(\tau), \ldots, N_J(\tau))$ as a continuous-time Markov chain. Because $G_m(\tau; \boldsymbol{\theta})$ is the probability of being in the state $(m, \boldsymbol{x})$ at time τ,

$$\int_0^{\infty} G_m(\tau; \boldsymbol{\theta})\, d\tau = \mathbb{E}\left[\int_0^{\infty} \mathbb{I}_{\{(N_0(\tau), \ldots, N_J(\tau)) = (m, \boldsymbol{x})\}}\, d\tau\right]$$

is the expected occupation time of that state. The process $(N_0(\tau), N_1(\tau), \ldots, N_J(\tau))$ is nondecreasing in each coordinate, so after leaving $(m, \boldsymbol{x})$ it cannot return. In this state, the outside-option process has rate 1, the products $j \in \mathcal{C}$ that remain in stock have rates θ_j, and the sold-out products cannot change any more. Hence the mean holding time in the state is $1/(1+\sum_{j\in\mathcal{C}}\theta_j)$, which yields (7). By Lemma 1, reaching $(m, \boldsymbol{x})$ is equivalent in distribution to observing terminal sales $\boldsymbol{x}$ after $T = m + \sum_{j=1}^{J} x_j$ customer choices.

Compared with the likelihood representation in (3), Lemma 2 already removes the summation over feasible choice trajectories. We next use this representation to derive tractable likelihood expressions under the two data settings.

3.1. The Known-Arrival Setting

For a feasible observation with $\sum_j x_j \le T$, Lemma 2 gives the likelihood

$$L^{\mathrm{K}}(\boldsymbol{v} \mid \boldsymbol{y}, \boldsymbol{x}, T) \coloneqq (1+\sum_{j\in\mathcal{C}}\theta_j)\int_0^\infty \pi_m(\tau)\prod_{j\in\mathcal{C}}\pi_{x_j}(\theta_j\tau)\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\theta_j\tau)\,d\tau, \tag{8}$$

where $s = \sum_{j=1}^{J} x_j$ is total product sales and $m = T - s$ is the observed number of no-purchase outcomes. The notation $L^{\mathrm{K}}(\boldsymbol{v} \mid \boldsymbol{y}, \boldsymbol{x}, T)$ emphasizes that the likelihood is a function of $\boldsymbol{v}$ (equivalently, $\boldsymbol{\theta}$), whereas $\boldsymbol{y}$, $\boldsymbol{x}$, and T are observed.

We next move the terms unrelated to τ out of the integral. Let $c = \sum_{j\in\mathcal{C}} x_j$ be the total sales of the products that remain in stock, and set $\rho_j = \theta_j/(1+\sum_{k\in\mathcal{C}}\theta_k)$ for $j \in \mathcal{A}$. Expanding the Poisson masses in (8) and changing variables to $u = (1+\sum_{j\in\mathcal{C}}\theta_j)\tau$ gives

$$\begin{aligned} L^{\mathrm{K}} &= \frac{(1+\sum_{j\in\mathcal{C}}\theta_j)\prod_{j\in\mathcal{C}}\theta_j^{x_j}}{m!\prod_{j\in\mathcal{C}}x_j!}\int_0^\infty e^{-(1+\sum_{j\in\mathcal{C}}\theta_j)\tau}\tau^{m+c}\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\theta_j\tau)\,d\tau \\ &= \frac{(1+\sum_{j\in\mathcal{C}}\theta_j)^{-m}\prod_{j\in\mathcal{C}}\rho_j^{x_j}}{m!\prod_{j\in\mathcal{C}}x_j!}\int_0^\infty u^{m+c}e^{-u}\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\rho_j u)\,du \\ &= \underbrace{\frac{(m+c)!}{m!\prod_{j\in\mathcal{C}}x_j!}(1+\sum_{j\in\mathcal{C}}\theta_j)^{-m}\prod_{j\in\mathcal{C}}\rho_j^{x_j}}_{C^{\mathrm{K}}(\boldsymbol{v})} \times \underbrace{\frac{1}{(m+c)!}\int_0^\infty u^{m+c}e^{-u}\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\rho_j u)\,du}_{R^{\mathrm{K}}(\boldsymbol{v})}. \end{aligned} \tag{9}$$

In the remaining integral, $u^{m+c}e^{-u}/(m+c)!$, $u>0$, is the density of $U \sim \mathrm{Gamma}(m+c+1, 1)$. Thus

$$R^{\mathrm{K}}(\boldsymbol{v}) = \mathbb{E}\left[\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\rho_j U)\right].$$

This leads to the following result.

THEOREM 1 **(Known-arrival likelihood).** *The likelihood of the known-arrival setting (8) can be written as*

$$L^{\mathrm{K}}(\boldsymbol{v} \mid \boldsymbol{y}, \boldsymbol{x}, T) = C^{\mathrm{K}}(\boldsymbol{v})R^{\mathrm{K}}(\boldsymbol{v}), \tag{10}$$

where

$$C^{\mathrm{K}}(\boldsymbol{v}) = \frac{(m+c)!}{m!\prod_{j\in\mathcal{C}} x_j!}\frac{\prod_{j\in\mathcal{C}}\rho_j^{x_j}}{(1+\sum_{j\in\mathcal{C}}\theta_j)^m}, \tag{11}$$

$$R^{\mathrm{K}}(\boldsymbol{v}) = \mathbb{E}_{U\sim\mathrm{Gamma}(m+c+1,1)}\left[\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\rho_j U)\right]. \tag{12}$$

This factorization is helpful because the terms are treated separately in the log-likelihood $\log L^{\mathrm{K}} = \log C^{\mathrm{K}} + \log R^{\mathrm{K}}$. When no product sells out, $\mathcal{B}=\varnothing$, $c=s$, $m+s=T$, and $R^{\mathrm{K}}=1$, so the likelihood reduces to C^{K}

$$L^{\mathrm{K}}(\boldsymbol{v}\mid\boldsymbol{y},\boldsymbol{x},T) = \frac{T!}{m!\prod_{j\in\mathcal{A}}x_j!}\left(\frac{1}{1+\sum_{k\in\mathcal{A}}\theta_k}\right)^m\prod_{j\in\mathcal{A}}\left(\frac{\theta_j}{1+\sum_{k\in\mathcal{A}}\theta_k}\right)^{x_j}.$$

This is the usual multinomial likelihood under the constant assortment $\mathcal{A}$, considering the possible permutations. When $\mathcal{B}\neq\varnothing$, the product $\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\rho_j u)$ is the probability that independent Poisson random variables with means $(\rho_j u)_{j\in\mathcal{B}}$ reach their respective inventory, and R^{K} averages this probability over the Gamma variable U. The factor R^{K} is therefore interpreted as the correction for the stockout-induced changes in the assortment.

Example 1. Consider the period in Figure 1, with $\boldsymbol{y}=(2,3,2)$, $T=8$, and $\boldsymbol{x}=(2,1,2)$. Products 1 and 3 sell out, so $\mathcal{B}=\{1,3\}$, $\mathcal{C}=\{2\}$, $m=3$, and $c=1$, and Theorem 1 gives

$$C^{\mathrm{K}}(\boldsymbol{v}) = \frac{4\theta_2}{(1+\theta_2)^4}, \qquad R^{\mathrm{K}}(\boldsymbol{v}) = \mathbb{E}_{U\sim\mathrm{Gamma}(5,1)}\left[\overline{F}_2\left(\frac{\theta_1 U}{1+\theta_2}\right)\overline{F}_2\left(\frac{\theta_3 U}{1+\theta_2}\right)\right].$$

The likelihood is thus a single one-dimensional expectation rather than a sum over the 1,680 purchase sequences consistent with the data.

3.2. Poisson-Arrival Setting

When the total number of arrivals is unobserved, the number of no-purchase customers is also latent, so the likelihood sums over its possible values:

$$\begin{aligned} L^{\mathrm{P}}(\lambda,\boldsymbol{v}\mid\boldsymbol{y},\boldsymbol{x}) &= \sum_{m=0}^{\infty}\pi_{s+m}(\lambda)\,\mathbb{P}_{\boldsymbol{v}}(\boldsymbol{X}=\boldsymbol{x}\mid\boldsymbol{y},\widetilde{T}=s+m) \\ &= \sum_{m=0}^{\infty}\pi_{s+m}(\lambda)L^{\mathrm{K}}(\boldsymbol{v}\mid\boldsymbol{y},\boldsymbol{x},T=s+m). \end{aligned} \tag{13}$$

Substituting (8) into (13) and interchanging the sum and integral lead to

$$L^{\mathrm{P}}(\lambda,\boldsymbol{v}\mid\boldsymbol{y},\boldsymbol{x}) = (1+\sum_{j\in\mathcal{C}}\theta_j)\int_0^{\infty}\sum_{m=0}^{\infty}\pi_{s+m}(\lambda)\pi_m(\tau)\prod_{j\in\mathcal{C}}\pi_{x_j}(\theta_j\tau)\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\theta_j\tau)\,d\tau.$$

The only term in the integral that involves m is $\sum_{m=0}^{\infty}\pi_{s+m}(\lambda)\pi_m(\tau)$. Expanding the Poisson PMF, we rewrite it as

$$\sum_{m=0}^{\infty}\pi_{s+m}(\lambda)\pi_m(\tau) = e^{-(\lambda+\tau)}\frac{\lambda^s}{s!}\Phi_s(\lambda\tau),$$

where

$$\Phi_s(z) := s! \sum_{m=0}^{\infty} \frac{z^m}{m!(s+m)!}, \qquad z \geq 0. \tag{14}$$

The factor $e^{-\lambda}\lambda^s/s!$ does not depend on τ, while $\Phi_s(\lambda\tau)$ retains the effect of summing over the latent no purchases.

Expanding the remaining Poisson masses and applying the same change of variables $u = (1 + \sum_{j\in\mathcal{C}} \theta_j)\tau$ gives

$$\begin{aligned} L^{\mathrm{P}}(\lambda, \boldsymbol{v} \mid \boldsymbol{y}, \boldsymbol{x}) &= \frac{e^{-\lambda}\lambda^s(1+\sum_{j\in\mathcal{C}}\theta_j)\prod_{j\in\mathcal{C}}\theta_j^{x_j}}{s!\prod_{j\in\mathcal{C}}x_j!} \int_0^\infty e^{-(1+\sum_{j\in\mathcal{C}}\theta_j)\tau}\tau^c \Phi_s(\lambda\tau)\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\theta_j\tau)\,d\tau \\ &= \frac{e^{-\lambda}\lambda^s\prod_{j\in\mathcal{C}}\rho_j^{x_j}}{s!\prod_{j\in\mathcal{C}}x_j!} \int_0^\infty u^c e^{-u}\Phi_s\left(\frac{\lambda u}{1+\sum_{j\in\mathcal{C}}\theta_j}\right)\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\rho_j u)\,du \\ &= \underbrace{e^{-\lambda}\lambda^s\frac{c!}{s!}\prod_{j\in\mathcal{C}}\frac{\rho_j^{x_j}}{x_j!}}_{C^{\mathrm{P}}(\lambda,\boldsymbol{v})} \times \underbrace{\frac{1}{c!}\int_0^\infty u^c e^{-u}\Phi_s\left(\frac{\lambda u}{1+\sum_{j\in\mathcal{C}}\theta_j}\right)\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\rho_j u)\,du}_{R^{\mathrm{P}}(\lambda,\boldsymbol{v})}. \end{aligned}$$

In the remaining integral, $u^c e^{-u}/c!$ is the density of $U \sim \mathrm{Gamma}(c+1, 1)$. Therefore,

$$R^{\mathrm{P}}(\lambda, \boldsymbol{v}) = \mathbb{E}\left[\Phi_s\left(\frac{\lambda U}{1+\sum_{j\in\mathcal{C}}\theta_j}\right)\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\rho_j U)\right].$$

Compared with (9), the additional factor $\Phi_s(\lambda u/(1+\sum_{j\in\mathcal{C}}\theta_j))$ remains inside the integral because it depends on u. Thus, it preserves the one-dimensional expectation, giving the following result.

THEOREM 2 **(Poisson-arrival likelihood)**. *The likelihood of the Poisson-arrival setting* (13) *can be written as*

$$L^{\mathrm{P}}(\lambda, \boldsymbol{v} \mid \boldsymbol{y}, \boldsymbol{x}) = C^{\mathrm{P}}(\lambda, \boldsymbol{v}) R^{\mathrm{P}}(\lambda, \boldsymbol{v}), \tag{15}$$

where

$$C^{\mathrm{P}}(\lambda, \boldsymbol{v}) = e^{-\lambda}\lambda^s\frac{c!}{s!}\prod_{j\in\mathcal{C}}\frac{\rho_j^{x_j}}{x_j!}, \tag{16}$$

$$R^{\mathrm{P}}(\lambda, \boldsymbol{v}) = \mathbb{E}_{U\sim\mathrm{Gamma}(c+1,1)}\left[\Phi_s\left(\frac{\lambda U}{1+\sum_{j\in\mathcal{C}}\theta_j}\right)\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\rho_j U)\right]. \tag{17}$$

The main difference from the known-arrival setting is that R^{P} contains an additional factor associated with the latent no purchases. The tail factors $\overline{F}_{y_j}(\rho_j U)$ account for sold-out products as in R^{K}, whereas $\Phi_s(\lambda U/(1+\sum_{j\in\mathcal{C}}\theta_j))$ accounts for the unobserved total number of arrivals. For an observed sales vector $\boldsymbol{x}$ with no sold-out products ($\mathcal{B} = \varnothing$), every customer faced the same assortment $\mathcal{A}$ throughout the period. The likelihood therefore reduces to $L^{\mathrm{P}} = \prod_{j\in\mathcal{A}}\pi_{x_j}(\lambda\theta_j/(1+\sum_{k\in\mathcal{A}}\theta_k))$, the same probability of observing $\boldsymbol{x}$ as in a model without inventory constraints.

4. The Gradient of the Log-Likelihood and Its Computation

The exact likelihood in Section 3 reduces to one-dimensional integration. For estimation, we further need to evaluate the corresponding gradients repeatedly and accurately. We first derive analytic expressions for the gradient of the log-likelihood in the known- and Poisson-arrival settings, and then develop a numerical procedure for evaluating these gradients.

4.1. The Gradient of the Log-Likelihood

In both the known-arrival and Poisson-arrival settings, by Theorems 1 and 2, the likelihood takes the form $L = CR$. Our goal is to evaluate the derivative of $\log L = \log C + \log R$, with respect to the parameters. The term $\log C$ and its gradient have closed forms. As for the gradient of $\log R$, we use the known-arrival setting as an example to show how to obtain a tractable form.

Recall that

$$R^{\mathrm{K}} = \frac{1}{(m+c)!}\int_0^\infty u^{m+c}e^{-u}\prod_{j\in\mathcal{B}}\overline{F}_{y_j}(\rho_j u)\,du.$$

The parameters $\boldsymbol{v}$ appear in ρ_j inside $\overline{F}_{y_j}$. Lemma A.2 in Appendix A justifies differentiating under the integral. Differentiating $\log R^{\mathrm{K}}$ with respect to v_j yields

$$\begin{aligned}\partial_{v_j}\log R^{\mathrm{K}} &= \frac{1}{R^{\mathrm{K}}(m+c)!}\int_0^\infty u^{m+c}e^{-u}\partial_{v_j}\prod_{k\in\mathcal{B}}\overline{F}_{y_k}(\rho_k u)\,du \\ &= \frac{1}{R^{\mathrm{K}}(m+c)!}\int_0^\infty u^{m+c}e^{-u}\prod_{k\in\mathcal{B}}\overline{F}_{y_k}(\rho_k u)\frac{\partial\log(\prod_{k\in\mathcal{B}}\overline{F}_{y_k}(\rho_k u))}{\partial v_j}\,du \qquad (18)\end{aligned}$$

where we use $f'(x) = f(x)(\log f(x))'$ in the second equality. The logarithm helps us break down the product. For $k\in\mathcal{B}$ and $j\in\mathcal{A}$, we have

$$\frac{\partial}{\partial v_j}\log\overline{F}_{y_k}(\rho_k u) = \frac{\rho_k u\,\pi_{y_k-1}(\rho_k u)}{\overline{F}_{y_k}(\rho_k u)}\left(\mathbb{I}_{\{j=k\}} - \rho_j\mathbb{I}_{\{j\in\mathcal{C}\}}\right).$$

Moreover, note that the remaining factors in the integrand in (18) satisfy

$$\frac{1}{R^{\mathrm{K}}(m+c)!}\int_0^\infty u^{m+c}e^{-u}\prod_{k\in\mathcal{B}}\overline{F}_{y_k}(\rho_k u)\,du = 1,$$

because of the definition of R^{K}. This normalization defines a separate random variable W with density

$$f_W(u) = \frac{u^{m+c}e^{-u}\prod_{k\in\mathcal{B}}\overline{F}_{y_k}(\rho_k u)}{R^{\mathrm{K}}(m+c)!}, \qquad u>0, \qquad (19)$$

then (18) can be compactly expressed as

$$\frac{\partial\log R^{\mathrm{K}}}{\partial v_j} = \mathbb{E}_W\left[\frac{\partial}{\partial v_j}\log\left\{\prod_{k\in\mathcal{B}}\overline{F}_{y_k}(\rho_k u)\right\}\Big|_{u=W}\right].$$

For $j\in\mathcal{C}$, the explicit factor contributes $\partial_{v_j}\log C^{\mathrm{K}} = x_j - (m+c)\rho_j$; for $j\notin\mathcal{C}$, that derivative is zero. Combining these terms gives

THEOREM 3 **(Gradient for known arrivals)**. *In the known-arrival setting, for each product j with positive initial inventory,*

$$\frac{\partial \log L^{\mathrm{K}}}{\partial v_j} = \begin{cases} \mathbb{E}_W \left[\dfrac{\rho_j W \, \pi_{y_j - 1}(\rho_j W)}{\overline{F}_{y_j}(\rho_j W)} \right], & j \in \mathcal{B}, \\ x_j - \rho_j \left\{ c + m + \displaystyle\sum_{k \in \mathcal{B}} \mathbb{E}_W \left[\dfrac{\rho_k W \, \pi_{y_k - 1}(\rho_k W)}{\overline{F}_{y_k}(\rho_k W)} \right] \right\}, & j \in \mathcal{C}. \end{cases} \tag{20}$$

Here, recall that $\mathcal{B}$ and $\mathcal{C}$ denote the sold-out products and the products that remain in stock, respectively, $c = \sum_{j \in \mathcal{C}} x_j$, $m = T - s$, and $\rho_j = \theta_j / (1 + \sum_{k \in \mathcal{C}} \theta_k)$. The random variable W has the density given in (19), while $\pi_k(\mu)$ and $\overline{F}_k(\mu)$ denote the probability mass function and tail probability of a Poisson random variable.

