A market discovery algorithm to estimate a general class of non-parametric choice models

APPENDIX

Garrett van Ryzin

Graduate School of Business, Columbia University, New York, NY 10027

e-mail: gjv1@columbia.edu

Gustavo Vulcano

Leonard N. Stern School of Business, New York University, New York, NY 10012,

e-mail: gvulcano@stern.nyu.edu

A1 Proofs of technical results

Proof of Proposition 1

First, we show that the following constraint qualification holds: $\mathbf{0} \in \operatorname{int} h(X, Y)$, where

$$h(X,Y) = \left\{ (a, \boldsymbol{b}),\ a \in \mathcal{R},\ \boldsymbol{b} \in \mathcal{R}^T,\ \text{with } a = \sum_{i=1}^{K} x_i - 1,\ \boldsymbol{b} = A\boldsymbol{x} - \boldsymbol{y},\ \text{for } \boldsymbol{x}, \boldsymbol{y} \geq 0 \right\}$$

The condition is equivalent to showing

$$\exists\, \epsilon > 0 \text{ s.t. } ||(a, \boldsymbol{b})|| < \epsilon \Rightarrow (a, \boldsymbol{b}) \in h(X, Y).$$

We will show instead that given $\epsilon > 0$, there is a $(1 + T)$-dimensional hypercube centered at zero with edge length 2ϵ that is included in int $h(X, Y)$. This automatically implies that a ball of radius ϵ is included in int $h(X, Y)$. We argue below that all corners of the hypercube belong to int $h(X, Y)$. We pick ϵ small enough such that $\epsilon = \max\{\bar{\epsilon} > 0 : K \leq (1 - \bar{\epsilon})/\bar{\epsilon}\}$. The are two cases to consider:

1. $(a, \boldsymbol{b}) = (\epsilon, \pm\boldsymbol{\epsilon})$. Here, $a = \epsilon$ is a scalar, and $\boldsymbol{b} = \pm\boldsymbol{\epsilon} \in \mathcal{R}^T$. Vector $\boldsymbol{b}$ has a mix of positive and negative ϵ in its coordinates.

   Define $x_i = (1+\epsilon)/K$. Note that, $x_i \geq 0$, and $\sum_{i=1}^{K} x_i = 1+\epsilon$, so that $a = \sum_{i=1}^{K} x_i - 1 = \epsilon$. For $\boldsymbol{b}$, since $\boldsymbol{b} = A\boldsymbol{x} - \boldsymbol{y}$, we have $\boldsymbol{y} = A\boldsymbol{x} - \boldsymbol{b}$. For b_t such that $b_t = \epsilon$,

   $$y_t = (A\boldsymbol{x})_t - \epsilon \geq \frac{1+\epsilon}{K} - \epsilon \geq \frac{1-\epsilon}{K} - \epsilon \geq 0,$$

   where the first inequality follows from the fact that at least one type is compatible with the transaction in period t, and the last one follows from the definition of ϵ.

   On the other hand, if $b_t = -\epsilon$, then $y_t = (A\boldsymbol{x})_t + \epsilon \geq 0$.

   Hence, $(a, \boldsymbol{b}) \in h(X, Y)$.

2. $(a, \boldsymbol{b}) = (-\epsilon, \pm\boldsymbol{\epsilon})$. In this case, $a = -\epsilon$, and vector $\boldsymbol{b}$ has again a mix of positive and negative ϵ in its coordinates.

   Define $x_i = (1-\epsilon)/K$. Note that, $x_i \geq 0$, and $\sum_{i=1}^{K} x_i = 1-\epsilon$, so that $a = \sum_{i=1}^{K} x_i - 1 = -\epsilon$. For $\boldsymbol{b}$, since $\boldsymbol{b} = A\boldsymbol{x} - \boldsymbol{y}$, we have $\boldsymbol{y} = A\boldsymbol{x} - \boldsymbol{b}$. For b_t such that $b_t = \epsilon$,

$$y_t = (A\boldsymbol{x})_t - \epsilon \geq \frac{1-\epsilon}{K} - \epsilon \geq 0,$$

   where the first inequality follows from the fact that at least one type is compatible with the transaction in period t, and the second one follows from the definition of ϵ.

