Demand Estimation under the Multinomial Logit Model from Sales Transaction Data

Tarek Abdallah
Kellogg School of Management, Operations Department, Evanston,
tarek.abdallah@kellogg.northwestern.edu

Gustavo Vulcano
School of Business, Universidad Torcuato Di Tella, gvulcano@utdt.edu
Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina

APPENDIX A

Appendix A1: Proofs of Technical Results

Proof of Lemma 1. First notice that any optimal solution $(\boldsymbol{\beta}^*,\boldsymbol{\lambda}^*)$ for the log-likelihood problem $\ell_I(\boldsymbol{\beta},\boldsymbol{\lambda})$ with restricted market share should satisfy the KKT conditions (LICQ regularity condition holds) given by

$$\frac{\partial \ell_I(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial \beta_l}|_{(\boldsymbol{\beta}^*,\boldsymbol{\lambda}^*)} = \pi \exp(\beta_l^*), \qquad \text{for all } l\in\mathcal{N},$$
$$\lambda_t^* = m_t \frac{1+\sum_{i\in S_t}\exp(\beta_i^*)}{\sum_{j\in S_t}\exp(\beta_j^*)} \qquad \text{for all } t=1,\dots,T.$$
$$\sum_{i=1}^n \exp(\beta_i^*) = \tilde{s},$$

for some $\pi\in\mathbb{R}$ that is the Lagrange multiplier of the market share constraint.

Moreover, the partial derivative with respect to β_l is given by

$$\begin{aligned}\frac{\partial \ell_I(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial \beta_l} &= \sum_{t=1}^T \mathbb{1}_{\{l\in S_t\}}\left[-m_t\frac{\exp(\beta_l)}{1+\sum_{j\in S_t}\exp(\beta_j)} - \lambda_t \frac{\exp(\beta_l)}{\left(1+\sum_{j\in S_t}\exp(\beta_j)\right)^2}\right] + K_l \\ &= \sum_{t=1}^T \mathbb{1}_{\{l\in S_t\}}\left[-m_t\frac{\exp(\beta_l)}{1+\sum_{j\in S_t}\exp(\beta_j)}\left(1+\frac{\lambda_t}{m_t\left(1+\sum_{j\in S_t}\exp(\beta_j)\right)}\right)\right] + K_l.\end{aligned}$$

Evaluating the partial derivative at $(\boldsymbol{\beta}^*,\boldsymbol{\lambda}^*)$ where $\lambda_t^* = m_t\frac{1+\sum_{j\in S_t}\exp(\beta_j^*)}{\sum_{j\in S_t}\exp(\beta_j^*)}$, we get

$$\begin{aligned}\frac{\partial \ell_I(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial \beta_l}|_{(\boldsymbol{\beta}^*,\boldsymbol{\lambda}^*)} &= \sum_{t=1}^T \mathbb{1}_{\{l\in S_t\}}\left[-m_t\frac{\exp(\beta_l^*)}{1+\sum_{j\in S_t}\exp(\beta_j^*)}\left(1+\frac{1}{\sum_{j\in S_t}\exp(\beta_j^*)}\right)\right] + K_l \\ &= \sum_{t=1}^T \mathbb{1}_{\{l\in S_t\}}\left[-m_t\frac{\exp(\beta_l^*)}{\sum_{j\in S_t}\exp(\beta_j^*)}\right] + K_l,\end{aligned}$$

where the last equality is the same as the partial derivative of $\ell_{MNL}(\boldsymbol{\beta})$. As a result, the KKT conditions of $\ell_I(\boldsymbol{\beta},\boldsymbol{\lambda})$ reduce to

$$\frac{\partial \ell_{MNL}(\boldsymbol{\beta}^*)}{\partial \beta_l}|_{\boldsymbol{\beta}^*} = \pi \exp(\beta_l^*), \qquad \text{for all } l\in\mathcal{N}, \tag{A1}$$
$$\lambda_t^* = m_t \frac{1+\sum_{i\in S_t}\exp(\beta_i^*)}{\sum_{j\in S_t}\exp(\beta_j^*)} \qquad \text{for all } t=1,\dots,T,$$
$$\sum_{i=1}^n \exp(\beta_i^*) = \tilde{s},$$

for some $\pi \in \mathbb{R}$. These are the same KKT conditions for $\ell_{MNL}(\beta^*)$ with the added condition on λ_t^*.

We next show that $\pi = 0$. Summing equation (A1) for $j = 1, \ldots, n$, we obtain

$$\sum_{t=1}^{T} -m_t \left[\frac{\sum_{j\in S_t} \exp(\beta_j^*)}{\sum_{j\in S_t} \exp(\beta_j^*)}\right] + \sum_{j=1}^{n} K_j = \pi \sum_{j=1}^{N} \exp(\beta_j^*)$$
$$\Leftrightarrow 0 = \pi \tilde{s}.$$

Since $\tilde{s} > 1$, then $\pi = 0$.

