Online Appendix: Assortment Optimization in the Presence of Contextual Effect

Appendix A: Exact Representation of the Fully Extended Attraction and Random Regret Models

This appendix proves the exact-recovery claims of Section 2.1. Both rely on a simple log-score principle: whenever a competing model's normalized score has a logarithm consisting of an own term plus directed pairwise terms, CMNL recovers it exactly by setting its deterministic utility equal to that log-score. Recall that the CMNL utility is $u_i(\mathcal{S}) = \mu_i + \sum_{j\in\mathcal{S}\setminus\{i\}} \alpha_{ji}$ and the choice probability is $P_i(\mathcal{S}) = \exp(u_i(\mathcal{S}))/\sum_{h\in\mathcal{S}} \exp(u_h(\mathcal{S}))$, with $\alpha_{ii} = 0$.

LEMMA EC.1 **(Log-score embedding)**. *Suppose a choice model assigns positive scores $W_i(\mathcal{S}) > 0$ and probabilities $P_i(\mathcal{S}) = W_i(\mathcal{S})/\sum_{h\in\mathcal{S}} W_h(\mathcal{S})$. If the log-score decomposes as $\log W_i(\mathcal{S}) = b_i + \sum_{j\in\mathcal{S}\setminus\{i\}} c_{ji}$, then CMNL recovers the model exactly by setting $\mu_i = b_i$ and $\alpha_{ji} = c_{ji}$ for $j \neq i$.*

Proof. Under this parameterization, $u_i(\mathcal{S}) = b_i + \sum_{j\in\mathcal{S}\setminus\{i\}} c_{ji} = \log W_i(\mathcal{S})$, so $\exp(u_i(\mathcal{S})) = W_i(\mathcal{S})$ and the CMNL probabilities coincide with $P_i(\mathcal{S})$. □

Fully extended attraction model. The fully extended attraction model of Cooper and Nakanishi (1988) (also called the attraction model with differential cross-competitive effects) specifies the attraction of item i as $A_i(\mathcal{S}) = \exp(\bar{a}_i)\prod_{k=1}^{K}\prod_{j\in\mathcal{S}} f_k(X_{kj})^{\beta_{kij}}$ with shares $s_i(\mathcal{S}) = A_i(\mathcal{S})/\sum_{h\in\mathcal{S}} A_h(\mathcal{S})$, where X_{kj} is the value of marketing variable k for item j, $f_k(\cdot) > 0$ is a fixed transformation, and β_{kij} is the cross-competitive effect of j's variable k on i's attraction. Taking logarithms,

$$\log A_i(\mathcal{S}) = \underbrace{\bar{a}_i + \sum_{k=1}^{K} \beta_{kii} \log f_k(X_{ki})}_{\text{own term for } i} + \sum_{j\in\mathcal{S}\setminus\{i\}} \underbrace{\sum_{k=1}^{K} \beta_{kij} \log f_k(X_{kj})}_{\text{directed contribution from } j},$$

which has the decomposition form of Lemma EC.1. Hence CMNL recovers the model exactly by setting $\mu_i = \bar{a}_i + \sum_k \beta_{kii} \log f_k(X_{ki})$ and $\alpha_{ji} = \sum_k \beta_{kij} \log f_k(X_{kj})$ for $j \neq i$, since then $u_i(\mathcal{S}) = \log A_i(\mathcal{S})$. The MCI variant ($f_k(X) = X$) gives $\alpha_{ji} = \sum_k \beta_{kij} \log X_{kj}$, and the MNL-attraction variant ($f_k(X) = e^X$) gives $\alpha_{ji} = \sum_k \beta_{kij} X_{kj}$.

Smooth random regret minimization. The smooth random regret minimization model of Chorus (2010) defines the systematic regret of item i as $\widetilde{R}_i(\mathcal{S}) = \sum_{j\in\mathcal{S}\setminus\{i\}} \sum_{m=1}^{M} \ln\big(1 + \exp[\beta_m(x_{jm} - x_{im})]\big)$, where x_{im} is item i's value on attribute m and β_m the attribute weight, and generates the probability $P_i^{\mathrm{RRM}}(\mathcal{S}) = \exp(-\widetilde{R}_i(\mathcal{S}))/\sum_{h\in\mathcal{S}} \exp(-\widetilde{R}_h(\mathcal{S}))$. Writing $r_{ji} = \sum_m \ln\big(1 + \exp[\beta_m(x_{jm} - x_{im})]\big)$ for the pairwise regret that comparison with j imposes on i, the log-score is $\log W_i^{\mathrm{RRM}}(\mathcal{S}) = -\widetilde{R}_i(\mathcal{S}) = \sum_{j\in\mathcal{S}\setminus\{i\}}(-r_{ji})$, again of the form in Lemma EC.1 with a zero own term. Hence CMNL recovers this model exactly by setting $\mu_i = 0$ and $\alpha_{ji} = -r_{ji}$ for $j \neq i$; any additive alternative-specific regret constant c_i is absorbed by instead setting $\mu_i = -c_i$. In both cases the model is a structured sub-family of CMNL, obtained by constraining the context-effect matrix to the indicated feature-based form.

Appendix B: Details on Exact MILP and other heuristics

Theorem 1 has shown that (3) does not allow any constant ratio approximation, which demonstrates the fundamental difficulty of solving it. We present the details of implementable methods for solving (3) in practice, including a mixed-integer programming formulation and two heuristics.

Mixed Integer Programming Formulation Due to the celebrated advances in state-of-the-art solvers for solving mixed-integer programming problems, reformulating (3) into a mixed-integer linear programming problem can enable us to solve it exactly for moderate sizes. We first introduce an additional continuous vector $\boldsymbol{y} \in \mathbb{R}_+^n$, where y_i represents the choice probability of the i-th product and is defined as

$$\begin{aligned} y_i &= \frac{x_i \exp\left(\mu_i + \sum_{j=1}^n x_j \alpha_{ji}\right)}{1 + \sum_{k=1}^n x_k \exp\left(\mu_k + \sum_{j=1}^n x_j \alpha_{jk}\right)} \\ \iff \quad & y_i + y_i \sum_{k=1}^n x_k \exp\left(\mu_k + \sum_{j=1}^n x_j \alpha_{jk}\right) = x_i \exp\left(\mu_i + \sum_{j=1}^n x_j \alpha_{ji}\right) \\ \iff \quad & y_i + y_i \sum_{k=1}^n x_k \left(\theta_k \prod_{j=1}^n (1 + x_j \gamma_{jk})\right) = x_i \left(\theta_i \prod_{j=1}^n (1 + x_j \gamma_{ji})\right), \end{aligned} \tag{EC.1}$$

where the last equivalence follows because we define $\theta_i := \exp(\mu_i)$ and $\gamma_{ji} := \exp(\alpha_{ji}) - 1$. Now we use the big-M method to linearize (EC.1). More precisely, we will use the big-M method to linearize each $x_i \prod_{j=1}^n (1 + x_j \gamma_{ji})$. Define $z_{i0} = x_i$ and $z_{i\ell} = x_i \prod_{j=1}^{\ell} (1 + x_j \gamma_{ji})$ for any $\ell \in [n]$, where we can express them recursively by

$$z_{i\ell} = z_{i(\ell-1)}(1 + x_\ell \gamma_{\ell i}) = z_{i(\ell-1)} + \gamma_{\ell i} z_{i(\ell-1)} x_\ell.$$

