Demand Estimation With Foldable Menu: A Case Study of China's Tobacco Industry

Hui Liu*
Department of Economics
University of Toronto
Please See Current Version Here

October 30, 2021

*Contact email: yihui@mail.utoronto.ca. I thank Yao Luo, Loren Brandt, and Victor Aguirregabiria for their invaluable guidance on this project. This paper has also benefited from comments from Eduardo Souza-Rodrigues, Yanyou Chen, Peter Morrow, Jie Fang, Hammad Shaikh, Etienne Makdissi, Marc-Antoine Chatelain, Connor Campbell, Heski Bar-Isaac, Matthew Mitchell, as well as the participants at the University of Toronto Industrial Organization and Business Economics seminars. In addition, I would like to thank Geoffrey T. Fong, Anne C.K. Quah, Steve Xu, Gang Meng, and Ce Shang for providing access to International Tobacco Control survey data and insight into the tobacco industry. All errors that remain are my own.

Abstract

Information about product availabilities is necessary for the demand estimation of products with varying availabilities, but such data is not always observable. When firms endogenously choose product assortments to maximize expected profit, we can recover unobservable product availabilities from firms' profit maximization conditions. Searching for optimal assortment[1] can be very time-consuming because the total number of assortments increases exponentially with the number of products. I show firms only need to choose from a foldable menu, wherein the number of assortments equals the number of products. I provide an empirical application of the foldable menu model on China's tobacco industry. My results show that ignoring varying product availabilities leads to underestimated price elasticity (0.38 vs. 1.2) and thus false tax policy implications. Making all cigarettes available has positive effects on consumer welfare (+6.75%) and cigarette sales (+12.11%) but negatively affects wholesale profits (-5.95%).
Keywords: Structural Demand Estimation, Varying Product Availabilities, Profit Maximization Condition, Fordable Menu, Tobacco Industry

JEL codes: D11, D12, L66

1 Introduction

I estimate product demand to predict consumer response to changes in product price, consumer income, or other factors that impact demand. It's a cornerstone of economics that informs optimal firm decision-making choice and enables researchers to assess the social welfare effects of public policy in different market environments.

Many recent studies about demand estimation for differentiated products build on the framework of Berry et al. (1995) (BLP hereafter). Most of these studies implicitly assume that the set of available choices is the same for each consumer. Although this assumption is convenient for model construction, numerous studies (Gruen et al. (2002); Quan and Williams (2018); Misra (2008)) show that this may be a poor approximation of reality for many retail environments. It is well accepted that ignoring the variation in product availability will lead to biased demand estimation and false inference of the impact of public policies. To correct the bias, researchers need to incorporate product availabilities into

[1] Optimal assortment is the assortment that maximizes firm expected profit.

their demand estimations. However, information about product availabilities is not always available.

When product prices are fixed in each market,[2] e.g., due to government regulations in the tobacco industry or wine industry, I show that firms can endogenously choose product assortments to maximize expected profits. Then, I can recover optimal assortments from firms' profit maximization conditions to obtain information about product availabilities.

I show that firms only need to choose assortments from {1}, {1,2},...,{1,2,...,J}, i.e., a "foldable menu," when the products do not have identical price-cost margins and are ordered from high to low unit margin. Firms should only offer product 1 to low price-sensitive consumers and provide more options for higher price-sensitive consumers. This finding is important because searching for optimal assortments from all possible assortments can be very time-consuming, especially when the number of products is large, because the total number of assortments increases exponentially with the number of products.[3] Furthermore, the optimal assortment for each consumer is uniquely determined by the consumer's deterministic price sensitivity[4] and a set of individual-specific cutoffs of deterministic price sensitivity. Estimating the foldable menu model requires only one additional step on top of the BLP framework: classifying each consumer into their individual-specific optimal assortment. Lastly, I prove that the contraction mapping proposed in BLP also works in the foldable menu model and employ it to simplify the estimation of the foldable menu model.

This paper presents several contributions:

1. I provide a new approach for demand estimation with varying product availabilities without relying on observable availabilities. Unlike the existing papers either relying on observable availabilities (Draganska et al. (2009), Ciliberto et al. (2018), Iaria (2014), Li et al. (2018), Musalem (2015), Shah et al. (2015), Aguirregabiria et al.(2020)), or using market-level "All Commodity Value" (ACV)[5] data to simulate product availability

[2]Product prices can vary across different markets.

[3]When there are 100 differentiated products, the total number of assortments is 2^{100}.

[4]This is the portion of consumer price sensitivity that can be deterministically estimated by firms.

[5]"All Commodity Value" is the standard metric that brand managers and other practitioners use to

distributions (Tenn and Yun (2008); Bruno and Vilcassim (2008)), I choose to derive unobservable product availabilities from firms' profit maximization conditions.

2. I show that firms only need to choose from the foldable menu, when the products do not have identical price-cost margins. Since the total number of assortments increases exponentially with the number of products, this finding significantly simplifies the estimation process.

3. I prove that the contraction mapping proposed in BLP also works in my revised framework for demand estimation. Following the BLP model in allowing consumers to have random coefficients for product characteristics, the estimation objective is to find demand parameters that minimize the distance between observed market shares and the model-predicted ones. Searching for parameters altogether is difficult because both predicted market shares and price sensitivity cutoffs (for classifying consumers into their optimal assortments) solve non-linear systems of equations. Therefore, preserving the contraction mapping is key to simplifying the estimation process.

4. I illustrate the estimation method using a simulated dataset and provide an empirical application of the model on China's tobacco industry.

I estimate cigarette demand at the tier level. In China, cigarettes are categorized in Tier I-V from high to low prices. My empirical results show that compared with the empirical price elasticity (1.14) estimated following the exogenous cigarette price increase in 2015, ignoring varying product availabilities leads to underestimated elasticity (0.38). The foldable menu model can correct the estimation bias of the misspecified model and get estimated price elasticity (1.2) close to the empirical one. The intuition is that consumers have inelastic demand for products with few substitutes. We falsely attribute the limited substitutions of products to lower price elasticities if we ignore the limited availabilities of some products.

quantify a product's retail distribution.

Due to the underestimated price elasticities, tax policies based on the model ignoring varying availabilities have false implications.

Using the estimated model parameters, I perform a counterfactual experiment to quantify changes in the sales of each cigarette tier when all tiers are available. Such an experiment effectively disentangles the effect of varying product availability from all other factors - a feat that cannot be accomplished without the foldable menu model. The experiment results show that when all tiers are available, Tiers I and II cigarettes will experience a decrease in sales, Tiers III - V cigarettes will see sales increase, for an overall increase in total cigarette sales by 12.11%. However, wholesale cigarette profits will drop by about 5.95%. Finally, given that all consumers can make their individual-specific optimal choice among all tiers, total consumer welfare will therefore increase by 6.75%.