Therefore, $\boldsymbol{\beta}^*$ is a stationary point. Moreover, since $\ell_I(\boldsymbol{\beta}, \boldsymbol{\lambda}^*) = \text{constant} + \ell_{MNL}(\boldsymbol{\beta})$, then their respective log-likelihoods share the same set of maximizers, $\boldsymbol{\beta}^*$, which completes the proof. $\square$

Proof of Lemma 2. Let $\alpha = \log\left(\frac{\tilde{s}}{\sum_{i=1}^{n} \exp(\beta_i)}\right)$. We have

$$\begin{aligned}
\ell_{MNL}\left(R(\boldsymbol{\beta}, \tilde{s})\right) &= \sum_{j=1}^{n} K_j\left(\alpha + \beta_j\right) - \sum_{t=1}^{T} m_t \log\left(\sum_{i\in S_t} \exp(\alpha + \beta_i)\right) \\
&= \sum_{j=1}^{n} K_j \alpha + \sum_{j=1}^{n} K_j\left(\beta_j\right) - \sum_{t=1}^{T} m_t \alpha - \sum_{t=1}^{T} m_t \log\left(\sum_{i\in S_t} \exp(\beta_i)\right) \\
&= \sum_{j=1}^{n} K_j \beta_j - \sum_{t=1}^{T} m_t \log\left(\sum_{i\in S_t} \exp(\beta_i)\right) = \ell_{MNL}(\boldsymbol{\beta}).
\end{aligned}$$

The last equality is due to the fact that $\sum_{j=1}^{n} K_j = \sum_{t=1}^{T} m_t$, which are the total number of purchases in the data.

Furthermore, we have

$$\sum_{i=1}^{N} \exp\left(\log\left(\frac{\tilde{s}}{\sum_{i=1}^{n} \exp(\beta_i)}\right) + \beta_i^*\right) = \frac{\tilde{s}}{\sum_{i=1}^{n} \exp(\beta_i)} \sum_{i=1}^{N} \exp(\beta_i^*) = \tilde{s},$$

and clearly $R(\boldsymbol{\beta}, \tilde{s}) = \boldsymbol{\beta}$ if and only if $\sum_{j=1}^{n} \exp(\beta_j) = \tilde{s}$. $\square$

Proof of Proposition 1. By adapting the arguments described in Hunter (2004) to exploit the special structure of the decomposed problem in Lemma 1, we show that analogous uniqueness conditions discussed by him for the MNL estimation problem from complete data with no market share constraint (see Assumption 3 and Lemma 2 therein) also hold for the joint log-likelihood problem subject to the market share constraint. ((i)⇒(ii)) We first show that if $G(\mathcal{N}, E)$ is a strongly connected graph, then optimal solution for the joint log-likelihood problem subject to the market share constraint has to be bounded.

Notice that since $1 < \tilde{s} < \infty$, then any solution for which there exists an item l such that $\beta_l \to +\infty$ is infeasible. Likewise, a solution where $\beta_j \to -\infty$ for all items $j = 1, \ldots, n$ is also infeasible. Hence, the only case when an unbounded solution may is if there exists a non-empty strict subset $\mathcal{K} \subset \mathcal{N}$ such that $\beta_j \to -\infty$ all $j \in \mathcal{K}$. Denote, by $\overline{\mathcal{K}} = \mathcal{N} \backslash \mathcal{K} \neq \emptyset$ for which $-\infty < \beta_j < \infty$ for all $j \in \overline{\mathcal{K}}$.

Since $G(\mathcal{N}, E)$ is strongly connected, then there exists an edge that goes from $\mathcal{K}$ to $\overline{\mathcal{K}}$. Consider the set S_t which defines that edge. We have that $S_t \cap \mathcal{K} \neq \emptyset$, $S_t \cap \overline{\mathcal{K}} \neq \emptyset$, and $z_{lt} \geq 1$ for some item $l \in S_t \cap \mathcal{K}$. Therefore, we get

$$\begin{aligned}
\lim_{||\boldsymbol{\beta}||\to\infty} \ell_I\left(\boldsymbol{\beta}, \lambda\right) &\leq \lim_{||\boldsymbol{\beta}||\to\infty} \log\left(\frac{m_t!}{\prod_{i\in S_t} z_{it}!} \frac{\exp(\beta_l)}{\sum_{i\in S_t\cap\mathcal{K}} \exp(\beta_i) + \sum_{j\in S_t\cap\overline{\mathcal{K}}} \exp(\beta_j)}\right) \\
&= -\infty,
\end{aligned}$$