Define $w_{i\ell} = z_{i(\ell-1)} x_\ell$ for each $\ell \in [n]$, where we can linearize each of them using McCormick representation:

$$w_{i\ell} \le z_{i(\ell-1)}, \quad w_{i\ell} \le M_{i\ell} x_\ell, \quad w_{i\ell} \ge 0, \quad w_{i\ell} \ge z_{i(\ell-1)} - M_{i\ell}(1 - x_\ell),$$

where $M_{i\ell} = \prod_{j=1}^{\ell-1} (1 + \gamma_{ji} \mathbb{I}(\gamma_{ji} > 0))$. Besides, we define $\xi_{ik0} = y_i x_k$, and $\xi_{ik\ell} = y_i x_k \prod_{j=1}^{\ell} (1 + x_j \gamma_{jk})$ for any $\ell \in [n]$, where again we can express them recursively by

$$\xi_{ik\ell} = \xi_{ik(\ell-1)} (1 + x_\ell \gamma_{\ell k}) = \xi_{ik(\ell-1)} + \gamma_{jk} \xi_{ik(\ell-1)} x_\ell.$$

Define $\eta_{ik\ell} = \xi_{ik(\ell-1)} x_\ell$ for each $\ell \in [n]$. Again, we can linearize each of them (together with ξ_{ik0}) using McCormick representation:

$$\xi_{ik0} \le y_i, \quad \xi_{ik0} \le x_k, \quad \xi_{ik0} \ge 0, \quad \xi_{ik0} \ge y_i - (1 - x_k),$$

$$\eta_{ik0} \le \xi_{ik(\ell-1)}, \quad \eta_{ik\ell} \le M_{k\ell} x_\ell, \quad \eta_{ik\ell} \ge 0, \quad \eta_{ik\ell} \ge \xi_{ik(\ell-1)} - M_{k\ell}(1 - x_\ell).$$

Combining all those transformations of variables and linearization, we can solve (3) using the following MIP formulation:

$$\begin{aligned} \max_{\substack{\boldsymbol{x} \in \{0,1\}^n,\ \boldsymbol{y} \in [0,1]^n, \\ \boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\xi}, \boldsymbol{\eta} \ge \boldsymbol{0}}} \quad & \sum_{i=1}^n r_i y_i \\ \text{s.t.} \quad & y_i + \sum_{k=1}^n \theta_k \xi_{ikn} = \theta_i z_{in}, \quad \forall i \in [n], \\ & z_{i0} = x_i, \quad z_{i\ell} = z_{i(\ell-1)} + \gamma_{\ell i} w_{i\ell}, \quad \forall i \in [n],\ \ell \in [n], \\ & w_{i\ell} \le z_{i(\ell-1)}, \quad w_{i\ell} \le M_{i\ell} x_\ell, \quad w_{i\ell} \ge 0, \quad w_{i\ell} \ge z_{i(\ell-1)} - M_{i\ell}(1 - x_\ell), \quad \forall i \in [n],\ \ell \in [n], \\ & \xi_{ik\ell} = \xi_{ik(\ell-1)} + \gamma_{\ell k} \eta_{ik\ell}, \quad \forall i \in [n],\ k \in [n],\ \ell \in [n], \\ & \xi_{ik0} \le y_i, \quad \xi_{ik0} \le x_k, \quad \xi_{ik0} \ge 0, \quad \xi_{ik0} \ge y_i - (1 - x_k), \quad \forall i \in [n],\ k \in [n], \\ & \eta_{ik\ell} \le \xi_{ik(\ell-1)}, \quad \eta_{ik\ell} \le M_{k\ell} x_\ell, \quad \eta_{ik\ell} \ge 0, \quad \eta_{ik\ell} \ge \xi_{ik(\ell-1)} - M_{k\ell}(1 - x_\ell), \quad \forall i \in [n],\ k \in [n],\ \ell \in [n]. \end{aligned} \tag{EC.2}$$

The first equality constraint is identical to (EC.1), and the remaining constraints follow from our transformation of variables and linearization. We note that other practical constraints, such as the cardinality constraint and the knapsack constraint on the assortment size, can also be added to (EC.2). After that, the problem remains a mixed-integer programming with only one additional linear constraint, which does not significantly complicate the solving procedure.

Revenue-Ordered Assortment. Some intuitive and simple heuristics can also perform well in practice. For notational convenience, we assume the products are indexed such that $r_1 \geq r_2 \geq \cdots \geq r_n$. One widely used approach is the *Revenue-ordered* heuristic, which finds the optimal assortment in the revenue-ordered assortments, i.e., among assortments $[1], [2], \ldots, [n]$. Such an algorithm can be easily executed by evaluating the objective of all revenue-ordered assortments and choose the best among them. This heuristic is known to yield the optimal assortment under the MNL model (Talluri and van Ryzin 2004) and certain variants of the Mixed MNL model (Rusmevichientong et al. 2014). It also demonstrates strong empirical performance under several recently proposed choice models (Wang and Sahin 2018). However, under our CMNL model, which incorporates context effects, the optimal assortment may no longer follow a revenue-ordered structure. This can occur when adding a high-revenue product significantly reduces the choice probabilities of other products. Nevertheless, one might expect this heuristic to perform reasonably well when context effects are weak or when high-revenue products exert positive context effects on others.

Local Search. The local search heuristic starts from an initial assortment $\mathcal{S}_0$ (which may be empty, the full set, or a warm start from another method) and iteratively seeks an improvement by exploring *neighborhoods* around the current assortment. For iteration ℓ with incumbent $\mathcal{S}$, the k-move neighborhood $\mathcal{N}_k(\mathcal{S})$ (Line 5) contains all subsets that differ from $\mathcal{S}$ by at most k inclusions/exclusions, i.e., those with symmetric difference size $|\mathcal{S}' \triangle \mathcal{S}| \leq k$. It is optional that the algorithm uses a layered strategy: it first attempts $k = K_{\min}$ moves; only if no improvement is found does it escalate to $k = K_{\min} + 1$, and so on, up to a user-chosen $K_{\max}$. At the first k that yields an improvement, it accepts the best neighbor and restarts the search from $k = K_{\min}$. The procedure terminates when no neighbor with $k \leq K_{\max}$ improves the objective $R(\cdot)$ of (3), returning the incumbent as a locally optimal solution. Details of the algorithm are shown in Algorithm EC.1.

The worst-case number of candidate sets per iteration is $\sum_{i=1}^{K} \binom{n}{i}$, which is $O(n^K)$. However, for most of the time, small K's might already yield good performances, for example, $K = 1, 2, 3$. In such cases, the heuristic enjoys a low computational burden.

Appendix C: General-Instance Performance Test

This appendix reports the performance test on general, unstructured CMNL instances, complementing the structured experiments of Section 4.3. Here every pairwise context effect α_{ji} is drawn at random, so none of the specialized algorithms of Section 3 applies and exact benchmarks are available only at moderate n. The no-purchase option has fixed utility $u_0 = 0$; for each product i we draw the baseline utility $\mu_i \sim N(0, 1)$ i.i.d. and the revenue $r_i \sim U[1, 10]$ i.i.d.; and for each ordered pair (j, i) we draw $\alpha_{ji} \sim N(0, 0.5)$ i.i.d. We test $n \in \{5, 10, 20, 30, 40, 50\}$ with 100 instances per size.

We use the optimal solution given by the MILP for instances where $n \leq 30$. The true optimal For $n \geq 40$ enumeration is out of reach, and we use the best assortment found by LS$(1, 3)$ as the benchmark. We also report the runtime of the MILP (EC.2), with a 10-minute time limit. The heuristics under evaluation are the best revenue-ordered assortment

Algorithm EC.1 Local Search with (Layered) K-Move Neighborhoods

Input: Universe of products $\mathcal{N}$ with $|\mathcal{N}| = n$; objective $R(\cdot)$ of (3); initial assortment $\mathcal{S}_0 \subseteq \mathcal{N}$; minimum move size, maximum move size $K_{\min},\ K_{\max} \in \mathbb{Z}_+$.