My work relates to several papers that propose methods to deal with varying product availabilities in demand estimation of differentiated products, but their models differ substantially from mine. There are two types of related literature characterized by the observability of product availabilities. When product availabilities are unobservable, researchers (Tenn and Yun (2008); Bruno and Vilcassim (2008)) rely on market-level ACV data to simulate product availability distributions and treat the distributions as exogenous. Both studies show that incorporating the product availability distributions into the demand estimation can help correct the estimation bias of models that ignore varying availabilities. However, both studies acknowledge that ACV is not an ideal variable for product availability distributions and that assuming the distributions to be exogenous may not be appropriate. When product availabilities are observable, firms choose to offer a product in some markets but not in others because of product selection on unobservable components of demand. For example, firms are more likely to offer products in markets with a higher average willingness to pay and where expected profits are positive. Studies (Draganska et al. (2009), Ciliberto et al. (2018), Iaria (2014), Li et al. (2018), Musalem (2015), Shah et al. (2015), Aguirregabiria et al.(2020)) choose to bring together the demand estimation with a game of market or product entry to solve the issue of product selection on unobservables.

My paper also relates to several studies using firms' profit maximization conditions to recover important unobservable variables. For example, Luo et al. (2018) recovers unknown consumer types, and D'Haultfœuille et al. (2019) recovers unobservable auto sale prices to analyze price discrimination. In this paper, I use firms' profit maximization conditions to recover unobservable product availabilities.

Two other seemingly related but very different papers are Gandhi et al. (2020) and Dubé et al. (2021). They deal with demand estimation with differentiated products when all products are available, but some products have zero sales and thus zero-valued market shares. The BLP framework requires all products included in the estimation to have positive market shares, so conceptually, observing products with zero-valued market share would invalidate BLP as the underlying model of demand. These two studies develop new methods to extend the BLP framework by allowing products to have zero-valued market shares.

The remainder of the paper is organized as follows: Section 2 briefly describes the institutional background of China's tobacco industry. Section 3 introduces my theoretical model and presents the model properties. Section 4 discusses the estimation strategy. Section 5 summarizes China's tobacco industry data, applies the model to estimate cigarette demand in China, and performs post-estimation analyses. Section 6 concludes with recommendations for future research.

2 Institutional Background

Two conditions are necessary to ensure that firms only need to choose from the foldable menu: 1. Firms have no control over product prices but can endogenously choose product assortment to maximize profit. 2. Products do not have identical price-cost margins. The particular market structure of China's tobacco industry shown in Figure 1 and several industry facts make China's tobacco industry ideal to apply for the foldable menu model:

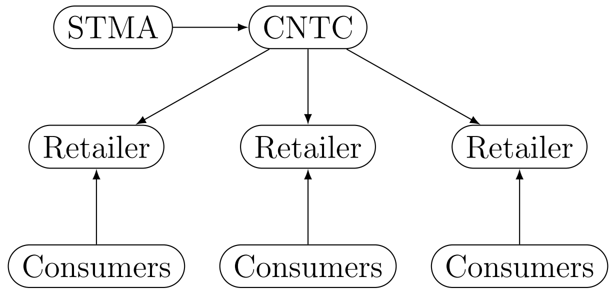

Figure 1: Market structure of China's tobacco industry. STMA sets cigarette prices. CNTC distributes cigarettes to retailers, from where consumers buy cigarettes.

1. The whole industry is supervised by the State Tobacco Monopoly Administration (STMA) and the Chinese National Tobacco Corporation (CNTC), the commercial counterpart of STMA. STMA is the regulatory agency in charge of enforcing related policies for tobacco and cigarette products and directly sets and controls retail cigarette prices. The distribution sector of CNTC (firms in the foldable menu model and cigarette wholesalers in China) is in charge of cigarette sales, has no control over cigarette prices, but can distribute varying cigarette assortments to licensed cigarette retailers.

2. I estimate cigarette demand at tier level. In China, cigarettes are categorized in Tier I-V from high to low prices. High-priced cigarettes have margin rates no lower than low-priced cigarettes at wholesale and retail levels. These industry facts ensure that cigarette tiers do not have identical price-cost margins.

3. I observe that all cigarette tiers are available at the province level, but one well-known fact in China's tobacco markets is that it it hard for smokers to find low-priced cigarettes (under ¥10) in recent years.[6] So it must be that low-priced cigarettes are available in some but not all retailers. This fact is consistent with the theoretical result of the foldable menu model that firms should make low-tier products only available in stores with high price-sensitive consumers (e.g., low-income communities).

[6]https://xw.qq.com/cmsid/20200327A0PNUJ00.

4. Selling cigarettes online is illegal in China. Also, data from the International Tobacco Control (ITC) China surveys show that 3 out of 4 smokers buy cigarettes from local stores, including convenience stores, gas stations, smoke shops and street vendors. The searching and transportation costs may be an important reason why Chinese smokers choose to buy cigarettes locally. Buying cigarettes locally ensures that consumers choose cigarettes within the given assortments rather than search across different stores.

3 Theoretical Model

3.1 The Unobserved Assortment Discrimination Model

I first present my theoretical model. The approach is identical to the BLP model, except for two differences: 1. Product prices are exogenous. 2. Firms endogenously choose product assortments to maximize expected profits from each of a finite number of heterogeneous groups of consumers. So in my model, different groups of consumers face different product assortments, which are unobservable to researchers. While in the standard BLP model, all consumers have the same access to all products.

For readability, I introduce the model using a single market and omit the market subscript at this moment. One firm serves the market and distributes various product assortments, selecting from J differentiated products to I stores. Each store i has a group of consumers with the same known demographies and unknown heterogeneities to the firm, and I use a representative consumer to represent the group of consumers. For abuse of notation, let consumer i be the representative consumer of store i.

The firm has no control over product prices but can distribute any product assortment to each store. The representative consumer can only choose products from the given assortment. Product availabilities vary because different stores have different product assortments. The utility of consumer i from choosing product j is assumed to be a linear function of product characteristics:

$$U_{ij} = X_j \beta_i + \xi_j - \alpha_i p_j + \epsilon_{ij},$$

where X_j is a K-dimensional (row) vector of observable product characteristics, ξ_j represents the valuation of unobserved product characteristics. β_i is the vector of coefficients measuring consumer $i's$ preferences for observable product characteristics, and α_i represents consumer $i's$ price sensitivity. Consumer i chooses product j if and only if

$$U_{ij} \geq U_{ir}, \text{ for } j, r \in \{0, \mathcal{J}_i\},$$

where $\mathcal{J}_i$ is the product assortment available in store i. Alternatives $r \in \{0, \mathcal{J}_i\}$ represent purchases of competing differentiated products. Alternative $r = 0$, or the outside alternative, represents the option of not purchasing any of those products, and I normalize the utility as $u_{i0} = \epsilon_{i0}$.

I make the following parametric assumption about individual heterogeneity. Specifically, the individual coefficients can be decomposed linearly into a mean, an individual-specific deviation from the mean and a deviation related to individual characteristics as follows:

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Pi D_i + \Sigma v_i[7], \quad v_i \sim N(0, I_{K+1}),$$

where K is the dimensionality of the observed product characteristics, D_i is a $d \times 1$ vector of demographic variables, Π is a $(K+1) \times d$ matrix of coefficients that measures how taste characteristics vary with consumer demographics, and Σ is a scaling matrix. This specification allows the individual characteristics to consist of demographics that are "observed" and additional characteristics that are "unobserved," denoted by D_i and v_i, respectively.