   On the other hand, if $b_t = -\epsilon$, then $y_t = (A\boldsymbol{x})_t + \epsilon \geq 0$.

   Hence, $(a, \boldsymbol{b}) \in h(X, Y)$.

Since $\mathcal{L}_I(\boldsymbol{y})$ is concave and the feasible region is defined by affine functions, and given the above constraint qualification, then from Theorem 6.2.4 in Bazaraa et al. [4], the primal and dual objective functions coincide. Moreover, from Theorem 6.2.5 and its corollary therein, the primal-dual solution $(\boldsymbol{x}^*, \boldsymbol{y}^*, \beta^*, \boldsymbol{\mu}^*)$ is a saddle point of the Lagrangian. ∎

Proof of Proposition 2

We start by proving the following auxiliary result:

Lemma A1 *For a given offer set S and pairs (j, S), for all $j \in S$, there exists a partition of the set of types $\boldsymbol{\sigma}$, so that*

$$\sum_{j \in S} \sum_{i \in \mathcal{M}(j,S)} x_i = 1.$$

Proof. First, note that for any type $\sigma^{(i)}$, there exist $j \in S$ such that $\sigma^{(i)}(j) < \sigma^{(i)}(j')$, for all $j' \in S$, $j' \neq j$. So, every type should be assigned to one of the sets $\mathcal{M}(\cdot, S)$. Second, consider two pairs (j, S) and (j', S). If $i \in \mathcal{M}(j, S)$, then $i \notin \mathcal{M}(j', S)$. This is because if $i \in \mathcal{M}(j, S)$, then $\sigma^{(i)}(j) < \sigma^{(i)}(j')$, for all $j' \in S$, $j' \neq j$, and hence we cannot have $\sigma^{(i)}(j') < \sigma^{(i)}(j)$. ∎

Going back to the main result, we formulate a collection of related problems. Define continuous variables $z_{j,S}$ for each S listed at least once in the horizon $t = 1, \ldots, \hat{T}$. Let $b_{j,S}$ be the number of times that the pair (j, S) is observed along the horizon, and consider the set of problems:

$$\begin{aligned} \max_{\boldsymbol{z} \geq 0} \quad & \sum_{j=1}^{n} b_{j,S} \log z_{j,S} \\ \text{s.t.:} \quad & \sum_{j \in S} z_{j,S} = 1. \end{aligned} \tag{A1}$$

Note that from the hypothesis, for all $j \in S$, we have $b_{j,S} \geq 1$. Note also that the objective function is concave. The Karush-Kuhn-Tucker (KKT) conditions give unique maximizers

$$\hat{z}_{j,S} = \frac{b_{j,S}}{\sum_{j' \in S} b_{j',S}}, \quad \text{for each } j = 1, \ldots, n.$$

Take a solution $\boldsymbol{x}^*$ of problem (3). By construction, there is an associated solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ of problem (4). Take the system $A\boldsymbol{x}^* = \boldsymbol{y}^*$ and w.l.o.g. delete repeated rows, so that any row representing the pair (j, S) appears only once. Furthermore, rearrange all rows corresponding to the same offer set S so that they are listed consecutively in the system of equations. Let M be the number of different offer sets listed along the horizon, and let $S^{(1)}, \ldots, S^{(M)}$ be those offer sets, of corresponding cardinality $s_1, \ldots, s_M$. Let $\underline{\mathrm{T}} = \sum_{m=1}^{M} s_m$ be the number of rows in the new system that satisfies $\underline{\mathrm{A}}\boldsymbol{x}^* = \underline{\boldsymbol{y}}^*$, where $\underline{\boldsymbol{y}}^*$ is given by the nondeleted coordinates of the original vector $\boldsymbol{y}^*$.

