Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Online Learning for Constrained Assortment Optimization Under Markov Chain Choice Model

Shukai Li, Qi Luo, Zhiyuan Huang, Cong Shi

informs.
https://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Articles in Advance, pp. 1–30
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Crosscutting Areas

Online Learning for Constrained Assortment Optimization Under Markov Chain Choice Model

Shukai Li,[a] Qi Luo,[b] Zhiyuan Huang,[c] Cong Shi[d,*]

[a] Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois 60208; [b] Department of Business Analytics, University of Iowa, Iowa City, Iowa 52242; [c] Department of Management Science and Engineering, Tongji University, Shanghai 200092, China; [d] Management Science, Miami Herbert Business School, University of Miami, Coral Gables, Florida 33146

*Corresponding author

Contact: shukaili2024@u.northwestern.edu, https://orcid.org/0009-0005-1406-5803 (SL); qluo2@clemson.edu, https://orcid.org/0000-0002-4103-7112 (QL); huangzy@tongji.edu.cn, https://orcid.org/0000-0003-1284-2128 (ZH); congshi@bus.miami.edu, https://orcid.org/0000-0003-3564-3391 (CS)

Received: December 31, 2022
Revised: December 11, 2023; March 6, 2024
Accepted: March 7, 2024
Published Online in Articles in Advance: May 15, 2024

Area of Review: Market Analytics and Revenue Management

https://doi.org/10.1287/opre.2022.0693

Copyright: © 2024 INFORMS

Abstract. We study a dynamic assortment selection problem where arriving customers make purchase decisions among offered products from a universe of products under a Markov chain choice (MCC) model. The retailer only observes the assortment and the customer's single choice per period. Given limited display capacity, resource constraints, and no a priori knowledge of problem parameters, the retailer's objective is to sequentially learn the choice model and optimize cumulative revenues over a finite selling horizon. We develop a fast linear system based explore-then-commit (FastLinETC for short) learning algorithm that balances the tradeoff between exploration and exploitation. The algorithm can simultaneously estimate the arrival and transition probabilities in the MCC model by solving a linear system of equations and determining the near-optimal assortment based on these estimates. Furthermore, our consistent estimators offer superior computational times compared with existing heuristic estimation methods, which often suffer from inconsistency or a significant computational burden.

Funding: The research of Q. Luo is partially supported by the National Science Foundation [Grant CMMI-2308750]. The research of Z. Huang is partially supported by the Shanghai Sailing Program [Grant 22YF1451100 and the Fundamental Research Funds for the Central Universities]. The research of C. Shi is partially supported by Amazon [Research Award].

Keywords: online learning • assortment planning • Markov chain choice model • capacity • regret analysis

1. Introduction

Assortment optimization problems find important applications in both brick-and-mortar and online retailing. The decision maker selects a subset of products (a.k.a., an assortment) to offer to customers from a universe of N substitutable products to maximize the expected revenue. Effective assortment management improves operational efficiency and provides better customer coverage and purchasing experiences. By integrating advanced analytical methods with assortment optimization, retailers can significantly improve their revenue and negotiation leverage when selecting suppliers (Nip et al. 2021).

Discrete choice models are critical in capturing customers' preferences for offered products as well as their substitution relationship. In a Markov chain choice (MCC) model, each incoming consumer intends to purchase a specific product and will purchase it immediately if it is available. Otherwise, the customer transitions to an alternative product according to a transition probability matrix until she reaches an available one and purchases that product. The MCC model is a generalization of the multinomial logit (MNL) and the random consideration set choice models (Gallego and Lu 2021). It also provides a good approximation for all other random utility models (RUMs) under mild assumptions, such as probit, nested logit, and mixture of MNL (MMNL) models. In addition, because assortment decisions obtained from the static optimization problem depend on the prior knowledge of the discrete choice models (Gallego and Topaloglu 2019), the predicted revenue is sensitive to which choice model is used to define customers' purchase decisions. As a result, assortment selection using the MCC model is robust to model misspecification while being flexible in capturing customers' substitution behaviors (Berbeglia 2016).

Another important merit of the MCC model is that it can be used as an approximate model when the true underlying model is known but the associated assortment optimization problem is intractable (e.g., MMNL and nested logit models). Blanchet et al. (2016) proved that if the MCC model is a good approximation to the underlying model, then the derived solution is likewise near-optimal. The MCC model's transition probabilities also provide useful information regarding the similarity of offered products. According to a recent computational study by Berbeglia et al. (2022), the MCC model has a significant advantage in balancing the modeling accuracy and computational complexity compared with other listed choice models.

This work studies the dynamic assortment optimization problem under the MCC model. Because the demand parameters are unknown a priori, the retailer needs to simultaneously learn customers' preferences from the purchase data and optimize cumulative revenues over a selling horizon T. The current work considers the case of uniform-price items without inventory constraints, and the product prices are fixed throughout the selling horizon. Customers choose the preferred product to maximize their expected utilities, and the retailer only observes purchased items. These are standard assumptions in the dynamic assortment literature to address the role of assortment in balancing information collection and revenue maximization (Sauré and Zeevi 2013, Agrawal et al. 2019). Solving the dynamic assortment optimization problem is critical for promoting new products whose demand models can only be estimated with abundant historical data. This is particularly true when the product life cycle is short. Our interest lies in developing a family of explore-then-commit algorithms that can automatically transition from the information collection (exploration) phase to the revenue maximization (exploitation) phase.

There is an extensive body of literature on online learning algorithms for the dynamic assortment selection problems under models such as MNL (Rusmevichientong et al. 2010; Sauré and Zeevi 2013; Agrawal et al. 2017, 2019; Wang et al. 2018) and nested logit models (Chen et al. 2021a). However, these algorithms typically rely on a specific structure that cannot be readily generalized for other models and may be sensitive to model selection errors. For instance, the MNL model assumes independence from irrelevant alternatives and is found optimistic in estimating recaptured demands (Gallego et al. 2015). Here "recapture" describes the situation where demand is redirected to a different available product, in contrast to "spill" that refers to the loss of demand due to competition or customers choosing not to purchase. Because the MCC model serves as a generalization or approximation for various RUMs, developing new algorithms for learning optimal assortment under the MCC model can capture more sophisticated purchasing behavior and reduce the likelihood of model misspecification.

However, designing learning algorithms for the MCC model is inherently challenging for the following reasons. The first challenge is parameter estimation. The MCC model captures the choice problem by two sets of parameters: (i) the arrival probability vector λ that characterizes the initial preference for all products when customers enter the system and (ii) the transition probability matrix ρ that characterizes the probability of choosing each substitution when the preferred product is not included in the assortment, including leaving the system without a purchase. Hence, the number of parameters grows quadratically with the number of products. The salient substitution between preferred products introduces strong correlations between observations but these substitutions are unobservable to the retailer. Because the choice probabilities' expressions include the inverse of transition submatrices (see (SS-1)–(SS-2)), there are no trivial unbiased estimators based on observing final purchase decisions. The maximum likelihood estimation (MLE) method cannot be directly applied to the MCC model due to the nonconvexity of objective. Second, even assuming that the MCC model's parameters are known a priori, choice probabilities and average revenues in the MCC model cannot be expressed as a simple functional form of the model parameters (Blanchet et al. 2016). Optimizing the assortment selection under simple constraints such as limited display spaces is shown to be NP-hard by Désir et al. (2020). Finally, balancing between exploration and exploitation in online learning is intriguing, considering that parameter estimation and optimization are already challenging tasks on their own. This work aims to address these challenges by leveraging the structural properties of the MCC model.

1.1. Key Results and Contributions

We propose the first online learning algorithm, named fast linear system based explore-then-commit policy (FastLinETC), for dynamic constrained assortment selection under the MCC model. The name originates from the algorithm's fast exploration, which involves testing only $O(N^2)$ assortments, and our estimation techniques, which revolve around transforming a set of linear equations of choice probabilities to recover the model parameters.

Our performance measure is the *cumulative regret* over a selling horizon of T periods, which is defined as the difference in expected revenue between a clairvoyant policy (with access to all the MCC parameters a priori) and our policy (without knowing these parameters). The clairvoyant policy could be optimal or near-optimal. On condition that the underlying (clairvoyant) model admits a polynomial-time algorithm that gives the exact optimal assortment, we use the exact optimal solution. By contrast, if the underlying model only admits an α-approximation algorithm, we use the α-approximate solution instead. We obtain the following two sets of regret bounds.

i. When the exact optimal assortment is computable, our algorithm admits a regret bound of $O(\mathrm{poly}(N)T^{2/3}\log T)$, where $\mathrm{poly}(N)$ is a polynomial function of N (Theorem 1). The exact optimal assortment under constraints is computable in some special settings, for example, when the MCC model is reduced to an MNL model and the constraint set is total-unimodular (TU) (Davis et al. 2013), or when the MCC model is reduced to a general attraction model and the constraint set is cardinality-based (Wang 2013). The most related lower bound result for a constrained MNL model is $\Omega(\sqrt{T})$ (Chen and Wang 2018). How to close this gap in constrained MCC models remains an open research question (which is discussed in more details in Sections 4 and 7). We also remark here that explore-then-commit strategies yield a $\Omega(T^{2/3})$ lower regret bound for general batched bandits (Perchet et al. 2016).

ii. When the exact optimal assortment is not computable, our algorithm admits an α-regret bound of $O(\mathrm{poly}(N)\log T)$ (Theorem 2), where there exists an α-approximation algorithm ($\alpha < 1$) for the static constrained MCC assortment optimization problem. The regret bound is much improved because we are using a weaker clairvoyant benchmark.

In deriving the previous main results, we make the following main contributions:

i. We design a dynamic constrained assortment selection algorithm under the MCC model, which is general enough to envelop the general attraction model (including the MNL model) and meanwhile provides a satisfactory approximation for more advanced models that may not even have extant learning algorithms (for instance, the MMNL model).

Our algorithmic framework is versatile in coping with any TU-constrained assortment optimization problems, such as cardinality, joint display and assortment, and capacity constraints. When the static constrained MCC assortment optimization problem is NP-hard, we use α-optimal assortments as the clairvoyant benchmark. The learning algorithm integrates α-approximation algorithms into online optimization, which gives the near-optimal solution sequentially under the estimated MCC parameters.

ii. Our learning algorithm solves a compounded system of linear equations repeatedly based on *batches* of offered assortments and customer choices and updates the parameter estimator sequentially. We develop an analytical method based on batch-to-batch sampling to quantify the estimation error from the so-called "chain of estimators", involving novel techniques to approximate revenue rate functions via matrix operations. In establishing α-regret bounds, we find the optimality gap of near-optimal assortments introduces an instance-independent "regret-free region" around the true parameters, which improves the scaling to $O(\log T)$.

The learning algorithm carefully balances the exploration and exploitation phases to obtain an instance-independent scaling of $O(\mathrm{poly}(N))$ (rather than combinatorially many assortments). By comparison, Gupta and Hsu (2020) gave the best-known sample complexity $O(N^2)$, but this complexity was only achieved in a stationary setting with oracles for purchasing probabilities. Also, compared with the state-of-the-art expectation-maximization (EM) method for parameter estimation (Şimşek and Topaloglu 2018), which incurs an increasing computational burden as T grows and does not always guarantee convergence, our consistent estimators enjoy provable concentration bounds and superior efficiency with minimal dependence on T.

iii. Our analysis does not require a *suboptimality gap* typically assumed in the literature (i.e., the revenue gap between the unique optimal assortment and all other assortments, which is often unknown in practice). We show that the use of suboptimality gaps in explore-then-commit learning literature (Sauré and Zeevi 2013, Gallego and Lu 2021) can be viewed as a special case of our α-regret analysis based on near-optimal clairvoyant policies (Section 5).

1.2. Literature Review

The MCC model is a flexible scheme for characterizing customers' purchase decisions between offered products with close substitutes. Our research is closely related to a growing body of literature on MCC assortment optimization.

1.2.1. Static MCC Assortment Selection Problem.

The MCC model was formally introduced in the seminal work of Blanchet et al. (2016), whereas Zhang and Cooper (2005) briefly demonstrated this new type of choice process in the context of airline revenue management. Blanchet et al. (2016) discovered that the *unconstrained* assortment optimization problem is equivalent to an optimal stopping problem. Berbeglia (2016) generalized that any MCC model can be converted to a RUM model by a random walk argument. Given the special structure of the unconstrained assortment selection under the MCC model, the optimal solution can be formulated as a linear program (LP) (Feldman and Topaloglu 2017, Gallego and Topaloglu 2019).

The *constrained* assortment selection under the MCC model is generally APX-hard, that is, not possible to approximate better than a constant factor even when all listed prices are uniform. Désir et al. (2020) and Udwani (2021), respectively, proposed $(1/2 - \varepsilon)$-approximation algorithms for the cardinality-constrained and the capacity-constrained assortment optimization problems. By contrast, this research copes with the assortment selection problem concerning a general category of linear constraints.

Because the MCC model approximates well-known models such as nested logit and MMNL models, there is continuing research interest in incorporating it into the network revenue management problem (Feldman and Topaloglu 2017), pricing problem (Dong et al. 2019), joint pricing and inventory problem (Gallego and Kim 2020), and joint assortment and inventory planning problem (El Housni et al. 2021). There are also variations of the MCC model. Ragain and Ugander (2016) proposed the pairwise choice Markov chain model that considered the transition probability to alternative products following the stationary distribution of a continuous-time Markov chain on the set of alternatives. Nip et al. (2021) proposed a variation called the single-transition choice model that limits the number of times that a customer visits any product during the customer's choice process. The platform can control which products to recommend to avoid departure without purchasing after one transition.

1.2.2. Parameter Estimation in MCC Models. The parameter estimation of MCC models is substantially more challenging than other RUM choice models such as MNL. Blanchet et al. (2016) proposed a straightforward parameter estimation approach that offered all products first and then the all-but-one assortments, which is not feasible with any display specific constraints. Two alternative strategies circumvent offering such large-cardinality assortments. Şimşek and Topaloglu (2018) developed an EM algorithm by converting the unobservable log-likelihood function to a complete log-likelihood function with a closed-form expression. However, the estimator is not guaranteed to be consistent or even to converge to a unique limit point; hence, this approach is not applicable to online learning. Moreover, the theoretical results in Şimşek and Topaloglu (2018) required that the MCC model included self-loops, which contradictorily made the MCC model not identifiable and the estimation problem ill defined; see the proof and counterexamples in Appendix C.1. Gupta and Hsu (2020) extended the all-but-one assortment idea to a parameter recovery algorithm under limited-sized assortments and noise-less choice probabilities. Offering consecutive assortments with the carnality of n^* and n^*+1 can reduce the sample complexity to $O(N^2)$ with $n^* \leqslant N/2$.

1.2.3. Online Learning for Dynamic Assortment Optimization. Learning algorithms for assortment optimization are a growing body of literature. The information on customers' preferences is unknown and needs to be learned over the selling horizon with a proper balance between exploration and exploitation. The standard learning approach cannot be directly applied in this context because (a) the expected reward of each assortment is not independent of other assortments (correlated actions), and (b) the online algorithm needs to select the best assortment among a large number of alternatives (combinatorial complexity). Efficient online learning algorithms for more tractable choice models such as MNL models are well studied in the recent literature. Rusmevichientong et al. (2010) and Sauré and Zeevi (2013) considered the explore-then-commit approach with a preset transition threshold over the selling horizon. With additional parameter identification assumptions, Sauré and Zeevi (2013) showed an asymptotic $O(N\log T)$ regret bound. More advanced algorithms have baked this exploration-exploitation tradeoff into the multiarmed bandit (MAB) paradigm (Auer et al. 2002). Agrawal et al. (2017) considered a Thompson sampling–based algorithm for the dynamic assortment selection with a fixed cardinality constraint K that achieved a regret bound of $O(\sqrt{NT}\log TK)$. Agrawal et al. (2019) customized the upper confidence bound (UCB) approach to the dynamic assortment optimization problem and achieved a regret bound of $\tilde{O}(\sqrt{NT})$. (Here $\tilde{O}(\cdot)$ hides any logarithmic factors.) Chen et al. (2021b) relaxed the regret bound on the dependence of N. These algorithms match the lower bound $\Omega(\sqrt{NT})$ for any regret-minimization assortment optimization under the MNL choice model (Chen and Wang 2018). Learning algorithms for assortment optimization under more general settings have also been studied, for example, contextual bandit (Bernstein et al. 2019; Oh and Iyengar 2019; Chen et al. 2020, 2021b; Kallus and Udell 2020), other choice models (Chen et al. 2021a), or joint operational decisions (Miao and Chao 2021).

Because both assortment optimization and parameter estimation for the MCC model are challenging in general, adaptive assortment selection with demand learning remains an open question in the literature. Gallego and Lu (2021) proposed a forward-backward greedy heuristic method for the *unconstrained* assortment selection under the MCC model. Based on this new heuristic, they developed an explore-then-commit learning algorithm that achieved a regret bound of $O(r_{\max}N^2\log T)$, where $r_{\max}$ was the maximal single-product revenue. Their algorithm was based on a binary comparison of revenues and avoided the issue of parameter estimation. However, their heuristic and learning algorithms were only applicable to the unconstrained case, as the regret analysis was conditioned on the observation that locally optimal assortments were globally optimal for the unconstrained assortment with the MCC model. Moreover, their learning algorithm required knowledge of the *suboptimality gap* at the beginning of the selling horizon. This information is however unavailable a priori in practice. Motivated by these limitations, this work aims to develop new learning algorithms for more general dynamic assortment optimization problems.

1.3. Organization

The remainder of this paper is organized as follows. Section 2 formulates the dynamic assortment selection problem under the MCC model. Section 3 proposes a dynamic assortment algorithm with demand learning. Section 4 leverages the special structure of the optimal assortment and provides performance bounds of the assortment algorithm. Section 5 discusses the connection between suboptimality gap and subregret. Section 6 conducts comparative numerical experiments with benchmark results. Section 7 presents our concluding remarks.

For ease of presentation, we introduce a notation $\text{vec}(A)$ that converts a matrix A into a vector composed of all the *rows* of matrix A. For all $x \in \mathbb{R}$, $[x]_+ := \max\{x, 0\}$. The acronym i.i.d. stands for "independent and identically distributed." The acronym w.r.t. stands for "with respect to." The complexity class of APX-hard refers to a set of all NP-optimization problems for which a c-approximation algorithm exists (with c constant).

2. Problem Description and Model Formulation

2.1. Assortment Selection Under the MCC Model

A retailer determines what should be carried in the assortment from a set of products $\mathcal{N} := \{1, 2, \ldots, N\}$. We represent customers' no-purchase alternative by a product 0. For any product subset $S \subseteq \mathcal{N}$, we define the extended subset $S_+ := S \cup \{0\}$. For example, $\mathcal{N}_+ = \mathcal{N} \cup \{0\} = \{0, 1, \ldots, N\}$ represent all products plus the no purchase option. For any assortment $S \subseteq \mathcal{N}$ and product $i \in \mathcal{N}_+$, $\pi(i, S)$ denotes the probability that a customer purchases product i. We use $r_i \in [r_{\min}, r_{\max}]$ to denote the revenue for selling one unit of product $i \in \mathcal{N}$, where $r_{\max}$ and $r_{\min}$ are two constants such that $0 < r_{\min} \leqslant r_{\max}$. We assume $r_{\min} > 0$ because products yielding zero revenue are unlikely to be offered by sellers, and consequently, they will not emerge in the market or be considered by customers.

The revenue vector $r := (r_0, r_1, \ldots, r_N)$ with $r_0 = 0$ for the no-purchase option.

We model the demand's spill and recapture as a Markov chain. Each arriving consumer has a first-choice product $i \in \mathcal{N}_+$ with probability λ_i and purchases product i as long as it is available in the current assortment; otherwise, the demand is redirected to product $j \in \mathcal{N}_+$ (including no purchase) with probability ρ_{ij}. Next, the customer behaves as if the customer's first-choice demand was product j: she purchases j if available and otherwise repeats the redirection. Product 0 represents the no-purchase option and is always available for customers. Therefore, transition from product 0 to other products is trivial and we can assume $\rho_{0i} = 1 - \rho_{00} = 0\,(i \in \mathcal{N})$. We also assume $\rho_{ii} = 0\,(i \in \mathcal{N})$, that is, there are no self-loops, because (i) we are interested in the eventual alternative product's distribution when a transition occurs, and (ii) any transition matrix with self-loops can be transformed into an equivalent matrix without self-loops. A transition matrix with self-loops is not identifiable as shown in Appendix C.1.

We define a vectorization operation $\text{vec}(\cdot)$ that maps any subset of $\{\lambda_i\}_{i\in\mathcal{N}} \cup \{\rho_{ij}\}_{i,j\in\mathcal{N}}$ into a column vector in the order of $\lambda_i (i \in \mathcal{N})$ and then $\rho_{ij} (i, j \in \mathcal{N})$, where ρ_{ij} is sorted row-wise (i.e., index i-wise). Particularly, let $\theta := \text{vec}(\{\lambda_i\}_{i\in\mathcal{N}} \cup \{\rho_{ij}\}_{i,j\in\mathcal{N}})$. The MCC model is completely characterized by a tuple $(\mathcal{N}, r, \theta)$: once we know θ, all arrival probabilities $\{\lambda_i\}_{i\in\mathcal{N}_+}$ for first-choice demand and all transition probabilities $\{\rho_{ij}\}_{i,j\in\mathcal{N}_+}$ for demand redirection are known, for example, $\lambda_0 = 1 - \sum_{i\in\mathcal{N}}\lambda_i$.

2.2. Static MCC Assortment Optimization Problem

Consider an MCC model $(\mathcal{N}, r, \theta)$ with all parameters known. The retailer determines an assortment to maximize revenues. We first analyze the revenue associated with every assortment. Given an assortment S, we compute the choice probabilities by solving a system of linear equations: we first compute $u(i, S; \theta)(i \in \mathcal{N}_+)$, the *average visit times* to product i under assortment S, by solving

$$u(j, S; \theta) = \lambda_j + \sum_{i\in\mathcal{N}_+} \mathbb{1}\{i \notin S_+\} \cdot u(i, S; \theta)\rho_{ij}, \quad j \in \mathcal{N}_+. \tag{SS-1}$$

The average visit times to product $i \in \mathcal{N}_+$ count the expected number of times that the customer's demand is redirected to product i (including the first-choice demand) during the customer's purchase decision process. For an unavailable product $i \in \mathcal{N} \backslash S$, the average visit times $u(i, S; \theta)$ may be greater than one because a customer's demand may be redirected to this product for multiple times. However, for an available product $i \in S_+$, the average visit times $u(i, S; \theta)$ must lie in $[0, 1]$ because, once a customer's demand is redirected to product i, the preference transition stops and the customer ends up with purchasing product i. This also indicates that the *choice probability* $\pi(i, S; \theta)$ equals the average visit times for every product $i \in S_+$:

$$\pi(i, S; \theta) = \mathbb{1}\{i \in S_+\} \cdot u(i, S; \theta), \quad i \in \mathcal{N}_+. \tag{SS-2}$$

Thus, the (*average*) *single-sale revenue* $r(S; \theta)$ under assortment S can be obtained by

$$r(S; \theta) = \sum_{i\in\mathcal{N}} r_i \pi(i, S; \theta). \tag{SS-3}$$

We consider the assortment optimization problem possibly constrained and let the set $\mathcal{S} \subseteq 2^{\mathcal{N}}$ denote the *possible assortments*. Several naturally arising constraints over the offered assortments include *cardinality constraints*, (i.e., $|S| \leqslant \bar{s}$), *capacity constraints*, (i.e., each product has a weight w_i, and the assortment is restricted to those with total weights $\sum_{i\in S} w_i \leqslant W$), *partition matroid constraints*, (i.e., the products are partitioned into segments, and the assortment has an upper bound on the number of products from each segment), and *joint display and assortment constraints*, (i.e., the assortment should include the display segment of each product).

We let $S^*(\theta)$ be the assortment maximizing the single-sale revenue under constraint $\mathcal{S}$:

$$S^*(\theta) = \arg\max_{S \in \mathcal{S}} r(S; \theta). \qquad \text{(SS-4)}$$

When (SS-4) has multiple solutions, we let $S^*(\theta)$ be an arbitrary assortment that maximizes the single-sale revenue $r(S;\theta)$. In the unconstrained scenario, that is, $\mathcal{S} = 2^{\mathcal{N}}$, the assortment selection (SS-4) under the MCC model can be reformulated into an optimal stopping time problem. The optimal assortment can be obtained from an LP; see theorem 5.1 and lemma 5.2 in Blanchet et al. (2016). In the constrained scenario, the exact optimal assortment is computable in some special settings, for example, when the MCC model is reduced to an MNL model and the constraint set is total-unimodular (TU) (Davis et al. 2013), or when the MCC model is reduced to a general attraction model and the constraint set is cardinality based (Wang 2013).

When the MCC model is not reduced, however, Désir et al. (2020) showed that the assortment selection (SS-4) could be APX-hard even under simple cardinality constraints and a uniform r. Because the exact optimal assortment $S^*(\cdot)$ can be uncomputable (in polynomial time), the retailer may instead offer an *α-optimal assortment* ($\alpha \in (0,1)$), the definition of which is consistent with the approximation algorithm literature including Désir et al. (2020) and Udwani (2021).

Definition 1 (α-Optimal Assortments). For the MCC model $(\mathcal{N}, r, \theta)$ with possible assortments $\mathcal{S}$, an *α-optimal assortment* denoted by $S^\alpha(\theta)$ is any assortment in $\mathcal{S}$ such that

$$r(S^\alpha(\theta);\theta) \geqslant \alpha r(S^*(\theta);\theta).$$

The collection of α-optimal assortments is denoted by

$$\mathcal{S}^\alpha(\theta) := \{S \in \mathcal{S} \mid r(S;\theta) \geqslant \alpha r(S^*(\theta);\theta)\}.$$

For instance, if the possible assortments $\mathcal{S}$ are subject to cardinality or capacity constraints, the approximation ratio α may be set to $\frac{1}{2} - \varepsilon$ for any $\varepsilon > 0$. The approximation algorithms are provided by Désir et al. (2020) and Udwani (2021) (see Appendix B.2). Thus, this paper treats α as a predefined constant associated with $\mathcal{S}$.

2.3. Online MCC Assortment Optimization Problem

The online assortment optimization assumes that the information about the product set $\{\mathcal{N}, r, \mathcal{S}\}$ is known but the MCC parameter θ is not known a priori. During the selling horizon, the set of products offered to each customer and the customers' purchased products are observable while the transition path that a customer follows in the MCC model is not observable. Let T denote the total number of customers that arrive during the selling horizon with one customer per period. We index customers and their arrival periods as $t \in \mathcal{T} := \{1, 2, \ldots, T\}$ and use two terms alternately throughout this study. Products' revenues are assumed to be fixed. The retailer's (*average*) *cumulative revenue* is

$$\mathsf{R}_P(T,\theta) := \mathbb{E}\sum_{t=1}^{T} r(S_t;\theta).$$

Here P denotes the retailer's nonanticipating assortment selection policy that is defined as follows: let $\{S_t\}_{t\in\mathcal{T}} \in \mathcal{S}^T$ be an assortment process, and Z_i^t be customer t's purchase decision regarding product $i \in \mathcal{N}$ with $Z_i^t = 1$ representing purchasing product i and $Z_i^t = 0$ otherwise. Particularly, $Z_0^t = 1$ represents purchasing nothing and $Z_0^t = 0$ otherwise. Let $Z^t := (Z_0^t, Z_1^t, \ldots, Z_N^t)$ be the purchase decision vector of customer t. Let $\mathcal{F}_t := \sigma((S_u, Z^u)_{1 \leqslant u \leqslant t}), t \in \mathcal{T}$ be the filtration associated with the assortment process and purchase decisions of customers $\{0, 1, \ldots, t\}$, and $\mathcal{F}_0 = \varnothing$. Let $\mathcal{P}$ be the collection of nonanticipating assortment policies, that is, any assortment policy $P \in \mathcal{P}$ is a mapping from past histories to possible assortment decisions $\{S_t\}_{t\in\mathcal{T}} \in \mathcal{S}^T$ such that S_t is $\mathcal{F}_{t-1}$-measurable for all $t \in \mathcal{T}$.

Because the MCC parameter θ is not known a priori, the assortment selection policy maximizing cumulative revenues is not obtainable. We analyze the performance of any online assortment selection policy via *regret*, that is, the difference between the cumulative revenue earned by an oracle who knows parameter θ and that earned by a retailer who is uncertain about θ. Depending on the computability of the static assortment optimization problem (SS-4), we define two types of (average) cumulative regret.

