Online Appendix

Discrete Choice Models with Piecewise Linear Utility: Modeling, Estimation and Pricing

Chenxu Ke Ruxian Wang Zifeng Zhao

In this electronic companion, we provide the technical proofs (Section §1), additional details on the approximation algorithm (Section §1.1), additional remarks (Section §2), additional numerical results on real data applications (Section §3), and additional examples (Section §4) for this paper.

§1. Technical Proofs

Proof of Proposition 1. Under the piecewise MNL model, we take the derivatives of the choice probabilities d_i defined in (2) and d_j with respect to τ_i for any $i, j \in S$. It is straightforward to have $\partial d_i/\partial \tau_i = \partial d_j/\partial \tau_i = 0$ for $\tau_i > p_i$ due to $(p_i - \tau_i)_+ = 0$. For $\tau_i < p_i$, we have $\partial d_i/\partial \tau_i = -(\beta_{1i} - \beta_{2i})d_i(1 - d_i) < 0$ and $\partial d_j/\partial \tau_i = (\beta_{1i} - \beta_{2i})d_i d_j > 0$ under Type I piecewise MNL model due to $\beta_{1i} > \beta_{2i}$ for all $i \in S$. By contrast, we immediately have $\partial d_i/\partial \tau_i > 0$ and $\partial d_j/\partial \tau_i < 0$ under Type II piecewise MNL model. In addition, at $\tau_i = p_i$, d_i and d_j are non-differentiable with respect to τ_i because $\frac{\partial d_i}{\partial \tau_i}\Big|_{p_i^-} \neq \frac{\partial d_i}{\partial \tau_i}\Big|_{p_i^+}$ and $\frac{\partial d_j}{\partial \tau_i}\Big|_{p_i^-} \neq \frac{\partial d_j}{\partial \tau_i}\Big|_{p_i^+}$. ☐

Proof of Proposition 2. Consider the derivatives of the choice probabilities d_i defined in (2) and d_j with respect to β_{1i} and β_{2i} for any $i, j \in S$. It is easy to show that $\partial d_i/\partial \beta_{1i} = d_i(1 - d_i)[(p_i - \tau_i)_+ - p_i] < 0$ and $\partial d_j/\partial \beta_{1i} = -d_i d_j[(p_i - \tau_i)_+ - p_i] > 0$ for any $\tau_i > 0$ and $p_i \in (0, \infty)$. Moreover, we have $\partial d_i/\partial \beta_{2i} = -(p_i - \tau_i)_+ d_i(1 - d_i) \leq 0$ and $\partial d_j/\partial \beta_{2i} = d_i d_j (p_i - \tau_i)_+ \geq 0$. In particular, the equalities hold for $p_i \leq \tau_i$. ☐

Proof of Corollary 1. Let $R = \sum_{i \in S}(p_i - c_i)d_i$ denote the total expected profit. For any product $i \in S$, if $\tau_i > p_i$, it is straightforward to have $\partial R/\partial \tau_i = 0$. While if $\tau_i < p_i$, we take the derivative of R with respect to τ_i and have $\partial R/\partial \tau_i = (\beta_{1i} - \beta_{2i})d_i[R - (p_i - c_i)]$. It is clear that $\partial R/\partial \tau_i < 0$ for $p_i - c_i > R$, and $\partial R/\partial \tau_i > 0$ for $p_i - c_i < R$ under Type I piecewise MNL. The opposite holds for Type II piecewise MNL. ☐

Proof of Corollary 2. For any product $i \in S$, we consider the derivatives of the total expected profit ($R = \sum_{i \in S}(p_i - c_i)d_i$) with respect to the price sensitivities. In particular, in terms of β_{1i}, we have $\partial R/\partial \beta_{1i} = d_i[(p_i - \tau_i)_+ - p_i][(p_i - c_i) - R]$. It is clear that $\partial R/\partial \beta_{1i} < 0$ for $p_i - c_i > R$ and $\partial R/\partial \beta_{1i} > 0$ for $p_i - c_i < R$ with any $\tau_i > 0$ and $p_i \in (0, \infty)$. Moreover, for β_{2i}, we have $\partial R/\partial \beta_{2i} = -d_i(p_i - \tau_i)_+[(p_i - c_i) - R]$. Clearly, $\partial R/\partial \beta_{2i} \leq 0$ for $p_i - c_i > R$ and $\partial R/\partial \beta_{2i} \geq 0$ for $p_i - c_i < R$. Note that when $\tau_i \geq p_i$, the change in β_{2i} does not affect the total profit. ☐

Proof of Corollary 3. For any product $i \in S$, if $\beta_{1i} = \beta_1$, $\beta_{2i} = \beta_2$ and $\tau_i = \tau$, the choice probability for product $i \in S$ is expressed as

$$d_i(S, \mathbf{p}; \tau) = \frac{\exp(\alpha_i - \beta_1 p_i + (\beta_1 - \beta_2)(p_i - \tau)_+)}{1 + \sum_{j \in S} \exp(\alpha_j - \beta_1 p_j + (\beta_1 - \beta_2)(p_j - \tau)_+)}.$$

If $\tau > p_{\max} = \max_{j \in S} p_j$, it is straightforward to have $\partial d_i/\partial \tau = 0$ for any product $i \in S$ with price $p_i \in (0, \tau)$. If $\tau < p_i$, we have

$$d_i = \frac{\exp(\alpha_i - \beta_2 p_i)/\exp((\beta_1 - \beta_2)\tau)}{1 + \sum_{i \in S_1} \exp(\alpha_i - \beta_1 p_i) + \sum_{i \in S_2} \exp(\alpha_i - \beta_2 p_i)/\exp((\beta_1 - \beta_2)\tau)},$$

where $S_1 = \{i \in S | p_i < \tau\}$ and $S_2 = \{i \in S | p_i \geq \tau\}$. Then, it is easy to derive that $\partial d_i/\partial \tau < 0$ for any product $i \in S$ with price $p_i \in (\tau, \infty)$ under Type I piecewise MNL model. The opposite holds for Type II piecewise MNL. Note that d_i is non-differentiable at $\mathbf{p}$ (not only p_i) for any $i \in S$. If $p_i < \tau < p_{\max}$, we have

$$d_i = \frac{\exp(\alpha_i - \beta_1 p_i)}{1 + \sum_{i \in S_1} \exp(\alpha_i - \beta_1 p_i) + \sum_{j \in S_2} \exp(\alpha_i - \beta_2 p_i)/\exp((\beta_1 - \beta_2)\tau)}.$$

Obviously, $\partial d_i/\partial\tau > 0$ holds for any product $i \in S$ with price $p_i < \tau < p_{\max}$ under Type I piecewise MNL model. The opposite holds for Type II piecewise MNL. We also notice that both the left and right derivatives exist but not the same where τ is equal to any $p_i, i \in S$. $\square$

Proof of Corollary 4. Under the piecewise MNL model with homogeneous price sensitivities β_1 and β_2, we consider the derivatives of $d_i(S,\mathbf{p};\tau)$ with respect to β_1 and β_2 respectively. In particular,

$$\frac{\partial d_i}{\partial \beta_1} = d_i \cdot [-p_i + (p_i-\tau)_+ - \sum_{j\in S}(-p_j + (p_j-\tau)_+)d_j] = d_i \cdot [-p_i + (p_i-\tau)_+ + \sum_{j\in S_1} p_j d_j + \sum_{j\in S_2} \tau d_j].$$

Thus, we have $\frac{\partial d_i}{\partial \beta_1} > 0$, if $p_i < \sum_{j\in S_1} p_j d_j + \tau\sum_{j\in S_2} d_j < \tau$, otherwise $\frac{\partial d_i}{\partial \beta_1} < 0$ if $p_i > \sum_{j\in S_1} p_j d_j + \tau\sum_{j\in S_2} d_j$, and $\frac{\partial d_i}{\partial \beta_1} = 0$ if $p_i = \sum_{j\in S_1} p_j d_j + \tau\sum_{j\in S_2} d_j$. Moreover, $\frac{\partial d_i}{\partial \beta_2} = -d_i \cdot [(p_i-\tau)_+ - \sum_{j\in S_2}(p_j-\tau)d_j]$. In particular, if $\tau \geq \max_{j\in S} p_j$, we have $\frac{\partial d_i}{\partial \beta_2} = 0$ for each i, that is, the choice probability of each product is independent of β_2. If $\tau < \max_{j\in S} p_j$, we have $\frac{\partial d_i}{\partial \beta_2} > 0$ for $p_i < \tau + \sum_{j\in S_2}(p_j-\tau)d_j$; otherwise, we have $\frac{\partial d_i}{\partial \beta_2} < 0$ for $p_i > \tau + \sum_{j\in S_2}(p_j-\tau)d_j$, and $\frac{\partial d_i}{\partial \beta_2} = 0$ for $p_i = \tau + \sum_{j\in S_2}(p_j-\tau)d_j$.

Moreover, we also consider the derivatives of $d_0(S,\mathbf{p};\tau)$ with respect to β_1 and β_2 respectively:

$$\frac{\partial d_0}{\partial \beta_1} = d_0 \cdot \sum_{j\in S}[p_j - (p_j-\tau)_+]d_j > 0, \text{ and } \frac{\partial d_0}{\partial \beta_2} = d_0 \cdot \sum_{j\in S}(p_j-\tau)_+ d_j \geq 0.$$

In particular, we have $\partial d_0/\partial\beta_2 = 0$ if $\tau \geq \max_{j\in S} p_j$. $\square$

Proof of Corollary 5. First, we consider the case with an identical threshold τ under Type I piecewise MNL. For $\tau > p_{\max}$, it is straightforward to have $\partial R/\partial\tau = 0$. Otherwise, we take the derivative of R with respect to τ, and have $\partial R/\partial\tau = (\beta_1-\beta_2)[R\cdot\sum_{j\in S_2} d_j - \sum_{j\in S_2}(p_j-c_j)d_j]$. We can see that $\partial R/\partial\tau < 0$ for $\sum_{j\in S_2}(p_j-c_j)d_j > R\cdot\sum_{j\in S_2} d_j$, and $\partial R/\partial\tau > 0$ for $\sum_{j\in S_2}(p_j-c_j)d_j < R\cdot\sum_{j\in S_2} d_j$. Then, the opposite holds clearly for Type II piecewise MNL. For both types of piecewise MNL, since each d_i is non-differentiable with respect to τ at $\mathbf{p}$ (by the proof of Corollary 3), we can infer that R is non-differentiable at $\mathbf{p}$. Although both the left and right derivatives exist when τ is equal to any p_i, $i \in S$, they are not the same. $\square$

Proof of Proposition 3. For a fixed $\tau \in \Theta_\tau$, the Hessian matrix of $L_G(\eta,\tau)$ with respect to η takes the form

$$\frac{\partial^2}{\partial\eta\partial\eta^\top} L_G(\eta,\tau) = -\frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{s}\mathbb{P}(i|\mathbf{z}_g,\mathbf{p}_g;\eta,\tau)(\mathbf{x}_{g,i}(\tau)-\bar{\mathbf{x}}_g(\tau))(\mathbf{x}_{g,i}(\tau)-\bar{\mathbf{x}}_g(\tau))^\top,$$

where $\bar{\mathbf{x}}_g(\tau) = \sum_{i=1}^{s}\mathbb{P}(i|\mathbf{z}_g,\mathbf{p}_g;\eta,\tau)\mathbf{x}_{g,i}(\tau)$, which also implicitly depends on η. To show $L_G(\eta,\tau)$ is strictly concave, we only need to prove that $\frac{\partial^2}{\partial\eta\partial\eta^\top}L_G(\eta,\tau)$ is negative definite.

Note that all $\mathbb{P}(i|\mathbf{z}_g,\mathbf{p}_g;\eta,\tau)$ are positive for any $(\eta,\tau)\in\Theta$, making $\frac{\partial^2}{\partial\eta\partial\eta^\top}L_G(\eta,\tau)$ a negative semi-definite matrix. Thus, $\frac{\partial^2}{\partial\eta\partial\eta^\top}L_G(\eta,\tau)$ is negative definite as long as it has full rank. Define $D'(\tau,\eta)$ as the $sG\times(d+1)$ matrix whose rows are $\mathbf{x}_{g,i}(\tau)-\bar{\mathbf{x}}_g(\tau)$ for $g=1,\ldots,G$ and $i=1,\ldots,s$. Clearly, $\frac{\partial^2}{\partial\eta\partial\eta^\top}L_G(\eta,\tau)$ is of full rank if $D'(\tau,\eta)$ is of rank $d+1$. To conclude, it is easy to see that for any η, the two matrices $D'(\tau,\eta)$ and $D(\tau)$ share the same rank. $\square$

Proof of Theorem 1. The first claim can be established based on Berge's maximum theorem. For clarity, we provide a more concrete proof here. To show $\widehat{\eta}(\tau)$ is continuous with respect to $\tau\in\Theta_\tau$, it is equivalent to show that for any sequence $\{\tau_n\}\to\tau$, we have $\widehat{\eta}(\tau_n)\to\widehat{\eta}(\tau)$. We prove by contradiction. Suppose $\widehat{\eta}(\tau_n)\not\to\widehat{\eta}(\tau)$, since $\widehat{\eta}(\tau_n)\in\Theta_\eta$ and Θ_η is compact, by Bolzano-Weierstrass theorem, there is a convergent subsequence $\{\tau_n'\}\subseteq\{\tau_n\}$ such that $\widehat{\eta}(\tau_n')\to\eta^*\neq\widehat{\eta}(\tau)$. By definition, we have $L(\widehat{\eta}(\tau),\tau_n')\leq L(\widehat{\eta}(\tau_n'),\tau_n')$. By continuity of $L(\eta,\tau)$, we have that in the limit, $L(\widehat{\eta}(\tau),\tau)\leq L(\eta^*,\tau)$. However, by Assumption 1, $L_G(\eta,\tau)$ has a unique maximizer $\widehat{\eta}(\tau)$, which implies that $\eta^*=\widehat{\eta}(\tau)$. Thus, there is a contradiction and the first claim is proved.