where the inequality holds because the term in parenthesis is less than or equal to one term of the likelihood function. The last equality is due to the fact that $\beta_l \to -\infty$, $\sum_{j\in S_t\cap\overline{\mathcal{K}}}\exp(\beta_j) \to$ a finite number, and $\sum_{j\in S_t\cap\mathcal{K}}\exp(\beta_j) \to 0$. Therefore, any optimal solution has a bounded $\boldsymbol{\beta}$ vector. Moreover, it is straightforward to show if there exists t such that $\lambda_t \to 0$ or $\lambda_t \to +\infty$ then $\ell_I(\boldsymbol{\beta},\lambda) \to -\infty$. Hence, any optimal solution is necessarily bounded. We next show that the optimal solution is unique.

Assume that there exists two bounded optimal solutions $(\bar{\boldsymbol{\beta}},\bar{\boldsymbol{\lambda}}) \neq (\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\lambda}})$. It follows from Lemma 1, that both $\bar{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ are global optimal solutions to $\max_\beta\{\ell_{MNL}$ s.t. $\sum_{i=1}^n \exp(\beta_i) = \tilde{s}\}$. Consider the convex combination of the two solutions given by $\alpha\bar{\boldsymbol{\beta}} + (1-\alpha)\tilde{\boldsymbol{\beta}}$ for some $\alpha \in (0,1)$.From Lemma 2, we have that $R(\alpha\bar{\boldsymbol{\beta}} + (1-\alpha)\tilde{\boldsymbol{\beta}},\tilde{s})$ is a feasible solution to the market share constraint and

$$\ell_{MNL}\Big(R(\alpha\bar{\boldsymbol{\beta}} + (1-\alpha)\tilde{\boldsymbol{\beta}},\tilde{s})\Big) = \ell_{MNL}\Big(\alpha\bar{\boldsymbol{\beta}} + (1-\alpha)\tilde{\boldsymbol{\beta}}\Big) \tag{A2}$$

At the same time,

$$\begin{aligned}
&\ell_{MNL}(\alpha\bar{\boldsymbol{\beta}} + (1-\alpha)\tilde{\boldsymbol{\beta}})\\
&= \alpha\sum_{j=1}^n K_j\bar{\beta}_j + (1-\alpha)\sum_{j=1}^n K_j\tilde{\beta}_j - \sum_{t=1}^T m_t \log\left[\sum_{i\in S_t}\exp\left(\bar{\beta}_i\right)^\alpha \exp\left(\tilde{\beta}_i\right)^{(1-\alpha)}\right]\\
&\geq \alpha\sum_{j=1}^n K_j\bar{\beta}_j + (1-\alpha)\sum_{j=1}^n K_j\tilde{\beta}_j - \sum_{t=1}^T m_t \log\left[\left(\sum_{i\in S_t}\exp\left(\bar{\beta}_i\right)\right)^\alpha \left(\sum_{i\in S_t}\exp\left(\tilde{\beta}_i\right)\right)^{(1-\alpha)}\right] \qquad \text{(A3)}\\
&= \alpha\sum_{j=1}^n K_j\bar{\beta}_j - \alpha\sum_{t=1}^T m_t\log\left(\sum_{i\in S_t}\exp\left(\bar{\beta}_i\right)\right) + (1-\alpha)\sum_{j=1}^n K_j\tilde{\beta}_j - (1-\alpha)\sum_{t=1}^T m_t\log\left(\sum_{i\in S_t}\exp\left(\tilde{\beta}_i\right)\right)\\
&= \alpha\ell_{MNL}(\bar{\boldsymbol{\beta}}) + (1-\alpha)\ell_{MNL}(\tilde{\boldsymbol{\beta}})\\
&= \ell_{MNL}(\bar{\boldsymbol{\beta}}) = \ell_{MNL}(\tilde{\boldsymbol{\beta}}).
\end{aligned}$$

Inequality (A3) is due to Holder's inequality, which states:

$$\sum_{i\in S} x_i\, y_i \leq \left(\sum_{i\in S} x_i^p\right)^{1/p}\left(\sum_{i\in S} y_i^q\right)^{1/q}, \quad \text{for} \quad p,q \in (1,\infty),\ 1/p + 1/q = 1.$$

In our case, we define $x_i = \exp(\bar{\beta}_i)^\alpha$, and $y_i = \exp(\tilde{\beta}_i)^{1-\alpha}$, with $p = 1/\alpha$ and $q = 1/(1-\alpha)$. It now follows from (A3) and (A2), that inequality (A3) must hold with equality since otherwise $\bar{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ are not optimal. However, inequality (A3) holds with equality if and only if there exists $\delta \in \mathbb{R}$ such that $\exp(\bar{\beta}_i) = \delta\exp(\tilde{\beta}_i)$ for all $i \in \mathcal{N}$.

