Online Supplement for

Estimating Primary Demand for a Heterogeneous-Groups Product Category under Hierarchical Consumer Choice Model

Haengju Lee and Yongsoon Eun

Daegu Gyeongbuk Institute of Science and Technology

333 Techno Jungang-daero, Hyeonpung-myeon, Dalseong-gun, Daegu, 711-873, Republic of Korea

July 3, 2015

A Heterogeneous Choice Hierarchy

An arriving consumer follows the brand-primary hierarchy with probability q^B and the type-primary hierarchy with probability $q^T = 1 - q^B$. We let the utility of a consumer following the brand-primary be

$$U_{ij}^B = u_{ij}^B + \epsilon_i^B + \epsilon_{ij}^B$$

for the purchase of a product with brand i and type j . Likewise, we let the corresponding utility of a consumer following the type-primary be

$$U_{ij}^T = u_{ij}^T + \epsilon_j^T + \epsilon_{ij}^T.$$

Let μ^B be the inter-brand similarity, μ^T be the inter-type similarity, and $\boldsymbol{\mu} = (\mu^B, \mu^T)$. The preference of product with brand i and type j is defined by $v_{ij}^B = e^{u_{ij}^B}$ for the brand-primary process and $v_{ij}^T = e^{u_{ij}^T}$ for the type-primary process. The preference vectors are defined by $\mathbf{v}^B = (v_{ij}^B, \forall j \in B^{B=i}, i = 1, \dots, K^B)$, $\mathbf{v}^T = (v_{ij}^T, \forall j \in B^{T=i}, i = 1, \dots, K^T)$, and $\mathbf{v} = (\mathbf{v}^B, \mathbf{v}^T)$. We denote the number of brands by K^B and the number of types by K^T . $S_t^{B=i}$ is the set of available types of brand i , and $S_t^{T=j}$ is the set of available brands of type j in period t . $B^{B=i}$ is the set of all types of brand i and $B^{T=j}$ is the set of all brands of type j .

If a consumer is in the brand-primary process, the probability of no-purchase in period t is $P_0^B(\mathbf{S}_t; \mathbf{v}^B, \mu^B) = \frac{1}{\sum_{m=1}^{K^B} (\sum_{l \in S_t^{B=m}} v_{ml}^B)^{\mu^B} + 1}$ and the probability of purchasing a product with brand

i and type j is $P_{ij}^B(\mathbf{S}_t; \mathbf{v}^B, \mu^B) = \frac{v_{ij}^B \cdot (\sum_{l \in S_t^{B=i}} v_{il}^B)^{\mu^B - 1}}{\sum_{m=1}^{K^B} (\sum_{l \in S_t^{B=m}} v_{ml}^B)^{\mu^B} + 1}$. If he is in the type-primary process, the

probability of no-purchase in period t is $P_0^T(\mathbf{S}_t; \mathbf{v}^T, \mu^T) = \frac{1}{\sum_{m=1}^{K^T} (\sum_{l \in S_t^{T=m}} v_{lm}^T)^{\mu^T} + 1}$ and the proba-

bility of purchasing a product with brand i and type j is $P_{ij}^T(\mathbf{S}_t; \mathbf{v}^T, \mu^T) = \frac{v_{ij}^T \cdot (\sum_{l \in S_t^{T=j}} v_{lj}^T)^{\mu^T - 1}}{\sum_{m=1}^{K^T} (\sum_{l \in S_t^{T=m}} v_{lm}^T)^{\mu^T} + 1}$.

The primary demand of brand i and type j , denoted by X_{ijt} , can be further broken down into two terms in the mixture model: the primary demand from the brand-primary process, denoted by X_{ijt}^B , and that from the type-primary process, denoted by X_{ijt}^T . Therefore,

$$X_{ijt} = X_{ijt}^B + X_{ijt}^T.$$

We can also write the primary demand of no-purchase such as

$$X_{0t} = X_{0t}^B + X_{0t}^T.$$

If the values of primary demands X_{ijt}^B , X_{ijt}^T , X_{0t}^B , and X_{0t}^T are available, the likelihood function based on the primary demands is written by

$$\begin{aligned} LL(\mathbf{v}, \boldsymbol{\mu}, q^B) = & \sum_{i=1}^{K^B} \sum_{j \in B^{B=i}} N_{ij}^B \ln P_{ij}^B(\mathbf{B}; \mathbf{v}^B, \mu^B) + N_{ij}^T \ln P_{ij}^T(\mathbf{B}; \mathbf{v}^T, \mu^T) \\ & + N_0^B \ln P_0^B(\mathbf{B}; \mathbf{v}^B, \mu^B) + N_0^T \ln P_0^T(\mathbf{B}; \mathbf{v}^T, \mu^T) \\ & + \left(\sum_{i=1}^{K^B} \sum_{j \in B^{B=i}} N_{ij}^B + N_0^B \right) \cdot \ln q_B + \left(\sum_{j=1}^{K^T} \sum_{i \in B^T=j} N_{ij}^B + N_0^B \right) \cdot \ln(1 - q_B), \quad (\text{A.1}) \end{aligned}$$

where $N_{ij}^B = \sum_{t=1}^T X_{ijt}^B$, $N_{ij}^T = \sum_{t=1}^T X_{ijt}^T$, $N_0^B = \sum_{t=1}^T X_{0t}^B$, and $N_0^T = \sum_{t=1}^T X_{0t}^T$.

A.1 E-Step for Mixture Model

Applying the similar notations defined in Section 4.1, we can compute the conditional expected value of the primary demand of brand i and type j in period t . If a product of brand i and type j is not available (i.e., $j \in B^{B=i} \setminus S_t^{B=i}$), we have $\hat{X}_{ijt}^B = \hat{\lambda}_t \cdot \hat{q}^B \cdot P_{ij}^B(\mathbf{B}; \hat{\mathbf{v}}^B, \hat{\mu}^B)$, $\hat{X}_{ijt}^T = \hat{\lambda}_t \cdot (1 - \hat{q}^B) \cdot P_{ij}^T(\mathbf{B}; \hat{\mathbf{v}}^T, \hat{\mu}^T)$, and $\sum_{i=1}^{K^B} \sum_{j \in S_t^{B=i}} z_{ijt} = \hat{\lambda}_t \cdot (1 - P_0(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B))$, where $P_0(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B) \equiv \hat{q}^B \cdot P_0^B(\mathbf{S}_t; \hat{\mathbf{v}}^B, \hat{\mu}^B) + (1 - \hat{q}^B) \cdot P_0^T(\mathbf{S}_t; \hat{\mathbf{v}}^T, \hat{\mu}^T)$. Combining the equations, we obtain

$$\hat{X}_{ijt}^B = \frac{\hat{q}^B P_{ij}^B(\mathbf{B}; \hat{\mathbf{v}}^B, \hat{\mu}^B)}{1 - P_0(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B)} \sum_{i=1}^{K^B} \sum_{j \in S_t^{B=i}} z_{ijt}$$

and

$$\hat{X}_{ijt}^T = \frac{(1 - \hat{q}^B) \cdot P_{ij}^T(\mathbf{B}; \hat{\mathbf{v}}^T, \hat{\mu}^T)}{1 - P_0(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B)} \sum_{i=1}^{K^B} \sum_{j \in S_t^{B=i}} z_{ijt}.$$