The second claim follows from an application of the implicit function theorem. Specifically, at any $\tau^*\in\Theta_\tau\setminus\mathbf{p}$, the log-likelihood function $L(\eta,\tau)$ is continuously differentiable with respect to (η,τ). By the first order condition, we have that

$$\frac{\partial}{\partial\eta}L(\eta,\tau^*)|_{\eta=\widehat{\eta}(\tau^*)} = 0\in\mathbb{R}^{d+1}.$$

Furthermore, the Jacobian of $\frac{\partial}{\partial \eta}L(\eta,\tau)$ with respect to η evaluated at $\eta=\widehat{\eta}(\tau^*)$ is in fact the Hessian matrix $\frac{\partial^2}{\partial\eta\partial\eta^\top}L(\eta,\tau^*)|_{\eta=\widehat{\eta}(\tau^*)}$, which is invertible by Assumption 1. Thus, by the implicit function theorem, there exists an open neighborhood U containing τ^* such that there exists a unique continuously differentiable function $g(\tau):U\to\mathbb{R}^{d+1}$ such that $g(\tau^*)=\widehat{\eta}(\tau^*)$ and $\frac{\partial}{\partial\eta}L(\eta,\tau)|_{\eta=g(\tau)}=0$ for all $\tau\in U$. By Assumption 1, for any τ, due to the strict concavity, $\widehat{\eta}(\tau)$ is the unique value such that $\frac{\partial}{\partial\eta}L(\eta,\tau)|_{\eta=g(\tau)}=0$. This implies that $g(\tau)=\widehat{\eta}(\tau)$ for all $\tau\in\Theta_\tau\setminus\mathbf{p}$. Therefore, $\widehat{\eta}(\tau)$, and thus $L(\widehat{\eta}(\tau),\tau)$ is a differentiable function with respect to τ. We finish the proof by the chain rule

$$\frac{\partial}{\partial\tau}L(\widehat{\eta}(\tau),\tau)=\frac{\partial}{\partial\eta}L(\eta,\tau)|_{\eta=\widehat{\eta}(\tau)}\frac{\partial}{\partial\tau}\widehat{\eta}(\tau)+\frac{\partial}{\partial\tau}L(\eta,\tau)|_{\eta=\widehat{\eta}(\tau)}=\frac{\partial}{\partial\tau}L(\eta,\tau)|_{\eta=\widehat{\eta}(\tau)}. \qquad \square$$

Proof of Proposition 4. Suppose the true parameter is $\theta^o=(\eta^o,\tau^o)$. We first show that, given observations $\{\mathbf{z}_g,\mathbf{p}_g\}_{g=1}^G$, for any $\theta=(\eta,\tau)\neq\theta^o$, if $D(\tau,\tau^o)$ is of rank $d+2$, there exists at least one consumer g and one item i such that $\mathbb{P}(i|\mathbf{z}_g,\mathbf{p}_g;\eta^o,\tau^o)\neq\mathbb{P}(i|\mathbf{z}_g,\mathbf{p}_g;\eta,\tau)$. We prove by contradiction. It is straightforward to show that $\mathbb{P}(i|\mathbf{z}_g,\mathbf{p}_g;\eta^o,\tau^o)=\mathbb{P}(i|\mathbf{z}_g,\mathbf{p}_g;\eta,\tau)$ for all consumers $g=1,\dots,G$, and all items $i=1,\dots,s$ is equivalent to $D(\tau^o)\eta^o=D(\tau)\eta$, where $D(\tau)$ is defined in Proposition 3. Suppose $\tau^o=\tau$, we then have $D(\tau)(\eta^o-\eta)=0$, which by assumption implies that $\eta=\eta^o$. Thus, we must have $\tau^o\neq\tau$. Rearranging the equation, we have $D(\tau^o,\tau)[(\eta^o_{-\lambda^o}-\eta_{-\lambda})^\top,\lambda^o,-\lambda]^\top=0$, where $\eta_{-\lambda}:=(\alpha,\beta_1)$. By assumption, this holds only if $\eta^o_{-\lambda^o}=\eta_{-\lambda}$ and $\lambda^o=\lambda=0$, i.e., there is no segmentation effect in price sensitivities, which is a contradiction. Assumption 2(b) follows by noting that $\mathbb{P}(i|\mathbf{z}_g,\mathbf{p}_g;\eta,\tau)$ is a continuous function of the feature vector $\mathbf{x}_g=(vec(\mathbf{z}_g)^\top,\mathbf{p}_g^\top)^\top$. $\square$

Proof of Theorem 2. We first establish a uniform convergence result for the log-likelihood function $L_G(\theta)=L_G(\eta,\tau)$, which serves as the basis for proving the consistency result. Denote the population log-likelihood as $L(\theta)=\mathbb{E}(\log(\mathbb{P}(c_g|\mathbf{z}_g,\mathbf{p}_g;\theta)))$. Specifically, we show

$$\sup_{\theta\in\Theta}|L_G(\theta)-L(\theta)|\to_p 0,\ \text{as } G\to\infty, \tag{EC.1}$$

by verifying the high-level assumptions of Lemma 2.4 in Newey and McFadden (1994), which gives a generic uniform law of large number (ULLN). First, it is easy to see that $\mathbb{P}(c_g|\mathbf{z}_g,\mathbf{p}_g;\theta)$ and thus $\log(\mathbb{P}(c_g|\mathbf{z}_g,\mathbf{p}_g;\theta))$ is a continuous function of θ. Second, by the compactness of Θ and the boundedness of the feature space, there is a lower bound on $\mathbb{P}(c_g|\mathbf{z}_g,\mathbf{p}_g;\theta)$ and thus there exists a sufficiently large constant $C>0$ such that $\sup_{\theta\in\Theta}|\log(\mathbb{P}(c_g|\mathbf{z}_g,\mathbf{p}_g;\theta))|<C$ for any $(c_g,\mathbf{z}_g,\mathbf{p}_g)$. Hence, by Lemma 2.4 in Newey and McFadden (1994), the population log-likelihood $L(\theta)$ is a continuous function of θ and the uniform convergence result (EC.1) holds. Furthermore, by Assumption 2 and the information inequality, we have that $L(\theta)$ is uniquely maximized at $\theta^o=(\eta^o,\tau^o)$.

The rest of the proof follows from standard arguments. To conserve space, we directly invoke Theorem 2.1 in Newey and McFadden (1994), which provides a consistency result for a generic MLE under high-level regularity conditions. Specifically, based on our arguments above, it can be easily seen that all conditions (i)-(iv) of Theorem 2.1 in Newey and McFadden (1994) are verified for $\widehat{\theta}$. Thus, we have that the MLE is consistent where $\widehat{\theta}\to_p\theta^o$ as $G\to\infty$. $\square$

Proof of Theorem 3.

[Step I]: We first show that the approximate first order condition holds, i.e. $m_G(\widehat{\theta})=o_p(1/\sqrt{G})$. Denote

$$m_G^1(\theta)=\frac{1}{G}\sum_{g=1}^G\left(\mathbf{x}_{g,c_g}(\tau)-\sum_{i=1}^s\frac{\exp(\mathbf{x}_{g,i}(\tau)^\top\eta)}{\sum_{j=1}^s\exp(\mathbf{x}_{g,j}(\tau)^\top\eta)}\mathbf{x}_{g,i}(\tau)\right),$$

$$m_G^2(\theta)=\frac{1}{G}\sum_{g=1}^G\left(-\lambda\mathbb{I}(p_{g,c_g}>\tau)+\sum_{i=1}^s\frac{\exp(\mathbf{x}_{g,i}(\tau)^\top\eta)}{\sum_{j=1}^s\exp(\mathbf{x}_{g,j}(\tau)^\top\eta)}\lambda\mathbb{I}(p_{g,i}>\tau)\right).$$

We have $m_G(\theta)=[m_G^1(\theta),m_G^2(\theta)]$. By the differentiability of $L_G(\theta)$ with respect to η, we have that $m_G^1(\widehat{\theta})\equiv 0$. Thus, we only need to prove that $m_G^2(\widehat{\theta})=o_p(1/\sqrt{G})$. Define

$$m_G^3(\theta)=\frac{1}{G}\sum_{g=1}^G\left(-\lambda\mathbb{I}(p_{g,c_g}\geq\tau)+\sum_{i=1}^s\frac{\exp(\mathbf{x}_{g,i}(\tau)^\top\eta)}{\sum_{j=1}^s\exp(\mathbf{x}_{g,j}(\tau)^\top\eta)}\lambda\mathbb{I}(p_{g,i}\geq\tau)\right).$$

Note that $m_G^2(\theta)$ is the left derivative and $m_G^3(\theta)$ is the right derivative of $L_G(\theta)$ with respect to τ. By the fact that $L_G(\widehat{\theta})$ is maximized at $\widehat{\theta}$, we have that $m_G^2(\widehat{\theta}) \geq 0$ and $m_G^3(\widehat{\theta}) \leq 0$. Moreover, we have that

$$m_G^2(\widehat{\theta}) = m_G^3(\widehat{\theta}) + \frac{1}{G}\sum_{g=1}^{G}\left(\lambda\mathbb{I}(p_{g,c_g}=\tau) - \sum_{i=1}^{s}\frac{\exp(\mathbf{x}_{g,i}(\tau)^\top\eta)}{\sum_{j=1}^{s}\exp(\mathbf{x}_{g,j}(\tau)^\top\eta)}\lambda\mathbb{I}(p_{g,i}=\tau)\right) \leq \lambda\cdot s/G,$$

where the inequality follows from Assumption 2 (i.e. continuity of the price distribution) and that $m_G^3(\widehat{\theta}) \leq 0$. Thus, we have that $0 \leq m_G^2(\widehat{\theta}) \leq \lambda s/G$ and $m_G(\widehat{\theta}) = o_p(1/\sqrt{G})$.

[Step II]: Next, we follow the basic steps in Andrews (1994) (Section 3.2) to establish the asymptotic normality of $\widehat{\theta}$. Define $m^*(\theta) = \mathbb{E}(m_G(\theta)) = \mathbb{E}(m(\theta|c_g,\mathbf{z}_g,\mathbf{p}_g))$. Importantly, though $m_G(\theta)$ is not a continuous function, due to the smoothing from the expectation operator, it is easy to verify that $m^*(\theta)$ is a continuously differentiable function with respect to θ. In addition, we can further verify that $m^*(\theta^o) \equiv 0$.

Thus, taking a Taylor expansion of $m^*(\theta^o)$ around $\widehat{\theta}$, we have that $0 = \sqrt{G}m^*(\theta^o) = \sqrt{G}m^*(\widehat{\theta}) - \frac{\partial}{\partial\theta}m^*(\widetilde{\theta})\sqrt{G}(\widehat{\theta} - \theta^o)$, where $\widetilde{\theta} = c\widehat{\theta} + (1-c)\theta^o$ for some $c \in (0,1)$. Rearranging the above equation, we further have

$$\begin{aligned}\frac{\partial}{\partial\theta}m^*(\widetilde{\theta})\sqrt{G}(\widehat{\theta}-\theta^o) =& \sqrt{G}(m^*(\widehat{\theta}) - m_G(\widehat{\theta})) + \sqrt{G}(m_G(\widehat{\theta}) - m_G(\theta^o)) + \sqrt{G}(m_G(\theta^o) - m^*(\theta^o)) \\ =& v_G(\theta^o) - v_G(\widehat{\theta}) + o_p(1) + \sqrt{G}m_G(\theta^o). \end{aligned} \tag{EC.2}$$

By Lemma 1, the process $v_G(\theta)$ is stochastically equicontinuous. Since $\widehat{\theta} \to_p \theta^o$, we have that $v_G(\widehat{\theta}) - v_G(\theta^o) = o_p(1)$. By the classical CLT, we have that $\sqrt{G}m_G(\theta^o) \to_d N(0,\Sigma)$, where $\Sigma = \mathbb{E}(m(\theta^o|c_g,\mathbf{z}_g,\mathbf{p}_g)m(\theta^o|c_g,\mathbf{z}_g,\mathbf{p}_g)^\top)$.

To finish the proof, we only need to show that $\frac{\partial}{\partial\theta}m^*(\widetilde{\theta}) \to_p -\Sigma$. Note that by the continuity of $\frac{\partial}{\partial\theta}m^*(\theta)$ with respect to θ, we have that $\frac{\partial}{\partial\theta}m^*(\widetilde{\theta}) \to_p \frac{\partial}{\partial\theta}m^*(\theta^o)$. Thus, we need to show $\frac{\partial}{\partial\theta}m^*(\theta^o) = -\Sigma$, which is essentially the information matrix equality and can be directly verified by elementary algebra. We omit it to conserve space.

Finally, by the compactness of the parameter space Θ and the boundedness of $(\mathbf{z}_g,\mathbf{p}_g)$, there exists a constant $C>0$ such that for any $(\mathbf{z}_g,\mathbf{p}_g)$, we have $\sup_{\theta\in\Theta}\|m(\theta|c_g,\mathbf{z}_g,\mathbf{p}_g)\| < C$. Thus, similar to the proof of Theorem 2, by uniform law of large number, we have $\widehat{\Sigma}_G \to_p \Sigma$. □

LEMMA 1 **(Stochastic equicontinuity)**. *Denote* $\mathbf{w}_g = (c_g,\mathbf{z}_g,\mathbf{p}_g)$ *and* $m(\theta|\mathbf{w}_g) = m(\theta|c_g,\mathbf{z}_g,\mathbf{p}_g)$. *The class of function* $\mathcal{M} = \{m(\theta|\cdot):\theta\in\Theta\}$ *satisfies Pollard's entropy condition with a constant* $C>0$. *Thus, the sequence of empirical processes* $\{v_G(\cdot):G\geq 1\}$ *is stochastically equicontinuous.*

Proof of Lemma 1. First, note that stochastic equicontinuity of a vector-valued empirical process follows from the stochastic equicontinuity of each element of the empirical process. Thus, to establish the stochastic equicontinuity of $\{v_G(\cdot):G\geq 1\}$, we only need to establish it for each scalar component of $\{v_G(\cdot):G\geq 1\}$.

To do so, we invoke Theorem 1 in Andrews (1994), which requires the verification of Assumptions A, B and C in Section 4 of Andrews (1994). Then, we verify Assumption A, by showing that the class of function $\mathcal{M} = \{m(\theta|\cdot):\theta\in\Theta\}$ satisfies Pollard's entropy condition with a constant $C>0$. Assumptions B and C can then be verified trivially.

To verify Assumption A, we further invoke Theorem 2 and Theorem 3 of Andrews (1994). By Theorem 3, if two classes of functions $\mathcal{M}_1$ and $\mathcal{M}_2$ satisfy Pollard's entropy condition with constants C_1 and C_2, respectively, the class of functions resulting from addition $\mathcal{M}_1\oplus\mathcal{M}_2 = \{m+m':m\in\mathcal{M}_1, m'\in\mathcal{M}_2\}$ and multiplication $\mathcal{M}_1\mathcal{M}_2 = \{m\cdot m': m\in\mathcal{M}_1, m'\in\mathcal{M}_2\}$ also satisfies Pollard's entropy condition with constants C_1+C_2 and C_1C_2, respectively.