However, since both $\bar{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ satisfy the market share constraint, then $\sum_{i=1}^n \exp(\bar{\beta}_i) = \sum_{i=1}^n \exp(\tilde{\beta}_i) = \tilde{s}$, hence $\delta = 1$. That is, $\bar{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$ and from Lemma 1 it has to be the case that $\bar{\boldsymbol{\lambda}} = \tilde{\boldsymbol{\lambda}}$ which is a contradiction.

((ii)⇒(i)) We will show the equivalent implication (¬(ii)⇒ ¬(i)). Suppose that $G(\mathcal{N},E)$ is not strongly connected but there exists a unique bounded optimal solution $(\boldsymbol{\beta}^*,\boldsymbol{\lambda}^*)$ to the joint log-likelihood problem with market share constraint. Then there exist two nodes i and j such that there is no directed path from i to j. Let V_i denote the set of nodes that can be reached from i, and define $V_i^c = \mathcal{N}\setminus V_i$. Then, we have that $i \in V_i$ and $j \in V_i^c$, and by construction there is no directed edge from V_i to V_i^c. We now consider two possible cases regarding the existence or not of a directed edge from V_i^c to V_i.

Case 1 (No directed edge from V_i^c to V_i). Suppose there is no directed edge from V_i^c to V_i. We will now construct a solution that violates the uniqueness assumption of $(\boldsymbol{\beta}^*, \boldsymbol{\lambda}^*)$.

Define a new vector $\tilde{\boldsymbol{\beta}}$ such that

$$\tilde{\beta}_l = \begin{cases} \beta_l^* & \text{if } l \in V_i, \\ c + \beta_l^* & \text{if } l \in V_i^c, \end{cases}$$

for any $c \in \mathbb{R}$, $c \neq 0$.

Since V_i and V_i^c are disconnected, then, by construction, for every t, we have $S_t \subseteq V_i$ or $S_t \subseteq V_i^c$. Consequently, we have

$$\frac{\mathcal{L}_{MNL}\left(\boldsymbol{\beta}^*\right)}{\mathcal{L}_{MNL}\left(\tilde{\boldsymbol{\beta}}\right)} = \frac{\Pi_{t=1}^T \Pi_{k\in S_t}\left(\frac{\exp(\beta_k^*)}{\sum_{l\in S_t}\exp(\beta_l^*)}\right)^{z_{kt}}}{\Pi_{t=1}^T \Pi_{k\in S_t}\left(\frac{\exp(\tilde{\beta}_k)}{\sum_{l\in S_t}\exp(\tilde{\beta}_l)}\right)^{z_{kt}}}$$
$$= \Pi_{t=1, S_t \subseteq V_i^c}^T \Pi_{k\in S_t}\left(\frac{\frac{\exp(\beta_k^*)}{\sum_{l\in S_t}\exp(\beta_l^*)}}{\frac{\exp(c+\beta_k^*)}{\sum_{l\in S_t}\exp(c+\beta_l^*)}}\right)^{z_{kt}} = 1.$$

Next, let $\bar{\boldsymbol{\beta}} = R(\tilde{\boldsymbol{\beta}}, \tilde{s})$. We have that $\bar{\boldsymbol{\beta}} \neq \tilde{\boldsymbol{\beta}}$ and following from Lemma 2 we obtain

$$\ell_{MNL}(\bar{\boldsymbol{\beta}}) = \ell_{MNL}(\tilde{\boldsymbol{\beta}}) = \ell_{MNL}(\boldsymbol{\beta}^*).$$

However, from Lemma 1, we have that $\boldsymbol{\beta}^* \in \arg\max\left\{\boldsymbol{\beta} : \ell_{MNL} \text{ s.t. } \sum_{j=1}^n \exp(\beta_j) = \tilde{s}\right\}$. Hence, $\bar{\boldsymbol{\beta}} \neq \boldsymbol{\beta}^*$ is also an optimal solution. Now, let $\bar{\lambda}_t = m_t \frac{1+\sum_{i\in S_t}\exp(\bar{\beta}_i)}{\sum_{i\in S_t}\exp(\bar{\beta}_i)}$, it follows again from Lemma 1 that $(\bar{\boldsymbol{\beta}}, \bar{\lambda}) \neq (\boldsymbol{\beta}, \boldsymbol{\lambda})$ is an optimal solution to the joint log-likelihood problem with market share constraint which is a contradiction.

Case 2 (Directed edge from V_i^c to V_i) Assume there exists a directed edge (h,k). In this case, there exists S_t such that $h \in (S_t \cap V_i^c)$ and $k \in (S_t \cap V_i)$. However, by construction, for all such S_t we have $z_{kt} = 0$ for all $k \in S_t \cap V_i$. Otherwise, there would be a directed edge from V_i to V_i^c.