Output: A locally optimal assortment $\widehat{\mathcal{S}}$.

$\mathcal{S} \leftarrow \mathcal{S}_0,\ \ell \leftarrow 0$
while true **do**
  improved $\leftarrow$ **false**
  for $k = K_{\min}$ **to** $K_{\max}$ **do** ▷ Set $K_{\min} = K_{\max}$ for no layered search
    $\mathcal{N}_k(\mathcal{S}) \leftarrow \{\ \mathcal{S}' \subseteq \mathcal{N}\ :\ 1 \le |\mathcal{S}' \triangle \mathcal{S}| \le k\ \}$
    $\mathcal{S}^\star \leftarrow \arg\max_{\mathcal{S}' \in \mathcal{N}_k(\mathcal{S})}\ F(\mathcal{S}')$ ▷ Break ties arbitrarily
    if $F(\mathcal{S}^\star) > F(\mathcal{S})$ **then**
      $\mathcal{S} \leftarrow \mathcal{S}^\star;\ \ell \leftarrow \ell + 1;$ *improved* $\leftarrow$ **true**
      break ▷ Accept best neighbor and restart from $k = K_{\min}$
    end if
  end for
  if not *improved* **then**
    return $\widehat{\mathcal{S}} \leftarrow \mathcal{S}$ ▷ No improving neighbor of size $\le K_{\max}$
  end if
end while

and local search (Algorithm EC.1) with move sizes $(K_{\min}, K_{\max}) = (1,1), (1,2), (1,3)$. Table EC.1 reports, for each method, the average runtime per instance and, for the heuristics, the average and maximum optimality gap (relative to the true optimum for $n \le 30$ and to LS$(1,3)$ otherwise) together with the percentage of instances in which the heuristic attains this benchmark.

The key observations from Table EC.1 are as follows. Against the true optimum, the revenue-ordered heuristic is extremely fast but has large optimality gaps (average 3.7%–9.3%, worst cases up to 37%) and essentially never finds the optimum beyond $n = 10$, whereas local search performs much better but degrades as n grows: LS$(1,3)$ recovers the optimum in 98% of instances at $n = 5$, 71% at $n = 20$, and 58% at $n = 30$. The seemingly better performance of LS$(1,1)$ and LS$(1,2)$ at $n \in \{40, 50\}$ is because the best known assortment there is LS$(1,3)$. This shows the value of the certified optima available at scale in the structured experiments of Section 4.3. We note that the lack of structure in those instances make them harder, thus the heuristics perform worse than on the structured instances. However, the qualitative conclusions are consistent with those structured experiments.

Appendix D: Empirical Evaluation of the CMNL Model on Other Product Categories

In this section, we report results for several other categories from the PDI Technologies dataset: frozen food, milk, and Sprite. In all these categories, the CMNL model is *not identifiable*, because each category has some products which are

		n					
Group	Metric	5	10	20	30	40	50
MILP	Avg. time (s)	0.051	2.733	57.73	603.28	—	—
Rev-ordered	Avg. time (s)	0.000	0.000	0.000	0.000	0.000	0.001
	Avg. gap	3.68%	7.33%	9.25%	9.15%	7.28%	6.32%
	Max gap	21.3%	36.9%	25.4%	26.5%	26.2%	19.6%
	% optimal	49%	9%	0%	0%	0%	0%
LS$(1,1)$	Avg. time (s)	0.000	0.001	0.001	0.003	0.005	0.009
	Avg. gap	0.24%	0.55%	1.71%	1.88%	0.22%	0.11%
	Max gap	15.7%	22.1%	17.0%	17.1%	7.5%	4.1%
	% optimal	93%	81%	49%	35%	42%	46%
LS$(1,2)$	Avg. time (s)	0.000	0.002	0.003	0.008	0.017	0.028
	Avg. gap	0.05%	0.41%	1.45%	1.80%	0.01%	0.04%
	Max gap	4.2%	22.1%	14.4%	17.1%	0.3%	3.0%
	% optimal	98%	89%	64%	50%	76%	82%
LS$(1,3)$	Avg. time (s)	0.000	0.004	0.012	0.047	0.129	0.263
	Avg. gap	0.05%	0.07%	1.12%	1.57%	—	—
	Max gap	4.2%	2.8%	14.4%	15.3%	—	—
	% optimal	98%	94%	71%	58%	—	—

Table EC.1 Performance of the MILP, local search, and revenue-ordered assortments on general synthetic CMNL instances (100 instances per n). Optimality gaps are computed relative to the true optimum obtained by MILP for $n \leq 30$ and relative to LS$(1,3)$ for $n \in \{40, 50\}$; "% optimal" is the share of instances in which the method attains this benchmark. A dash (—) for MILP indicates that it exceeds the 10-minute time limit; for LS$(1,3)$, that it is itself the benchmark.

almost always offered in all stores. However, with L_2-regularization, a stable estimator is still available. Data filtration and train-test split methods similar to Section 4.2 are applied. In these categories, although the CMNL model is not identifiable, it still achieves good out-of-sample performance compared to benchmarks with L_2-regularization. The results are qualitatively similar to those reported in the main draft. The CMNL-3 model outperform the CMNL model in some metrics. However, we can see the improvement is small relative to the improvement of the CMNL model compared with benchmarks, although CMNL-3 is significantly more complex than then CMNL and have significantly larger number of parameters.

Appendix E: A Feature-Level CMNL and an Empirical Study on Expedia Data

Throughout the paper the context effects α_{ji} are free per-pair parameters, estimable when a fixed set of products recurs across many assortments, as in the convenience-store dataset of Section 4.2. In attribute-rich applications where products do not recur, one can instead estimate the base utilities and context effects using product attributes, as anticipated by the remark following Proposition 6. This appendix empirically investigates the featurized CMNL on the public Expedia hotel-search data, which is used in prior literature, e.g. Aouad et al. (2023).

E.1. Data, Estimation, and Metrics

We use the public *Personalize Expedia Hotel Searches* dataset, and pre-processing applied same as in Aouad et al. (2023). Each search is a choice event: the offered set is the list of displayed properties, and the choice is the booked property or the no-purchase option. For each displayed hotel we build a feature vector following the specification of

		Log-Likelihood		Cross-Entropy		MRR		NDCG	
Model	**Parameter Count**	Value	Rank	Value	Rank	Value	Rank	Value	Rank
MNL	65	-251858	7	1.7083	7	0.6864	8	0.8929	9
GAM	130	-252276	10	1.7111	10	0.6863	10	0.8929	9
MMNL-2	130	-251599	6	1.7065	6	0.6895	6	0.8941	7
MMNL-5	325	-251409	4	1.7052	4	0.6906	5	0.8950	5
Markov	4225	-252025	9	1.7094	9	0.6889	7	0.8942	6
RAsN	8580	-279236	11	1.8939	11	0.6572	11	0.8769	11
CMNL	4225	-250624	2	1.6999	2	0.6986	2	**0.8982**	**1**
CMNL-3	135265	**-250615**	**1**	**1.6998**	**1**	**0.6988**	**1**	0.8981	2
CMNL_target_i	130	-251528	5	1.7060	5	0.6924	4	0.8955	4
CMNL_source_j	130	-251858	7	1.7083	7	0.6864	8	0.8930	8
CMNL_factorized	325	-251210	3	1.7039	3	0.6965	3	0.8979	3

Table EC.2 Performance comparison of choice models in category frozen food with 65 products and 0.42 million total transactions. For each metric, the best-performing model's results are highlighted in bold. Rank of models according to each metric is also reported.