Thus, we only need to verify that each individual component of $m(\theta|\cdot)$ satisfies Pollard's entropy condition. Recall

$$m(\theta|\mathbf{w}_g) = m(\theta|c_g,\mathbf{z}_g,\mathbf{p}_g) = \begin{bmatrix}\mathbf{x}_{g,c_g}(\tau) \\ -\lambda\mathbb{I}(p_{g,c_g}>\tau)\end{bmatrix} - \sum_{i=1}^{s}\frac{\exp(\mathbf{x}_{g,i}(\tau)^\top\eta)}{\sum_{j=1}^{s}\exp(\mathbf{x}_{g,j}(\tau)^\top\eta)}\begin{bmatrix}\mathbf{x}_{g,i}(\tau) \\ -\lambda\mathbb{I}(p_{g,i}>\tau)\end{bmatrix},$$

where $\mathbf{x}_{g,i}(\tau) = [\mathbf{z}_{g,i}^\top, -p_{g,i}, p_{g,i}(\tau)]^\top \in \mathbb{R}^{d+1}$ for $i=1,\ldots,s$ with $p_{g,i}(\tau) = \max(0, p_{g,i}-\tau)$. In the following, we verify that $\mathbf{x}_{g,c_g}(\tau)$, $\mathbb{I}(p_{g,c_g}>\tau)$, and $\frac{\exp(\mathbf{x}_{g,i}(\tau)^\top\eta)}{\sum_{j=1}^{s}\exp(\mathbf{x}_{g,j}(\tau)^\top\eta)}$ satisfy Pollard's entropy condition with some constant $C>0$.

Specifically, we show that these functions are either Type I or Type II class of functions as defined in Section 4 of Andrews (1994). To start, we have $\mathbf{x}_{g,c_g}(\tau) = \sum_{i=1}^{s}\mathbb{I}(c_g=i)\mathbf{x}_{g,i}(\tau)$. The indicator function $\mathbb{I}(c_g=i)$ is a Type I(b) class function. The component $[\mathbf{z}_{g,i}^\top, -p_{g,i}]$ is a Type I(a) class function. In addition, we can verify that the component $p_{g,i}(\tau) = \max(0, p_{g,i} - \tau)$ is 1-Lipschitz in τ where $|p_{g,i}(\tau) - p_{g,i}(\tau')| \leq |\tau - \tau'|$ for any $p_{g,i}$ and any $\tau, \tau' \in \Theta_\tau$. Thus, $p_{g,i}(\tau)$ is a Type II class function. Furthermore, by the boundedness of the feature space, we know that there exists a constant C such that $\|\mathbf{x}_{g,i}(\tau)\| \leq C$ for all $i = 1, \dots, s$ and $\tau \in \Theta_\tau$. Thus, by Theorem 2 of Andrews (1994), $\mathbf{x}_{g,c_g}(\tau)$ satisfies Pollard's entropy condition with a constant envelope function sC.

For $\mathbb{I}(p_{g,c_g} > \tau)$, we have $\mathbb{I}(p_{g,c_g} > \tau) = \sum_{i=1}^s \mathbb{I}(p_{g,i} > \tau)$. The class of indicator functions $\{\mathbb{I}(\cdot > \tau)\}$ is a Type I(b) class of function, by Theorem 2 of Andrews (1994), we have that $\mathbb{I}(p_{g,c_g} > \tau)$ satisfies Pollard's entropy condition with a constant envelope function s.

Finally, for $q_i(\theta; \mathbf{w}_g) := \exp(\mathbf{x}_{g,i}(\tau)^\top \eta) / \sum_{j=1}^s \exp(\mathbf{x}_{g,j}(\tau)^\top \eta)$, we show it is C-Lipschitz continuous with respect to θ and thus a Type II class of functions. Define $f_i(y_1, y_2, \dots, y_s)$ as the first order derivative of the function $\exp(y_i)/\sum_{j=1}^s \exp(y_j)$. By Taylor expansion, we have $|q_i(\theta; \mathbf{w}_g) - q_i(\theta'; \mathbf{w}_g)| = f_i(y_1^*, y_2^*, \dots, y_s^*)^\top [(\mathbf{x}_{g,1}(\tau)^\top \eta, \dots, \mathbf{x}_{g,s}(\tau)^\top \eta) - (\mathbf{x}_{g,1}(\tau')^\top \eta', \dots, \mathbf{x}_{g,s}(\tau')^\top \eta')]$, where y_i^* is between $\mathbf{x}_{g,i}(\tau)^\top \eta$ and $\mathbf{x}_{g,i}(\tau')^\top \eta'$. By compactness of the parameter space Θ and the boundedness of the feature space, there exists a constant $C > 0$ such that $\|f_i(y_1^*, \dots, y_s^*)\| < C$ for all $\mathbf{w}_g$ and $\theta \in \Theta$. The result follows by elementary algebra and the fact that $p_{g,i}(\tau)$ is a 1-Lipschitz function of τ. Thus, by Theorem 2 of Andrews (1994), $q_i(\theta; \mathbf{w}_g)$ satisfies Pollard's entropy condition with a constant envelope function C.

By Theorem 1 of Andrews (1994), we have $\{v_G(\cdot) : G \geq 1\}$ is stochastically equicontinuous. □

Proof of Theorem 4. The proof of Theorem 4 focuses on the analysis of the log-likelihood $L_G(\theta)$, for both the scenarios of under-estimation and over-estimation of the order K^o.

We first define notations used in the proof. Denote $\Theta_\tau^K = \{\tau_1, \dots, \tau_K | \tau_i - \tau_{i-1} > \delta, i = 1, \dots, K+1\}$ as the parameter space for the K thresholds, where we define $\tau_0 = p_{\min}$, $\tau_{K+1} = p_{\max}$ and $\delta > 0$ is an arbitrarily small positive constant. Given K thresholds of price sensitivities, $\tau_1 < \tau_2 < \cdots < \tau_K$, the nominal utility of item i for consumer g is given by

$$u_{g,i} = \alpha^\top \mathbf{z}_{g,i} - \beta_1 p_{g,i} + \sum_{k=1}^K \lambda_k (p_{g,i} - \tau_k)_+ = \alpha^\top \mathbf{z}_{g,i} - \beta_1 p_{g,i} + \sum_{k=1}^K \lambda_k p_{g,i}(\tau_k), \text{ for } i = 1, 2, \dots, s. \tag{EC.3}$$

Collect the model parameters as $\lambda^K = (\lambda_1, \dots, \lambda_K)$ and $\tau^K = (\tau_1, \dots, \tau_K)$ and denote $\theta^K = (\alpha, \beta_1, \lambda^K, \tau^K)$. We further denote the implied choice probability as $\mathbb{P}(c_g | \mathbf{z}_g, \mathbf{p}_g; \theta^K) = \mathbb{P}(c_g | \mathbf{z}_g, \mathbf{p}_g; \alpha, \beta_1, \lambda^K, \tau^K)$ and denote $\widehat{\theta}^K$ as the MLE, which is the maximizer of the log-likelihood function $L_G^K(\theta^K) = 1/G \sum_{g=1}^G \log(\mathbb{P}(c_g | \mathbf{z}_g, \mathbf{p}_g; \theta^K))$. In addition, we collect the true parameters for the true model with K^o thresholds as $\lambda^o = (\lambda_1^o, \dots, \lambda_{K^o}^o)$ and $\tau^o = (\tau_1^o, \dots, \tau_{K^o}^o)$. Denote $\theta^o = (\alpha^o, \beta_1^o, \lambda^o, \tau^o)$ and $\widehat{\theta}$ as its MLE.

We first show that probability of under-estimation of the order K^o goes to 0 as $G \to \infty$. For $K = 0, 1, \dots, K^o - 1$, denote θ_K^o as the pseudo true parameter, which minimizes the Kullback-Leibler distance between the true model and the piecewise MNL model with K thresholds. Based on the same arguments as that in the proof of Theorem 2, we can show that for any fixed $K = 0, 1, \dots, K^o - 1$, the MLE $\widehat{\theta}^K$ converges to the pseudo true parameter θ_K^o. By uniform law of large number, we have that the log-likelihood function $L_G^K(\widehat{\theta}^K) \to_p \mathbb{E}(\log(\mathbb{P}(c_g | \mathbf{z}_g, \mathbf{p}_g; \theta_K^o))$. In addition, we have at $K = K^o$, $L_G(\widehat{\theta}) \to_p \mathbb{E}(\log(\mathbb{P}(c_g | \mathbf{z}_g, \mathbf{p}_g; \theta^o))$. By Assumption 2 and information inequality, we have that $\mathbb{E}(\log(\mathbb{P}(c_g | \mathbf{z}_g, \mathbf{p}_g; \theta_K^o)) < \mathbb{E}(\log(\mathbb{P}(c_g | \mathbf{z}_g, \mathbf{p}_g; \theta^o))$. Since the penalty of BIC is of order $\log G/G \to 0$, it is clear that under-estimation of order K approaches probability 0 as $G \to \infty$.

The proof of over-estimation error is more involved. The reason is that, for $K = K^o + 1, \dots, K_{max}$, the true model with $K = K^o$ is nested as a special case. To see this, consider the concrete example of $K = K^o + 1$. Set $\lambda_*^{K^o+1} = (\lambda^o, 0)$ and $\tau_*^{K^o+1} = (\tau^o, \tau_{K^o+1})$ for any $\tau_{K^o+1} > \tau_{K^o}^o$, it is easy to see that

$$\mathbb{P}(c_g | \mathbf{z}_g, \mathbf{p}_g; \alpha^o, \beta_1^o, \lambda_*^{K^o+1}, \tau_*^{K^o+1}) \equiv \mathbb{P}(c_g | \mathbf{z}_g, \mathbf{p}_g; \theta^o), \text{ for any } c_g, \mathbf{z}_g, \mathbf{p}_g.$$

More generally, for a subset $S \subseteq \{1, 2, \dots, K\}$, denote $S^c = \{1, 2, \dots, K\} \setminus S$ and denote $\tau_S^K = \{\tau_i^K | i \in S\}$ and $\lambda_S^K = \{\lambda_i^K | i \in S\}$. The above nesting phenomenon holds for any $\theta^K = (\alpha^o, \beta_1^o, \lambda^K, \tau^K)$ where there exists a set S

of cardinality K^o such that $\tau_S^K = \tau^o$, $\lambda_S^K = \lambda^o$, and $\lambda_{S^c}^K = 0$. In particular, the threshold parameter $\tau_{S^c}^K$ becomes irrelevant as $\lambda_{S^c}^K = 0$, which is an important property that will be used later. Given the MLE $\widehat{\theta}^K = (\widehat{\alpha}, \widehat{\beta}_1, \widehat{\lambda}^K, \widehat{\tau}^K)$. Define $\widehat{S} = \cup_{i=1}^{K^o} \{\arg\min_{j=1,\ldots,K} |\widehat{\tau}_j^K - \tau_i^o|\}$ and define $\widehat{\lambda}^{Ko}$ as the parameter vector with $\widehat{\lambda}_{\widehat{S}}^{Ko} = \lambda^o$ and $\widehat{\lambda}_{\widehat{S}^c}^{Ko} = 0$. Based on the same arguments (i.e. stochastic continuity) as that in the proof of Theorem 3, we can readily show that

$$|\widehat{\alpha} - \alpha^o| = O_p(1/\sqrt{G}),\ |\widehat{\beta}_1 - \beta_1^o| = O_p(1/\sqrt{G}),\ |\widehat{\lambda}^K - \widehat{\lambda}^{Ko}| = O_p(1/\sqrt{G}),\ \text{and } |\widehat{\tau}_{\widehat{S}}^K - \tau^o| = O_p(1/\sqrt{G}). \qquad \text{(EC.4)}$$

In other words, the convergence rate of MLE is still $O_p(1/\sqrt{G})$.

We now analyze the behavior of the log-likelihood function $L_G^K(\theta^K)$. Note that the log-likelihood function $L_G^K(\theta^K)$ is not differentiable with respect to τ^K, which prevents the use of Taylor expansion. To bypass this difficulty, we employ the "deletion" strategy in Feder (1975). Specifically, it is easy to see that $L_G^K(\theta^K)$ is twice continuously differentiable in a neighborhood covering both $\widehat{\tau}_{\widehat{S}}^K$ and τ^o, if there is no observation $(c_g, \mathbf{z}_g, \mathbf{p}_g)$ such that $p_{g,i}$ lives between $\widehat{\tau}_{\widehat{S},i}^K$ and τ_i^o for some $i = 1, 2, \ldots, s$.

To achieve so, following the strategy in Feder (1975), we artificially remove all observations $(c_g, \mathbf{z}_g, \mathbf{p}_g)$ such that $p_{g,i}$ lives in $\left[\tau_i^o - \log\log G/\sqrt{G}, \tau_i^o + \log\log G/\sqrt{G}\right]$ for some $i = 1, 2, \ldots, s$. Note that since $|\widehat{\tau}_{\widehat{S}}^K - \tau^o| = O_p(1/\sqrt{G})$, this ensures that with probability going to 1, all observations with $p_{g,i}$ living between $\widehat{\tau}_{\widehat{S},i}^K$ and τ_i^o will be removed.

Denote R_G as the set of indices for all observations such that $p_{g,i}$ lives in $\left[\tau_i^o - \log\log G/\sqrt{G}, \tau_i^o + \log\log G/\sqrt{G}\right]$ for some $i = 1, 2, \ldots, s$. Denote $L_{G1}^K(\theta^K)$ as the log-likelihood function with all observations in R_G removed and define $L_{G2}^K(\theta^K) = L_G^K(\theta^K) - L_{G1}^K(\theta^K)$. To summarize, $L_{G1}^K(\theta^K)$ now is a twice continuously differentiable function with respect to $(\widehat{\alpha}, \widehat{\beta}_1, \widehat{\lambda}^K, \widehat{\tau}_{\widehat{S}}^K)$ thanks to the deletion strategy.

We show that this deletion strategy has minimal impact on the log-likelihood function by establishing an upper bound on the number of observations removed, i.e. the cardinality of R_G. Specifically, for each observation $(c_g, \mathbf{z}_g, \mathbf{p}_g)$, the probability of $p_{g,i}$ falling into $\left[\tau_i^o - \log\log G/\sqrt{G}, \tau_i^o + \log\log G/\sqrt{G}\right]$ for some $i = 1, \ldots, s$ is upper bounded by $C \log\log G/\sqrt{G}$ for some constant $C > 0$. Thus, by Chernoff's inequality on sum of independent Bernoulli random variables (Vershynin 2018), we have that $\mathbb{P}(|R_G| \geq 2C\sqrt{G}\log\log G) \leq 2\exp(-c\sqrt{G}\log\log G) \to 0$ for some constant $c > 0$. Thus, with probability approaching 1, the deletion strategy at most removes $2C\sqrt{G}\log\log G$ observations among all G observations. We are now ready to bound the over-estimation error. Specifically, we show that

$$L_G^K(\widehat{\theta}^K) - L_G(\widehat{\theta}) \leq L_G^K(\widehat{\theta}^K) - L_G(\theta^o) \leq O_p(\log\log G/G),$$

where the first inequality follows from the definition of MLE $\widehat{\theta}$. In other words, the gain in log-likelihood brought by a larger $K > K^o$ is of smaller order than the added BIC penalty $\log G/G$, which then concludes the proof.