[7]Alternative expression:

$$\begin{cases} \alpha_i = \alpha + \Pi_\alpha D_i + \sigma_\alpha v_{i\alpha}, \ v_{i\alpha} \sim N(0, I), \\ \beta_i = \beta + \Pi_\beta D_i + \Sigma_\beta v_{i\beta}, \ v_{i\beta} \sim N(0, I_K). \end{cases}$$

The firm has full information about product price-cost margins π, observed product characteristics X, and consumer demographics D, and can estimate $\{\alpha, \beta, \Pi, \Sigma, \xi\}$ from consumers' repeat purchases. However, the firm only knows the distributions of individual heterogeneities v and idiosyncratic error terms ϵ. As is typical in the literature, ϵ follows the extreme value type I distribution. Let $\vec{J} = \{1, 2, ..., J\}$ be the complete assortment. For any subset $\mathcal{J}_i \subseteq \vec{J}$, the expected profit from consumer i is

$$\Pi_{i,\mathcal{J}_i} = \int \sum_{j \in \mathcal{J}_i} \pi_j \underbrace{\frac{\exp(X_j \beta_i + \xi_j - \alpha_i p_j)}{1 + \sum_{j \in \mathcal{J}_i} \exp(X_j \beta_i + \xi_j - \alpha_i p_j)}}_{\text{choice probability}} dP_v(v_i),$$

where $P_v(v_i)$ is the population density of v_i. The firm should offer the assortment with the highest expected profit to consumer i, i.e., the optimal assortment, denoted as $\mathcal{J}_i^*$. I obtain information about product availabilities after finding $\mathcal{J}_i^*$, i.e., use firms' profit maximization conditions to recover unobservable product availabilities.

Searching for $\mathcal{J}_i^*$ among all possible assortments is computational burdensome because the total number of assortments increases exponentially with the number of products. The following section demonstrates that firms only need to choose from a foldable menu, wherein the number of assortments equals the number of products, when the following two assumptions hold.

3.2 Foldable Menu

Assumption 1 The product prices are exogenous.

Assumption 2 The products do not have identical price-cost margins.

I derive the foldable menu model based on Assumptions 1 and 2, allowing me to theoretically eliminate product assortments that can never be optimal and ensure that each representative consumer's optimal assortment is unique. The unique optimal assortment is the key to recovering unobserved product availabilities in each store.

Assumption 1 holds mostly in highly regulated industries, such as tobacco, wine, etc. For example, in China's tobacco industry, STMA directly sets and controls cigarette prices by setting wholesale and retail margin rates, which I will explain in detail in the following section of data and variables.

Assumption 2 holds for product assortments according to quality levels (or tiers) because mostly top-tier products are more expensive and have higher margin rates than low-tier products. For example, In this study, I estimate China's cigarette demand at the cigarette tier level. Cigarettes are categorized in Tier I-V from high to low prices. High-priced cigarettes have margin rates higher than or equal to low-priced cigarettes at wholesale and retail levels. These industry facts ensure that cigarette tiers do not have identical price-cost margins. A similar example is that a grocery store's assortment may include a store brand of ice cream, a mid-tier national brand (e.g., Dreyer Ice Cream), and a top-tier (premium) national brand (e.g., Haagen Dazs ice cream).

For the cases of product assortments not based on quality levels and when some products do have the same price-cost margin, researchers need to categorize these products into a tier and estimate demand for the tier instead of individual products.

Without loss of generality, let the products be ordered such that $\pi_1 > \pi_2 > ... > \pi_J > \pi_0 = 0$. Based on Assumption 1, there are three Lemmas.

Lemma 1: Any assortment that does not contain top-tier products cannot be optimal.

Proof See appendix A.1.

Intuition Same as in the BLP model, all offered products have positive sales probabilities in my model. Whenever the firm adds a more profitable product to an assortment, the sales probability of the product increases from zero to positive and the overall sales likelihood increases.

Illustration Using the case of three products, there are seven possible non-empty assortments: {1}, {2}, {3}, {1,2}, {1,3}, {2,3}, {1,2,3}. Among them, {2}, {3} and {2,3} are eliminated by Lemma 1, as shown in Figure 2. I draw Figure 2 using the price vector $p = \{4, 3, 2\}$, price-cost margin vectot $\pi = \{2, 1, 0.5\}$, and consumers' preferences captured by the fixed utilities of products $\xi = \{2, 1.6, 1.2\}$. The y values represent expected profits of various product assortments from consumers with different price sensitiviities.[8]

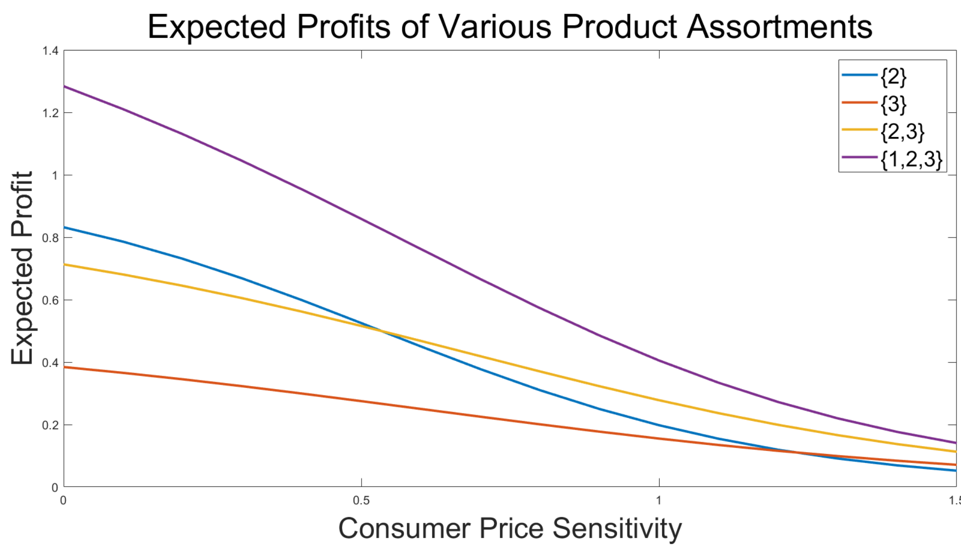

Figure 2: Graph illustration for Lemma 1, with the setup $p = \{4, 3, 2\}$, $\pi = \{2, 1, 0.5\}$, and consumers' preferences captured by the fixed utilities of products $\xi = \{2, 1.6, 1.2\}$.

Lemma 2: Let $\vec{\mathbb{J}} = \{\vec{1}, \vec{2}, ..., \vec{J}\}$, and let $\mathcal{J}_i \in \vec{\mathbb{J}}$ be the optimal assortment among $\vec{\mathbb{J}}$

[8]I draw Figure 2 using the simulated data without including unobserved individual heterogeneity. The underline idea and proofs of the Lemmas are the same for the cases with or without unobserved individual heterogeneity.

to consumer i. There exist $J-1$ cutoffs of $\bar{\alpha}_i$,[9] $\{c_{i1}, c_{i2}, ..., c_{i,J-1}\}$, such that

$$\mathcal{J}_i = \begin{cases} \vec{1}, & \textbf{if } \bar{\alpha}_i < c_{i1}, \\ \vec{m}, & \textbf{if } c_{i,m-1} \leq \bar{\alpha}_i < c_{im} \textbf{ for m = 2,3,...,J-1}, \\ \vec{J}, & \textbf{if } \bar{\alpha}_i \geq c_{i,J-1}. \end{cases}$$

Proof See appendix A.2.