This result reflects how the parameters enter the likelihood. For a sold-out product, the derivative with respect to v_j is the expected derivative of its log Poisson tail probability with respect to v_j, with the integration variable held fixed. For a product that remains in stock, changing v_j changes both the explicit factor and the common denominator $1 + \sum_{k \in \mathcal{C}} \theta_k$, and hence changes the tail argument of every sold-out product. These effects produce the observed-sales term x_j and the shared subtraction in (20). Products with zero initial inventory do not enter the likelihood, so their corresponding log-likelihood derivatives are zero.

The representation also simplifies gradient computation. For a given observation of a period, all gradients use the same scalar random variable W. Once the expectations associated with the sold-out products are evaluated, their sum determines the gradient of every product that remains in stock. Thus, no separate numerical evaluation of the expectations is carried out for each component of $\boldsymbol{v}$. Section 4.2 uses this structure to evaluate the likelihood and its gradient simultaneously.

Under Poisson arrivals, the additional factor Φ_s in Theorem 2 contributes to the attraction gradients and to the derivative with respect to the log arrival rate. The same differentiation argument applies. Define a separate random variable $\bar{W}$ whose density incorporates Φ_s:

$$f_{\bar{W}}(u) = \frac{u^c e^{-u} \Phi_s(\lambda u / (1 + \sum_{j \in \mathcal{C}} \theta_j)) \prod_{k \in \mathcal{B}} \overline{F}_{y_k}(\rho_k u)}{R^{\mathrm{P}} c!}, \qquad u > 0, \tag{21}$$

In addition, define

$$\psi_s(z) := z \frac{d}{dz} \log \Phi_s(z), \quad z > 0, \qquad \psi_s(0) = 0, \qquad \overline{m} := \mathbb{E}_{\bar{W}} \left[\psi_s \left(\frac{\lambda \bar{W}}{1 + \sum_{j \in \mathcal{C}} \theta_j} \right) \right].$$

The quantity $\overline{m}$ replaces m in the attraction gradients below. Appendix A provides the derivation.

THEOREM 4 **(Gradient for Poisson arrivals)**. *In the Poisson-arrival setting,*

$$\frac{\partial \log L^{\mathrm{P}}}{\partial v_j} = \begin{cases} \mathbb{E}_{\bar{W}} \left[\dfrac{\rho_j \bar{W} \, \pi_{y_j - 1}(\rho_j \bar{W})}{\overline{F}_{y_j}(\rho_j \bar{W})} \right], & j \in \mathcal{B}, \\ x_j - \rho_j \left\{ c + \overline{m} + \displaystyle\sum_{k \in \mathcal{B}} \mathbb{E}_{\bar{W}} \left[\dfrac{\rho_k \bar{W} \, \pi_{y_k - 1}(\rho_k \bar{W})}{\overline{F}_{y_k}(\rho_k \bar{W})} \right] \right\}, & j \in \mathcal{C}. \end{cases}$$

In addition, for $\eta = \log\lambda$,

$$\frac{\partial \log L^{\mathrm{P}}}{\partial \eta} = \sum_{j=1}^{J} x_j - \lambda + \overline{m}.$$

Both settings therefore require only one-dimensional expectations. The next subsection describes their deterministic evaluation and the resulting optimization problem.

4.2. Computation of the Gradient

The remaining computation has four components: approximating these expectations using numerical integration, evaluating the Poisson tail probability $\overline{F}$, evaluating the functions Φ_s and ψ_s, and how to implement the gradient-based likelihood maximization. Each component can be handled using standard numerical methods, while the specific structure of our likelihood allows these calculations to be implemented efficiently. We briefly describe the numerical methods used for the first three components below. For the last component, we use the L-BFGS-B method (Liu and Nocedal 1989) with a standard line search for step-size selection.

Generalized Gauss–Laguerre method for numerical integration. The expectation in the likelihood can be written generally as $\int_u u^\alpha e^{-u} f(u)du$. This form is naturally suited to generalized Gauss–Laguerre quadrature (NIST Digital Library of Mathematical Functions 2026, Section 3.5(v)). For a chosen Q, the method uses Q nodes and corresponding weights to approximate the integral by a finite weighted sum. The quadrature approximation introduces numerical integration error, and increasing Q provides a finer numerical approximation at a higher computational cost. Appendix B.1 gives the implementation of the generalized Gauss–Laguerre method for our problem.

Evaluation of the Poisson tails. The Poisson tail $\overline{F}_y(\mu)$ equals the regularized lower incomplete gamma function. We evaluate it using `torch.special.gammainc` (PyTorch Documentation 2026). We jointly evaluate the Poisson tail probability $\overline{F}_{y_{j,n}}(\mu)$ and the derivative $\partial \log \overline{F}_{y_{j,n}}(\mu)/\partial \log \mu$ appearing in the gradient formulas, where $\mu = \rho_{j,n} u$ and $\rho_{j,n} = \theta_j/(1+\sum_{k\in\mathcal{C}_n}\theta_k)$ for $j \in \mathcal{A}_n$. Appendix B.2 provides the numerical details for evaluating these quantities.

Evaluation of Φ_s *and* ψ_s. In the Poisson-arrival setting, both quantities involve an infinite sum. (See their definitions in (14) and after (21).) The ratios of successive series terms eventually fall below one and continue to decrease, allowing the omitted tails to be bounded by geometric sums, then it is sufficient to approximate them with the sum of a truncated sequence. Appendix B.3 documents the details and provides an error analysis.

Computational complexity. For observation n, computing the total sales $s_n = \sum_{j=1}^{J} x_{j,n}$, the sales $c_n = \sum_{j\in\mathcal{C}_n} x_{j,n}$ of the products that remain in stock, the common denominator $V_n = 1+\sum_{j\in\mathcal{C}_n}\theta_j$, and the remaining closed-form terms in the gradient requires $O(J)$ time. The remaining work is performed over the Q nodes in the numerical integration. At each node, product-specific calculations are needed only for the sold-out products $j \in \mathcal{B}_n$, while all gradient components share the same normalized

weights for approximating expectations under W or $\bar{W}$. Under Poisson arrivals, evaluating Φ and ψ adds one scalar calculation per node. Counting each scalar-function evaluation as one operation, the gradient evaluation requires

$$O\left(NJ + Q\sum_{n=1}^{N}(1+|\mathcal{B}_n|)\right) \leq O(NQJ). \tag{22}$$

For a fixed Q, this number of operations is linear in both N and J. The arithmetic cost of each scalar evaluation also depends on the number of series terms required. The computation is lighter when fewer products sell out, because fewer Poisson tails need to be evaluated. For $N = 500$, $J = 100$, and $Q = 128$, the average time per gradient iteration is approximately 0.038 seconds under known arrivals and 0.080 seconds under Poisson arrivals on one NVIDIA RTX 5090 GPU.

5. Theoretical Analysis

The gradient expressions in Theorems 3 and 4 make gradient-based methods computationally efficient in each iteration. We next study the curvature of the log-likelihood and the identification of the model parameters, which provide conditions for the convergence of the gradient-based methods.

5.1. Known Arrivals

We first consider the known-arrival setting. Theorem 1 factorizes the likelihood into two terms. Because we maximize the log-likelihood, it suffices to establish the concavity of $\log C^{\mathrm{K}}$ and $\log R^{\mathrm{K}}$. We focus on these two terms *given the observation of a single selling period*. We can show that the term $\log C^{\mathrm{K}}$ is (weakly) concave, while $\log R^{\mathrm{K}}$ is strictly concave when at least one product sells out. On the other hand, if no product sells out, then $\log C^{\mathrm{K}}$ is strictly concave. These properties yield the following result.

THEOREM 5 **(Global concavity in the known-arrival setting)**. *For any known-arrival observation $(\boldsymbol{y}, \boldsymbol{x}, T)$ with $T \geq 1$, the sample log-likelihood $\log L^{\mathrm{K}}(\boldsymbol{v})$ is jointly strictly concave in the log attractions of products with positive initial inventory, provided at least one such product exists.*

For a single-period observation, the log-likelihood is strictly increasing in v_j whenever product j sells out. Consequently, a finite maximizer does not exist if at least one product sells out in that period. After aggregating the data from multiple periods, the sample log-likelihood is strictly concave in the parameters of products initially stocked in at least one period with $T_n \geq 1$. Therefore, any finite stationary point is the unique global maximizer over these parameters.

Because identifiability concerns whether different parameter vectors generate different terminal-sales distributions, it is a weaker condition than global concavity. Indeed, if the log-likelihood is strictly concave, then it leads to identifiability. This observation leads to the following result.

COROLLARY 1 **(Identifiability in the known-arrival setting)**. *For any fixed $(\boldsymbol{y}, T)$ with $T \geq 1$, the parameters θ_j of products with positive initial inventory are identifiable from the distribution of terminal sales $\boldsymbol{X}$.*

5.2. Poisson Arrivals

Unlike the known-arrival setting, the Poisson-arrival log-likelihood need not be globally concave in $(\eta, \boldsymbol{v})$ for a single observation. Lemma A.7 in Appendix A demonstrates this using a zero-sale observation, showing that nonconcavity can arise even without an observed stockout.

5.2.1. Identifiability Despite this nonconcavity, the parameters are still identifiable from the terminal-sales distribution. Under Poisson arrivals, the main identification challenge is separating the unknown arrival rate λ from product attractions using only terminal sales. Before any stockout, product j is purchased at rate $\lambda\theta_j/(1+\sum_{k=1}^{J}\theta_k)$. Hence, the pre-stockout sales dynamics depend on λ and $\boldsymbol{\theta}$ only through these product-specific purchase rates, making the two components indistinguishable before stockouts occur. For instance, when there is one product ($J=1$), we have the terminal sales $X_1 \overset{d}{=} \min\{y_1, Z\}$, where $Z \sim \mathrm{Poi}\left(\frac{\lambda\theta_1}{1+\theta_1}\right)$. Thus, the terminal-sales distribution identifies only $\lambda\theta_1/(1+\theta_1)$ as a whole.

For $J \geq 2$, stockout-induced substitution provides additional information. To separate the product purchase rates from the remaining parameter, we reparametrize:

$$\alpha_j = \frac{\lambda\theta_j}{1+\sum_{k=1}^{J}\theta_k}, \qquad \alpha_+ = \sum_{j=1}^{J}\alpha_j, \qquad \delta = \frac{\lambda}{1+\sum_{k=1}^{J}\theta_k}. \tag{23}$$

Here α_j is the purchase rate of product j before any stockout, and δ is the corresponding no-purchase rate. The transformation is one-to-one: $\lambda = \alpha_+ + \delta$ and $\theta_j = \alpha_j/\delta$. Hence identifying $(\lambda, \boldsymbol{\theta})$ is equivalent to identifying $(\boldsymbol{\alpha}, \delta)$. The zero-sale probability identifies the total pre-stockout purchase rate α_+:

$$\mathbb{P}(\boldsymbol{X} = \mathbf{0} \mid \boldsymbol{y}) = e^{-\alpha_+}. \tag{24}$$

For any product j with $y_j \geq 2$, selling exactly one unit of product j and nothing else does not cause a stockout, so

$$\mathbb{P}(X_j = 1,\ X_k = 0 \text{ for } k \neq j \mid \boldsymbol{y}) = \alpha_j e^{-\alpha_+}. \tag{25}$$

Thus, α_j is identified for every product with at least two units of initial inventory.

If $y_j \geq 2$ for every product, (25) identifies the entire vector $\boldsymbol{\alpha}$. With $\boldsymbol{\alpha}$ fixed, Lemma A.5 in Appendix A shows that expected total sales are strictly decreasing in δ. Since this expectation is determined by the terminal-sales distribution, δ is also identified.

When $y_j = 1$, in Appendix A we show that, the probability of the event that only product j is sold is strictly increasing in both α_j and δ, holding α_+ fixed. If two parameter vectors generate the same terminal-sales distribution and $\delta' > \delta$, equality of their one-sale probabilities requires $\alpha_j' < \alpha_j$ for every product with $y_j = 1$, while (25) fixes α_j for every product with $y_j \geq 2$. This contradicts the equality of α_+ implied by (24). Reversing the two parameter vectors rules out $\delta' < \delta$, so δ is identified, and the same strict monotonicity identifies the remaining components of $\boldsymbol{\alpha}$.

THEOREM 6 **(Identifiability under Poisson arrivals).** *Suppose $J \geq 2$ and fix any $\boldsymbol{y} \in \mathbb{Z}^J_{++}$. Then $(\lambda, \boldsymbol{v}) \in \mathbb{R}_{++} \times \mathbb{R}^J$ is identifiable from the conditional terminal-sales distribution.*

Theorem 6 establishes identifiability without observed stockout times, variation in initial inventories, or the total customer arrivals. This differs from identifiability based on observed assortment variation. Abdallah and Vulcano (2021) study MNL estimation from period-level sales with observed assortments and external market-share information. Under a strong-connectivity condition, they establish identifiability of preference parameters but not of period-specific arrival rates. Here, the assortment changes induced by unobserved stockouts help identifying both the common Poisson arrival rate and all product attractions.

5.2.2. Local Concavity As noted at the beginning of Section 5.2, the log-likelihood of a single period need not be concave under Poisson arrivals. Unlike in the known-arrival setting, the sample log-likelihood therefore comes with no global concavity guarantee, and a stationary point reached by a gradient method need not be a maximizer. We show that this difficulty disappears near the true value once enough selling periods are observed: there, the average log-likelihood is strongly concave with high probability. Recall that a twice-differentiable function is called γ-*strongly concave* on a convex set if its second derivative along every unit direction is at most $-\gamma$ on that set.

Strong concavity of the average log-likelihood on a neighborhood has three consequences for gradient-based methods. First, the neighborhood contains at most one stationary point, which then maximizes the log-likelihood over the neighborhood. A gradient method that converges to a point in the neighborhood has therefore found this maximizer rather than, say, a saddle point. Second, the gradient bounds the distance to the maximizer over the neighborhood: at any point of the neighborhood, this distance is at most the norm of the gradient of the average log-likelihood divided by γ. Within the neighborhood, a small gradient thus guarantees that the numerical solution is close to the maximizer, and numerical errors in the gradient, such as those from the quadrature in Section 4.2, lead to proportionally small errors in the solution. Third, because the second derivatives are also bounded on the neighborhood (see Appendix A), gradient ascent with a backtracking line search converges linearly once its iterates are close enough to a stationary point in the neighborhood (Boyd and Vandenberghe 2004, Section 9.3).

THEOREM 7 **(Local strong concavity under Poisson arrivals).** *Let $\boldsymbol{\beta} = (\eta, \boldsymbol{v})$ with $\eta = \log \lambda$, and let $\boldsymbol{\beta}_0$ be its true value. Suppose that $J \geq 2$, that the initial inventories are nonrandom and take values in a finite set, that the terminal sales are independent across selling periods, and that the fraction of periods in which every product is initially stocked is bounded away from zero as $N \to \infty$. Then there exist constants $r, \gamma > 0$, independent of N, such that, with probability approaching 1 as $N \to \infty$, the average log-likelihood $N^{-1} \sum_{n=1}^{N} \log L^{\mathrm{P}}(\lambda, \boldsymbol{v} \mid \boldsymbol{y}_n, \boldsymbol{x}_n)$ is γ-strongly concave in $\boldsymbol{\beta}$ on the ball $\{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2 \leq r\}$.*

Because r and γ do not depend on N, the neighborhood and its curvature do not shrink as more selling periods are observed, so the guarantees above do not weaken as data accumulate. The proof in Appendix A builds on the argument for Theorem 6. For a period in which every product is initially stocked, we can show the Fisher information is nonsingular at $\boldsymbol{\beta}_0$, so its expected log-likelihood is strictly concave (Lemma A.6). Because the periods in which every product is initially stocked make up a positive fraction of the sample, the average log-likelihood inherits this curvature at $\boldsymbol{\beta}_0$ as N grows, and the concavity extends to a neighborhood of $\boldsymbol{\beta}_0$.

6. Numerical Study

In this section, we evaluate the accuracy and computational cost of our method and other benchmark methods, when available, in a large extensive numerical study.

6.1. Experimental Setup

Data generation. Each replication generates one dataset of N selling periods for J products. Within each replication, we draw product attractions while fixing the no-purchase probability under the initial assortment at $p_0 = 0.2$:

$$\theta_j = \frac{1-p_0}{p_0}\frac{\exp(z_j)}{\sum_{\ell=1}^{J}\exp(z_\ell)}, \qquad z_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1).$$

Thus $\sum_j \theta_j = 4$. In each period, we generate the total number of customer arrivals. We use a parameter κ to control the number of customers relative to the number of products: we let $T_n \equiv \lceil \kappa J \rceil$ in the known-arrival setting and $\widetilde{T}_n \sim \text{Poi}(\lambda)$ with $\lambda = \kappa J$ in the Poisson-arrival setting. To generate the initial inventory $y_{j,n}$, we use

$$U_{j,n} \overset{\text{iid}}{\sim} \text{Uniform}(0,2), \qquad y_{j,n} = \max\left\{1, \left\lceil c_s \kappa J \frac{\theta_j}{1+\sum_{l=1}^{J}\theta_l} U_{j,n} \right\rceil\right\}. \tag{26}$$

The term $\kappa J \theta_j (1+\sum_{l=1}^{J}\theta_l)^{-1}$ is the baseline expected demand for product j when all products remain available. The multipliers $U_{j,n}$ create variation in stocking levels across selling periods, while c_s controls the inventory availability.

For each dataset, we measure the realized stockout rate by $\frac{1}{NJ}\sum_{n=1}^{N}|\mathcal{B}_n|$, the fraction of product–period observations ending in a stockout. For a target stockout rate ζ, we calibrate c_s using independent pilot simulations and then generate the sample with the calibrated multiplier. The details are provided in Appendix C. The hyperparameters (J, N, κ, ζ) control data generation; each combination defines an experimental setting.

Experimental settings and benchmarks. The main experiment uses $J \in \{10, 50, 100\}$, $\kappa \in \{5, 10\}$, $\zeta \in \{0.25, 0.50\}$, and $N \in \{200, 500\}$, with 10 replications per experimental setting. Within each replication, all methods use the same generated dataset.