Denote $\widehat{\eta}^K = (\widehat{\alpha}, \widehat{\beta}_1, \widehat{\lambda}^K)$ and $\widehat{\eta}^{Ko} = (\alpha^o, \beta_1^o, \widehat{\lambda}^{Ko})$. By the above discussion, we have

$$\begin{aligned}
&L_G^K(\widehat{\theta}^K) - L_G(\theta^o) \\
=&L_G^K(\widehat{\eta}^K, \widehat{\tau}_{\widehat{S}}^K, \widehat{\tau}_{\widehat{S}^c}^K) - L_G^K(\widehat{\eta}^{Ko}, \tau^o, \widehat{\tau}_{\widehat{S}^c}^K) \\
=&L_{G1}^K(\widehat{\eta}^K, \widehat{\tau}_{\widehat{S}}^K, \widehat{\tau}_{\widehat{S}^c}^K) - L_{G1}^K(\widehat{\eta}^{Ko}, \tau^o, \widehat{\tau}_{\widehat{S}^c}^K) + L_{G2}^K(\widehat{\eta}^K, \widehat{\tau}_{\widehat{S}}^K, \widehat{\tau}_{\widehat{S}^c}^K) - L_{G2}^K(\widehat{\eta}^{Ko}, \tau^o, \widehat{\tau}_{\widehat{S}^c}^K),
\end{aligned}$$

where the first equality follows from the nesting phenomenon. In the following, we further bound terms in the last equation. First, by a Taylor expansion of $L_{G1}^K(\widehat{\eta}^{Ko}, \tau^o, \widehat{\tau}_{\widehat{S}^c}^K)$ around $(\widehat{\eta}^K, \widehat{\tau}_{\widehat{S}}^K, \widehat{\tau}_{\widehat{S}^c}^K)$, we have that

$$\begin{aligned}
&L_{G1}^K(\widehat{\eta}^K, \widehat{\tau}_{\widehat{S}}^K, \widehat{\tau}_{\widehat{S}^c}^K) - L_{G1}^K(\widehat{\eta}^{Ko}, \tau^o, \widehat{\tau}_{\widehat{S}^c}^K) \\
=&\frac{\partial}{\partial\eta\partial\tau_{\widehat{S}}} L_{G1}^K(\widehat{\eta}^K, \widehat{\tau}_{\widehat{S}}^K, \widehat{\tau}_{\widehat{S}^c}^K) \begin{bmatrix} \widehat{\eta}^K - \widehat{\eta}^{Ko} \\ \widehat{\tau}_{\widehat{S}}^K - \tau^o \end{bmatrix} - \frac{1}{2}[\widehat{\eta}^K - \widehat{\eta}^{Ko}, \widehat{\tau}_{\widehat{S}}^K - \tau^o] \frac{\partial^2}{\partial\eta^2\partial\tau_{\widehat{S}}^2} L_{G1}^K(\widetilde{\eta}^K, \widetilde{\tau}_{\widehat{S}}^K, \widehat{\tau}_{\widehat{S}^c}^K) \begin{bmatrix} \widehat{\eta}^K - \widehat{\eta}^{Ko} \\ \widehat{\tau}_{\widehat{S}}^K - \tau^o \end{bmatrix} \\
=&O_p(\sqrt{G}\log\log G/G) O_p(1/\sqrt{G}) + O_p(1/\sqrt{G}) O_p(1/\sqrt{G}) = O_p(\log\log G/G),
\end{aligned}$$

where $\widetilde{\eta}^K$ and $\widetilde{\tau}_{\widehat{S}}^K$ are between $\widehat{\eta}^K$ and $\widehat{\eta}^{K_o}$, and $\widehat{\tau}_{\widehat{S}}^K$ and τ^o, respectively.

To establish the second equality, first, by the same argument as that in the proof of Theorem 3 (Step I), we can show that the approximate first order condition holds for $L_G^K(\widehat{\theta}^K)$, i.e. we have $m_G^K(\widehat{\theta}^K) = o_p(1/\sqrt{G})$, where $m_G^K(\widehat{\theta}^K)$

denotes the left derivative of $L_G^K(\widehat{\theta}^K)$ with respect to $(\widehat{\eta}^K, \widehat{\tau}_{\widehat{S}}^K)$. By definition of $L_{G1}^K(\widehat{\theta}^K)$ and the compactness of parameter space and boundedness of feature space, we have

$$\frac{\partial}{\partial\eta\partial\tau_{\widehat{S}}} L_{G1}^K(\widehat{\eta}^K, \widehat{\tau}_{\widehat{S}}^K, \widehat{\tau}_{\widehat{S}^c}^K) - m_G^K(\widehat{\theta}^K) = O_p(\sqrt{G}\log\log G)/G,$$

where the difference stems from the removal of up to $O(\sqrt{G}\log\log G)$ observations in R_G. Thus, we have $\frac{\partial}{\partial\eta\partial\tau_{\widehat{S}}} L_{G1}^K(\widehat{\eta}^K, \widehat{\tau}_{\widehat{S}}^K, \widehat{\tau}_{\widehat{S}^c}^K) = O_p(\sqrt{G}\log\log G/G)$. Similarly, by the compactness of parameter space and boundedness of feature space, we have that $\frac{\partial^2}{\partial\eta^2\partial\tau_{\widehat{S}}^2} L_{G1}^K(\widetilde{\eta}^K, \widetilde{\tau}_{\widehat{S}}^K, \widehat{\tau}_{\widehat{S}^c}^K) = O_p(1)$. Together with (EC.4), the equality follows.

For the second term, by the compactness of parameter space and boundedness of feature space, it is clear that $\log(\mathbb{P}(c_g|\mathbf{z}_g, \mathbf{p}_g; \theta^K))$ is C-Lipschitz continuous with respect to θ^K for any $(c_g, \mathbf{z}_g, \mathbf{p}_g)$ with some constant $C > 0$. Thus,

$$L_{G2}^K(\widehat{\eta}^K, \widehat{\tau}_{\widehat{S}}^K, \widehat{\tau}_{\widehat{S}^c}^K) - L_{G2}^K(\widehat{\eta}^{Ko}, \tau^o, \widehat{\tau}_{\widehat{S}^c}^K) \le \frac{1}{G} C\sqrt{G}\log\log G \times \Big(|\widehat{\eta}^K - \widehat{\eta}^{Ko}| + |\widehat{\tau}_{\widehat{S}}^K - \tau^o|\Big) = O_p(\log\log G/G).$$

We conclude the proof by noting that there is a finite number of $K > K^o$ and thus all probability bounds above continue to hold uniformly for all $K = K^o + 1, \ldots, K_{max}$. □

Proof of Proposition 5. Under the piecewise MNL model, given x, τ_i, for each $i \in S$, we have

$$h_i(x, p_i) = \begin{cases} h_{1i}(x, p_i), & p_i < \tau_i, \\ h_{1i}(x, p_i) = h_{2i}(x, p_i), & p_i = \tau_i, \\ h_{2i}(x, p_i), & p_i > \tau_i. \end{cases}$$

Recall that $h_{1i}(x, p_i) = (p_i - c_i - x)\exp\big(\alpha_i - \beta_{1i}p_i\big)$ and $h_{2i}(x, p_i) = (p_i - c_i - x)\exp\big(\alpha_i - \beta_{1i}p_i + (\beta_{1i} - \beta_{2i})(p_i - \tau_i)\big)$ are two unimodal functions with maximizers at $m_{1i}(x) = x + c_i + 1/\beta_{1i}$ and $m_{2i}(x) = x + c_i + 1/\beta_{2i}$ respectively.

Under Type I piecewise MNL model, we have $m_{1i}(x) < m_{2i}(x)$. Then, we will show the different structures of $h_i(x, p_i)$ depending on the relative position between τ_i and the above two maximizers. (a) If $\tau_i \le m_{1i}(x)$, $h_i(x, p_i)$ (i.e., $h_{1i}(x, p_i)$) first increases in p_i for $p_i \le \tau_i$, and then $h_i(x, p_i)$ (i.e., $h_{2i}(x, p_i)$) increases in p_i for $\tau_i < p_i \le m_{2i}(x)$, and decreases in p_i thereafter. (b) If $\tau_i \ge m_{2i}(x)$, $h_i(x, p_i)$ (i.e., $h_{1i}(x, p_i)$) first increases in p_i for $p_i \le m_{1i}(x)$, and then it decreases in p_i for $m_{1i}(x) < p_i < \tau_i$, and thereafter $h_i(x, p_i)$ (i.e., $h_{2i}(x, p_i)$) decreases in p_i. (c) If $m_{1i}(x) < \tau_i < m_{2i}(x)$, we have that $h_i(x, p_i)$ (i.e., $h_{1i}(x, p_i)$) first increases in p_i for $p_i \le m_{1i}(x)$, and then it decreases in p_i for $m_{1i}(x) < p_i < \tau_i$, and thereafter $h_i(x, p_i)$ (i.e., $h_{2i}(x, p_i)$) increases again in p_i for $\tau_i < p_i \le m_{2i}(x)$, and decreases in p_i for $p_i > m_{2i}(x)$. Thus, $h_i(x, p_i)$ is unimodal for $\tau_i \le m_{1i}(x)$ or $\tau_i \ge m_{2i}(x)$, while bimodal for $m_{1i}(x) < \tau_i < m_{2i}(x)$. Now, we complete the proof of Proposition 5. □

Proof of Theorem 5. Following the analysis in the proof of Proposition 5, under Type I piecewise MNL, $h_i(x) = \max_{p_i} h_i(x, p_i) = \max_{p_i}(p_i - c_i - x)\exp\big(\alpha_i - \beta_{1i}p_i + (\beta_{1i} - \beta_{2i})(p_i - \tau_i)_+\big)$ can be explicitly given by

$$h_i(x) = \begin{cases} h_{2i}(x), & \tau_i \le m_{1i}(x), \\ \max\{h_{1i}(x), h_{2i}(x)\}, & m_{1i}(x) < \tau_i < m_{2i}(x), \\ h_{1i}(x), & \tau_i \ge m_{2i}(x), \end{cases}$$

where $h_{1i}(x) = 1/\beta_{1i}\exp\big(\alpha_i - \beta_{1i}(x + c_i) - 1\big)$, and $h_{2i}(x) = 1/\beta_{2i}\exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x + c_i) - 1\big)$.

Regarding the case of $\tau_i \in (m_{1i}(x), m_{2i}(x))$, there exists a unique threshold for any given x, denoted by $\tau_i(x) \in (m_{2i}(x), m_{1i}(x))$. At $\tau_i = \tau_i(x) = (\log(\beta_{1i}) - \log(\beta_{2i}))/(\beta_{1i} - \beta_{2i}) + c_i + x$, we have $h_{1i}(x) = h_{2i}(x)$. More specifically, if $m_{1i}(x) < \tau_i < \tau_i(x)$, we have $h_{2i}(x) > h_{1i}(x)$; otherwise, we have $h_{2i}(x) < h_{1i}(x)$ if $\tau_i(x) < \tau_i < m_{2i}(x)$. As a result, for $\tau_i < \tau_i(x)$, $h_i(x, p_i)$ achieves its maximum $h_{2i}(x)$ at $p_i = m_{2i}(x)$; for $\tau_i > \tau_i(x)$, $h_i(x, p_i)$ achieves its maximum $h_{1i}(x)$ at $p_i = m_{1i}(x)$; otherwise, we have $h_i(x) = h_{1i}(x) = h_{2i}(x)$, which implies that both $p_i = m_{1i}(x)$ and $p_i = m_{2i}(x)$ are optimal for product i with $\tau_i = \tau_i(x)$. Immediately, we specify $h_i(x)$ as follows:

$$h_i(x) = \begin{cases} h_{2i}(x), & \tau_i < \tau_i(x), \\ h_{1i}(x) = h_{2i}(x), & \tau_i = \tau_i(x), \\ h_{1i}(x), & \tau_i > \tau_i(x). \end{cases}$$

Next, we focus on finding the optimal x^*. We observe that there exists a one-to-one positive relationship between $\tau_i(x)$ and x by the expression $\tau_i(x) = (\log(\beta_{1i}) - \log(\beta_{2i}))/(\beta_{1i} - \beta_{2i}) + c_i + x$. Then, for each product i, we have

$$h_i(x) = \begin{cases} h_{1i}(x), & x < \tau_i - c_i - (\log(\beta_{1i}) - \log(\beta_{2i}))/(\beta_{1i} - \beta_{2i}), \\ h_{1i}(x) = h_{2i}(x), & x = \tau_i - c_i - (\log(\beta_{1i}) - \log(\beta_{2i}))/(\beta_{1i} - \beta_{2i}), \\ h_{2i}(x), & x > \tau_i - c_i - (\log(\beta_{1i}) - \log(\beta_{2i}))/(\beta_{1i} - \beta_{2i}). \end{cases}$$

We observe that each $h_i(x)$ is a continuous decreasing function with respect to x. Then, the optimal total expected profit can be found by solving the following single variable function regarding x: $x = \sum_{i \in S} h_i(x)$. Notice that the left hand-side(LHS) of the above equation is strictly increasing in x, while the right hand-side(RHS) is decreasing in x. Therefore, we can find the unique solution, denoted by x^*, which is the optimal expected profit of the firm, by such as binary search. Immediately, we can obtain the optimal price(s) for each product under Type I piecewise MNL. □

Proof of Proposition 6. Similar to the reasoning process in the proof of Proposition 5, we will characterize the structure of $h_i(x, p_i)$ under Type II piecewise MNL model in three cases, based on the relative position between τ_i and maximizers $m_{1i}(x)$ and $m_{2i}(x)$. Note that we have $m_{1i}(x) > m_{2i}(x)$ under Type II piecewise MNL model. (a) If $\tau_i \le m_{2i}(x)$, $h_i(x, p_i)$ (i.e., $h_{1i}(x, p_i)$) first increases in p_i for $p_i \le \tau_i$, and then $h_i(x, p_i)$ (i.e., $h_{2i}(x, p_i)$) increases in p_i for $\tau_i < p_i \le m_{2i}(x)$, and decreases in p_i thereafter. (b) If $\tau_i \ge m_{1i}(x)$, $h_i(x, p_i)$ (i.e., $h_{1i}(x, p_i)$) first increases in p_i for $p_i \le m_{1i}(x)$, and then it decreases in p_i for $m_{1i}(x) < p_i < \tau_i$, and thereafter $h_i(x, p_i)$ (i.e., $h_{2i}(x, p_i)$) decreases in p_i. (c) If $m_{2i}(x) < \tau_i < m_{1i}(x)$, we have that $h_i(x, p_i)$ (i.e., $h_{1i}(x, p_i)$) first increases in p_i for $p_i \le \tau_i$, and then $h_i(x, p_i)$ (i.e., $h_{2i}(x, p_i)$) decreases in p_i for $p_i > \tau_i$. Thus, $h_i(x, p_i)$ is unimodal in all three cases under the Type II piecewise MNL model, but with different maximizers. Now, we complete the proof of Proposition 6. □