		Log-Likelihood		Cross-Entropy		MRR		NDCG	
Model	**Parameter Count**	Value	Rank	Value	Rank	Value	Rank	Value	Rank
MNL	62	-1437743	8	2.0172	6	0.7679	9	0.9377	9
GAM	124	-1437741	6	2.0172	6	0.7679	9	0.9377	9
MMNL-2	124	-1435926	5	2.0147	5	0.7720	5	0.9390	5
MMNL-5	310	-1435063	4	2.0135	4	0.7747	3	0.9400	3
Markov	3844	-1440555	10	2.0212	10	0.7715	6	0.9387	6
RAsN	7812	-1483894	11	2.0820	11	0.7243	11	0.9200	11
CMNL	3844	**-1431404**	**1**	**2.0084**	**1**	0.7753	2	0.9405	2
CMNL-3	117304	-1431632	2	2.0087	2	**0.7754**	**1**	**0.9407**	**1**
CMNL_target_i	124	-1437913	9	2.0175	9	0.7704	7	0.9385	7
CMNL_source_j	124	-1437741	6	2.0173	8	0.7680	8	0.9378	8
CMNL_factorized	310	-1432807	3	2.0104	3	0.7744	4	0.9399	4

Table EC.3 Performance comparison of choice models in category milk with 62 products and 1.40 million total transactions. For each metric, the best-performing model's results are highlighted in bold. Rank of models according to each metric is also reported.

Aouad et al. (2023) (nightly and log historical price, two location scores, star and review scores, brand and promotion indicators, a display-position one-hot, search-level controls, and a price $\times$ booking-window interaction). We follow the same data pre-processing as their paper: focusing on randomized-display searches (which removes display-position endogeneity), and the five most active country websites. The data is split into train/test of 75/25. At this scale a free CMNL is infeasible, since the five sites contain tens of thousands of distinct hotels that rarely recur, so identifiability condition required by Proposition 7 is not satisfied. So, we estimate only the featurized models. We report the same goodness-of-fit metrics as in Section 4.2: test log-likelihood, cross-entropy, MRR, and NDCG.

E.2. The Featurized CMNL

Let $q_{ti} \in \mathbb{R}^K$ be the attribute vector of product i at search event t (e.g., price, star rating, review score, brand). For notational brevity, we drop subscript t. The featurized CMNL parameterizes

$$\mu_i = \theta^\top q_i, \qquad \alpha_{ji} = \phi^\top g(q_j, q_i),$$

		Log-Likelihood		Cross-Entropy		MRR		NDCG	
Model	**Parameter Count**	Value	Rank	Value	Rank	Value	Rank	Value	Rank
MNL	29	-1318080	9	1.6697	7	0.9444	4	0.9731	8
GAM	58	-1318070	7	1.6697	7	0.9444	4	0.9731	8
MMNL-2	58	-1316772	6	1.6681	6	0.9441	8	0.9733	6
MMNL-5	145	-1315394	4	1.6663	4	0.9438	9	0.9734	5
Markov	841	-1328927	11	1.6834	11	0.9407	10	0.9716	10
RAsN	1740	-1319356	10	1.6713	10	0.9405	11	0.9716	10
CMNL	841	-1313409	2	1.6638	2	0.9447	3	0.9737	2
CMNL-3	11803	**-1312959**	**1**	**1.6635**	**1**	0.9442	7	0.9737	2
CMNL_target_i	58	-1316568	5	1.6678	5	0.9448	2	0.9736	4
CMNL_source_j	58	-1318077	8	1.6697	7	0.9444	4	0.9732	7
CMNL_factorized	145	-1314752	3	1.6655	3	**0.9450**	**1**	**0.9738**	**1**

Table EC.4 Performance comparison of choice models in category Sprite with 29 products and 1.60 million total transactions. For each metric, the best-performing model's results are highlighted in bold. Rank of models according to each metric is also reported.

where $g : \mathbb{R}^K \times \mathbb{R}^K \to \mathbb{R}^{K_{\text{eff}}}$ is a fixed attribute-pair transformation and (θ, ϕ) are the parameters. The product utility is $u_i(\mathcal{S}) = \theta^\top q_i + \phi^\top \sum_{j \in \mathcal{S}\setminus\{i\}} g(q_j, q_i)$, and the choice probabilities follow (2). Because g is fixed, (μ, A) is an affine image of (θ, ϕ), so by the remark following Proposition 6 the (regularized) log-likelihood is concave in (θ, ϕ). The featurized model has only $K + K_{\text{eff}}$ parameters, against the $n + n(n-1)$ of the free CMNL, a significant reduction when the number of attributes is far smaller than the number of products, and the device that makes context-effect estimation feasible when products do not recur, so that a free A and the identifiability conditions of Proposition 7 are unavailable.

We use a context matrix built from a few salient attributes. With $\Delta p_{ji} = p_j - p_i$ the price gap and $\Delta s_{ji} = s_j - s_i$ the star-rating gap (price and star standardized),

$$g(q_j, q_i) = \big(\Delta p_{ji},\, -|\Delta p_{ji}|,\, \Delta s_{ji},\, -|\Delta s_{ji}|,\, \mathbf{1}[j \succeq i],\, \mathbf{1}[i \succeq j],\, \mathbf{1}[\text{brand}_j = \text{brand}_i]\big),$$

so $K_{\text{eff}} = 7$, where $j \succeq i$ means j dominates i on price, star rating, and review score. The signed gaps capture reference-price / value-for-money and compromise effects (a positive coefficient on $\sum_j \Delta p_{ji}$ rewards interior, less extreme prices), the proximity terms capture similarity-based cannibalization, the dominance indicators capture attraction/decoy effects, and the brand indicator captures within-brand interference. In addition to the full set of contextual features, we also estimate a price-only variant, which keeps only $(\Delta p_{ji}, -|\Delta p_{ji}|)$.

E.3. Results

Table EC.5 reports the out-of-sample fit of the context-less featurized MNL (26 parameters) and the price-only (28 parameters) and full (33 parameters) featurized CMNL at each of the five most active sites, individually and pooled. Adding feature-based context effects improves the MNL on cross-entropy at all five sites and on the ranking metrics (MRR and NDCG) at four of the five, by margins comparable to the MNL-versus-Exponomial gaps reported by Aouad et al. (2023) on the same data, the price-only context already captures much of the gain.