Intuition The firm should only offer the top tier (most profitable) product to low price-sensitive consumers because they are highly likely to buy it. If the firm provides them with more options, their probability of buying the most profitable product will drop. On the other hand, the firm should offer more options to more price-sensitive consumers, because they are more likely to buy nothing if only presented with the most profitable one.

Illustration For consumer i with $\bar{\alpha}_i \geq c_{i2}$, $\{1\}$ and $\{1{,}2\}$ are eliminated by Lemma 2, as shown in Figure 3.

Here, I show the existence of multiple optimal assortments when Assumption 1 does not hold. As shown in Appendix A.2, the price sensitivity cutoffs are the solutions to the set of equations (3). The firm uses the cutoffs to determine the optimal assortment to each consumer i. When two consecutive products have the same price-cost margin, for example, $\pi_1 = \pi_2$, the equation for c_{i1} has no real solution. For other products, e.g., when $\pi_2 = \pi_3$, we will get $c_{i2} = c_{i3}$, so that both $\vec{2}$ and $\vec{3}$ are optimal assortments among $\vec{\mathbb{J}}$ for consumer i with $c_{i1} \leq \bar{\alpha}_i < c_{i2} = c_{i3}$. In this case, we are not able to recover the complete information about product availabilities. Furthermore, selecting one of the multiple optimal assortments

[9] $\bar{\alpha}_i = \alpha + \Pi_\alpha D_i$, consumer i's deterministic price sensitivity.

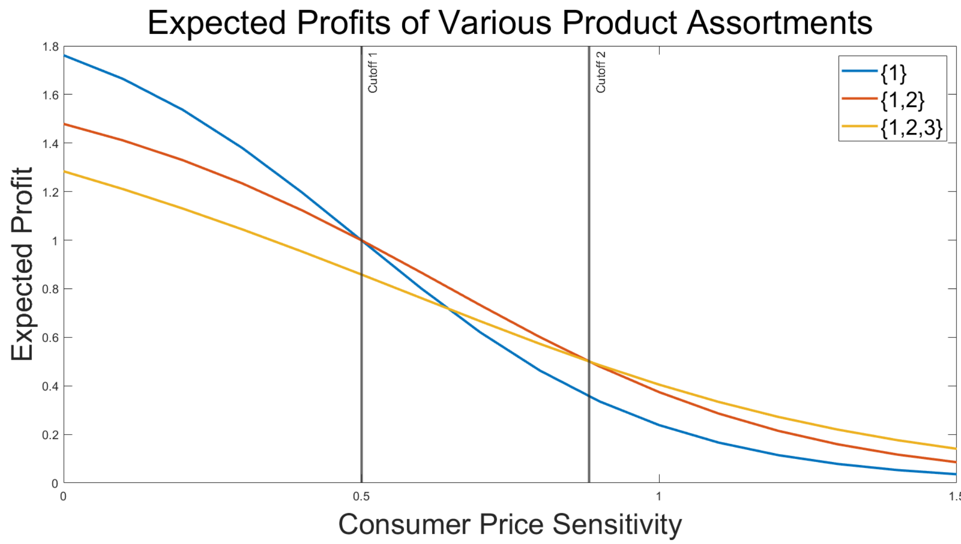

Figure 3: Graph illustration for Lemma 2, with the setup $p = \{4, 3, 2\}$, $\pi = \{2, 1, 0.5\}$, and consumers' preferences captured by the fixed utilities of products $\xi = \{2, 1.6, 1.2\}$.

presents a new problem for model estimation, as different optimal assortments lead to different predicted market shares of the products.

Lemma 3: When $\mathcal{J}_i$ is optimal among $\vec{\mathbb{J}}$, $\mathcal{J}_i$ strictly dominates any subset n_s lacking internal products of $\vec{n} \in \vec{\mathbb{J}}$ (not product 1 or n).

Proof See appendix A.3.

Intuition I get n_s by taking away one or more moderately profitable products from $\vec{n}$, which will make n_s less profitable than $\mathcal{J}_i$ when $\mathcal{J}_i$ is optimal among $\vec{\mathbb{J}}$.

Illustration For consumer i with $\bar{\alpha}_i \geq c_{i2}$, {1,3} is eliminated by Lemma 3, as shown in Figure 4.

Together, these three Lemmas prove that, when $\mathcal{J}_i$ is optimal among $\vec{\mathbb{J}}$ for consumer i, $\mathcal{J}_i$ is also optimal among all possible assortments for consumer i, i.e., $\mathcal{J}_i = \mathcal{J}_i^*$. See Figure 5 for the elimination process when $\bar{\alpha}_i \geq c_{i2}$. This now leads to Theorem 1 as a conclusion of

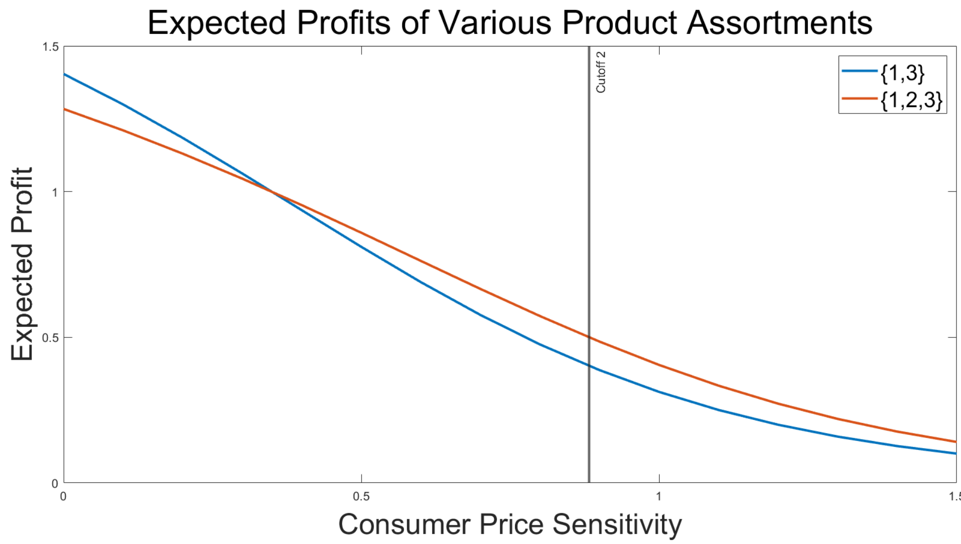

Figure 4: Graph illustration for Lemma 3, with the setup $p = \{4, 3, 2\}$, $\pi = \{2, 1, 0.5\}$, and consumers' preferences captured by the fixed utilities of products $\xi = \{2, 1.6, 1.2\}$.

the three Lemmas.