Throughout the numerical study, we refer to our proposed likelihood-based estimator as `Exact-MLE`. In the known-arrival setting, we use the following benchmark methods: (1) `Naïve MNL` ignores stockout-induced assortment changes and fits the MNL model using the initial assortment throughout the selling period. (2) `SK-SAA` (Shao and Kleywegt 2020) approximates the likelihood by sampling latent stockout orders. (3) For small experimental settings such as $J = 10$, we report `DA-MCMC` (Wan et al. 2018), a computationally intensive Bayesian method that explicitly samples the latent stockout trajectory. Under Poisson arrivals, `SK-SAA` provides the closest likelihood-based benchmark and jointly estimates the arrival rate and choice parameters by sampling latent stockout histories. We also compare with `Naïve MNL` when it is supplied with the true arrival rate λ, because it is otherwise unidentifiable, and with `Exact-MLE` under observed arrival counts to quantify the value of arrival information.

Computation and metrics. We impose a 600-second time limit per replication on all methods except `DA-MCMC` . Due to its computational cost, `DA-MCMC` is evaluated only for $J = 10$ and is run without a time limit, as some of the larger instances require more than 600 seconds. If a method runs over time, we report "–" in the table. The experiments are run on one NVIDIA RTX 5090 GPU.

For each method, let $\widehat{\theta}_j$ denote its estimate of θ_j, and write $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_J)$. We report the root mean square of productwise relative errors, rRMSE_θ, and the relative ℓ_2 error, RelErr_θ:

$$\mathrm{rRMSE}_\theta = \left\{ \frac{1}{J} \sum_{j=1}^{J} \left(\frac{\widehat{\theta}_j - \theta_j}{\theta_j} \right)^2 \right\}^{1/2}, \qquad \mathrm{RelErr}_\theta = \frac{\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2}{\|\boldsymbol{\theta}\|_2}.$$

The first metric weights products equally when averaging squared relative errors; the second measures error relative to the norm of the attraction vector. Under Poisson arrivals, writing $\widehat{\lambda}$ for the estimated arrival rate, we additionally report the relative absolute error $\mathrm{RelErr}_\lambda = |\widehat{\lambda} - \lambda|/\lambda$. Errors are computed within each replication and then averaged across replications.

6.2. Known Arrivals

Table 2 reports estimation errors and estimation times under known arrivals. Overall, our method `Exact-MLE` achieves the best accuracy in the most settings and is computationally efficient. We summarize the patterns below. First, the estimation error of `Naïve MNL` comes from ignoring the stockout substitution. It doesn't diminish as the data size increases. Second, when the Bayesian method `DA-MCMC` can compute a result within the time limit in the most settings, it is comparable to our method `Exact-MLE` . However, `DA-MCMC` is not practical computationally even for instances of mild scales. Third, in terms of scalability, `SK-SAA` is comparable to our method, with around 55%

$\zeta = 0.25$

J	κ	N	Exact-MLE			SK-SAA			DA-MCMC			Naïve MNL		
			rRMSE$_\theta$	RelErr$_\theta$	Estimation time (s)	rRMSE$_\theta$	RelErr$_\theta$	Estimation time (s)	rRMSE$_\theta$	RelErr$_\theta$	Estimation time (s)	rRMSE$_\theta$	RelErr$_\theta$	Estimation time (s)
10	5	200	**0.0510**	**0.0335**	0.75	0.0517	0.0336	0.95	0.0522	0.0338	158.5	0.1443	0.1693	0.30
10	5	500	**0.0309**	**0.0191**	0.91	0.0314	0.0202	1.36	0.0313	0.0192	386.1	0.1402	0.1605	0.32
10	10	200	**0.0381**	**0.0264**	0.77	0.0383	0.0269	0.93	0.0402	0.0267	152.0	0.1381	0.1513	0.32
10	10	500	0.0269	**0.0200**	0.98	0.0276	0.0203	1.38	**0.0266**	0.0202	382.2	0.1377	0.1595	0.31
50	5	200	**0.0600**	**0.0251**	1.00	0.0601	0.0256	1.58		–		0.1341	0.1389	0.34
50	5	500	**0.0397**	**0.0165**	1.39	0.0402	0.0185	2.02		–		0.1318	0.1372	0.35
50	10	200	**0.0417**	**0.0192**	1.03	0.0429	0.0232	1.47		–		0.1334	0.1430	0.33
50	10	500	**0.0270**	**0.0120**	1.40	0.0278	0.0136	2.01		–		0.1279	0.1357	0.35
100	5	200	**0.0579**	**0.0252**	1.10	0.0586	0.0276	1.88		–		0.1379	0.1345	0.34
100	5	500	**0.0360**	**0.0160**	1.60	0.0364	0.0170	2.40		–		0.1311	0.1310	0.33
100	10	200	**0.0404**	**0.0165**	1.16	0.0420	0.0199	2.18		–		0.1286	0.1280	0.37
100	10	500	**0.0251**	**0.0109**	1.70	0.0262	0.0145	2.70		–		0.1251	0.1283	0.38

$\zeta = 0.50$

J	κ	N	Exact-MLE			SK-SAA			DA-MCMC			Naïve MNL		
			rRMSE$_\theta$	RelErr$_\theta$	Estimation time (s)	rRMSE$_\theta$	RelErr$_\theta$	Estimation time (s)	rRMSE$_\theta$	RelErr$_\theta$	Estimation time (s)	rRMSE$_\theta$	RelErr$_\theta$	Estimation time (s)
10	5	200	**0.0536**	**0.0373**	0.79	0.0552	0.0415	1.02	0.0539	0.0381	241.6	0.2834	0.3082	0.32
10	5	500	**0.0352**	0.0221	0.95	0.0369	0.0266	1.48	0.0360	**0.0209**	586.8	0.2766	0.3001	0.31
10	10	200	0.0402	0.0231	0.80	0.0441	0.0346	1.00	**0.0398**	**0.0230**	247.0	0.2979	0.3270	0.31
10	10	500	0.0261	0.0152	1.00	0.0283	0.0293	1.55	**0.0257**	**0.0149**	624.2	0.2937	0.3257	0.31
50	5	200	**0.0564**	**0.0301**	1.02	0.0597	0.0345	1.62		–		0.2723	0.2850	0.35
50	5	500	**0.0361**	**0.0173**	1.41	0.0374	0.0269	2.13		–		0.2735	0.2908	0.34
50	10	200	**0.0461**	**0.0210**	1.15	0.0518	0.0327	1.77		–		0.2648	0.2762	0.33
50	10	500	**0.0285**	**0.0153**	1.60	0.0350	0.0296	2.39		–		0.2653	0.2818	0.32
100	5	200	**0.0600**	**0.0268**	1.25	0.0636	0.0362	2.56		–		0.2587	0.2734	0.35
100	5	500	**0.0389**	**0.0169**	1.85	0.0420	0.0266	3.04		–		0.2563	0.2707	0.36
100	10	200	**0.0443**	**0.0198**	1.24	0.0534	0.0366	2.45		–		0.2606	0.2709	0.37
100	10	500	**0.0288**	**0.0125**	1.87	0.0362	0.0263	3.18		–		0.2596	0.2705	0.40

Table 2 Mean estimation error and estimation time under known arrivals.

Notes. Boldface denotes the smallest mean estimation error in each row; `DA-MCMC` is reported only for $J = 10$ because of its estimation time.

more estimation time on average. However, its estimation error is larger than ours especially for large J. This is because `SK-SAA` approximates the likelihood by sampling latent stockout vectors, whereas `Exact-MLE` evaluates the exact one-dimensional representation by deterministic quadrature. Fourth, the estimation error of `Exact-MLE` changes little between the two target stockout rates ($\zeta = 0.25$ and $\zeta = 0.50$).

6.3. Poisson Arrivals

Table 3 reports estimation errors and estimation times under Poisson arrivals. For Poisson arrivals, `SK-SAA` is the only benchmark that is applicable to this setting. To enrich the comparison, we include `Naïve MNL`, which treats λ as known and ignores the stockout-induced assortment changes. The true value of λ needs to be provided for `Naïve MNL` to be identifiable. We also fit `Exact-MLE` after revealing the realized counts $\{\widetilde{T}_n\}_{n=1}^N$ from the same Poisson-generated data. Conditional on these counts, the fit uses the known-arrival likelihood. This comparison measures the effect of observing arrival counts on estimation error.

The advantage of `Exact-MLE` relative to `SK-SAA` is more pronounced under Poisson arrivals, for both estimation accuracy and computational efficiency. In particular, the reduction of estimation time

Panel A: Target stockout rate $\zeta = 0.25$

J	κ	N	Exact-MLE				SK-SAA				Exact-MLE known arrivals	Naïve MNL true λ
			rRMSE$_\theta$	RelErr$_\theta$	RelErr$_\lambda$	Estimation time (s)	rRMSE$_\theta$	RelErr$_\theta$	RelErr$_\lambda$	Estimation time (s)	rRMSE$_\theta$	rRMSE$_\theta$
10	5	200	0.3893	**0.3808**	**0.0605**	1.55	**0.3883**	0.3913	0.1743	14.95	0.0588	0.1476
10	5	500	**0.1862**	**0.1826**	**0.0350**	1.62	0.4038	0.4057	0.1976	39.46	0.0342	0.1427
10	10	200	**0.2436**	**0.2435**	**0.0623**	1.60	0.2808	0.2801	0.0930	12.78	0.0455	0.1473
10	10	500	**0.1038**	**0.1044**	**0.0230**	1.62	0.2366	0.2350	0.0815	36.25	0.0231	0.1435
50	5	200	0.4433	0.4399	0.0785	4.15	**0.2593**	**0.2541**	**0.0740**	30.35	0.0547	0.1332
50	5	500	**0.2774**	**0.2705**	**0.0497**	4.08	0.3596	0.3593	0.1425	88.74	0.0347	0.1342
50	10	200	**0.2195**	**0.2110**	**0.0463**	5.10	0.2391	0.2319	0.0827	34.22	0.0416	0.1366
50	10	500	**0.0959**	**0.0880**	**0.0223**	4.92	0.1780	0.1752	0.0546	92.28	0.0248	0.1305
100	5	200	0.3111	0.3015	0.0658	7.22	**0.2003**	**0.1958**	**0.0633**	41.19	0.0606	0.1315
100	5	500	0.2577	0.2569	**0.0565**	8.12	**0.2210**	**0.2198**	0.0709	101.21	0.0376	0.1275
100	10	200	**0.1966**	**0.1920**	**0.0505**	10.46	0.2851	0.2817	0.1096	43.68	0.0438	0.1340
100	10	500	**0.1351**	**0.1320**	**0.0313**	11.06	0.3327	0.3316	0.1232	131.53	0.0288	0.1311

Panel B: Target stockout rate $\zeta = 0.50$

J	κ	N	Exact-MLE				SK-SAA				Exact-MLE known arrivals	Naïve MNL true λ
			rRMSE$_\theta$	RelErr$_\theta$	RelErr$_\lambda$	Estimation time (s)	rRMSE$_\theta$	RelErr$_\theta$	RelErr$_\lambda$	Estimation time (s)	rRMSE$_\theta$	rRMSE$_\theta$
10	5	200	**0.1880**	**0.1795**	**0.0560**	1.49	0.3637	0.3714	0.1824	15.79	0.0614	0.3134
10	5	500	**0.1668**	**0.1596**	**0.0438**	1.49	0.4220	0.4308	0.2186	33.92	0.0366	0.3086
10	10	200	**0.1372**	**0.1297**	**0.0290**	1.92	0.2506	0.2463	0.1032	19.96	0.0428	0.3059
10	10	500	**0.1184**	**0.1110**	**0.0259**	1.94	0.2866	0.2884	0.1168	38.34	0.0283	0.3106
50	5	200	**0.1568**	**0.1447**	**0.0435**	4.72	0.3970	0.3881	0.2403	31.04	0.0683	0.2686
50	5	500	**0.0994**	**0.0873**	**0.0230**	5.30	0.5314	0.5412	0.4141	81.11	0.0409	0.2645
50	10	200	**0.1125**	**0.0999**	**0.0240**	6.31	0.3241	0.3293	0.1988	32.87	0.0480	0.2610
50	10	500	**0.0673**	**0.0597**	**0.0142**	6.66	0.4957	0.5043	0.3291	106.07	0.0301	0.2663
100	5	200	**0.1810**	**0.1636**	**0.0330**	9.53	0.4305	0.4324	0.3078	42.67	0.0669	0.2638
100	5	500	**0.0929**	**0.0782**	**0.0175**	9.94	0.6583	0.6622	0.5944	113.81	0.0405	0.2582
100	10	200	**0.1198**	**0.1077**	**0.0298**	10.91	0.4060	0.4124	0.2134	44.77	0.0431	0.2568
100	10	500	**0.0860**	**0.0787**	**0.0212**	10.94	0.5686	0.5751	0.3791	132.87	0.0279	0.2557

Table 3 Poisson-arrival results averaged over ten replications.

Notes. The last two columns use observed arrivals and the true arrival rate, respectively. We only highlight the error between `Exact-MLE` and `SK-SAA` .

is more than 90% and that of estimation error is more than 50%. When total arrivals are unobserved, `SK-SAA` must also sum over possible arrival counts while approximating the likelihood associated with each count. `Exact-MLE` sums out the unobserved arrival count analytically and retains a one-dimensional likelihood representation. Compared to `Exact-MLE` with known arrivals, the estimation error does increase significantly (more than 300%). It highlights the challenge of not observing the total arrivals, and this challenge is fundamental.

6.4. Large-Scale Experiments

Table 4 reports results from the large-scale experiments using $J, N \in \{500, 1000, 2000\}$, $\kappa = 5$, and $\zeta \in \{0.25, 0.50\}$, with 10 replications per experimental setting. We fit only `Exact-MLE` under both known and Poisson arrivals, as `SK-SAA` and `DA-MCMC` cannot be computed within the time limit. Under known arrivals, estimation error is nearly unchanged as J increases at a fixed N, whereas increasing N consistently reduces the error. At $J = N = 2000$, mean rRMSE is approximately 0.019, with mean estimation times of 48.77 and 66.90 seconds for $\zeta = 0.25$ and 0.50, respectively. Thus, the

Panel A: Known arrivals

		Target stockout rate $\zeta = 0.25$			Target stockout rate $\zeta = 0.50$		
J	N	rRMSE_θ	RelErr_θ	Estimation time (s)	rRMSE_θ	RelErr_θ	Estimation time (s)
500	500	0.0375	0.0144	2.96	0.0377	0.0166	4.18
500	1000	0.0269	0.0102	5.05	0.0271	0.0118	7.20
500	2000	0.0190	0.0075	9.10	0.0190	0.0084	13.59
1000	500	0.0372	0.0152	8.24	0.0389	0.0160	10.51
1000	1000	0.0263	0.0106	14.76	0.0277	0.0114	20.18
1000	2000	0.0185	0.0073	27.92	0.0192	0.0079	38.48
2000	500	0.0389	0.0137	13.88	0.0385	0.0155	17.55
2000	1000	0.0273	0.0097	25.96	0.0272	0.0110	34.44
2000	2000	0.0191	0.0069	48.77	0.0192	0.0078	66.90

Panel B: Poisson arrivals

		Target stockout rate $\zeta = 0.25$				Target stockout rate $\zeta = 0.50$			
J	N	rRMSE_θ	RelErr_θ	RelErr_λ	Estimation time (s)	rRMSE_θ	RelErr_θ	RelErr_λ	Estimation time (s)
500	500	0.1533	0.1415	0.0314	31.77	0.1465	0.1381	0.0305	47.49
500	1000	0.1016	0.0920	0.0215	37.49	0.0962	0.0900	0.0203	60.23
500	2000	0.0370	0.0284	0.0058	50.37	0.0549	0.0500	0.0116	93.30
1000	500	0.0376	0.0145	0.0006	53.58	0.0442	0.0246	0.0047	81.83
1000	1000	0.0273	0.0113	0.0007	63.80	0.0365	0.0246	0.0054	99.76
1000	2000	0.0190	0.0076	0.0005	93.43	0.0279	0.0210	0.0049	207.74
2000	500	0.0376	0.0148	0.0007	110.70	0.0441	0.0253	0.0048	149.96
2000	1000	0.0267	0.0105	0.0006	142.24	0.0347	0.0227	0.0048	338.84
2000	2000	0.0189	0.0075	0.0005	285.07	0.0277	0.0207	0.0048	565.57

Table 4 Mean estimation error and estimation time for `Exact-MLE` in the large-scale experiments.

increase in parameter dimension does not materially degrade per-product accuracy in this design, and `Exact-MLE` remains computationally practical at the largest tested scale.

Under Poisson arrivals, mean rRMSE also decreases with N. Estimation time is higher at the larger target stockout rate $\zeta = 0.50$. But even at $J = N = 2000$, the average estimation time is 285.07 seconds for $\zeta = 0.25$ and 565.57 seconds for $\zeta = 0.50$. These results show that the one-dimensional likelihood expression makes `Exact-MLE` very robust and scalable.

7. Concluding Remarks and Future Directions

This paper develops maximum-likelihood estimation for MNL demand when only initial inventories and terminal product sales are observed, while customer choices, stockout times, and the assortments faced by individual customers remain latent. Because we can express the (log-)likelihood function and its gradient with a one-dimensional integral, we are able to use standardized numerical recipes to maximize the likelihood efficiently, even for large-scale problems.

Several directions remain for future research. First, the continuous-time representation is made possible by the IIA property of the MNL choice model. It is natural to ask whether other choice

models, including observed covariates or customer heterogeneity, retain a tractable likelihood. A one-dimensional representation is unlikely, but we may obtain a representation with a much lower dimension than the latent space of choice sequences. Second, we have seen fruitful results in inventory management and demand estimation when applying the continuous-time view to dynamic stockout substitution, developed in Zhou et al. (2026). An exciting direction is to explore other operational levers that can be analyzed thanks to this viewpoint.

References

Abdallah T, Vulcano G (2021) Demand estimation under the multinomial logit model from sales transaction data. *Manufacturing & Service Operations Management* 23(5):1196–1216.

Agrawal N, Smith SA (1996) Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics* 43(6):839–861.

Akchen YC, Caro F (2026) On size substitution and its role in assortment and inventory planning. *Manufacturing & Service Operations Management* 28(2):624–642.

Anupindi R, Dada M, Gupta S (1998) Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science* 17(4):406–423.

Aouad A, Segev D (2023) The stability of MNL-based demand under dynamic customer substitution and its algorithmic implications. *Operations Research* 71(4):1216–1249.

Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge, UK: Cambridge University Press).

Chen B, Chao X (2020) Dynamic inventory control with stockout substitution and demand learning. *Management Science* 66(11):5108–5127.

Chen X, Adulyasak Y, Cordeau JF (2026) Bayesian learning of demand and stockout-based substitution for risk-aware inventory optimization. *Available at SSRN 7444520* .