The fitted context coefficients are interpretable and significant. Table EC.6 reports the model coefficients ϕ across all five sites individually and pooled together (price and star standardized, 1 s.d. of price $\approx$ \$146). We can observe

Site	Observations	Model	Parameter Count	LL	CE	MRR	NDCG
		MNL	26	−13960.6	0.7953	0.3768	0.5148
5	52,664	CMNL price	28	**−13795.0**	**0.7859**	0.3754	0.5138
		CMNL full	33	−13857.0	0.7894	**0.3821**	**0.5197**
		MNL	26	−2495.3	0.8424	0.3475	0.4910
14	8,885	CMNL price	28	**−2430.4**	**0.8205**	0.3499	0.4933
		CMNL full	33	−2441.7	0.8243	**0.3725**	**0.5115**
		MNL	26	−1628.7	0.8361	0.3845	0.5215
15	5,846	CMNL price	28	−1610.7	0.8269	0.3881	0.5242
		CMNL full	33	**−1600.8**	**0.8217**	**0.3938**	**0.5287**
		MNL	26	−1226.8	0.8397	0.3477	0.4919
24	4,382	CMNL price	28	−1219.2	0.8345	**0.3496**	**0.4932**
		CMNL full	33	**−1206.9**	**0.8261**	0.3436	0.4891
		MNL	26	−727.2	0.6106	0.4073	0.5355
32	3,574	CMNL price	28	**−720.2**	**0.6047**	0.3915	0.5244
		CMNL full	33	−721.5	0.6058	**0.4119**	**0.5390**
		MNL	26	−20002.2	0.7964	0.3721	0.5109
Pooled	75,351	CMNL price	28	**−19670.4**	**0.7832**	0.3712	0.5103
		CMNL full	33	−19676.7	0.7834	**0.3798**	**0.5177**

Table EC.5 Out-of-sample fit of the featurized models at the five most active Expedia sites. For each site and metric, the best of the three models is in bold.

the price effect: the positive coefficient on the price gap $p_j - p_i$ means a hotel's choice utility rises when its displayed competitors are more expensive (a reference-price / value-for-money context effect). The utility of a hotel is lowered when higher-rated or dominating competitors exist, and the dominance pair interaction is consistent with the attraction/decoy effect ($j \succeq i$ negative, $i \succeq j$ positive). All seven coefficients are significant at the 1% level in the pooled estimate.

Context term	Site 5	Site 14	Site 15	Site 24	Site 32	**Pooled**
price gap $p_j - p_i$	$+0.381^{**}$	$+0.589^{**}$	$+0.169$	$+0.305^{*}$	$+0.347^{*}$	$+0.374^{**}$
	(+9.2)	(+4.9)	(+1.6)	(+2.3)	(+2.1)	(+10.9)
price proximity $-\lvert p_j - p_i\rvert$	$+0.307^{**}$	$+0.459^{**}$	$+0.139$	$+0.337^{*}$	$+0.280$	$+0.307^{**}$
	(+7.5)	(+3.9)	(+1.3)	(+2.5)	(+1.8)	(+9.0)
star gap $s_j - s_i$	-0.319^{**}	-0.331^{**}	-0.314^{**}	-0.386^{**}	-0.467^{**}	-0.329^{**}
	(−12.1)	(−4.8)	(−4.0)	(−3.6)	(−3.3)	(−14.9)
star proximity $-\lvert s_j - s_i\rvert$	$+0.061^{**}$	$+0.123^{*}$	$+0.096$	$+0.117$	$+0.353^{**}$	$+0.078^{**}$
	(+3.3)	(+2.6)	(+1.8)	(+1.6)	(+3.6)	(+5.0)
j dominates i	-0.374^{**}	-0.450^{**}	-0.285^{**}	-0.296^{**}	-0.454^{**}	-0.374^{**}
	(−16.3)	(−7.2)	(−4.2)	(−3.6)	(−3.9)	(−19.1)
i dominates j	$+0.058^{**}$	$+0.044$	-0.036	-0.049	$+0.082$	$+0.049^{**}$
	(+4.3)	(+1.3)	(−0.8)	(−0.9)	(+1.3)	(+4.2)
same brand	-0.184^{**}	-0.142^{**}	-0.166^{**}	-0.215^{**}	-0.150	-0.179^{**}
	(−12.4)	(−3.6)	(−3.2)	(−3.6)	(−2.0)	(−14.5)

Table EC.6 Featurized-CMNL context coefficients ϕ per site and pooled (z-scores in parentheses). *: $|z| > 1.96$; **: $|z| > 2.58$.

Because price enters both the base utility (through θ) and the context effect (through ϕ), the featurized CMNL shows two aspects relevant to pricing and product design: a hotel's own features, and the context effect from the features of other offered hotels. Joint assortment, pricing, and feature optimization under the featurized CMNL is a natural direction for future work.

Appendix F: Technical Proof

Proof of Thorem 1 Let us conduct an approximation ratio preserved reduction from the maximum independent set (MIS) problem, which is well known to not allow any polynomial-time approximation algorithm to achieve a factor within $O\left(1/n^{1-\delta}\right)$ for any constant $\delta > 0$ unless NP $\subseteq$ BPP (Williamson and Shmoys 2011), using the antagonistic model.

MIS Problem: Given an undirected graph $G = (V, E)$, MIS calls for a vertex set $S \subseteq V$ with the maximum cardinality such that no pair of nodes in S is connected by some edge in E.

Instance Construction: We need to construct an instance of (3) to solve MIS problem. Let

$$[n] = V, \quad r_i = 1, \quad \mu_i = -10\log(n), \quad \alpha_{ji} = \begin{cases} -10\log(n) & e_{ji} \in E, \\ 0 & e_{ji} \notin E, \end{cases} \quad \forall i, j \in [n].$$

Note that the constructed r_i, μ_i, α_{ji}'s are all of polynomial size. In the following, we will use V and $[n]$ (i.e., the vertex set and the product set) interchangeably. To prove the correctness of such a reduction, we need to make sure that solving (3) with the above constructed instance equivalently solves the given MIS problem. To prove this, we should ensure two criteria:

(**C1**) For any set $\mathcal{S} \subseteq V$ that is not an independent set, let $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ be the maximum independent set that is included in $\mathcal{S}$. Then, our constructed instance of (3) prefers $\tilde{\mathcal{S}}$ over $\mathcal{S}$ since it generates higher revenue.

(**C2**) For two independent sets $\mathcal{S}_1, \mathcal{S}_2 \subseteq V$, as long as $|\mathcal{S}_1| > |\mathcal{S}_2|$, then our constructed instance of (3) prefers $\mathcal{S}_1$ over $\mathcal{S}_2$ since it generates higher revenue.

The first criterion guarantees that our constructed instance of (3) will always choose the assortment that corresponds to an independent set in G. The second criterion guarantees that our constructed instance of (3) will always select the assortment corresponding to the independent set with the maximum size.

We first check (**C1**). For any set $\mathcal{S} \subseteq V$ that is not an independent set, let $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ be the maximum independent set that is included in $\mathcal{S}$. On one hand, the revenue generated by assortment $\tilde{\mathcal{S}}$ is $\frac{|\tilde{\mathcal{S}}|/n^{10}}{1+|\tilde{\mathcal{S}}|/n^{10}}$. On the other hand, the revenue generated by $\mathcal{S}$ is $\frac{\sum_{i\in\mathcal{S}} \exp\left(\sum_{j\in\mathcal{S}} \alpha_{ji}\right)/n^{10}}{1+\sum_{k\in\mathcal{S}} \exp\left(\sum_{j\in\mathcal{S}} \alpha_{jk}\right)/n^{10}}$. We want to show

$$\frac{|\tilde{\mathcal{S}}|/n^{10}}{1+|\tilde{\mathcal{S}}|/n^{10}} > \frac{\sum_{i\in\mathcal{S}} \exp\left(\sum_{j\in\mathcal{S}} \alpha_{ji}\right)/n^{10}}{1+\sum_{k\in\mathcal{S}} \exp\left(\sum_{j\in\mathcal{S}} \alpha_{jk}\right)/n^{10}} \quad \Longleftrightarrow \quad |\tilde{\mathcal{S}}| > \sum_{i\in\mathcal{S}} \exp\left(\sum_{j\in\mathcal{S}} \alpha_{ji}\right), \tag{EC.3}$$