Theorem 1: Let $\mathcal{J}_i^*$ be the optimal assortment to consumer i. Then

$$\mathcal{J}_i^* = \begin{cases} \vec{1}, & \textbf{if } \bar{\alpha}_i < c_{i1}, \\ \vec{m}, & \textbf{if } c_{i,m-1} \leq \bar{\alpha}_i < c_{im} \textbf{ for m = 2,3,...,J-1}, \\ \vec{J}, & \textbf{if } \bar{\alpha}_i \geq c_{i,J-1}. \end{cases}$$

Given Theorem 1, the number of optimal assortments drops to the number of products. The firm only needs to choose from J assortments of the foldable menu, instead of the number of all product assortments, 2^J.

Unlike the standard BLP model that assumes that all consumers have access to the complete assortment $\vec{J}$ with the choice probability

$$f_{ij}(v_i, D_i, p, X; \theta, \vec{J}) = \frac{\exp(X_j\beta_i + \xi_j - \alpha_i p_j)}{1 + \sum_{k \in \vec{J}} \exp(X_k\beta_i + \xi_k - \alpha_i p_k)},$$

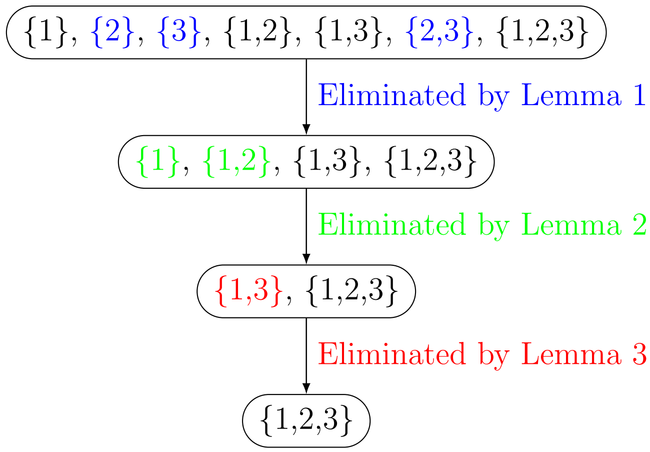

Figure 5: Elimination Process When $\bar{\alpha}_i \geq c_{i2}$

under the foldable menu model, each consumer is given a unique optimal assortment $\mathcal{J}_i^*$, and the choice probability is therefore

$$f_{ij}(v_i, D_i, p, X; \theta, \mathcal{J}_i^*) = \begin{cases} \frac{\exp(X_j\beta_i + \xi_j - \alpha_i p_j)}{1+\sum_{k\in\mathcal{J}_i^*}\exp(X_k\beta_i+\xi_k-\alpha_i p_k)}, & \text{if } j \in \mathcal{J}_i^* \\ 0, & \text{otherwise} \end{cases}.$$

My estimation method below relies both on the demand and supply models. In summary, firms endogenously choose product assortments to maximize expected profits; I recover unobservable product availabilities from firms' profit maximization conditions; consumers make utility-maximizing choices among available products.

4 Estimation Strategy

As is typical in the BLP literature, we need observations in multiple markets for demand estimation.[10] Suppose we observe $l = 1, ..., L$ markets, each with i $= 1, ..., I_l$ stores and representative consumers. Denote the observed market shares in market l as $s_l^{obs} = \{s_{1l}^{obs}, s_{2l}^{obs}, ..., s_{Jl}^{obs}\}$. The estimation objective is to find demand parameters that minimize the distance between predicted market shares and their observed counterparts. The first step in the estimation is

[10]In the following application, I observe cigarette sales in all of China's provinces spanning the years 2011 to 2016.

to calculate predicted market shares.

4.1 Predicted Market Shares

The utility of consumer i from choosing product j in market l can be expressed as:

$$\begin{aligned}
U_{ijl} &= X_j\beta_i + \xi_j + \Delta\xi_{jl} - \alpha_i p_{jl} + \epsilon_{ijl} \\
&= X_j(\beta + \Pi_\beta D_i + \Sigma_\beta v_{i\beta}) + \xi_j + \Delta\xi_{jl} - (\alpha + \Pi_\alpha D_i + \sigma_\alpha v_{i\alpha})p_{jl} + \epsilon_{ijl} \\
&= \delta_{jl}(X_j, p_{jl}, \xi_j, \Delta\xi_{jl}; \theta_1) + \mu_{ijl}(X_j, p_{jl}, D_i, v_i; \theta_2) + \epsilon_{ijl},
\end{aligned}$$

where subscript l is the market identifier, ξ_j is the mean fixed utility of product j in all markets and $\Delta\xi_{jl}$ represents random demand shock of product j in market l. $\delta_{jl} = X_j\beta + \xi_j + \Delta\xi_{jl}$ and $\mu_{ijl} = X_j(\Pi_\beta D_i + \Sigma_\beta v_{i\beta}) - (\alpha + \Pi_\alpha D_i + \sigma_\alpha v_{i\alpha})p_{jl}$. The utility is now expressed as being comprised of a mean preference, represented by δ_{jl}, and all other part, $\mu_{ijl} + \epsilon_{ijl}$; $\theta = (\theta_1, \theta_2)$ is the vector containing all parameters of the model, where $\theta_1 = \beta$ and $\theta_2 = (\alpha, \Pi, \Sigma)$. This setting is different from the BLP literature which including the part αp_{jl} into the main utility and write as $\delta_{jl} = X_j\beta + \xi_j + \Delta\xi_{jl} - \alpha p_{jl}$. The reason for the difference is that product prices are endogenous in the BLP model, but exogenous in the foldable menu model.

Under the foldable menu model, each consumer is given a unique optimal assortment $\mathcal{J}_i^*$, and the choice probability is therefore

$$f_{ijl}(v_i, D_i, p_l, X, \delta_l; \theta_2, \mathcal{J}_i^*) = \begin{cases} \frac{\exp(\delta_{jl}+\mu_{ijl})}{1+\sum_{k\in\mathcal{J}_i^*}\exp(\delta_{kl}+\mu_{ikl})}, & \text{if } j \in \mathcal{J}_i^* \\ 0, & \text{otherwise} \end{cases}.$$

Note that $\mathcal{J}_i^*$ is determined according to the cutoffs of $\bar{\alpha}_i$, $\{c_{i1}, c_{i2}, ..., c_{i,J-1}\}$, and the product price-cost margins are required for calculating the cutoffs. Therefore to estimate the foldable menu model, researchers need to observe the price-cost margins.[11] Here I present

[11] China's cigarette margin rates and excess taxes are available from previous studies. So I can calculate

my assumption for the model estimation.

Assumption 2 The product price-cost margins are observable by researchers.

Let $P_v(v_i)$ be the population density of v_i. The market share of product j in market l conditional only on product characteristics is

$$s_{jl}(p_l, \pi_l, X, \delta_l; \theta_2, P_v) = \int f_{ijl}(v_i, D_i, p_l, X, \delta_l; \theta_2, \mathcal{J}_i^*) dP_v(v_i) \tag{1}$$

for the foldable menu model. Then, the estimation objective is finding θ to minimize the distance between the model predicted market shares $s(p, \pi, X, \delta; \theta_2, P_v)$ and their observed counterparts, s^{obs}.