Proof of Theorem 6. Following the analysis in the proof of Proposition 6, and based on the comparison between τ_i and two maximizers $m_{1i}(x) = x + c_i + 1/\beta_{1i}$ and $m_{2i}(x) = x + c_i + 1/\beta_{2i}$, $h_i(x)$ can be explicitly given by

$$h_i(x) = \begin{cases} h_{2i}(x), & \tau_i \le m_{2i}(x), \\ (\tau_i - c_i - x)\exp(\alpha_i - \beta_{1i}\tau_i), & m_{2i}(x) < \tau_i < m_{1i}(x), \\ h_{1i}(x), & \tau_i \ge m_{1i}(x), \end{cases}$$

under Type II piecewise MNL, where $h_{1i}(x) = 1/\beta_{1i} \exp\big(\alpha_i - \beta_{1i}(x + c_i) - 1\big)$, and $h_{2i}(x) = 1/\beta_{2i} \exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x + c_i) - 1\big)$. Then, we find each $h_i(x)$ is a continuous and decreasing function in x under Type II piecewise MNL, so by the similar arguments, we can obtain the optimal total profit x^* and immediately the optimal price p_i^* for each product $i \in S$, as shown in Theorem 6. □

Proof of Proposition 7. (1) Under Type I piecewise MNL, the optimal total expected profit can be expressed as

$$x^* = \sum_{i \in S_1} 1/\beta_{1i} \exp\big(\alpha_i - \beta_{1i}(x^* + c_i) - 1\big) + \sum_{i \in S_2} 1/\beta_{2i} \exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x^* + c_i) - 1\big),$$

where S_1 includes the products charging optimal price $p_i^* = m_{1i}(x^*) = x^* + c_i + 1/\beta_{1i}$ and S_2 includes the products charging optimal price $p_i^* = m_{2i}(x^*) = x^* + c_i + 1/\beta_{2i}$, with a little bit of notation abuse. Note that for the products with two optimal price solutions, they are grouped in S_2.

For any $\tau_i > \tau_i(x^*) = (\log(\beta_{1i}) - \log(\beta_{2i}))/(\beta_{1i} - \beta_{2i}) + c_i + x^*$, the corresponding product i is included in S_1, so we immediately obtain that $\partial x^*/\partial \tau_i = 0$. Therefore, the optimal price for each product is unchanged, i.e., $\partial p_i^*/\partial \tau_i = 0$ and $\partial p_j^*/\partial \tau_i = 0$, for $i, j \in S$. For $\tau_i < \tau_i(x^*) = (\log(\beta_{1i}) - \log(\beta_{2i}))/(\beta_{1i} - \beta_{2i}) + c_i + x^*$, we have the following derivative:

$$\frac{\partial x^*}{\partial \tau_i} = -\frac{(\beta_{1i} - \beta_{2i})}{\beta_{2i}} \cdot \frac{\exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x^* + c_i) - 1\big)}{\mathcal{A}_I} < 0,$$

where $\mathcal{A}_I = 1 + \sum_{i \in S_1} \exp\big(\alpha_i - \beta_{1i}(x^* + c_i) - 1\big) + \sum_{i \in S_2} \exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x^* + c_i) - 1\big)$. Thus, as the price threshold τ_i decreases, the total profit x^* increases, and the optimal price increases for each product, i.e., $\partial x^*/\partial \tau_i < 0$, and $\partial p_i^*/\partial \tau_i < 0$, $\partial p_j^*/\partial \tau_i < 0$.

Note that the total expected profit is non-differentiable at $\tau_i = \tau_i(x^*) = (\log(\beta_{1i}) - \log(\beta_{2i}))/(\beta_{1i} - \beta_{2i}) + c_i + x^*$ because its left derivative is not equal to its right derivative although both exist.

(2) Under Type II piecewise MNL model, the optimal total expected profit can be expressed as

$$\begin{aligned} x^* \;=\; & \sum_{i \in S_1} 1/\beta_{1i} \exp\big(\alpha_i - \beta_{1i}(x^* + c_i) - 1\big) + \sum_{i \in S_2} 1/\beta_{2i} \exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x^* + c_i) - 1\big) \\ & + \sum_{i \in S_3} (\tau_i - c_i - x^*) \exp\big(\alpha_i - \beta_{1i}\tau_i\big), \end{aligned}$$

where S_1 includes the products charging optimal price $p_i^* = m_{1i}(x^*) = x^* + c_i + 1/\beta_{1i}$, S_2 includes the products charging optimal price $p_i^* = m_{2i}(x^*) = x^* + c_i + 1/\beta_{2i}$, and S_3 includes the products charging optimal price $p_i^* = \tau_i$.

For any $\tau_i > m_{1i}(x^*)$, the corresponding product i is included in S_1, so we immediately obtain that $\partial x^*/\partial \tau_i = 0$. Therefore, the optimal price for each product is unchanged, i.e., $\partial p_i^*/\partial \tau_i = 0$ and $\partial p_j^*/\partial \tau_i = 0$, for $i, j \in S$.

For $\tau_i < m_{2i}(x^*)$ and $m_{2i}(x^*) \le \tau_i \le m_{1i}(x^*)$ we have the following derivatives, respectively:

$$\begin{aligned} \frac{\partial x^*}{\partial \tau_i} &= -\frac{(\beta_{1i} - \beta_{2i})}{\beta_{2i}} \cdot \frac{\exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x^* + c_i) - 1\big)}{\mathcal{A}_{II}} > 0; \\ \frac{\partial x^*}{\partial \tau_i} &= \big[1 - \beta_{1i}(\tau_i - c_i - x^*)\big] \cdot \frac{\exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x^* + c_i) - 1\big)}{\mathcal{A}_{II}} > 0; \end{aligned}$$

where $\mathcal{A}_{II} = 1 + \sum_{i \in S_1} \exp\big(\alpha_i - \beta_{1i}(x^* + c_i) - 1\big) + \sum_{i \in S_2} \exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x^* + c_i) - 1\big) + \sum_{i \in S_3} \exp\big(\alpha_i - \beta_{1i}\tau_i\big)$. Thus, for $\tau_i < m_{1i}(x^*)$, as the price threshold τ_i increases, the optimal total profit x^* increases, and the optimal price increases for each product, i.e., $\partial x^*/\partial \tau_i > 0$, and $\partial p_i^*/\partial \tau_i > 0$, $\partial p_j^*/\partial \tau_i > 0$. Note that the optimal total expected profit is differentiable at $\tau_i = m_{1i}(x^*)$ and $\tau_i = m_{2i}(x^*)$. ☐

Proof of Proposition 8. (1) Following the analysis in the proof of Proposition 7, under Type I piecewise MNL, for any $\tau_i > \tau_i(x^*)$, we take the derivatives of the optimal total profit with respect to β_{1i} and β_{2i} respectively, and have

$$\frac{\partial x^*}{\partial \beta_{1i}} = -\frac{1 + \beta_{1i}(x^* + c_i)}{\beta_{1i}^2} \cdot \frac{\exp\big(\alpha_i - \beta_{1i}(x^* + c_i) - 1\big)}{\mathcal{A}_I} < 0, \quad \frac{\partial x^*}{\partial \beta_{2i}} = 0.$$

Then, for $\tau_i < \tau_i(x^*)$, we have

$$\begin{aligned} \frac{\partial x^*}{\partial \beta_{1i}} &= -\frac{\tau_i}{\beta_{2i}} \cdot \frac{\exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x^* + c_i) - 1\big)}{\mathcal{A}_I} < 0, \\ \frac{\partial x^*}{\partial \beta_{2i}} &= -\frac{1/\beta_{2i} - \big(\tau_i - (x^* + c_i)\big)}{\beta_{2i}} \cdot \frac{\exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x^* + c_i) - 1\big)}{\mathcal{A}_I}. \end{aligned}$$

Since $\tau_i < \tau_i(x^*) = (\log(\beta_{1i}) - \log(\beta_{2i}))/(\beta_{1i} - \beta_{2i}) + c_i + x^*$, that is, $\tau_i - (c_i + x^*) < (\log(\beta_{1i}) - \log(\beta_{2i}))/(\beta_{1i} - \beta_{2i})$. By comparing $1/\beta_{2i}$ and $(\log(\beta_{1i}) - \log(\beta_{2i}))/(\beta_{1i} - \beta_{2i})$, we have

$$1/\beta_{2i} - (\log(\beta_{1i}) - \log(\beta_{2i}))/(\beta_{1i} - \beta_{2i}) = \frac{\beta_{1i} - \beta_{2i} - \beta_{2i}(\log(\beta_{1i}) - \log(\beta_{2i}))}{\beta_{2i}(\beta_{1i} - \beta_{2i})}.$$

We find the numerator $\beta_{1i} - \beta_{2i} - \beta_{2i}(\log(\beta_{1i}) - \log(\beta_{2i}))$ decreases in β_{2i}, and it achieves its minimum 0 when $\beta_{2i} = \beta_{1i}$. Then, we can show $\frac{\partial x^*}{\partial \beta_{2i}} < 0$ in this case. According to the structure of the optimal prices for all products, the effects of price sensitivities on the optimal prices can be straightforwardly obtained.

(2) Under Type II piecewise MNL, for any $\tau_i > m_{2i}(x^*)$, it is straightforward to have $\frac{\partial x^*}{\partial \beta_{2i}} = 0$. Then, we take the derivatives of the optimal total profit with respect to β_{1i} for $m_{2i}(x^*) < \tau_i < m_{1i}(x^*)$ and $\tau_i > m_{1i}(x^*)$, and have

$$\frac{\partial x^*}{\partial \beta_{1i}} = -\tau_i(\tau_i - c_i - x^*) \cdot \frac{\exp\big(\alpha_i - \beta_{1i}\tau_i\big)}{\mathcal{A}_{II}} < 0, \quad \frac{\partial x^*}{\partial \beta_{1i}} = -\frac{1 + \beta_{1i}(x^* + c_i)}{\beta_{1i}^2} \cdot \frac{\exp\big(\alpha_i - \beta_{1i}(x^* + c_i) - 1\big)}{\mathcal{A}_{II}} < 0.$$

Then, for $\tau_i < m_{2i}(x^*)$, we have

$$\begin{aligned} \frac{\partial x^*}{\partial \beta_{1i}} &= -\frac{\tau_i}{\beta_{2i}} \cdot \frac{\exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x^* + c_i) - 1\big)}{\mathcal{A}_{II}} < 0, \\ \frac{\partial x^*}{\partial \beta_{2i}} &= -\frac{1/\beta_{2i} - \big(\tau_i - (x^* + c_i)\big)}{\beta_{2i}} \cdot \frac{\exp\big(\alpha_i - (\beta_{1i} - \beta_{2i})\tau_i - \beta_{2i}(x^* + c_i) - 1\big)}{\mathcal{A}_{II}} < 0. \end{aligned}$$

Immediately, according to the structure of the optimal prices for all products, the effects of price sensitivities on the optimal prices can be straightforwardly obtained. $\square$

Proof of Proposition 9. The proof can be derived directly from Theorem 6 and is therefore omitted to avoid repetition. $\square$

Proof of Theorem 7. Suppose $p_i < p_j \le \tau$. The proof for the other direction is analogous. Define $f(p) = \exp(\alpha_i - p) + \exp(\alpha_j - p)$. Note that $f(p)$ is monotonic and $f(p_i) > \exp(\alpha_i - p_i) + \exp(\alpha_j - p_j) > f(p_j)$. Therefore there exists a $p^* \in (p_i, p_j)$ such that $f(p^*) = \exp(\alpha_i - p^*) + \exp(\alpha_j - p^*) = \exp(\alpha_i - p_i) + \exp(\alpha_j - p_j)$. Importantly, this implies that $p^* \le \tau$ as well. Define $g(p) = \exp(p)$, which is convex. We have that

$$
\begin{aligned}
&\frac{\exp(\alpha_i - p_i)}{\exp(\alpha_i - p_i) + \exp(\alpha_j - p_j)} g(p_i) + \frac{\exp(\alpha_j - p_j)}{\exp(\alpha_i - p_i) + \exp(\alpha_j - p_j)} g(p_j) \\
=&\frac{\exp(\alpha_i) + \exp(\alpha_j)}{\exp(\alpha_i - p_i) + \exp(\alpha_j - p_j)} = \frac{\exp(\alpha_i) + \exp(\alpha_j)}{\exp(\alpha_i - p^*) + \exp(\alpha_j - p^*)} = \exp(p^*) = g(p^*).
\end{aligned}
$$

By the convexity of $g(p)$, we know that

$$
p^* > \frac{\exp(\alpha_i - p_i)}{\exp(\alpha_i - p_i) + \exp(\alpha_j - p_j)} p_i + \frac{\exp(\alpha_j - p_j)}{\exp(\alpha_i - p_i) + \exp(\alpha_j - p_j)} p_j,
$$

which implies that $p^* \cdot (\exp(\alpha_i - p^*) + \exp(\alpha_j - p^*)) > \exp(\alpha_i - p_i) p_i + \exp(\alpha_j - p_j) p_j$. Then, we can immediately derive the results in Theorem 7. $\square$

§1.1. Additional Details of the Approximation Algorithm

In this section, we provide details of the block dynamic programming (DP) component for the approximation algorithm proposed in Section 4.5 of the main text, which is used to solve the assortment optimization problem (18). Our block DP algorithm is adapted from Désir et al. (2022), which solves capacitated assortment optimization under mixture of MNL. However, important modifications are made to accommodate the block cardinality constraints in (18).

[Block Dynamic Programming] For notational simplicity, in the following, we assume the cost $c_i = 0$ for all $i \in S$ in (18). We remark that this can be done without loss of generality, as our block DP can readily handle the case of general c_i's. We denote $w_{m,o} = \sum_{i \in S} a_{m,i}(b_u^{(i)})$ and $W_{m,o} = \sum_{i \in S} a_{m,i}(b_l^{(i)})$. Denote $r_{m,o} = \sum_{i \in S} b_l^{(i)} a_{m,i}(b_u^{(i)})$ and $R_{m,o} = \sum_{i \in S} b_u^{(i)} a_{m,i}(b_l^{(i)})$.