where the equivalency holds because $\frac{y'}{1+y'} > \frac{y}{1+y}$ when $y' > y \geq 0$. To prove the inequality, we first re-express $\sum_{i\in\mathcal{S}} \exp\left(\sum_{j\in\mathcal{S}} \alpha_{ji}\right)$ as follows

$$\sum_{i\in\mathcal{S}} \exp\left(\sum_{j\in\mathcal{S}} \alpha_{ji}\right) = \underbrace{\sum_{i\in\tilde{\mathcal{S}}} \exp\left(\sum_{j\in\mathcal{S}\setminus\tilde{\mathcal{S}}} \alpha_{ji}\right)}_{\text{(I)}} + \underbrace{\sum_{i\in\mathcal{S}\setminus\tilde{\mathcal{S}}} \exp\left(\sum_{j\in\mathcal{S}} \alpha_{ji}\right)}_{\text{(II)}},$$

where we decompose the summation as two parts, one for $\tilde{\mathcal{S}}$ and one for $\mathcal{S}\backslash\tilde{\mathcal{S}}$, and use the fact that $\tilde{\mathcal{S}}$ is an independent set such that no pair of product in $\tilde{\mathcal{S}}$ can have context effects by our constructed instance to simplify (I). Our next step is to upper bound (I) and (II) separately. Concerning (I), we note that there exists at least one $i \in \tilde{\mathcal{S}}$ that is connected to some $j \in \mathcal{S}\backslash\tilde{\mathcal{S}}$ such that $(i,j) \in E$, otherwise we can augment $\tilde{\mathcal{S}}$ with any $j \in \mathcal{S}\backslash\tilde{\mathcal{S}}$ to get a larger independent set included in $\mathcal{S}$, contradicting the maximality of $\tilde{\mathcal{S}}$. Hence,

$$\text{(I)} \leq \left(|\tilde{\mathcal{S}}| - 1\right) + \exp\left(-10\log(n)\right) = \left(|\tilde{\mathcal{S}}| - 1\right) + 1/n^{10}. \tag{EC.4}$$

Besides, for any $i \in \mathcal{S}\backslash\tilde{\mathcal{S}}$, there exists some $j \in \tilde{\mathcal{S}}$ such that $(i,j) \in E$, otherwise we can augment $\tilde{\mathcal{S}}$ with i to get a larger independent set included in $\mathcal{S}$, violating the maximality of $\tilde{\mathcal{S}}$. Hence,

$$\text{(II)} \leq \left(|\mathcal{S}| - |\tilde{\mathcal{S}}|\right) \exp\left(-10\log(n)\right) = \left(|\mathcal{S}| - |\tilde{\mathcal{S}}|\right)/n^{10}. \tag{EC.5}$$

Combining (EC.4) and (EC.5), we have that

$$\sum_{i\in\mathcal{S}} \exp\left(\sum_{j\in\mathcal{S}} \alpha_{ji}\right) \leq \left(|\tilde{\mathcal{S}}| - 1\right) + \left(|\mathcal{S}| - |\tilde{\mathcal{S}}| + 1\right)/n^{10} \leq \left(|\tilde{\mathcal{S}}| - 1\right) + 1/n^9 < |\tilde{\mathcal{S}}|.$$

Consequently, (EC.3) is checked and (**C1**) is verified.

We next check (**C2**). Given two independent set $\mathcal{S}_1, \mathcal{S}_2 \subseteq V$ with $|\mathcal{S}_1| > |\mathcal{S}_2|$, the revenue generated by the corresponding assortments are $\frac{|\mathcal{S}_1|/n^{10}}{1+|\mathcal{S}_1|/n^{10}}$ and $\frac{|\mathcal{S}_2|/n^{10}}{1+|\mathcal{S}_2|/n^{10}}$, where the former one is larger than the later one as $\frac{y'}{1+y'} > \frac{y}{1+y}$ when $y' > y \geq 0$. Hence, (**C2**) is verified.

Suppose $\mathcal{S}^*$ is the optimal assortment of the constructed instance. By (**C1**) and (**C2**), $\mathcal{S}^*$ is a maximum independent set. Let $\hat{\mathcal{S}}$ be an assortment with revenue greater than or equal to $\rho(n)$ times that of $\mathcal{S}^*$. Greedily construct a maximal independent set $\bar{\mathcal{S}} \subseteq \hat{\mathcal{S}}$; the same argument as in (**C1**), which uses only maximality, shows that the revenue of $\bar{\mathcal{S}}$ is higher than that of $\hat{\mathcal{S}}$. Since $\bar{\mathcal{S}}$ and $\mathcal{S}^*$ are independent sets,

$$\rho(n) \leq \frac{|\bar{\mathcal{S}}|}{|\mathcal{S}^*|}\frac{n^{10}+|\mathcal{S}^*|}{n^{10}+|\bar{\mathcal{S}}|} \leq \left(1+\frac{1}{n^9}\right)\frac{|\bar{\mathcal{S}}|}{|\mathcal{S}^*|} \quad \Longrightarrow \quad \frac{|\bar{\mathcal{S}}|}{|\mathcal{S}^*|} \geq \frac{\rho(n)}{1+1/n^9}.$$

Thus, the reduction preserves the approximation ratio up to the factor $1 + 1/n^9 \leq 2$, and hence preserves the stated asymptotic inapproximability ratio. This concludes our proof. □

Proof of Proposition 1 Let us conduct a reduction (which is not ratio-preserved) from the maximum independent set (MIS) problem, which is well-known to be NP-hard, using the synergistic model.

MIS Problem: Given an undirected graph $G = (V, E)$, MIS calls for a vertex set $S \subseteq V$ with the maximum cardinality such that no pair of nodes in S is connected by some edge in E.

Instance Construction: We need to construct an instance of (3) with $\alpha_{ij} \geq 0$ for all $i, j \in [n]$ to solve the MIS problem. Let the grand product set be $[n+1]$, and

$$[n] = V, \quad r_i = 0, \ \forall i \in [n], \quad r_{n+1} = 1, \quad \mu_i = -10\log(n), \ \forall i \in [n], \quad \mu_{n+1} = 0,$$
$$\alpha_{ji} = \begin{cases} 20\log(n) & e_{ji} \in E, \\ 0 & e_{ji} \notin E, \end{cases} \forall i, j \in [n], \quad \alpha_{j(n+1)} = \frac{1}{n}, \ \alpha_{(n+1)j} = 0, \ \forall j \in [n].$$

Note that the constructed r_i, μ_i, α_{ji}'s are all of polynomial size. In the following, we will use V and $[n]$ (i.e., the vertex set and the product set) interchangeably. To prove the correctness of such a reduction, we need to make sure that solving (3) with the above constructed instance equivalently solves the given MIS problem. We first note that item $n+1$ will always be picked into our optimal assortment; otherwise, the resulting objective will be 0. Therefore, as long as we can ensure that the products picked from $[n]$ into our optimal assortment constitute the maximum independent set of the MIS problem, we achieve our task. This requires us to ensure the following two criteria are satisfied:

(**C1**) For any set $\mathcal{S} \subseteq V$ that is not an independent set, let $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ be the maximum independent set that is included in $\mathcal{S}$. Then, our constructed instance of (3) prefers $\tilde{\mathcal{S}} \cup \{n+1\}$ over $\mathcal{S} \cup \{n+1\}$ since it generates higher revenue.

(**C2**) For two independent sets $\mathcal{S}_1, \mathcal{S}_2 \subseteq V$, as long as $|\mathcal{S}_1| > |\mathcal{S}_2|$, then our constructed instance of (3) prefers $\mathcal{S}_1 \cup \{n+1\}$ over $\mathcal{S}_2 \cup \{n+1\}$ since it generates higher revenue.