4.2 Contraction Mapping

Directly matching $s(p, \pi, X, \delta; \theta_2, P_v)$ with the observed market shares is difficult because $s(p, \pi, X, \delta; \theta_2, P_v)$ solves a nonlinear system, as shown in equation (1). Estimating the foldable menu model is even more challenging because deriving the price sensitivity cutoffs calls for solving another nonlinear system. In this section, I prove that the contraction mapping proposed in BLP is also applicable in my framework so that I can concentrate out the parameters θ_1 using the contraction mapping and focus on searching for the rest of the parameters (θ_2) nonlinearly.

Theorem 2 Let $h(\delta) = \delta + ln(s^{obs}) - ln(s(p, \pi, X, \delta; \theta_2, P_v))$; $h(\delta)$ is a contraction of modulus less than one.

Proof See appendix B.

cigarette after-tax price-cost margins in this application on China's tobacco industry.

When D_i is observed by the econometrician, $\bar{\alpha}_i$ is known for any given parameters α and Π_α, However, the integral for the predicted market shares over v_i does not have a closed-form expression and has a dimension of $K + 1$, and is thus difficult to calculate. Following BLP, I use a simulation estimator to replace the population density, $P_v(v)$, with the empirical distribution obtained from a set of ns pseudo-random draws from P_v, say $(v_1, ..., v_{ns})$, and calculate

$$s_{jl}(p_l, \pi_l, X, \delta_l; \theta_2, P_{ns}) = \frac{1}{ns} \sum_{i=1}^{ns} f_{ijl}(v_i, D_i, p, X, \delta_l; \theta_2, \mathcal{J}_i^*). \tag{2}$$

For random values of θ_2 and random starting values of δ, I can calculate the predicted market shares using equation (2) and then find converged δ using the contraction mapping.

4.3 Identification

After determining the converged δ_{jl} of product j in market l using the contraction mapping, note δ_{jl} is the dependent variable of the regression equation $\delta_{jl} = X_j\beta + \xi_j + \Delta\xi_{jl}$. I can run an OLS regression of δ on X and the product dummies to get the estimates of β and ξ. Predicted residuals of the regression are $\Delta\xi_{jl}$, which represent random demand shocks. Because product prices are exogenous in the foldable menu model, the random demand shocks cannot be correlated with product prices. Identification is achieved by searching for θ_2 to minimize the absolute value of the covariance between the random demand shocks and the exogenous product prices. Now it should be clear why I exclude the term αp from the mean utilities. Because if I include αp in the mean utilities, by including p as an explanatory variable of the OLS regression, I already assume p is uncorrelated with the residuals of the regression. Then the above identification assumption will lose its identification power.

5 Empirical Application

In this section, I apply the foldable menu model to China's tobacco industry. Aggregate sales of cigarette tiers are observed at the provincial level. Thus, a market represents one province in this application, and a product represents one cigarette tier. A firm is one provincial CNTC responsible for distributing various cigarette assortments to different retailers. Each retailer has a representative consumer whose observable demographic (real disposable income) and mean preference for each cigarette tier are known to the provincial CNTC. However, with regards to the representative consumer's random preferences for cigarette tiers and random price sensitivities, the provincial CNTC only knows their distributions.

5.1 Data and Variables

My data come from multiple sources. This section describes the data sources and how the study variables are defined and constructed.

5.1.1 Cigarette sales and market shares

I collect provincial-level data on the five tiers of cigarette sales from the 2011-2016 China Tobacco Yearbooks. For each year, data are available for each of China's 31 provincial-level administrative divisions (provinces hereafter), which gives me sales data of size $5(tiers) \times 31(markets) \times 6(years) = 930$. The summary statistics of cigarette tier sales are shown in Appendix C.

To incorporate the outside option of not smoking, I need to estimate China's potential cigarette market size, defined as the number of cigarettes that would be consumed if the entire population aged 15 and above were smokers. According to the Global Market Information Database, smoking prevalence among adults in China was approximately 28.1% in 2011, 28% in 2012 and 2013, and 27.9% in 2014-2016. Then, I construct the measure of market share relative to the potential cigarette market size by multiplying the smoking prevalence rates with their corresponding percentages of cigarette tier sales. I use these constructed market shares as the observed market shares in the foldable menu model. Table 1 summarizes the observed market shares. Tiers I and II cigarettes have clear upward market share trajectories, while Tiers IV and V have clear downward trajectories. Tier III's market shares increase significantly from 2011 to 2012 and become relatively stable after 2012.

Table 1: Summary Statistics of Observed Market Shares (%)

Tier \ Year	2011	2012	2013	2014	2015	2016
I	3.68 (1.88)	4.19 (1.54)	4.77 (1.74)	5.41 (1.94)	5.73 (2.27)	5.62 (2.54)
II	1.77 (1.04)	2.14 (1.12)	2.47 (1.26)	2.88 (1.43)	3.32 (1.47)	3.71 (1.56)
III	10.69 (2.93)	12.56 (2.63)	12.85 (2.62)	12.45 (2.31)	12.2 (2.39)	12.05 (2.54)
IV	8.22 (2.66)	6.37 (2.08)	5.87 (2.13)	5.44 (2.04)	5.08 (2.17)	5.02 (2.4)
V	3.74 (1.47)	2.75 (0.94)	2.04 (0.78)	1.73 (0.77)	1.58 (0.78)	1.51 (0.78)
N	31	31	31	31	31	31

Note: Observed market shares are constructed by multiplying the smoking prevalence rates with the corresponding percentages of cigarette tier sales. Standard deviations are in parentheses.

5.1.2 Cigarette Tier Prices and Profits

Cigarette retail prices and wholesale profits are needed in the demand estimation of the foldable menu model. Goodchild and Zheng (2018) select five brands to represent each cigarette tier, with these brands having large market shares within their respective tier. They provide average wholesale and retail prices of these cigarette brands in 2014-2016, as shown in Table 2.

Table 2: Average nominal prices and margins

Tier	Retail Price		Wholesale Price		Wholesale Margin	
	2011-2014	2015-2016	2011-2014	2015-2016	2011-2014	2015-2016
I	24.5	26.5	21.8	23	3.75	3.65
II	13	14	11.6	12.3	1.59	1.49
III	9.5	10	8.3	8.8	1.14	1.04
IV	5	5.5	4.5	4.8	0.48	0.38
V	2.5	3	2.3	2.4	0.17	0.07

Note: Prices are the average prices of each tier's five cigarette brands of as selected by Goodchild and Zheng (2018). Prices change due to tax increases in 2015.

Two exogenous tax increases in the industry occurred in May 2009 and May 2015. Details of the tax increases are summarized in Appendix D. STMA did not allow the 2009 tax increase transfer to raise cigarette prices,[12] but allowed wholesale cigarette prices to increase to coincide with the tax increase in May 2015. Thus cigarette prices change before and after 2015 in Table 2, and I argue cigarette wholesalers have no control over cigarette prices in China.