Next, if a product of brand i and type j is available (i.e., $j \in S_t^{B=i}$), we have $\hat{X}_{ijt}^B = \hat{\lambda}_t \cdot \hat{q}^B \cdot P_{ij}^B(\mathbf{B}; \hat{\mathbf{v}}^B, \hat{\mu}^B)$, $\hat{X}_{ijt}^T = \hat{\lambda}_t \cdot (1 - \hat{q}^B) \cdot P_{ij}^T(\mathbf{B}; \hat{\mathbf{v}}^T, \hat{\mu}^T)$, and $z_{ijt} = \hat{\lambda}_t \cdot P_{ij}(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B)$, where $P_{ij}(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B) \equiv \hat{q}^B \cdot P_{ij}^B(\mathbf{S}_t; \hat{\mathbf{v}}^B, \hat{\mu}^B) + (1 - \hat{q}^B) \cdot P_{ij}^T(\mathbf{S}_t; \hat{\mathbf{v}}^T, \hat{\mu}^T)$. From these equations, we have

$$\hat{X}_{ijt}^B = \frac{\hat{q}^B \cdot P_{ij}^B(\mathbf{B}; \hat{\mathbf{v}}^B, \hat{\mu}^B)}{P_{ij}(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B)} \cdot z_{ijt}$$

and

$$\hat{X}_{ijt}^T = \frac{(1 - \hat{q}^B) \cdot P_{ij}^T(\mathbf{B}; \hat{\mathbf{v}}^T, \hat{\mu}^T)}{P_{ij}(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B)} \cdot z_{ijt}.$$

Finally, the conditional expected value of the primary demand of no-purchase can be derived from $\hat{X}_{0t}^B = \hat{\lambda}_t \cdot \hat{q}_B \cdot P_0^B(\mathbf{B}; \hat{\mathbf{v}}^B, \hat{\mu}^B)$, $\hat{X}_{0t} = \hat{\lambda}_t \cdot P_0(\mathbf{B}; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B)$, $\hat{X}_{0t} + \sum_{i=1}^{K^B} \hat{X}_{ij} = \hat{\lambda}_t$, and $s = 1 - P_0(\mathbf{B}; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B)$, where s is the market share. We have

$$\hat{X}_{0t}^B = \frac{\hat{q}^B \cdot P_0^B(\mathbf{B}; \hat{\mathbf{v}}^B, \hat{\mu}^B)}{P_0(\mathbf{B}; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B)} \cdot \hat{X}_{0t},$$

and

$$\hat{X}_{0t}^T = \hat{X}_{0t} - \hat{X}_{0t}^B,$$

where $\hat{X}_{0t} = \frac{1-s}{s} \sum_{i=1}^{K^B} \sum_{j \in B^{B=i}} \hat{X}_{ijt}$. We summarize the results in the following lemma.

Lemma A.1. *The conditional expected value of the primary demand of product from brand i and type j in period t following the brand-primary process is*

$$\hat{X}_{ijt}^B = \begin{cases} \frac{\hat{q}^B \cdot P_{ij}^B(\mathbf{B}; \hat{\mathbf{v}}^B, \hat{\mu}^B)}{1 - P_0(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B)} \sum_{m=1}^{K^B} \sum_{l \in S_t^{B=m}} z_{mlt}, & \text{for all } j \in B^{B=i} \setminus S_t^{B=i} \\ \frac{\hat{q}^B \cdot P_{ij}^B(\mathbf{B}; \hat{\mathbf{v}}^B, \hat{\mu}^B)}{P_{ij}(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B)} \cdot z_{ijt}, & \text{for all } j \in S_t^{B=i} \end{cases}$$

and that following the type-primary process is

$$\hat{X}_{ijt}^T = \hat{X}_{ijt}^B \cdot \frac{(1 - \hat{q}^B) \cdot P_{ij}^T(\mathbf{B}; \hat{\mathbf{v}}^T, \hat{\mu}^T)}{\hat{q}^B \cdot P_{ij}^B(\mathbf{B}; \hat{\mathbf{v}}^B, \hat{\mu}^B)}.$$

The conditional expected primary demand of no-purchase following the brand primary process is

$$\hat{X}_{0t}^B = \frac{\hat{q}^B \cdot P_0^B(\mathbf{B}; \hat{\mathbf{v}}^B, \hat{\mu}^B)}{P_0(\mathbf{B}; \hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B)} \cdot \hat{X}_{0t}$$

and that following the type primary process is

$$\hat{X}_{0t}^T = \hat{X}_{0t} - \hat{X}_{0t}^B,$$

where $\hat{X}_{0t} = \frac{1-s}{s} \sum_{i=1}^{K^B} \sum_{j \in B^{B=i}} \hat{X}_{ijt}$ and s is the exogenously given market share of the store.

The conditional expected log-likelihood in Equation (A.1) for a given parameter set $(\hat{\mathbf{v}}, \hat{\mu}, \hat{q}^B)$ is

$$\hat{LL}(\mathbf{v}, \mu, q^B) = \hat{LL}^B(\mathbf{v}^B, \mu^B) + \hat{LL}^T(\mathbf{v}^T, \mu^T) + \hat{LL}^q(q^B), \quad (\text{A.2})$$

where

$$\begin{aligned} \hat{LL}^B(\mathbf{v}^B, \mu^B) &= \sum_{i=1}^{K^B} \sum_{j \in B^{B=i}} \hat{N}_{ij}^B \cdot \ln v_{ij}^B + (\mu^B - 1) \sum_{i=1}^{K^B} \left(\sum_{j \in B^{B=i}} \hat{N}_{ij}^B \right) \cdot \ln \sum_{j \in B^{B=i}} v_{ij}^B \\ &\quad - \left(\sum_{i=1}^{K^B} \sum_{j \in B^{B=i}} \hat{N}_{ij}^B + \hat{N}_0^B \right) \cdot \ln \left(\sum_{i=1}^{K^B} \left(\sum_{j \in B^{B=i}} v_{ij}^B \right)^{\mu^B} + 1 \right), \end{aligned} \quad (\text{A.3})$$

$$\begin{aligned} \hat{LL}^T(\mathbf{v}^T, \mu^T) &= \sum_{j=1}^{K^T} \sum_{i \in B^{T=j}} \hat{N}_{ij}^T \cdot \ln v_{ij}^T + (\mu^T - 1) \sum_{j=1}^{K^T} \left(\sum_{i \in B^{T=j}} \hat{N}_{ij}^T \right) \cdot \ln \sum_{i \in B^{T=j}} v_{ij}^T \\ &\quad - \left(\sum_{j=1}^{K^T} \sum_{i \in B^{T=j}} \hat{N}_{ij}^T + \hat{N}_0^T \right) \cdot \ln \left(\sum_{j=1}^{K^T} \left(\sum_{i \in B^{T=j}} v_{ij}^T \right)^{\mu^T} + 1 \right), \end{aligned} \quad (\text{A.4})$$

and

$$\hat{LL}^q(q^B) = \left(\sum_{i=1}^{K^B} \sum_{j \in B^{B=i}} \hat{N}_{ij}^B + \hat{N}_0^B \right) \cdot \ln q_B + \left(\sum_{j=1}^{K^T} \sum_{i \in B^{T=j}} \hat{N}_{ij}^T + \hat{N}_0^T \right) \cdot \ln(1 - q_B). \quad (\text{A.5})$$