Next, we define the matrix $C \in \{0,1\}^{M \times \underline{\mathrm{T}}}$ as a matrix with blocks of consecutive ones in each row, with $c_{m,j} = 1$ if $j \in S_m$; and $c_{m,j} = 0$, otherwise. The optimal solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$ is still feasible for the system $(C\underline{\mathrm{A}})\boldsymbol{x}^* = C\underline{\boldsymbol{y}}^*$, with $\boldsymbol{y}^*$ projected onto $\underline{\boldsymbol{y}}^*$. Row m in the LHS corresponds to:

$$\sum_{j \in S_m} \sum_{i \in \mathcal{M}_m(j, S_m)} x_i = 1, \text{ from Lemma A1 above.} \tag{A2}$$

The RHS corresponds to

$$\sum_{j \in S_m} \underline{\mathrm{y}}_j = 1,$$

where the equality holds from (A2).

Regrouping the terms in the objective function of (4), and using the separability of the objective function with respect to the offer sets S_t, we get the collection of equivalent problems (A1), and solutions $\underline{\boldsymbol{y}}^*$ for them. Since the solution for each of these problems is unique, our original problem (4) has a unique solution $\boldsymbol{y}^*$. ∎

Proof of Proposition 3

Given a connected graph $G = (V, E)$, with nodes $V = \{1, 2, \ldots, v\}$, $v \geq 2$, and arcs $E \subset \{(i,j) \in V \times V, i < j\}$, a *maximum independent set* of G is a subset $V' \subset V$ such that there is no arc between two nodes in V', and V' is of maximum cardinality.

Let I be an instance of maximum independent set. We will construct an instance J of problem (11) corresponding to I. There will be $n = v$ products and $T = v$ binary decision variables $w_1, \ldots, w_v$. The variable w_i represents node $i \in V$. For each node $i \in V$, we consider the set of adjacent nodes $\Gamma(i) = \{j : (i,j) \in E\}$. Then, for every node $j \in V$, we define the "transaction" $j_t := j$ and the availability set $S_t := \Gamma(j) \bigcup \{i, 0\}$. Finally, we set the coefficients $\mu_t^* := 1$.

Note that in the optimum of J, we cannot have $w_{i_0} = w_{j_0} = 1$, for an arc $(i_0, j_0) \in E$. Arguing by contradiction: Take an arc $(i_0, j_0) \in E$. If $w_{i_0} = 1$, and since $j_0 \in \Gamma(i_0)$, then $\sigma(i_0) < \sigma(j_0)$. Similarly, if $w_{j_0} = 1$, and since $i_0 \in \Gamma(j_0)$, then $\sigma(j_0) < \sigma(i_0)$, which gives an inconsistency in the order σ. Therefore, the output of the optimization problem (11) for instance J is an independent set of maximum cardinality. ∎

A2 Complementary result

Proposition A1 *The function $\mathcal{L}_I(\boldsymbol{x}, \lambda)$ in the optimization problem (14) is not quasiconcave in general.*

Proof. We exhibit an example where the objective function in (14) is not quasiconcave over the feasible set. Consider a case where there are three customer types and, given the product availability and transaction data, the objective function is:

$$\begin{aligned}\mathcal{L}_I(\boldsymbol{x}, \lambda) \quad = \quad & 3(\log \lambda + \log x_1) + \log \lambda + \log(x_2 + x_3) + \log \lambda + \log x_3 \\ & +2\log(\lambda(x_2 + x_3) + 1 - \lambda) + 100\log(\lambda(x_1 + x_2) + 1 - \lambda) + 3\log(1 - \lambda).\end{aligned}$$

Take the following two points $(\boldsymbol{x}, \lambda)$: The first one is $(\boldsymbol{x}^1, \lambda^1) = (0.08, 0.15, 0.77, 0.1)$, with $\mathcal{L}_I(\boldsymbol{x}^1, \lambda^1) = -12.07$; the second one is $(\boldsymbol{x}^2, \lambda^2) = (0.07, 0.64, 0.29, 0.7)$, with $\mathcal{L}_I(\boldsymbol{x}^2, \lambda^2) = -16.747$. If we take $\alpha = 0.4$, we get an intermediate point $(\boldsymbol{x}^\alpha, \lambda^\alpha) = \alpha(\boldsymbol{x}^1, \lambda^1) + (1-\alpha)(\boldsymbol{x}^2, \lambda^2) = (0.074, 0.444, 0.482, 0.46)$, with $\mathcal{L}_I(\boldsymbol{x}^\alpha, \lambda^\alpha) = -17.316$, which is smaller than the previous two. Therefore, quasiconcavity is violated. ∎