Firms are provincial cigarette wholesalers, and their economic objective is to maximize wholesale after-tax profit. Assuming firms' fixed costs, including advertising, management, etc., are not affected by cigarette sales, their after-tax profits are determined by each cigarette tier's sales volume and unit after-tax price-cost margin. Calculating the wholesale unit after-tax price-cost margins requires, in addition to cigarette taxes, China's cigarette pricing mechanism and cigarette wholesale margin rates, which I collect from Gao et al. (2012) and Zheng (2018), respectively. The detailed calculation process of wholesale after-tax price-cost margins is shown in Appendix E. The last two columns of Table 2 summarize the unit after-tax cigarette wholesale margins. As stated in Goodchild and Zheng (2018), the increase in the wholesale price of cigarettes matches the change in the ad valorem excise tax rate from 5% to 11% in 2015, indicating that the new specific excise tax of ¥0.10/pack was absorbed into wholesale margins, rather than being passed along to consumers. Thus, wholesale profits for 2015 and 2016 equal the values for 2011-2014 minus the new specific excise tax of ¥0.10/pack.

One concern here is that the cigarette prices and profits listed above are at the national level, but the cigarette sales and market shares are at the provincial level. Given that the composition of cigarette tier sales varies across provinces and years, the weighted average cigarette tier prices and profits also vary across provinces and years. Unfortunately, as far

[12] STMA issued a document, entitled Notice of Adjusting Cigarettes Allocation Price (2009, No. 180), also effective May 1, 2009, which declared that wholesale cigarette prices should remain the same nationwide as before the excise tax adjustment. Given that the excise tax is collected at the producer and wholesale levels but not at the retail level, retail cigarette prices will remain the same if the wholesale cigarette prices remain the same.

as I know, the detailed compositions of provincial cigarette sales data are not available, and thus I assume the nominal prices and profits of cigarette tiers are the same in all provinces. However, we can easily extend this model to incorporate different compositions of cigarette tiers when more detailed datasets are available.

5.1.3 CPI and Disposable Income

The nominal prices and profits of cigarette tiers summarized in Table 2 are the same across the country, but tobacco product consumer price indexes (CPIs hereafter) vary across provinces and years, which provides variation in real cigarette tier prices, and the variation is necessary for the estimation of price sensitivity parameters. I collect provincial-level tobacco product CPIs from the 2011-2016 China Statistics Yearbooks and deflate nominal retail prices and wholesale profits using the tobacco product CPIs to obtain real prices and profits.

I use consumers' disposable incomes as their observable demographic to cigarette wholesalers because, intuitively, high-income consumers are more likely to buy top-tier cigarettes. I collect disposable income data from the 2011-2016 Provincial Statistical Yearbooks for each province. The yearbooks report the overall average disposable income and the average income for each income quintile of each province and year. The income quintile data is necessary because provinces with similar overall average incomes may have very different income dispersions. When I generate random consumer incomes using the quintile values, distributions of the generated incomes are closer to the actual income distributions than the randomly generated ones that only use the overall average incomes.

As with cigarette prices, incomes reported in the Provincial Statistical Yearbooks are also in nominal terms. I additionally collect the overall CPIs of each province and year and use them to deflate nominal income to obtain real income. Table 3 summarizes the overall CPIs, tobacco product CPIs, nominal provincial mean disposable incomes, and real provincial mean disposable incomes in these years.

Table 3: Summary Statistics of CPI and Disposable Income

Variable \ Year	2011	2012	2013	2014	2015	2016
Overall CPI	105.49 (0.38)	108.38 (0.65)	111.44 (1.03)	113.69 (1.19)	115.34 (1.39)	117.43 (1.59)
Tobacco Product CPI	102.72 (1.06)	105.66 (2.26)	106.20 (2.42)	105.60 (2.47)	108.00 (2.53)	109.56 (2.69)
Nominal Mean Income (¥1000)	20.61 (53.63)	23.22 (58.44)	25.28 (64.49)	27.54 (70.09)	29.9 (75.63)	32.29 (82.61)
Real Mean Income (¥1000)	19.54 (51.12)	21.43 (54.31)	22.7 (58.43)	24.24 (62.15)	25.93 (65.37)	27.49 (69.42)
N	31	31	31	31	31	31

Note: Data are provided by China Statistics Yearbooks and Provincial Statistics Yearbooks 2011-2016. 2010 is the base year for CPIs. Standard deviations are in parentheses.

I include real disposable income in the model because the market shares of Grade A cigarettes (Tier I and II) are positively correlated with provincial mean disposable incomes. I run an OLS regression of the observed market shares of Grade A cigarettes (sum of Tier I and II cigarette market shares) on the log value of real mean disposable income to test the correlation. The coefficient of income is positive and statistically significant at 1% significance level, as shown in Table 4. On average, 1% higher real mean disposable incomes are associated with 5.6 percentage point larger market shares of Grade A cigarettes. However, the regression result may overstate the effect of income on Grade A cigarette market shares. Because in the foldable menu model, high-priced cigarettes are more affordable for high-income consumers, and high-income consumers are more likely to be given limited assortments containing only high-priced cigarettes.

5.2 Estimation

When I estimate the model using the actual dataset, the utility function of a consumer i who chooses a Tier j cigarette in province l and year t is

$$U_{ijlt} = \xi_{jlt} + \sigma_j v_{ij} - \alpha_i p_{jlt} + \epsilon_{ijlt},$$

Table 4: OLS Regression Results

	Observed market share of Grade A cigarettes
Log mean disposable income	0.0562***
	(0.0075)
Constant	-0.4875***
	(0.075)
N	186

Robust standard errors are in parentheses .

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

where $\xi_{jlt} = \bar{\xi}_{jl} + \Delta\xi_{jlt}$ represents the mean fixed utility of Tier j cigarettes in province l and year t; ξ_{jlt} consists of a time-invariant part $\bar{\xi}_{jl}$ and a time-variant part $\Delta\xi_{jlt}$; $\Delta\xi_{jlt}$ represents the random demand shock to the mean fixed utility in different years; $\alpha_i = \alpha - \alpha_{inc} \times inc_i + \sigma_\alpha v_{i\alpha}$. For each province-year, I generate 10,000 random incomes using the provincial income quintile data and then deflate them using the overall CPIs to obtain real incomes; inc_i is the natural log of the simulated real income in units of ¥10,000 to capture the fact that the marginal effect of income on price sensitivity is larger for low-income consumers than high-income consumers. Overall, the mean fixed utilities, disposable income distributions, real retail prices, real wholesale profits, and individual heterogeneities determine the varying market shares of cigarette tiers in different provinces and years. Smokers in rich provinces buy more Grade A cigarettes because they are less price-sensitive and have a lower chance of finding low-priced cigarettes in their local stores because of the foldable menu.

Nominal cigarette prices are exogenous and set by STMA. Variations of real cigarette prices in different provinces and years are only caused by the tobacco product CPIs, which are assumed to be uncorrelated with the random demand shocks. Given $\theta_2 = \{\alpha, \alpha_{inc}, \sigma_\alpha, \sigma_j \text{ for } j = 1, 2, ..., 5\}$, I can obtain converged ξ using the contraction mapping and then run an OLS regression of ξ on the province-tier dummies to get the estimates of $\bar{\xi}$. Predicted residuals of the regression represent random demand shocks $\Delta\xi$. Identification is achieved by searching for θ_2 to minimize the absolute value of the covariance between the random demand shocks and real cigarette prices.