For each m, we define the geometric grids $\Gamma_{m,\epsilon} = \{r_{m,o}(1+\epsilon)^l, l = 0, 1, \ldots, L_{m1}\}$ and $\Delta_{m,\epsilon} = \{w_{m,o}(1+\epsilon)^l, l = 0, 1, \ldots, L_{m2}\}$, where $L_{m1} = \lceil \log(R_{m,o}/r_{m,o}) / \log(1+\epsilon) \rceil$ and $L_{m2} = \lceil \log(W_{m,o}/w_{m,o}) / \log(1+\epsilon) \rceil$. Note that $\Gamma_{m,\epsilon}$ and $\Delta_{m,\epsilon}$ are used to discretize/approximate the ranges of $\sum_{i \in S} \sum_{n=1}^{N_i} p_n^{(i)} a_{m,i}(p_n^{(i)}) x_n^{(i)}$ and $\sum_{i \in S} \sum_{n=1}^{N_i} a_{m,i}(p_n^{(i)}) x_n^{(i)}$ for segment m in (18), respectively. Denote $\Gamma_\epsilon = \Gamma_{1,\epsilon} \times \Gamma_{2,\epsilon} \times \cdots \times \Gamma_{M,\epsilon}$ and $\Delta_\epsilon = \Delta_{1,\epsilon} \times \cdots \times \Delta_{M,\epsilon}$. Given a guess $(\mathbf{h}, \mathbf{g}) \in \Gamma_\epsilon \times \Delta_\epsilon$, we discretize the coefficients as

$$
\tilde{r}_{m,i,n} = \left\lfloor \frac{p_n^{(i)} a_{m,i}(p_n^{(i)})}{\epsilon h_m / s} \right\rfloor \text{ and } \tilde{w}_{m,i,n} = \left\lceil \frac{a_{m,i}(p_n^{(i)})}{\epsilon g_m / s} \right\rceil, \text{ for all } m, i, n.
$$

Denote $\tilde{\mathbf{r}}_{i,n} = (\tilde{r}_{1,i,n}, \tilde{r}_{2,i,n}, \ldots, \tilde{r}_{M,i,n})$ and $\tilde{\mathbf{w}}_{i,n} = (\tilde{w}_{1,i,n}, \tilde{w}_{2,i,n}, \ldots, \tilde{w}_{M,i,n})$.

Recall $\mathbf{x}_i = (x_1^{(i)}, \ldots, x_{N_i}^{(i)}) \in \{0,1\}^{N_i}$ is the decision variable for item i and $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_s)$. Recall the offer set $S = \{1, 2, \ldots, s\}$. In the following, we give a block dynamic program that can find a feasible assortment $\mathbf{x}$ (given its existence) such that for all $m \in [M]$, we have that

$$
\sum_{i \in S} \sum_{n=1}^{N_i} p_n^{(i)} a_{m,i}(p_n^{(i)}) x_n^{(i)} \ge h_m \text{ and } \sum_{i \in S} \sum_{n=1}^{N_i} a_{m,i}(p_n^{(i)}) x_n^{(i)} \le g_m; \text{ and } \sum_{n=1}^{N_i} x_n^{(i)} = 1 \text{ for all } i \in S. \tag{EC.5}
$$

Set $I = \lfloor s/\epsilon \rfloor - s$ and $J = \lceil s/\epsilon \rceil + s$. For a positive integer z, denote $[z] = \{1, \ldots, z\}$. Denote $[I]_+ = \{0\} \cup [I]$. For each $(\mathbf{i}, \mathbf{j}, k) \in [I]_+^M \times [J]^M \times [s]$, define $F(\mathbf{i}, \mathbf{j}, k)$ as the feasibility indicator of the existence of an assortment $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ for items $\{1, 2, \ldots, k\}$ such that for all $m \in [M]$,

$$
\sum_{i=1}^{k} \sum_{n=1}^{N_i} \tilde{r}_{m,i,n} x_n^{(i)} \ge i_m \text{ and } \sum_{i=1}^{k} \sum_{n=1}^{N_i} \tilde{w}_{m,i,n} x_n^{(i)} \le j_m,
$$

where we denote $\mathbf{i}=(i_1,i_2,\dots,i_M)$ and $\mathbf{j}=(j_1,j_2,\dots,j_M)$.

To preserve the block cardinality constraint for each $i\in S=\{1,2,\dots,s\}$, we compute $F(\mathbf{i},\mathbf{j},k)$ for $(\mathbf{i},\mathbf{j},k)\in[I]_+^M\times[J]^M\times[s]$ using the following recursion

$$F(\mathbf{i},\mathbf{j},k+1)=\min\left\{F(\mathbf{i}-\tilde{\mathbf{r}}_{k+1,n},\mathbf{j}-\tilde{\mathbf{w}}_{k+1,n},k),\ n=1,\dots,N_{k+1}\right\}. \tag{EC.6}$$

The initial condition is set as

$$F(\mathbf{i},\mathbf{j},1)=\begin{cases}0 & \text{if } \mathbf{i}\le\tilde{\mathbf{r}}_{1,n} \text{ and } \mathbf{j}\ge\tilde{\mathbf{w}}_{1,n} \text{ for some } n=1,\dots,N_1,\\ \infty & \text{otherwise.}\end{cases}$$

For boundary conditions, by the definition of F, for any $\mathbf{i}$ with negative entries, we set $F(\mathbf{i},\mathbf{j},k)=F(\mathbf{i}\vee\mathbf{0},\mathbf{j},k)$. Moreover, for any $\mathbf{j}$ with non-positive entries, we set $F(\mathbf{i},\mathbf{j},k)=\infty$.

The dynamic program with recursion (EC.6) moves forward in a block-wise fashion where each block consists of different prices for the same item i (collected in $\mathcal{P}_i$). It thus naturally imposes the block cardinality constraint that for each item one and only one price can be picked. Note that this can be easily extended to the scenario of joint pricing and assortment optimization by further including $F(\mathbf{i},\mathbf{j},k)$ itself in the minimization set of the recursion in (EC.6). Assuming $F(\mathbf{i},\mathbf{j},k+1)=0$, denote $n_{k+1}=\arg\min_{n=1,\dots,N_{k+1}}F(\mathbf{i}-\tilde{\mathbf{r}}_{k+1,n},\mathbf{j}-\tilde{\mathbf{w}}_{k+1,n},k)$ (when there is a tie, choose n_{k+1} arbitrarily). Based on the recursion (EC.6), the associated assortment (i.e. price decisions) of $F(\mathbf{i},\mathbf{j},k+1)$ is the union of $p_{n_{k+1}}^{(k+1)}$ (i.e. $x_{n_{k+1}}^{(k+1)}=1$) and the associated assortment of $F(\mathbf{i}-\tilde{\mathbf{r}}_{k+1,n_{k+1}},\mathbf{j}-\tilde{\mathbf{w}}_{k+1,n_{k+1}},k)$.

Denote $\mathbf{I}=(I,\dots,I)$ and $\mathbf{J}=(J,\dots,J)$. Proposition EC.1 states that the proposed block DP algorithm can find an assortment $\mathbf{x}$ that satisfies (EC.5) with a controllable approximation error.

Proposition EC.1. *For any guess* $\mathbf{h},\mathbf{g}$*, if there exists a feasible* $\dot{\mathbf{x}}$ *such that* (EC.5) *is satisfied, then* $F(\mathbf{I},\mathbf{J},s)=0$*. Moreover, if* $F(\mathbf{I},\mathbf{J},s)=0$*, then the DP finds an* $\mathbf{x}$ *such that for all* $m\in[M]$*, we have that*

$$\sum_{i\in S}\sum_{n=1}^{N_i}p_n^{(i)}a_{m,i}(p_n^{(i)})x_n^{(i)}\ge h_m(1-2\epsilon)\ \textit{and}\ \sum_{i\in S}\sum_{n=1}^{N_i}a_{m,i}(p_n^{(i)})x_n^{(i)}\le g_m(1+2\epsilon);\ \textit{and}\ \sum_{n=1}^{N_i}x_n^{(i)}=1\ \textit{for all}\ i\in S.$$

Denote $\mathbf{x}(\mathbf{h},\mathbf{g})$ as the assortment returned by the block DP for $(\mathbf{h},\mathbf{g})$ in $\Gamma_\epsilon\times\Delta_\epsilon$ (given that $F(\mathbf{I},\mathbf{J},s)=0$) and denote $R(\mathbf{x}(\mathbf{h},\mathbf{g}))$ as its revenue. The optimal assortment (and thus the optimal price vector) given by our algorithm is therefore defined as the $\mathbf{x}^*$ that achieves $\max_{(\mathbf{h},\mathbf{g})\in\Gamma_\epsilon\times\Delta_\epsilon}R(\mathbf{x}(\mathbf{h},\mathbf{g}))$. As shown in Theorem 8 of the main text, $R(\mathbf{x}^*)\ge(1-10\epsilon)R^*$, where the proof is built on Proposition 11 and Proposition EC.1.

§1.1.1. Technical proofs of our approximation algorithm

Proof of Proposition 11. By definition of $\mathcal{W}_{m,i}$, for each $a_{m,i}(p_i^*)$, there is an $l_{m,i}\in\{0,1,\dots,L_{m,i}-1\}$ such that

$$w_{m,i}(1+\epsilon)^{l_{m,i}}\le a_{m,i}(p_i^*)\le w_{m,i}(1+\epsilon)^{l_{m,i}+1}\wedge W_{m,i}.$$

Define $p_{m,i}^o=a_{m,i}^{-1}\left(w_{m,i}(1+\epsilon)^{l_{m,i}}\right)$, we have that $p_{m,i}^o\in\mathcal{P}_{m,i}$ and $p_{m,i}^o\ge p_i^*$. Define $p_i^o=\min_{m=1,\dots,M}p_{m,i}^o$, we have that $p_i^o\ge p_i^*$ and $p_i^o\in\mathcal{P}_i$. Importantly, note that by the monotone decreasing and left-continuous property of $a_{m,i}(\cdot)$, it holds that $w_{m,i}(1+\epsilon)^{l_{m,i}}\le a_{m,i}(p_{m,i}^o)\le a_{m,i}(p_i^o)\le a_{m,i}(p_i^*)$ for all $m\in\{1,2,\dots,M\}$. Therefore, we have that

$$1\le\frac{a_{m,i}(p_i^*)}{a_{m,i}(p_i^o)}\le\frac{a_{m,i}(p_i^*)}{w_{m,i}(1+\epsilon)^{l_{m,i}}}\le 1+\epsilon.$$

This implies that

$$R(p_1^o,p_2^o,\dots,p_n^o)=\sum_{m=1}^M\theta_m\frac{\sum_{i\in S}(p_i^o-c_i)a_{m,i}(p_i^o)}{1+\sum_{i\in S}a_{m,i}(p_i^o)}\ge\sum_{m=1}^M\theta_m\frac{\sum_{i\in S}(p_i^*-c_i)a_{m,i}(p_i^*)/(1+\epsilon)}{1+\sum_{i\in S}a_{m,i}(p_i^*)}=\frac{1}{1+\epsilon}R^*\ge(1-\epsilon)R^*.$$

This completes the proof. □

Proof of Proposition EC.1. Suppose there exists a feasible $\dot{\mathbf{x}}$ such that (EC.5) is satisfied. Define

$$\mathbf{i}_s=\sum_{i\in S}\sum_{n=1}^{N_i}\tilde{\mathbf{r}}_{i,n}\dot{x}_n^{(i)}\ \text{and}\ \mathbf{j}_s=\sum_{i\in S}\sum_{n=1}^{N_i}\tilde{\mathbf{w}}_{i,n}\dot{x}_n^{(i)}.$$

Note that for all $m \in [M]$, at $\mathbf{i}_s$, we have

$$\sum_{i\in S}\sum_{n=1}^{N_i} \tilde{r}_{m,i,n}\dot{x}_n^{(i)} = \sum_{i\in S}\sum_{n=1}^{N_i} \left\lfloor \frac{p_n^{(i)} a_{m,i}(p_n^{(i)})}{\epsilon h_m/s} \right\rfloor \dot{x}_n^{(i)} \ge \sum_{i\in S}\sum_{n=1}^{N_i} \left\{ \frac{p_n^{(i)} a_{m,i}(p_n^{(i)})}{\epsilon h_m/s} - 1 \right\} \dot{x}_n^{(i)} \ge s/\epsilon - s \ge I,$$

where the second to last inequality follows from (EC.5) and that $\sum_{i\in S}\sum_{n=1}^{N_i} \dot{x}_n^{(i)} = s$.

Similarly, for all $m \in [M]$, at $\mathbf{j}_s$, we have

$$\begin{aligned}\sum_{i\in S}\sum_{n=1}^{N_i} \tilde{w}_{m,i,n}\dot{x}_n^{(i)} &= \sum_{i\in S}\sum_{n=1}^{N_i} \left\lceil \frac{a_{m,i}(p_n^{(i)})}{\epsilon g_m/s} \right\rceil \dot{x}_n^{(i)} \le \sum_{i\in S}\sum_{n=1}^{N_i} \left\{ \frac{a_{m,i}(p_n^{(i)})}{\epsilon g_m/s} + 1 \right\} \dot{x}_n^{(i)} \\ &\le s/\epsilon + s \le J.\end{aligned}$$

Therefore, we have that $\mathbf{i}_s \ge \mathbf{I}$ and $\mathbf{j}_s \le \mathbf{J}$, which implies that $F(\mathbf{I}, \mathbf{J}, s) = 0$. Moreover, using the same argument, it is easy to verify that $F(\mathbf{i}_k \wedge \mathbf{I}, \mathbf{j}_k, k) = 0$ for $k = 1, 2, \ldots, s-1$, where we define

$$\mathbf{i}_k = \sum_{i=1}^{k}\sum_{n=1}^{N_i} \tilde{\mathbf{r}}_{i,n}\dot{x}_n^{(i)} \text{ and } \mathbf{j}_k = \sum_{i=1}^{k}\sum_{n=1}^{N_i} \tilde{\mathbf{w}}_{i,n}\dot{x}_n^{(i)}.$$