2.3.1. Cumulative Regret for Exact Optimal Assortment.

Suppose the exact optimal assortment $S^*(\cdot)$ in (SS-4) is computable. The retailer should repeatedly offer the revenue-maximizing assortment $S^*(\theta)$ when the parameter θ is known a priori. The maximal cumulative revenue is

$$\mathsf{R}^*(T,\theta) = \sum_{t=1}^{T} r(S_t^*(\theta);\theta) = T \cdot r(S^*(\theta);\theta).$$

When θ is not known a priori, the exact optimal assortment should be used as a benchmark, and our primary objective is to derive an assortment selection policy $P \in \mathcal{P}$ that minimizes the *cumulative regret*

$$\mathsf{Reg}_P(T,\theta) := \mathsf{R}^*(T,\theta) - \mathsf{R}_P(T,\theta).$$

2.3.2. Cumulative α-Regret for Near-Optimal Assortment.

When the exact optimal assortment $S^*(\cdot)$ is not computable, the retailers may offer α-optimal assortments to customers and obtain the following cumulative revenue when the parameter θ is known a priori:

$$\mathsf{R}^{\alpha}(T,\theta) = \sum_{t=1}^{T} r(S_t^{\alpha}(\theta);\theta).$$

Here $S_t^{\alpha}(\theta) \in \mathcal{S}^{\alpha}(\theta)$ may vary with t and may not be unique because the α-optimal assortment in the static assortment optimization problem (SS-4) may have multiple solutions. In our context, our primary objective is to design an assortment selection policy minimizing the *cumulative α-regret*

$$\begin{aligned}\mathsf{Reg}_P^{\alpha}(T,\theta) &:= \inf_{S_t^{\alpha}(\theta)\in\mathcal{S}^{\alpha}(\theta),\, t\in\mathcal{T}} \mathsf{R}^{\alpha}(T,\theta) - \mathsf{R}_P(T,\theta)\\ &= \inf_{S_t^{\alpha}(\theta)\in\mathcal{S}^{\alpha}(\theta),\, t\in\mathcal{T}} \sum_{t=1}^{T} r(S_t^{\alpha}(\theta);\theta) - \mathsf{R}_P(T,\theta).\end{aligned}$$

Here operation $\inf\{\cdot\}$ is due to the possible existence of multiple α-optimal assortments. The cumulative α-regret measures the revenue loss due to the unknown MCC parameter and compares the implemented policy against the worst α-optimal assortments. If the implemented policy is purely comprised of α-optimal assortments, the cumulative α-regret is at most zero.

Remark 1 (Near-Optimal Benchmark). Using a weaker tractable benchmark as a substitute for exact optimal benchmark is not uncommon in the learning literature, particularly when the clairvoyant optimal policy is not computable. For example, Zhong et al. (2022) considered learning algorithms for the scheduling problem in multiclass many server queues with abandonment, whose optimal policy was intractable even when the model parameters were known a priori. The authors used a benchmark policy of simple prioritization rule that was asymptotically optimal as the arrival rates and number of servers approach infinity.

The α-regret is also closely related with the concept of suboptimality gap in the explore-then-commit learning literature. As detailed in Section 5, the use of suboptimality gap can be viewed as a special case of our α-regret analysis.

3. FastLinETC Algorithm

3.1. Challenges and Overview

This section outlines an explore-then-commit algorithm for online assortment selection under the MCC model. Algorithms under an MNL model are provided in Rusmevichientong et al. (2010) and Sauré and Zeevi (2013). For our algorithm design under the MCC model, there are two central problems: (i) how to identify a separation period τ that divides the selling horizon $\mathcal{T}$ into an exploration phase and an exploitation phase and (ii) how to estimate the MCC parameters conditional on customers' final purchase decisions in the exploration phase.

The estimation of the arrival probability vector λ and transition matrix ρ is challenging because the choice probabilities' expressions (SS-1)–(SS-2) involve the inverse of submatrices of ρ and thereby the associated log-likelihood function may not be concave. Şimşek and Topaloglu (2018) proposed an EM algorithm based on a concave incomplete log-likelihood function. However, the EM estimator may have multiple limit points in the exploration phase and has no consistency guarantee. Moreover, the theoretical results in Şimşek and Topaloglu (2018) required that the MCC model included self-loops, which contradictorily made the MCC model lose identifiability as shown in Appendix C.1. Thus, using the EM estimator cannot help derive sublinear regret bounds.

We propose a parameter recovery method (E-1)–(E-9) with consistency guarantee and sub-Gaussian concentration bounds (Lemma 5). To establish that, we assume the possible assortment set $\mathcal{S}$ has a subset such that, by presenting the included assortments to customers and observing purchase decisions, a consistent estimator of θ can be constructed:

Assumption 1. *There exist* (i) $n^* \in \{2,3,\ldots,\lfloor\frac{N}{2}\rfloor\}$, (ii) *two assortments* $S_{\cap}, S'_{\cap} \subseteq \mathcal{N}$ *such that* $S_{\cap} \cap S'_{\cap} = \emptyset$, $|S_{\cap}| = |S'_{\cap}| = n^* - 1$, *and* (iii) *assortment collections* $\mathcal{S}_0, \tilde{\mathcal{S}}_0 \subseteq \mathcal{S}$ *such that*

$$\begin{aligned}\mathcal{S}_0 &:= \{S \cup \{k\} \mid k \in \mathcal{N}\backslash S, S \in \{S_{\cap}, S'_{\cap}\}\},\\ \tilde{\mathcal{S}}_0 &:= \{S \cup \{k\} \cup \{j\} \mid j \in \mathcal{N}\backslash(S \cup \{k\}),\\ &\qquad k \in \mathcal{N}\backslash S, S \in \{S_{\cap}, S'_{\cap}\}\}.\end{aligned}$$

Assumption 1 is easy to verify under commonly used assortment constraints. For example, if $\mathcal{S}$ is defined via a cardinality constraint such that $\mathcal{S} := \{S \subseteq \mathcal{N} \,|\, |S| \leqslant \bar{s}\}$, Assumption 1 is equivalent to assuming $\bar{s} \geqslant 3$. If $\mathcal{S}$ is defined via a capacity constraint such that $\mathcal{S} := \{S \subseteq \mathcal{N} \mid \sum_{i\in S} w_i \leqslant W\}$, Assumption 1 is equivalent to assuming $w_{(2)} + w_{(N-1)} + w_{(N)} \leqslant W$, where $w_{(i)}$ denotes the weight of the ith product in the increasing order of all products. These specific constraints are considered because Désir et al. (2020) and Udwani (2021) have provided constant-factor approximation algorithms for static constrained assortment optimization problems.

Our parameter estimation method is based on presenting the assortments in $\mathcal{S}_0 \cup \tilde{\mathcal{S}}_0$. Let $d := |\tilde{\mathcal{S}}_0| + |\mathcal{S}_0|$ and $d_0 := |\mathcal{S}_0|$. Then $d \leqslant N^2$ and $d_0 \leqslant 2N$; that is, we use $O(N^2)$ different assortments for estimation. Different from other RUMs such as MNL models (Agrawal et al. 2017, 2019), the parameter estimation of the MCC model cannot be performed with observations from repeatedly offering a single assortment. For example, suppose we use a single assortment $S \subset \mathcal{N}$ to estimate the MCC parameters. Then the transition from product $i \in S$ to other products will never occur, and the resulting transition parameters $\{\rho_{ij}\}_{j\in\mathcal{N}}$ cannot be identified. Under the unconstrained setting ($\mathscr{S} = 2^{\mathcal{N}}$), using only $O(N)$ assortments is possible, and Appendix E shows how this relaxation simplifies our learning algorithm and regret analysis.

We weave the estimation techniques into online optimization and derive the separation period τ based on our estimators' sub-Gaussian concentration bounds. During the exploration phase from period 1 to τ, the retailer repeatedly presents assortments in $\mathscr{S} \cup \tilde{\mathscr{S}}_0$ and observes customers' purchase decisions. At the end of this phase, a consistent MCC parameter estimator $\tilde{\theta}^\tau$ is generated based on the collected data. Next, a *recommended assortment* is computed based on $\tilde{\theta}^\tau$, which will be offered during the remaining exploitation phase from period $\tau+1$ to T. We prove the consistency of estimator $\tilde{\theta}^\tau$ obtained from the exploration phase in Section 4. Because the recommended assortment is computed based on $\tilde{\theta}^\tau$, the latter's consistency ensures the former's convergence in probability to the exact optimal (respectively, α-optimal) assortment if the exact optimal assortment $S^*(\cdot)$ is computable (respectively, otherwise).

3.2. Description of FastLinETC

3.2.1. Inputs and Initialization. Given the following information about products and selling horizon $\{N, r, \mathscr{S}, T\}$, we have two inputs prior to implementing the learning algorithm: (i) a separation period τ that splits the exploration and exploitation phases and (ii) an assortment function $S_P(\cdot)$ that maps the exploration estimator $\tilde{\theta}^\tau$ to an exact optimal/near-optimal assortment. In the following description, we keep the flexibility in specifying τ and $S_P(\cdot)$, which depend on the availability of exact optimal/near-optimal solutions in the static assortment optimization and yield different regret bounds (Theorems 1 and 2). For example, if $S^*(\cdot)$ is computable for any estimated parameter, we let $\tau = \lceil T^{\frac{2}{3}} \log T \rceil$ and set function $S_P(\cdot)$ to be $S^*(\cdot)$ throughout the selling horizon. This combination will yield a cumulative regret bounded by $O(\text{poly}(N) T^{\frac{2}{3}} \log T)$.

3.2.2. Exploration Phase. In the exploration phase, we repeatedly present the assortments in $\mathscr{S}_0 \cup \tilde{\mathscr{S}}_0$, where the assortment set $\mathscr{S}_0 \cup \tilde{\mathscr{S}}_0$ is defined in Assumption 1, and each presented assortment has the cardinality of n^* or $n^* + 1$.

Example 1. For better understanding of this exploration phase, we provide a simple illustrative example with $\mathcal{N} = \{1,2,3,4\}$. Following Assumption 1, we can define the exploration assortments as $\tilde{\mathscr{S}}_0 \cup \mathscr{S}_0 = \{\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{1,2,3\},\{1,2,4\},\{1,3,4\},\{2,3,4\}\}$ and denote $\tilde{\mathscr{S}}_0 \cup \mathscr{S}_0$ as $\{A_0, A_1, \ldots, A_8\}$ accordingly. Then in the exploration phase, the algorithm will sequentially offer $A_0, A_1, \ldots, A_8, A_0, A_1, \ldots, A_8, \ldots$ to customers until the separation period τ. □

At the separation period τ, we obtain MCC parameter estimators $\hat{\theta}^\tau$ and $\tilde{\theta}^\tau$ as follows. We first define the trivial parameters' indices $I_0 := \{(i,i) \mid i \in \mathcal{N}\}$ since $\rho_{ii} \equiv 0$, $i \in \mathcal{N}$. (I_0 may include more indices in $\mathcal{N}^2$, such as in a sparse transition matrix, $\rho_{ij} \equiv 0$ for i, j belonging to different product categories.) Then the MCC parameter θ is divided into trivial and nontrivial elements such that $\theta = \text{vec}(\theta_0, \theta_{++})$, where the trivial parameter $\theta_0 := \text{vec}(\{\rho_{ij}\}_{(i,j)\in I_0})$ and the nontrivial parameter $\theta_{++} := \text{vec}(\{\lambda_i\}_{i\in\mathcal{N}} \cup \{\rho_{ij}\}_{(i,j)\in\mathcal{N}^2\setminus I_0})$. Then θ_0 is naturally estimated by

$$\hat{\theta}_0^\tau := \text{vec}(\{\hat{\rho}_{ij}^\tau \equiv 0\}_{(i,j)\in I_0}). \tag{E-1}$$

Nontrivial parameter θ_{++} is estimated by solving a system of linear equations due to intrinsic properties of the MCC model. Based on the collected data in the exploration phase including offered assortments $\{S_t\}_{t\leqslant\tau}$ and customers' purchase decisions $\{Z^t\}_{t\leqslant\tau} = \{(Z_0^t, Z_1^t, \ldots, Z_N^t)\}_{t\leqslant\tau}$, we estimate the nontrivial parameter θ_{++} in four steps.

Step 1: The choice probabilities $\pi(i,S)$ for every $i \in \mathcal{N}_+$ and $S \in \mathscr{S}_0 \cup \tilde{\mathscr{S}}_0$ are estimated by

$$\hat{\pi}^\tau(i,S) := \frac{\sum_{t=1}^{\tau} \mathbb{1}\{S_t = S, Z_i^t = 1\}}{\sum_{t=1}^{\tau} \mathbb{1}\{S_t = S\}}, \quad i \in \mathcal{N}_+, S \in \mathscr{S}_0 \cup \tilde{\mathscr{S}}_0. \tag{E-2}$$

Here the denominator denotes the frequency that assortment S is offered, and the numerator denotes the frequency that product i is purchased under assortment S. Therefore, the resulting fraction can be used as an estimator for choice probability $\pi(i,S)$. In the following steps, we use these choice probability estimators to construct a set of linear equations, from which we recover the parameter θ_{++}.

Step 2: We define intermediate variables for all $i, j \in \mathcal{N}_+, S \in \mathscr{S}_0$ as

$$\hat{\pi}^\tau(j,S|i) := \begin{cases} 1, & \text{if } i = j, \\ \dfrac{\hat{\pi}^\tau(j,S) - \hat{\pi}^\tau(j,S\cup\{i\})}{\hat{\pi}^\tau(i,S\cup\{i\})}, & \text{if } i \in \mathcal{N}\setminus S, \\ 0, & \text{if } i \in S_+\setminus\{j\}. \end{cases} \tag{E-3}$$

Here $\hat{\pi}^\tau(j,S|i)$ estimates the probability of purchasing product j from assortment S conditional on that the first-choice demand is product i.

Step 3: θ_{++} is estimated (recovered) by minimizing squared residuals of the following linear equations:

$$\sum_{k\in\mathcal{N}} [\hat{\pi}^\tau(j,S|k) - \hat{\pi}^\tau(j,S|0)]\hat{\rho}_{ik}^\tau = [\hat{\pi}^\tau(j,S|i) - \hat{\pi}^\tau(j,S|0)], \quad i \in \mathcal{N}\setminus S, j \in S_+, S \in \mathscr{S}_0, \tag{E-4}$$

$$\sum_{k\in\mathcal{N}} [\hat{\pi}^\tau(j,S|k) - \hat{\pi}^\tau(j,S|0)]\hat{\lambda}_k^\tau = [\hat{\pi}^\tau(j,S) - \hat{\pi}^\tau(j,S|0)], \quad j \in S_+, S \in \mathscr{S}_0. \tag{E-5}$$

Note that the predefined trivial parameter estimator $\hat{\theta}_0^\tau = \text{vec}(\{\hat{\rho}_{ij}^\tau \equiv 0\}_{(i,j)\in I_0})$ is plugged into (E-4) before computing the solutions above. For the convenience of exposition, we rewrite events (E-4)–(E-5) using a matrix notation:

$$\{\hat{\theta}^{\tau}_{++} : (E-4),(E-5)\} \Longleftrightarrow \{\hat{\theta}^{\tau}_{++} : \hat{X}^{\tau}\hat{\theta}^{\tau}_{++} = \hat{Y}^{\tau}\}, \quad \text{(E-6)}$$

where entries of $\hat{X}^{\tau}$ and $\hat{Y}^{\tau}$ are defined by coefficients in (E-4) and (E-5). Minimizing squared residuals gives the following estimator for θ_{++}:

$$\begin{aligned}\hat{\theta}^{\tau}_{++} &:= \mathrm{vec}(\{\hat{\lambda}^{\tau}_i\}_{i\in\mathcal{N}} \cup \{\hat{\rho}^{\tau}_{ij}\}_{(i,j)\in\mathcal{N}^2\setminus I_0})\\ &= (\hat{X}^{\tau^{\top}}\hat{X}^{\tau})^{-1}(\hat{X}^{\tau^{\top}}\hat{Y}^{\tau}), \quad \text{(E-7)}\end{aligned}$$

Combining (E-1) and (E-7), we obtain the following estimator for θ:

$$\hat{\theta}^{\tau} := \mathrm{vec}(\hat{\theta}^{\tau}_0, \hat{\theta}^{\tau}_{++}) = \mathrm{vec}(\{\hat{\lambda}^{\tau}_i\}_{i\in\mathcal{N}} \cup \{\hat{\rho}^{\tau}_{ij}\}_{(i,j)\in\mathcal{N}^2}). \quad \text{(E-8)}$$

Because $\hat{\theta}^{\tau}$ is computed from (E-4)–(E-5), it may not be a valid MCC parameter in space Θ. We use the following rounded estimator as the final estimator:

$$\tilde{\theta}^{\tau} := \arg\min_{\theta'\in\Theta} \|\theta' - \hat{\theta}^{\tau}\|_1. \quad \text{(E-9)}$$

3.2.3. Exploitation Phase.

In the exploitation phase, a *recommended assortment* $S_P(\tilde{\theta}^{\tau})$ is offered based on estimator $\tilde{\theta}^{\tau}$ until period T.

The explore-then-commit learning algorithm for the dynamic assortment optimization problem under the MCC model is summarized in Algorithm 1.

Algorithm 1 (FastLinETC $P(N, T, \tau, S_P(\cdot))$)

```
Input: integer $\tau$ and assortment function $S_P(\cdot)$.
Output: offered assortments $\{S_t\}_{t=1}^{T}$.
Phase 1. Exploration:
Define an arbitrary order for the exploration assortments $\mathcal{S}_0 \cup \tilde{\mathcal{S}}_0$ satisfying Assumption 1 and denote them as $\{A_0, A_1, \ldots, A_{d-1}\}$.
for $t \in \{1, 2, \ldots, \tau\}$ do
    Define $k_t := (t-1) \mod d$, and offer $S_t = A_{k_t}$ to customer $t$.
    Observe the customer purchase decisions $Z^t = (Z^t_0, Z^t_1, \ldots, Z^t_N)$.
end for
Compute choice probability estimators $\hat{\pi}^{\tau}(i, S)$ for all $i \in \mathcal{N}_+, S \in \mathcal{S}_0 \cup \tilde{\mathcal{S}}_0$ via (E-2).
Compute conditional choice probability estimators $\hat{\pi}^{\tau}(j, S|i)$ for all $i, j \in \mathcal{N}_+, S \in \mathcal{S}_0$ via (E-3).
Compute the linear equation system's coefficients $\hat{X}^{\tau}$ and $\hat{Y}^{\tau}$ via (E-4), (E-5), and (E-6).
Compute the MCC parameter estimator $\hat{\theta}^{\tau}$ via (E-1), (E-7), and (E-8).
Compute the rounded MCC parameter estimator $\tilde{\theta}^{\tau}$ via (E-9).
Phase 2. Exploitation:
To all remaining $T - \tau$ customers, offer $S_P(\tilde{\theta}^{\tau})$.
```

4. Performance Analysis of FastLinETC

Our main results are two instance-independent upper bounds on the policy regret associated with Algorithm 1. Given the assortment constraints and possible assortments $\mathcal{S}$, if the exact optimal assortment is computable, an order-of-$T^{\frac{2}{3}}\log T$ policy (i.e., $\tau \in O(T^{\frac{2}{3}}\log T)$) will yield a cumulative regret of $O(\mathrm{poly}(N)T^{\frac{2}{3}}\log T)$. Otherwise, if near-optimal assortments are computable, an order-of-$\log T$ policy will yield a cumulative α-regret of $O(\mathrm{poly}(N)\log T)$. Here $\alpha \in (0,1)$ is a predefined constant associated with $\mathcal{S}$ for the benchmark of constrained assortment optimization as outlined in Section 2.3.

To establish our results, we make the following assumption on parameter space.

Assumption 2. *For the MCC model $(\mathcal{N}, r, \theta)$, θ belongs to a space $\Theta := \Theta_{\lambda} \times \prod_{i=1}^{N}\Theta_{\rho_{i\cdot}}$, where*

$$\Theta_{\lambda} := \left\{\frac{1}{\sum_{k=0}^{N} a_k}\cdot(a_1, a_2, \ldots, a_N)\,\middle|\, a_0 = 1, a_j \in [\underline{a}_j, \overline{a}_j], j \in \mathcal{N}\right\},$$

$$\Theta_{\rho_{i\cdot}} := \left\{\frac{1}{\sum_{k=0}^{N} b_{ik}}\cdot(b_{i1}, b_{i2}, \ldots, b_{iN})\,\middle|\, b_{i0} = 1, b_{ij} \in [\underline{b}_{ij}, \overline{b}_{ij}], j \in \mathcal{N}\right\}, \quad i \in \mathcal{N},$$

and the known constants $\{\underline{a}_i, \overline{a}_i\}_{i\in\mathcal{N}} \cup \{\underline{b}_{ij}, \overline{b}_{ij}\}_{i,j\in\mathcal{N}} \subseteq \mathbb{R}_+$ satisfy the following two conditions:

(No self-loops) *for all $i \in \mathcal{N}$, $\underline{b}_{ii} = \overline{b}_{ii} = 0$; and*

(Bounded attraction) *$0 < \underline{a}_i \leqslant \overline{a}_i < \infty\,(i \in \mathcal{N})$, and $\overline{b}_{ij} < \infty\,(i, j \in \mathcal{N})$.*

In this assumption, the unknown $\{a_i\}_{i\in\mathcal{N}}$ and $\{b_{ij}\}_{i,j\in\mathcal{N}}$ determine the parameter $\theta = \mathrm{vec}(\{\lambda_i\}_{i\in\mathcal{N}} \cup \{\rho_{ij}\}_{i,j\in\mathcal{N}})$: given $\{a_i\}_{i\in\mathcal{N}}$ and $\{b_{ij}\}_{i,j\in\mathcal{N}}$, we can compute $\lambda_i = \frac{a_i}{\sum_{k=0}^{N} a_k}$ and $\rho_{ij} = \frac{b_{ij}}{\sum_{k=0}^{N} b_{ik}}$ for all $i, j \in \mathcal{N}$. Correspondingly, the ranges of $\{a_i\}_{i\in\mathcal{N}}$ and $\{b_{ij}\}_{i,j\in\mathcal{N}}$ define the parameter space for θ.

Our main results are formally stated as follows.

Theorem 1 (Regret of Order-Of-$T^{\frac{2}{3}}\log T$ Policy). *Suppose Assumptions 1 and 2 hold and the exact optimal assortments $S^*(\cdot)$ are computable under possible assortments $\mathcal{S}$. Let $\mu > 0$ be an arbitrary constant. There exist $\kappa_1, T_1 \in O(\mathrm{poly}(N))$ such that by letting $\tau = \lceil \mu T^{\frac{2}{3}}\log T\rceil$, $S_P = S^*(\cdot)$, and policy P_1 be defined by Algorithm 1, the regret associated with policy P_1 at any time $T \geqslant T_1$ is bounded as*

$$\mathrm{Reg}_{P_1}(T, \theta) \leqslant \kappa_1 T^{\frac{2}{3}}\log T,$$

where κ_1 and T_1 are constants independent of the MCC parameter θ.

Theorem 2 (Regret of Order-Of-$\log T$ Policy). *Suppose Assumptions 1* and *2 hold and the $\gamma\alpha$-optimal assortments $S^{\gamma\alpha}(\cdot)$ are computable under possible assortments $\mathcal{S}$ where $\alpha \in (0,1)$ and $\gamma \in (1, \frac{1}{\alpha})$. There exist $\psi, \kappa_2, T_2 \in O(\mathrm{poly}(N))$*

such that by letting $\tau = \lceil \psi \log T \rceil$, $S_P = S^{\gamma\alpha}(\cdot)$, *and policy* P_2 *be defined by Algorithm 1, the* α*-regret associated with policy* P_2 *at any time* $T \geqslant T_2$ *is bounded as*

$$\mathsf{Reg}^{\alpha}_{P_2}(T, \theta) \leqslant \kappa_2 \log T,$$

where ψ, κ_2, *and* T_2 *are constants independent of the MCC parameter* θ.

We next remark on motivation and weakness of Assumption 2, justify the use of $\gamma\alpha$-optimal assortments in Theorem 2, and discuss the regret upper bounds in Theorems 1 and 2.

Remark 2 (On Assumption 2). The attraction parameters $\{a_i\}_{i \in \mathcal{N}}$ and $\{b_{ij}\}_{i,j \in \mathcal{N}}$ play the same role as the attraction parameters in the MNL model: They are introduced for convenience of estimation, and $a_i = \frac{\lambda_i}{\lambda_0}\,(i \in \mathcal{N}_+)$, $b_{ij} = \frac{\rho_{ij}}{\rho_{i0}}(i \in \mathcal{N}, j \in \mathcal{N}_+)$. Because $a_0 = b_{i0} = 1$ for all $i \in \mathcal{N}$, the fractions defining Θ have strictly positive denominators and are well defined. The main analysis can be similarly developed to other closed parameter spaces satisfying conditions (i) and (ii) under Assumption 2. These two conditions are not strong. Condition (i) has been discussed in Section 2.1, which is necessary for model identifiability as shown in Appendix C.1. Condition (ii) essentially requires that the arrival probabilities $\{\lambda_i\}_{i \in \mathcal{N}_+}$ and the transit-to-no-purchase probabilities $\{\rho_{i0}\}_{i \in \mathcal{N}}$ are lower bounded. (For counterexample, consider $i, j \in \mathcal{N}, i \neq j$. $\rho_{i0} = \frac{1}{\sum_{k \in \mathcal{N}_+} b_{ik}}$ will approach 0 if b_{ij} approaches ∞. $\lambda_i = \frac{a_i}{\sum_{k \in \mathcal{N}_+} a_k}$ will approach zero if a_j approaches ∞ or a_i approaches zero.) The lower bounded $\{\rho_{i0}\}_{i \in \mathcal{N}}$ avoids "infinite loops" in the MCC model because the customer's probability of being absorbed into the no-purchase option is above zero after each transition. Overall, we only require $O(N)$ entries of transition matrix ρ lower bounded and allow ρ to be sparse (e.g., transitions only occurring within the same product category). This condition is significantly weaker than Şimşek and Topaloglu (2018) assuming that all $\Omega(N^2)$ entries of ρ and λ are lower bounded.

Remark 3 (Use of $\gamma\alpha$-Optimal Assortments). In Theorem 2, we provide $\gamma\alpha$-optimal assortments in the exploitation phase while we analyze the α-regret and use the α-optimal assortments as benchmark. The improvement of approximation ratios from α to $\gamma\alpha$ is insignificant because the approximation ratios of near-optimal assortments often have open ranges and thereby, we can find a valid improvement factor $\gamma > 1$ for all ratio α. For example, Désir et al. (2020) and Udwani (2021), respectively, gave an $(1/2 - \varepsilon)$-approximation algorithm for assortment optimization under cardinality or capacity constraints with $\varepsilon > 0$ being an arbitrary small constant. Let the benchmark take 0.45-optimal assortments (i.e., $\varepsilon = 0.05$). Then 0.49-optimal assortments are obtainable. Here we can set the predefined constant $\alpha = 0.45$ and the improvement factor $\gamma = \frac{0.49}{0.45}$. The α-regret bound in Theorem 2 depends on factor γ as $\kappa_2, T_2 \in \tilde{O}\left(\frac{1}{(\gamma-1)^2}\right)$; see Section 4.4. In fact, the use of suboptimality gap $\Delta_{\min}$ in explore-then-commit learning literature (Sauré and Zeevi 2013, Gallego and Lu 2021) can be viewed as a special case of analyzing α-regret while providing $\gamma\alpha$-optimal solutions where $\alpha = 1 - \frac{\Delta_{\min}}{2r_{\max}}$ and $\gamma = \frac{1}{\alpha}$. A detailed discussion is in Section 5.