The first criterion guarantees that our constructed instance of (3) will always choose the assortment that corresponds to an independent set in G, plus item $n+1$. The second criterion guarantees that our constructed instance of (3) will always select the assortment corresponding to the independent set with the maximum size, plus item $n+1$.

We first check (**C1**). For any set $\mathcal{S} \subseteq V$ that is not an independent set, let $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ be the maximum independent set that is included in $\mathcal{S}$. The revenue generated by the assortment $\tilde{\mathcal{S}} \cup \{n+1\}$ is

$$\frac{\exp\left(|\tilde{\mathcal{S}}|/n\right)}{1+\exp\left(|\tilde{\mathcal{S}}|/n\right)+|\tilde{\mathcal{S}}|/n^{10}} > \frac{|\tilde{\mathcal{S}}|/n}{1+|\tilde{\mathcal{S}}|/n+|\tilde{\mathcal{S}}|/n^{10}} \geq \frac{1}{1+n+1/n^9}, \tag{EC.6}$$

where the first inequality follows from the fact that $\exp(x) > x+1 > x$ for any $x > 0$ and $\frac{y'}{C+y'} \geq \frac{y}{C+y}$ for any $C > 0$ and $y' > y$, and the second inequality follows by letting $|\tilde{\mathcal{S}}| = 1$ in the numerator and $|\tilde{\mathcal{S}}| = n$ in the denominator. On the other hand, the revenue generated by the assortment $\mathcal{S} \cup \{n+1\}$ is

$$\begin{aligned} & \frac{\exp\left(|\mathcal{S}|/n\right)}{1+\exp\left(|\mathcal{S}|/n\right)+\sum_{i\in\mathcal{S}}\exp\left(\sum_{j\in\mathcal{S}}\alpha_{ji}\right)/n^{10}} \\ \leq\ & \frac{\exp\left(|\mathcal{S}|/n\right)}{1+\exp\left(|\mathcal{S}|/n\right)+n^{10}} \\ \leq\ & \frac{1}{2/e+n^{10}/e}, \end{aligned} \tag{EC.7}$$

where the first inequality follows because there exists at least one (i,j) pair in $\mathcal{S}$ such that $(i,j) \in E$, and the second inequality follows by plugging in $|\mathcal{S}| = n$ in the numerator and $|\mathcal{S}| = 0$ in the denominator. Since $\frac{1}{2/e+n^{10}/e} < \frac{1}{1+n+1/n^9}$ when $n \geq 2$, by (EC.6) and (EC.7), we conclude that $\tilde{\mathcal{S}} \cup \{0\}$ generates higher value than $\mathcal{S} \cup \{0\}$ and is more preferred.

We then check (**C2**). For two independent sets $\mathcal{S}_1, \mathcal{S}_2 \subseteq V$ with $|\mathcal{S}_1| > |\mathcal{S}_2|$, where we want to ensure that the revenue generated by $\mathcal{S}_1 \cup \{n+1\}$ is larger than that generated by $\mathcal{S}_2 \cup \{n+1\}$, i.e.,

$$\begin{aligned} & \frac{\exp\left(|\mathcal{S}_1|/n\right)}{1+\exp\left(|\mathcal{S}_1|/n\right)+|\mathcal{S}_1|/n^{10}} > \frac{\exp\left(|\mathcal{S}_2|/n\right)}{1+\exp\left(|\mathcal{S}_2|/n\right)+|\mathcal{S}_2|/n^{10}} \\ \iff\ & \exp\left(|\mathcal{S}_1|/n\right)\left(1+|\mathcal{S}_2|/n^{10}\right) > \exp\left(|\mathcal{S}_2|/n\right)\left(1+|\mathcal{S}_1|/n^{10}\right) \\ \iff\ & \exp\left((|\mathcal{S}_1|-|\mathcal{S}_2|)/n\right) > \frac{1+|\mathcal{S}_1|/n^{10}}{1+|\mathcal{S}_2|/n^{10}}. \end{aligned}$$

To establish the above inequality, first observe

$$\exp\left((|\mathcal{S}_1|-|\mathcal{S}_2|)/n\right) \geq \exp\left(1/n\right) > 1+1/n, \tag{EC.8}$$

where the first inequality follows since $|\mathcal{S}_1| > |\mathcal{S}_2|$, and the second inequality follows because $\exp(x) > 1+x$ for $x > 0$. Besides, observe

$$\frac{1+|\mathcal{S}_1|/n^{10}}{1+|\mathcal{S}_2|/n^{10}} < 1+1/n^9, \tag{EC.9}$$

where the inequality follows by plugging in $|\mathcal{S}_1| = n$ in the numerator and $|\mathcal{S}_2| = 0$ in the denominator. Since $1+1/n > 1+1/n^9$, combining (EC.8) and (EC.9) we conclude that the revenue generated by $\mathcal{S}_1 \cup \{n+1\}$ is larger than that generated by $\mathcal{S}_2 \cup \{n+1\}$. This proves the correctness of our reduction and concludes our proof. □

Proof of Proposition 4 We conduct a reduction from the partition problem, which is known to be NP-complete.

Partition Problem: Given $a_1, \ldots, a_n \in \mathbb{Z}_+$, decide whether there exists $S \subseteq [n]$ such that $\sum_{i \in S} a_i = \sum_{i \notin S} a_i$.

Instance Construction: Consider the case when $K = 1$. Define $b_i = \frac{2a_i}{\sum_{j=1}^n a_j}$ for each $i \in [n]$. Let $r_i = 1$, $\mu_i = \log(b_i)$, $f_{i1} = b_i$, and $h_{1i} = -1$ for all $i \in [n]$. Therefore, (3) reduces to be

$$\begin{aligned} & \max_{\boldsymbol{x} \in \{0,1\}^n} && \frac{\sum_{i=1}^n b_i x_i \exp\left(-\sum_{j=1}^n b_j x_j\right)}{1 + \sum_{i=1}^n b_i x_i \exp\left(-\sum_{j=1}^n b_j x_j\right)} \\ \iff & \max_{\boldsymbol{x} \in \{0,1\}^n} && \sum_{i=1}^n b_i x_i \exp\left(-\sum_{j=1}^n b_j x_j\right) \\ \iff & \max_{\boldsymbol{x} \in \{0,1\}^n} && \log\left(\sum_{i=1}^n b_i x_i\right) - \sum_{i=1}^n b_i x_i. \end{aligned}$$

Since the function $\log(y) - y$ is concave in x and has a unique maximizer at $y^* = 1$, then if there exists a polynomial time algorithm that can solve (3) optimally for the above constructed instance and return $\boldsymbol{x}^*$, then we can check whether $\sum_{i=1}^n b_i x_i^* = \frac{1}{2}\sum_{i=1}^n b_i = \frac{1}{2}\sum_{i=1}^n \frac{2a_i}{\sum_{j=1}^n a_j} = 1$ to decide whether there exists a valid partition for $\{a_i\}_{i \in [n]}$ or not. Hence, even if Assumption 3 is made and $K = 1$, (3) is still NP-hard. □

Proof of Proposition 6 The total log-likelihood $\mathcal{LL}(\boldsymbol{\mu}, A)$ is a sum of terms, one for each observation t. A sum of concave functions is concave. Therefore, it is sufficient to prove that the log-likelihood contribution from a single observation is a concave function of the parameters. Let us consider an arbitrary observation and drop the subscript t for notational simplicity. The single-observation log-likelihood is

$$\ell(\boldsymbol{\mu}, A) = \sum_{i=1}^n y_i \left(\mu_i + \sum_{j=1}^n x_j \alpha_{ji}\right) - \log\left(1 + \sum_{k=1}^n x_k \exp\left(\mu_k + \sum_{j=1}^n x_j \alpha_{jk}\right)\right).$$