Cho S, Ferguson M, Im J, Pekgün P (2024) Robust demand estimation with customer choice-based models for sales transaction data. *Production and Operations Management* 33(9):1801–1820.

Cho S, Ferguson M, Pekgün P, Vakhutinsky A (2023) Estimating personalized demand with unobserved no-purchases using a mixture model: An application in the hotel industry. *Manufacturing & Service Operations Management* 25(4):1245–1262.

Conlon CT, Mortimer JH (2013) Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics* 5(4):1–30.

Fisher M, Vaidyanathan R (2014) A demand estimation procedure for retail assortment optimization with results from implementations. *Management Science* 60(10):2401–2415.

Honhon D, Gaur V, Seshadri S (2010) Assortment planning and inventory decisions under stockout-based substitution. *Operations Research* 58(5):1364–1379.

Hopp WJ, Xu X (2008) A static approximation for dynamic demand substitution with applications in a competitive market. *Operations Research* 56(3):630–645.

Jain A, Rudi N, Wang T (2015) Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Operations Research* 63(1):134–150.

Karabatı S, Tan B, Öztürk ÖC (2009) A method for estimating stock-out-based substitution rates by using point-of-sale data. *IIE Transactions* 41(5):408–420.

Kök AG, Fisher ML (2007) Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research* 55(6):1001–1021.

Lee J, Gaur V, Muthulingam S, Swisher GF (2016) Stockout-based substitution and inventory planning in textbook retailing. *Manufacturing & Service Operations Management* 18(1):104–121.

Letham B, Letham LM, Rudin C (2016) Bayesian inference of arrival rate and substitution behavior from sales transaction data with stockouts. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1695–1704 (ACM).

Li A, Talluri K, Tekin M (2025) Estimating demand with unobserved no-purchases on revenue-managed data. *Manufacturing & Service Operations Management* 27(1):161–180.

Liang AJ, Jasin S, Uichanco J (2021) Assortment and inventory planning under dynamic substitution with MNL model: An LP approach and an asymptotically optimal policy. Technical report, University of Michigan, Ann Arbor, MI.

Liang Y, Mao X, Wang S (2026) Online joint assortment–inventory optimization under MNL choices. *Operations Research* 74(4):1909–1931.

Liu DC, Nocedal J (1989) On the limited memory BFGS method for large scale optimization. *Mathematical Programming* 45(1–3):503–528.

Liu Z, Rusmevichientong P, Wang M (2026) When less data is better: Eliminating stockout-induced bias in choice estimation. Available at SSRN 7106418.

Mahajan S, van Ryzin G (2001) Stocking retail assortments under dynamic consumer substitution. *Operations Research* 49(3):334–351.

Mersereau AJ (2015) Demand estimation from censored observations with inventory record inaccuracy. *Manufacturing & Service Operations Management* 17(3):335–349.

Musalem A, Olivares M, Bradlow ET, Terwiesch C, Corsten D (2010) Structural estimation of the effect of out-of-stocks. *Management Science* 56(7):1180–1197.

Nahmias S (1994) Demand estimation in lost sales inventory systems. *Naval Research Logistics* 41(6):739–757.

Newman JP, Ferguson ME, Garrow LA, Jacobs TL (2014) Estimation of choice-based models using sales data from a single firm. *Manufacturing & Service Operations Management* 16(2):184–197.

NIST Digital Library of Mathematical Functions (2026) Gauss quadrature. National Institute of Standards and Technology, URL `https://dlmf.nist.gov/3.5`, Section 3.5(v), accessed September 14, 2026.

PyTorch Documentation (2026) `torch.special.gammainc`. PyTorch Special Functions Documentation, URL `https://docs.pytorch.org/docs/stable/special.html#torch.special.gammainc`, accessed September 14, 2026.

Shao H, Kleywegt AJ (2020) Joint estimation of discrete choice model and arrival rate with unobserved stock-out events. arXiv preprint arXiv:2003.02313.

Şimşek AS, Topaloglu H (2018) Technical note—An expectation-maximization algorithm to estimate the parameters of the markov chain choice model. *Operations Research* 66(3):748–760.

Smith SA, Agrawal N (2000) Management of multi-item retail inventory systems with demand substitution. *Operations Research* 48(1):50–64.

Steeneck D, Eng-Larsson F, Jauffred F (2022) Estimating lost sales for substitutable products with uncertain on-shelf availability. *Manufacturing & Service Operations Management* 24(3):1578–1594.

Subramanian S, Harsha P (2021) Demand modeling in the presence of unobserved lost sales. *Management Science* 67(6):3803–3833.

Talluri K, Tekin M (2025) Estimation using marginal competitor sales information. *Journal of Operations Management* 71(5):588–610.

Talluri K, van Ryzin G (2004) Revenue management under a general discrete choice model of consumer behavior. *Management Science* 50(1):15–33.

van Ryzin G, Vulcano G (2017) Technical note—An expectation-maximization method to estimate a rank-based choice model of demand. *Operations Research* 65(2):396–407.

Vulcano G, van Ryzin G, Ratliff R (2012) Estimating primary demand for substitutable products from sales transaction data. *Operations Research* 60(2):313–334.

Wan M, Huang Y, Zhao L, Deng T, Fransoo JC (2018) Demand estimation under multi-store multi-product substitution in high density traditional retail. *European Journal of Operational Research* 266(1):99–111.

Wang X, Xu W (2025) Joint estimation of the arrival rate and customer taste coefficients from censored transactional data. *Production and Operations Management* 34(9):2853–2872.

Zhang J, Ma W, Topaloglu H (2025) Technical note—Leveraging the degree of dynamic substitution in assortment and inventory planning. *Operations Research* 73(3):1248–1259.

Zheng Y, Chen YF, Hedenstierna CPT, Yang J (2026) Offline learning and optimization for multi-product inventory management with stockout-based substitution. *Production and Operations Management* 35(9):3272–3291.

Zhou Z, Wang T, Zhang J (2026) Optimal inventory allocation for indifferent goods under dynamic substitution. *Operations Research* 74(4):2001–2012.

Zhou Z, Zhang J (2026) Approximations to stockout-based substitution under MNL model. Available at SSRN 6427181.

Appendix A: Proofs

Throughout this appendix, we mostly consider a single selling period and suppress the index n. We also write $V = 1 + \sum_{k\in\mathcal{C}} \theta_k$, so that $\rho_j = \theta_j / V$ for $j \in \mathcal{A}$.

We begin with Lemma 1, which is due to Zhou et al. (2026); we include a short proof for completeness because the proof of Lemma 2 builds on it. The idea is to show that the choices (C_t) in the discrete-time system and the labels (D_t) in the continuous-time system are generated by the same rule: given the past, the next choice follows the MNL probabilities over the products that have not sold out.

Proof of Lemma 1: We first check that each D_t is well defined. Since $N_{\mathrm{tot}} \geq N_0 = \Pi_0$ and $\Pi_0(\tau) \to \infty$ almost surely, every σ_t is finite almost surely. Moreover, because independent Poisson processes almost surely have no common jump times, every jump of N_{tot} has size one, and exactly one of $N_0, \ldots, N_J$ jumps at each σ_t.

Next, fix $h \geq 0$ and set $\sigma_0 = 0$. The set of products that have not been sold out for the $(h+1)$st arrival,

$$\widetilde{S}_{h+1} = \{j \in [J] : N_j(\sigma_h) < y_j\},$$

is determined by $(D_1, \ldots, D_h)$. By the strong Markov property at the stopping time σ_h, the processes $(\Pi_j(\sigma_h + \tau) - \Pi_j(\sigma_h))_{\tau\geq 0}$, $j = 0, \ldots, J$, are independent Poisson processes with rates θ_j, independent of the history up to σ_h. The $(h+1)$st choice is made by whichever of Π_0 and Π_j, $j \in \widetilde{S}_{h+1}$, jumps first after σ_h. Since these first jump times are independent exponential random variables with rates 1 and $(\theta_j)_{j\in\widetilde{S}_{h+1}}$,

$$\mathbb{P}(D_{h+1} = j \mid D_1, \ldots, D_h) = \begin{cases} \dfrac{\theta_j}{1 + \sum_{k\in\widetilde{S}_{h+1}} \theta_k}, & j \in \widetilde{S}_{h+1}, \\ \dfrac{1}{1 + \sum_{k\in\widetilde{S}_{h+1}} \theta_k}, & j = 0, \\ 0, & j \in [J] \setminus \widetilde{S}_{h+1}. \end{cases} \tag{A.1}$$

In the discrete-time system, by (2), customer $h+1$ faces $S_{h+1} = \{j \in [J] : y_j - \sum_{i=1}^{h} \mathbb{I}_{\{C_i = j\}} > 0\}$, the same function of $(C_1, \ldots, C_h)$ as $\widetilde{S}_{h+1}$ is of $(D_1, \ldots, D_h)$. Hence, by (1), C_{h+1} given $(C_1, \ldots, C_h)$ has the same conditional distribution as D_{h+1} given $(D_1, \ldots, D_h)$ in (A.1). Since the joint distribution of the first t choices is the product of these conditional distributions, (5) follows. □

For the proof of Lemma 2, we first compute the expected time that the continuous-time system spends in the state $(m, \boldsymbol{x})$ in two ways: by integrating the probability $G_m(\tau; \boldsymbol{\theta})$ of being in the state at time τ, and by multiplying the probability of reaching the state by the expected holding time there. We then show that reaching the state is equivalent to observing the terminal sales $\boldsymbol{x}$ after $T = m + s$ customers, where $s = \sum_{j=1}^{J} x_j$, which justifies the likelihood (8).

Proof of Lemma 2: Write $\boldsymbol{N}(\tau) = (N_0(\tau), \ldots, N_J(\tau))$, and let $\tau_{\mathrm{in}} = \inf\{\tau \geq 0 : \boldsymbol{N}(\tau) = (m, \boldsymbol{x})\}$ and $\tau_{\mathrm{out}} = \inf\{\tau > \tau_{\mathrm{in}} : \boldsymbol{N}(\tau) \neq (m, \boldsymbol{x})\}$ be the times at which the continuous-time system enters and leaves $(m, \boldsymbol{x})$, with $\inf \varnothing = \infty$. The event that $\boldsymbol{N}$ reaches $(m, \boldsymbol{x})$ is $H = \{\tau_{\mathrm{in}} < \infty\}$.

We first verify (6). For fixed $\tau > 0$, $N_0(\tau), \ldots, N_J(\tau)$ are independent, and $\mathbb{P}(N_0(\tau) = m) = \pi_m(\tau)$. For $j \in \mathcal{C}$, $x_j < y_j$, so $N_j(\tau) = x_j$ if and only if $\Pi_j(\tau) = x_j$, which has probability $\pi_{x_j}(\theta_j \tau)$. For $j \in \mathcal{B}$, $x_j = y_j$, so

$N_j(\tau) = x_j$ if and only if $\Pi_j(\tau) \geq y_j$, which has probability $\overline{F}_{y_j}(\theta_j \tau)$. Multiplying these probabilities gives (6).

Next, let $O = \int_0^\infty \mathbb{I}_{\{\boldsymbol{N}(\tau)=(m,\boldsymbol{x})\}}\, d\tau$ be the total time that $\boldsymbol{N}$ spends in $(m, \boldsymbol{x})$. By Tonelli's theorem,

$$\mathbb{E}[O] = \int_0^\infty \mathbb{P}\big(\boldsymbol{N}(\tau) = (m,\boldsymbol{x})\big)\, d\tau = \int_0^\infty G_m(\tau;\boldsymbol{\theta})\, d\tau. \tag{A.2}$$

On the other hand, every coordinate of $\boldsymbol{N}$ is nondecreasing and right-continuous, so $\boldsymbol{N}$ never returns to $(m, \boldsymbol{x})$ after leaving it; thus $O = \tau_{\text{out}} - \tau_{\text{in}}$ on H, and $O = 0$ on H^c. In state $(m, \boldsymbol{x})$, only jumps of Π_0 and of Π_j, $j \in \mathcal{C}$, change $\boldsymbol{N}$, because the other products are at their caps. Hence, by the strong Markov property at τ_{in}, the holding time $\tau_{\text{out}} - \tau_{\text{in}}$ is, conditional on H, exponential with rate $V = 1 + \sum_{j\in\mathcal{C}} \theta_j$, and

$$\mathbb{E}[O] = \mathbb{P}(H)\, \mathbb{E}[\tau_{\text{out}} - \tau_{\text{in}} \mid H] = \frac{\mathbb{P}(H)}{V}. \tag{A.3}$$

Equating (A.2) and (A.3) gives (7).

Finally, we show that $\mathbb{P}(H) = \mathbb{P}(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{y}, T)$ with $T = m + s$. Since every jump of N_{tot} has size one, $N_{\text{tot}}(\tau) = T$ exactly when $\tau \in [\sigma_T, \sigma_{T+1})$, and $\boldsymbol{N}$ is constant on this interval. As the coordinates of $(m, \boldsymbol{x})$ sum to T, it follows that $H = \{\boldsymbol{N}(\sigma_T) = (m, \boldsymbol{x})\}$, the event that the first T choices $(D_1, \ldots, D_T)$ consist of m no-purchase outcomes and x_j purchases of each product j. By Lemma 1, $(D_1, \ldots, D_T)$ has the same distribution as $(C_1, \ldots, C_T)$, so $\mathbb{P}(H) = \mathbb{P}(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{y}, T)$, which justifies the likelihood (8). □

We next prove the gradient formulas in Theorems 3 and 4. In both settings, the likelihood factorizes as $L = CR$. The factor C is explicit and can be differentiated directly, whereas the factor R is an integral, which we differentiate under the integral sign. We first establish two lemmas: Lemma A.1 shows the properties of the Poisson tail $\overline{F}_y$ and of the functions Φ_s and ψ_s that we need, and Lemma A.2 uses these properties to justify the differentiation.

LEMMA A.1 **(Properties of the scalar functions).** *(a) For $y \geq 1$ and $\mu > 0$, $\overline{F}'_y(\mu) = \pi_{y-1}(\mu)$. Consequently, $d \log \overline{F}_y(\mu)/d \log \mu = h_y(\mu)$, where*

$$h_y(\mu) = \frac{\mu \pi_{y-1}(\mu)}{\overline{F}_y(\mu)} = \frac{y \pi_y(\mu)}{\overline{F}_y(\mu)} \in (0, y).$$

(b) For $s \in \mathbb{Z}_+$ and $z \geq 0$, $\Phi'_s(z) = \Phi_{s+1}(z)/(s+1)$,

$$1 \leq \Phi_s(z) \leq e^{2\sqrt{z}}, \qquad \text{and} \qquad 0 \leq \psi_s(z) \leq \frac{z}{s+1}.$$

Proof of Lemma A.1: For part (a), $\overline{F}_y(\mu) = 1 - \sum_{\ell=0}^{y-1} \pi_\ell(\mu)$ and $\pi'_\ell = \pi_{\ell-1} - \pi_\ell$ with $\pi_{-1} \equiv 0$, so the derivative telescopes to $\overline{F}'_y(\mu) = \pi_{y-1}(\mu)$, and $d \log \overline{F}_y(\mu)/d \log \mu = \mu \pi_{y-1}(\mu)/\overline{F}_y(\mu)$. The second expression for h_y follows from $\mu \pi_{y-1}(\mu) = y \pi_y(\mu)$, and $0 < h_y(\mu) < y$ because $\overline{F}_y(\mu) = \sum_{\ell \geq y} \pi_\ell(\mu) > \pi_y(\mu) > 0$.

For part (b), the power series (14) has an infinite radius of convergence, so it can be differentiated term by term:

$$\Phi'_s(z) = s! \sum_{m=1}^{\infty} \frac{z^{m-1}}{(m-1)!\,(s+m)!} = s! \sum_{m=0}^{\infty} \frac{z^m}{m!\,(s+1+m)!} = \frac{\Phi_{s+1}(z)}{s+1}.$$

Since the coefficient of z^m in Φ_{s+1} is $(s+1)/(s+1+m) \leq 1$ times that in Φ_s, we have $\Phi_{s+1} \leq \Phi_s$, and hence, for $z > 0$ (the case $z = 0$ being trivial),

$$0 \leq \psi_s(z) = \frac{z \Phi'_s(z)}{\Phi_s(z)} = \frac{z}{s+1} \cdot \frac{\Phi_{s+1}(z)}{\Phi_s(z)} \leq \frac{z}{s+1}.$$

Iterating the comparison gives $\Phi_s(z) \le \Phi_0(z) = \sum_{m\ge 0} z^m/(m!)^2$, and since $1/(m!)^2 = \binom{2m}{m}/(2m)! \le 4^m/(2m)!$,

$$\Phi_0(z) \le \sum_{m=0}^{\infty} \frac{(2\sqrt{z})^{2m}}{(2m)!} = \cosh(2\sqrt{z}) \le e^{2\sqrt{z}}.$$

Finally, $\Phi_s(z) \ge 1$ because the series has constant term one and nonnegative coefficients. □

The proofs below use the derivatives of $\log \rho_k$, $k \in \mathcal{B}$, and of $\log z$, where $z = \lambda u/V$ is the argument of Φ_s in (17) and $u > 0$ is held fixed. Since $\log \rho_k = v_k - \log V$, $\log z = \eta + \log u - \log V$, and $\partial \log V/\partial v_j = \rho_j \mathbb{I}_{\{j\in\mathcal{C}\}}$, we have, for $k \in \mathcal{B}$ and $j \in [J]$,

$$\frac{\partial \log \rho_k}{\partial v_j} = \mathbb{I}_{\{j=k\}} - \rho_j \mathbb{I}_{\{j\in\mathcal{C}\}}, \qquad \frac{\partial \log \rho_k}{\partial \eta} = 0, \qquad \frac{\partial \log z}{\partial v_j} = -\rho_j \mathbb{I}_{\{j\in\mathcal{C}\}}, \qquad \frac{\partial \log z}{\partial \eta} = 1. \tag{A.4}$$

These derivatives do not depend on u and lie in $[-1, 1]$, because $\rho_j < 1$ for $j \in \mathcal{C}$ and $\mathcal{B} \cap \mathcal{C} = \varnothing$.