Remark 4 (On *poly*(N) Regret and Its Unconstrained Relaxation). Our regret upper bounds scale polynomially in N. This is ideal because it avoids a combinatorial complexity of assortment selection. The polynomial order of N in both Theorems 1 and 2 has two sources: first, in the exploration phase, we use $O(N^2)$ different assortments, that is, $|\mathscr{S}_0 \cup \tilde{\mathscr{S}}_0| \in O(N^2)$. Second, when we estimate the MCC parameter θ, we solve a system of linear Equations (E-4)–(E-5) with $O(N^2)$ rows and $O(N^2)$ columns. If the problem is unconstrained, that is, $\mathscr{S} = 2^{\mathcal{N}}$, then both the number of exploration assortments and the size of linear equations for estimating θ will be reduced significantly; see the simplified algorithm in Appendix E. Consequently, the order of N in regret bounds is also reduced. For example, the regret coefficient κ_1 in Theorem 1 will reduce its order from $O(N^{3.5})$ to $O(N^{2.5})$; see Expressions (6) versus (E.4).

Remark 5 (Estimation Efficiency and Consistency). Technically, to estimate the MCC parameters, we devise a new batch-to-batch sampling method and derive the desired concentrations bounds involving novel techniques to approximate stationary distributions via matrix operations; see Lemma 5. Compared with the state-of-the-art EM method for parameter estimation (Şimşek and Topaloglu 2018) that suffers from unguaranteed convergence and increasing computational burden in T, our consistent estimators enjoy sub-Gaussian concentration bounds and superior computational times with almost negligible dependence on T. In our computational experiments, our estimators are at least 1,000 times more efficient.

In the static assortment optimization, the exact optimal assortment is computable only for the unconstrained setting (Blanchet et al. 2016) or the MCC model's special cases, for example, the MNL model under a TU constraint set (Davis et al. 2013) and the general attraction model under a cardinality constraint set (Wang 2013). The most related lower bound result is $\Omega(\sqrt{T})$ for a constrained MNL model (Chen and Wang 2018). There is a gap between our $T^{2/3}\log T$ upper bound and the known $\sqrt{T}$ lower bound. We would like to make the following comments.

i. The estimation for MCC parameters is far more challenging than traditional MNL models (because it captures both the initial preference and the substitution between each pair of products and requires $\Omega(N)$ different assortments for parameter estimation). It is an open research question if the lower bound still remains on the order of $\sqrt{T}$, which can be worthy of further investigation.

ii. There could be variants of our algorithm (e.g., involving multiple batches of exploration and exploitation) that could potentially improve our upper bound, but the design and analysis would be extremely challenging. For instance, there are two key difficulties in using upper confidence bound (UCB) based multi-armed bandit techniques. First, one has to carefully define the notion of "arms" and construct the reward confidence radius associated with each arm. The products cannot be treated as independent arms because the customer choice observed on offering a product i also depends on other products in the same assortment S. It is also computationally inefficient to treat all possible assortments $S \in \mathscr{S}$ as arms. We cannot define arms based on each product's attraction value as in an MNL model (Agrawal et al. 2019) because customers' choice probabilities are jointly decided by the arrival probabilities of each product and the transition probabilities between products. Second, given that the constrained assortment optimization problem often lacks computable solutions (in polynomial time), we can expect that it is even harder to compute the so-called "optimistic" constrained assortment over the entire parameter space defined via confidence radii in every iteration.

iii. Moreover, a simple explore-then-commit strategy is arguably more deployable in real-world settings (in terms of implementation and communication to various nontechnical stakeholders). Our upper bound matches the $T^{2/3}$ lower regret bound of batched bandits under explore-then-commit strategies (Perchet et al. 2016) up to a logarithmic factor.

In the "more general" static assortment optimization, the exact optimal assortment under the MCC model is not typically computable, and only near-optimal assortments are available. Because a retailer in practice has to resort to near-optimal assortments rather than exact optimal assortments (which are not computable) in such a scenario, the normal regret based on exact optimal assortments is almost impossible to quantify. Instead, our α-regret is a more practical measure for assortment selection policies. Our analysis of α-regret-based learning algorithms departs from previous assortment studies based on a normal regret definition (Agrawal et al. 2019, Chen et al. 2021a, Gallego and Lu 2021). Because the α-regret uses an α-optimal assortment as our clairvoyant policy, the designed learning algorithm converges to such a weaker benchmark faster than that in Theorem 1, and the derived regret bound is improved. In Theorem 2, we find the optimality gap of near-optimal assortments introduces an instance-independent "regret-free region" (Lemma 4) surrounding the true parameters, and the convergence rate of our batch-to-batch estimator to this region is on the order of $\log T$, resulting in a better regret upper bound.

4.1. Proof Outline

We outline proofs of Theorems 1 and 2 by showing sub-Gaussian concentration bounds of $\tilde{\theta}^{\tau}$ and linear bounds of *exploitation optimality gap w.r.t.* $\tilde{\theta}^{\tau}$. Here the exploitation optimality gap denotes the single-sale revenue difference between the exact optimal assortment calculated with the true parameter θ and that calculated with the estimator $\tilde{\theta}^{\tau}$, i.e., $|r(S^*(\theta);\theta) - r(S^*(\tilde{\theta}^{\tau});\theta)|$. A flowchart is shown in Figure 1.

First, Section 4.2 studies the smoothness of the objective function through the lens of Lipschitz continuity. For any assortment S, the single-sale revenue $r(S;\theta)$ is Lipschitz continuous $w.r.t.$ θ. Moreover, we can find a Lipschitz constant C_L that is independent of S and decreases as $\mathscr{S}$ becomes smaller (Lemma 2). The Lipschitz continuity of single-sale revenue implies a linear bound of the exploitation optimality gap w.r.t. the estimator $\tilde{\theta}^{\tau}$ ((i) of Lemma 4). Moreover, when analyzing near-optimal assortments and α-regret, the Lipschitz continuity introduces an instance-independent regret-free region: any $\gamma\alpha$-optimal assortment based on a $\tilde{\theta}^{\tau}$ that is within the d^*-neighborhood of true parameter θ is α-optimal under θ and thereby the associated α-regret is zero. This distance d^* is independent of θ ((ii) of Lemma 4).

Next, Section 4.3 shows the consistency and sub-Gaussian concentration bounds of the estimator $\tilde{\theta}^{\tau}$ (Lemma 5). We show the following properties of estimators: (i) $\tilde{\theta}^{\tau}$ is Lipschitz continuous w.r.t. the estimated choice probability $\hat{\pi}^{\tau}(i,S)$ for any $i \in \mathcal{N}_+$ and $S \in \mathscr{S}_0 \cup \tilde{\mathscr{S}}_0$ due to the structure of (E-1)–(E-9); and (ii) $\hat{\pi}^{\tau}(i,S)$ for $i \in \mathcal{N}_+, S \in \mathscr{S}_0 \cup \tilde{\mathscr{S}}_0$ is a consistent estimator. Following empirical distributions of independent samples, we derive concentration bounds for $\hat{\pi}^{\tau}(i,S)$ and $\tilde{\theta}^{\tau}$.

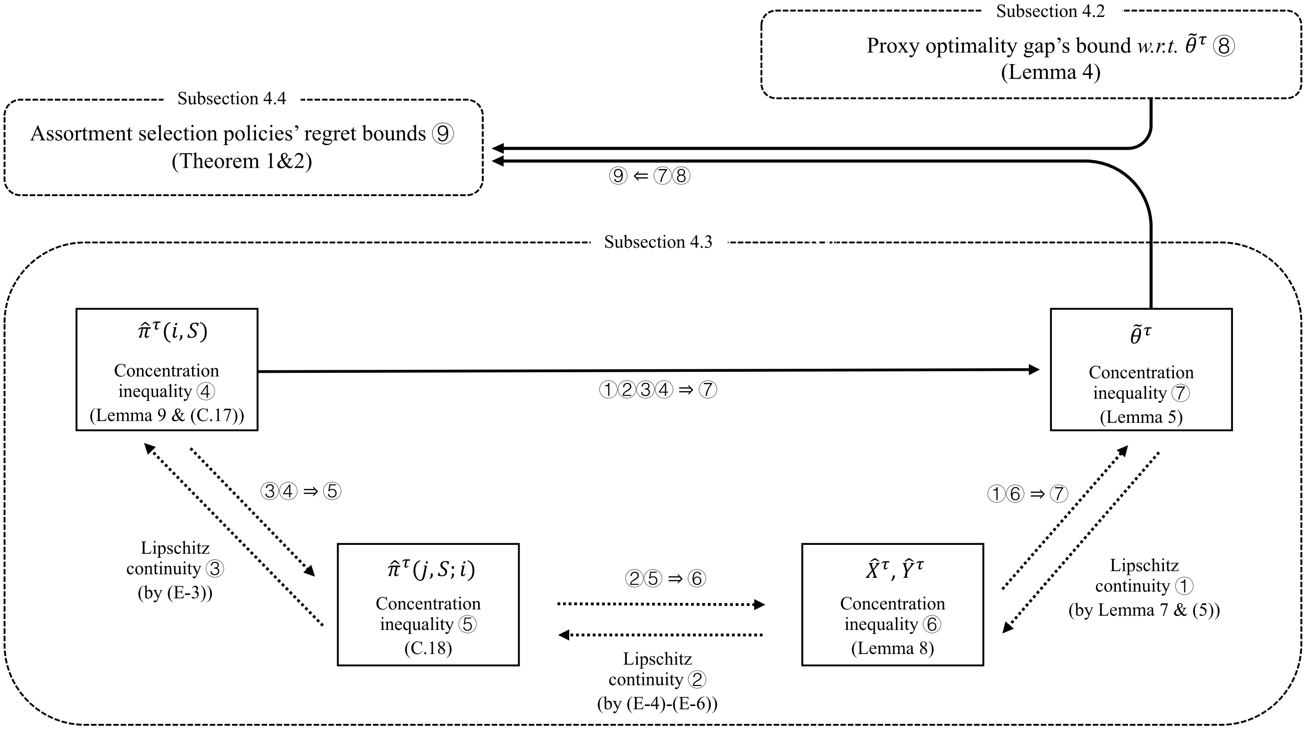
Finally, we derive regret bounds in Section 4.4 using the linear bounds and concentration bounds obtained previously. These policy regret bounds are obtained by balancing and minimizing the cumulative regret of the exploration and exploitation phases.

4.2. Lipschitz Continuity

We first show that the single-sale revenue $r(S;\theta)$ is the unique solution of (SS-1)–(SS-3) and then derive its Lipschitz constant w.r.t. θ, based on which we prove that the exploitation optimality gap has a linear bound w.r.t. $\tilde{\theta}^{\tau}$. For convenience of analysis, we define the following constants. First, we define the following lower bound for the maximal single-sale revenue in the worst case of $\theta \in \Theta$:

$$\underline{r} := \max_{S\in\mathscr{S}} \sum_{i\in S} \frac{\underline{a}_i r_i}{1+\sum_{k\in\mathcal{N}} \overline{a}_k} = \frac{\max_{S\in\mathscr{S}} \sum_{i\in S} \underline{a}_i r_i}{1+\sum_{k\in\mathcal{N}} \overline{a}_k}.$$

Figure 1. Flowchart of Proving Theorems 1 and 2 for Assortment Selection Policies' Regret Bounds



Notes. The numbers in circles index the results presented in this figure. Solid arrows highlight main proofs. Dotted arrows represent supplementary proofs.

Indeed, for any assortment $S \in \mathscr{S}$ and product $i \in S$, r_i is the revenue for selling one unit of product i, and $\frac{\underline{a}_i}{1+\sum_{k\in\mathcal{N}}\bar{a}_k} \leqslant \lambda_i$ is a lower bound for the probability of product i being purchased. Thus, $\underline{r}$ returns a lower bound of maximal revenue. Next, we define the minimal absorbing probability:

$$\eta := \min_{i\in\mathcal{N}} \inf_{\theta\in\Theta} \rho_{i0} = \frac{1}{1+\max_{i\in\mathcal{N}}\sum_{k\in\mathcal{N}}\bar{b}_{ik}},$$

that is, the minimal transition probability to state 0 (no purchase) from any other state (product).

4.2.1. Computation of Single-Sale Revenue. We verify that (SS-1)–(SS-3) return the choice probabilities and single-sale revenues.

Lemma 1. *For the MCC model* $(\mathcal{N}, r, \theta)$ *with a parameter space* Θ, *equations* (SS-1)–(SS-3) *have a unique solution and return the average visit times, the choice probabilities, and the single-sale revenue for every* $\theta \in \Theta$ *and* $S \subseteq \mathcal{N}$.

Proof. The average visit times and choice probabilities are well defined according to the following observation. For any $\theta = \mathrm{vec}(\{\lambda_i\}_{i\in\mathcal{N}} \cup \{\rho_{ij}\}_{i,j\in\mathcal{N}}) \in \Theta$ and $S \subseteq \mathcal{N}$, we have $\rho_{00} = 1$ and $\rho_{i0} \geqslant \eta > 0$ for all $i \in \mathcal{N}$ by the definition of η. Thus, state 0, that is, no purchase, is an absorbing state, and the probability of being not absorbed after t transitions is upper bounded by $(1-\eta)^t$. Thus, the average visit times and choice probabilities are well defined and finite.

Equation (1) in Şimşek and Topaloglu (2018) shows that the average visit times and choice probabilities should satisfy (SS-1)–(SS-2). Then to prove Lemma 1, it is sufficient to show that the system of linear equations (SS-1) has a unique solution. Let us write (SS-1) into the matrix notation with

$$\begin{aligned} &\boldsymbol{u} := (u(i,S;\theta))_{i\in S}, && \boldsymbol{\lambda} := (\lambda_i)_{i\in S}, && \boldsymbol{P} := (\rho_{ij})_{i\in\mathcal{N}\setminus S, j\in S}, \\ &\bar{\boldsymbol{u}} := (u(i,S;\theta))_{i\in\mathcal{N}\setminus S}, && \bar{\boldsymbol{\lambda}} := (\lambda_i)_{i\in\mathcal{N}\setminus S}, && \bar{\boldsymbol{P}} := (\rho_{ij})_{i,j\in\mathcal{N}\setminus S}. \end{aligned} \tag{1}$$

Then (SS-1) can be written as

$$\bar{\boldsymbol{u}} = \bar{\boldsymbol{\lambda}} + \bar{\boldsymbol{P}}^\top \bar{\boldsymbol{u}}, \tag{SS-M1}$$

$$\boldsymbol{u} = \boldsymbol{\lambda} + \boldsymbol{P}^\top \bar{\boldsymbol{u}}, \tag{SS-M2}$$

$$u(0,S;\theta) = \left(1 - \sum_{i\in\mathcal{N}} \lambda_i\right) + \sum_{i\in\mathcal{N}\setminus S} u(i,S;\theta)\rho_{i0}. \tag{SS-M3}$$

Thus, (SS-1) will have a unique solution if $(\boldsymbol{I} - \bar{\boldsymbol{P}}^\top)$ is invertible. For every vector $\boldsymbol{x} \in \mathbb{R}_+^{|\mathcal{N}\setminus S|}, \|\bar{\boldsymbol{P}}^\top \boldsymbol{x}\|_1 \leqslant (1-\eta)\|\boldsymbol{x}\|_1$ because $\sum_{j\in\mathcal{N}\setminus S}\rho_{ij} \leqslant 1 - \rho_{i0} \leqslant 1-\eta$ for all $i \in \mathcal{N}\setminus S$. Then for every $\boldsymbol{x} \in \mathbb{R}_+^{|\mathcal{N}\setminus S|}$ and thereby every $\boldsymbol{x} \in \mathbb{R}^{|\mathcal{N}\setminus S|}$, we have $\lim_{t\to\infty}(\bar{\boldsymbol{P}}^\top)^t \boldsymbol{x} = \boldsymbol{0}$. Thus, $(\boldsymbol{I} - \bar{\boldsymbol{P}}^\top)^{-1}$ exists and equals $\boldsymbol{I}+\bar{\boldsymbol{P}}^\top+(\bar{\boldsymbol{P}}^\top)^2+\cdots$ Thus, the solution of (SS-1) is unique. □

Last, recall (SS-M1)–(SS-M2) and let $\boldsymbol{r}:=(r_i)_{i\in S}$. Then we can write the single-sale revenue $r(S;\theta)$ as a closed-form function

$$r(S;\theta)=\boldsymbol{r}^\top[\boldsymbol{\lambda}+\boldsymbol{P}^\top(\boldsymbol{I}-\bar{\boldsymbol{P}}^\top)^{-1}\bar{\boldsymbol{\lambda}}]. \tag{SS-M4}$$

4.2.2. Lipschitz Continuity and Exploitation Optimality Gap. We now prove the Lipschitz continuity of the single-sale revenue $r(S;\theta)$ w.r.t. θ.

Lemma 2. *For the MCC model $(\mathcal{N},r,\theta)$ with parameter space Θ and possible assortments $\mathscr{S}$, there exists a Lipschitz constant C_L such that $|r(S;\theta_2)-r(S;\theta_1)|\leqslant C_L\,\|\theta_2-\theta_1\|_1$ for every $\theta_1,\theta_2\in\Theta$ and $S\in\mathscr{S}$.*

We use the following inequalities, the detailed proof of which is included in Appendix C.

Lemma 3. *For all $\boldsymbol{y}\in\mathbb{R}_+^{|\mathcal{N}\setminus S|}$, $S\subseteq\mathcal{N}$, and $\theta\in\Theta$, we have that*

$$\|(\boldsymbol{I}-\bar{\boldsymbol{P}})^{-1}\boldsymbol{y}\|_\infty\leqslant\frac{\|\boldsymbol{y}\|_\infty}{\eta(\mathscr{S})}, \tag{2}$$

$$\|(\boldsymbol{I}-\bar{\boldsymbol{P}}^\top)^{-1}\boldsymbol{y}\|_\infty\leqslant\frac{\|\boldsymbol{y}\|_1}{\eta(\mathscr{S})}, \tag{3}$$

where the constant $\eta(\mathscr{S})$ is given by

$$\eta(\mathscr{S}):=\inf_{S\in\mathscr{S},\theta\in\Theta,i\in\mathcal{N}\setminus S}\rho_{i0}=\min_{S\in\mathscr{S},i\in\mathcal{N}\setminus S}\frac{1}{1+\sum_{k\in\mathcal{N}}\bar{b}_{ik}}.$$

Here $\eta(\mathscr{S})\geqslant\eta$ by definitions, and $\eta(\mathscr{S})$ can be interpreted as the lower bound of transition probabilities to state 0 (i.e., no purchase) from any other state (product) outside $S\in\mathscr{S}$. Thus, a small $\mathscr{S}$ due to strong assortment constraints may yield an $\eta(\mathscr{S})$ significantly greater than η.

Proof of Lemma 2. *Lipschitz constant induced by partial derivatives.* Given an assortment $S\in\mathscr{S}$, $r(S;\theta)$ is differentiable w.r.t. the parameter $\theta\in\Theta$ because $r(S;\theta)$ is a composite function of θ produced by (SS-1)–(SS-3).

Formally, recall (SS-M4). Because $\boldsymbol{\lambda}$, $\bar{\boldsymbol{\lambda}}$, $\boldsymbol{P}$, and $\bar{\boldsymbol{P}}$ are differentiable w.r.t. $\theta=\operatorname{vec}(\{\lambda_i\}_{i\in\mathcal{N}}\cup\{\rho_{ij}\}_{i,j\in\mathcal{N}})\in\Theta$ by definitions in (1), the composite function $r(S;\theta)$ is also differentiable w.r.t. θ.

We can define a tight Lipschitz constant C_L^* as the largest partial derivative of $r(S;\theta)$ w.r.t. elements of θ across all $S\in\mathscr{S}$ and $\theta\in\Theta$:

$$C_L^*:=\sup_{S\in\mathscr{S},\theta\in\Theta}\left\{\max_{i\in\mathcal{N}}\left|\frac{\partial r(S;\theta)}{\partial\lambda_i}\right|,\ \max_{(i,j)\in\mathcal{N}^2\setminus I_0}\left|\frac{\partial r(S;\theta)}{\partial\rho_{ij}}\right|\right\}.$$

The trivial parameters $\rho_{ij}\equiv0$ for all $(i,j)\in I_0$ and thus, they are immaterial to C_L^*.

4.2.2.1. Divide Partial Derivatives. We divide the partial derivatives inside the definition of C_L^* into five groups and calculate their bounds. We rewrite $C_L^*=\max\{C_{L_1},C_{L_2},C_{L_3},C_{L_4},C_{L_5}\}$, where each parameter is defined as

$$C_{L_1}:=\sup_{S\in\mathscr{S},\theta\in\Theta}\left\{\max_{i\in S}\left|\frac{\partial r(S;\theta)}{\partial\lambda_i}\right|\right\},$$

$$C_{L_2}:=\sup_{S\in\mathscr{S},\theta\in\Theta}\left\{\max_{i\in\mathcal{N}\setminus S}\left|\frac{\partial r(S;\theta)}{\partial\lambda_i}\right|\right\},$$

$$C_{L_3}:=\sup_{S\in\mathscr{S},\theta\in\Theta}\left\{\max_{(i,j)\in((\mathcal{N}\setminus S)\times\mathscr{S})\setminus I_0}\left|\frac{\partial r(S;\theta)}{\partial\rho_{ij}}\right|\right\},$$

$$C_{L_4}:=\sup_{S\in\mathscr{S},\theta\in\Theta}\left\{\max_{(i,j)\in(\mathcal{N}\setminus S)^2\setminus I_0}\left|\frac{\partial r(S;\theta)}{\partial\rho_{ij}}\right|\right\},$$

$$C_{L_5}:=\sup_{S\in\mathscr{S},\theta\in\Theta}\left\{\max_{(i,j)\in(S\times\mathcal{N})\setminus I_0}\left|\frac{\partial r(S;\theta)}{\partial\rho_{ij}}\right|\right\}.$$

Here $C_{L_5}\equiv0$ because, for all $(i,j)\in S\times\mathcal{N}$, $r(S;\theta)$ is independent of ρ_{ij} and $\partial r(S;\theta)/\partial\rho_{ij}=0$.

4.2.2.2. Bound Partial Derivatives by Groups. Next, we bound $C_{L_1},C_{L_2},C_{L_3},C_{L_4}$. Define $\bar{r}(\mathscr{S}):=\max_{S\in\mathscr{S},i\in S}r_i\leqslant r_{\max}$ as the maximal single-product revenue under possible assortments $\mathscr{S}$. According to (1) and (SS-M1)–(SS-M4), C_{L_1} satisfies

$$C_{L_1}=\sup_{S\in\mathscr{S},\theta\in\Theta}\left\|\frac{\partial r(S;\theta)}{\partial\boldsymbol{\lambda}}\right\|_\infty=\sup_{S\in\mathscr{S},\theta\in\Theta}\|\boldsymbol{r}\|_\infty=\bar{r}(\mathscr{S}).$$

Similarly, C_{L_2}-C_{L_4} can be bounded by

$$\begin{aligned}
C_{L_2}&=\sup_{S\in\mathscr{S},\theta\in\Theta}\left\|\frac{\partial r(S;\theta)}{\partial\bar{\boldsymbol{\lambda}}}\right\|_\infty=\sup_{S\in\mathscr{S},\theta\in\Theta}\|(\boldsymbol{I}-\bar{\boldsymbol{P}})^{-1}\boldsymbol{P}\boldsymbol{r}\|_\infty\\
&\leqslant\sup_{S\in\mathscr{S},\theta\in\Theta}\frac{\|\boldsymbol{P}\boldsymbol{r}\|_\infty}{\eta(\mathscr{S})}\leqslant\frac{\bar{r}(\mathscr{S})}{\eta(\mathscr{S})}, \qquad \text{(due to (2))}\\
C_{L_3}&=\sup_{S\in\mathscr{S},\theta\in\Theta}\left\|\frac{\partial r(S;\theta)}{\partial\boldsymbol{P}}\right\|_\infty=\sup_{S\in\mathscr{S},\theta\in\Theta}\|(\boldsymbol{I}-\bar{\boldsymbol{P}}^\top)^{-1}\bar{\boldsymbol{\lambda}}\boldsymbol{r}^\top\|_\infty\\
&=\sup_{S\in\mathscr{S},\theta\in\Theta}\{\|(\boldsymbol{I}-\bar{\boldsymbol{P}}^\top)^{-1}\bar{\boldsymbol{\lambda}}\|_\infty\cdot\|\boldsymbol{r}\|_\infty\}\\
&\leqslant\sup_{S\in\mathscr{S},\theta\in\Theta}\left\{\frac{\|\bar{\boldsymbol{\lambda}}\|_1}{\eta(\mathscr{S})}\cdot\bar{r}(\mathscr{S})\right\}\leqslant\frac{\bar{r}(\mathscr{S})}{\eta(\mathscr{S})}, \qquad \text{(due to (3))}\\
C_{L_4}&=\sup_{S\in\mathscr{S},\theta\in\Theta}\left\{\max_{(i,j)\in(\mathcal{N}\setminus S)^2\setminus I_0}\left|\frac{\partial r(S;\theta)}{\partial\rho_{ij}}\right|\right\}\\
&=\sup_{S\in\mathscr{S},\theta\in\Theta}\left\{\max_{(i,j)\in(\mathcal{N}\setminus S)^2\setminus I_0}|\bar{\boldsymbol{\lambda}}^\top(\boldsymbol{I}-\bar{\boldsymbol{P}})^{-1}\boldsymbol{E}_{ij}(\boldsymbol{I}-\bar{\boldsymbol{P}})^{-1}\boldsymbol{P}\boldsymbol{r}|\right\}\\
&\leqslant\sup_{S\in\mathscr{S},\theta\in\Theta}\{\|\bar{\boldsymbol{\lambda}}^\top(\boldsymbol{I}-\bar{\boldsymbol{P}})^{-1}\|_\infty\cdot\|(\boldsymbol{I}-\bar{\boldsymbol{P}})^{-1}\boldsymbol{P}\boldsymbol{r}\|_\infty\}\\
&\leqslant\sup_{S\in\mathscr{S},\theta\in\Theta}\left\{\frac{\|\bar{\boldsymbol{\lambda}}\|_1}{\eta(\mathscr{S})}\cdot\frac{\|\boldsymbol{P}\boldsymbol{r}\|_\infty}{\eta(\mathscr{S})}\right\}\leqslant\frac{\bar{r}(\mathscr{S})}{\eta^2(\mathscr{S})},
\end{aligned}$$

where the single-entry matrix $\boldsymbol{E}_{ij} \in \mathbb{R}^{|\mathcal{N}\setminus S|^2}$ has only one nonzero entry of value one, whose location is the same with that of ρ_{ij} in $\bar{\boldsymbol{P}}$, and the last row's inequality is due to (2) and (3) in Lemma 3.

Therefore, the tight Lipschitz constant C_L^* is bounded by

$$C_L^* = \max\{C_{L_1}, C_{L_2}, C_{L_3}, C_{L_4}, C_{L_5}\}$$
$$\leqslant \max\left\{\bar{r}(\mathscr{S}), \frac{\bar{r}(\mathscr{S})}{\eta(\mathscr{S})}, \frac{\bar{r}(\mathscr{S})}{\eta(\mathscr{S})}, \frac{\bar{r}(\mathscr{S})}{\eta^2(\mathscr{S})}, 0\right\} \leqslant \frac{\bar{r}(\mathscr{S})}{\eta^2(\mathscr{S})}.$$

Thus, Lemma 2 requires a Lipschitz constant of $\bar{r}(\mathscr{S})/\eta^2(\mathscr{S})$. □

Lemma 2 also implies a linear bound of the exploitation optimality gap for exact optimal assortments and a regret-free region for near-optimal assortments.