Let $\boldsymbol{\theta}$ be the vector containing all model parameters $(\mu_1, \ldots, \mu_n)$ and $(\alpha_{ji})_{j,i \in [n]}$. Let $\boldsymbol{u} = (u_1, \ldots, u_n)$ be the vector of deterministic utilities, where $u_k = \mu_k + \sum_{j=1}^n x_j \alpha_{jk}$. The mapping from the parameter vector $\boldsymbol{\theta}$ to the utility vector $\boldsymbol{u}$ is an affine transformation. We can write the single-observation log-likelihood as a composition of two functions: $\ell(\boldsymbol{\theta}) = f(\boldsymbol{u}(\boldsymbol{\theta}))$, where $\boldsymbol{u}(\boldsymbol{\theta})$ is the affine map from parameters to utilities, and $f(\boldsymbol{u}) = \sum_{i=1}^n y_i u_i - \log(1 + \sum_{k=1}^n x_k e^{u_k})$. A property of concave functions is that their concavity is preserved under composition with an affine map. Thus, if we can show that $f(\boldsymbol{u})$ is a concave function of $\boldsymbol{u}$, it follows that $\ell(\boldsymbol{\mu}, A)$ is a concave function of $(\boldsymbol{\mu}, A)$. The concavity of $f(\boldsymbol{u})$ follows directly from the convexity of log-sum-exp functions (see, for example, 3.1.5 of Boyd and Vandenberghe 2004). □

Proof of Proposition 7 For ease of exposition, we drop the subscript t and focus on any single customer. Suppose $\boldsymbol{\theta} = (\boldsymbol{\mu}, A)$, and let $\boldsymbol{\theta}_i := \left(\mu_i, \{\alpha_{ji}\}_{j \neq i}\right) \in \mathbb{R}^n$. For any offer set $\boldsymbol{x} \in \mathcal{X}_i$, dividing $p_i(\boldsymbol{x}; \boldsymbol{\theta})$ by $p_0(\boldsymbol{x}; \boldsymbol{\theta})$ in equation (2) yields the *log-odds linearization*

$$\log \frac{p_i(\boldsymbol{x}; \boldsymbol{\theta})}{p_0(\boldsymbol{x}; \boldsymbol{\theta})} = \mu_i + \sum_{j \neq i} x_j \, \alpha_{ji} = \langle \boldsymbol{d_i}(\boldsymbol{x}), \, \boldsymbol{\theta}_i \rangle, \qquad \text{for all } \boldsymbol{x} \in \mathcal{X}_i. \tag{EC.10}$$

- **$(b) \Rightarrow (a)$: full rank implies identifiability.** Assume $\operatorname{rank}(V_i) = n$ for each i. Suppose $\boldsymbol{\theta}$ and $\boldsymbol{\theta}' = (\boldsymbol{\mu}', A')$ satisfy $p_i(\boldsymbol{x}; \boldsymbol{\theta}) = p_i(\boldsymbol{x}; \boldsymbol{\theta}')$ for all $i \in [n] \cup \{0\}$ and all $\boldsymbol{x}$. Then, for each i and each $\boldsymbol{x} \in \mathcal{X}_i$, we have,

$$\langle \boldsymbol{d}_i(\boldsymbol{x}), \, \boldsymbol{\theta}_i \rangle = \log \frac{p_i(\boldsymbol{x}; \boldsymbol{\theta})}{p_0(\boldsymbol{x}; \boldsymbol{\theta})} = \log \frac{p_i(\boldsymbol{x}; \boldsymbol{\theta}')}{p_0(\boldsymbol{x}; \boldsymbol{\theta}')} = \langle \boldsymbol{d}_i(\boldsymbol{x}), \, \boldsymbol{\theta}_i' \rangle \quad \text{for all } \boldsymbol{x} \in \mathcal{X}_i.$$

Stacking over $\boldsymbol{d_i}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}_i$ gives $D_i\boldsymbol{\theta}_i = D_i\boldsymbol{\theta}'_i$, hence $D_i(\boldsymbol{\theta}_i - \boldsymbol{\theta}'_i) = \mathbf{0}$. Because $\mathrm{rank}(D_i) = n$, the nullspace of D_i is trivial, so $\boldsymbol{\theta}_i = \boldsymbol{\theta}'_i$ for every i. Therefore $\boldsymbol{\mu} = \boldsymbol{\mu}'$ and each column of A equals the corresponding column of A', i.e., $\boldsymbol{\theta} = \boldsymbol{\theta}'$.

- $(\boldsymbol{a}) \Rightarrow (\boldsymbol{b})$ **: lack of full rank implies non-identifiability.** Assume that $\mathrm{rank}(D_{i^\star}) < n$ for some $i^\star \in [n]$. Then there exists a nonzero vector $\boldsymbol{\delta} \in \mathbb{R}^n$ with $D_{i^\star}\boldsymbol{\delta} = \mathbf{0}$. Fix any parameter $\boldsymbol{\theta} = (\boldsymbol{\mu}, A)$ and construct $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\mu}}, \tilde{A})$ by modifying only the $i^\star$-th coefficient vector:

$$\tilde{\boldsymbol{\theta}}_{i^\star} := \boldsymbol{\theta}_{i^\star} + t\boldsymbol{\delta} \quad \text{for some } t \in \mathbb{R}, \qquad \tilde{\boldsymbol{\theta}}_i := \boldsymbol{\theta}_i \ \text{ for all } i \neq i^\star.$$

For any offered set $\boldsymbol{x}$:

—If $x_{i^\star} = 0$, item $i^\star$ does not enter the denominator, so the entire vector of exponents $\{\mu_k + \sum_j x_j \alpha_{jk}\}_k$ is unchanged; thus $p_i(\boldsymbol{x}; \tilde{\boldsymbol{\theta}}) = p_i(\boldsymbol{x}; \boldsymbol{\theta})$ for all $i \in [n] \cup \{0\}$.

—If $x_{i^\star} = 1$, then

$$\langle \boldsymbol{d}_{i^\star}(\boldsymbol{x}), \tilde{\boldsymbol{\theta}}_{i^\star} \rangle = \langle \boldsymbol{d}_{i^\star}(\boldsymbol{x}), \boldsymbol{\theta}_{i^\star} \rangle + t \langle \boldsymbol{d}_{i^\star}(\boldsymbol{x}), \boldsymbol{\delta} \rangle = \langle \boldsymbol{d}_{i^\star}(\boldsymbol{x}), \boldsymbol{\theta}_{i^\star} \rangle,$$

because $D_{i^\star}\boldsymbol{d} = \mathbf{0}$. Hence the exponent for item $i^\star$ is unchanged for all $\boldsymbol{x} \in \mathcal{X}_{i^\star}$. Since no other item's coefficients were altered, the entire denominator and all numerators in the choice probability are the same under $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$. Therefore $p_i(\boldsymbol{x}; \tilde{\boldsymbol{\theta}}) = p_i(\boldsymbol{x}; \boldsymbol{\theta})$ for all $i \in [n] \cup \{0\}$ and all $\boldsymbol{x}$ with $\tilde{\boldsymbol{\theta}} \neq \boldsymbol{\theta}$, implying non-identifiability.

Combining the two directions establishes the equivalence of (a) and (b). $\square$