A.2 M-Step for Mixture Model

In the M-step, we find estimators maximizing the conditional expected log-likelihood given in Equation (A.2). Using Proposition 4.1, we find the unique optimal value of preference weight for a given $\boldsymbol{\mu} = (\mu^B, \mu^T)$. The optimal preference weight of a product with brand i and type j , v_{ij}^B , in the brand-primary process maximizes $\hat{L}^B(\mathbf{v}^B, \mu^B)$ in Equation (A.3). Similarly, the optimal value of that in the type-primary process maximizes $\hat{L}^T(\mathbf{v}^T, \mu^T)$ in Equation (A.4). Finally, the optimal value of q^B maximizes $\hat{L}^q(q^B)$. We state the result in the following proposition (we skip the proof because it can be easily derived by using Proposition 4.1).

Proposition A.1. *For a given vector of $\boldsymbol{\mu} = (\mu^B, \mu^T)$, the conditional expected log-likelihood function in the E-step is unimodal, and its unique optimal solution is*

$$(v_{ij}^B)^* = \frac{\hat{N}_{ij}^B}{\sum_{l \in B^{B=i}} \hat{N}_{il}^B} \cdot \left(\frac{\sum_{l \in B^{B=i}} \hat{N}_{il}^B}{\hat{N}_0^B} \right)^{1/\mu^B}, \forall j \in B^{B=i}, i = 1, \dots, K^B.$$

$$(v_{ij}^T)^* = \frac{\hat{N}_{ij}^T}{\sum_{l \in B^{T=j}} \hat{N}_{lj}^T} \cdot \left(\frac{\sum_{l \in B^{T=j}} \hat{N}_{lj}^T}{\hat{N}_0^T} \right)^{1/\mu^T}, \forall j \in B^{B=i}, i = 1, \dots, K^B,$$

$$(q^B)^* = \frac{\sum_{m=1}^{K^B} \sum_{l \in B^{B=m}} \hat{N}_{ml}^B + \hat{N}_0^B}{\sum_{m=1}^{K^B} \sum_{l \in B^{B=m}} \hat{N}_{ml} + \hat{N}_0},$$

A.3 Numerical Example for Mixture Model

Assume that there are two brands (brand A and brand B) and two types (type 1 and type 2). An arriving consumer is in the brand-primary process with probability $q^B = 0.5$ and the type-primary process with probability $1 - q^B = 0.5$. The parameters are $(v_{A1}^B, v_{A2}^B, v_{B1}^B, v_{B2}^B) = (1, 0.5, 1, 0.5)$, $\mu^B = 0.3$ for the brand-primary process and $(v_{A1}^T, v_{A2}^T, v_{B1}^T, v_{B2}^T) = (1, 1, 0.5, 0.5)$, $\mu^T = 0.3$ for the type-primary process. We assume a homogeneous arrival rate $\lambda = 50$ over $T = 15$. Table A.1 shows the simulated data with these demand parameters.

We first run the algorithm of EM_MNL on the simulated data set. The incomplete data log-likelihood value is -124.1033. Next, we run the algorithm of EM_NMNL. The incomplete data log-likelihood value is -123.9187 under the brand-primary model and -122.1895 under the type-primary model. Hence, it is predicted that the data set follows the type-primary model. Finally, we run the EM algorithm based on the mixture hierarchy model, denoted by EM_MIXTURE. The incomplete data log-likelihood value is -120.4523. The number of parameters is 19 for EM_MNL, 20 for EM_NMNL, and 26 for EM_MIXTURE. The likelihood ratio index between EM_MNL and EM_NMNL is 3.8274. This is less than the five percent quantile of the chi-squared value (the degree of freedom is 1) of 3.8415. We next compute the likelihood ratio index between EM_MNL and EM_MIXTURE. The value is 7.3020 that is less than the five percent quantile of the chi-squared value (the degree of freedom is 7) of 14.0671. Therefore, the likelihood ratio test supports the MNL model. We also consider AIC and BIC. The number of observations is $|data| = 462$. The AIC is 286.2066 for EM_MNL, 284.3792 for EM_NMNL, and 292.9046 for EM_MIXTURE. The BIC is 364.7823 for EM_MNL, 367.0905 for EM_NMNL, and 400.4293 for EM_MIXTURE. AIC supports

Table A.1: Simulated purchases and no-purchases over 15 periods for a product category with two brands and two types in the heterogeneous choice hierarchy model, brand-primary process with 0.5 probability and type-primary process with 0.5 probability (NA denotes non-availability)

Observable data (purchases)																
Brand	Type	Periods														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	1	10	10	5	11	13	15	15	NA	NA	10	18	12	12	10	6
	2	12	6	5	9	7	9	6	13	9	NA	NA	11	5	11	12
B	1	5	6	9	5	10	10	4	11	13	11	12	NA	NA	7	3
	2	7	3	6	8	5	6	2	5	4	7	16	10	15	NA	NA
Non-observable data																
No-purchases		14	16	18	15	13	8	14	17	12	11	21	21	22	17	18
λ		48	41	43	48	48	48	41	46	38	39	67	54	54	45	39

the NMNL model with the type-primary process, but BIC supports the MNL model. In general, we can see that the MNL model is good enough for this data set. Because an arriving customer follows the brand-primary process with 0.5 probability and the type-primary process with 0.5 probability, the resulting substitution pattern has no distinct group effect.

B NMNL Model with Higher Dimensions

We consider the hierarchy with three stages by adding a third dimension. In the first stage, consumers select a product group to purchase or no-purchase. If they choose product group k ($k \neq 0$), they next select a subgroup m to purchase. Finally, they select product j to purchase from available products, denoted by S_{kmt} . In this case, the utility of a consumer is (Ben-Akiva and Lerman [?])

$$U_{kmj} = u_{kmj} + \epsilon_k + \epsilon_{km} + \epsilon_{kmj}.$$

ϵ_{kmj} are independent and identically distributed zero-mean Gumbel random variables with scale parameter 1. ϵ_{km} are distributed so that $\max_{i \in B_{km}} U_{kmi}$ are Gumbel distributed with scale parameter μ_2 , and ϵ_k are distributed so that $\max_{n \in B_k, i \in B_{kn}} U_{kni}$ are Gumbel distributed with scale parameter μ_1 where $\mu_1 \leq \mu_2 \leq 1$. We denote $\boldsymbol{\mu} = (\mu_1, \mu_2)$. The probability of purchasing product j from product subgroup m (in the second stage) and product group k (in the first stage) is given by the product of three conditional probabilities, each of which is a logit model (Ben-Akiva and Lerman [?]).