Lemma 4. *Consider the MCC model $(\mathcal{N}, r, \theta)$ with parameter space Θ and possible assortments $\mathscr{S}$. For any $\theta \in \Theta$ and any estimator $\hat{\theta} \in \Theta$, we have the following:*

i. ***(Exploitation optimality gap)*** *$|r(S^*(\hat{\theta}); \theta) - r(S^*(\theta); \theta)|$ is bounded by $2C_L\|\hat{\theta} - \theta\|_1$.*

ii. ***(Regret-free region)*** *If $\|\hat{\theta} - \theta\|_1 \leqslant \frac{\underline{r}\alpha(\gamma-1)}{C_L(1+\alpha)}$, then $S^{\gamma\alpha}(\hat{\theta}) \in \mathscr{S}^{\alpha}(\theta)$ where $\alpha \in (0, 1), \gamma \in (1, \frac{1}{\alpha})$. If $\|\hat{\theta} - \theta||_1 \leqslant \frac{\underline{r}(1-\alpha)}{C_L(1+\alpha)}$, then $S^*(\hat{\theta}) \in \mathscr{S}^{\alpha}(\theta)$. In other words, a $\gamma\alpha$-optimal assortment under $\hat{\theta}$ is an α-optimal assortment under θ if $\|\hat{\theta} - \theta||_1 \leqslant \frac{\underline{r}\alpha(\gamma-1)}{C_L(1+\alpha)}$; an exact-optimal assortment under $\hat{\theta}$ is an α-optimal assortment under θ if $\|\hat{\theta} - \theta\|_1 \leqslant \frac{\underline{r}(1-\alpha)}{C_L(1+\alpha)}$.*

Proof. The first statement (i) holds true because

$$\begin{aligned}
&|r(S^*(\hat{\theta}); \theta) - r(S^*(\theta); \theta)| = r(S^*(\theta); \theta) - r(S^*(\hat{\theta}); \theta)\\
&= [r(S^*(\theta); \theta) - r(S^*(\theta); \hat{\theta})] + [r(S^*(\theta); \hat{\theta}) - r(S^*(\hat{\theta}); \hat{\theta})]\\
&\quad + [r(S^*(\hat{\theta}); \hat{\theta}) - r(S^*(\hat{\theta}); \theta)]\\
&\leqslant [r(S^*(\theta); \theta) - r(S^*(\theta); \hat{\theta})] + [r(S^*(\hat{\theta}); \hat{\theta}) - r(S^*(\hat{\theta}); \theta)]\\
&\hspace{18em} (\text{by } r(S^*(\theta); \hat{\theta}) \leqslant r(S^*(\hat{\theta}); \hat{\theta}))\\
&\leqslant 2C_L\|\hat{\theta} - \theta\|_1. \hspace{12em} (\text{by Lemma 2}).
\end{aligned}$$

The first half of (ii) holds true because

$$\begin{aligned}
&\frac{r(S^{\gamma\alpha}(\hat{\theta}); \theta)}{r(S^*(\theta); \theta)} \geqslant \frac{r(S^{\gamma\alpha}(\hat{\theta}); \hat{\theta}) - C_L\|\hat{\theta} - \theta\|_1}{r(S^*(\theta); \hat{\theta}) + C_L\|\hat{\theta} - \theta\|_1}\\
&\geqslant \frac{r(S^{\gamma\alpha}(\hat{\theta}); \hat{\theta}) - C_L\|\hat{\theta} - \theta\|_1}{r(S^*(\hat{\theta}); \hat{\theta}) + C_L\|\hat{\theta} - \theta\|_1} \qquad (\text{by Lemma 2})\\
&\geqslant \frac{\frac{r(S^{\gamma\alpha}(\hat{\theta}); \hat{\theta})}{r(S^*(\hat{\theta}); \hat{\theta})} - \frac{C_L\|\hat{\theta} - \theta\|_1}{r(S^*(\hat{\theta}); \hat{\theta})}}{1 + \frac{C_L\|\hat{\theta} - \theta\|_1}{r(S^*(\hat{\theta}); \hat{\theta})}}\\
&\geqslant \frac{\gamma\alpha - \frac{C_L\|\hat{\theta} - \theta\|_1}{\underline{r}}}{1 + \frac{C_L\|\hat{\theta} - \theta\|_1}{\underline{r}}} \qquad (\text{by dividing } r(S^*(\hat{\theta}); \hat{\theta}),\\
&\hspace{16em} \text{and } r(S^*(\hat{\theta}); \hat{\theta}) \geqslant \underline{r})\\
&\geqslant \frac{\gamma\alpha - \frac{\alpha(\gamma-1)}{1+\alpha}}{1 + \frac{\alpha(\gamma-1)}{1+\alpha}} \geqslant \alpha. \quad \left(\text{by } \|\hat{\theta} - \theta\|_1 \leqslant \frac{\underline{r}\alpha(\gamma - 1)}{C_L(1 + \alpha)}\right).
\end{aligned}$$

Similarly, the second half of (ii) holds true because

$$\begin{aligned}
\frac{r(S^*(\hat{\theta}); \theta)}{r(S^*(\theta); \theta)} &\geqslant \frac{r(S^*(\hat{\theta}); \hat{\theta}) - C_L\|\hat{\theta} - \theta\|_1}{r(S^*(\theta); \hat{\theta}) + C_L\|\hat{\theta} - \theta\|_1}\\
&\geqslant \frac{r(S^*(\hat{\theta}); \hat{\theta}) - C_L\|\hat{\theta} - \theta\|_1}{r(S^*(\hat{\theta}); \hat{\theta}) + C_L\|\hat{\theta} - \theta\|_1} \quad (\text{by Lemma 2})\\
&\geqslant \frac{1 - \frac{C_L\|\hat{\theta} - \theta\|_1}{r(S^*(\hat{\theta}); \hat{\theta})}}{1 + \frac{C_L\|\hat{\theta} - \theta\|_1}{r(S^*(\hat{\theta}); \hat{\theta})}} \geqslant \frac{1 - \frac{C_L\|\hat{\theta} - \theta\|_1}{\underline{r}}}{1 + \frac{C_L\|\hat{\theta} - \theta\|_1}{\underline{r}}}\\
&\geqslant \frac{1 - \frac{1-\alpha}{1+\alpha}}{1 + \frac{1-\alpha}{1+\alpha}} \geqslant \alpha.\\
&\left(\text{by } r(S^*(\hat{\theta}); \hat{\theta}) \geqslant \underline{r}, \|\hat{\theta} - \theta\|_1 \leqslant \frac{\underline{r}(1 - \alpha)}{C_L(1 + \alpha)}\right). \quad \square
\end{aligned}$$

4.3. Estimator Consistency and Concentration Bounds

The estimation strategy in (E-1)–(E-9) provides a concentration inequality of estimator $\tilde{\theta}^{\tau}$.

Lemma 5 (Consistency and Sub-Gaussian Concentration Bounds). *Consider the MCC model $(\mathcal{N}, r, \theta)$ with the parameter space Θ. For every $\theta \in \Theta$, the estimator $\tilde{\theta}^{\tau}$ defined through (E-1)–(E-9) is consistent such that $\tilde{\theta}^{\tau} \overset{p}{\longrightarrow} \theta$ as τ goes to ∞, where "$\overset{p}{\longrightarrow}$" denotes convergence in probability. Moreover, there exist constants $\omega, \phi, \zeta_0 \in \mathbb{R}_{++}$ independent of θ such that for all $\zeta \in (0, \zeta_0 N)$,*

$$\mathbb{P}[\|\tilde{\theta}^{\tau} - \theta\|_1 > \zeta] \leqslant \phi N^2 e^{-\frac{\omega\tau\zeta^2}{N^7}}, \qquad \tau \in \mathbb{N}. \tag{4}$$

The detailed proof and how to obtain (ω, ϕ, ζ_0) are given in Appendix C. The main idea is to develop the following "chain of estimators" (Figure 1):

- (E-9) indicates that the distances from the two estimators $\tilde{\theta}^{\tau}$ and $\hat{\theta}^{\tau}$ to the true parameter θ satisfy

$$\|\tilde{\theta}^{\tau} - \theta\|_1 \leqslant \|\tilde{\theta}^{\tau} - \hat{\theta}^{\tau}\|_1 + \|\hat{\theta}^{\tau} - \theta\|_1 \leqslant 2\|\hat{\theta}^{\tau} - \theta\|_1. \tag{5}$$

- (E-1) and (E-7)–(E-8) indicate that $\hat{\theta}^{\tau}$ is Lipschitz continuous *w.r.t.* the intermediate variables $(\hat{X}^{\tau}, \hat{Y}^{\tau})$.
- (E-4)–(E-6) indicate that the intermediate variables $(\hat{X}^{\tau}, \hat{Y}^{\tau})$ are Lipschitz continuous w.r.t. the estimated conditional choice probabilities $\{\hat{\pi}^{\tau}(j, S|i) | i, j \in \mathcal{N}_+, S \in \mathscr{S}_0\}$.
- (E-3) indicates that the estimated conditional choice probabilities $\{\hat{\pi}^{\tau}(j, S|i) | i, j \in \mathcal{N}_+, S \in \mathscr{S}_0\}$ are Lipschitz continuous *w.r.t.* the estimated choice probabilities $\{\hat{\pi}^{\tau}(i, S) | i \in \mathcal{N}_+, S \in \mathscr{S}_0 \cup \tilde{\mathscr{S}}_0\}$.
- For any fixed $S \in \mathscr{S}_0 \cup \tilde{\mathscr{S}}_0$, by the definition of (E-2), the estimated choice probabilities $\{\hat{\pi}^{\tau}(i, S)\}_{i \in \mathcal{N}_+}$ form an empirical distribution and thus follow the Dvoretzky–

Kiefer–Wolfowitz inequality (Kosorok 2006), which is a sub-Gaussian concentration bound. Moreover, $\{\hat{\pi}^\tau(i,S) \mid i \in \mathcal{N}_+, S \in \mathscr{S}_0 \cup \tilde{\mathscr{S}}_0\}$ are consistent.

Applying the chain rule to these estimators, $\tilde{\theta}^\tau$ is a consistent estimator of the MCC parameter θ. Moreover, combining the above Lipschitz continuity properties along the estimator chain, we can obtain a concentration inequality of $\tilde{\theta}^\tau$ based on that of $\{\hat{\pi}^\tau(i,S) \mid i \in \mathcal{N}_+, S \in \mathscr{S}_0 \cup \tilde{\mathscr{S}}_0\}$.

Remark 6 (Sampling Strategy in Exploration Phase). We elaborate on the rationale behind assigning the exploration assortments $\mathscr{S}_0 \cup \tilde{\mathscr{S}}_0$ equal weights to offer in Algorithm 1. Our MCC parameter estimator $\tilde{\theta}^\tau$ is constructed as a highly nonlinear function of choice probability estimators $\{\hat{\pi}^\tau(i,S)\}_{i\in\mathcal{N}_+, S\in\mathscr{S}_0\cup\tilde{\mathscr{S}}_0}$; see (E-1)–(E-9), which first constructs a set of linear equations using both the MCC parameter and the choice probabilities, and then inverses the coefficient matrix to recover the MCC parameter. To obtain an accurate estimator $\tilde{\theta}^\tau$ from the exploration phase, Algorithm 1 minimizes the supremum error of choice probability estimators $\{\hat{\pi}^\tau(i,S)\}_{i\in\mathcal{N}_+, S\in\mathscr{S}_0\cup\tilde{\mathscr{S}}_0}$. By the Dvoretzky–Kiefer–Wolfowitz inequality, we observe that the concentration bound of any choice probability estimator $\hat{\pi}^\tau(i,S)$ is completely determined by the frequency of assortment S offered and is independent of the true choice probability $\pi(i,S)$; see (C.17) in Appendix C.3. This suggests that offering the exploration assortments with equal weights is as efficient as alternative sampling strategies that minimize the supremum error of estimators $\{\hat{\pi}^\tau(i,S)\}_{i\in\mathcal{N}_+, S\in\mathscr{S}_0\cup\tilde{\mathscr{S}}_0}$.

4.4. Proof of the Performance of FastLinETC

4.4.1. Regret for the Exact Optimal Assortment. We prove the regret bounds in Theorem 1 and show that the following constants T_1 and κ_1 in these bounds are independent of $\theta \in \Theta$.

$$\kappa_1 := 2r_{\max}\mu\Big(1+\frac{\phi N^2}{\mu^2}\Big) + 3\omega^{-\frac{1}{2}}C_L N^3\Big(1+\frac{3N}{\mu}+2N\Big)^{\frac{1}{2}},$$
$$T_1 := \max\left\{2, 2^{\frac{3}{2}}\mu^{-\frac{3}{2}},\ \zeta_0^{-3}\omega^{-\frac{3}{2}}N^6\Big(1+\frac{3N}{\mu}+2N\Big)^{\frac{3}{2}}\right\}. \qquad (6)$$

Proof. Let us define $\phi^* := \phi N^2, \omega^* := \omega N^{-7}, \zeta_0^* := \zeta_0 N$ so that (4) can be written as

$$\mathbb{P}[\|\tilde{\theta}^\tau - \theta\|_1 > \zeta] \leqslant \phi^* e^{-\omega^*\tau\zeta^2}, \qquad \zeta \in (0, \zeta_0^*), \tau \in \mathbb{N}, \quad (7)$$

and we have $\kappa_1 \geqslant 2r_{\max}\mu + 2r_{\max}\mu^{-1}\phi^* + 3(\frac{3}{\mu}+2)^{\frac{1}{2}}\omega^{*-\frac{1}{2}}C_L$, $T_1 \geqslant (\frac{3}{\mu}+2)^{\frac{3}{2}}\zeta_0^{*-3}\omega^{*-\frac{3}{2}}$.

Because $T \geqslant T_1 \geqslant \max\left\{2, 2^{\frac{3}{2}}\mu^{-\frac{3}{2}}\right\}$, we have $\tau = \lceil \mu T^{\frac{2}{3}}\log T \rceil \geqslant \|\mu T_1^{\frac{2}{3}} \log T_1\| \geqslant \|2\log 2\| \geqslant 2$, which further indicates $\sqrt{\frac{\log\tau}{\omega^*\tau}} > 0$. In addition, we have

$$\begin{aligned}\sqrt{\frac{\log\tau}{\omega^*\tau}} &= \sqrt{\frac{\log\lceil\mu T^{\frac{2}{3}}\log T\rceil}{\omega^*\|\mu T^{\frac{2}{3}}\log T\|}} \leqslant \sqrt{\frac{\log(2\mu T^{\frac{2}{3}}\log T)}{\omega^*\mu T^{\frac{2}{3}}\log T}} \leqslant \sqrt{\frac{\log\mu T^3}{\omega^*\mu T^{\frac{2}{3}}\log T}}\\ &\leqslant \sqrt{\frac{3\log T+\log\mu}{\omega^*\mu T^{\frac{2}{3}}\log T}} \leqslant \sqrt{\frac{3\log T}{\omega^*\mu T^{\frac{2}{3}}\log T}+\frac{\log\mu}{\omega^*\mu T^{\frac{2}{3}}\log T}}\\ &\leqslant \sqrt{\frac{3}{\omega^*\mu T^{\frac{2}{3}}}+\frac{2}{\omega^* T^{\frac{2}{3}}}} \leqslant \Big(\frac{3}{\mu}+2\Big)^{\frac{1}{2}}\cdot\omega^{*-\frac{1}{2}}T^{-\frac{1}{3}} \leqslant \zeta_0^*.\\ &\left(\text{due to} T \geqslant T_1 \geqslant \Big(\frac{3}{\mu}+2\Big)^{\frac{3}{2}}\zeta_0^{*-3}\omega^{*-\frac{3}{2}}\right).\end{aligned} \qquad (8)$$

Because $\sqrt{\frac{\log\tau}{\omega^*\tau}} \in (0,\zeta_0^*)$, we can plug $\zeta = \sqrt{\frac{\log\tau}{\omega^*\tau}}$ into (7), which yields

$$\mathbb{P}\left[\|\tilde{\theta}^\tau - \theta\|_1 > \sqrt{\frac{\log\tau}{\omega^*\tau}}\right] \leqslant \phi^*\tau^{-1}. \qquad (9)$$

Then the regret of policy P_1 can be divided into three parts: (i) for $t \leqslant \tau$, the regret associated with every single customer is bounded by $r_{\max}$; (ii) for $t > \tau$ such that $\|\tilde{\theta}^\tau - \theta\|_1 > \sqrt{\frac{\log\tau}{\omega^*\tau}}$, the regret associated with every single customer is also bounded by $r_{\max}$; and (iii) for $t > \tau$ such that $\|\tilde{\theta}^\tau - \theta\|_1 \leqslant \sqrt{\frac{\log\tau}{\omega^*\tau}}$, the regret associated with every single customer is bounded by $2C_L\sqrt{\frac{\log\tau}{\omega^*\tau}}$ due to Lemma 4. Therefore, the regret of policy P_1 is bounded by

$$\begin{aligned}\mathrm{Reg}_{P_1}(T,\theta) &\leqslant r_{\max}\tau + r_{\max}(T-\tau)\cdot\mathbb{P}\left[\|\tilde{\theta}^\tau-\theta\|_1 > \sqrt{\frac{\log\tau}{\omega^*\tau}}\right]\\ &\quad + 2C_L\sqrt{\frac{\log\tau}{\omega^*\tau}}(T-\tau)\cdot\mathbb{P}\left[\|\tilde{\theta}^\tau-\theta\|_1 \leqslant \sqrt{\frac{\log\tau}{\omega^*\tau}}\right]\\ &\leqslant r_{\max}\tau + r_{\max}(T-\tau)\cdot\phi^*\tau^{-1}\\ &\quad + 2C_L\sqrt{\frac{\log\tau}{\omega^*\tau}}\cdot(T-\tau) && \text{(due to (9))}\\ &\leqslant r_{\max}\tau + r_{\max}\phi^* T\tau^{-1} + 2C_L T\sqrt{\frac{\log\tau}{\omega^*\tau}}\\ &\leqslant r_{\max}\lceil\mu T^{\frac{2}{3}}\log T\rceil + r_{\max}\phi^* T\lceil\mu T^{\frac{2}{3}}\log T\rceil^{-1}\\ &\quad + 2C_L T\sqrt{\frac{\log\tau}{\omega^*\tau}} && (\text{due to } \tau = \lceil\mu T^{\frac{2}{3}}\log T\rceil)\\ &\leqslant 2r_{\max}\mu T^{\frac{2}{3}}\log T + r_{\max}\phi^* T(\mu T^{\frac{2}{3}}\log T)^{-1}\\ &\quad + 2C_L T\cdot\Big(\frac{3}{\mu}+2\Big)^{\frac{1}{2}}\cdot\omega^{*-\frac{1}{2}}T^{-\frac{1}{3}} && \text{(due to (8))}\\ &\leqslant 2r_{\max}\mu T^{\frac{2}{3}}\log T + 2r_{\max}\mu^{-1}\phi^* T^{\frac{2}{3}}\log T\\ &\quad + 2\Big(\frac{3}{\mu}+2\Big)^{\frac{1}{2}}\omega^{*-\frac{1}{2}}C_L T^{\frac{2}{3}}\\ &\leqslant \Big(2r_{\max}\mu + 2r_{\max}\mu^{-1}\phi^*\\ &\quad + 3\Big(\frac{3}{\mu}+2\Big)^{\frac{1}{2}}\omega^{*-\frac{1}{2}}C_L\Big)T^{\frac{2}{3}}\log T \leqslant \kappa_1 T^{\frac{2}{3}}\log T. \quad \square\end{aligned}$$

4.4.2. Regret for the α-Optimal Assortment. We prove the α-regret bound in Theorem 2 and that the following constants ψ, T_2, and κ_2 in Theorem 2 are independent of $\theta \in \Theta$.

$$\kappa_2 := r_{\max} + r_{\max}\phi N^2 + \frac{r_{\max}N^7}{\omega\zeta'2}, \quad \psi := \frac{N^7}{\omega\zeta'2},$$

$$\zeta' := \min\left\{\frac{\underline{r}\alpha(\gamma-1)}{C_L(1+\alpha)}, \zeta_0 N\right\},$$

$$T_2 := h(\psi) = \inf\{x \geqslant 3 \mid \psi \log y \leqslant y,\ \forall y \geqslant x\}. \tag{10}$$

By definition, function $h(\psi) \in O(\psi^{1+\epsilon})$ for all $\epsilon > 0$. Thus, $T_2 \in O(N^{7(1+\epsilon)})$.

Proof. We first rewrite (10) into $\psi = \frac{1}{\omega^*\zeta'2}$, $\zeta' = \min\left\{\frac{\underline{r}\alpha(\gamma-1)}{C_L(1+\alpha)}, \zeta_0^*\right\}$, and $\kappa_2 = r_{\max} + r_{\max}\phi^* + r_{\max}\psi$. With $\tau = \lceil \psi \log T \rceil$, inequality $T \geqslant T_2$ indicates that $T \geqslant \max\{\tau, 3\}$. In addition, concentration Inequality (7) indicates that for all $\tau \in \mathbb{N}$, we have

$$\mathbb{P}[\|\tilde{\theta}^\tau - \theta\|_1 > \zeta'] \leqslant \phi^* e^{-\omega^*\tau\zeta'2}. \tag{11}$$

Thus, the α-regret of policy P_2 can be divided into three parts: (i) for $t \leqslant \tau$, the α-regret associated with every single customer is bounded by $r_{\max}$; (ii) for $t > \tau$ such that $\|\tilde{\theta}^\tau - \theta\|_1 > \zeta'$, the α-regret associated with every single customer is also bounded by $r_{\max}$; and (iii) for $t > \tau$ such that $\|\tilde{\theta}^\tau - \theta\|_1 \leqslant \zeta'$, the α-regret associated with every single customer is zero because $\|\tilde{\theta}^\tau - \theta\|_1 \leqslant \zeta' \leqslant \frac{\underline{r}\alpha(\gamma-1)}{C_L(1+\alpha)}$; that is, $\tilde{\theta}^\tau$ falls into the regret-free region. Thereby, $S_t = S^{\gamma\alpha}(\tilde{\theta}^\tau)(t > \tau)$ will be an α-optimal assortment under the true parameter θ according to Lemma 4. Thus, the α-regret of policy P_2 is bounded by

$$\begin{aligned}
\mathsf{Reg}^{\alpha}_{P_2}(T,\theta) &\leqslant r_{\max}\tau + r_{\max}(T-\tau)\cdot\mathbb{P}[\|\tilde{\theta}^\tau - \theta\|_1 > \zeta'] \\
&\leqslant r_{\max}\tau + r_{\max}T\cdot\phi^* e^{-\omega^*\tau\zeta'2} \qquad \text{(due to (11))} \\
&\leqslant r_{\max}\lceil\psi\log T\rceil + r_{\max}T\cdot\phi^* e^{-\omega^*\|\psi\log T\|\zeta'2} \\
&\leqslant r_{\max}\psi\log T + r_{\max} + r_{\max}T\cdot\phi^* e^{-\omega^*\psi\log T\zeta'2} \\
&\leqslant r_{\max}\psi\log T + r_{\max} + r_{\max}\phi^* T^{1-\omega^*\psi\zeta'2} \\
&\leqslant r_{\max}\psi\log T + r_{\max} + r_{\max}\phi^* \left(\text{due to } \psi = \frac{1}{\omega^*\zeta'2}\right) \\
&\leqslant (r_{\max}\psi + r_{\max} + r_{\max}\phi^*)\log T \\
&\leqslant \kappa_2 \log T. \qquad \text{(due to } T \geqslant 3). \quad \square
\end{aligned}$$

5. Connection Between Suboptimality Gap and Subregret

In this section, we demonstrate that the use of suboptimality gap in explore-then-commit learning literature can be viewed as a special case of our α-regret analysis. Let $\Delta_{\min}$ denote the suboptimality gap in our static MCC assortment selection problem:

$$\Delta_{\min} := \inf_{\theta\in\Theta}\left(r(S^*(\theta);\theta) - \max_{S\in\mathcal{S}\setminus\{S^*(\theta)\}} r(S;\theta)\right),$$

which represents the revenue gap between the best and the second-best assortments across all θ. We assume (i) the optimal solution $S^*(\theta)$ is unique and computable for $\theta \in \Theta$, and (ii) $\Delta_{\min} > 0$, which is a common assumption for explore-then-commit learning literature that uses suboptimality gap (Sauré and Zeevi 2013, Gallego and Lu 2021). Then we find the following equivalence between the α-optimality/α-regret and the exact optimality/regret.

Proposition 1. *By letting $\alpha = 1 - \frac{\Delta_{\min}}{2r_{\max}}$, any α-optimal assortment S is also exact optimal:*

$$\mathcal{S}^\alpha(\theta) = \{S^*(\theta)\}, \quad \theta \in \Theta. \tag{12}$$

For any policy P, its α-regret equals its regret:

$$\mathsf{Reg}^{\alpha}_{P}(T,\theta) = \mathsf{Reg}_{P}(T,\theta). \tag{13}$$

Proof. (12) is because for any $S \in \mathcal{S}^\alpha(\theta)$, we have $r(S;\theta) \geqslant \alpha\cdot r(S^*(\theta);\theta) \geqslant (1 - \frac{\Delta_{\min}}{2r_{\max}})\cdot r(S^*(\theta);\theta) \geqslant r(S^*(\theta);\theta) - \Delta_{\min}\cdot\frac{r(S^*(\theta);\theta)}{2r_{\max}} > r(S^*(\theta);\theta) - \Delta_{\min}$. because the revenue gap between S and $S^*(\theta)$ is smaller than $\Delta_{\min}$, $S = S^*(\theta)$. Combining (12) with the definition of regret,

$$\begin{aligned}
\mathsf{Reg}^{\alpha}_{P}(T,\theta) &= \inf_{S^\alpha_t(\theta)\in\mathcal{S}^\alpha(\theta),\, t\in\mathcal{T}} \mathsf{R}^\alpha(T,\theta) - \mathsf{R}_P(T,\theta) \\
&= \mathsf{R}^*(T,\theta) - \mathsf{R}_P(T,\theta) = \mathsf{Reg}_P(T,\theta). \quad \square
\end{aligned}$$

We let $\gamma = \frac{1}{\alpha}$ and write $S^*(\theta)$ as $S^{\gamma\alpha}(\theta)$ with slight abuse of notation. Together with Theorem 2, Proposition 1 suggests that by providing $\gamma\alpha$-optimal assortment (i.e., the exact optimal assortment) in the exploitation phase and deriving an order-of-$\log T$ α-regret, we can eventually bound the (normal) regret as $O(\log T)$. Formally, we have the following.

Proposition 2. *Suppose Assumptions 1 and 2 hold. By letting $\tau = \lceil\psi\log T\rceil$, $S_P = S^*(\cdot)$, and policy P_2' be defined by Algorithm 1, the regret associated with policy P_2' at any time $T \geqslant T_2$ is bounded as*

$$\mathsf{Reg}_{P_2'}(T,\theta) \leqslant \kappa_2 \log T,$$

where ψ, κ_2, and T_2 are constants defined in (10) with $\alpha = 1 - \frac{\Delta_{\min}}{2r_{\max}}$ and $\gamma = \frac{1}{\alpha}$.

The proof is similar to that for Theorem 2 and is provided in Appendix D. The regret bound in Proposition 2 matches the explore-then-commit learning literature that uses suboptimality gap to obtain an order-of-$\log T$ regret bound (Sauré and Zeevi 2013, Gallego and Lu 2021).

Moreover, as discussed in Remark 3, the separation period τ and regret coefficient κ_2 are both in the order of $\tilde{O}\left(\frac{1}{(\gamma-1)^2}\right)$. Because of the definitions of γ and α, we have

$\tau, \kappa_2 \in \tilde{O}\left(\frac{1}{\Delta_{\min}^2}\right)$. Then we conclude that our results, that is, the regret $\mathsf{Reg}_{P'_2}(T,\theta) \in O(\log T)$, the separation period $\tau \in \tilde{O}\left(\frac{1}{\Delta_{\min^2}}\right)$, and the regret coefficient $\kappa_2 \in \tilde{O}\left(\frac{1}{\Delta_{\min^2}}\right)$, exactly match that of Gallego and Lu (2021) based on assuming available suboptimality gap $\Delta_{\min}$.

As argued in Gallego and Lu (2021, p. 13), given that the suboptimality gap is often unavailable, "in practice we can artificially choose a small value for $\Delta_{\min}$." By doing so, any α-optimal assortment with some $\alpha > 1 - \frac{\Delta_{\min}}{r_{\max}}$ will be considered as good as the exact optimal assortment, and its associated regret will be omitted. This is essentially to use α-optimal solutions as a benchmark policy and use α-regret in place of normal regret, whereas the exact-optimal solutions are actually computable and applied in the exploitation phase (i.e., $S^*(\tilde{\theta}_\tau)$ instead of $S^\alpha(\tilde{\theta}_\tau)$ is provided).

6. Numerical Simulation

We conduct a numerical study of the FastLinETC algorithm. Section 6.1 presents the performance of our least square estimator for the MCC parameters. Section 6.2 investigates the performance of the proposed policy measured by regret.

6.1. Performance in Parameter Estimation

We use the least square (LS) estimator in Algorithm 1 and show its consistency in Lemma 5. Here we consider the practical performance of the least square estimator in a set of numerical examples. We also present the performance of the EM algorithm proposed in Şimşek and Topaloglu (2018) as a baseline. Although convergence analysis for an EM algorithm is challenging in general (Balakrishnan et al. 2017), the EM algorithm under the MCC model has appealing performance in various practical instances (Berbeglia et al. 2022).