LEMMA A.2 **(Differentiation in the integral)**. *The functions R^{K} and R^{P} are continuously differentiable in $\boldsymbol{v}$ and in $(\eta, \boldsymbol{v})$, respectively, and, for any coordinate ϑ of the corresponding parameter vector,*

$$\frac{\partial \log R^{\mathrm{K}}}{\partial \vartheta} = \mathbb{E}_W \left[\frac{\partial}{\partial \vartheta} \log \left\{ \prod_{k\in\mathcal{B}} \overline{F}_{y_k}(\rho_k u) \right\} \Bigg|_{u=W} \right], \tag{A.5}$$

$$\frac{\partial \log R^{\mathrm{P}}}{\partial \vartheta} = \mathbb{E}_{\bar{W}} \left[\frac{\partial}{\partial \vartheta} \log \left\{ \Phi_s\left(\frac{\lambda u}{V}\right) \prod_{k\in\mathcal{B}} \overline{F}_{y_k}(\rho_k u) \right\} \Bigg|_{u=\bar{W}} \right], \tag{A.6}$$

where u is held fixed in the differentiation and W and $\bar{W}$ have the densities (19) and (21).

The idea behind Lemma A.2 is simple. If $R = \int_0^\infty f(u)\,du$ for a positive function f that depends on the parameters, then, formally, $\partial_\vartheta R = \int_0^\infty f(u)\,\partial_\vartheta \log f(u)\,du$. Dividing by R turns the right-hand side into an expectation under the density f/R, which is how W and $\bar{W}$ arise. The proof only needs to justify the interchange of differentiation and integration.

Proof of Lemma A.2: We give the proof for R^{P}; for R^{K}, replace Φ_s by one and c by $m + c$. By (17), $R^{\mathrm{P}} = \int_0^\infty f(u)\,du$ with the positive integrand

$$f(u) = \frac{u^c e^{-u}}{c!}\,\Phi_s(z) \prod_{k\in\mathcal{B}} \overline{F}_{y_k}(\rho_k u), \qquad z = \frac{\lambda u}{V}.$$

By the chain rule, the definition of ψ_s, and Lemma A.1(a),

$$\frac{\partial \log f(u)}{\partial \vartheta} = \psi_s(z)\frac{\partial \log z}{\partial \vartheta} + \sum_{k\in\mathcal{B}} h_{y_k}(\rho_k u)\frac{\partial \log \rho_k}{\partial \vartheta}, \tag{A.7}$$

so, by Lemma A.1 and (A.4), $|\partial \log f(u)/\partial\vartheta| \le z + \sum_{k\in\mathcal{B}} y_k$.

Fix a parameter value and a bounded open ball $\mathcal{N}$ around it, and let $\bar{\lambda} = \sup_{\mathcal{N}} \lambda < \infty$, so that $z \le \bar{\lambda} u$ on $\mathcal{N}$ because $V \ge 1$. Since $\Phi_s(z) \le e^{2\sqrt{z}}$ and $\overline{F}_{y_k} \le 1$, for every parameter value in $\mathcal{N}$ and every $u > 0$,

$$\left|\frac{\partial f(u)}{\partial \vartheta}\right| = f(u)\left|\frac{\partial \log f(u)}{\partial \vartheta}\right| \le \frac{u^c e^{-u}}{c!}\, e^{2\sqrt{\bar{\lambda} u}} \left(\bar{\lambda} u + \sum_{k\in\mathcal{B}} y_k\right) =: g(u).$$

Because $2\sqrt{\bar{\lambda} u} \le u/2 + 2\bar{\lambda}$, the function g is bounded by a multiple of $u^c(1+u)e^{-u/2}$ and is therefore integrable. By the mean value theorem, g also bounds the difference quotients of $f(u)$ in ϑ on $\mathcal{N}$, so, by dominated convergence,

$$\frac{\partial R^{\mathrm{P}}}{\partial \vartheta} = \int_0^\infty \frac{\partial f(u)}{\partial \vartheta}\,du = \int_0^\infty f(u)\,\frac{\partial \log f(u)}{\partial \vartheta}\,du,$$

which, again by dominated convergence, is continuous on $\mathcal{N}$ because $\partial f(u)/\partial\vartheta$ is continuous in the parameters. Dividing by $R^{\mathrm{P}} > 0$ and noting that f/R^{P} is the density (21) of $\bar{W}$ gives (A.6). For R^{K}, the dominating function $\left(\sum_{k\in\mathcal{B}} y_k\right) u^{m+c} e^{-u}/(m+c)!$ works for all parameter values. □

Proof of Theorem 3: By (11) and $\log\rho_k = v_k - \log V$, we have $\log C^{\mathrm{K}} = \text{constant} + \sum_{k\in\mathcal{C}} x_k v_k - (m+c)\log V$. By Lemmas A.1(a) and A.2, and since the derivatives in (A.4) do not depend on u,

$$\frac{\partial \log R^{\mathrm{K}}}{\partial v_j} = \sum_{k\in\mathcal{B}} \mathbb{E}_W\left[h_{y_k}(\rho_k W)\right] \frac{\partial \log \rho_k}{\partial v_j}. \tag{A.8}$$

If $j \in \mathcal{B}$, then $\log C^{\mathrm{K}}$ does not depend on v_j, and $\partial \log\rho_k/\partial v_j = \mathbb{I}_{\{j=k\}}$ for $k \in \mathcal{B}$, so $\partial \log L^{\mathrm{K}}/\partial v_j = \mathbb{E}_W[h_{y_j}(\rho_j W)]$, which is the first case of (20). If $j \in \mathcal{C}$, then $\partial \log C^{\mathrm{K}}/\partial v_j = x_j - (m+c)\rho_j$ and $\partial \log\rho_k/\partial v_j = -\rho_j$ for every $k \in \mathcal{B}$, so adding $\partial \log R^{\mathrm{K}}/\partial v_j = -\rho_j \sum_{k\in\mathcal{B}} \mathbb{E}_W[h_{y_k}(\rho_k W)]$ gives the second case. □

Because $h_y > 0$, the first case of (20) also shows that the log-likelihood of a single period is strictly increasing in v_j whenever product j sells out in that period, as noted in Section 5.1.

Proof of Theorem 4: The proof parallels that of Theorem 3. By (16), $\log C^{\mathrm{P}} = \text{constant} - e^{\eta} + s\eta + \sum_{k\in\mathcal{C}} x_k v_k - c\log V$. Substituting (A.7) into (A.6) gives, for any coordinate ϑ of $(\eta, \boldsymbol{v})$,

$$\frac{\partial \log R^{\mathrm{P}}}{\partial \vartheta} = \overline{m}\,\frac{\partial \log z}{\partial \vartheta} + \sum_{k\in\mathcal{B}} \mathbb{E}_{\bar{W}}\left[h_{y_k}(\rho_k \bar{W})\right] \frac{\partial \log \rho_k}{\partial \vartheta}, \tag{A.9}$$

where $\overline{m} = \mathbb{E}_{\bar{W}}[\psi_s(\lambda \bar{W}/V)]$. If $j \in \mathcal{B}$, neither C^{P} nor z depends on v_j, so $\partial \log L^{\mathrm{P}}/\partial v_j = \mathbb{E}_{\bar{W}}[h_{y_j}(\rho_j \bar{W})]$ as in the known-arrival setting. If $j \in \mathcal{C}$, then $\partial \log C^{\mathrm{P}}/\partial v_j = x_j - c\rho_j$ and $\partial \log z/\partial v_j = \partial \log\rho_k/\partial v_j = -\rho_j$ for every $k \in \mathcal{B}$, so

$$\frac{\partial \log L^{\mathrm{P}}}{\partial v_j} = x_j - \rho_j \left\{ c + \overline{m} + \sum_{k\in\mathcal{B}} \mathbb{E}_{\bar{W}}\left[h_{y_k}(\rho_k \bar{W})\right] \right\}.$$

Products outside $\mathcal{A}$ do not enter the likelihood. Finally, $\partial \log C^{\mathrm{P}}/\partial\eta = s - \lambda$, $\partial \log z/\partial\eta = 1$, and $\partial \log\rho_k/\partial\eta = 0$, so $\partial \log L^{\mathrm{P}}/\partial\eta = s - \lambda + \overline{m}$. □

We now turn to Theorem 5. Since $\log L^{\mathrm{K}} = \log C^{\mathrm{K}} + \log R^{\mathrm{K}}$, we establish joint strict concavity by considering the two terms separately. The explicit term $\log C^{\mathrm{K}}$ is concave and becomes strictly concave when no product sells out, in which case $R^{\mathrm{K}} = 1$. When at least one product sells out, we establish strict concavity of $\log R^{\mathrm{K}}$ in two steps: we first show that its integrand is jointly strictly log-concave in the active log attractions and the integration variable u, and then apply the following lemma to show that integrating out u preserves strict log-concavity.

LEMMA A.3 **(Integration preserves log-concavity).** *Let $f : \mathbb{R}^J \times (0,\infty) \to (0,\infty)$ be continuous and jointly log-concave, and suppose that $F(\boldsymbol{v}) = \int_0^\infty f(\boldsymbol{v}, u)\,du < \infty$ for every $\boldsymbol{v} \in \mathbb{R}^J$. Then $\log F$ is concave on $\mathbb{R}^J$. If f is jointly strictly log-concave, then $\log F$ is strictly concave.*

Proof of Lemma A.3: Fix $\boldsymbol{v}_0, \boldsymbol{v}_1 \in \mathbb{R}^J$ and $t \in (0,1)$, and let $\boldsymbol{v}_t = (1-t)\boldsymbol{v}_0 + t\boldsymbol{v}_1$; we show that $F(\boldsymbol{v}_t) \geq F(\boldsymbol{v}_0)^{1-t} F(\boldsymbol{v}_1)^t$. The idea is to pair the points that split $F(\boldsymbol{v}_0)$ and $F(\boldsymbol{v}_1)$ at the same fraction. For $i \in \{0,1\}$, the function $K_i(w) = \int_0^w f(\boldsymbol{v}_i, u)\,du$ is continuously differentiable on $(0,\infty)$ with $K_i'(w) = f(\boldsymbol{v}_i, w) > 0$ and

increases from 0 to $F(\boldsymbol{v}_i)$. Hence, for each $a \in (0,1)$, there is a unique $u_i(a) > 0$ with $K_i(u_i(a)) = aF(\boldsymbol{v}_i)$, and, by the inverse function theorem, u_i is a continuously differentiable increasing bijection from $(0,1)$ onto $(0,\infty)$ with

$$f\big(\boldsymbol{v}_i, u_i(a)\big)\, u_i'(a) = F(\boldsymbol{v}_i), \qquad i = 0,1. \tag{A.10}$$

The convex combination $u_t = (1-t)u_0 + tu_1$ is again such a bijection, so the change of variables $u = u_t(a)$ gives

$$\begin{aligned} F(\boldsymbol{v}_t) &= \int_0^1 f\big(\boldsymbol{v}_t, u_t(a)\big)\, u_t'(a)\, da \\ &\geq \int_0^1 f\big(\boldsymbol{v}_0, u_0(a)\big)^{1-t} f\big(\boldsymbol{v}_1, u_1(a)\big)^t\, u_0'(a)^{1-t} u_1'(a)^t\, da \\ &= \int_0^1 F(\boldsymbol{v}_0)^{1-t} F(\boldsymbol{v}_1)^t\, da = F(\boldsymbol{v}_0)^{1-t} F(\boldsymbol{v}_1)^t. \end{aligned}$$

The inequality uses the log-concavity of f at $(\boldsymbol{v}_t, u_t(a)) = (1-t)(\boldsymbol{v}_0, u_0(a)) + t(\boldsymbol{v}_1, u_1(a))$ and the weighted arithmetic–geometric mean inequality $u_t'(a) \geq u_0'(a)^{1-t} u_1'(a)^t$; the second equality follows from (A.10). If f is jointly strictly log-concave and $\boldsymbol{v}_0 \neq \boldsymbol{v}_1$, the log-concavity inequality is strict for every $a \in (0,1)$. Hence $F(\boldsymbol{v}_t) > F(\boldsymbol{v}_0)^{1-t} F(\boldsymbol{v}_1)^t$. □

Proof of Theorem 5: We work with the active log attractions $(v_j)_{j\in\mathcal{A}}$, since products outside $\mathcal{A}$ do not enter the likelihood. By (11),

$$\log C^{\mathrm{K}} = \text{constant} + \sum_{j\in\mathcal{C}} x_j v_j - (m+c)\log\left(1 + \sum_{j\in\mathcal{C}} e^{v_j}\right),$$

which is concave because the log-sum-exp function is convex and $m + c \geq 0$.

If $\mathcal{B} = \varnothing$, then $\mathcal{C} = \mathcal{A}$, $m + c = T \geq 1$, and $R^{\mathrm{K}} = 1$. Recall that $\rho_j = \theta_j / V = e^{v_j}/(1 + \sum_{k\in\mathcal{A}} e^{v_k})$ in this case. The Hessian entries of $\log C^{\mathrm{K}}$ are

$$\frac{\partial^2 \log C^{\mathrm{K}}}{\partial v_j \partial v_k} = -T\left(\mathbb{I}_{\{}j = k\}\rho_j - \rho_j\rho_k\right), \qquad j,k \in \mathcal{A}.$$

Since $\rho_j > 0$ and $\sum_{k\in\mathcal{A}} \rho_k < 1$, the diagonal entries are negative and satisfy

$$-\frac{\partial^2 \log C^{\mathrm{K}}}{\partial v_j^2} - \sum_{k\in\mathcal{A}\setminus\{j\}} \left|\frac{\partial^2 \log C^{\mathrm{K}}}{\partial v_j \partial v_k}\right| = T\rho_j\left(1 - \sum_{k\in\mathcal{A}} \rho_k\right) > 0.$$

Thus the Hessian is symmetric, strictly diagonally dominant, and negative definite. Hence $\log L^{\mathrm{K}} = \log C^{\mathrm{K}}$ is jointly strictly concave.

Then we consider $\mathcal{B} \neq \varnothing$. By (12), $R^{\mathrm{K}}(\boldsymbol{v}) = \int_0^\infty f(\boldsymbol{v}, u)\, du \leq 1$ with the positive and continuous integrand

$$f(\boldsymbol{v}, u) = \frac{u^{m+c} e^{-u}}{(m+c)!} \prod_{j\in\mathcal{B}} \overline{F}_{y_j}\big(\rho_j(\boldsymbol{v})u\big),$$

so, by Lemma A.3, it suffices to show that f is jointly strictly log-concave in $(\boldsymbol{v}, u)$. The key property is that the probability that a Poisson random variable reaches a given level is log-concave in the logarithm of its mean: for $y \geq 1$, the function $q_y(w) = \log \overline{F}_y(e^w)$ is strictly increasing and strictly concave. To see this,

let $\mu = e^w$. By Lemma A.1(a), $q_y'(w) = h_y(\mu) > 0$, and differentiating $\log h_y(\mu) = \log y + \log \pi_y(\mu) - \log \overline{F}_y(\mu)$ with respect to w, using $d\log \pi_y(\mu)/d\log \mu = y - \mu$, gives

$$q_y''(w) = h_y(\mu)\big\{y - \mu - h_y(\mu)\big\}.$$

Hence q_y is strictly concave if $h_y(\mu) > y - \mu$, which holds immediately when $\mu \geq y$. When $0 < \mu < y$,

$$\frac{\overline{F}_y(\mu)}{\pi_y(\mu)} = \sum_{\ell=0}^{\infty} \frac{\mu^\ell}{(y+1)(y+2)\cdots(y+\ell)} < \sum_{\ell=0}^{\infty} \left(\frac{\mu}{y}\right)^\ell = \frac{y}{y-\mu},$$

so $h_y(\mu) = y\pi_y(\mu)/\overline{F}_y(\mu) > y - \mu$.

Now,

$$\log f(\boldsymbol{v}, u) = \text{constant} + (m+c)\log u - u + \sum_{j \in \mathcal{B}} q_{y_j}\big(\log \rho_j(\boldsymbol{v}) + \log u\big).$$

Fix two distinct points $(\boldsymbol{v}_0, u_0)$ and $(\boldsymbol{v}_1, u_1)$, and let $(\boldsymbol{v}_t, u_t) = (1-t)(\boldsymbol{v}_0, u_0) + t(\boldsymbol{v}_1, u_1)$ for $t \in (0,1)$. For each $j \in \mathcal{B}$, joint concavity of $\log \rho_j(\boldsymbol{v}) + \log u$ and strict monotonicity and concavity of q_{y_j} give

$$\begin{aligned}
&q_{y_j}\big(\log \rho_j(\boldsymbol{v}_t) + \log u_t\big) \\
&\quad \geq q_{y_j}\big((1-t)(\log \rho_j(\boldsymbol{v}_0) + \log u_0) + t(\log \rho_j(\boldsymbol{v}_1) + \log u_1)\big) \\
&\quad \geq (1-t) q_{y_j}\big(\log \rho_j(\boldsymbol{v}_0) + \log u_0\big) + t q_{y_j}\big(\log \rho_j(\boldsymbol{v}_1) + \log u_1\big).
\end{aligned}$$

The first inequality can hold with equality only if $u_0 = u_1$ and $\boldsymbol{v}_0, \boldsymbol{v}_1$ agree on $\mathcal{C}$, because $\log u$ and the negative log-sum-exp term are strictly concave. In that case, since the two points are distinct, they must differ in v_j for some $j \in \mathcal{B}$, making the second inequality strict by the strict concavity of q_{y_j}.

Since $(m+c)\log u - u$ is concave, it follows that

$$\log f(\boldsymbol{v}_t, u_t) > (1-t)\log f(\boldsymbol{v}_0, u_0) + t \log f(\boldsymbol{v}_1, u_1).$$

Thus $\log f$ is jointly strictly concave. By Lemma A.3, $\log R^{\mathrm{K}}$ is strictly concave, and adding the concave term $\log C^{\mathrm{K}}$ proves that $\log L^{\mathrm{K}}$ is jointly strictly concave. □

We then prove the Poisson-arrival results in Section 5.2. The sales-dynamics and monotonicity lemmas support both the identifiability result and the local strong concavity result. For these arguments, we use the parameters $(\boldsymbol{\alpha}, \delta)$ defined in (23), with $\lambda = \alpha_+ + \delta$ and $\theta_j = \alpha_j/\delta$. The identification arguments concern a fixed positive inventory vector; the local concavity result also allows periods in which some products are not initially stocked.

LEMMA A.4 **(Sales dynamics under Poisson arrivals)**. *Represent a selling period by the interval $[0,1]$. Under Poisson arrivals, the terminal sales have the distribution of the sales by time one in the continuous-time Markov chain in which, whenever the set A of products in stock is nonempty, the next sale occurs at rate*

$$\Lambda_A(\delta) = \alpha_A \frac{\alpha_+ + \delta}{\alpha_A + \delta}, \qquad \textit{where } \alpha_A = \sum_{k \in A} \alpha_k, \tag{A.11}$$

and is of product $j \in A$ with probability α_j/α_A; sales stop once $A = \varnothing$. Moreover, $\Lambda_{[J]}(\delta) = \alpha_+$, and $\Lambda_A(\delta)$ is strictly decreasing in δ for every nonempty proper subset A of $[J]$.