Define a new vector $\tilde{\boldsymbol{\beta}}$ such that

$$\tilde{\beta}_l = \begin{cases} \beta_l^* & l \in V_i \\ c + \beta_l^* & l \in V_i^c \end{cases}$$

for any $c > 0$. Therefore we have,

$$\frac{\mathcal{L}_{MNL}\left(\boldsymbol{\beta}^*\right)}{\mathcal{L}_{MNL}\left(\tilde{\boldsymbol{\beta}}\right)} = \frac{\Pi_{t=1}^T \Pi_{k\in S_t}\left(\frac{\exp(\beta_k^*)}{\sum_{l\in S_t}\exp(\beta_l^*)}\right)^{z_{kt}}}{\Pi_{t=1}^T \Pi_{k\in S_t}\left(\frac{\exp(\tilde{\beta}_k)}{\sum_{l\in S_t}\exp(\tilde{\beta}_l)}\right)^{z_{kt}}}$$
$$= \Pi^T_{\substack{t=1: S_t\cap V_i^c \neq \phi \\ S_t \cap V_i \neq \phi}} \Pi_{k\in S_t\cap V_i^c}\left(\frac{\frac{\exp(\beta_k^*)}{\sum_{l\in S_t}\exp(\beta_l^*)}}{\frac{\exp(c+\beta_k^*)}{\sum_{m\in S_t\cap V_i}\exp(\beta_m^*)+\sum_{l\in S_t\cap V_i^c}\exp(c+\beta_l^*)}}\right)^{z_{kt}} < 1,$$

where the last strict inequality is due to the fact that the denominator is strictly increasing in c. Next, let $\bar{\boldsymbol{\beta}} = R(\tilde{\boldsymbol{\beta}}, \tilde{s})$. We have that $\bar{\boldsymbol{\beta}} \neq \tilde{\boldsymbol{\beta}}$ and following from Lemma 2 we obtain

$$\ell_{MNL}(\bar{\boldsymbol{\beta}}) = \ell_{MNL}(\tilde{\boldsymbol{\beta}}) > \ell_{MNL}(\boldsymbol{\beta}^*),$$

where $\bar{\boldsymbol{\beta}}$ is feasible to the market share constraint. However, from Lemma 1, we have that $\boldsymbol{\beta}^* \in \arg\max\left\{\boldsymbol{\beta} : \ell_{MNL} \text{ s.t. } \sum_{j=1}^n \exp(\beta_j) = \tilde{s}\right\}$ which is a contradiction. $\square$

Proof of Proposition 2. Let $\mathcal{S}$ be the set of assortments that can be generated with positive probability. Assume for a contradiction that the model is not identifiable. Following Lemma 2.2 in (Newey and McFadden 1994, Chapter 36), there exists $\boldsymbol{\beta}' \neq \boldsymbol{\beta}$ such that $\mathbb{P}_i(S;\boldsymbol{\beta}) = \mathbb{P}_i(S;\boldsymbol{\beta}')$ for all $i \in S$ and all $S \in \mathcal{S}$ that satisfy the market share constraint $\sum_{j=1}^n \exp(\beta_j) = \sum_{j=1}^n \exp(\beta'_j) = \tilde{s}$.

Assume that

$$\mathbb{P}_i(S;\boldsymbol{\beta}) = \mathbb{P}_i(S;\boldsymbol{\beta}') \text{ for all } i \in S, \text{ and for all } S \in \mathcal{S}$$

It now follows from the above expression that

$$\begin{gathered}\mathbb{P}_i(S;\boldsymbol{\beta})/\mathbb{P}_j(S;\boldsymbol{\beta}) = \mathbb{P}_i(S;\boldsymbol{\beta}')/\mathbb{P}_j(S;\boldsymbol{\beta}') \text{ for all } i,j \in S, \text{ and for all } S \in \mathcal{S}\\ \Leftrightarrow \exp(\beta_i - \beta_j) = \exp\left(\beta'_i - \beta'_j\right) \text{ for all } i,j \in S, \text{ and for all } S \in \mathcal{S}\\ \Leftrightarrow \beta_i - \beta_j = \beta'_i - \beta'_j \text{ for all } i,j \in S, \text{ and for all } S \in \mathcal{S}.\end{gathered}$$

However, since $\mathcal{S}$ generates a strongly connected graph with a positive probability, then the assortments overlap, giving

$$\beta_i - \beta_j = \beta'_i - \beta'_j \text{ for all } i,j = 1,\ldots,n. \tag{A4}$$