In the experiments, we consider an MCC model with randomly generated parameters. We create the assortment collection in Assumption 1 with two arbitrary singletons as $S_\cap, S'_\cap$ and evenly generate samples using each assortment in the collection. To evaluate the estimation performance, we adopt the log-likelihood of independent out-of-sample test data as a criterion (Şimşek and Topaloglu 2018, Berbeglia et al. 2022). The test data are generated using random assortments where each product has a 0.5 probability to be offered (and we resample if the assortment contains no products). We use 10,000 samples to evaluate the performance of each estimate. The estimators are assessed by the average performance in two different settings with $n = 5$ and $n = 10$ products, respectively. Those tests are repeated 30 and 15 times, respectively, and the latter case uses less repetition due to computational burdens.

From the experiments, we find that the LS estimator has advantages over the EM algorithm in computational time and large sample performance. The disadvantage of the EM algorithm is induced by its iterative structure and the local search heuristic nature. Because the MCC model is favorable among all choice models in large data volume situations (Berbeglia et al. 2022), the LS estimator's computational advantages will be appreciated along with the proposed explore-then-commit policy.

The experiment results are shown in Table 1. It presents the log-likelihood of the estimates using different numbers of samples. In the $n = 5$ case, we observe that the EM estimate has better performance when the sample sizes are smaller ($10^3 \sim 10^4$ samples), whereas the LS estimate converges faster as the sample size increases. Similarly, in the $n = 10$ case, the EM algorithm requires fewer samples to achieve an acceptable (compared with the LS estimator with a small sample size) out-of-sample likelihood, but it cannot further improve as the sample size grows. Conversely, the LS estimator shows a better convergence performance.

Table 1 shows the computation time of the estimators. We observe that the EM algorithm requires more than 10^4 times of computation time than the LS estimator in both $n = 5$ and $n = 10$ cases. Moreover, the computation time of the LS estimator is almost independent of the sample size, whereas the EM algorithm has increasing computation time as the sample size grows.

6.2. Performance in Cumulative Regret

We illustrate the performance of the FastLinETC algorithm regarding the problems with computable and uncomputable optimal assortments. In particular, for the computable optimal assortment case, we consider the unconstrained assortment optimization that can be solved by LP (the LP formulation in Blanchet et al. (2016) is provided in Appendix B.1). For the uncomputable optimal assortment case, we consider the cardinality-constrained assortment optimization problem and use the approximation method in Désir et al. (2020) to obtain α-optimal solutions (the full description is provided in Appendix B.2). The two problems are expected to reflect the respective regret bounds in Theorems 1 and 2.

To begin with, we illustrate the regret bound in Theorem 1 through an unconstrained problem with $n = 10$ products. In our experiment, we show the practical performance of the explore-then-commit algorithm with varying selling horizon T. The performance of the learning algorithms is evaluated over 100 replications of randomly generated problem parameters. Specifically, each replication shares a fixed revenue vector with $r_i \in [2,3]$, $i \in \mathcal{N}$ and independently generates parameters λ_i $(i \in \mathcal{N})$ and ρ_{ij} $((i,j) \in \mathcal{N}^2 \backslash I_0)$ using uniform distributions with proper normalization. We use $\tau = \mu T^{2/3}$ with

Table 1. Performance of LS Algorithm Against EM Algorithm for Estimating the MCC Parameters

		LS algorithm (this work)			EM algorithm		
Samples ($\times 10^3$)	Number of products N	Log-likelihood	Runtime (s)	Mean error	Log-likelihood	Runtime (s)	Mean error
1	5	−35,134	0.001	0.980	−26,239	13.9	0.479
2	5	−24,303	0.001	0.369	−21,497	29.1	0.212
3	5	−22,368	0.001	0.261	−20,081	42.9	0.132
5	5	−20,158	0.001	0.136	−19,422	70.5	0.094
7	5	−19,246	0.001	0.084	−18,771	100.9	0.058
10	5	−19,160	0.001	0.080	−18,676	144.7	0.053
20	5	−18,324	0.001	0.032	−18,576	305.5	0.047
30	5	−18,266	0.001	0.029	−18,340	453.9	0.034
50	5	−17,940	0.001	0.011	−18,097	732.8	0.020
70	5	−17,894	0.001	0.009	−18,082	998.2	0.019
100	5	−17,863	0.001	0.007	−18,144	1,463.4	0.022
10	10	−51,086	0.022	0.533	−39,233	164.6	0.177
20	10	−45,733	0.003	0.372	−38,446	333.4	0.153
30	10	−43,199	0.002	0.296	−37,371	501.4	0.121
50	10	−39,740	0.002	0.192	−37,602	836.2	0.128
70	10	−38,333	0.003	0.150	−37,278	1,169.7	0.118
100	10	−36,305	0.002	0.089	−37,291	1,673.1	0.119
200	10	−35,104	0.005	0.053	−37,989	3,347.2	0.139
300	10	−34,308	0.005	0.029	−37,045	5,018.3	0.111
500	10	−33,921	0.008	0.018	−37,460	8,371.7	0.124
700	10	−33,736	0.015	0.012	−37,840	11,811.2	0.135
1,000	10	−33,643	0.017	0.009	−36,877	16,852.3	0.106

constant $\mu > 0$ to approximate the separation period in Theorem 1 and the selling horizon T ranging from 10^6 to 10^8. Although Theorem 1 would be valid for any constant value of μ, it is natural to enhance practical performance by incorporating problem scale and the selling horizon T into the selection of μ. In our experiment, we set the value of μ as $\mu \approx \frac{\log(N^2)}{\log T}$, which remains "asymptotic constant." Given our experimental setting, we observed that $\frac{\log(N^2)}{\log T} \approx 0.3$ and therefore we consider three values around 0.3 for μ, namely $\mu = 0.15, 0.3, 0.45$. To illustrate the efficiency of the separation scheme, we also implement the algorithm using $\mu T^{1/2}$ and $\mu T^{3/4}$ as the separation periods to provide performance baselines. In Figure 2, we present the cumulative regret over increasing selling horizons in Figure 2, (a)–(c), and summarize the proportion of exploration periods of all cases in Figure 2(d). From Figure 2, (a)–(c), we observe that all three choices of separation periods can provide sublinear cumulative regret rates, while the performance with $\tau = \mu T^{2/3}$ provides the smallest total regret. The diminishing exploration ratios in all three cases in Figure 2(d) validate the slow growth of the cumulative regret.

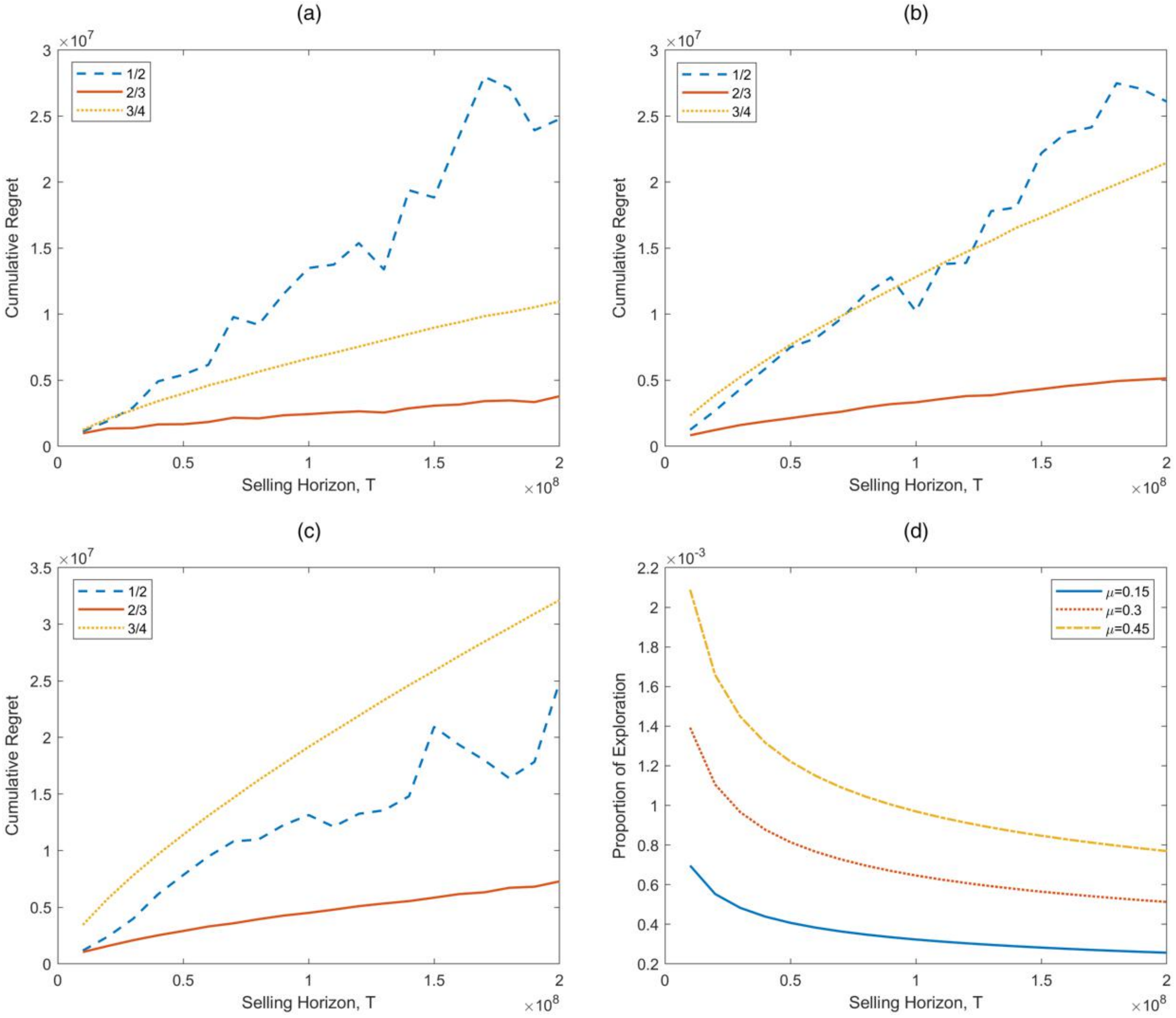
Next, we consider two capacity-constrained problems with $n = 10$ and $n = 20$ products, respectively. For each case, we use a fixed revenue vector with $r_i \in (0, 1)$, $i \in \mathcal{N}$, and generate random λ_i $(i \in \mathcal{N})$ and ρ_{ij} $((i, j) \in \mathcal{N}^2 \backslash I_0)$ using uniformly distributed nonzero entries with proper normalization. Because the capacity-constrained problem cannot be solved directly, we use an $\alpha = 0.4$-optimal approximation method in the optimization steps in Algorithm 1 and obtain the optimal assortments by enumeration for regret computation. We set $\psi = 350$ and present the simulation results based on 100 replications in Figure 3. Figure 3, (a) and (b), presents the average α-regret for the random instances with $n = 10$ and $n = 20$, respectively. The regret are of order-log T in both cases, as predicted in Theorem 2. (To better illustrate the order-of-log T rate, we provided rescaled plots in Appendix F.) Compared with the previous case, the proportion of exploration periods (customers) diminishes much faster as the number of samples increases. The exploration to selling horizon ratio ranges from 0.48% with $T = 10^5$ to 0.06% with $T = 10^6$. In all instances we considered, the final assortment provided by our algorithm achieves a near-zero α-regret.

7. Conclusion

We studied the dynamic constrained assortment selection problem under the MCC model and proposed the first online learning algorithm FastLinETC to minimize the cumulative regret over a selling horizon. Our results are particularly important because the MCC model is general enough to encompass the general attraction model (including the MNL model) and at the same time provides a good approximation for more advanced models that have no existing learning algorithms (for instance, the MMNL model).

A key future research direction is to develop a simultaneously-explore-and-exploit or learning-while-doing algorithm for the dynamic assortment planing problem under MCC model and obtain a better regret

Figure 2. (Color online) Performance of Algorithm 1 in Unconstrained Revenue Maximization Problems over Increasing Selling Horizons

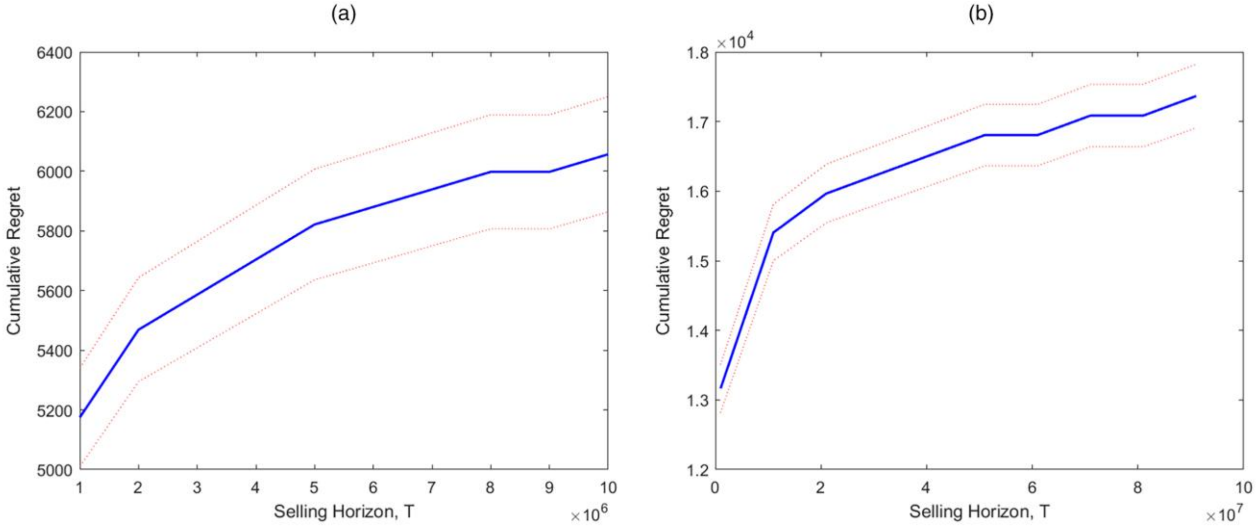


Notes. (a) The cumulative regret in the case with $\mu = 0.15$. (b) The cumulative regret in the case with $\mu = 0.3$. (c) The cumulative regret in the case with $\mu = 0.45$. (d) The proportion of exploration periods.

bound, for example, in the order of $\tilde{O}(\sqrt{T})$. We envision that the hardness of parameter estimation lies at the core of designing such an adaptive learning algorithm. Section 3.1 mentioned that a consistent parameter estimation for MCC model requires multiple exploration assortments, implying that suboptimal assortments are included. Remark 6 suggests equal sampling weights across exploration assortments. These observations indicate that, in an adaptive learning algorithm, the frequency of suboptimal assortment is proportional to the selling horizon length T and the total regret grows linearly with T. On an intuitive level, an adaptive learning algorithm can yield a better regret bound only if (a) the model parameter estimator becomes more accurate as time goes by, whereas (b) the optimal assortment is assigned an increasing sampling weight as time goes by and this weight of optimal assortment should approach one; one example is the UCB algorithm for MNL model in Agrawal et al. (2019). For the MCC model, however, the goals of (a) and (b) are conflicting. Continuously offering the optimal assortment alone is not sufficient to guarantee an increasingly accurate estimator, particularly for transition matrix ρ.

We close this paper by pointing out two plausible directions for future research. First, our model assumes the presence of substitution effects between every pair of products. Consequently, we need to estimate $O(N^2)$ parameters, which adds complexity to the design of learning policies. However, in practice, substitutions may exhibit sparsity and only occur between products belonging to the same subgroups, such as substitutions within the same brand. An interesting direction is to identify the substitution sparsity and use the sparsity to simplify the policy design and reduce policy regret. Second, another interesting direction is to incorporate the pricing or inventory planning decisions into the

Figure 3. (Color online) Performance of Algorithm 1 in Constrained Revenue Maximization Problems over Increasing Selling Horizons



Notes. (a) Ten products. (b) Twenty products. The solid lines represent the mean regret, and the dotted lines represent the estimated 95% confidence intervals for the simulation results.

assortment problem under the MCC model and develop efficient learning algorithms.

Acknowledgments

The authors thank the area editor Professor Ilan Lobel, associate editor, and anonymous referees for detailed and constructive comments that significantly improved the content and exposition of this paper.

Appendix A. Summary of Major Notation

Table A.1 summarizes the major mathematical notation used in the manuscript.

Appendix B. Optimization Algorithm for Static Assortment Optimization

B.1. Unconstrained MCC Assortment Optimization

Blanchet et al. (2016) proposed an approach using LP to obtain the optimal assortment for the MCC model. The approach considers the following LP

$$\begin{aligned} \min_{g} \ & \sum_{i \in \mathcal{N}} g_i \\ s.t. \ & g_i \geq r_i, \text{ for } i \in \mathcal{N}, \\ & g_i \geq \sum_{j \in \mathcal{N}} \rho_{ij} g_j, \text{ for } i \in \mathcal{N}. \end{aligned} \tag{B.1}$$

Table A.1. Major Notation and Their Definitions

Notation	Definition
$A_1, A_2, \ldots$	Assortments offered in the exploration phase
a_i	Attraction parameter associated with λ_i
$\underline{a}_i, \overline{a}_i$	Lower and upper bounds of a_i
α	Optimality ratio of approximation algorithms
b_{ij}	Attraction parameter associated with ρ_{ij}
$\underline{b}_{ij}, \overline{b}_{ij}$	Lower and upper bounds of b_{ij}
$C_L, (C_L^*)$	(Tight) Lipschitz constant of $r(S, \theta)$ $w.r.t.$ θ
$C_{L_1}, C_{L_2}, \ldots, C_{L_5}$	Tight Lipschitz constants of $r(S, \theta)$ $w.r.t.$ five different groups of elements in θ
d, d_0	Dimensions of $\mathscr{S}_0 \cup \tilde{\mathscr{S}}_0$ and $\mathscr{S}_0$
$\Delta_{\min}$	Suboptimality gap (i.e., the revenue gap between the unique optimal solution and the other solutions)
$\eta, (\eta(\mathscr{S}))$	Minimal transition probability to state 0 from any other state (outside any $S \in \mathscr{S}$)
γ	Multiplier for optimality ratio
I_0	Indices of trivial elements in MCC parameter θ
$\kappa_1, T_1, \kappa_2, T_2$	Constants used to construct regret bounds of P_1 and P_2
$\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}}$	Vectors of λ_i with respectively $i \in S$ and $i \in \mathcal{N} \backslash S$
λ_i	Arrival probability to state i in the MCC model
N	Total number of products
$\mathcal{N}$	Set of all products
n^*	Arbitrary integer outlined in Assumption 1

Table A.1. (Continued)

Notation	Definition
ω, ω^*	Constants used to construct the concentration inequality of $\tilde{\theta}^\tau$
P	Assortment selection policy
$\boldsymbol{P}, \bar{\boldsymbol{P}}$	Matrices of ρ_{ij} with, respectively, $i \in \mathcal{N} \backslash S, j \in S$, and $i, j \in \mathcal{N} \backslash S$
P_1, μ	Order of $T^{2/3}\log T$ policy and its associated parameter
P_2, ψ	Order of $\log T$ policy and its associated parameter
ϕ, ϕ^*	Constants used to construct the concentration inequality of $\tilde{\theta}^\tau$
$\pi(i,S), \hat{\pi}^\tau(i,S)$	Probability that a customer purchases product i from assortment S and its estimator
$\pi(i,S;\theta)$	Probability that a customer purchases product i from assortment S under parameter θ
$\pi(j,S\|i), \hat{\pi}^\tau(j,S\|i)$	Probability that a customer purchases product j from assortment S conditional on that the first-choice demand is product i, and its estimator
r	Single-product revenue vector with $i \in \mathcal{N}_+$
$r(S;\theta)$	Average single-sale revenue under assortment S and parameter θ
$\mathsf{R}^*(T,\theta)$	T-period revenue of the optimal policy under parameter θ
$\boldsymbol{r}$	Vector of r_i with $i \in S$
$\mathsf{R}^\alpha(T,\theta)$	T-period revenue of a policy comprised of α-optimal assortments under parameter θ
$r_{\max}, (\overline{r}(\mathcal{S}))$	Maximal single-product revenue (under possible assortments $\mathcal{S}$)
$r_{\min}$	Minimal single-product revenue
$\mathsf{Reg}_P^\alpha(T,\theta)$	Cumulative T-period α-regret of policy P under parameter θ
$\mathsf{Reg}_P(T,\theta)$	Cumulative T-period regret of policy P under parameter θ
ρ_{ij}	Transition probability from state i to j in the MCC model
r_i	Single-product revenue of product i
$\mathsf{R}_P(T,\theta)$	T-period revenue of policy P under parameter θ
$\underline{r}$	Lower bound of the maximal single-sale revenue in the worst case of θ
S	Subset of products
$\mathcal{S}$	Set of possible assortments
$\mathcal{S}_0, \tilde{\mathcal{S}}_0$	Assortments offered in the exploration phase
$S_\cap, S'_\cap$	Pivot sets outlined in Assumption 1
S_+	Subset of products, including the no purchase option
$S_P(\cdot)$	Assortment function in Algorithm 1 that maps any parameter θ to an assortment
S_t	Assortment offered in period t
$S^*(\theta), (S_t^*(\theta))$	Exact optimal assortment under parameter θ (offered in period t)
$S^\alpha(\theta), (S_t^\alpha(\theta))$	α-Optimal assortment under parameter θ (offered in period t)
$\mathcal{S}^\alpha(\theta)$	Collection of α-optimal assortments under parameter θ
T	Selling horizon length
t	Index of period
τ	Separation period
$\mathcal{T}$	Set of all periods
Θ	Parameter space for θ
$\theta, \hat{\theta}^\tau, \tilde{\theta}^\tau$	MCC parameter, estimator, and rounded estimator
$\theta_0, \hat{\theta}_0^\tau$	MCC parameter (trivial elements) and estimator
$\theta_{++}, \hat{\theta}_{++}^\tau$	MCC parameter (nontrivial elements) and estimator
$u(i,S;\theta)$	Average visit times to product i under assortment S and parameter θ
$\boldsymbol{u}, \bar{\boldsymbol{u}}$	Vectors of $u(i,S;\theta)$ with respectively $i \in S$ and $i \in \mathcal{N} \backslash S$
$X, \hat{X}^\tau$	Coefficient matrix constructed by $\pi(j,S\|i)$ to compute θ_{++} and its estimator
$Y, \hat{Y}^\tau$	Inhomogeneous vector constructed by $\pi(j,S\|i)$ to compute θ_{++} and its estimator
ζ'	Upper bound for a "proper" ζ, used to construct regret bounds
ζ_0, ζ_0^*	Upper bound for a "proper" ζ, used to construct concentration inequality of $\tilde{\theta}^\tau$
$Z^t, (Z_i^t)$	Purchase decision vector of customer t (regarding product i)

With the optimal solution g^* from (B.1), the optimal assortment is given by $S^* = \{i : r_i = g_i^*\}$. Two alternative methods for solving the unconstrained assortment optimization are respectively proposed by Blanchet et al. (2016) and Gallego and Lu (2021).

B.2. Approximation Algorithm for Cardinality-Constrained MCC Assortment Optimization.

Algorithm B.1 presents the $(1/2 - \epsilon)$-approximation algorithm for cardinality-constrained MCC assortment optimization proposed by Désir et al. (2020). The algorithm selects a product per iteration and stops when the assortment hits the cardinality bound $\bar{s}$. We use S_t to denote the set of selected items at step t with $S_0 = \emptyset$ and use C_t to denote the consideration set with $C_0 = \mathcal{N}$ According to Désir et al. (2020), we set the tuning parameter $\beta = 1/2$.

The adjusted revenue of item i with regard to assortment S is defined as

$$r_i^S = \begin{cases} r_i - \sum_{j \in S} \rho_{ij} r_j, \text{ if } i \notin S, \\ 0, \text{ if } i \in S. \end{cases} \tag{B.2}$$

The adjusted transition probabilities with regard to assortment S are defined as

$$\rho_{ij}^{S} = \begin{cases} 1, \text{ if } i \in S, j = 0, \\ 0, \text{ if } i \in S, j \neq 0, \\ \rho_{ij}, \text{ otherwise}. \end{cases} \tag{B.3}$$

With the adjusted transition probabilities, we use θ^S to denote the adjusted MCC parameters with regard to assortment S. Finally, the adjusted revenue of assortment $\tilde{S}$ with regard to S is given by

$$r^S(\tilde{S};\theta) = \sum_{i \in \tilde{S}} r_i^S \pi(i, \tilde{S}; \theta^S). \tag{B.4}$$

Algorithm B.1 ($(1/2 - \epsilon)$-Approximation Algorithm for $(\mathcal{N}, r, \theta)$ with Cardinality Constraint $\bar{s}$)

```
Compute the unconstrained optimal assortment revenue, $r(U^*;\theta)$, using (B.1), and set
    $$B_j = \frac{\bar{s}}{N} r(U^*;\theta)(1+\epsilon)^j, j = 1, \ldots, J, \quad (B.5)$$
where $J = \min\{j \in \mathbb{N} | B_j \geqslant r(U^*;\theta)\}$.
for $j \in \{1, 2, \ldots, J\}$ do
    Set $t = 1$, $S_0 = \varnothing$, and $C_0 = \mathcal{N}$.
    while $|S_{t-1}| < \bar{s}$ and $C_{t-1} \neq \varnothing$ do
        Set $C_t = \left\{ i \in \mathcal{N} \setminus S_{t-1} | r^{S_{t-1}}(\{i\};\theta) \geqslant \beta \frac{B_j}{\bar{s}} \right\}$, where $r^{S_{t-1}}(\{i\};\theta)$ follows (B.4).
        $S_t = S_{t-1} \cup \{i^*\}$, where $i^* = \arg\max_{i \in C_t} r_i^{S_{t-1}}$ (breaking ties arbitrarily), where $r_i^{S_{t-1}}$
        follows (B.2). Let $t := t + 1$.
    end while
    Let $S^j := S_t$.
end for
Return $S^{1/2-\epsilon} = S^{j^*}$ where $j^* = \arg\max_{j \in \{1,\ldots,J\}} r(S^j;\theta)$.
```

Appendix C. Proofs for Technical Results in Section 4

C.1. Relaxation of Assumption 2 in MCC Models

We can transform any MCC model with self-loops, that is, a transition matrix $\{\hat{\rho}_{ij}\}_{i,j \in \mathcal{N}_+}$ where $\hat{\rho}_{ii} > 0$ for some $i \in \mathcal{N}$, into an MCC model without self-loops by reparameterization. Given any arrival probability vector $\{\lambda_i\}_{i \in \mathcal{N}_+}$, this transition matrix $\{\hat{\rho}_{ij}\}_{i,j \in \mathcal{N}_+}$ defines a discrete-time Markov chain $\hat{\Phi}(t)(t \in \mathbb{N})$ for a customer's process of transitioning to preferred products. Here t indexes the number of transitions; the preferred product may not change after one transition due to $\hat{\rho}_{ii} > 0$ for some $i \in \mathcal{N}$. $\hat{\Phi}(t)$ represents the preferred product after t transitions. Particularly, the customer stops and purchases whenever the customer's preferred product is available from the offered assortment S or leave with no-purchase.

The reparametrized MCC model without self-loops can be defined as follows. Let $n \in \mathbb{N}$ denote the number of preference changes (to a different product) and $\Phi(n)$ denote the preferred product after n changes. Then $\Phi(n)$ is a Markov chain embedded in $\hat{\Phi}(t)$, whose transition matrix is denoted by $\{\rho_{ij}\}_{i,j \in \mathcal{N}_+}$, and it satisfies the no self-loop assumption (i.e., $\rho_{ii} = 0, i \in \mathcal{N}$). This embedded Markov chain $\Phi(n)$ yields identical choice probabilities to the original Markov chain $\hat{\Phi}(t)$ for any assortment S. This is because for any sample path of the original Markov chain $\hat{\Phi}(n)$, where the customer stops (a product in S or the no-purchase) is the same with the corresponding sample path of embeded Markov chain $\Phi(t)$. Let the reparametrized MCC model's transition matrix be $\{\rho_{ij}\}_{i,j \in \mathcal{N}_+}$ and arrival probability vector remain $\{\lambda_i\}_{i \in \mathcal{N}_+}$. Then the reparametrized MCC model's choice probabilities are identical to that of the original MCC model for any assortment S.