Lemma B.1. (Ben-Akiva and Lerman [?]) *The probability of no-purchase is*

$$P_0(\mathbf{B}; \mathbf{v}, \boldsymbol{\mu}) = \frac{1}{\sum_{l=1}^K (\sum_{n \in B_l} (\sum_{i \in B_{ln}} v_{lni})^{\mu_2})^{\mu_1/\mu_2} + 1}.$$

The probability of purchasing product j from product subgroup m and product group k is given by

$$P_{kmj}(\mathbf{B}; \mathbf{v}, \boldsymbol{\mu}) = Pr(j|km) \cdot Pr(m|k) \cdot Pr(k),$$

where

$$Pr(j|km) = \frac{v_{kmj}}{\sum_{i \in B_{km}} v_{kmi}},$$

$$Pr(m|k) = \frac{(\sum_{i \in B_{km}} v_{kmi})^{\mu_2}}{\sum_{n \in B_k} (\sum_{i \in B_{kn}} v_{kni})^{\mu_2}},$$

$$Pr(k) = \frac{(\sum_{n \in B_k} (\sum_{i \in B_{kn}} v_{kni})^{\mu_2})^{\mu_1/\mu_2}}{\sum_{l=1}^K (\sum_{n \in B_l} (\sum_{i \in B_{ln}} v_{lni})^{\mu_2})^{\mu_1/\mu_2} + 1}.$$

The E-step of the EM algorithm can be easily extended using Lemma 3.1. Hence, we only focus on the optimal solution in the M-step of the EM algorithm. The conditional expected log-likelihood corresponding to Equation (7) is now

$$\begin{aligned} \hat{LL}(\mathbf{v}, \mu) &= \sum_{l=1}^K \sum_{n \in B_l} \sum_{i \in B_{ln}} \hat{N}_{lni} \cdot \ln v_{lni} + (\mu_2 - 1) \sum_{l=1}^K \sum_{n \in B_l} \left(\sum_{i \in B_{ln}} \hat{N}_{lni} \right) \cdot \ln \sum_{i \in B_{ln}} v_{lni} \\ &\quad + \left(\frac{\mu_1}{\mu_2} - 1 \right) \cdot \sum_{l=1}^K \left(\sum_{n \in B_l} \sum_{i \in B_{ln}} \hat{N}_{lni} \right) \cdot \ln \sum_{n \in B_l} \left(\sum_{i \in B_{ln}} v_{lni} \right)^{\mu_2} \\ &\quad - \frac{\hat{N}_0}{1-s} \cdot \ln \left(\sum_{l=1}^K \left(\sum_{n \in B_l} \left(\sum_{i \in B_{ln}} v_{lni} \right)^{\mu_2} \right)^{\mu_1/\mu_2} + 1 \right). \end{aligned} \quad (\text{B.1})$$

We can obtain the closed-form solution for the preference vector, which is a function of μ_1 and μ_2 , if we consider both μ_1 and μ_2 are constant in calculating the critical solution of Equation (B.1). We state the result in the following proposition.

Proposition B.1. *For given (μ_1, μ_2) , the optimal solution of the expected log-likelihood function in the E-step is*

$$v_{kmj}^* = \frac{\hat{N}_{kmj}}{\sum_{i \in B_{km}} \hat{N}_{kmi}} \cdot \left(\frac{\sum_{i \in B_{km}} \hat{N}_{kmi}}{\sum_{n \in B_k} \sum_{i \in B_{kn}} \hat{N}_{kni}} \right)^{1/\mu_2} \cdot \left(\frac{\sum_{n \in B_k} \sum_{i \in B_{kn}} \hat{N}_{kni}}{\hat{N}_0} \right)^{1/\mu_1}.$$

By comparing the result for the two-stage choice hierarchy in Proposition 3.1 with that of the above proposition, we can extend the result to the choice hierarchy with any number of stages to have the optimal solution in the M-step. The optimal solution for the preference weight of the product with n -stage choice hierarchy is the multiplication of n terms. The m th term of the multiplication corresponds to m th subgroup in the choice hierarchy and its base is the ratio of the total primary demand of its subgroup and the total primary demand of its parent's group (let the parent's group be the subgroup in the right above in the choice hierarchy) and its exponent is $1/\mu_m$, where $\mu_1 \leq \mu_2 \leq \dots \leq \mu_{n-1} \leq 1$. Note that the first term of the multiplication corresponds to the first (or the highest) stage and its base is the ratio of the total primary demand of its group and the total primary demand of no-purchase.

B.1 Numerical Example for NMNL with Three Stages

Assume that a product category can be grouped by brand, color, and type. There are two brands (brand A and brand B), two colors (color a and color b) and 2 types (type 1 and type 2). An arriving consumer chooses a product in the order of brand, color, and type. The parameters are $(v_{Aa1}, v_{Aa2}, v_{Ab1}, v_{Ab2}, v_{Ba1}, v_{Ba2}, v_{Bb1}, v_{Bb2}) = (1, 0.5, 1, 0.5, 1, 0.5, 1, 0.5)$, $\mu_1 = 0.3$, and $\mu_2 = 0.5$. We assume a homogeneous arrival rate $\lambda = 100$ over $T = 15$, Table B.1 shows the simulated data with the demand parameters.

Table B.1: Simulated purchases and no-purchases over 15 periods for a product category with two brands, two colors, and two types under the three-stage choice hierarchy in the order of brand, color, type (NA denotes non-availability)

Observable data (purchases)																	
Brand	Color	Type	Periods														
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	a	1	16	15	6	11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
		2	7	7	3	7	14	13	10	11	13	19	16	13	17	19	17
	b	1	10	11	10	14	14	16	16	NA	NA	NA	NA	NA	NA	NA	NA
		2	4	3	8	5	7	9	3	13	19	10	14	15	25	15	27
B	a	1	9	7	11	11	15	10	11	17	22	14	NA	NA	NA	NA	NA
		2	8	9	8	7	5	6	7	10	10	9	12	7	21	16	17
	b	1	14	9	15	13	9	17	7	12	18	10	14	20	20	NA	NA
		2	8	7	4	7	7	7	14	5	8	3	6	6	10	26	12
Non-observable data																	
No-purchases			22	20	25	22	27	20	19	26	23	18	27	24	32	30	33
λ			98	88	90	97	98	98	87	94	113	83	89	85	125	106	106

We first run the algorithm of EM_MNL on the simulated data set. The incomplete data log-likelihood value is -241.5788. Next, we run the EM_NMNL (assuming $\mu_2 = 1$). The incomplete data log-likelihood value is -235.5245. Finally, we run the EM algorithm based on the hierarchy with three stages, denoted by EM_3NMNL. The incomplete data log-likelihood value is -231.5799. The number of parameters is 23 for EM_MNL, 24 for EM_NMNL, and 25 for EM_3NMNL. The likelihood ratio index between EM_MNL and EM_NMNL is 12.1086. This is greater than the five percent quantile of the chi-squared value (the degree of freedom is 1) of 3.8415. We next compute the likelihood ratio index between EM_NMNL and EM_3NMNL. The value is 7.8892 that is greater than 3.8415. Therefore, the likelihood ratio supports the NMNL model with three stages. We consider AIC and BIC. The number of observations is $|data| = 1089$. The value of AIC is 529.1576 for EM_MNL, 519.0490 for EM_NMNL, and 513.1598 for EM_3NMNL. The value of BIC is 643.9969 for EM_MNL, 638.8814 for EM_NMNL, and 637.9852 for EM_3NMNL. Therefore, both AIC and BIC support the NMNL model with three stages.