Proof of Lemma A.4: Let customers arrive on $[0,1]$ according to a Poisson process with rate λ and choose according to (1); the number of customers is then $\mathrm{Poi}(\lambda)$, and given this number, their choices are as in Section 2. While the set of products in stock is A, each arriving customer buys product $j \in A$ with probability $\theta_j/(1+\sum_{k\in A}\theta_k)$, independently of the past, so, by Poisson thinning, product j sells at rate

$$\frac{\lambda\theta_j}{1+\sum_{k\in A}\theta_k} = \frac{(\alpha_+ + \delta)\,\alpha_j/\delta}{1+\alpha_A/\delta} = \alpha_j \frac{\alpha_+ + \delta}{\alpha_A + \delta},$$

where we used $\lambda = \alpha_+ + \delta$ and $\theta_k = \alpha_k/\delta$. Summing over $j \in A$ gives the total rate $\Lambda_A(\delta)$, and the next sale is of product j with probability α_j/α_A. Finally, $\Lambda_{[J]}(\delta) = \alpha_+$, and $\Lambda_A'(\delta) = \alpha_A(\alpha_A - \alpha_+)/(\alpha_A+\delta)^2 < 0$ whenever $0 < \alpha_A < \alpha_+$, that is, whenever A is a nonempty proper subset of $[J]$. □

Lemma A.4 clarifies the role of δ. Before the first stockout, product j sells at rate α_j, which does not depend on δ. After a stockout, δ determines how much of the demand for the sold-out products is redirected to the remaining products: as $\delta \to 0$, $\Lambda_A(\delta) \to \alpha_+$, so all of this demand is redirected, whereas as $\delta \to \infty$, $\Lambda_A(\delta) \to \alpha_A$, so none of it is. Hence information about δ comes only from the post-stockout sales dynamics, including the absence of further sales. In particular, when $J = 1$, nothing remains in stock after the stockout, so the product sells at rate α_1 until it sells out, and $X_1 \overset{d}{=} \min\{y_1, Z\}$ with $Z \sim \mathrm{Poi}(\alpha_1)$. This distribution depends on (λ, θ_1) only through $\alpha_1 = \lambda\theta_1/(1+\theta_1)$ and not on δ, which proves the claim for $J = 1$ in Section 5.2. For $J \geq 2$, the next lemma shows that, with $\boldsymbol{\alpha}$ fixed, a larger δ leads to fewer sales.

LEMMA A.5 **(Monotonicity in δ).** *Fix $\boldsymbol{\alpha} \in \mathbb{R}^J_{++}$ and $\boldsymbol{y} \geq \mathbf{1}$, let $\boldsymbol{X}(\delta)$ denote the terminal sales under $(\boldsymbol{\alpha}, \delta)$, and let $\mu_{\boldsymbol{y}}(\boldsymbol{\alpha}, \delta) = \mathbb{E}[X_+(\delta)]$ be the expected total sales, where $X_+(\delta) = \sum_j X_j(\delta)$.*

(a) For $0 < \delta_1 < \delta_2$, $\boldsymbol{X}(\delta_1)$ and $\boldsymbol{X}(\delta_2)$ can be coupled so that $\boldsymbol{X}(\delta_1) \geq \boldsymbol{X}(\delta_2)$ componentwise.

(b) If $J \geq 2$, then $\mu_{\boldsymbol{y}}(\boldsymbol{\alpha}, \delta)$ is continuously differentiable and $\partial_\delta \mu_{\boldsymbol{y}}(\boldsymbol{\alpha}, \delta) < 0$ for every $\delta > 0$.

Proof of Lemma A.5: For part (a), we construct the chain of Lemma A.4 for all δ simultaneously. Let $y_+ = \sum_j y_j$, and draw the products sold, $j_1, \ldots, j_{y_+}$, by choosing $j_{h+1} = j$ with probability α_j/α_{A_h} for $j \in A_h$, where A_h is the set of products still in stock after the first h sales. Independently, draw $\xi_1, \ldots, \xi_{y_+} \overset{\text{iid}}{\sim} \mathrm{Exp}(1)$, and, for each $\delta > 0$, let the qth sale occur at time

$$\tau_q(\delta) = \sum_{h=0}^{q-1} \frac{\xi_{h+1}}{\Lambda_{A_h}(\delta)}, \qquad q = 1, \ldots, y_+. \tag{A.12}$$

By Lemma A.4, the sales by time one, $X_j(\delta) = \sum_q \mathbb{I}_{\{j_q = j,\ \tau_q(\delta) \leq 1\}}$, have the distribution of the terminal sales under $(\boldsymbol{\alpha}, \delta)$. Only the sale times depend on δ, and each $\tau_q(\delta)$ is nondecreasing in δ because each $\Lambda_{A_h}(\delta)$ is nonincreasing. Since the sales completed by time one are the first $X_+(\delta)$ sales of the common sequence, it follows that $X_+(\delta_2) \leq X_+(\delta_1)$ and $\boldsymbol{X}(\delta_2) \leq \boldsymbol{X}(\delta_1)$.

For part (b), let $J \geq 2$ and $0 < \delta_1 < \delta_2$, and fix a product k. On the event $\mathcal{E}_k = \{j_1 = \cdots = j_{y_k} = k\}$, product k sells out at the y_kth sale while the other products remain in stock, so $\tau_{y_k} = \sum_{h=1}^{y_k} \xi_h/\alpha_+$ does not depend on δ, and the (y_k+1)st sale, which exists because $J \geq 2$, follows after the time $\xi_{y_k+1}/\Lambda_{-k}(\delta)$, where

$$\Lambda_{-k}(\delta) := \Lambda_{[J]\setminus\{k\}}(\delta) = (\alpha_+ - \alpha_k)\frac{\alpha_+ + \delta}{\alpha_+ - \alpha_k + \delta}$$

is strictly decreasing in δ. The nonnegative integer $X_+(\delta_1) - X_+(\delta_2)$ is positive on $\mathcal{E}_k \cap \{\tau_{y_k+1}(\delta_1) \le 1 < \tau_{y_k+1}(\delta_2)\}$, and $\tau_{y_k+1}(\delta_1) \le \tau_{y_k+1}(\delta_2)$, so

$$\mu_{\boldsymbol{y}}(\boldsymbol{\alpha}, \delta_1) - \mu_{\boldsymbol{y}}(\boldsymbol{\alpha}, \delta_2) \ge \mathbb{P}\big(\mathcal{E}_k,\ \tau_{y_k+1}(\delta_1) \le 1\big) - \mathbb{P}\big(\mathcal{E}_k,\ \tau_{y_k+1}(\delta_2) \le 1\big). \tag{A.13}$$

The event $\mathcal{E}_k$ has probability $(\alpha_k/\alpha_+)^{y_k}$ and is independent of (ξ_h), and τ_{y_k} has density $\alpha_+^{y_k} u^{y_k-1} e^{-\alpha_+ u}/(y_k - 1)!$, so conditioning on $\tau_{y_k} = u$ gives

$$\mathbb{P}\big(\mathcal{E}_k,\ \tau_{y_k+1}(\delta) \le 1\big) = \frac{\alpha_k^{y_k}}{(y_k-1)!} \int_0^1 u^{y_k-1} e^{-\alpha_+ u} \Big\{1 - e^{-\Lambda_{-k}(\delta)(1-u)}\Big\}\, du, \tag{A.14}$$

whose derivative with respect to δ is

$$\frac{\alpha_k^{y_k} \Lambda'_{-k}(\delta)}{(y_k-1)!} \int_0^1 u^{y_k-1}(1-u) e^{-\alpha_+ u - \Lambda_{-k}(\delta)(1-u)}\, du < 0.$$

The finite-state sales process has transition rates that are smooth in $\delta > 0$, so $\mu_{\boldsymbol{y}}$ is continuously differentiable. Dividing (A.13) by $\delta_2 - \delta_1$ and letting $\delta_2 \downarrow \delta_1 = \delta$ gives

$$\partial_\delta \mu_{\boldsymbol{y}}(\boldsymbol{\alpha}, \delta) \le \frac{\partial}{\partial \delta} \mathbb{P}\big(\mathcal{E}_k, \tau_{y_k+1}(\delta) \le 1\big) < 0.$$

This proves part (b). □

We are now ready to prove Theorem 6. The proof first identifies α_+, as well as α_j for every product with $y_j \ge 2$, from the probabilities of selling nothing and of selling exactly one unit. It then identifies δ in one of two ways. If every product has $y_j \ge 2$, we use Lemma A.5. Otherwise, some product has $y_j = 1$; selling one unit of it causes a stockout, so its one-sale probability depends on δ.

Proof of Theorem 6: Let $p_{\boldsymbol{0}} = \mathbb{P}(\boldsymbol{X} = \boldsymbol{0} \mid \boldsymbol{y})$ and $p_j = \mathbb{P}(\boldsymbol{X} = \boldsymbol{e}_j \mid \boldsymbol{y})$, where $\boldsymbol{e}_j$ is the jth unit vector. By Lemma A.4, while all products are in stock, sales occur at rate α_+ and each sale is of product j with probability α_j/α_+, so $p_{\boldsymbol{0}} = e^{-\alpha_+}$ identifies α_+. If $y_j \ge 2$, then $\{\boldsymbol{X} = \boldsymbol{e}_j\}$ is the event that the first sale, at some time $u \in [0,1]$, is of product j and that no further sale occurs, at the unchanged rate α_+, during $(u,1]$; hence $p_j = \int_0^1 \alpha_j e^{-\alpha_+ u} e^{-\alpha_+(1-u)}\, du = \alpha_j e^{-\alpha_+}$, which identifies α_j.

If $y_j \ge 2$ for every product, then $\boldsymbol{\alpha}$ is identified, and so is δ, because the expected total sales $\mu_{\boldsymbol{y}}(\boldsymbol{\alpha}, \delta)$ are strictly decreasing in δ by Lemma A.5(b).

Otherwise, consider a product j with $y_j = 1$, and write $a = \alpha_j$. Once product j is sold, it is sold out, and the other products sell at rate $\Lambda_{-j} = (\alpha_+ - a)(\alpha_+ + \delta)/(\alpha_+ - a + \delta)$, so, as above,

$$p_j = \varphi(a, \delta) := a \int_0^1 e^{-\alpha_+ u}\, e^{-\Lambda_{-j}(1-u)}\, du, \tag{A.15}$$

where, given α_+, the function φ is the same for every product with unit inventory. Since $0 < a < \alpha_+$ because $J \ge 2$,

$$\frac{\partial \Lambda_{-j}}{\partial a} = -\frac{\delta(\alpha_+ + \delta)}{(\alpha_+ - a + \delta)^2} < 0, \qquad \frac{\partial \Lambda_{-j}}{\partial \delta} = -\frac{a(\alpha_+ - a)}{(\alpha_+ - a + \delta)^2} < 0,$$

so φ is strictly increasing in both a and δ; intuitively, after product j sells out, the probability that nothing else is sold is larger when fewer customers substitute, that is, when δ is larger. Now suppose that $(\boldsymbol{\alpha}', \delta')$ induces the same distribution of $\boldsymbol{X}$ as $(\boldsymbol{\alpha}, \delta)$, so that $\alpha'_+ = \alpha_+$ and $\alpha'_k = \alpha_k$ whenever $y_k \ge 2$. If $\delta' > \delta$,

then, for every product j with $y_j = 1$, $\varphi(\alpha'_j, \delta') = p_j = \varphi(\alpha_j, \delta) < \varphi(\alpha_j, \delta')$, so $\alpha'_j < \alpha_j$, and summing over all products gives $\alpha'_+ < \alpha_+$, a contradiction. By symmetry, $\delta' < \delta$ is also impossible, so $\delta' = \delta$, and then $\alpha'_j = \alpha_j$ for every product with $y_j = 1$. In both cases, $(\boldsymbol{\alpha}, \delta)$, and hence $(\lambda, \boldsymbol{v})$, is identified. □

We next turn to Theorem 7. Its proof rests on the curvature of the expected log-likelihood of a single period at the true parameter, which we establish first. Recall that $\boldsymbol{\beta} = (\eta, \boldsymbol{v})$ with $\eta = \log \lambda$ and that $\boldsymbol{\beta}_0$ is its true value. We write $L^{\mathrm{P}}(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{x})$ for the likelihood (13) evaluated at $(\lambda, \boldsymbol{v}) = (e^{\eta}, \boldsymbol{v})$. For a fixed initial inventory $\boldsymbol{y} \in \mathbb{Z}^J_+$, let $p_{\boldsymbol{\beta}}(\boldsymbol{x} \mid \boldsymbol{y}) = L^{\mathrm{P}}(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{x})$ denote the probability of the terminal sales $\boldsymbol{x}$, and define the expected log-likelihood

$$\mathscr{L}^{\mathrm{P}}_{\boldsymbol{y}}(\boldsymbol{\beta}) = \mathbb{E}_{\boldsymbol{\beta}_0}\left[\log L^{\mathrm{P}}(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{X}) \,\middle|\, \boldsymbol{y}\right] = \sum_{\boldsymbol{x}} p_{\boldsymbol{\beta}_0}(\boldsymbol{x} \mid \boldsymbol{y}) \log p_{\boldsymbol{\beta}}(\boldsymbol{x} \mid \boldsymbol{y}),$$

where the sum is over the finitely many feasible sales vectors, that is, those with $0 \le x_j \le y_j$ for every product j. Each feasible $\boldsymbol{x}$ has positive probability; for example, exactly $\sum_j x_j$ customers may arrive and buy these units. Moreover, $p_{\boldsymbol{\beta}}(\boldsymbol{x} \mid \boldsymbol{y})$ is infinitely differentiable in $\boldsymbol{\beta}$. Indeed, by Lemma A.4 and its proof, it is the time-one transition probability from $\mathbf{0}$ to $\boldsymbol{x}$ of the continuous-time Markov chain on the feasible sales vectors that moves from $\boldsymbol{x}'$ to $\boldsymbol{x}' + \boldsymbol{e}_j$ at rate $\lambda\theta_j / (1 + \sum_{k : x'_k < y_k} \theta_k)$ whenever $x'_j < y_j$, where $\boldsymbol{e}_j$ is the jth unit vector. It is therefore an entry of the exponential of a finite generator matrix whose entries are smooth in $\boldsymbol{\beta}$, and the matrix exponential is a real-analytic function of the matrix entries.

Lemma A.6 **(Expected log-likelihood curvature).** *(a) For every $\boldsymbol{y} \in \mathbb{Z}^J_+$, $\nabla^2 \mathscr{L}^{\mathrm{P}}_{\boldsymbol{y}}(\boldsymbol{\beta}_0) \preceq 0$.*
(b) If $J \ge 2$ and $\boldsymbol{y} \in \mathbb{Z}^J_{++}$, then $\nabla^2 \mathscr{L}^{\mathrm{P}}_{\boldsymbol{y}}(\boldsymbol{\beta}_0) \prec 0$.

In statistical terms, $-\nabla^2 \mathscr{L}^{\mathrm{P}}_{\boldsymbol{y}}(\boldsymbol{\beta}_0)$ is the Fisher information of a period with initial inventory $\boldsymbol{y}$, and part (b) states that it is nonsingular. To prove the lemma, we express $\boldsymbol{d}^\top \nabla^2 \mathscr{L}^{\mathrm{P}}_{\boldsymbol{y}}(\boldsymbol{\beta}_0) \boldsymbol{d}$ as a negative weighted sum of squared directional derivatives of terminal-sales probabilities. We then apply the argument for Theorem 6 to first order to show that, for every $\boldsymbol{d} \neq \mathbf{0}$, at least one of these derivatives is nonzero at $\boldsymbol{\beta}_0$.

Proof of Lemma A.6: We first prove part (a). Fix $\boldsymbol{y} \in \mathbb{Z}^J_+$. Since the sum defining $\mathscr{L}^{\mathrm{P}}_{\boldsymbol{y}}$ is finite and its terms are smooth, it can be differentiated term by term, with the weights $p_{\boldsymbol{\beta}_0}(\boldsymbol{x} \mid \boldsymbol{y})$ held fixed. Hence, for every direction $\boldsymbol{d} \in \mathbb{R}^{J+1}$,

$$\begin{aligned} \boldsymbol{d}^\top \nabla^2 \mathscr{L}^{\mathrm{P}}_{\boldsymbol{y}}(\boldsymbol{\beta}_0) \boldsymbol{d} &= \sum_{\boldsymbol{x}} \boldsymbol{d}^\top \nabla^2 p_{\boldsymbol{\beta}_0}(\boldsymbol{x} \mid \boldsymbol{y}) \boldsymbol{d} - \sum_{\boldsymbol{x}} \frac{\{\boldsymbol{d}^\top \nabla p_{\boldsymbol{\beta}_0}(\boldsymbol{x} \mid \boldsymbol{y})\}^2}{p_{\boldsymbol{\beta}_0}(\boldsymbol{x} \mid \boldsymbol{y})} \\ &= -\sum_{\boldsymbol{x}} \frac{\{\boldsymbol{d}^\top \nabla p_{\boldsymbol{\beta}_0}(\boldsymbol{x} \mid \boldsymbol{y})\}^2}{p_{\boldsymbol{\beta}_0}(\boldsymbol{x} \mid \boldsymbol{y})} \le 0. \end{aligned}$$

The first equality evaluates $\nabla^2 \log p = \nabla^2 p / p - \nabla p \nabla p^\top / p^2$ at $\boldsymbol{\beta}_0$, where the weight $p_{\boldsymbol{\beta}_0}(\boldsymbol{x} \mid \boldsymbol{y})$ cancels one factor of p. The second equality holds because $\sum_{\boldsymbol{x}} \nabla^2 p_{\boldsymbol{\beta}}(\boldsymbol{x} \mid \boldsymbol{y}) = \mathbf{0}$, which follows by differentiating $\sum_{\boldsymbol{x}} p_{\boldsymbol{\beta}}(\boldsymbol{x} \mid \boldsymbol{y}) = 1$ twice. This proves part (a).

We next prove part (b), so let $J \ge 2$ and $\boldsymbol{y} \in \mathbb{Z}^J_{++}$. We express the zero- and one-sale probabilities through the pre-stockout purchase rates α_j, their sum α_+ and the no-purchase rate δ in (23), which satisfy $\lambda = \alpha_+ + \delta$ and $\theta_j = \alpha_j / \delta$.