It is not hard to obtain the reparametrized transition matrix. Consider the ith row. If $i \in \mathcal{N}$ and $\hat{\rho}_{ii} > 0$, we have $\rho_{ii} = 0$ and $\rho_{ij} = \frac{\hat{\rho}_{ij}}{1 - \hat{\rho}_{ii}}$ for $j \neq i$; otherwise, $\rho_{ij} = \hat{\rho}_{ij}$ for $j \in \mathcal{N}_+$. We demonstrate this reparameterization in Example C.1.

Example C.1. Assume that the number of products is $n = 2$ and the arrival probability vector $\{\lambda_i\}_{i \in \mathcal{N}_+}$ is $(0, 0.5, 0.5)$. The original transition matrix $\{\hat{\rho}_{ij}\}_{i,j \in \mathcal{N}_+}$ with self-loops and the reparametrized transition matrix $\{\rho_{ij}\}_{i,j \in \mathcal{N}_+}$ without self-loops are

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix},$$

respectively. Using (SS-1) and (SS-2), we find that the choice probabilities are the same for transition matrices $\{\rho_{ij}\}_{i,j \in \mathcal{N}_+}$ and $\{\hat{\rho}_{ij}\}_{i,j \in \mathcal{N}_+}$, given any assortment $S = \varnothing, \{1\}, \{2\}$, or $\{1, 2\}$. For instance, if $S = \{1\}$, then the choice probability vector for products 0, 1, and 2 is $(0.25, 0.75, 0)$ under both transition matrices.

This example shows that an MCC model with self-loops has identical choice probabilities with the reparametrized MCC model without self-loops for any assortment S. Moreover, if both models A and B have self-loops and both can be reparametrized into a model C without self-loops, then models A and B have identical choice probabilities for any assortment S. As a result, the MCC model cannot be identified.

C.2. Proof of Lemma 3

We first prove (2) using the following equality in the supremum norm:

$$\|(\boldsymbol{I} - \bar{\boldsymbol{P}})^{-1}\boldsymbol{y}\|_\infty = \sup_{\boldsymbol{x} \in \mathcal{D}^{|\mathcal{N} \setminus S|}} |\boldsymbol{x}^\top (\boldsymbol{I} - \bar{\boldsymbol{P}})^{-1} \boldsymbol{y}|, \tag{C.1}$$

where $\mathcal{D}^k := \{\boldsymbol{x} \in \mathbb{R}_+^k | \|\boldsymbol{x}\|_1 = 1\}$ is the collection of stochastic vectors in $\mathbb{R}^k (k \in \mathbb{N})$. To bound $|\boldsymbol{x}^\top (\boldsymbol{I} - \bar{\boldsymbol{P}})^{-1} \boldsymbol{y}|$ for an arbitrary $\boldsymbol{x} \in \mathcal{D}^{|\mathcal{N} \setminus S|}$, we construct the following Markov chain with states $(\mathcal{N} \setminus S)_+ = \{0\} \cup (\mathcal{N} \setminus S)$. The transition matrix is $\boldsymbol{Q}$ indexed by $i, j \in (\mathcal{N} \setminus S)_+$, where $\boldsymbol{Q}_{i,j \in \mathcal{N} \setminus S} = \bar{\boldsymbol{P}}$ and $\boldsymbol{Q}_{0,j \in \mathcal{N} \setminus S} = \boldsymbol{x}^\top$. The Markov chain transition matrix $\boldsymbol{Q}$ is uniquely specified, and it has a unique stationary distribution because (i) $\boldsymbol{Q}_{i0} = 1 - \sum_{j \in \mathcal{N} \setminus S} \rho_{ij} \geqslant \rho_{i0} \geqslant \eta(\mathcal{S}) \geqslant \eta > 0, \forall i \in \mathcal{N} \setminus S$ and (ii) the communication class that contains state 0 forms an irreducible Markov chain with a unique stationary distribution that equals the unique stationary distribution associated with transition matrix $\boldsymbol{Q}$.

Let a stochastic vector $(z_0, \boldsymbol{z}) \in \mathbb{R}_+^{1 + |\mathcal{N} \setminus S|}$, where $\boldsymbol{z} = (z_i)_{i \in \mathcal{N} \setminus S}$, denote the unique stationary distribution associated with $\boldsymbol{Q}$. Then, $(z_0, \boldsymbol{z})$ must satisfy the balance equations:

$$z_0 \boldsymbol{x} + \bar{\boldsymbol{P}}^\top \boldsymbol{z} = \boldsymbol{z}, \tag{C.2}$$

$$\sum_{i \in \mathcal{N} \setminus S} z_i \boldsymbol{Q}_{i0} = z_0. \tag{C.3}$$

According to (C.2), we have $\boldsymbol{x}^\top(\boldsymbol{I}-\bar{\boldsymbol{P}})^{-1} = \boldsymbol{z}^\top/z_0$. According to (C.3), we have the following lower bound for z_0:

$$z_0 \geqslant \left(\min_{i\in\mathcal{N}\setminus S} \boldsymbol{Q}_{i0}\right)\cdot \sum_{i\in\mathcal{N}\setminus S} z_i \geqslant \eta(\mathscr{S})\cdot(1-z_0) \quad\Rightarrow\quad z_0 \geqslant \frac{\eta(\mathscr{S})}{1+\eta(\mathscr{S})}. \tag{C.4}$$

Because $\boldsymbol{z}$ is a substochastic vector, we have

$$|\boldsymbol{z}^\top\boldsymbol{y}| \leqslant \|\boldsymbol{z}\|_1 \cdot \|\boldsymbol{y}\|_\infty \leqslant (1-z_0)\|\boldsymbol{y}\|_\infty. \tag{C.5}$$

Combining (C.4) and (C.5) yield

$$|\boldsymbol{x}^\top(\boldsymbol{I}-\bar{\boldsymbol{P}})^{-1}\boldsymbol{y}| = \frac{|\boldsymbol{z}^\top\boldsymbol{y}|}{z_0} \leqslant \frac{(1-z_0)\|\boldsymbol{y}\|_\infty}{z_0} \leqslant \frac{\|\boldsymbol{y}\|_\infty}{\eta(\mathscr{S})}.$$

Plugging this into (C.1) yields (2).

We next prove (3). (C.3) and (C.4) also indicate for all $\boldsymbol{x}\in\mathcal{D}^{|\mathcal{N}\setminus S|}$, $\|(\boldsymbol{I}-\bar{\boldsymbol{P}}^\top)^{-1}\boldsymbol{x}\|_\infty = \|\boldsymbol{x}^\top(\boldsymbol{I}-\bar{\boldsymbol{P}})^{-1}\|_\infty = \|\frac{\boldsymbol{z}^\top}{z_0}\|_\infty \leqslant \frac{1-z_0}{z_0} \leqslant \frac{1}{\eta(\mathscr{S})}$. Therefore, for all $\boldsymbol{y}\in\mathbb{R}_+^{|\mathcal{N}\setminus S|}, S\subseteq\mathcal{N}$, and $\theta\in\Theta$, we have

$$\|(\boldsymbol{I}-\bar{\boldsymbol{P}}^\top)^{-1}\boldsymbol{y}\|_\infty \leqslant \frac{\|\boldsymbol{y}\|_1}{\eta(\mathscr{S})},$$

which is exactly (3). □

C.3. Proof of Lemma 5

We prove the results in four steps. First, we show that $\tilde{\theta}^\tau$ defines a consistent estimator. Second, we prove the Lipschitz continuity of $\hat{\theta}^\tau$ w.r.t. $(\hat{X}^\tau, \hat{Y}^\tau)$. Next, we prove a concentration inequality for $(\hat{X}^\tau, \hat{Y}^\tau)$. Finally, we prove the concentration inequalities for $\hat{\theta}^\tau$ and $\tilde{\theta}^\tau$ by combining results from Steps 2 and 3.

C.3.1. Step 1: Consistency. According to theorem 1 in Gupta and Hsu (2020), given any $\theta\in\Theta$, the following equality holds:

$$\sum_{k\in\mathcal{N}}[\pi(j,S|k)-\pi(j,S|0)]\cdot\rho_{ik} = [\pi(j,S|i)-\pi(j,S|0)],$$
$$i\in\mathcal{N}\setminus S, j\in S_+, S\in\mathscr{S}_0, \tag{C.6}$$

$$\sum_{k\in\mathcal{N}}[\pi(j,S|k)-\pi(j,S|0)]\cdot\lambda_k = [\pi(j,S)-\pi(j,S|0)],$$
$$j\in S_+, S\in\mathscr{S}_0, \tag{C.7}$$

where

$$\pi(j,S|i) := \begin{cases} 1 & \text{if } i=j, \\ \dfrac{\pi(j,S)-\pi(j,S\cup\{i\})}{\pi(i,S\cup\{i\})} & \text{if } i\in\mathcal{N}\setminus S, \\ 0 & \text{if } i\in S_+\setminus\{j\}. \end{cases} \tag{C.8}$$

$$\pi(i,S) = \pi(i,S;\theta), \quad i\in\mathcal{N}_+, S\in\tilde{\mathscr{S}}_0\cup\mathscr{S}_0. \tag{C.9}$$

Similar to the matrix formulation of (E-4)–(E-5) into (E-6), we can also write (C.6)–(C.7) into a matrix equation of nontrivial parameter $\theta_{++} = \text{vec}(\{\lambda_i\}_{i\in\mathcal{N}} \cup \{\rho_{ij}\}_{(i,j)\in\mathcal{N}^2\setminus I_0})$ after plugging in trivial parameter $\theta_0 = \text{vec}(\{\rho_{ij}\}_{(i,j)\in I_0}) \equiv 0$:

$$\{\theta_{++} : (C.6), (C.7)\} \Longleftrightarrow \{\theta_{++} : X\theta_{++} = Y\}, \tag{C.10}$$

where entries of X and Y are defined by rewriting coefficients in (C.6)–(C.7) into matrices. The following result shows that (C.10) returns a unique solution for θ_{++}.

Lemma C.1 (Lemma 7 and Lemma 8 from Gupta and Hsu 2020). *Consider the MCC model $(\mathcal{N}, r, \theta)$ with parameter space Θ. For every $\theta\in\Theta$, define matrices X and Y through (C.6)–(C.10). Then X has a full column rank, and for all $\theta\in\Theta$, we have*

$$\theta_{++} = (X^\top X)^{-1}(X^\top Y).$$

Proof. According to lemma 7 and lemma 8 in Gupta and Hsu (2020), the matrix X has a full column rank under Assumption 2. Unlike the full matrix X in Gupta and Hsu (2020), we have reduced nonessential parameters $\lambda_0 = 1-\sum_{i\in\mathcal{N}}\lambda_i$ and $\{\rho_{i0} = 1-\sum_{k\in\mathcal{N}}\rho_{ik}\}_{i\in\mathcal{N}}$ from our coefficient matrix X via elementary column operations, but the full rank property still holds. Thus, (C.10) has a unique solution θ_{++}. □

Compared with θ_{++} exactly satisfying all equalities in (C.6) and (C.7), estimators $\hat{\theta}^\tau_{++}$ with minimal squared residuals may not satisfy all equalities in (E-4) and (E-5). Lemma C.1 implies that if the intermediate estimators $\hat{X}^\tau = X$ and $\hat{Y}^\tau = Y$, then by definition (E-7), $\hat{\theta}^\tau_{++} = (\hat{X}^{\tau^\top}\hat{X}^\tau)^{-1}(\hat{X}^{\tau^\top}\hat{Y}^\tau) = (X^\top X)^{-1}(X^\top Y) = \theta_{++}$. Moreover, because the choice probability estimators $\hat{\pi}^\tau(i,S)$ $(i\in\mathcal{N}_+, S\in\mathscr{S}_0\cup\tilde{\mathscr{S}}_0)$, the conditional choice probability estimators $\hat{\pi}^\tau(j,S|i)$ $(i,j\in\mathcal{N}_+, S\in\mathscr{S}_0)$, and the linear equation system's coefficients $(\hat{X}^\tau, \hat{Y}^\tau)$ are all consistent under norm $\|\cdot\|_\infty$, $\hat{\theta}^\tau_{++}$ as a composite estimator is also consistent. Additionally, the trivial parameter estimator $\hat{\theta}^\tau_0 \equiv \theta_0 \equiv 0$ is naturally consistent. Thus, estimator $\hat{\theta}^\tau = \text{vec}(\hat{\theta}^\tau_0, \hat{\theta}^\tau_{++})$ is consistent.

Finally, by the definition of $\tilde{\theta}^\tau$ in (E-9) and the distance bound in (5), $\tilde{\theta}^\tau$ is also consistent, that is, $\tilde{\theta}^\tau \xrightarrow{p} \theta$ as τ goes to ∞.

C.3.2. Step 2: Lipschitz Continuity in Intermediate Estimators. In this section, we further develop results in Step 1 and show that the error of $\tilde{\theta}^\tau$ is Lipschitz continuous in that of $(\hat{X}^\tau, \hat{Y}^\tau)$.

Let us first define the following constants independent of $\theta\in\Theta$:

$$L_\Lambda := \inf_{\theta\in\Theta}\Lambda_{\min}(X^\top X), \qquad U_{\mathcal{XY}} := \sup_{\theta\in\Theta}\|X^\top Y\|_2,$$
$$U_{\mathcal{X}} := \sup_{\theta\in\Theta}\|X\|_2, \qquad U_{\mathcal{Y}} := \sup_{\theta\in\Theta}\|Y\|_2.$$

Here, the matrices X and Y are defined through (C.6)–(C.10) and are dependent on θ. By Lemma C.1, $\Lambda_{\min}(X^\top X) > 0$ for all $\theta\in\Theta$. Because Θ is compact, $L_\Lambda = \inf_{\theta\in\Theta}\Lambda_{\min}(X^\top X) > 0$. We also define L_π as a lower bound for the choice probabilities $\{\pi(i,S\cup\{i\}) | i\in\mathcal{N}\setminus S, S\in\mathscr{S}_0\}$, which is independent of $\theta\in\Theta$. For example, because $\pi(i,S\cup\{i\}) \geqslant \lambda_i$, we can let $L_\pi = \min_{i\in\mathcal{N}}\inf_{\theta\in\Theta}\{\lambda_i\} = \min_{i\in\mathcal{N}}\frac{\underline{a}_i}{1+\sum_{k\in\mathcal{N}}\overline{a}_k+(\underline{a}_i-\overline{a}_i)}$. Last, let $\dim(M)$ denote the total number of elements in a matrix M.

We have the following Lipschitz continuity result.

Lemma C.2. *For the MCC model $(\mathcal{N}, r, \theta)$ with parameter space Θ, define matrices X and Y through (C.6)–(C.10) with a fixed $\theta\in\Theta$. There exist constants δ_1 and C_E independent of $\theta\in\Theta$ such that for every $\hat{X}\in\mathbb{R}^{\dim(X)}$ and every $\hat{Y}\in\mathbb{R}^{\dim(Y)}$, if $\|(\hat{X},\hat{Y})-(X,Y)\|_2 \leqslant \delta_1$, then $\hat{X}$ has a full column rank and*

$$\|(\hat{X}^\top\hat{X})^{-1}(\hat{X}^\top\hat{Y}) - (X^\top X)^{-1}(X^\top Y)\|_2 \leqslant C_E\|(\hat{X},\hat{Y})-(X,Y)\|_2.$$

Particularly, we can define δ_1 and C_E as

$$\delta_1 := \min\left\{U_{\mathcal{X}}, \frac{L_\Lambda}{6U_{\mathcal{X}}}\right\},$$

$$C_E := \max\left\{\frac{2\sqrt{2}U_{\mathcal{Y}}}{L_\Lambda} + \frac{6\sqrt{2}U_{\mathcal{XY}}U_{\mathcal{X}}}{L_\Lambda^2}, \frac{4\sqrt{2}U_{\mathcal{X}}}{L_\Lambda}\right\}.$$

Proof. Let $\Delta_{\mathcal{X}} := \hat{X} - X$ and $\Delta_{\mathcal{Y}} := \hat{Y} - Y$. We first prove the smallest eigenvalue perturbation:

$$|\Lambda_{\min}(X^\top X) - \Lambda_{\min}(\hat{X}^\top \hat{X})| \leqslant \|\Delta_{\mathcal{X}}\|_2^2 + 2\|\Delta_{\mathcal{X}}\|_2\|X\|_2. \qquad \text{(C.11)}$$

Let "$M \succcurlyeq 0$" denote that the matrix M is positive semidefinite, $\text{vec}(M)$ denote the vectorization of M, and operation $M_1 \circ M_2$ denote $\mathit{trace}(M_1^\top M_2) = \text{vec}(M_1)^\top \text{vec}(M_2)$. Then we have

$$\begin{aligned}\Lambda_{\min}(X^\top X) &= \max\{t \in \mathbb{R} : X^\top X - tI \succcurlyeq 0\}\\ &= \min\{Z \circ (X^\top X) : \mathit{trace}(Z) = 1, Z \succcurlyeq 0\},\\ \Lambda_{\min}(\hat{X}^\top \hat{X}) &= \max\{t \in \mathbb{R} : \hat{X}^\top \hat{X} - tI \succcurlyeq 0\}\\ &= \min\{Z \circ (\hat{X}^\top \hat{X}) : \mathit{trace}(Z) = 1, Z \succcurlyeq 0\},\end{aligned}$$

where the second equality of each row is due to the duality of semidefinite programming. Let $Z^* = \text{argmin}\{Z \circ (X^\top X) : \mathit{trace}(Z) = 1, Z \succcurlyeq 0\}$. Then,

$$\begin{aligned}\Lambda_{\min}(\hat{X}^\top \hat{X}) &= \min\{Z \circ (\hat{X}^\top \hat{X}) : \mathit{trace}(Z) = 1, Z \succcurlyeq 0\} \leqslant Z^* \circ (\hat{X}^\top \hat{X})\\ &\leqslant Z^* \circ (X^\top X) + Z^* \circ (\Delta_{\mathcal{X}}^\top X + X^\top \Delta_{\mathcal{X}} + \Delta_{\mathcal{X}}^\top \Delta_{\mathcal{X}})\\ &\qquad (\text{due to } \hat{X} = X + \Delta_{\mathcal{X}})\\ &\leqslant \Lambda_{\min}(X^\top X) + \text{vec}(Z^*)^\top \text{vec}(\Delta_{\mathcal{X}}^\top X + X^\top \Delta_{\mathcal{X}} + \Delta_{\mathcal{X}}^\top \Delta_{\mathcal{X}})\\ &\leqslant \Lambda_{\min}(X^\top X) + \|Z^*\|_2 \cdot (\|\Delta_{\mathcal{X}}^\top X\|_2 + \|X^\top \Delta_{\mathcal{X}}\|_2 + \|\Delta_{\mathcal{X}}^\top \Delta_{\mathcal{X}}\|_2)\\ &\leqslant \Lambda_{\min}(X^\top X) + (2\|\Delta_{\mathcal{X}}^\top X\|_2 + \|\Delta_{\mathcal{X}}^\top \Delta_{\mathcal{X}}\|_2), \qquad \text{(C.12)}\end{aligned}$$

where the last inequality is because of $\|Z^*\|_2 \leqslant 1$:

$$\begin{aligned}\|Z^*\|_2^2 &= \text{vec}(Z^*)^\top \text{vec}(Z^*) = \mathit{trace}(Z^* Z^*) = \mathit{trace}(\mathit{diag}(Z^*)\mathit{diag}(Z^*))\\ &\leqslant \mathit{trace}(\mathit{diag}(Z^*)) \leqslant 1 \quad (\text{due to } \mathit{diag}(Z^*) \geqslant 0, \mathit{trace}(\mathit{diag}(Z^*))\\ &\qquad = \mathit{trace}(Z^*) = 1). \qquad \text{(C.13)}\end{aligned}$$

Similarly, we can define $\hat{Z}^* = \text{argmin}\{Z \circ (\hat{X}^\top \hat{X}) : \mathit{trace}(Z) = 1, Z \succcurlyeq 0\}$. Then,

$$\begin{aligned}\Lambda_{\min}(X^\top X) &= \min\{Z \circ (X^\top X) : \mathit{trace}(Z) = 1, Z \succcurlyeq 0\} \leqslant \hat{Z}^* \circ (X^\top X)\\ &\leqslant \hat{Z}^* \circ (\hat{X}^\top \hat{X}) + \hat{Z}^* \circ (-\Delta_{\mathcal{X}}^\top X - X^\top \Delta_{\mathcal{X}} + \Delta_{\mathcal{X}}^\top \Delta_{\mathcal{X}})\\ &\qquad (\text{due to } X = \hat{X} - \Delta_{\mathcal{X}})\\ &\leqslant \Lambda_{\min}(\hat{X}^\top \hat{X}) + \text{vec}(\hat{Z}^*)^\top \text{vec}(-\Delta_{\mathcal{X}}^\top X - X^\top \Delta_{\mathcal{X}} + \Delta_{\mathcal{X}}^\top \Delta_{\mathcal{X}})\\ &\leqslant \Lambda_{\min}(\hat{X}^\top \hat{X}) + \|\hat{Z}^*\|_2 \cdot (\|\Delta_{\mathcal{X}}^\top X\|_2 + \|X^\top \Delta_{\mathcal{X}}\|_2 + \|\Delta_{\mathcal{X}}^\top \Delta_{\mathcal{X}}\|_2)\\ &\leqslant \Lambda_{\min}(\hat{X}^\top \hat{X}) + (2\|\Delta_{\mathcal{X}}^\top X\|_2 + \|\Delta_{\mathcal{X}}^\top \Delta_{\mathcal{X}}\|_2), \qquad \text{(C.14)}\end{aligned}$$

where the last inequality is because of $\|\hat{Z}^*\|_2 \leqslant 1$ with exactly the same reason of (C.13). Now, combining (C.12) and (C.14) yields (C.11):

$$\begin{aligned}|\Lambda_{\min}(X^\top X) - \Lambda_{\min}(\hat{X}^\top \hat{X})| &\leqslant 2\|\Delta_{\mathcal{X}}^\top X\|_2 + \|\Delta_{\mathcal{X}}^\top \Delta_{\mathcal{X}}\|_2\\ &\leqslant \|\Delta_{\mathcal{X}}\|_2^2 + 2\|\Delta_{\mathcal{X}}\|_2\|X\|_2.\end{aligned}$$

Next, we show that $\hat{X}$ has a full column rank by proving that $\hat{X}^\top \hat{X}$ has a positive smallest eigenvalue and thus is invertible:

$$\begin{aligned}\Lambda_{\min}\{\hat{X}^\top \hat{X}\} &\geqslant \Lambda_{\min}\{X^\top X\} - \|\Delta_{\mathcal{X}}\|_2^2 - 2\|\Delta_{\mathcal{X}}\|_2\|X\|_2 \quad (\text{due to (C.11)})\\ &\geqslant L_\Lambda - 3U_{\mathcal{X}}\|\Delta_{\mathcal{X}}\|_2 \quad (\text{due to } \|\Delta_{\mathcal{X}}\|_2 \leqslant \delta_1 \leqslant U_{\mathcal{X}})\\ &\geqslant \frac{L_\Lambda}{2}. \quad \left(\text{due to } \|\Delta_{\mathcal{X}}\|_2 \leqslant \delta_1 \leqslant \frac{L_\Lambda}{6U_{\mathcal{X}}}\right)\end{aligned}$$

Last, we prove the Lipschitz continuity in Lemma C.2. Let $\Lambda_{\max}(\cdot)$ represents taking the largest eigenvalue. Then we have

$$\begin{aligned}&\|(\hat{X}^\top \hat{X})^{-1}(\hat{X}^\top \hat{Y}) - (X^\top X)^{-1}(X^\top Y)\|_2\\ &\leqslant \|(\hat{X}^\top \hat{X})^{-1}(\hat{X}^\top \hat{Y}) - (\hat{X}^\top \hat{X})^{-1}(X^\top Y)\|_2\\ &\quad + \|(\hat{X}^\top \hat{X})^{-1}(X^\top Y) - (X^\top X)^{-1}(X^\top Y)\|_2\\ &\leqslant \|(\hat{X}^\top \hat{X})^{-1}(\hat{X}^\top \hat{Y} - X^\top Y)\|_2\\ &\quad + \|(\hat{X}^\top \hat{X})^{-1}(X^\top X - \hat{X}^\top \hat{X})(X^\top X)^{-1}(X^\top Y)\|_2\\ &\leqslant \Lambda_{\max}(\hat{X}^\top \hat{X})^{-1} \cdot \|\hat{X}^\top \hat{Y} - X^\top Y\|_2\\ &\quad + \Lambda_{\max}(\hat{X}^\top \hat{X})^{-1} \cdot \|X^\top X - \hat{X}^\top \hat{X}\|_2 \cdot \Lambda_{\max}(X^\top X)^{-1} \cdot \|X^\top Y\|_2\\ &\leqslant \Lambda_{\min}^{-1}(\hat{X}^\top \hat{X}) \cdot \|\Delta_{\mathcal{X}}^\top Y + X^\top \Delta_{\mathcal{Y}} + \Delta_{\mathcal{X}}^\top \Delta_{\mathcal{Y}}\|_2\\ &\quad + \Lambda_{\min}^{-1}(\hat{X}^\top \hat{X}) \cdot \|X^\top X - \hat{X}^\top \hat{X}\|_2 \cdot \Lambda_{\min}^{-1}(X^\top X) \cdot U_{\mathcal{XY}}\\ &\leqslant \frac{2}{L_\Lambda} \cdot (\|Y\|_2\|\Delta_{\mathcal{X}}\|_2 + \|X\|_2\|\Delta_{\mathcal{Y}}\|_2 + \|\Delta_{\mathcal{X}}\|_2\|\Delta_{\mathcal{Y}}\|_2)\\ &\quad + \frac{2U_{\mathcal{XY}}}{L_\Lambda^2} \cdot (\|\Delta_{\mathcal{X}}\|_2^2 + 2\|\Delta_{\mathcal{X}}\|_2\|X\|_2)\\ &\leqslant \frac{2}{L_\Lambda} \cdot (U_{\mathcal{Y}}\|\Delta_{\mathcal{X}}\|_2 + 2U_{\mathcal{X}}\|\Delta_{\mathcal{Y}}\|_2) + \frac{2U_{\mathcal{XY}}}{L_\Lambda^2} \cdot (3U_{\mathcal{X}}\|\Delta_{\mathcal{X}}\|_2)\\ &\leqslant \left(\frac{2U_{\mathcal{Y}}}{L_\Lambda} + \frac{6U_{\mathcal{XY}}U_{\mathcal{X}}}{L_\Lambda^2}\right) \cdot \|\Delta_{\mathcal{X}}\|_2 + \left(\frac{4U_{\mathcal{X}}}{L_\Lambda}\right) \cdot \|\Delta_{\mathcal{Y}}\|_2\\ &\leqslant \max\left\{a\frac{2U_{\mathcal{Y}}}{L_\Lambda} + \frac{6U_{\mathcal{XY}}U_{\mathcal{X}}}{L_\Lambda^2}, \frac{4U_{\mathcal{X}}}{L_\Lambda}\right\} \cdot \sqrt{2}\|(\hat{X} - X, \hat{Y} - Y)\|_2\\ &\leqslant C_E \cdot \|(\hat{X} - X, \hat{Y} - Y)\|_2. \quad \square\end{aligned}$$

C.3.3. Step 3: Concentration Inequality of Intermediate Estimators.

We show that the intermediate estimators $(\hat{X}^\tau, \hat{Y}^\tau)$ are converging with the following rate.

Lemma C.3 (Intermediate Estimator Concentration). *For the MCC model $(\mathcal{N}, r, \theta)$ with parameter space Θ, define matrices X and Y through (C.6)–(C.10) with a fixed $\theta \in \Theta$. Also define the intermediate estimators $(\hat{X}^\tau, \hat{Y}^\tau)$ through (E-2)–(E-6). There exist constants $\omega_{\mathcal{XY}}, C_{\mathcal{XY}}, \zeta_{\mathcal{XY}} \in \mathbb{R}_{++}$ independent of θ such that*

$$\mathbb{P}[\|(\hat{X}^\tau, \hat{Y}^\tau) - (X, Y)\|_2 > \zeta] \leqslant C_{\mathcal{XY}} \cdot e^{-\omega_{\mathcal{XY}} \cdot \tau\zeta^2},$$
$$\forall \zeta \in (0, \zeta_{\mathcal{XY}}), \tau \in \mathbb{N}.$$

Particularly, $\omega_{\mathcal{XY}}$, $C_{\mathcal{XY}}$, and $\zeta_{\mathcal{XY}}$ can be defined as

$$C_{\mathcal{XY}} := 5N^2, \qquad \omega_{\mathcal{XY}} := \frac{L_\pi^4}{2^8(n^* + 1)N^5}, \qquad \zeta_{\mathcal{XY}} := \frac{4\sqrt{2n^* N^3}}{L_\pi}.$$

Our proof uses the Dvoretzky-Kiefer-Wolfowitz inequality for discrete distributions.