C Second Derivative of Incomplete Data Log-likelihood

The incomplete data log-likelihood function is given in Equation (3).

$$\log \mathcal{L} = \sum_{t=1}^T \left[m_t \log \lambda_t - \lambda_t \cdot (1 - P_0(\mathbf{S}_t; \mathbf{v}, \mu)) - \sum_{k=1}^K \sum_{j \in S_{kt}} (\log(z_{kjt}!) - z_{kjt} \cdot \log P_{kj}(\mathbf{S}_t; \mathbf{v}, \mu)) \right].$$

For the notational simplicity, we write $P_0(\mathbf{S}_t; \mathbf{v}, \mu)$ and $P_{kj}(\mathbf{S}_t; \mathbf{v}, \mu)$ by P_{0t} and P_{kjt} respectively. We also denote the equation in the square bracket as $\log \mathcal{L}_t$ so that we have $\log \mathcal{L} = \sum_{t=1}^T \log \mathcal{L}_t$. All the second derivative of $\log \mathcal{L}_t$ can be obtained by summing up $\log \mathcal{L}_t$ over all the periods. The

second derivative of $\log \mathcal{L}_t$ with respect to parameters are given by

$$\begin{aligned} \frac{\partial^2 \log \mathcal{L}_t}{\partial \mu^2} &= (2\lambda_t P_{0t} + m_t) P_{0t}^2 \left(\sum_{k=1}^K \left(\sum_{j \in S_{kt}} v_{kj} \right)^\mu \cdot \log \sum_{j \in S_{kt}} v_{kj} \right)^2 \\ &\quad - (\lambda_t P_{0t} + m_t) P_{0t} \sum_{k=1}^K \left(\sum_{j \in S_{kt}} v_{kj} \right)^\mu \cdot \left(\log \sum_{j \in S_{kt}} v_{kj} \right)^2, \\ \frac{\partial^2 \log \mathcal{L}_t}{\partial v_{kj} \partial v_{kl}} &= \begin{cases} (2\lambda_t P_{0t} + m_t) \left(\frac{\mu P_{kjt}}{v_{kj}} \right)^2 - (\lambda_t P_{0t} + m_t) \frac{\mu(\mu-1) P_{kjt}}{v_{kj} \sum_{l \in S_{kt}} v_{kl}} + \frac{(1-\mu) \sum_{l \in S_{kt}} z_{klt}}{(\sum_{l \in S_{kt}} v_{kl})^2} & \text{if } j, l \in S_{kt} \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

$$\frac{\partial^2 \log \mathcal{L}_t}{\partial v_{kj}^2} = \begin{cases} \frac{\partial^2 \log \mathcal{L}}{\partial v_{kj} \partial v_{kl}} - \frac{z_{kjt}}{v_{kj}^2} & \text{if } j \in S_{kt} \\ 0 & \text{otherwise,} \end{cases}$$

$$\frac{\partial^2 \log \mathcal{L}_t}{\partial \mu v_{kj}} = \begin{cases} (2\lambda_t P_{0t} + m_t) P_{0t} \frac{\mu P_{kjt}}{v_{kj}} \sum_{m=1}^K (\sum_{l \in S_{ml}} v_{ml})^\mu \log \sum_{l \in S_{ml}} v_{ml} \\ \quad - (\lambda_t P_{0t} + m_t) \frac{P_{kjt}}{v_{kj}} (\mu \log \sum_{l \in S_{ml}} v_{kl} + 1) + \frac{\sum_{l \in S_{kt}} z_{klt}}{\sum_{l \in S_{kt}} v_{kl}} & \text{if } j \in S_{tk} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\frac{\partial^2 \log \mathcal{L}_t}{\partial v_{kj} \partial v_{ml}} = \begin{cases} (2\lambda_t + m_t) \frac{\mu^2 P_{kjt} P_{mlt}}{v_{kj} v_{ml}} & \text{if } j \in S_{kt}, l \in S_{mt} \\ 0 & \text{otherwise.} \end{cases}$$

D Proofs

Proof of Lemma 3.1

See Chapter 10 of Ben-Akiva and Lerman [?].

Proof of Lemma 3.2

Let $\mathbf{S}'_t$ be same as $\mathbf{S}_t$ except that product h from product group k is unavailable in $\mathbf{S}'_t$ while it is available in $\mathbf{S}_t$. Now we investigate the changes of purchase probabilities of products from product group k and products from product group $m (\neq k)$. The ratio of purchase probabilities of product j from product group k under $\mathbf{S}'_t$ and $\mathbf{S}_t$ is

$$\frac{P_{kj}(\mathbf{S}'_t; \mathbf{v}, \mu)}{P_{kj}(\mathbf{S}_t; \mathbf{v}, \mu)} = \frac{\sum_{n=1}^K (\sum_{p \in S_{nt}} v_{np})^\mu + 1}{\sum_{n=1}^K (\sum_{p \in S'_{nt}} v_{np})^\mu + 1} \cdot \left(\frac{\sum_{i \in S_{ki}} v_{ki} - v_{kh}}{\sum_{i \in S_{ki}} v_{ki}} \right)^{\mu-1},$$

and that of product l from a different product group $m (\neq k)$ is

$$\frac{P_{ml}(\mathbf{S}'_t; \mathbf{v}, \mu)}{P_{ml}(\mathbf{S}_t; \mathbf{v}, \mu)} = \frac{\sum_{n=1}^K (\sum_{p \in S_{nt}} v_{np})^\mu + 1}{\sum_{n=1}^K (\sum_{p \in S'_{nt}} v_{np})^\mu + 1}.$$

Therefore, $\frac{P_{kj}(\mathbf{S}'_t; \mathbf{v}, \mu)}{P_{kj}(\mathbf{S}_t; \mathbf{v}, \mu)} > \frac{P_{ml}(\mathbf{S}'_t; \mathbf{v}, \mu)}{P_{ml}(\mathbf{S}_t; \mathbf{v}, \mu)}$ if $\mu < 1$, whereas $\frac{P_{kj}(\mathbf{S}'_t; \mathbf{v}, \mu)}{P_{kj}(\mathbf{S}_t; \mathbf{v}, \mu)} = \frac{P_{ml}(\mathbf{S}'_t; \mathbf{v}, \mu)}{P_{ml}(\mathbf{S}_t; \mathbf{v}, \mu)}$ if $\mu = 1$. Note that $\frac{P_{kj}(\mathbf{S}'_t; \mathbf{v}, \mu)}{P_{kj}(\mathbf{S}_t; \mathbf{v}, \mu)}$ does not depend on product j , which means that the percentage increases of purchase probabilities are same for all products in product group k . In addition, $\frac{P_{ml}(\mathbf{S}'_t; \mathbf{v}, \mu)}{P_{ml}(\mathbf{S}_t; \mathbf{v}, \mu)}$ depends on neither product l nor product group $m (\neq k)$, which means that the percentage increases of purchase probabilities are same for all products in product groups other than product group k . Therefore, in the NMNL model with $\mu < 1$, an availability of a product results in higher percentage increase of purchase probabilities for products in the same product group than those for products in different product groups. However, if products are homogeneous as in the MNL model (i.e., $\mu = 1$), the percentage increases of purchase probabilities are all same.