Since both $\boldsymbol{\beta}$ and $\boldsymbol{\beta}'$ satisfy the market share constraint, we obtain

$$\begin{gathered}\sum_{i=1}^n \exp(\beta_i) = \tilde{s},\\ \sum_{i=1}^n \exp(\beta_i - \beta_j) = \tilde{s}\exp(-\beta_j) \text{ for all } j = 1,\ldots,n, \text{ and}\\ \sum_{i=1}^n \exp(\beta'_i - \beta'_j) = \tilde{s}\exp(-\beta'_j) \text{ for all } j = 1,\ldots,n.\end{gathered}$$

Hence, it follows from (A4), that

$$\begin{gathered}\tilde{s}\exp(-\beta_j) = \tilde{s}\exp(-\beta'_j) \text{ for all } j = 1,\ldots,n\\ \Leftrightarrow \beta_j = \beta'_j \text{ for all } j = 1,\ldots,n.\end{gathered}$$

This implies that $\boldsymbol{\beta} = \boldsymbol{\beta}'$, which is a contradiction. □

Proof of Proposition 3. We start by proving the convergence of the loglikelihood function. Notice that by construction the following inequality holds,

$$\ell_{MNL}(\boldsymbol{\beta}^{(k_i)}) = g(\boldsymbol{\beta}^{(k_i)}|\boldsymbol{\beta}^{(k_i)}) \leq g(H(\boldsymbol{\beta}^{(k_i)})|\boldsymbol{\beta}^{(k_i)}) \leq \ell_{MNL}(H(\boldsymbol{\beta}^{(k_i)})) = \ell_{MNL}(M(\boldsymbol{\beta}^{(k_i)})) \leq \ell_{MNL}(\boldsymbol{\beta}^{(k_{i+1})}). \tag{A5}$$

The first equality follows from $g(\cdot|\boldsymbol{\beta}^{(k_i)})$ being the minorizer of $\ell_{MNL}(\cdot)$ with equality at $\boldsymbol{\beta}^{(k_i)}$. The maximizer of $g(\cdot|\boldsymbol{\beta}^{(k_i)})$ is $H(\boldsymbol{\beta}^{(k_i)})$, determining the next inequality. Next, $g(\cdot|\boldsymbol{\beta}^{(k_i)})$ is again a minorizer of $\ell_{MNL}(\cdot)$, whereas the last equality holds from Lemma 2. Finally, the last inequality follows from the fact that $M(\boldsymbol{\beta}^{(k_i)}) = \boldsymbol{\beta}^{(k_i+1)}$ and the loglikelihood is non-decreasing at each iteration. Hence, the sequence of the loglikelihood function generated by the MM algorithm is monotonic and is bounded above by zero and therefore is guaranteed to converge.

Next, denote by $\boldsymbol{\beta}^* = \lim_{i\to\infty} \boldsymbol{\beta}^{(k_i)}$ the limit point of a subsequence $\boldsymbol{\beta}^{(k_1)}, \boldsymbol{\beta}^{(k_2)}, \ldots$, then the result is obtained by taking the limit in (A5).

By taking the limit of the subsequence, we obtain that $\ell_{MNL}(M(\boldsymbol{\beta}^\star)) = \ell_{MNL}(\boldsymbol{\beta}^\star)$. This implies that $\boldsymbol{\beta}^\star$ is a stationary point that is feasible to the market share constraint since the differentiable minorizing function is tangent to the log-likelihood function at the current iterate. However, since $\ell_{MNL}(\cdot)$ is concave then it is a global optimal solution. Finally, the optimality with respect to $\ell_I(\boldsymbol{\beta}, \boldsymbol{\lambda})$ follows from Lemma 1.

Proof. of Proposition 4. Consider a sequence of MM iterations $\{\boldsymbol{\beta}^{(k)}\}$ re-scaled via $M(\cdot)$ such that they are feasible for the restricted market share problem. Since $G(\mathcal{N}, E)$ is strongly connected and $0 < s < 1$, it follows from Proposition 1 and Lemma 1 that $\ell_{MNL}(\boldsymbol{\beta})$ has a unique optimal solution which is a unique stationary point. From Proposition 3, we know that if there is a convergence point for the algorithm, in this case it must be the unique global optimal solution. So we are left to show that the algorithm indeed converges.