Lemma C.4 (Theorem 11.5 in Kosorok 2006). *For any i.i.d. sample $Z_1,\dots,Z_n$ with distribution $F(s)$ and empirical distribution $\hat{F}_n(s) := \frac{\sum_{i=1}^n 1\{Z_i \leqslant s\}}{n}$,*

$$\mathbb{P}\left[\sup_{s\in\mathbb{R}} |\hat{F}_n(s) - F(s)| > \zeta\right] \leqslant 2e^{-2n\zeta^2}, \quad \forall \zeta > 0, n \in \mathbb{N}.$$

Here $F(s)$ can be any continuous distribution or any discrete one with at most countable discontinuities.

Proof. Because $(\hat{X}^\tau, \hat{Y}^\tau)$ (respectively (X, Y)) is composed by $\{\hat{\pi}^\tau(i,S) | i \in \mathcal{N}_+, S \in \mathcal{S}_0 \cup \tilde{\mathcal{S}}_0\}$ (respectively, $\{\pi(i,S) | i \in \mathcal{N}_+, S \in \mathcal{S}_0 \cup \tilde{\mathcal{S}}_0\}$), it is useful to first explore the concentration inequality of $\{\hat{\pi}^\tau(i,S) | i \in \mathcal{N}_+, S \in \mathcal{S}_0 \cup \tilde{\mathcal{S}}_0\}$. For a fixed $\theta \in \Theta$ and a fixed $S \in \mathcal{S}_0 \cup \tilde{\mathcal{S}}_0$, the stochastic vector $\{\pi(i,S)\}_{i\in\mathcal{N}_+}$ forms a discrete distribution $F(s) := \sum_{i\in\mathcal{N}_+} \pi(i,S) 1\{i \leqslant s\}$. Then, $\{\hat{\pi}^\tau(i,S)\}_{i\in\mathcal{N}_+}$ can be viewed as jump sizes in the empirical distribution $\hat{F}_{n(\tau)}(s) := \sum_{i\in\mathcal{N}_+} 1\{i \leqslant s\} \cdot \frac{\sum_{t=1}^\tau 1\{S_t = S, Z_i^t = 1\}}{\sum_{t=1}^\tau 1\{S_t = S\}} = \sum_{i\in\mathcal{N}_+} 1\{i \leqslant s\}$ $\hat{\pi}^\tau(i,S)$, where $n(\tau)$ denotes the sample size $\sum_{t=1}^\tau 1\{S_t = S\}$, and $n(\tau) \geqslant \lfloor \frac{\tau}{d} \rfloor$ according to Algorithm 1. Then, $\hat{F}_{n(\tau)}(s)$ is an empirical distribution formed by an i.i.d. sample of size $n(\tau)$ with distribution $F(s)$. Thus, according to Lemma C.4,

$$\mathbb{P}\left[\sup_{s\in\mathbb{R}} |\hat{F}_{n(\tau)}(s) - F(s)| > \zeta\right] \leqslant 2e^{-2n(\tau)\zeta^2}, \quad \forall \zeta > 0, \tau \in \mathbb{N}. \tag{C.15}$$

For every $i \in \mathcal{N}_+$, we have

$$\begin{aligned} |\hat{\pi}^\tau(i,S) - \pi(i,S)| &= |(\hat{F}_{n(\tau)}(i) - \hat{F}_{n(\tau)}(i-1)) - (F(i) - F(i-1))| \\ &\leqslant |\hat{F}_{n(\tau)}(i) - F(i)| + |\hat{F}_{n(\tau)}(i-1) - F(i-1)|. \end{aligned} \tag{C.16}$$

Plugging (C.16) into (C.15) yields

$$\begin{aligned} &\mathbb{P}\left[\max_{i\in\mathcal{N}_+} |\hat{\pi}^\tau(i,S) - \pi(i,S)| > \zeta\right] \\ &= 1 - \mathbb{P}\left[\max_{i\in\mathcal{N}_+} |\hat{\pi}^\tau(i,S) - \pi(i,S)| \leqslant \zeta\right] \\ &\leqslant 1 - \mathbb{P}\left[\sup_{s\in\mathbb{R}} |\hat{F}_{n(\tau)}(s) - F(s)| \leqslant \frac{\zeta}{2}\right] \\ &\leqslant 1 - [1 - 2e^{-n(\tau)\frac{\zeta^2}{2}}] \leqslant 2e^{-n(\tau)\frac{\zeta^2}{2}} \\ &\leqslant 2e^{-\lfloor\frac{\tau}{d}\rfloor\frac{\zeta^2}{2}} \leqslant 2e^{-\|\frac{\tau}{N^2}\|\frac{\zeta^2}{2}}, \quad \forall \zeta > 0, \tau \in \mathbb{N}. \end{aligned} \tag{C.17}$$

Consider the union of the previous concentration inequalities across all $S \in \mathcal{S}_0 \cup \tilde{\mathcal{S}}_0$. We have that

$$\begin{aligned} &\mathbb{P}\left[\max_{S\in\mathcal{S}_0\cup\tilde{\mathcal{S}}_0} \max_{i\in\mathcal{N}_+} |\hat{\pi}^\tau(i,S) - \pi(i,S)| > \zeta\right] \\ &\leqslant \sum_{S\in\mathcal{S}_0\cup\tilde{\mathcal{S}}_0} \mathbb{P}\left[\max_{i\in\mathcal{N}_+} |\hat{\pi}^\tau(i,S) - \pi(i,S)| > \zeta\right] \\ &\leqslant 2N^2 e^{-\lfloor\frac{\tau}{N^2}\rfloor\frac{\zeta^2}{2}}, \quad \forall \zeta > 0, \tau \in \mathbb{N}. \end{aligned} \tag{C.18}$$

Next, we consider the concentration inequality of $\{\hat{\pi}^\tau(j,S|i) | i,j \in \mathcal{N}_+, S \in \mathcal{S}_0\}$. According to (E-3) and (C.8), the estimation errors occur only when $i \in \mathcal{N} \backslash S$. Recall that we have $\pi(i, S \cup \{i\}) \geqslant L_\pi$ for all $i \in \mathcal{N} \backslash S, S \in \mathcal{S}_0, \theta \in \Theta$. Then we can outline the concentration inequality when errors of $\{\hat{\pi}^\tau(i,S) | i \in \mathcal{N}_+, S \in \mathcal{S}_0 \cup \tilde{\mathcal{S}}_0\}$ are below $\frac{L_\pi}{2}$: for every $\zeta \leqslant \frac{L_\pi}{2}$, if $\max_{S\in\mathcal{S}_0\cup\tilde{\mathcal{S}}_0} \max_{i\in\mathcal{N}_+} |\hat{\pi}^\tau(i,S) - \pi(i,S)| \leqslant \zeta$, then for all $j \in \mathcal{N}_+, i \in \mathcal{N} \backslash S, S \in \mathcal{S}_0$, we have

$$\begin{aligned} &|\hat{\pi}^\tau(j,S|i) - \pi(j,S|i)| \\ &= \left|\frac{\hat{\pi}^\tau(j,S) - \hat{\pi}^\tau(j,S\cup\{i\})}{\hat{\pi}^\tau(i,S\cup\{i\})} - \frac{\pi(j,S) - \pi(j,S\cup\{i\})}{\pi(i,S\cup\{i\})}\right| \\ &\leqslant \left|\frac{\hat{\pi}^\tau(j,S) - \hat{\pi}^\tau(j,S\cup\{i\})}{\hat{\pi}^\tau(i,S\cup\{i\})} - \frac{\hat{\pi}^\tau(j,S) - \hat{\pi}^\tau(j,S\cup\{i\})}{\pi(i,S\cup\{i\})}\right| \\ &\quad + \left|\frac{\hat{\pi}^\tau(j,S) - \hat{\pi}^\tau(j,S\cup\{i\})}{\pi(i,S\cup\{i\})} - \frac{\pi(j,S) - \pi(j,S\cup\{i\})}{\pi(i,S\cup\{i\})}\right| \\ &\leqslant |\hat{\pi}^\tau(j,S) - \hat{\pi}^\tau(j,S\cup\{i\})| \cdot \left|\frac{1}{\hat{\pi}^\tau(i,S\cup\{i\})} - \frac{1}{\pi(i,S\cup\{i\})}\right| \\ &\quad + \left|\frac{1}{\pi(i,S\cup\{i\})}\right| \cdot |[\hat{\pi}^\tau(j,S) - \hat{\pi}^\tau(j,S\cup\{i\})] \\ &\quad - [\pi(j,S) - \pi(j,S\cup\{i\})]| \\ &\leqslant \left|\frac{1}{\hat{\pi}^\tau(i,S\cup\{i\})} - \frac{1}{\pi(i,S\cup\{i\})}\right| \\ &\quad + \frac{1}{L_\pi} \cdot |[\hat{\pi}^\tau(j,S) - \pi(j,S)] - [\hat{\pi}^\tau(j,S\cup\{i\}) - \pi(j,S\cup\{i\})]| \\ &\qquad\qquad (\text{due to } \hat{\pi}^\tau(j,S), \hat{\pi}^\tau(j,S\cup\{i\}) \in [0,1]; \pi(i,S\cup\{i\}) \geqslant L_\pi) \\ &\leqslant \left|\frac{\hat{\pi}^\tau(i,S\cup\{i\}) - \pi(i,S\cup\{i\})}{\hat{\pi}^\tau(i,S\cup\{i\})\pi(i,S\cup\{i\})}\right| \\ &\quad + \frac{|\hat{\pi}^\tau(j,S) - \pi(j,S)| + |\hat{\pi}^\tau(j,S\cup\{i\}) - \pi(j,S\cup\{i\})|}{L_\pi} \\ &\leqslant \frac{\zeta}{\left(\frac{L_\pi^2}{2}\right)} + \frac{2\zeta}{L_\pi} \qquad \left(\text{due to } \hat{\pi}^\tau(i,S\cup\{i\}) \geqslant \pi(i,S\cup\{i\}) - \zeta \right. \\ &\qquad\qquad \left. \geqslant L_\pi - \frac{L_\pi}{2} \geqslant \frac{L_\pi}{2}\right) \\ &\leqslant \left(\frac{2 + 2L_\pi}{L_\pi^2}\right)\zeta \leqslant \frac{4}{L_\pi^2}\zeta. \qquad (\text{due to } L_\pi \leqslant 1) \end{aligned}$$

Additionally, when $i \notin \mathcal{N} \backslash S, |\hat{\pi}^\tau(j,S|i) - \pi(j,S|i)| = 0$. Thus, for all $j,i \in \mathcal{N}_+, S \in \mathcal{S}_0$, we have $|\hat{\pi}^\tau(j,S|i) - \pi(j,S|i)| \leqslant \frac{4}{L_\pi^2}\zeta$. Using this property and (C.18), we have

$$\begin{aligned} &\mathbb{P}\left[\max_{S\in\mathcal{S}_0} \max_{i,j\in\mathcal{N}_+} |\hat{\pi}^\tau(j,S|i) - \pi(j,S|i)| > \zeta\right] \\ &\leqslant \mathbb{P}\left[\max_{S\in\mathcal{S}_0\cup\tilde{\mathcal{S}}_0} \max_{i\in\mathcal{N}_+} |\hat{\pi}^\tau(i,S) - \pi(i,S)| > \frac{L_\pi^2}{4}\zeta\right] \\ &\leqslant 2N^2 e^{-\frac{L_\pi^4}{32}\lfloor\frac{\tau}{N^2}\rfloor\zeta^2}, \quad \forall \zeta \in \left(0, \frac{2}{L_\pi}\right), \forall \tau \in \mathbb{N}. \end{aligned} \tag{C.19}$$

Last, recall the definition of $(\hat{X}^\tau, \hat{Y}^\tau)$ via differencing $\{\hat{\pi}^\tau(j,S|i) | i,j \in \mathcal{N}_+, S \in \mathcal{S}_0\}$ and $\{\hat{\pi}^\tau(i,S) | i \in \mathcal{N}_+, S \in \mathcal{S}_0 \cup \tilde{\mathcal{S}}_0\}$ in

(E-4)–(E-6). We have

$$
\begin{aligned}
&\mathbb{P}[\|(\hat{X}^{\tau},\hat{Y}^{\tau})-(X,Y)\|_2>\zeta]\\
&\leqslant \mathbb{P}\left[\|(\hat{X}^{\tau},\hat{Y}^{\tau})-(X,Y)\|_\infty>\frac{\zeta}{\sqrt{2(n^*+1)N^3}}\right]\\
&\qquad\qquad\qquad\qquad\text{(due to the dimension of } (X,Y))\\
&\leqslant \mathbb{P}\left[\max_{S\in\mathscr{S}_0}\max_{i,j\in\mathcal{N}_+}|\hat{\pi}^{\tau}(j,S|i)-\pi(j,S|i)|>\frac{\zeta}{2\sqrt{2(n^*+1)N^3}}\right]+\\
&\quad \mathbb{P}\left[\max_{S\in\mathscr{S}_0\cup\tilde{\mathscr{S}}_0}\max_{i\in\mathcal{N}_+}|\hat{\pi}^{\tau}(i,S)-\pi(i,S)|>\frac{\zeta}{2\sqrt{2(n^*+1)N^3}}\right]\\
&\leqslant 2N^2e^{-\frac{L_\pi^4}{2^8(n^*+1)N^3}\lfloor\frac{\tau}{N^2}\rfloor\zeta^2}+2N^2e^{-\frac{1}{2^4(n^*+1)N^3}\|\frac{\tau}{N^2}\|\zeta^2}\\
&\leqslant 4N^2e^{-\frac{L_\pi^4}{2^8(n^*+1)N^3}\lfloor\frac{\tau}{N^2}\rfloor\zeta^2},\\
&\qquad\forall\zeta\in\left(0,\frac{4\sqrt{2n^*N^3}}{L_\pi}\right)\subset\left(0,\frac{4\sqrt{2(n^*+1)N^3}}{L_\pi}\right),\ \forall\tau\in\mathbb{N}.
\end{aligned}
$$

To remove the floor operation, we have

$$
\begin{aligned}
&\mathbb{P}[\|(\hat{X}^{\tau},\hat{Y}^{\tau})-(X,Y)\|_2>\zeta]\leqslant 4N^2e^{-\frac{L_\pi^4}{2^8(n^*+1)N^3}\lfloor\frac{\tau}{N^2}\rfloor\zeta^2}\\
&\leqslant 4N^2e^{-\frac{L_\pi^4}{2^8(n^*+1)N^3}\left(\frac{\tau}{N^2}-1\right)\zeta^2}\leqslant 4N^2e^{-\frac{L_\pi^4}{2^8(n^*+1)N^5}\tau\zeta^2+\frac{L_\pi^4}{2^8(n^*+1)N^3}\zeta^2}\\
&\leqslant 4N^2e^{-\frac{L_\pi^4}{2^8(n^*+1)N^5}\tau\zeta^2+\frac{L_\pi^2}{2^3}}\qquad\left(\text{due to }\zeta\leqslant\frac{4\sqrt{2n^*N^3}}{L_\pi}\right)\\
&\leqslant 4e^{\frac{L_\pi^2}{8}}N^2e^{-\frac{L_\pi^4}{2^8(n^*+1)N^5}\tau\zeta^2}\leqslant 4e^{\frac{1}{8}}N^2e^{-\frac{L_\pi^4}{2^8(n^*+1)N^5}\tau\zeta^2}\\
&\leqslant 5N^2e^{-\frac{L_\pi^4}{2^8(n^*+1)N^5}\tau\zeta^2},\qquad\forall\zeta\in\left(0,\frac{4\sqrt{2n^*N^3}}{L_\pi}\right),\ \forall\tau\in\mathbb{N},
\end{aligned}
$$

which is exactly the concentration bound in Lemma C.3. □

C.3.4. Step 4: Concentration Inequality of Main Estimators.

We now combine results from Steps 2 and 3 and develop them into a concentration inequality for $\tilde{\theta}^{\tau}$. Recall Lemmas C.2 and C.3. Fix any $\theta\in\Theta$ and its corresponding matrices X and Y defined through (C.6)–(C.10). For all $\tau\in\mathbb{N}$, if $\|(\hat{X}^{\tau},\hat{Y}^{\tau})-(X,Y)\|_2\leqslant\zeta\leqslant\delta_1$, then $\|\hat{\theta}^{\tau}-\theta\|_2=\|(\hat{X}^{\tau^\top}\hat{X}^{\tau})^{-1}(\hat{X}^{\tau^\top}\hat{Y}^{\tau})-(X^\top X)^{-1}(X^\top Y)\|_2\leqslant C_E\zeta$ by Lemma C.2. Then, by Lemma C.3, for every $\zeta\in(0,\min\{C_E\delta_1,C_E\zeta_{\mathcal{XY}}\})$,

$$
\mathbb{P}[\|\hat{\theta}^{\tau}-\theta\|_2>\zeta]\leqslant\mathbb{P}\left[\|(\hat{X}^{\tau},\hat{Y}^{\tau})-(X,Y)\|_2>\frac{\zeta}{C_E}\right]\leqslant C_{\mathcal{XY}}\cdot e^{-\frac{\omega_{\mathcal{XY}}}{C_E^2}\cdot\tau\zeta^2},\qquad\forall\tau\in\mathbb{N}.
$$

Recall (5). We have that for every $\zeta\in(0,2NC_E\min\{\delta_1,\zeta_{\mathcal{XY}}\})=\left(0,2NC_E\min\left\{\delta_1,\frac{4\sqrt{2n^*N^3}}{L_\pi}\right\}\right)$,

$$
\begin{aligned}
\mathbb{P}[\|\tilde{\theta}^{\tau}-\theta\|_1>\zeta]&\leqslant\mathbb{P}\left[\|\hat{\theta}^{\tau}-\theta\|_1>\frac{\zeta}{2}\right]\\
&\leqslant\mathbb{P}\left[\|\hat{\theta}^{\tau}-\theta\|_2>\frac{\zeta}{2N}\right]\leqslant C_{\mathcal{XY}}\cdot e^{-\frac{\omega_{\mathcal{XY}}}{4N^2C_E^2}\cdot\tau\zeta^2}\\
&\leqslant 5N^2e^{-\frac{L_\pi^4}{2^{10}(n^*+1)C_E^2N^7}\cdot\tau\zeta^2},\qquad\forall\tau\in\mathbb{N}.
\end{aligned}
$$

Then we can define let $\phi:=5$, $\omega:=\frac{L_\pi^4}{2^{10}(n^*+1)C_E^2}$, and $\zeta_0:=2C_E\min\left\{\delta_1,\frac{4\sqrt{2n^*N^3}}{L_\pi}\right\}$. Then we have $\mathbb{P}[\|\tilde{\theta}^{\tau}-\theta\|_1>\zeta]\leqslant\phi N^2e^{-\omega N^{-7}\tau\zeta^2}$, $\forall\zeta\in(0,\zeta_0N),\tau\in\mathbb{N}$. This completes the proof of Lemma 5.

Appendix D. Proof for Proposition 2

Proof. With γ specified as $\frac{1}{\alpha}$, we have $\zeta'=\min\left\{\frac{r(1-\alpha)}{C_L(1+\alpha)},\zeta_0^*\right\}$. The α-regret of policy P_2' can be divided into three parts: (*a*) for $t\leqslant\tau$, the α-regret associated with every single customer is bounded by $r_{\max}$; (*b*) for $t>\tau$ such that $\|\tilde{\theta}^{\tau}-\theta\|_1>\zeta'$, the α-regret associated with every single customer is also bounded by $r_{\max}$; and (*c*) for $t>\tau$ such that $\|\tilde{\theta}^{\tau}-\theta\|_1\leqslant\zeta'$, the α-regret associated with every single customer is zero because $\|\tilde{\theta}^{\tau}-\theta\|_1\leqslant\zeta'\leqslant\frac{r(1-\alpha)}{C_L(1+\alpha)}$, that is, $\tilde{\theta}^{\tau}$ falls into the regret-free region. Thereby, $S_t=S^{\gamma\alpha}(\tilde{\theta}^{\tau})=S^*(\tilde{\theta}^{\tau})(t>\tau)$ will be an α-optimal assortment under the true parameter θ according to Lemma 4. Thus, the α-regret of policy P_2' is bounded by

$$
\begin{aligned}
\mathsf{Reg}^{\alpha}_{P_2'}(T,\theta)&\leqslant r_{\max}\tau+r_{\max}(T-\tau)\cdot\mathbb{P}[\|\tilde{\theta}^{\tau}-\theta\|_1>\zeta']\\
&\leqslant r_{\max}\tau+r_{\max}T\cdot\phi^*e^{-\omega^*\tau\zeta'2}\qquad\text{(due to (11))}\\
&\leqslant r_{\max}\lceil\psi\log T\rceil+r_{\max}T\cdot\phi^*e^{-\omega^*\|\psi\log T\|\zeta'2}\\
&\leqslant r_{\max}\psi\log T+r_{\max}+r_{\max}T\cdot\phi^*e^{-\omega^*\psi\log T\zeta'2}\\
&\leqslant r_{\max}\psi\log T+r_{\max}+r_{\max}\phi^*T^{1-\omega^*\psi\zeta'2}\\
&\leqslant r_{\max}\psi\log T+r_{\max}+r_{\max}\phi^*\quad\left(\text{due to }\psi=\frac{1}{\omega^*\zeta'2}\right)\\
&\leqslant(r_{\max}\psi+r_{\max}+r_{\max}\phi^*)\log T\leqslant\kappa_2\log T.\\
&\qquad\qquad\qquad\qquad\text{(due to } T\geqslant 3).
\end{aligned}
$$

Because of (13), we have $\mathsf{Reg}_{P_2'}(T,\theta)=\mathsf{Reg}^{\alpha}_{P_2'}(T,\theta)\leqslant\kappa_2\log T$, which is exactly the result of Proposition 2. □

Appendix E. Simplified Algorithms for Unconstrained Model

This section discusses the unconstrained setting (i.e., $\mathscr{S}=2^{\mathcal{N}}$) and show how this condition will simplify our algorithm and regret bounds.

We mainly modify the exploration phase, during which we repeatedly present the assortments in $\mathscr{S}_U:=\{\mathcal{N}\}\cup\{\mathcal{N}\backslash\{i\}\}_{i\in\mathcal{N}}$. Each presented assortment has the cardinality of $N-1$ or N. The number of assortments $d_U:=|\mathscr{S}_U|=N+1$.

Example E.1. Suppose $\mathcal{N}=\{1,2,3,4\}$. Then $\mathscr{S}_U=\{\{1,2,3,4\},\{1,2,3\},\{1,2,4\},\{1,3,4\},\{2,3,4\}\}$, which we denote as $\{A_0,A_1,\ldots,A_4\}$ accordingly. In the exploration phase, the algorithm will sequentially offer $A_0,A_1,\ldots,A_4,A_0,A_1,\ldots,A_4,\ldots$ to customers until the separation period τ. □

At the separation period τ, θ_0 is naturally estimated by

$$
\hat{\theta}_0^{\tau}:=\text{vec}(\{\hat{\rho}_{ij}^{\tau}\equiv 0\}_{(i,j)\in I_0}).\tag{E.1}
$$

Nontrivial parameter θ_{++} is estimated in two steps.

Step 1: The choice probabilities are estimated by

$$
\hat{\pi}^{\tau}(i,S):=\frac{\sum_{t=1}^{\tau}\mathbb{1}\{S_t=S,Z_i^t=1\}}{\sum_{t=1}^{\tau}\mathbb{1}\{S_t=S\}},\quad i\in\mathcal{N}_+,S\in\mathscr{S}_U.\tag{E.2}
$$

Step 2: The arrival and transition probabilities are estimated as

$$
\begin{aligned}
&\hat{\lambda}_i^{\tau}:=\hat{\pi}^{\tau}(i,\mathcal{N}),\ i\in\mathcal{N};\\
&\hat{\rho}_{ij}^{\tau}:=\frac{\hat{\pi}^{\tau}(j,\mathcal{N}\backslash\{i\})-\hat{\pi}^{\tau}(j,\mathcal{N})}{\hat{\pi}^{\tau}(i,\mathcal{N})},\ (i,j)\in\mathcal{N}^2\backslash I_0.
\end{aligned}\tag{E.3}
$$

Step 3: Recall $\theta_{++} := \mathrm{vec}(\{\lambda_i\}_{i\in\mathcal{N}} \cup \{\rho_{ij}\}_{(i,j)\in\mathcal{N}^2\backslash I_0})$. θ_{++} is naturally estimated as

$$\hat{\theta}_{++}^{\tau} := \mathrm{vec}(\{\hat{\lambda}_i^{\tau}\}_{i\in\mathcal{N}} \cup \{\hat{\rho}_{ij}^{\tau}\}_{(i,j)\in\mathcal{N}^2\backslash I_0}). \tag{E.4}$$

Combining (E.1) and (E.4), we obtain the following rounded estimator $\tilde{\theta}^{\tau}$ for θ:

$$\hat{\theta}^{\tau} := \mathrm{vec}(\hat{\theta}_0^{\tau}, \hat{\theta}_{++}^{\tau}), \qquad \tilde{\theta}^{\tau} := \underset{\theta'\in\Theta}{\mathrm{argmin}} \|\theta' - \hat{\theta}^{\tau}\|_1. \tag{E.5}$$

The modified FastLinETC under the unconstrained MCC model is summarized in Algorithm E.1.

Algorithm E.1 (Modified FastLinETC $P_U(N,T,\tau)$ for Unconstrained MCC Model)

```
Input: integer $\tau$.
Output: offered assortments $\{S_t\}_{t=1}^{T}$.
*Phase* 1. Exploration:
Define the exploration assortments $\mathscr{S}_U := \{\mathcal{N}\} \cup \{\mathcal{N}\backslash\{i\}\}_{i\in\mathcal{N}}$ and denote them as $\{A_0, A_1, \ldots, A_N\}$.
**for** $t \in \{1,2,\ldots,\tau\}$ **do**
    Define $k_t := (t-1) \bmod d_U$, and offer $S_t = A_{k_t}$ to customer $t$.
    Observe the customer purchase decisions $Z^t = (Z_0^t, Z_1^t, \ldots, Z_N^t)$.
**end for**
Compute choice probability estimators $\hat{\pi}^{\tau}(i,S)$ for all $i \in \mathcal{N}_+, S \in \mathscr{S}_U$ via (E.2).
Compute arrival and transition probability estimators $\{\hat{\lambda}_i^{\tau}\}_{i\in\mathcal{N}}$ and $\{\hat{\rho}_{ij}^{\tau}\}_{(i,j)\in\mathcal{N}^2\backslash I_0}$ via (E.3).
Compute the rounded MCC parameter estimator $\tilde{\theta}^{\tau}$ via (E.1), (E.4), and (E.5).
*Phase* 2. Exploitation:
To all remaining $T-\tau$ customers, offer $S^*(\tilde{\theta}^{\tau})$.
```

We have the following instance-independent upper bounds on the policy regret associated with Algorithm E.1.

Theorem E.1 (Regret of Unconstrained Policy). *Suppose Assumptions 1 and 2 hold and the possible assortments $\mathscr{S} = 2^{\mathcal{N}}$. Let $\nu > 0$ be an arbitrary constant. There exist $\kappa_3, T_3 \in O(\mathrm{poly}(N))$ such that by letting $\tau = \lceil \nu T^{\frac{2}{3}} \log T \rceil$ and policy P_3 be defined by Algorithm E.1, the regret associated with policy P_3 at any time $T \geqslant T_3$ is bounded as*

$$\mathrm{Reg}_{P_3}(T,\theta) \leqslant \kappa_3 T^{\frac{2}{3}} \log T,$$

where κ_3 and T_3 are constants independent of the MCC parameter θ.