Proof of Proposition 4.1

The first order equation with respect to a parameter v_{kj} is given by

$$\frac{\partial \hat{L}(\mathbf{v}, \mu)}{\partial v_{kj}} = \frac{\hat{N}_{kj}}{v_{kj}} + \frac{\mu - 1}{\sum_{l \in B_k} v_{kl}} \cdot \left(\sum_{l \in B_k} \hat{N}_{kl} \right) - \frac{\hat{N}_0}{1 - s} \cdot \frac{\mu \cdot (\sum_{l \in B_k} v_{kl})^{\mu-1}}{\sum_{m=1}^K (\sum_{l \in B_m} v_{ml})^{\mu} + 1}.$$

The first order solution v_{kj}^* satisfies the following equation.

$$\frac{\hat{N}_{kj}}{v_{kj}^*} = \frac{1 - \mu}{\sum_{l \in B_k} v_{kl}^*} \cdot \left(\sum_{l \in B_k} \hat{N}_{kl} \right) + \frac{\hat{N}_0}{1 - s} \cdot \frac{\mu (\sum_{l \in B_k} v_{kl}^*)^{\mu-1}}{\sum_{m=1}^K (\sum_{l \in B_m} v_{ml}^*)^{\mu} + 1},$$

which implies that $\frac{\hat{N}_{kj}}{v_{kj}^*}$ is identical for all $j \in B_k$. We let $v_{kj}^* = m_k^* \cdot \hat{N}_{kj}$ to have

$$(m_k^*)^{\mu} \left(\sum_{l \in B_k} \hat{N}_{kl} \right)^{\mu-1} = \frac{(1 - s) \cdot (\sum_{m=1}^K (m_m^* \sum_{l \in B_m} \hat{N}_{ml})^{\mu} + 1)}{\hat{N}_0}.$$

The above equation implies that $(m_k^*)^{\mu} (\sum_{l \in B_k} \hat{N}_{kl})^{\mu-1}$ is identical for all $k = 1, \dots, K$. We let $(m_k^*)^{\mu} \cdot (\sum_{l \in B_k} \hat{N}_{kl})^{\mu-1} = M^*$. Then we have

$$M^* = \frac{1}{\hat{N}_0}$$

and

$$v_{kj}^* = \frac{\hat{N}_{kj}}{\sum_{l \in B_k} \hat{N}_{kl}} \cdot \left(\frac{\sum_{l \in B_k} \hat{N}_{kl}}{\hat{N}_0} \right)^{1/\mu}.$$

Now we prove the unimodality. The second cross partial derivatives are:

$$\begin{aligned} \frac{\partial^2 \hat{L}(\mathbf{v}, \mu)}{\partial v_{kj}^2} &= -\frac{\hat{N}_{kj}}{v_{kj}^2} + \frac{1 - \mu}{(\sum_{l \in B_k} v_{kl})^2} \cdot \left(\sum_{l \in B_k} \hat{N}_{kl} \right) \\ &\quad - \frac{\hat{N}_0}{1 - s} \cdot \left(\frac{\mu \cdot (\mu - 1) (\sum_{l \in B_k} v_{kl})^{\mu-2}}{\sum_{m=1}^K (\sum_{l \in B_m} v_{ml})^{\mu} + 1} - \frac{(\mu \cdot (\sum_{l \in B_k} v_{kl})^{\mu-1})^2}{(\sum_{m=1}^K (\sum_{l \in B_m} v_{ml})^{\mu} + 1)^2} \right), \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 \hat{LL}(\mathbf{v}, \mu)}{\partial v_{kj} \partial v_{kl}} &= \frac{1 - \mu}{(\sum_{l \in B_k} v_{kl})^2} \cdot (\sum_{l \in B_k} \hat{N}_{kl}) \\ &\quad - \frac{\hat{N}_0}{1 - s} \cdot \left(\frac{\mu \cdot (\mu - 1) (\sum_{l \in B_k} v_{kl})^{\mu-2}}{\sum_{m=1}^K (\sum_{l \in B_m} v_{ml})^\mu + 1} - \frac{(\mu \cdot (\sum_{l \in B_k} v_{kl})^{\mu-1})^2}{(\sum_{m=1}^K (\sum_{l \in B_m} v_{ml})^\mu + 1)^2} \right), \end{aligned}$$

and

$$\frac{\partial^2 \hat{LL}(\mathbf{v}, \mu)}{\partial v_{kj} \partial v_{ml}} = \frac{\hat{N}_0}{1 - s} \cdot \frac{\mu^2 \cdot (\sum_{l \in B_k} v_{kl})^{\mu-1} \cdot (\sum_{l \in B_m} v_{ml})^{\mu-1}}{(\sum_{m=1}^K (\sum_{l \in B_m} v_{ml})^\mu + 1)^2}.$$

Let $\mathbf{H}(\mathbf{v}, \mu)$ be the Hessian of the conditional expected log-likelihood function. Define $n = \sum_{k=1}^K |B_k|$, which stands for the total number of products in the full assortment. For $\mathbf{x} \in \mathcal{R} \setminus$ and $\mathbf{x} \neq 0$,

$$\begin{aligned} &\mathbf{x}^T \mathbf{H}(\mathbf{v}, \mu) \mathbf{x} \\ &= \sum_{k=1}^K \frac{1 - \mu}{(\sum_{l \in B_k} v_{kl})^2} \cdot (\sum_{l \in B_k} \hat{N}_{kl}) \cdot (\sum_{l \in B_k} x_{kl})^2 \\ &\quad - \frac{\hat{N}_0}{1 - s} \cdot \frac{\mu \cdot (\mu - 1) \sum_{k=1}^K (\sum_{l \in B_k} v_{kl})^{\mu-2}}{\sum_{m=1}^K (\sum_{l \in B_m} v_{ml})^\mu + 1} (\sum_{l \in B_k} x_{kl})^2 - \sum_{k=1}^K \sum_{j \in B_k} \hat{N}_{kj} \frac{x_{kj}^2}{v_{kj}^2} \\ &\quad + \mu^2 \cdot \frac{\hat{N}_0}{1 - s} \cdot \left(\frac{\sum_{k=1}^K (\sum_{l \in B_k} v_{kl})^{\mu-1} \cdot \sum_{l \in B_k} x_{kl}}{(\sum_{m=1}^K (\sum_{l \in B_m} v_{ml})^\mu + 1)} \right)^2. \end{aligned}$$