Fix a nonzero direction $\boldsymbol{d}$ and suppose, for a contradiction, that $\boldsymbol{d}^\top \nabla^2 \mathscr{L}_{\boldsymbol{y}}^{\mathrm{P}}(\boldsymbol{\beta}_0)\boldsymbol{d} = 0$. Every denominator in the curvature identity is positive, so every squared term must be zero. Consider the path $\boldsymbol{\beta}(t) = \boldsymbol{\beta}_0 + t\boldsymbol{d}$, and use a dot to denote differentiation at $t = 0$. The zero-curvature assumption therefore gives

$$\dot{p}_{\boldsymbol{x}} = \frac{d}{dt} p_{\boldsymbol{\beta}(t)}(\boldsymbol{x} \mid \boldsymbol{y})\Big|_{t=0} = \boldsymbol{d}^\top \nabla p_{\boldsymbol{\beta}_0}(\boldsymbol{x} \mid \boldsymbol{y}) = 0$$

for every feasible $\boldsymbol{x}$. The quantities $\dot{\alpha}_j$, $\dot{\delta}$, and $\dot{\alpha}_+$ denote derivatives along this same path. We will show that they must all be zero, and then use the inverse parameter relations to obtain a contradiction.

The zero-sale outcome is feasible, so its probability $p_{\boldsymbol{0}}$ also has zero derivative. All products stay in stock until the first sale, which occurs at rate α_+, so $p_{\boldsymbol{0}} = e^{-\alpha_+}$, and hence

$$0 = \dot{p}_{\boldsymbol{0}} = -e^{-\alpha_+}\dot{\alpha}_+.$$

The exponential factor is positive, so $\dot{\alpha}_+ = 0$. Next, let p_j denote the probability of selling exactly one unit of product j and no other products. When $y_j \geq 2$, this event causes no stockout, and $p_j = \alpha_j e^{-\alpha_+}$. Therefore,

$$0 = \dot{p}_j = e^{-\alpha_+}(\dot{\alpha}_j - \alpha_j \dot{\alpha}_+) = e^{-\alpha_+}\dot{\alpha}_j,$$

which gives $\dot{\alpha}_j = 0$ for every product with at least two units of inventory.

If every product has at least two units, all purchase-rate derivatives are now zero. It remains to determine $\dot{\delta}$. Expected total sales are given by $\mu_{\boldsymbol{y}}(\boldsymbol{\alpha}, \delta) = \sum_{\boldsymbol{x}} (\sum_j x_j) p_{\boldsymbol{x}}$. Since every $\dot{p}_{\boldsymbol{x}} = 0$, this expectation also has zero derivative. Moreover, $\mu_{\boldsymbol{y}}$ is smooth in $(\boldsymbol{\alpha}, \delta)$, because each $p_{\boldsymbol{x}}$ is smooth in $\boldsymbol{\beta}$ and $\boldsymbol{\beta}$ is a smooth function of $(\boldsymbol{\alpha}, \delta)$. Since all $\dot{\alpha}_j = 0$, the chain rule gives

$$0 = \sum_{\boldsymbol{x}} \left(\sum_j x_j \right) \dot{p}_{\boldsymbol{x}} = \dot{\mu}_{\boldsymbol{y}} = \partial_\delta \mu_{\boldsymbol{y}}\, \dot{\delta}.$$

Lemma A.5(b) gives $\partial_\delta \mu_{\boldsymbol{y}} < 0$, and therefore $\dot{\delta} = 0$.

Suppose instead that at least one product has unit inventory. For such a product, its one-sale event depletes its inventory, so the probability also depends on the purchase rates after that stockout. Equation (A.15) writes this probability as $p_j = \varphi(\alpha_j, \delta)$ when α_+ is held fixed. Differentiating (A.15) under the integral sign, which is justified because the integrand is smooth on the compact interval $[0, 1]$, gives the partial derivatives with respect to the first and second arguments,

$$\varphi_a(a, \delta) = \int_0^1 e^{-\alpha_+ u - \Lambda_{-j}(1-u)} \left\{ 1 - a(1-u) \frac{\partial \Lambda_{-j}}{\partial a} \right\} du,$$
$$\varphi_\delta(a, \delta) = -a \frac{\partial \Lambda_{-j}}{\partial \delta} \int_0^1 (1-u)\, e^{-\alpha_+ u - \Lambda_{-j}(1-u)}\, du.$$

Both are strictly positive, because $\partial \Lambda_{-j}/\partial a < 0$ and $\partial \Lambda_{-j}/\partial \delta < 0$ by the proof of Theorem 6. Along our parameter path, $\dot{\alpha}_+ = 0$, so changes in α_+ contribute no term to the first derivative of p_j. Consequently,

$$0 = \dot{p}_j = \varphi_a(\alpha_j, \delta)\dot{\alpha}_j + \varphi_\delta(\alpha_j, \delta)\dot{\delta}.$$

If $\dot{\delta} > 0$, this equality requires $\dot{\alpha}_j < 0$ for every unit-inventory product. If $\dot{\delta} < 0$, it requires $\dot{\alpha}_j > 0$ for every such product. Both possibilities contradict $\dot{\alpha}_+ = \sum_j \dot{\alpha}_j = 0$, because the purchase-rate derivatives of all other

products are already zero. Thus $\dot{\delta} = 0$, and the displayed equality then gives $\dot{\alpha}_j = 0$ for each unit-inventory product.

In both cases, $\dot{\alpha}_j = 0$ for every product and $\dot{\delta} = 0$. To return to the original parameters, use

$$\eta = \log(\alpha_+ + \delta), \qquad v_j = \log \alpha_j - \log \delta.$$

Differentiating along the same path gives

$$\dot{\eta} = \frac{\dot{\alpha}_+ + \dot{\delta}}{\alpha_+ + \delta} = 0, \qquad \dot{v}_j = \frac{\dot{\alpha}_j}{\alpha_j} - \frac{\dot{\delta}}{\delta} = 0, \quad j \in [J].$$

Thus $\dot{\boldsymbol{\beta}} = \mathbf{0}$. However, the definition $\boldsymbol{\beta}(t) = \boldsymbol{\beta}_0 + t\boldsymbol{d}$ gives $\dot{\boldsymbol{\beta}} = \boldsymbol{d} \neq \mathbf{0}$, a contradiction. The expected directional curvature is therefore strictly negative in every nonzero direction, which proves part (b). □

We prove Theorem 7 in three steps. First, we use the preceding lemma and the positive frequency of periods in which all products are initially stocked to establish a uniform negative bound on the average expected Hessian. Second, we show that the sample Hessian at the true parameter concentrates around its expectation. Finally, continuity of the Hessians extends this negative curvature to a fixed neighborhood of the true parameter.

Proof of Theorem 7: For period n, write $L_n^{\mathrm{P}}(\boldsymbol{\beta}) = L^{\mathrm{P}}(\boldsymbol{\beta} \mid \boldsymbol{y}_n, \boldsymbol{x}_n)$, so that the average log-likelihood in the theorem is $\mathcal{L}_N^{\mathrm{P}}(\boldsymbol{\beta}) = N^{-1} \sum_{n=1}^{N} \log L_n^{\mathrm{P}}(\boldsymbol{\beta})$. Since the second derivative of $\mathcal{L}_N^{\mathrm{P}}$ at $\boldsymbol{\beta}$ along a unit direction $\boldsymbol{d}$ is $\boldsymbol{d}^\top \nabla^2 \mathcal{L}_N^{\mathrm{P}}(\boldsymbol{\beta}) \boldsymbol{d}$, we need to find $r, \gamma > 0$, independent of N, such that, with probability approaching one, $\nabla^2 \mathcal{L}_N^{\mathrm{P}}(\boldsymbol{\beta}) \preceq -\gamma I_{J+1}$ at every point of the ball $\{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|_2 \leq r\}$. We first obtain a negative bound for the expected Hessian at $\boldsymbol{\beta}_0$. Let $\mathcal{Y}$ denote the finite support of the initial inventories, and write $H_{\boldsymbol{y}} = \nabla^2 \mathscr{L}_{\boldsymbol{y}}^{\mathrm{P}}(\boldsymbol{\beta}_0)$, which equals $\mathbb{E}_{\boldsymbol{\beta}_0}[\nabla^2_{\boldsymbol{\beta}} \log L^{\mathrm{P}}(\boldsymbol{\beta}_0 \mid \boldsymbol{y}, \boldsymbol{X}) \mid \boldsymbol{y}]$ because the expectation is a finite sum. By Lemma A.6, $H_{\boldsymbol{y}} \preceq 0$ for every $\boldsymbol{y}$, and, since $J \geq 2$, $H_{\boldsymbol{y}} \prec 0$ whenever $\boldsymbol{y} \in \mathbb{Z}_{++}^J$. The assumed positive proportion of periods stocking every product ensures that $\mathcal{Y} \cap \mathbb{Z}_{++}^J$ is nonempty. Since this set is finite and each $-H_{\boldsymbol{y}}$ with $\boldsymbol{y} \in \mathcal{Y} \cap \mathbb{Z}_{++}^J$ is positive definite, the smallest eigenvalue among these matrices, denoted by γ_0, is positive. Thus $H_{\boldsymbol{y}} \preceq -\gamma_0 I_{J+1}$ for every $\boldsymbol{y} \in \mathcal{Y} \cap \mathbb{Z}_{++}^J$.

By assumption, there exists $\epsilon_0 > 0$ such that at least $\epsilon_0 N$ periods stock every product for all sufficiently large N. Set $\gamma = \epsilon_0 \gamma_0 / 2$. Then, for all sufficiently large N,

$$\begin{aligned}
\frac{1}{N} \sum_{n=1}^{N} H_{\boldsymbol{y}_n} &\preceq \frac{1}{N} \sum_{\substack{1 \leq n \leq N \\ \boldsymbol{y}_n \in \mathbb{Z}_{++}^J}} H_{\boldsymbol{y}_n} \\
&\preceq -\gamma_0 \frac{\#\{n \leq N : \boldsymbol{y}_n \in \mathbb{Z}_{++}^J\}}{N} I_{J+1} \\
&\preceq -\epsilon_0 \gamma_0 I_{J+1} = -2\gamma I_{J+1}.
\end{aligned}$$

The first inequality omits the remaining periods, whose expected Hessians are negative semidefinite; the second uses $H_{\boldsymbol{y}} \preceq -\gamma_0 I_{J+1}$ for every $\boldsymbol{y} \in \mathcal{Y} \cap \mathbb{Z}_{++}^J$; and the third uses the assumed proportion of such periods. We will bound the sampling fluctuation and the change in the Hessian over a neighborhood by $\gamma/2$ each, leaving a negative bound of $-\gamma I_{J+1}$.

We next control the sampling fluctuation at $\boldsymbol{\beta}_0$. A feasible inventory–sales pair $(\boldsymbol{y}, \boldsymbol{x})$ consists of an initial inventory vector $\boldsymbol{y} \in \mathcal{Y}$ and a terminal-sales vector $\boldsymbol{x} \in \mathbb{Z}_+^J$ satisfying $0 \le x_j \le y_j$ for every product j. For each $\boldsymbol{y}$, there are $\prod_{j=1}^{J}(y_j+1)$ such sales vectors. Since $\mathcal{Y}$ is finite, the total number of feasible pairs is also finite. Let $\|\cdot\|_{\mathrm{op}}$ denote the matrix operator norm induced by the Euclidean norm. For each feasible pair, the likelihood is positive and twice continuously differentiable in a neighborhood of $\boldsymbol{\beta}_0$. Its logarithm therefore has a Hessian with finite entries at $\boldsymbol{\beta}_0$, and this matrix has a finite operator norm. Taking the maximum over the finitely many feasible pairs gives

$$M := \max_{\substack{\boldsymbol{y}\in\mathcal{Y},\, \boldsymbol{x}\in\mathbb{Z}_+^J \\ 0\le \boldsymbol{x}\le \boldsymbol{y}}} \left\|\nabla^2_{\boldsymbol{\beta}} \log L^{\mathrm{P}}(\boldsymbol{\beta}_0 \mid \boldsymbol{y}, \boldsymbol{x})\right\|_{\mathrm{op}} < \infty.$$

The constant M depends only on the fixed parameter $\boldsymbol{\beta}_0$ and the inventory support $\mathcal{Y}$, and is independent of N and the realized sample.

Define

$$\Delta_n := \nabla^2_{\boldsymbol{\beta}} \log L^{\mathrm{P}}_n(\boldsymbol{\beta}_0) - H_{\boldsymbol{y}_n}.$$

Because the inventories are nonrandom and the terminal sales are independent across periods, the matrices Δ_n are independent and have zero means. Since the squared Frobenius norm is the sum of the squared matrix entries, the variance identity gives

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\beta}_0}\|\Delta_n\|_{\mathrm{F}}^2 &= \mathbb{E}_{\boldsymbol{\beta}_0}\left\|\nabla^2_{\boldsymbol{\beta}} \log L^{\mathrm{P}}_n(\boldsymbol{\beta}_0)\right\|_{\mathrm{F}}^2 - \|H_{\boldsymbol{y}_n}\|_{\mathrm{F}}^2 \\ &\le (J+1)M^2,\end{aligned}$$

where the last inequality uses $\|\cdot\|_{\mathrm{F}}^2 \le (J+1)\|\cdot\|_{\mathrm{op}}^2$ for $(J+1)\times(J+1)$ matrices. Independence and zero means eliminate the cross terms when expanding the squared Frobenius norm of the sum. Therefore,

$$\mathbb{E}_{\boldsymbol{\beta}_0}\left\|\frac{1}{N}\sum_{n=1}^{N}\Delta_n\right\|_{\mathrm{op}}^2 \le \mathbb{E}_{\boldsymbol{\beta}_0}\left\|\frac{1}{N}\sum_{n=1}^{N}\Delta_n\right\|_{\mathrm{F}}^2 = \frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}_{\boldsymbol{\beta}_0}\|\Delta_n\|_{\mathrm{F}}^2 \le \frac{(J+1)M^2}{N}.$$

Markov's inequality applied to the squared operator norm now yields

$$\mathbb{P}_{\boldsymbol{\beta}_0}\left(\left\|\frac{1}{N}\sum_{n=1}^{N}\Delta_n\right\|_{\mathrm{op}} > \frac{\gamma}{2}\right) \le \frac{4(J+1)M^2}{N\gamma^2} \longrightarrow 0.$$

A symmetric matrix is bounded above by its operator norm times the identity. Thus, on the event where this norm is at most $\gamma/2$, for all sufficiently large N,

$$\begin{aligned}\nabla^2_{\boldsymbol{\beta}}\mathcal{L}^{\mathrm{P}}_N(\boldsymbol{\beta}_0) &= \frac{1}{N}\sum_{n=1}^{N}H_{\boldsymbol{y}_n} + \frac{1}{N}\sum_{n=1}^{N}\Delta_n \\ &\preceq -2\gamma I_{J+1} + \frac{\gamma}{2}I_{J+1} = -\frac{3\gamma}{2}I_{J+1}.\end{aligned}$$

It remains to extend this bound to a fixed neighborhood of $\boldsymbol{\beta}_0$. For each feasible inventory–sales pair, continuity of the log-likelihood Hessian at $\boldsymbol{\beta}_0$ gives a positive radius within which its change in operator norm is at most $\gamma/2$. Taking the minimum of these radii over the finitely many pairs gives a common $r > 0$, independent of N and the realized sample, such that

$$\max_{\substack{\boldsymbol{y}\in\mathcal{Y},\, \boldsymbol{x}\in\mathbb{Z}_+^J \\ 0\le \boldsymbol{x}\le \boldsymbol{y}}} \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_2\le r} \left\|\nabla^2_{\boldsymbol{\beta}} \log L^{\mathrm{P}}(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{x}) - \nabla^2_{\boldsymbol{\beta}} \log L^{\mathrm{P}}(\boldsymbol{\beta}_0 \mid \boldsymbol{y}, \boldsymbol{x})\right\|_{\mathrm{op}} \le \frac{\gamma}{2}.$$

Averaging over periods and applying the triangle inequality yields, for every sample,

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_2\le r} \left\|\nabla^2_{\boldsymbol{\beta}}\mathcal{L}^{\mathrm{P}}_N(\boldsymbol{\beta}) - \nabla^2_{\boldsymbol{\beta}}\mathcal{L}^{\mathrm{P}}_N(\boldsymbol{\beta}_0)\right\|_{\mathrm{op}} \le \frac{\gamma}{2}.$$

Consequently, on the same event used above, every $\boldsymbol{\beta}$ with $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_2 \le r$ satisfies

$$\begin{aligned}\nabla^2_{\boldsymbol{\beta}}\mathcal{L}^{\mathrm{P}}_N(\boldsymbol{\beta}) &\preceq \nabla^2_{\boldsymbol{\beta}}\mathcal{L}^{\mathrm{P}}_N(\boldsymbol{\beta}_0) + \frac{\gamma}{2} I_{J+1} \\ &\preceq -\frac{3\gamma}{2} I_{J+1} + \frac{\gamma}{2} I_{J+1} = -\gamma I_{J+1}.\end{aligned}$$

This event has probability approaching one, and both r and γ are independent of N, which completes the proof. □

The proof also yields the two properties used in the discussion of gradient-based methods in Section 5.2.2. First, the Hessian of $\mathcal{L}^{\mathrm{P}}_N$ is bounded on the ball for every sample: the Hessian of each period's log-likelihood lies within $\gamma/2$ of its value at $\boldsymbol{\beta}_0$, whose operator norm is at most M, so $\|\nabla^2_{\boldsymbol{\beta}}\mathcal{L}^{\mathrm{P}}_N(\boldsymbol{\beta})\|_{\mathrm{op}} \le M + \gamma/2$ whenever $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_2 \le r$. Second, suppose that $\mathcal{L}^{\mathrm{P}}_N$ is γ-strongly concave on the ball, and let $\widehat{\boldsymbol{\beta}}$ be its maximizer over the ball, which exists because the ball is compact. Fix $\boldsymbol{\beta}$ in the ball and consider the derivative of $t \mapsto \mathcal{L}^{\mathrm{P}}_N(\widehat{\boldsymbol{\beta}} + t(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}))$ on $[0,1]$. At $t=0$, this derivative equals $\nabla\mathcal{L}^{\mathrm{P}}_N(\widehat{\boldsymbol{\beta}})^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}) \le 0$, because $\widehat{\boldsymbol{\beta}}$ maximizes $\mathcal{L}^{\mathrm{P}}_N$ over the convex ball. Its own derivative is at most $-\gamma\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\|_2^2$ on $[0,1]$, because the segment from $\widehat{\boldsymbol{\beta}}$ to $\boldsymbol{\beta}$ lies in the ball. Integrating over $[0,1]$ gives $\nabla\mathcal{L}^{\mathrm{P}}_N(\boldsymbol{\beta})^\top(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}) \le -\gamma\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\|_2^2$, and the Cauchy–Schwarz inequality yields $\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\|_2 \le \|\nabla\mathcal{L}^{\mathrm{P}}_N(\boldsymbol{\beta})\|_2/\gamma$.