Proof. Because the exploitation optimality gap in Lemma 4 still holds, we only need to construct a new concentration inequality for $\tilde{\theta}^{\tau}$ in place of Lemma 5. We start from concentration inequalities for $\{\hat{\pi}^{\tau}(i,S)\}_{i\in\mathcal{N}_+, S\in\mathscr{S}_U}$. Using the same argument for (C.18), we obtain

$$\begin{aligned}
&\mathbb{P}\left[\max_{S\in\mathscr{S}_U}\max_{i\in\mathcal{N}_+} |\hat{\pi}^{\tau}(i,S) - \pi(i,S)| > \zeta\right] \\
&\leqslant \sum_{S\in\mathscr{S}_U} \mathbb{P}\left[\max_{i\in\mathcal{N}_+} |\hat{\pi}^{\tau}(i,S) - \pi(i,S)| > \zeta\right] \\
&\leqslant (N+1)\cdot 2e^{-\lfloor\frac{\tau}{N+1}\rfloor\frac{\zeta^2}{2}}, \quad \forall \zeta > 0, \tau \in \mathbb{N}.
\end{aligned} \tag{E.6}$$

This indicates

$$\begin{aligned}
\mathbb{P}\left[\max_{i\in\mathcal{N}} |\hat{\lambda}_i^{\tau} - \lambda_i| > \zeta\right] &\leqslant \mathbb{P}\left[\max_{S\in\mathscr{S}_U}\max_{i\in\mathcal{N}_+} |\hat{\pi}^{\tau}(i,S) - \pi(i,S)| > \zeta\right] \\
&\leqslant (N+1)\cdot 2e^{-\lfloor\frac{\tau}{N+1}\rfloor\frac{\zeta^2}{2}}, \quad \forall \zeta > 0, \tau \in \mathbb{N}.
\end{aligned} \tag{E.7}$$

Next we consider the concentration inequality of $\{\hat{\rho}_{ij}^{\tau} | (i,j) \in \mathcal{N}^2\backslash I_0\}$. We define $L_\lambda := \inf_{\theta\in\Theta}\min_{i\in\mathcal{N}}\lambda_i$, which is a constant independent of $\theta \in \Theta$. Recall that $\pi(i,\mathcal{N}) = \lambda_i \geqslant L_\lambda$ for all $i \in \mathcal{N}, \theta \in \Theta$. We also define assortment $A_i := \mathcal{N}\backslash\{i\} \in \mathscr{S}_U, i \in \mathcal{N}$. Then we can outline the concentration inequality for $\hat{\rho}_{ij}^{\tau}$ when errors of $\{\hat{\pi}^{\tau}(i,S) | i \in \mathcal{N}_+, S \in \mathscr{S}_U\}$ are below $\frac{L_\lambda}{2}$: for every $\zeta \leqslant \frac{L_\lambda}{2}$, if $\max_{S\in\mathscr{S}_U}\max_{i\in\mathcal{N}_+} |\hat{\pi}^{\tau}(i,S) - \pi(i,S)| \leqslant \zeta$, then for all $(i,j) \in \mathcal{N}^2\backslash I_0$, we have

$$\begin{aligned}
|\hat{\rho}_{ij}^{\tau} - \rho_{ij}| &= \left|\frac{\hat{\pi}^{\tau}(j,A_i) - \hat{\pi}^{\tau}(j,\mathcal{N})}{\hat{\pi}^{\tau}(i,\mathcal{N})} - \frac{\pi(j,A_i) - \pi(j,\mathcal{N})}{\pi(i,\mathcal{N})}\right| \\
&\leqslant \left|\frac{\hat{\pi}^{\tau}(j,A_i) - \hat{\pi}^{\tau}(j,\mathcal{N})}{\hat{\pi}^{\tau}(i,\mathcal{N})} - \frac{\hat{\pi}^{\tau}(j,A_i) - \hat{\pi}^{\tau}(j,\mathcal{N})}{\pi(i,\mathcal{N})}\right| \\
&\quad + \left|\frac{\hat{\pi}^{\tau}(j,A_i) - \hat{\pi}^{\tau}(j,\mathcal{N})}{\pi(i,\mathcal{N})} - \frac{\pi(j,A_i) - \pi(j,\mathcal{N})}{\pi(i,\mathcal{N})}\right| \\
&\leqslant |\hat{\pi}^{\tau}(j,A_i) - \hat{\pi}^{\tau}(j,\mathcal{N})| \cdot \left|\frac{1}{\hat{\pi}^{\tau}(i,\mathcal{N})} - \frac{1}{\pi(i,\mathcal{N})}\right| \\
&\quad + \left|\frac{[\hat{\pi}^{\tau}(j,A_i) - \hat{\pi}^{\tau}(j,\mathcal{N})] - [\pi(j,A_i) - \pi(j,\mathcal{N})]}{\pi(i,\mathcal{N})}\right| \\
&\leqslant \left|\frac{1}{\hat{\pi}^{\tau}(i,\mathcal{N})} - \frac{1}{\pi(i,\mathcal{N})}\right| \\
&\quad + \frac{|[\hat{\pi}^{\tau}(j,A_i) - \pi(j,A_i)] - [\hat{\pi}^{\tau}(j,\mathcal{N}) - \pi(j,\mathcal{N})]|}{L_\lambda} \\
&\qquad (\text{due to } \hat{\pi}^{\tau}(j,A_i), \hat{\pi}^{\tau}(j,\mathcal{N}) \in [0,1]; \pi(i,\mathcal{N}) \geqslant L_\lambda) \\
&\leqslant \left|\frac{\hat{\pi}^{\tau}(i,\mathcal{N}) - \pi(i,\mathcal{N})}{\hat{\pi}^{\tau}(i,\mathcal{N})\pi(i,\mathcal{N})}\right| \\
&\quad + \frac{|\hat{\pi}^{\tau}(j,A_i) - \pi(j,A_i)| + |\hat{\pi}^{\tau}(j,\mathcal{N}) - \pi(j,\mathcal{N})|}{L_\lambda} \\
&\leqslant \frac{\zeta}{\left(\frac{L_\lambda^2}{2}\right)} + \frac{2\zeta}{L_\lambda} \leqslant \left(\frac{2 + 2L_\lambda}{L_\lambda^2}\right)\zeta \leqslant \frac{4}{L_\lambda^2}\zeta. \\
&\qquad \left(\text{due to } \hat{\pi}^{\tau}(i,\mathcal{N}) \geqslant \pi(i,\mathcal{N}) - \zeta \geqslant L_\lambda - \frac{L_\lambda}{2} \geqslant \frac{L_\lambda}{2}\right).
\end{aligned}$$

Using this property and (E.6), we have

$$\begin{aligned}
&\mathbb{P}\left[\max_{(i,j)\in\mathcal{N}^2\backslash I_0} |\hat{\rho}_{ij}^{\tau} - \rho_{ij}| > \zeta\right] \\
&\leqslant \mathbb{P}\left[\max_{S\in\mathscr{S}_U}\max_{i\in\mathcal{N}_+} |\hat{\pi}^{\tau}(i,S) - \pi(i,S)| > \frac{L_\lambda^2}{4}\zeta\right] \\
&\leqslant (N+1)\cdot 2e^{-\frac{L_\lambda^4}{32}\lfloor\frac{\tau}{N+1}\rfloor\zeta^2}, \quad \forall \zeta \in \left(0, \frac{2}{L_\lambda}\right), \ \forall \tau \in \mathbb{N}.
\end{aligned}$$

Combining this with (E.7), we have

$$
\begin{aligned}
\mathbb{P}[\|\hat{\theta}^{\tau} - \theta\|_1 > \zeta] &\leqslant \mathbb{P}\left[\|\hat{\theta}^{\tau} - \theta\|_\infty > \frac{\zeta}{N^2}\right] \\
&\qquad \text{(due to the dimensions of } \theta_{++}, \theta) \\
&\leqslant \mathbb{P}\left[\max_{i\in\mathcal{N}} |\hat{\lambda}_i^{\tau} - \lambda_i| > \frac{\zeta}{N^2}\right] \\
&\quad + \mathbb{P}\left[\max_{(i,j)\in\mathcal{N}^2\setminus I_0} |\hat{\rho}_{ij}^{\tau} - \rho_{ij}| > \frac{\zeta}{N^2}\right] \\
&\leqslant (N+1)\cdot 2e^{-\frac{1}{2N^4}\lfloor\frac{\tau}{N+1}\rfloor\zeta^2} + (N+1)\cdot 2e^{-\frac{L_\lambda^4}{32N^4}\|\frac{\tau}{N+1}\|\zeta^2} \\
&\leqslant 8Ne^{-\frac{L_\lambda^4}{32N^4}\lfloor\frac{\tau}{N+1}\rfloor\zeta^2}, \quad \forall\zeta\in\left(0, \frac{2N^2}{L_\lambda}\right),\ \forall\tau\in\mathbb{N}.
\end{aligned}
$$

To remove the floor operation, we have

$$
\begin{aligned}
&\mathbb{P}[\|\hat{\theta}^{\tau} - \theta\|_1 > \zeta] \leqslant 8Ne^{-\frac{L_\lambda^4}{32N^4}\lfloor\frac{\tau}{N+1}\rfloor\zeta^2} \leqslant 8Ne^{-\frac{L_\lambda^4}{32N^4}(\frac{\tau}{N+1}-1)\zeta^2} \leqslant 8Ne^{-\frac{L_\lambda^4}{32N^4}\frac{\tau}{N+1}\zeta^2 + \frac{L_\lambda^4}{32N^4}\zeta^2} \\
&\leqslant 8Ne^{\frac{L_\lambda^4}{32N^4}\zeta^2}\cdot e^{-\frac{L_\lambda^4}{32N^4}\frac{\tau}{N+1}\zeta^2} \leqslant 8e^{\frac{L_\lambda^2}{8}}Ne^{-\frac{L_\lambda^4}{32N^4}\frac{\tau}{N+1}\zeta^2} \quad \left(\text{due to } \zeta < \frac{2N^2}{L_\lambda}\right) \\
&\leqslant 10Ne^{-\frac{L_\lambda^4}{64N^5}\tau\zeta^2}, \quad \forall\zeta\in\left(0, \frac{2N^2}{L_\lambda}\right), \forall\tau\in\mathbb{N}.
\end{aligned}
$$

Therefore, we have

$$
\mathbb{P}[\|\tilde{\theta}^{\tau} - \theta\|_1 > \zeta] \leqslant \mathbb{P}\left[\|\hat{\theta}^{\tau} - \theta\|_1 > \frac{\zeta}{2}\right] \leqslant 10Ne^{-\frac{L_\lambda^4}{2^8N^5}\tau\zeta^2}, \quad \forall\zeta\in\left(0, \frac{4N^2}{L_\lambda}\right),\ \forall\tau\in\mathbb{N}.
$$

By letting $\phi := 10, \omega := \frac{L_\lambda^4}{2^8}, \zeta_0 := \frac{4}{L_\lambda}$, we have the following concentration inequality in place of Lemma 5: For all $\zeta\in(0, \zeta_0N^2)$,

$$
\mathbb{P}[\|\tilde{\theta}^{\tau} - \theta\|_1 > \zeta] \leqslant \phi Ne^{-\frac{\omega\tau\zeta^2}{N^5}}, \qquad \tau\in\mathbb{N}. \tag{E.8}
$$

Last, we prove the regret bounds in Theorem E.1 and show that the following constants T_3 and κ_3 in these bounds are independent of $\theta\in\Theta$.

$$
\begin{aligned}
&\kappa_3 := 2r_{\max}\nu + 2r_{\max}\nu^{-1}\phi N + 3\omega^{-\frac{1}{2}}C_LN^2\left(1 + \frac{3N}{\nu} + 2N\right)^{\frac{1}{2}}, \\
&T_3 := \max\left\{2, 2^{\frac{3}{2}}\nu^{-\frac{3}{2}}, \zeta_0^{-3}\omega^{-\frac{3}{2}}(1 + \frac{3N}{\nu} + 2N)^{\frac{3}{2}}\right\}.
\end{aligned} \tag{E.9}
$$

Let us define $\phi^* := \phi N, \omega^* := \omega N^{-5}, \zeta_0^* := \zeta_0N^2$ so that (E.2) can be written as

$$
\mathbb{P}[\|\tilde{\theta}^{\tau} - \theta\|_1 > \zeta] \leqslant \phi^* e^{-\omega^*\tau\zeta^2}, \qquad \zeta\in(0, \zeta_0^*), \tau\in\mathbb{N}, \tag{E.10}
$$

and we have $\kappa_3 \geqslant 2r_{\max}\nu + 2r_{\max}\nu^{-1}\phi^* + 3(\frac{3}{\nu}+2)^{\frac{1}{2}}\omega^{*-\frac{1}{2}}C_L$, $T_3 \geqslant (\frac{3}{\nu}+2)^{\frac{3}{2}}\zeta_0^{*-3}\omega^{*-\frac{3}{2}}$.

Because $T \geqslant T_3 \geqslant \max\left\{2, 2^{\frac{3}{2}}\nu^{-\frac{3}{2}}\right\}$, we have $\tau = \lceil\nu T^{\frac{2}{3}}\log T\rceil \geqslant \|\nu T_3^{\frac{2}{3}}\log T_3\| \geqslant \|2\log 2\| \geqslant 2$, which further indicates $\sqrt{\frac{\log\tau}{\omega^*\tau}} > 0$. In addition, we have

$$
\begin{aligned}
\sqrt{\frac{\log\tau}{\omega^*\tau}} &= \sqrt{\frac{\log\lceil\mu T^{\frac{2}{3}}\log T\rceil}{\omega^*\lceil\mu T^{\frac{2}{3}}\log T\rceil}} \leqslant \sqrt{\frac{\log(2\mu T^{\frac{2}{3}}\log T)}{\omega^*\mu T^{\frac{2}{3}}\log T}} \leqslant \sqrt{\frac{\log\mu T^3}{\omega^*\mu T^{\frac{2}{3}}\log T}} \\
&\leqslant \sqrt{\frac{3\log T + \log\mu}{\omega^*\mu T^{\frac{2}{3}}\log T}} \leqslant \sqrt{\frac{3\log T}{\omega^*\mu T^{\frac{2}{3}}\log T} + \frac{\log\mu}{\omega^*\mu T^{\frac{2}{3}}\log T}} \\
&\leqslant \sqrt{\frac{3}{\omega^*\mu T^{\frac{2}{3}}} + \frac{2}{\omega^* T^{\frac{2}{3}}}} \leqslant \left(\frac{3}{\mu}+2\right)^{\frac{1}{2}}\cdot\omega^{*-\frac{1}{2}}T^{-\frac{1}{3}} \leqslant \zeta_0^*. \\
&\qquad \left(\text{due to} T \geqslant T_1 \geqslant \left(\frac{3}{\mu}+2\right)^{\frac{3}{2}}\zeta_0^{*-3}\omega^{*-\frac{3}{2}}\right).
\end{aligned} \tag{E.11}
$$

Because $\sqrt{\frac{\log\tau}{\omega^*\tau}} \in (0, \zeta_0^*)$, we can plug $\zeta = \sqrt{\frac{\log\tau}{\omega^*\tau}}$ into (E.10), which yields

$$
\mathbb{P}\left[\|\tilde{\theta}^{\tau} - \theta\|_1 > \sqrt{\frac{\log\tau}{\omega^*\tau}}\right] \leqslant \phi^*\tau^{-1}. \tag{E.12}
$$

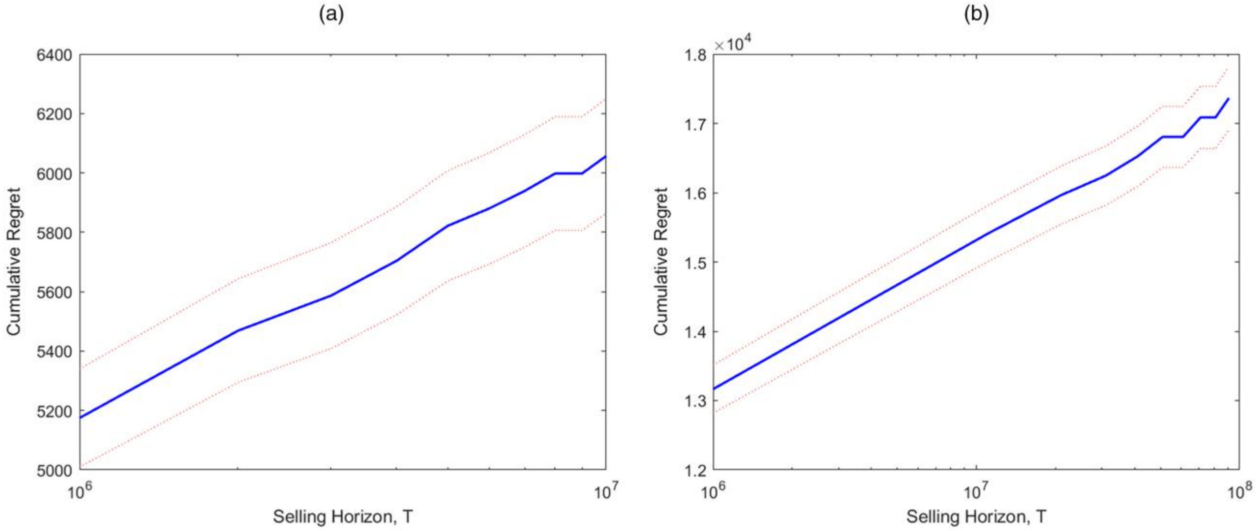
Then the regret of policy P_3 can be divided into three parts: (*a*) for $t\leqslant\tau$, the regret associated with every single customer is bounded by $r_{\max}$; (*b*) for $t>\tau$ such that $\|\tilde{\theta}^{\tau} - \theta\|_1 > \sqrt{\frac{\log\tau}{\omega^*\tau}}$, the regret associated with every single customer is also bounded by $r_{\max}$; and (*c*) for $t>\tau$ such that $\|\tilde{\theta}^{\tau} - \theta\|_1 \leqslant \sqrt{\frac{\log\tau}{\omega^*\tau}}$, the regret associated with every single customer is bounded by $2C_L\sqrt{\frac{\log\tau}{\omega^*\tau}}$ due to Lemma 4. Therefore, the regret of policy P_3 is bounded by

$$
\begin{aligned}
\text{Reg}_{P_3}(T,\theta) &\leqslant r_{\max}\tau + r_{\max}(T-\tau)\cdot\mathbb{P}\left[\|\tilde{\theta}^{\tau} - \theta\|_1 > \sqrt{\frac{\log\tau}{\omega^*\tau}}\right] \\
&\quad + 2C_L\sqrt{\frac{\log\tau}{\omega^*\tau}}(T-\tau)\cdot\mathbb{P}\left[\|\tilde{\theta}^{\tau} - \theta\|_1 \leqslant \sqrt{\frac{\log\tau}{\omega^*\tau}}\right] \\
&\leqslant r_{\max}\tau + r_{\max}(T-\tau)\cdot\phi^*\tau^{-1} \\
&\quad + 2C_L\sqrt{\frac{\log\tau}{\omega^*\tau}}\cdot(T-\tau) \qquad \text{(due to (E.7))} \\
&\leqslant r_{\max}\tau + r_{\max}\phi^*T\tau^{-1} + 2C_LT\sqrt{\frac{\log\tau}{\omega^*\tau}} \\
&\leqslant r_{\max}\lceil\nu T^{\frac{2}{3}}\log T\rceil + r_{\max}\phi^*T\lceil\nu T^{\frac{2}{3}}\log T\rceil^{-1} \\
&\quad + 2C_LT\sqrt{\frac{\log\tau}{\omega^*\tau}} \qquad \text{(due to } \tau = \lceil\nu T^{\frac{2}{3}}\log T\rceil) \\
&\leqslant 2r_{\max}\nu T^{\frac{2}{3}}\log T + r_{\max}\phi^*T(\nu T^{\frac{2}{3}}\log T)^{-1} \\
&\quad + 2C_LT\cdot\left(\frac{3}{\nu}+2\right)^{\frac{1}{2}}\cdot\omega^{*-\frac{1}{2}}T^{-\frac{1}{3}} \qquad \text{(due to (E.6))} \\
&\leqslant 2r_{\max}\nu T^{\frac{2}{3}}\log T + r_{\max}\nu^{-1}\phi^*\frac{T^{\frac{1}{3}}}{\log T} \\
&\quad + 2\left(\frac{3}{\nu}+2\right)^{\frac{1}{2}}\omega^{*-\frac{1}{2}}C_LT^{\frac{2}{3}} \\
&\leqslant 2r_{\max}\nu T^{\frac{2}{3}}\log T + 2r_{\max}\nu^{-1}\phi^*T^{\frac{2}{3}}\log T \\
&\quad + 2\left(\frac{3}{\nu}+2\right)^{\frac{1}{2}}\omega^{*-\frac{1}{2}}C_LT^{\frac{2}{3}} \qquad \text{(due to } T \geqslant 2) \\
&\leqslant \left(2r_{\max}\nu + 2r_{\max}\nu^{-1}\phi^* + 3\left(\frac{3}{\nu}+2\right)^{\frac{1}{2}}\omega^{*-\frac{1}{2}}C_L\right)T^{\frac{2}{3}}\log T \\
&\leqslant \kappa_3T^{\frac{2}{3}}\log T. \quad \square
\end{aligned}
$$

Appendix F. Additional Plots

We change the x axes (selling horizon) in Figure 3 to log-scale. The linear trend in Figure F.1, (a) and (b), indicates that the cumulative regret is roughly $O(\log T)$, which is consistent with Theorem 2.

Figure F.1. (Color online) Performance of Algorithm 1 in Constrained Revenue Maximization Problems over Increasing Selling Horizons



Notes. (a) Ten products. (b) Twenty products. The selling horizons are plotted in log-scale. The solid lines represent the mean regret, and the dotted lines represent the estimated 95% confidence intervals for the simulation results.

References

Agrawal S, Avadhanula V, Goyal V, Zeevi A (2017) Thompson sampling for the MNL-bandit. Kale S, Shamir O, eds. *Proc. 30th Conf. Learning Theory* (PMLR, New York), 76–78.

Agrawal S, Avadhanula V, Goyal V, Zeevi A (2019) MNL-bandit: A dynamic learning approach to assortment selection. *Oper. Res.* 67(5):1453–1485.

Auer P, Cesa-Bianchi N, Freund Y, Schapire RE (2002) The nonstochastic multiarmed bandit problem. *SIAM J. Comput.* 32(1): 48–77.

Balakrishnan S, Wainwright MJ, Yu B (2017) Statistical guarantees for the EM algorithm: From population to sample-based analysis. *Ann. Statist.* 45(1):77–120.

Berbeglia G (2016) Discrete choice models based on random walks. *Oper. Res. Lett.* 44(2):234–237.

Berbeglia G, Garassino A, Vulcano G (2022) A comparative empirical study of discrete choice models in retail operations. *Management Sci.* 68(6):4005–4023.

Bernstein F, Modaresi S, Sauré D (2019) A dynamic clustering approach to data-driven assortment personalization. *Management Sci.* 65(5):2095–2115.

Blanchet J, Gallego G, Goyal V (2016) A Markov chain approximation to choice modeling. *Oper. Res.* 64(4):886–905.

Chen X, Wang Y (2018) A note on a tight lower bound for capacitated MNL-bandit assortment selection models. *Oper. Res. Lett.* 46(5):534–537.

Chen X, Wang Y, Zhou Y (2020) Dynamic assortment optimization with changing contextual information. *J. Machine Learn. Res.* 21:216–1.

Chen X, Wang Y, Zhou Y (2021b) Optimal policy for dynamic assortment planning under multinomial logit models. *Math. Oper. Res.* 46(4):1639–1657.

Chen X, Shi C, Wang Y, Zhou Y (2021a) Dynamic assortment planning under nested logit models. *Production Oper. Management* 30(1):85–102.

Davis J, Gallego G, Topaloglu H (2013) Assortment planning under the multinomial logit model with totally unimodular constraint structures. Technical report, Cornell University, Ithaca, NY.

Désir A, Goyal V, Segev D, Ye C (2020) Constrained assortment optimization under the Markov chain-based choice model. *Management Sci.* 66(2):698–721.

Dong J, Şimşek AS, Topaloglu H (2019) Pricing problems under the Markov chain choice model. *Production Oper. Management* 28(1):157–175.

El Housni O, Goyal V, Humair S, Mouchtaki O, Sadighian A, Wu J (2021) Joint assortment and inventory planning for heavy tailed demand. Technical report, Cornell Tech, New York.

Feldman JB, Topaloglu H (2017) Revenue management under the Markov chain choice model. *Oper. Res.* 65(5):1322–1342.

Gallego G, Kim S (2020) Joint pricing and inventory decisions for substitutable products. Technical report, Hong Kong University of Science and Technology, Hong Kong.

Gallego G, Lu W (2021) An optimal greedy heuristic with minimal learning regret for the Markov chain choice model. Technical report, Hong Kong University of Science and Technology, Hong Kong.

Gallego G, Topaloglu H (2019) *Revenue Management and Pricing Analytics* (Springer, New York).

Gallego G, Ratliff R, Shebalov S (2015) A general attraction model and sales-based linear program for network revenue management under customer choice. *Oper. Res.* 63(1):212–232.

Gupta A, Hsu D (2020) Parameter identification in Markov chain choice models. *Theoretical Comput. Sci.* 808:99–107.

Kallus N, Udell M (2020) Dynamic assortment personalization in high dimensions. *Oper. Res.* 68(4):1020–1037.

Kosorok MR (2006) *Introduction to Empirical Processes and Semiparametric Inference* (Springer, New York).

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

Nip K, Wang Z, Wang Z (2021) Assortment optimization under a single transition choice model. *Production Oper. Management* 30(7):2122–2142.

Oh M, Iyengar G (2019) Thompson sampling for multinomial logit contextual bandits. Wallach HM, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox EB, Garnett R, eds. *Proc. Adv. Neural Inform. Processing Systems: Annual Conf. Neural Inform. Processing Systems* (Curran Associates, Inc., Red Hook, NY), 3145–3155.

Perchet V, Rigollet P, Chassang S, Snowberg E (2016) Batched bandit problems. *Ann. Statist.* 44(2):660–681.

Ragain S, Ugander J (2016) Pairwise choice Markov chains. Lee DD, Sugiyama M, von Luxburg U, Guyon I, Garnett R, eds. *Proc. Adv. Neural Inform. Processing Systems: Annual Conf. Neural Inform. Processing Systems* (Curran Associates, Inc., Red Hook, NY), 3198–3206.

Rusmevichientong P, Shen ZJM, Shmoys DB (2010) Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Oper. Res.* 58(6):1666–1680.

Sauré D, Zeevi A (2013) Optimal dynamic assortment planning with demand learning. *Manufacturing Service Oper. Management* 15(3):387–404.

Şimşek AS, Topaloglu H (2018) An expectation-maximization algorithm to estimate the parameters of the Markov chain choice model. *Oper. Res.* 66(3):748–760.

Udwani R (2021) Submodular order functions and assortment optimization. Technical report, University of California, Berkeley, Berkeley, CA.

Wang R (2013) Assortment management under the generalized attraction model with a capacity constraint. *J. Revenue Pricing Management* 12(3):254–270.

Wang Y, Chen X, Zhou Y (2018) Near-optimal policies for dynamic multinomial logit assortment selection models. Bengio S, Wallach HM, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, eds. *Proc. Adv. Neural Inform. Processing Systems: Annual Conf. Neural Inform. Processing Systems* (Curran Associates, Inc., Red Hook, NY), 3105–3114.

Zhang D, Cooper WL (2005) Revenue management for parallel flights with customer-choice behavior. *Oper. Res.* 53(3): 415–431.

Zhong Y, Birge JR, Ward A (2022) Learning the scheduling policy in time-varying multiclass many server queues with abandonment. Technical report, University of Chicago, Chicago.

Shukai Li is a final-year PhD student in the Department of Industrial Engineering and Management Sciences, Northwestern University. His research interests include stochastic model analysis, reinforcement learning, and optimization, with applications to supply chain management and healthcare operations.

Qi Luo is an assistant professor in the Department of Business Analytics, Tippie College of Business, University of Iowa. His research interests include the design of reinforcement learning and large-scale optimization algorithms with applications to transportation systems and supply chain management.

Zhiyuan Huang is an assistant professor in the Department of Management Science and Engineering, School of Economics and Management, Tongji University, Shanghai, China. His research interests include efficient Monte Carlo simulation and data-driven optimization.

Cong Shi is an associate professor in the Department of Management Science, Miami Herbert Business School, University of Miami. His research interests lie in the design and analysis of online learning algorithms with applications to revenue management and supply chain management.