We first show an auxiliary result:

Lemma A1. *Given a strongly connected graph $G(\mathcal{N}, E)$ and $0 < \tilde{s} < \infty$, then, for every $c \in \mathbb{R}$, the set $\mathcal{A}(c) := \{\boldsymbol{\beta} : \ell_{MNL}(\boldsymbol{\beta}) \geq c, \sum_{i=1}^n \exp(\beta_i) = \tilde{s}\}$ is compact.*

Proof of Lemma A1 We have that

$$\mathcal{A}(c) = \{\boldsymbol{\beta} : \ell_{MNL}(\boldsymbol{\beta}) \geq c\} \cap \{\boldsymbol{\beta} : \sum_{i=1}^{n} \exp(\beta_i) \geq \tilde{s}\} \cap \{\boldsymbol{\beta} : \sum_{i=1}^{n} \exp(\beta_i) \leq \tilde{s}\}.$$

By the continuity of $\ell_{MNL}(\boldsymbol{\beta})$ and $\exp(\beta_i)$, we have that, for every $c \in \mathbb{R}$, the three sets on the right hand side are closed and so is their intersection. Hence, we are left to show that $\mathcal{A}(c)$ is bounded for any $c \in \mathbb{R}$. Assume for a contradiction that it is not, then there exists $\tilde{c} \in \mathbb{R}$ such that $\mathcal{A}(\tilde{c})$ is unbounded. Consequently, there exists a sequence $\{\boldsymbol{\beta}_i : i = 1, 2, \ldots\} \in \mathcal{A}(\tilde{c})$ with $||\boldsymbol{\beta}|| \to \infty$. However, we show that such a sequence cannot exist. We consider three different cases:

Case 1 ($\beta_i \to +\infty$ for some $i \in \mathcal{N}$): It is clear that such a sequence does not belong to $\mathcal{A}(\tilde{c})$ since it violates the market share constraint. In particular, we have

$$\lim_{||\boldsymbol{\beta}|| \to +\infty} \sum_{i=1}^{n} \exp(\beta_i) = +\infty \neq \tilde{s}.$$

Case 2 ($\beta_i \to -\infty$ for all $i \in \mathcal{N}$): Again this sequence violates the market share constraint where

$$\lim_{\beta_i \to -\infty} \sum_{i=1}^{n} \exp(\beta_i) = 0 \neq \tilde{s}.$$

We are left to eliminate the case when only a subset of the elements have $\beta \to -\infty$. Notice that such cases are not ruled out by Case 1 and 2. For this reason, define a new set $\underline{\mathcal{K}} \subsetneq \mathcal{N}$ where

$$\underline{\mathcal{K}} := \{i \in \mathcal{N} : \beta_i \to -\infty\}.$$

Case 3 ($\underline{\mathcal{K}} \neq \emptyset$): Let $\mathcal{K} := \mathcal{N} \backslash \underline{\mathcal{K}}$ where $\mathcal{K} \neq \emptyset$. Since $G(\mathcal{N}, E)$ is strongly connected, then there exists an edge that goes from $\mathcal{K}$ to $\underline{\mathcal{K}}$. Consider the set S_t that defines that edge. We have that $S_t \cap \mathcal{K} \neq \emptyset$, $S_t \cap \underline{\mathcal{K}} \neq \emptyset$, and $z_{lt} \geq 1$ for some item $l \in S_t \cap \underline{\mathcal{K}}$. Therefore, we get

$$\lim_{||\boldsymbol{\beta}|| \to \infty} \ell_{MNL}(\boldsymbol{\beta}) \leq \lim_{||\boldsymbol{\beta}|| \to \infty} \log\left(\frac{\exp(\beta_l)}{\sum_{i \in S_t \cap \mathcal{K}} \exp(\beta_i) + \sum_{k \in S_t \cap \underline{\mathcal{K}}} \exp(\beta_k)}\right) = -\infty,$$

where the inequality holds because the term in parenthesis is just one term of the likelihood function (1) and the equality is due to the fact that $\lim_{||\boldsymbol{\beta}||\to\infty}\sum_{i\in S_t\cap\mathcal{K}}\exp(\beta_i)>0$ and $\exp(\beta_l)\to 0$.

Hence, any sequence for which $||\boldsymbol{\beta}||\to\infty$, cannot belong to $\mathcal{A}(\tilde{c})$ which is a contradiction. Therefore, for every $c\in\mathbb{R}$, $\mathcal{A}(\tilde{c})$ is a closed and bounded set in $\mathbb{R}^n$ and hence compact. $\square$

From Lemma A1 the set $\mathcal{A}(c):=\{\boldsymbol{\beta}:\ell_{MNL}(\boldsymbol{\beta})\geq c,\sum_{i=1}^n\exp(\beta_i)=\tilde{s}\}$ is compact for every $c\in\mathbb{R}$. Therefore, by construction, the sequence of MM iterations $\{\boldsymbol{\beta}^{(k)}\}$ is bounded and has at least one convergent subsequence.