On the other hand, suppose $F(\mathbf{I}, \mathbf{J}, s) = 0$ and denote $\mathbf{x}$ as the associated assortment. For all $m \in [M]$, we have

$$s/\epsilon - 1 - s \le I \le \sum_{i\in S}\sum_{n=1}^{N_i} \tilde{r}_{m,i,n} x_n^{(i)} = \sum_{i\in S}\sum_{n=1}^{N_i} \left\lfloor \frac{p_n^{(i)} a_{m,i}(p_n^{(i)})}{\epsilon h_m/s} \right\rfloor x_n^{(i)} \le \sum_{i\in S}\sum_{n=1}^{N_i} \left\{ \frac{p_n^{(i)} a_{m,i}(p_n^{(i)})}{\epsilon h_m/s} \right\} x_n^{(i)},$$

which, by simple algebra, implies that $\sum_{i\in S}\sum_{n=1}^{N_i} p_n^{(i)} a_{m,i}(p_n^{(i)}) x_n^{(i)} \ge (1-2\epsilon) h_m$. Similarly, for all $m \in [M]$, we have

$$s/\epsilon + 1 + s \ge J \ge \sum_{i\in S}\sum_{n=1}^{N_i} \tilde{w}_{m,i,n} x_n^{(i)} = \sum_{i\in S}\sum_{n=1}^{N_i} \left\lceil \frac{a_{m,i}(p_n^{(i)})}{\epsilon g_m/s} \right\rceil x_n^{(i)} \ge \sum_{i\in S}\sum_{n=1}^{N_i} \left\{ \frac{a_{m,i}(p_n^{(i)})}{\epsilon g_m/s} \right\} x_n^{(i)}$$

which implies that $\sum_{i\in S}\sum_{n=1}^{N_i} a_{m,i}(p_n^{(i)}) x_n^{(i)} \le (1+2\epsilon) g_m$. This completes the proof. □

Proof of Theorem 8. For the price vector $(p_1^o, p_2^o, \ldots, p_s^o)$ in Proposition 11, denote $\dot{\mathbf{x}}$ as its associated assortment in $\mathcal{P}$. Clearly, there exist $\mathbf{l}_1 = (l_{1,1}, l_{1,2}, \ldots, l_{1,M})$ and $\mathbf{l}_2 = (l_{2,1}, l_{2,2}, \ldots, l_{2,M})$ such that for all $m \in [M]$, we have

$$\begin{aligned} r_{m,o}(1+\epsilon)^{l_{1,m}} &\le \sum_{i\in S} p_i^o a_{m,i}(p_i^o) = \sum_{i\in S}\sum_{n=1}^{N_i} p_n^{(i)} a_{m,i}(p_n^{(i)}) \dot{x}_n^{(i)} \le r_{m,o}(1+\epsilon)^{l_{1,m}+1}, \\ w_{m,o}(1+\epsilon)^{l_{2,m}} &\le \sum_{i\in S} a_{m,i}(p_i^o) = \sum_{i\in S}\sum_{n=1}^{N_i} a_{m,i}(p_n^{(i)}) x_n^{(i)} \le w_{m,o}(1+\epsilon)^{l_{2,m}+1}. \end{aligned}$$

Denote $\mathbf{h} = (h_1, \ldots, h_M)$ with $h_m = r_{m,o}(1+\epsilon)^{l_{1,m}}$ and $\mathbf{g} = (g_1, \ldots, g_M)$ with $g_m = w_{m,o}(1+\epsilon)^{l_{2,m}+1}$. By Proposition EC.1, for this pair of $\mathbf{h}, \mathbf{g}$, we have that $F(\mathbf{I}, \mathbf{J}, s) = 0$. Moreover, the associated $\tilde{\mathbf{x}}$ satisfies that for all $m \in [M]$,

$$\sum_{i\in S}\sum_{n=1}^{N_i} p_n^{(i)} a_{m,i}(p_n^{(i)}) \tilde{x}_n^{(i)} \ge h_m(1-2\epsilon) \text{ and } \sum_{i\in S}\sum_{n=1}^{N_i} a_{m,i}(p_n^{(i)}) \tilde{x}_n^{(i)} \le g_m(1+2\epsilon); \text{ and } \sum_{n=1}^{N_i} \tilde{x}_n^{(i)} = 1 \text{ for all } i \in S.$$

Therefore, we have that

$$\begin{aligned} R(\tilde{\mathbf{x}}) &= \sum_{m=1}^{M} \theta_m \sum_{i\in S}\sum_{n=1}^{N_i} \frac{p_n^{(i)} a_{m,i}(p_n^{(i)}) \tilde{x}_n^{(i)}}{1 + \sum_{i\in S}\sum_{n=1}^{N_i} a_{m,i}(p_n^{(i)}) \tilde{x}_n^{(i)}} \ge \sum_{m=1}^{M} \theta_m \frac{h_m(1-2\epsilon)}{1 + g_m(1+2\epsilon)} \\ &\ge \sum_{m=1}^{M} \theta_m \frac{\sum_{i\in S} p_i^o a_{m,i}(p_i^o)(1-2\epsilon)/(1+\epsilon)}{1 + \sum_{i\in S} a_{m,i}(p_i^o)(1+\epsilon)(1+2\epsilon)} \ge (1-8\epsilon) R(\dot{\mathbf{x}}) = (1-8\epsilon) R(p_1^o, p_2^o, \ldots, p_s^o) \ge (1-10\epsilon) R^*, \end{aligned}$$

where the last inequality follows from Proposition 11.

Computational Complexity. Recall that $w_o = \min_m w_{m,o} = \min_m \sum_{i\in S} a_{m,i}(b_u^{(i)})$ and $W_o = \max_m W_{m,o} = \max_m \sum_{i\in S} a_{m,i}(b_l^{(i)})$, and $r_o = \min_m r_{m,o} = \min_m \sum_{i\in S} b_l^{(i)} a_{m,i}(b_u^{(i)})$ and $R_o = \max_m R_{m,o} = \max_m \sum_{i\in S} b_u^{(i)} a_{m,i}(b_l^{(i)})$. The cardinality of Γ_ϵ is upper bounded by $O(L_1^M)$ with $L_1 = \log(R_o/r_o)/\epsilon$ and the cardinality of Δ_ϵ is upper bounded by $O(L_2^M)$ with $L_2 = \log(W_o/w_o)/\epsilon$. Define $w_{min} = \min_{m,i} w_{m,i} = \min_{m,i} a_{m,i}(b_u^{(i)})$ and $W_{max} = \max_{m,i} W_{m,i} = \max_{m,i} a_{m,i}(b_l^{(i)})$. For a given $(\mathbf{h}, \mathbf{g})$, the number of operations of the block DP is upper bounded by $O(s^{2M+1}/\epsilon^{2M+1} M \log(W_{max}/w_{min}))$. Therefore, in total the computational complexity is $O(M \log(R_o/r_o)^M \log(W_o/w_o)^M \log(W_{max}/w_{min}) s^{2M+1}/\epsilon^{4M+1})$.

§2. Additional Remarks

Remark EC.1 (Identification Condition) To build more intuition, in the following, we give a concrete example where Assumption 2(b) holds. We follow the setting and notations in Section 3 and consider a feature-free piecewise MNL model. In particular, we assume each consumer g is faced with an offer set $S_g = \{1, \ldots, s\}$, where the nominal utility of item 1 is $u_{g,1} \equiv 0$ (i.e. we designate item 1 as the outside option) and the nominal utility of item $i \in \{2, 3, \ldots s\}$ is $u_{g,i} = \alpha - \beta_1 p_{g,i} + \lambda(p_{g,i} - \tau)_+$. Denote $\mathbf{p}_g = (p_{g,2}, \ldots, p_{g,s})^\top$. Therefore, the choice probability of consumer g for item i under the piecewise MNL is

$$\mathbb{P}(i|\mathbf{p}_g; \alpha, \beta_1, \lambda) = \frac{\exp(u_{g,i})}{1 + \sum_{j=2}^{s} \exp(u_{g,j})}, \text{ for } i \in \{1, 2, \ldots, s\}. \tag{EC.7}$$

Given G consumers with observations $\{c_g, \mathbf{p}_g\}_{g=1}^{G}$, collect all price points into $P_G := \cup_{g=1}^{G} \{p_{g,2}, p_{g,3}, \ldots, p_{g,s}\}$. Denote P_G^* as the set of *unique* price points in P_G and denote $n = |P_G^*|$ as its cardinality. We re-index the price points in P_G^* as $P_G^* = \{p_1^*, p_2^*, \ldots, p_n^*\}$, where without loss of generality, we assume $p_1^* < p_2^* < \cdots < p_n^*$. Proposition EC.2 shows the piecewise MNL in (EC.7) is identifiable under mild requirements on P_G^* and Θ_τ.

PROPOSITION EC.2 **(Identification)**. *Suppose $n = |P_G^*| \geq 4$ (i.e. the dataset has at least 4 unique price points), for any parameter space $\Theta_\tau \subseteq [p_2^*, p_{n-1}^*]$, the piecewise MNL in* (EC.7) *is identifiable and Assumption 2(b) holds.*

Proof of Proposition EC.2. By Proposition 4, we only need to verify that under the condition of Proposition EC.2, we have for any pair $\{\tau, \tau', \tau \neq \tau'\} \subseteq \Theta_\tau$, the matrix $D(\tau, \tau')$ is of full rank. Recall $D(\tau, \tau')$ is the matrix whose rows are $[(\mathbf{z}_{g,i} - \mathbf{z}_{g,1})^\top, p_{g,i} - p_{g,1}, p_{g,i}(\tau) - p_{g,1}(\tau), p_{g,i}(\tau') - p_{g,1}(\tau')]^\top$ for $g = 1, \ldots, G$ and $i = 1, \ldots, s$.

For the feature-free piecewise MNL in (EC.7), we can write $\mathbf{z}_{g,1} = 0, p_{g,1} = 0$ (recall item 1 is the outside option) and we further have $\mathbf{z}_{g,i} = 1$ for $i = 2, \ldots, s$. Therefore, the rows of $D(\tau, \tau')$ are $[1, p_{g,i}, (p_{g,i} - \tau)_+, (p_{g,i} - \tau')_+]$ for $g = 1, \ldots, G$ and $i = 2, \ldots, s$. By the condition in Proposition EC.2, we have that $D(\tau, \tau')$ has a submatrix of the form

$$\text{sub-}D(\tau, \tau') = \begin{pmatrix} 1 & -p_1^* & 0 & 0 \\ 1 & -p_2^* & 0 & 0 \\ 1 & -p_{n-1}^* & (p_{n-1}^* - \tau)_+ & (p_{n-1}^* - \tau')_+ \\ 1 & -p_n^* & (p_n^* - \tau)_+ & (p_n^* - \tau')_+ \end{pmatrix}$$

By the fact that $p_1^* < p_2^* < p_{n-1}^* < p_n^*$, it is easy to verify that sub-$D(\tau, \tau')$ is of full rank, which implies that $D(\tau, \tau')$ is of full rank for any $\tau \neq \tau' \in \Theta_\tau$. Thus, by Proposition 4, the piecewise MNL in (EC.7) is identifiable. $\square$

Remark EC.2 (Optimization Set) Note that for any $a \in (0, A)$, the only difference between the optimization of $r_a(p, q)$ over $\{(p, q) : p \leq \tau < q\}$ and over $\{(p, q) : p \leq \tau \leq q\}$ occurs when $p^* \leq q^* = \tau$. In such case, we have that

$$\max_{\{(p,q): p \leq \tau < q\}} r_a(p, q) < r_a(p^*, \tau) \leq \frac{p^* \exp(-p^*)a + \tau \exp(-\tau)(A - a)}{1 + \exp(-p^*)a + \exp(-\tau)(A - a)} \leq \max_{p \in [0,\tau]} r_A(p) := \max_{p \in [0,\tau]} \frac{p \exp(-p)A}{1 + \exp(-p)A},$$

where the first inequality follows by assumption, the second inequality follows from the fact that

$$\frac{p^* \exp(-p^*)a}{1 + \exp(-p^*)a} \leq \frac{p^* \exp(-p^*)a + \tau \exp(-\tau)(A - a)}{1 + \exp(-p^*)a + \exp(-\tau)(A - a)},$$

since $p^* \leq q^* = \tau$, and the third inequality follows from the same argument as the one used in the proof of Theorem 7 (i.e. the optimal price of an MNL is a single price point). Therefore, if the optimal price q^* is achieved at τ, we know this policy is not optimal and will be dominated by another policy with a single price between $[0, \tau]$. Luckily, this policy has been considered in the sequence of optimizations, i.e. we set $S_1 = S$ $(a = A)$ and optimize over $p \in [0, \tau]$.

Remark EC.3 (Zero Utility) Take the consideration set model as an example. It is possible that $a_{m,i}(b_u^{(i)}) = 0$ due to the threshold effect. We remark that the proposed approximation algorithm works under such scenarios with simple modifications. In particular, in Step I, redefine $w_{m,i} = \min_{p \in [b_l^{(i)}, b_u^{(i)}]} \{a_{m,i}(p) : a_{m,i}(p) > 0\}$ (note that $w_{m,i} = a_{m,i}(\tau_{m,i})$) and modify the discretized price set as $\mathcal{P}_{m,i} = \{a_{m,i}^{-1}(w) : w \in \mathcal{W}_{m,i}\} \cup \{\tau_{m,i} + \epsilon\}$. With simple modifications of the proof of Proposition 11, it can be shown that the approximation result in Proposition 11 still holds for this modified price set $\mathcal{P}_{m,i}$. In Step II, redefine $w_{m,o}=\min_{i\in S} w_{m,i}$ and $r_{m,o}=\min_{i\in S} b_l^{(i)} w_{m,i}$ and conduct the block DP. The only scenario where this block DP cannot provide performance guarantee is when some segments $\mathcal{M}\subset\{1,2,\ldots,M\}$ are completely priced out (i.e. the attractiveness for all $i\in S$ is 0 for segment $m\in\mathcal{M}$). However, this can be solved via a simple fix where we further run a block DP but only on $\{1,2,\ldots,M\}\setminus\mathcal{M}$ with the price set $\mathcal{P}\cap\{p_i>\tau_{m,i}$ for all $i\in S, m\in\mathcal{M}\}$. Since there are at most 2^M possibilities for $\mathcal{M}$, this modified algorithm can provide the same performance guarantee as that in Theorem 8 with a computational complexity of $O(2^M M\log(R_o/r_o)^M\log(W_o/w_o)^M\log(W_{max}/w_{min})s^{2M+1}/\epsilon^{4M+1})$.

§3. Additional Real Data Analysis

Ketchup Data. The Ketchup dataset consists of the choices of 2798 consumers for 4 brands of ketchups. For each purchase, we have the information of whether each brand is on promotion display, whether each brand is featured on newspaper advertisement, and the price of each brand. For model flexibility, we include an intercept term for each brand. Table EC.2 gives the detailed estimation result for the Ketchup data.

Hotel Data. The hotel data in Bodea et al. (2009) collects booking records from five U.S. properties of a major hotel chain between March 12, 2007 and April 15, 2007. We focus our analysis on the hotel (Hotel1) with largest number of bookings. Following van Ryzin and Vulcano (2015), we preprocess the dataset to remove rooms that account for less than 4% of all transactions. In total, Hotel1 consists of complete booking records of 1272 consumers' choice behavior for 8 different room types (e.g. double beds, king, queen, suite, special type room). For each booking, the data further records the prices of all rooms offered to the consumer. For model flexibility, we further include an intercept term for each room to account for room-specific effects. Table EC.3 gives the estimation result.