Finally, we verify the nonconcavity claim using the zero-sale observation, for which the likelihood is explicit.

Lemma A.7 **(Nonconcavity despite identifiability).** *Under the conditions of Theorem 6, consider the zero-sale observation $\boldsymbol{x}=\mathbf{0}$, and let $\boldsymbol{v}=\bar{\boldsymbol{v}}+r\mathbf{1}$ for a fixed $\bar{\boldsymbol{v}}\in\mathbb{R}^J$, where $r\in\mathbb{R}$ and $\mathbf{1}$ is the all-ones vector. Then the Hessian of $\log L^{\mathrm{P}}$ with respect to (η,r) is indefinite at every (η,r), so the log-likelihood of a single period is not concave in $(\eta,\boldsymbol{v})$.*

Proof of Lemma A.7: By the proof of Theorem 6, $\log L^{\mathrm{P}} = -\alpha_+ = -e^{\eta}b(r)$, where $b(r)=\Theta(r)/\{1+\Theta(r)\}$, with $\Theta(r)=\sum_j e^{\bar{v}_j+r}$, is the probability that a customer buys some product when all products are in stock. Since $\Theta'=\Theta$, we have $b'=b(1-b)$ and $b''=b(1-b)(1-2b)$, so the Hessian with respect to (η,r) is

$$-e^{\eta}\begin{pmatrix} b & b(1-b) \\ b(1-b) & b(1-b)(1-2b)\end{pmatrix},$$

with determinant $e^{2\eta}\{b^2(1-b)(1-2b) - b^2(1-b)^2\} = -e^{2\eta}b^3(1-b) < 0$. A symmetric 2×2 matrix with a negative determinant is indefinite, and since the restriction of a concave function to an affine set is concave, $\log L^{\mathrm{P}}$ is not concave in $(\eta,\boldsymbol{v})$. □

Appendix B: Computational Methods for Gradient Evaluation

This appendix describes how we evaluate the likelihood and its gradient numerically, as summarized in Section 4.2. By Theorems 1–4, all the required quantities are built from one-dimensional expectations with respect to a Gamma distribution. Appendix B.1 approximates these expectations by generalized Gauss–Laguerre quadrature. Appendices B.2 and B.3 explain how we evaluate the scalar functions at the quadrature nodes, namely $\overline{F}_y$ and h_y, and Φ_s and ψ_s, and bound the resulting truncation errors.

B.1. Generalized Gauss–Laguerre Evaluation

Both R^{K} and R^{P}, as well as the expectations under W and $\bar{W}$, are built from integrals of the form $\int_0^\infty u^\alpha e^{-u} f(u)\,du/\alpha!$ for a nonnegative integer α. Generalized Gauss–Laguerre quadrature is designed for integrals with the weight $u^\alpha e^{-u}$. Let $(u_q^{(\alpha)}, w_q^{(\alpha)})_{q=1}^Q$ denote the nodes and weights of this quadrature rule, and set $\bar{w}_q^{(\alpha)} = w_q^{(\alpha)}/\alpha!$ to account for the normalizing constant of the Gamma density. Then

$$\frac{1}{\alpha!}\int_0^\infty u^\alpha e^{-u} f(u)\,du \approx \sum_{q=1}^{Q} \bar{w}_q^{(\alpha)} f(u_q^{(\alpha)}). \tag{B.1}$$

For observation n, let $L_n^r = C_n^r R_n^r$ denote the likelihood factorization in setting $r \in \{\mathrm{K}, \mathrm{P}\}$. The exponent of u in the quadrature weight is $\alpha_n^{\mathrm{K}} = m_n + c_n$, where $m_n = T_n - s_n$, under known arrivals and $\alpha_n^{\mathrm{P}} = c_n$ under Poisson arrivals. The corresponding Gamma shape parameter is one plus this exponent. Fix r and write $u_{nq} = u_q^{(\alpha_n^r)}$ and $\bar{w}_{nq} = \bar{w}_q^{(\alpha_n^r)}$. When $\mathcal{B}_n = \varnothing$, we use the closed-form likelihood and gradient, setting $\widehat{R}_{n,Q}^r = R_n^r$ and $\widehat{L}_{n,Q}^r = L_n^r$. The quadrature formulas below apply when $\mathcal{B}_n \neq \varnothing$.

To write both settings together, define

$$\Xi_{\mathrm{K},n}(u) = 1, \qquad \Xi_{\mathrm{P},n}(u) = \Phi_{s_n}\left(\frac{\lambda u}{1+\sum_{j\in\mathcal{C}_n}\theta_j}\right).$$

Treating scalar evaluations at the nodes as exact, form the positive unnormalized weights

$$\xi_{nq}^r = \bar{w}_{nq}\Xi_{r,n}(u_{nq}) \prod_{j\in\mathcal{B}_n} \overline{F}_{y_{j,n}}(\rho_{j,n}u_{nq}). \tag{B.2}$$

Their sum approximates the Gamma expectation, giving $\widehat{R}_{n,Q}^r = \sum_{q=1}^Q \xi_{nq}^r$ and $\widehat{L}_{n,Q}^r = C_n^r \widehat{R}_{n,Q}^r$. Normalizing these weights gives

$$\widetilde{w}_{nq}^r = \frac{\xi_{nq}^r}{\sum_{\ell=1}^Q \xi_{n\ell}^r}.$$

The unnormalized weights and their sum are evaluated on the log scale. The normalized weights $\widetilde{w}_{nq}^r$ approximate the distributions of W and $\bar{W}$ used in Theorems 3 and 4. In particular, the expectation for sold-out product j is approximated by $\sum_{q=1}^Q \widetilde{w}_{nq}^r h_{y_{j,n}}(\rho_{j,n}u_{nq})$, and under Poisson arrivals,

$$\widehat{\overline{m}}_{n,Q} = \sum_{q=1}^{Q} \widetilde{w}_{nq}^{\mathrm{P}} \psi_{s_n}\left(\frac{\lambda u_{nq}}{1+\sum_{j\in\mathcal{C}_n}\theta_j}\right).$$

Substituting these quantities into the two gradient theorems gives the gradient computed by quadrature.

Let ∇ denote the gradient with respect to $\boldsymbol{v}$ for $r = \mathrm{K}$ and with respect to $(\eta, \boldsymbol{v}^\top)^\top$ for $r = \mathrm{P}$. The nodes and the weights $\bar{w}_{nq}$ depend only on the exponent α_n^r, which is determined by the observation, and remain fixed as the parameters vary. Differentiating the finite sum therefore gives

$$\nabla \log \widehat{L}_{n,Q}^r = \nabla \log C_n^r + \sum_{q=1}^{Q} \widetilde{w}_{nq}^r \nabla \log\left\{\Xi_{r,n}(u_{nq}) \prod_{j\in\mathcal{B}_n} \overline{F}_{y_{j,n}}(\rho_{j,n}u_{nq})\right\}. \tag{B.3}$$

Thus, with exact scalar evaluations, the gradient computed by quadrature is exactly the gradient of the approximate log-likelihood $\log \widehat{L}_{n,Q}^r$. In other words, the optimizer works with a gradient that is consistent with the objective it maximizes.

Choice of the Gamma shape parameter. Under Poisson arrivals, the reported implementation increases the Gamma shape parameter by a fixed nonnegative integer to move the quadrature nodes toward the region contributing most to the integral. For any fixed integer $d \geq 0$,

$$\frac{1}{c_n!}\int_0^\infty u^{c_n} e^{-u} f(u)\, du = \frac{(c_n+d)!}{c_n!}\mathbb{E}_{U\sim\mathrm{Gamma}(c_n+d+1,1)}[U^{-d} f(U)]. \tag{B.4}$$

The factor U^{-d} and the factorial ratio exactly correct the change of Gamma density. The corresponding nodes and coefficients are

$$u_{nq}^{(d)} = u_q^{(c_n+d)}, \qquad \omega_{nq}^{(d)} = \bar{w}_q^{(c_n+d)} \frac{(c_n+d)!}{c_n!}(u_{nq}^{(d)})^{-d}.$$

Replacing u_{nq} and $\bar{w}_{nq}$ by these quantities throughout the quadrature formulas above gives the quadrature rule with the increased Gamma shape parameter. Because d is chosen before optimization and held fixed, the resulting nodes and coefficients remain parameter-independent, so (B.3) still holds. The coefficients $\omega_{nq}^{(d)}$ are positive but need not sum to one; normalized weights are obtained only after forming the unnormalized weights ξ_{nq}^r and dividing each by their sum.

B.2. Evaluation of Poisson Tails and Their Derivatives

For $y \geq 1$ and $\mu > 0$, the Poisson tail equals the regularized lower incomplete gamma function, $\overline{F}_y(\mu) = \{(y-1)!\}^{-1}\int_0^\mu t^{y-1}e^{-t}\, dt$. When this evaluation is numerically unstable for a small tail, we use the positive series

$$\overline{F}_y(\mu) = \pi_y(\mu)\sum_{\ell=0}^{\infty} a_\ell, \qquad a_0 = 1, \qquad a_{\ell+1} = a_\ell \frac{\mu}{y+\ell+1}. \tag{B.5}$$

This series follows from $\overline{F}_y(\mu) = \sum_{\ell\geq 0}\pi_{y+\ell}(\mu)$, because $a_\ell = \pi_{y+\ell}(\mu)/\pi_y(\mu)$. Since $\mu\pi_{y-1}(\mu) = y\pi_y(\mu)$, the same terms give $h_y(\mu) = y/\sum_{\ell=0}^\infty a_\ell$ without dividing two small probabilities.

The adjacent-term ratios decrease with ℓ. Once $r_L = \mu/(y+L+1) < 1$, we have $a_{L+k} \leq a_L r_L^k$ for $k \geq 1$, and hence

$$0 \leq \sum_{\ell>L} a_\ell \leq B_L := \frac{a_L r_L}{1-r_L}.$$

Writing $S_L = \sum_{\ell=0}^L a_\ell$, the exact sum lies between S_L and $S_L + B_L$. Taking logarithms and reciprocals gives

$$0 \leq \log\overline{F}_y(\mu) - \log\{\pi_y(\mu)S_L\} \leq \log(1 + B_L/S_L),$$
$$0 \leq \frac{y}{S_L} - h_y(\mu) \leq \frac{yB_L}{S_L^2}.$$

Thus one recurrence supplies both scalar quantities and explicit truncation bounds.

B.3. Stable Evaluation of Φ_s and ψ_s

For $s \in \mathbb{Z}_+$ and $z > 0$, write

$$t_\ell = \frac{s!z^\ell}{\ell!(s+\ell)!}, \qquad \Phi_s(z) = \sum_{\ell=0}^\infty t_\ell, \qquad \psi_s(z) = \frac{\sum_{\ell=0}^\infty \ell t_\ell}{\sum_{\ell=0}^\infty t_\ell}. \tag{B.6}$$

The ratio $t_{\ell+1}/t_\ell = z/\{(\ell+1)(s+\ell+1)\}$ decreases with ℓ, so the terms are nondecreasing up to their maximum and nonincreasing thereafter. We retain an interval $\ell_-, \ldots, \ell_+$ containing all indices attaining the maximum, including both indices when two adjacent terms tie, and use the same interval for both quantities. Define $S_0 = \sum_{\ell=\ell_-}^{\ell_+} t_\ell$ and $S_1 = \sum_{\ell=\ell_-}^{\ell_+} \ell t_\ell$.

To bound the omitted terms with $\ell > \ell_+$, let $r_+ = z/\{(\ell_+ + 1)(s + \ell_+ + 1)\} < 1$. Since $t_{\ell_+ + k} \le t_{\ell_+} r_+^k$ for $k \ge 1$, summing the geometric bounds for the omitted terms, both unweighted and weighted by their indices, gives

$$B_{0,\mathrm{R}} = \frac{t_{\ell_+} r_+}{1 - r_+}, \qquad B_{1,\mathrm{R}} = t_{\ell_+} \left\{ \frac{\ell_+ r_+}{1 - r_+} + \frac{r_+}{(1 - r_+)^2} \right\}.$$

For the omitted terms with $\ell < \ell_-$, when $\ell_- > 0$, let $r_- = \ell_-(s + \ell_-)/z < 1$. The bound $t_{\ell_- - k} \le t_{\ell_-} r_-^k$ for $1 \le k \le \ell_-$ gives

$$B_{0,\mathrm{L}} = \frac{t_{\ell_-} r_-}{1 - r_-}, \qquad B_{1,\mathrm{L}} = (\ell_- - 1) B_{0,\mathrm{L}},$$

because each of these omitted indices is at most $\ell_- - 1$. When $\ell_- = 0$, no terms are omitted below the retained interval, so both corresponding bounds are zero.

Let $B_i = B_{i,\mathrm{L}} + B_{i,\mathrm{R}}$ for $i = 0, 1$. The omitted denominator and numerator terms in (B.6) lie in $[0, B_0]$ and $[0, B_1]$, respectively. Taking logarithms, and subtracting the exact and truncated ratios using $\Phi_s(z) \ge S_0$, yields

$$\begin{aligned} 0 \le \log \Phi_s(z) - \log S_0 &\le \log\left(1 + \frac{B_0}{S_0}\right), \\ \left| \psi_s(z) - \frac{S_1}{S_0} \right| &\le \frac{B_1 + (S_1/S_0) B_0}{S_0}. \end{aligned} \tag{B.7}$$

Thus the approximations are $\widehat{\Phi}_s(z) = S_0$ and $\widehat{\psi}_s(z) = S_1/S_0$.

The reported implementation constructs the interval either by a forward recurrence from $\ell = 0$ or by expanding around a term attaining the maximum, with relative truncation tolerance $B_0/S_0 \le \varepsilon$, where $\varepsilon = 2 \times 10^{-15}$. Here B_0/S_0 is an upper bound on the omitted sum divided by the retained sum. The implementation evaluates the sums on the log scale, computing $\log \widehat{\Phi}_s(z)$ as the logarithm of the sum of the retained terms and $\widehat{\psi}_s(z)$ as the ratio S_1/S_0. At most 10,000 terms are retained. The stated truncation bound is met when this tolerance is satisfied; reaching the term cap alone does not ensure it. No separate tolerance is imposed on ψ_s, whose error is bounded by (B.7). At $z = 0$, set $\widehat{\Phi}_s(0) = \Phi_s(0) = 1$ and $\widehat{\psi}_s(0) = \psi_s(0) = 0$.

For a fixed retained interval, $z\, d \log S_0/dz = S_1/S_0$. This derivative identity applies while that interval remains fixed; the gradient identity in (B.3) assumes exact scalar evaluations.

Appendix C: Calibration of the Target Stockout Rate

The stockout rate is a key factor in the numerical study: it determines how often the assortment changes within a selling period, and hence how much stockout-induced substitution the data contain. To compare settings with different numbers of products and customers at the same stockout frequency, we fix the target stockout rate ζ and adjust the inventory multiplier c_s in (26) accordingly (Section 6.1). Because the stockout rate implied by a given c_s depends on the attractions, the arrivals, and the substitution within each period, it is not available in closed form, so we calibrate c_s by simulation, separately for each (J, κ) and $\zeta \in \{0.25, 0.50\}$. The calibration relies on the fact that, in a common simulation, larger inventories cannot lead to more stockouts. We first establish this fact and then describe the search.

LEMMA C.1 **(Monotonicity of stockouts in inventory)**. *Let $\boldsymbol{y}' \ge \boldsymbol{y} \ge \mathbf{1}$ be two inventory vectors. Generate the two sales processes from the continuous-time representation in Section 3, using the same Poisson processes $\Pi_0, \ldots, \Pi_J$ and the same number of customers T. Then every product that sells out under $\boldsymbol{y}'$ also sells out under $\boldsymbol{y}$.*

Proof of Lemma C.1: We use superscripts to indicate the initial inventory. Since $\min\{y'_j, \Pi_j(\tau)\} \geq \min\{y_j, \Pi_j(\tau)\}$ for every j and τ, (4) gives

$$N_{\text{tot}}^{\boldsymbol{y}'}(\tau) = \Pi_0(\tau) + \sum_j \min\{y'_j, \Pi_j(\tau)\} \geq \Pi_0(\tau) + \sum_j \min\{y_j, \Pi_j(\tau)\} = N_{\text{tot}}^{\boldsymbol{y}}(\tau).$$

In other words, choices accumulate at least as fast under $\boldsymbol{y}'$, so the Tth choice occurs no later: $\sigma_T^{\boldsymbol{y}'} \leq \sigma_T^{\boldsymbol{y}}$. If product j sells out under $\boldsymbol{y}'$, then $\Pi_j(\sigma_T^{\boldsymbol{y}'}) \geq y'_j$. Since Π_j is nondecreasing,

$$\Pi_j(\sigma_T^{\boldsymbol{y}}) \geq \Pi_j(\sigma_T^{\boldsymbol{y}'}) \geq y'_j \geq y_j,$$

so product j also sells out under $\boldsymbol{y}$. By Lemma 1, each of the two processes has the correct distribution of terminal sales, so the construction is a valid coupling. □

The same conclusion holds in the Poisson-arrival setting after conditioning on a common number of arrivals. We apply Lemma C.1 across candidate multipliers by reusing the inventory draws $U_{j,n}$ from (26), the arrival counts, and the underlying Poisson processes. For each fixed draw $U_{j,n}$, the generated inventory $y_{j,n}(c_s)$ is nondecreasing in c_s. Let $x_{j,n}(c_s)$ denote the corresponding terminal sales. For a pilot sample containing N_{pilot} periods, Lemma C.1 implies that the realized stockout rate

$$\widehat{\zeta}_{\text{pilot}}(c_s) = \frac{1}{N_{\text{pilot}} J} \sum_{n=1}^{N_{\text{pilot}}} \sum_{j=1}^{J} \mathbb{I}_{\{x_{j,n}(c_s) = y_{j,n}(c_s)\}}$$

is nonincreasing in c_s, and so is its expectation.

The search is performed on the log scale. Because the ceiling operation in (26) makes the pilot stockout rate a step function, bisection narrows the search interval; among the evaluated multipliers, we select the one whose pilot rate is closest to ζ.

Each candidate uses three pilot samples of 200 periods in the known-arrival setting and five pilot samples of 500 periods in the Poisson-arrival setting. For targets 0.25 and 0.50, the resulting main-sample stockout rates are 0.2480 and 0.5000 under known arrivals and 0.2499 and 0.4969 under Poisson arrivals. The main samples are generated independently of the pilot samples.