Now consider any convergent subsequence $\{\boldsymbol{\beta}^{(k_l)}\}$ and denote its limit point by $\mathcal{B}$. We have that the sequence of likelihood functions $\{\ell_{MNL}(\boldsymbol{\beta}^{(k_l)})\}$ is non-decreasing and bounded from above, hence it converges. In particular, we have

$$\lim_{l\to\infty}\ell_{MNL}(M(\boldsymbol{\beta}^{(k_l)}))=\lim_{l\to\infty}\ell_{MNL}(\boldsymbol{\beta}^{(k_l)}),$$

and by the continuity of $\ell_{MNL}(.)$ and $M(.)$ we get

$$\ell_{MNL}(M(\mathcal{B}))=\ell_{MNL}(\mathcal{B}).$$

As a result, $\mathcal{B}$ is a stationary point for ℓ_{MNL} and is feasible to the market share constraint. Therefore, $\mathcal{B}$ is the unique optimal solution to the constrained ℓ_{MNL} problem. Finally, since the subsequence $\{\boldsymbol{\beta}^{(k_l)}\}$ was chosen arbitrarily, then the whole sequence $\{\boldsymbol{\beta}^{(k)}\}$ converges to the unique optimal solution and the log-likelihood value is non-decreasing at each iteration. Finally, the fact that the MM estimates are optimal solutions to $\ell_I(\boldsymbol{\lambda},\boldsymbol{\beta})$ subject to the market share constraint follows from Lemma 1. $\square$

Appendix A2: Extension: Linear-in-parameters utility

We use the MM algorithm proposed in Jagabathula and Venkataraman (2020) to estimate the parameters of the MNL model in the presence of product features, and extend it to the censored data case. Specifically, we adapted their algorithm to estimate both $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$, as described in Figure A1.

Assume that there are $L\geq 1$ observable product attributes relevant for the utility specification. The utility for product $i\in\mathcal{N}$ can be written as

$$u_i=\beta_0+\sum_{l=1}^{L}\beta_l X_{il}+\epsilon_i,\quad\text{and}\quad u_0=\epsilon_0,$$

where u_0 is the utility of the no-purchase option whose deterministic part is normalized to 0 and ϵ_i are i.i.d draws from the standard Gumbel distribution.

We define $\tilde{q}_{it}(\boldsymbol{\beta}):=\mathbf{1}_{\{i\in S_t\}}\frac{\exp(\sum_{l=1}^{L}\beta_l X_{il})}{\sum_{j\in S_t}\exp(\sum_{l=1}^{L}\beta_l X_{jl})}$. The iteration map for the parameters $\boldsymbol{\beta}$ tries to match the observed and the predicted market share for each feature. In order for the algorithm to converge, the matrix of features $\boldsymbol{X}$ should be normalized such that the L_1 norm for the features of each product are less than 1. One approach is to normalize each feature by the number of features multiplied by the maximum value of the feature across the different products.

Figure A1 MM Algorithm for the joint estimation problem with covariates.

```
• Input: Transactional data $\{(\boldsymbol{z_t}, \boldsymbol{X_t}, S_t)\}_{t=1}^{T}$, market share $s$
• Set $\tilde{s} \leftarrow s/(1-s)$,
• Initialize $\boldsymbol{\beta}$,
• While Stopping criteria are not met,
    ○ $\tilde{q}_{it}(\boldsymbol{\beta}) \leftarrow \mathbf{1}_{\{i \in S_t\}} \frac{\exp(\sum_{l=1}^{L} \beta_l X_{il})}{\sum_{j \in S_t} \exp(\sum_{l=1}^{L} \beta_l X_{jl})}$   for $i \in \mathcal{N}$, and $t = 1, \ldots, T$,
    ○ $\beta_l \leftarrow \beta_l + \log\left(\frac{\sum_{j=1}^{n} K_j X_{jl}}{\sum_{t=1}^{T} m_t \sum_{j=1}^{n} X_{jl} \tilde{q}_{jt}(\boldsymbol{\beta})}\right)$   for $l = 1, \ldots, L$.
• End While
• $\tilde{s}_1 \quad \sum_{j=1}^{n} \exp(\sum_{l=1}^{L} \beta_l X_{jl})$,
• $\beta_0 \leftarrow \beta_0 + \log\left(\frac{\tilde{s}}{\tilde{s}_1}\right)$,
• Set: $\lambda_t \leftarrow m_t \frac{1 + \sum_{j \in S_t} \exp(\beta_0 + \sum_{l=1}^{L} \beta_l X_{jl})}{\sum_{j \in S_t} \exp(\beta_0 + \sum_{l=1}^{L} \beta_l X_{il})}$   for $t = 1, \ldots, T$.
• End
```

Note

Online Appendix B could be found in the following link.

References

Hunter, D. 2004. MM algorithms for generalized Bradley-Terry models. *Annals of Statistics* 384–406

Newey, W., D. McFadden. 1994. Large sample estimation and hypothesis testing. *Handbook of Econometrics* **4** 2111–2245

Jagabathula, S., A. Venkataraman. 2020. An MM algorithm for estimating the MNL model with product features. *Available at SSRN: https://ssrn.com/abstract=3733971*