Table EC.1 Summary of Estimation on Cracker Data

Model	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_3$	$\widehat{\alpha}_4$	$\widehat{\alpha}_5$	$-\widehat{\beta}_1$	$\widehat{\nu}$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\tau}_1$	$\widehat{\tau}_2$	Time (s)
MNL	-0.662	-0.169	1.793	0.092	0.496	-0.031						1.40
	(0.090)	(0.117)	(0.100)	(0.062)	(0.095)	(0.002)						
csMNL	-0.663	-0.169	1.792	0.092	0.495	-0.031						1.17
	(0.090)	(0.117)	(0.100)	(0.062)	(0.095)	(0.002)						
qMNL	-0.602	-0.117	1.839	0.090	0.464	-0.051	1.04×10^{-4}					4.02
	(0.095)	(0.121)	(0.103)	(0.062)	(0.096)	(0.009)	(4.63×10^{-5})					
rMNL	-0.756	-0.238	1.723	0.087	0.494	0.002		0.028	-0.038			1.80
	(0.091)	(0.108)	(0.090)	(0.061)	(0.095)	(0.001)		(0.003)	(0.003)			
pMNL1	-0.696	-0.253	1.723	0.066	0.138	-0.177		0.154		61.0		60.45
	(0.082)	(0.105)	(0.088)	(0.061)	(0.096)	(0.016)		(0.016)		(0.77)		
pMNL2	-1.225	-0.697	1.270	0.079	-0.035	-0.228		0.259	-0.070	63.0	89.0	377.31
	(0.088)	(0.111)	(0.094)	(0.062)	(0.096)	(0.016)		(0.018)	(0.008)	(0.61)	(2.11)	
pMNL2m	-0.883	-0.292	1.679	0.037	0.010	-0.217		0.216	-0.038	61.5	89.0	436.75
	(0.087)	(0.108)	(0.092)	(0.062)	(0.097)	(0.016)		(0.018)	(0.008)	(0.69)	(3.91)	
pqMNL	-1.232	-0.698	1.216	0.102	0.113	-0.716	4.86×10^{-3}	-0.225	-3.96×10^{-3}	88.9		474.29
	(0.089)	(0.114)	(0.096)	(0.062)	(0.096)	(0.051)	(3.78×10^{-4})	(0.015)	(4.16×10^{-4})	(0.81)		

Notes. The numbers in parentheses report the estimated standard errors. 89.83% of prices are larger than 61 dollars. The quartile of price is (79, 99, 109). $(\widehat{\alpha}_1,\widehat{\alpha}_2,\widehat{\alpha}_3)$ are intercept terms. $(\widehat{\alpha}_4,\widehat{\alpha}_5)$ are for the promotion display and newspaper featuring effects. The estimated thresholds of csMNL for the four cracker brands are $(\widehat{\tau}_1,\ldots,\widehat{\tau}_4)=(129,135,169,115)$.

We conduct the same 5-fold cross-validation to evaluate the out-of-sample prediction and price optimization performance of each model, and report the results averaged over the 5-fold cv in Table EC.4. Note that for price optimization, since different consumers have different offer sets (i.e. a subset of 8 room types), for each purchase, we

Table EC.2 Summary of Estimation on Ketchup Data

Model	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_3$	$\widehat{\alpha}_4$	$\widehat{\alpha}_5$	$-\widehat{\beta}_1$	$\widehat{\nu}$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\tau}_1$	$\widehat{\tau}_2$	Time (s)
MNL	1.354	1.501	2.426	0.876	0.909	-1.402						0.96
	(0.123)	(0.069)	(0.096)	(0.097)	(0.114)	(0.058)						
csMNL	1.346	1.506	2.417	0.877	0.907	-1.396						0.39
	(0.123)	(0.069)	(0.096)	(0.097)	(0.114)	(0.058)						
qMNL	1.321	1.564	2.443	0.852	0.938	-4.059	0.351					3.46
	(0.127)	(0.072)	(0.101)	(0.098)	(0.116)	(0.296)	(0.037)					
rMNL	1.293	1.585	2.504	0.831	1.026	-0.384		1.728	0.004			0.86
	(0.133)	(0.076)	(0.106)	(0.093)	(0.108)	(0.042)		(0.095)	(0.099)			
pMNL1	1.041	1.591	2.221	0.884	0.897	-4.074		3.155		3.03		15.39
	(0.134)	(0.077)	(0.107)	(0.095)	(0.111)	(0.284)		(0.286)		(0.03)		
pMNL2	1.127	1.592	2.280	0.873	0.907	-4.107		3.030	0.655	3.00	4.60	107.63
	(0.136)	(0.077)	(0.109)	(0.095)	(0.111)	(0.285)		(0.290)	(0.360)	(0.04)	(0.27)	
pqMNL	1.091	1.622	2.272	0.852	0.906	1.804	-1.141	3.873	1.244	2.99		132.06
	(0.137)	(0.081)	(0.110)	(0.096)	(0.111)	(4.780)	(1.010)	(1.205)	(1.007)	(0.08)		

Notes. 84.81% of prices are larger than 3.03 dollars. The quartile of price is (3.2, 3.7, 4.6). The average price is 3.86 dollars. $(\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3)$ are intercept terms. $(\widehat{\alpha}_4, \widehat{\alpha}_5)$ are for the promotion display and newspaper featuring effects. The estimated thresholds of csMNL for the four ketchup brands are $(\widehat{\tau}_1, \widehat{\tau}_2, \widehat{\tau}_3, \widehat{\tau}_4) = (5.2, 3.7, 6.1, 4.8)$.

Table EC.3 Summary of Estimation on Hotel Data

Model	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_3$	$\widehat{\alpha}_4$	$\widehat{\alpha}_5$	$\widehat{\alpha}_6$	$\widehat{\alpha}_7$	$-\widehat{\beta}_1$	$\widehat{\nu}$	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\tau}_1$	$\widehat{\tau}_2$
MNL	-15.5	-17.7	-15.1	-12.4	-16.1	-16.4	-9.4	-0.095					
	(0.64)	(0.75)	(0.64)	(0.52)	(0.65)	(0.66)	(0.42)	(0.004)					
csMNL	-15.6	-17.7	-15.1	-12.4	-16.1	-16.4	-9.4	-0.096					
	(0.64)	(0.75)	(0.65)	(0.53)	(0.65)	(0.66)	(0.42)	(0.004)					
qMNL	-14.8	-17.1	-14.3	-11.5	-15.3	-15.5	-8.6	-0.132	4.57×10^{-5}				
	(0.64)	(0.78)	(0.64)	(0.53)	(0.65)	(0.66)	(0.44)	(0.007)	(7.26×10^{-6})				
rMNL	-11.1	-13.4	-10.6	-7.9	-11.5	-12.0	-5.6	-0.104		0.004	-0.048		
	(0.40)	(0.46)	(0.40)	(0.37)	(0.42)	(0.45)	(0.36)	(0.001)		(0.003)	(0.003)		
pMNL1	-15.0	-17.4	-14.5	-11.6	-15.5	-15.8	-8.5	-0.104		0.023		451.0	
	(0.39)	(0.45)	(0.38)	(0.34)	(0.40)	(0.43)	(0.32)	(0.002)		(0.003)		(8.78)	
pMNL2	-15.1	-17.4	-14.6	-11.7	-15.6	-15.9	-8.7	-0.099		-0.014	0.033	399.0	451.0
	(0.39)	(0.45)	(0.39)	(0.35)	(0.40)	(0.44)	(0.32)	(0.003)		(0.008)	(0.008)	(21.0)	(7.70)
pqMNL	-14.9	-17.2	-14.4	-11.6	-15.4	-15.8	-8.6	-0.080	-3.11×10^{-5}	0.036	-1.31×10^{-5}	452.5	
	(0.13)	(0.18)	(0.11)	(0.12)	(0.14)	(0.19)	(0.21)	(0.020)	(2.68×10^{-5})	(0.007)	(3.51×10^{-5})	(7.22)	

Notes. The numbers in parentheses report the estimated standard errors. 30.32% of prices are larger than 451 dollars. The quartile of price is (351, 419, 471). The average price is 421.28 dollars. $(\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_7)$ are intercept terms. The estimated thresholds of csMNL for the eight room types are $(\widehat{\tau}_1, \ldots, \widehat{\tau}_8) = (499, 449, 549, 519, 565.5, 529, 579, 709)$.

randomly select one room type in the offer set as the no-purchase option. Again, all models from the pMNL family provide a notable improvement over MNL, rMNL and qMNL, while csMNL performs the worst. Overall, pMNL1 and pMNL2 are the clear winner with pMNL1 giving slightly better performance.

Comparison with rMNL. For each dataset, we further compute $C_p = \frac{1}{G}\sum_{g=1}^{G} \mathbb{I}(r(S_g, \mathbf{p}) \geq \tau)$, which measures the proportion of consumers where rMNL gives a reference price larger than the threshold τ of pMNL1. Interestingly, for Cracker and Ketchup, we have $C_p = 0.9997$ and $C_p = 0.9937$, while for Hotel, $C_p = 0.4057$. In other words, rMNL almost always gives a reference price larger than τ of pMNL1 for Cracker and Ketchup, which may explain its notable performance gap compared with pMNL1. For Hotel, the reference prices of rMNL aligns more closely with τ of pMNL1, which results in a smaller performance gap in terms of log-likelihood, although the out-of-sample prediction performance of rMNL is still notably worse.

Recall we set $r(S_g, \mathbf{p}) = \sum_{i \in S_g} p_{g,i}/s_g$ for rMNL, i.e. the mean price of all items in the offer set S_g, as it provides the best overall performance among different specifications of $r(S_g, \mathbf{p})$ such as median or lowest price. However, this specification is still significantly less effective than pMNL in terms of in-sample goodness of fit and out-of-sample prediction performance. The above analysis indicates that one reason for this phenomenon can be attributed to the fact that it is more beneficial to estimate the unknown threshold via a data-driven strategy as is in pMNL, which is more adaptive to different datasets, than a pre-specified function as in rMNL.

Table EC.4 Summary of Prediction and Price Optimization on Hotel Data

	Market Share Prediction				Revenue Prediction				O.O.S. Loglik		Price Optimization			
Model	hMAE	Imp.	hRMSE	Imp.	rMAE	Imp.	rRMSE	Imp.	Loglik	Imp.	$b=25$	Perc.	$b=50$	Perc.
MNL	11.49		22.05		45.47		91.18		-215.24		120922	98.6	127997	98.0
csMNL	11.60	-0.94	22.39	-1.52	45.99	-1.13	93.11	-2.07	-252.19	-14.65	111788	91.2	107155	82.0
qMNL	11.20	2.62	21.83	1.00	44.61	1.93	90.18	1.10	-212.40	1.34	121870	99.4	128328	98.2
rMNL	11.34	1.35	22.18	-0.59	45.07	0.87	92.02	-0.91	-210.02	2.75	118739	96.8	126812	97.1
pMNL1	11.06	3.91	21.58	2.20	44.05	3.23	89.05	2.39	-209.45	2.77	122622	100	130625	100
pMNL2	11.04	4.11	21.62	1.97	43.83	3.75	89.18	2.24	-208.73	3.12	122523	99.9	130494	99.9
pqMNL	11.17	2.90	21.69	1.68	44.43	2.35	89.37	2.03	-210.14	2.43	121853	99.4	128382	98.3

Notes. hMAE and hRMSE are reported in the scale of $\times 10^{-2}$. "Imp." shows relative improvement of a given model compared to MNL (in %). "Perc." shows the relative revenue of a model compared to pMNL1 (in %). The revenue from an average consumer is 378.60 dollars. The actual revenue achieved is 96316 dollars.

§4. Additional Examples

EXAMPLE EC.1. Consider $S = \{i, j\}$ with parameters: $\alpha_i = 15$, $\beta_{1i} = 3.5$, $\beta_{2i} = 0.8$; $\alpha_j = 6$, $\beta_{1j} = 1.5$, $\beta_{2j} = 0.5$. Set $c_i = c_j = 0$ and $(\tau_i, \tau_j) = (3.90, 5.5)$. We have the optimal profit $x^* = 3.36$ and $\tau_i(x^*) = 3.90$ and $\tau_j(x^*) = 4.46$. Thus, by Theorem 5, the optimal price for product j is $p_j^* = x^* + c_i + 1/\beta_{1j} = 4.02$ as $\tau_j > \tau_j(x^*)$. On the other hand, since $\tau_i = \tau_i(x^*)$, we have that both $p_i^* = x^* + c_i + 1/\beta_{2i} = 4.61$ and $p_i^* = x^* + c_i + 1/\beta_{1i} = 3.64$ are optimal for product i. In particular, when the firm sets the price at 3.64, the total market share of S is 0.92, and the profits of products i and j are 3.02 and 0.34, respectively. If the firm sets the price at 4.61, the total market share decreases to 0.76, and the profit contributions from products i and j are 2.42 and 0.94, respectively. Under both pricing strategies for product i, the firm can obtain its optimal profit $x^* = 3.36$. $\square$

EXAMPLE EC.2. Consider $S = \{1, 2, 3, 4\}$ with $\alpha_1 = 2.5, \alpha_2 = 2, \alpha_3 = 1.5, \alpha_4 = 0.5$ and $\theta_1 = 0.4, \theta_2 = 0.6$. (a). Set $\tau \geq 2.71$. The optimal solution is $p_1^* = p_2^* = p_3^* = p_4^* = 2.71 \leq \tau$ and $R^* = 1.71$. (b). Set $\tau = 1.90$, $p_1^* = p_2^* = p_3^* = p_4^* = 1.90 = \tau$, and $R^* = 1.51$. (c). Set $\tau = 1.60$, $p_1^* = p_2^* = 1.60 = \tau$, $p_3^* = p_4^* = 2.39$, and $R^* = 1.34$. (d). Set $\tau = 1.40$, $p_2^* = p_4^* = 1.40 = \tau$, $p_1^* = p_3^* = 2.42$, and $R^* = 1.24$. (e). Set $\tau = 0.70$, $p_1^* = p_2^* = p_3^* = p_4^* = 2.71 > \tau$, and $R^* = 1.03$. $\square$

Example EC.2 reveals the optimal profit decreases as τ decreases (since consumers in segment I grow more selective) and the optimal price switches from pricing all products at some $p^* \leq \tau$ (i.e. $S_1 = S$, offer segment I all items) to pricing all products at some $q^* > \tau$ (i.e. $S_1 = \varnothing$, ignore segment I).

References

Vershynin, R. 2018. *High-dimensional Probability: An Introduction with Applications in Data Science*, Cambridge University Press.

Feder, P. I. 1975. On asymptotic distribution theory in segmented regression problems–identified case. *The Annals of Statistics* **3**(1) 49–83.