Applying the preference weight $v_{kj}^*(\mu)$, we have

$$\begin{aligned} &\mathbf{x}^T \mathbf{H}(\mathbf{v}^*(\mu), \mu) \mathbf{x} \\ &= (1 - \mu^2) (\hat{N}_0)^{2/\mu} \sum_{k=1}^K (\sum_{l \in B_k} \hat{N}_{kl})^{1-2/\mu} \cdot (\sum_{l \in B_k} x_{kl})^2 - (\hat{N}_0)^{2/\mu} \sum_{k=1}^K (\sum_{l \in B_k} \hat{N}_{kl})^{2-2/\mu} \sum_{j \in B_k} \frac{x_{kj}^2}{\hat{N}_{kj}} \\ &\quad + \mu^2 (\hat{N}_0)^{2/\mu-1} \cdot (1 - s) \left(\sum_{k=1}^K (\sum_{m=1}^K \hat{N}_{kl})^{1-1/\mu} (\sum_{l \in B_k} x_{kl}) \right)^2. \end{aligned}$$

Diving $(\hat{N}_0)^{2/\mu-1}$ on both sides, we have

$$\begin{aligned}
& (\hat{N}_0)^{1-2/\mu} \cdot \mathbf{x}^T \mathbf{H}(\mathbf{v}^*(\mu), \mu) \mathbf{x} \\
= & (1 - \mu^2) \cdot \hat{N}_0 \sum_{k=1}^K \left(\sum_{l \in B_k} \hat{N}_{kl} \right)^{1-2/\mu} \cdot \left(\sum_{l \in B_k} x_{kl} \right)^2 - \hat{N}_0 \sum_{k=1}^K \left(\sum_{l \in B_k} \hat{N}_{kl} \right)^{2-2/\mu} \sum_{j \in B_k} \frac{x_{kj}^2}{\hat{N}_{kj}} \\
& + \mu^2 \cdot (1 - s) \left(\sum_{k=1}^K \left(\sum_{m=1}^K \hat{N}_{km} \right)^{1-1/\mu} \left(\sum_{l \in B_k} x_{kl} \right) \right)^2 \\
= & \hat{N}_0 \sum_{k=1}^K \left(\sum_{l \in B_k} \hat{N}_{kl} \right)^{1-2/\mu} \cdot \left[\left(\sum_{l \in B_k} x_{kl} \right)^2 - \left(\sum_{l \in B_k} \hat{N}_{kl} \right) \sum_{j \in B_k} \frac{x_{kj}^2}{\hat{N}_{kj}} \right] \\
& + \mu^2 \cdot (1 - s) \left[\left(\sum_{k=1}^K \left(\sum_{m=1}^K \hat{N}_{km} \right)^{1-1/\mu} \left(\sum_{l \in B_k} x_{kl} \right) \right)^2 - \frac{\hat{N}_0}{1-s} \sum_{k=1}^K \left(\sum_{l \in B_k} \hat{N}_{kl} \right)^{1-2/\mu} \cdot \left(\sum_{l \in B_k} x_{kl} \right)^2 \right]
\end{aligned}$$

Applying the Cauchy-Schwartz inequality, we have

$$\begin{aligned}
\left(\sum_{l \in B_k} x_{kl} \right)^2 &= \left(\sum_{l \in B_k} \sqrt{\hat{N}_{kj}} \cdot \frac{x_{kl}}{\sqrt{\hat{N}_{kj}}} \right)^2 \\
&\leq \left(\sum_{l \in B_k} \hat{N}_{kj} \right) \cdot \left(\sum_{l \in B_k} \frac{x_{kl}^2}{\hat{N}_{kj}} \right), \forall k \in N,
\end{aligned}$$

and

$$\begin{aligned}
\left(\sum_{k=1}^K \left(\sum_{l \in B_k} \hat{N}_{kl} \right)^{1-1/\mu} \sum_{l \in B_k} x_{kl} \right)^2 &= \left(\sum_{k=1}^K \left(\sum_{l \in B_k} \hat{N}_{kl} \right)^{1/2} \left(\sum_{l \in B_k} \hat{N}_{kl} \right)^{1/2-1/\mu} \sum_{l \in B_k} x_{kl} \right)^2 \\
&\leq \left(\sum_{k=1}^K \sum_{l \in B_k} \hat{N}_{kl} \right) \cdot \left(\sum_{k=1}^K \left(\sum_{l \in B_k} \hat{N}_{kl} \right)^{1-2/\mu} \left(\sum_{l \in B_k} x_{kl} \right)^2 \right) \\
&= \frac{s}{1-s} \cdot \hat{N}_0 \sum_{k=1}^K \left(\sum_{l \in B_k} \hat{N}_{kl} \right)^{1-2/\mu} \left(\sum_{l \in B_k} x_{kl} \right)^2.
\end{aligned}$$

Hence, we have

$$(\hat{N}_0)^{1-2/\mu} \cdot \mathbf{x}^T \mathbf{H}(\mathbf{v}^*(\mu), \mu) \mathbf{x} \leq -\mu^2(1-s) \hat{N}_0 \sum_{k=1}^K \left(\sum_{l \in B_k} \hat{N}_{kl} \right)^{1-2/\mu} \cdot \left(\sum_{l \in B_k} x_{kl} \right)^2 < 0,$$

which leads to $\mathbf{x}^T \mathbf{H}(\mathbf{v}^*(\mu), \mu) \mathbf{x} < 0$. For a given μ , the unique critical point of the preference vector is a local maximum. Because the conditional expected log-likelihood is unbounded from below (i.e., it goes to $-\infty$ as v_{kj} goes to zero), it is unimodal.

Proof of Proposition 4.2

$$1 - P_0(\mathbf{B}, \mathbf{v}^*(\mu), \mu) = \frac{\sum_{k=1}^K (\sum_{j \in B_k} v_{kj}^*(\mu))^\mu}{\sum_{k=1}^K (\sum_{j \in B_k} v_{kj}^*(\mu))^\mu + 1} = \frac{\sum_{k=1}^K \sum_{l \in B_k} \hat{N}_{kl}}{\sum_{k=1}^K \sum_{l \in B_k} \hat{N}_{kl} + \hat{N}_0} = s.$$

Proof of Proposition B.1

The first-order equation with respect to v_{kmj} of Equation (B.1) is given by

$$\begin{aligned} \frac{\partial \hat{L}L(\mathbf{v}, \mu)}{\partial v_{kmj}} &= \frac{\hat{N}_{kmj}}{v_{kmj}} + \frac{\mu_2 - 1}{\sum_{i \in B_{km}} v_{kmi}} \cdot \sum_{i \in B_{km}} \hat{N}_{kmi} \\ &\quad + (\mu_1 - \mu_2) \cdot \frac{(\sum_{i \in B_{km}} v_{kmi})^{\mu_2 - 1}}{\sum_{n \in B_k} (\sum_{i \in B_{kn}} v_{kni})^{\mu_2}} \cdot \sum_{n \in B_k} \sum_{i \in B_{kn}} \hat{N}_{kni} \\ &\quad - \frac{\hat{N}_0}{1 - s} \cdot \frac{\mu_1 (\sum_{n \in B_k} (\sum_{i \in B_{kn}} v_{kni})^{\mu_2})^{\mu_1 / \mu_2 - 1} (\sum_{i \in B_{km}} v_{kmi})^{\mu_2 - 1}}{\sum_{l=1}^K (\sum_{n \in B_l} (\sum_{i \in B_{ln}} v_{lni})^{\mu_2})^{\mu_1 / \mu_2} + 1}. \end{aligned}$$