A3 Supplement to numerical experiments

A3.1 DVD data from Amazon.com

Table A1 summarizes the MNL model fit to Amazon.com DVD sales data collected between July 1, 2005, and September 30, 2005, as reported by Farias et al. [16, Section 4.2.1]. The third column includes weights of the MNL model used by Farias et al., and that we use here in our synthetic-data-based MNL and CNL experiments. The fourth to seventh columns are perturbations of the second column and correspond to the weights of the products for the different segments considered in our LC-MNL model. For the latter, the proportions of the segments are $(0.3, 0.2, 0.2, 0.2, 0.1)$.

A3.2 Amazon Example: Complementary table

In the table below we compare the performance of the independent demand assumption with that of the set of types discovered by our MD algorithm when the training dataset contains $T = 500$ periods and the market share of the firm varies. We provide three measures: the RMSE, the maximum log-likelihood value, and the AIC$_c$. We also report the final number of types found by MD.

Product ID	Segment #1		Segment #2	Segment #3	Segment #4	Segment #5
	Nominal utility u_j	Weight	Weight	Weight	Weight	Weight
1	-4.513	0.010964595	0.009392069	0.011489518	0.015719293	0.015227003
2	-4.601	0.010044509	0.007369928	0.013246395	0.008167432	0.006489315
3	-4.790	0.008311871	0.004532987	0.01074264	0.011920438	0.007996986
4	-4.514	0.010950576	0.007529854	0.006514414	0.005760067	0.010145039
5	-4.312	0.013410778	0.010115117	0.017725151	0.013298268	0.012448655
6	-4.839	0.007911401	0.008657218	0.00414178	0.004253296	0.004400657
7	-4.888	0.007538734	0.004526155	0.006760084	0.007890648	0.009846412
8	-4.758	0.00858676	0.010697536	0.011981461	0.009767898	0.012651933
9	-4.553	0.010537923	0.009651463	0.012687673	0.010377687	0.011765564
10	-4.594	0.010111887	0.011051911	0.008709671	0.007349483	0.01008567
11	-4.553	0.010537428	0.009602063	0.010356076	0.011163975	0.009054309
12	-3.590	0.027601941	0.024818242	0.035120775	0.04065703	0.032972036
13	-4.739	0.008747872	0.012635604	0.011750436	0.012965096	0.009010166
14	-4.697	0.009119782	0.010188665	0.012826062	0.009666714	0.011076749
15	-4.707	0.009035462	0.004931018	0.011483989	0.005424855	0.00582217

Table A1: Description of the Amazon data. Segment #1 corresponds to the MNL fit used in Farias et al. [16]. Segments #2 through #5 are based on perturbations of Segment #1.

Ground-truth	Market share	Indep. demand perf.			MD performance			
		RMSE	LL value	AIC$_c$	# of types	RMSE	LL value	AIC$_c$
MNL	0.2	0.13	-249.39	529.77	15	0.13	-249.39	529.77
	0.4	0.08	-310.89	652.77	15	0.08	-310.89	652.77
	0.6	0.03	-449.54	930.07	17	0.05	-437.05	909.37
	0.8	0.11	-660.40	1,351.79	20	0.05	-581.94	1,205.63
CNL	0.2	0.05	-374.99	780.97	15	0.05	-374.99	780.97
	0.4	0.05	-512.42	1,055.83	17	0.05	-487.91	1,011.09
	0.6	0.13	-774.13	1,579.25	22	0.08	-624.66	1,295.44
	0.8	0.21	-915.04	1,861.07	25	0.08	-664.98	1,382.70
LC-MNL	0.2	0.12	-226.73	484.45	15	0.12	-226.73	488.45
	0.4	0.08	-344.56	720.11	15	0.08	-344.56	720.11
	0.6	0.03	-494.47	1,019.93	18	0.07	-468.83	975.08
	0.8	0.11	-653.84	1,338.67	20	0.05	-581.37	1,204.49

Table A2: Performance of the indep. demand assumption and the MD algorithm under different market share values. The initial number of types is $N = 15$. The number of in-sample periods is $T = 500$.