The first-order solution v_{kmj}^* satisfies the following equation.

$$\begin{aligned} \frac{\hat{N}_{kmj}}{v_{kmj}^*} &= \frac{1 - \mu_2}{\sum_{i \in B_{km}} v_{kmi}^*} \cdot \sum_{i \in B_{km}} \hat{N}_{kmi} \\ &\quad + (\mu_2 - \mu_1) \cdot \frac{(\sum_{i \in B_{km}} v_{kmi}^*)^{\mu_2 - 1}}{\sum_{n \in B_k} (\sum_{i \in B_{kn}} v_{kni}^*)^{\mu_2}} \cdot \sum_{n \in B_k} \sum_{i \in B_{kn}} \hat{N}_{kni} \\ &\quad + \frac{\hat{N}_0}{1 - s} \cdot \frac{\mu_1 (\sum_{n \in B_k} (\sum_{i \in B_{kn}} v_{kni}^*)^{\mu_2})^{\mu_1 / \mu_2 - 1} (\sum_{i \in B_{km}} v_{kmi}^*)^{\mu_2 - 1}}{\sum_{l=1}^K (\sum_{n \in B_l} (\sum_{i \in B_{ln}} v_{lni}^*)^{\mu_2})^{\mu_1 / \mu_2} + 1}. \end{aligned}$$

This means that $\frac{\hat{N}_{kmj}}{v_{kmj}^*}$ is identical for all product $j \in B_{km}$. We let $v_{kmj}^* = m_{km}^* \hat{N}_{kmj}$ to have

$$\begin{aligned} \frac{1}{m_{km}^*} &= \frac{1 - \mu_2}{m_{km}^*} + (\mu_2 - \mu_1) \frac{(m_{km}^*)^{\mu_2 - 1} (\sum_{i \in B_{km}} \hat{N}_{kmi})^{\mu_2 - 1}}{\sum_{n \in B_k} (m_{kn}^*)^{\mu_2} (\sum_{i \in B_{kn}} \hat{N}_{kni})^{\mu_2}} \cdot \sum_{n \in B_k} \sum_{i \in B_{kn}} \hat{N}_{kni} \\ &\quad + \frac{\hat{N}_0}{1 - s} \cdot \frac{\mu_1 (m_{km}^*)^{\mu_2 - 1} (\sum_{n \in B_k} (m_{kn}^*)^{\mu_2} (\sum_{i \in B_{kn}} \hat{N}_{kni})^{\mu_2})^{\mu_1 / \mu_2 - 1} (\sum_{i \in B_{km}} \hat{N}_{kmi})^{\mu_2 - 1}}{\sum_{l=1}^K (\sum_{n \in B_l} (m_{ln}^*)^{\mu_2} (\sum_{i \in B_{ln}} \hat{N}_{lni})^{\mu_2})^{\mu_1 / \mu_2} + 1} \end{aligned}$$

and

$$\begin{aligned} \frac{\mu_2}{(m_{km}^*)^{\mu_2} (\sum_{i \in B_{km}} \hat{N}_{kmi})^{\mu_2 - 1}} &= (\mu_2 - \mu_1) \frac{\sum_{n \in B_k} \sum_{i \in B_{kn}} \hat{N}_{kni}}{\sum_{n \in B_k} (m_{kn}^*)^{\mu_2} (\sum_{i \in B_{kn}} \hat{N}_{kni})^{\mu_2}} \\ &\quad + \frac{\hat{N}_0}{1 - s} \cdot \frac{\mu_1 (\sum_{n \in B_k} (m_{kn}^*)^{\mu_2} (\sum_{i \in B_{kn}} \hat{N}_{kni})^{\mu_2})^{\mu_1 / \mu_2 - 1}}{\sum_{l=1}^K (\sum_{n \in B_l} (m_{ln}^*)^{\mu_2} (\sum_{i \in B_{ln}} \hat{N}_{lni})^{\mu_2})^{\mu_1 / \mu_2} + 1}. \end{aligned}$$

The above equation implies that $(m_{km}^*)^{\mu_2} (\sum_{i \in B_{km}} \hat{N}_{kmi})^{\mu_2 - 1}$ is identical for all $m \in B_k$. We let $(m_{km}^*)^{\mu_2} (\sum_{i \in B_{km}} \hat{N}_{kmi})^{\mu_2 - 1} = m_k^*$

$$\frac{\mu_2}{m_k^*} = \frac{\mu_2 - \mu_1}{m_k^*} + \frac{\hat{N}_0}{1 - s} \cdot \frac{\mu_1 (m_k^* \sum_{n \in B_k} \sum_{i \in B_{kn}} \hat{N}_{kni})^{\mu_1 / \mu_2 - 1}}{\sum_{l=1}^K (m_l^* \sum_{n \in B_l} \sum_{i \in B_{ln}} \hat{N}_{lni})^{\mu_1 / \mu_2} + 1}.$$

We have

$$(m_k^*)^{\mu_1/\mu_2} \left(\sum_{n \in B_k} \sum_{i \in B_{kn}} \hat{N}_{kni} \right)^{\mu_1/\mu_2 - 1} = \frac{1-s}{\hat{N}_0} \cdot \left(\sum_{l=1}^K (m_l^* \sum_{n \in B_l} \sum_{i \in B_{ln}} \hat{N}_{lni})^{\mu_1/\mu_2} + 1 \right)$$

Because $(m_k^*)^{\mu_1/\mu_2} \left(\sum_{n \in B_k} \sum_{i \in B_{kn}} \hat{N}_{kni} \right)^{\mu_1/\mu_2 - 1}$ is identical for all $k = 1, \dots, K$, we let $(m_k^*)^{\mu_1/\mu_2} \left(\sum_{n \in B_k} \sum_{i \in B_{kn}} \hat{N}_{kni} \right)^{\mu_1/\mu_2 - 1} = M^*$. Then

$$\begin{aligned} M^* &= \frac{1-s}{\hat{N}_0} \cdot (M^* \sum_{l=1}^K \sum_{n \in B_l} \sum_{i \in B_{ln}} \hat{N}_{lni} + 1) \\ &= \frac{1}{\hat{N}_0} \cdot (M^* \cdot s \cdot \hat{N}_0 + 1 - s) \end{aligned}$$

Therefore,

$$M^* = \frac{1}{\hat{N}_0}$$

and

$$v_{kmj}^* = \frac{\hat{N}_{kmj}}{\sum_{i \in B_{km}} \hat{N}_{kmi}} \cdot \left(\frac{\sum_{i \in B_{km}} \hat{N}_{kmi}}{\sum_{n \in B_k} \sum_{i \in B_{kn}} \hat{N}_{kni}} \right)^{1/\mu_2} \cdot \left(\frac{\sum_{n \in B_k} \sum_{i \in B_{kn}} \hat{N}_{kni}}{\hat{N}_0} \right)^{1/\mu_1}.$$