I make the following application-specific simplifying assumptions. Note equation (3) for price sensitivity cutoffs has no closed-form expression because it's a multiple integral over $v_i \in R^6$. Thus, it is hard to code and calculate. One way to solve the equations is to simulate a large number of v_i, but it could be very time-consuming. Here, I present my first simplifying assumption.

Simplifying Assumption 1 Firms are not able to discriminate based on unobserved consumer heterogeneity.

A similar assumption is used in some previous BLP type studies. For example, D'Haultfœuille et al. (2019) allows consumers to be heterogeneous within a group classifying by the same known demographics to auto sellers, but assume the sellers are not able to discriminate based on this unobserved heterogeneity.

Given Simplifying Assumption 1, the cigarette wholesaler in province l and year t should calculate the ex-ante profit from consumer i and assortment $\mathcal{J}_i$ using

$$\Pi_{i,\mathcal{J}_i} = \sum_{j \in \mathcal{J}_i} \pi_j \underbrace{\frac{\exp(\xi_{jlt} - \bar{\alpha}_i p_{jlt})}{1 + \sum_{j \in \mathcal{J}_i} \exp(\xi_{jlt} - \bar{\alpha}_i p_{jlt})}}_{\text{choice probability}},$$

where $\bar{\alpha}_i = \alpha - \alpha_{inc} \times inc_i$ is the deterministic price sensitivity of consumer i, and the equations for the cutoffs of $\bar{\alpha}_i$ change to

$$\sum_{j=1}^{m} (\pi_{jlt} - \pi_{m+1,lt}) \exp(\xi_{jlt} - c_{mlt} p_{jlt}) - \pi_{m+1,lt} = 0,$$

for $m = 1, 2, 3, 4$. Now the set of c_{mlt} can be calculated using any nonlinear algorithm in MATLAB.

Also, note that I use a single set of c_{mlt} for all consumers in province l and year t, as this is an application for cigarette consumption and the only known demographic is their disposable income. I argue that regardless of whether a consumer is rich or poor, his mean preference for a given cigarette tier is the same. Individual deviations from the mean are captured by the term v_{ij}. Now, I present my second simplifying assumption.

Simplifying Assumption 2 The preference parameters do not depend on consumer demographics known to the firm.

Readers should note the Simplifying Assumption 2 does not apply to some industries, e.g., the automobile industry, where young professionals prefer car performance and families with multiple kids care more about internal capacity. To apply the foldable menu model to such industries, we need to classify consumers into groups according to their known demographics and calculate a set of price sensitivity cutoffs for each group of consumers.

I conduct a Monte Carlo experiment to test the performance of the estimation strategy for the simplified model. The data generation process, estimation procedure, and results are shown in Appendix F. I apply the same estimation method to the real dataset. Table 5 summarizes the estimation results using the actual data. The first column lists the estimation results using the foldable menu model with three product assortments by assuming the cigarettes of Tiers I, II and III are always available because the CNTC's strategy is to decrease the availability of Tiers IV and V cigarettes. The second column lists the estimation results using the standard foldable menu model with five cases. The results of the first two columns are similar because only small fractions of consumers are offered the very limited assortments {I} and {I,II}.[13] The last column represents the estimation results using the misspecified model ignoring varying cigarette availabilities, and I call it the random coefficient logit model.

When estimating cigarette demand using the foldable menu model (for both three and five cases), all coefficients have the expected sign and scales. For α, α_{inc}, σ_α, and σ_j for

[13] In my simulation, 2.36% consumers are offered {I} and 2.35% {I,II} on average.

Table 5: Estimation results using the actual dataset

Coefficient	Foldable menu (3 cases)	Foldable menu (5 cases)	Random coefficient logit
α	0.743	0.7671	0.1002
	(0.0211)	(0.023)	(0.004)
α_{inc}	0.2962	0.3058	0.0306
	(0.0084)	(0.0092)	(0.0018)
σ_α	0.1192	0.1266	0.0498
	(0.0034)	(0.0044)	(0.002)
σ_1	0.658	0.6534	0.9172
	(0.0048)	(0.0023)	(0.0012)
σ_2	0.7602	0.7581	0.7258
	(0.0025)	(0.0027)	(0.0032)
σ_3	0.9598	0.9598	0.9436
	(0.0007)	(0.0006)	(0.0008)
σ_4	0.5016	0.4799	0.7535
	(0.0064)	(0.0037)	(0.0032)
σ_5	0.7468	0.7413	0.8117
	(0.0034)	(0.005)	(0.0012)
$\bar{\xi}_1$	4.6684	4.6269	-1.801
	(1.2321)	(1.2953)	(0.3964)
$\bar{\xi}_2$	1.4857	1.6618	-2.8074
	(0.7981)	(0.8092)	(0.4852)
$\bar{\xi}_3$	2.0607	2.3248	-1.4363
	(0.5206)	(0.5074)	(0.2106)
$\bar{\xi}_4$	0.0855	0.1966	-2.4478
	(0.5513)	(0.5489)	(0.4331)
$\bar{\xi}_5$	-1.7463	-1.635	-3.7588
	(0.5115)	(0.4882)	(0.4907)

Note: For α, α_{inc}, σ_α, and σ_j for $j = 1, 2, ..., 5$, estimates are the mean of 20 estimations with bootstrap standard errors in parentheses; $\bar{\xi}_j$ for $j = 1, 2, ..., 5$ are the average values of the mean $\bar{\xi}_{jl}$ (of 20 estimations) over China's 31 provinces, with associated standard deviations in parentheses. All coefficients are statistically significance at 1% significance level.

$j = 1, 2, ..., 5$, the estimates are the mean of 20 estimations with bootstrap standard errors in parentheses, where each estimation uses a different simulation of individual heterogeneities v. I also calculate the mean $\bar{\xi}_{jl}$ of the 20 estimations; $\bar{\xi}_j$ for $j = 1, 2, ..., 5$ are the average values of the mean $\bar{\xi}_{jl}$ over China's 31 provinces. The mean fixed utilities are monotonic with the prices of cigarette tiers, except for Tier II, which has a lower mean fixed utility than Tier III on average, which may be a reason why Tier III is China's most popular cigarette tier.

On the other hand, when I estimate cigarette demand using the random coefficient logit model, the coefficients for price sensitivities and mean fixed utilities are underestimated, similar to the estimation results using the simulated dataset in Appendix F. The underestimated coefficients for price sensitivities lead to lower estimates of price elasticities, as shown in the following section.

5.3 Counterfactual Experiments

5.3.1 Price elasticities

Estimating demand elasticities is the standard post demand estimation practice. I summarize the own-price and cross-price demand elasticities for each of the five cigarette tiers, which measure the percent change in overall sales (of all provinces and years) for a particular cigarette tier if its price increases by 1% but prices of other tiers remain unchanged. Table 6 summarizes the estimates using the foldable menu model and the misspecified model ignoring varying product availabilities, i.e., the random coefficient logit model.

The price elasticities estimated using the foldable menu model are relatively high compared to previous estimates of price elasticities that assume cigarettes to be homogeneous (e.g., Hu and Mao (2002); Lance et al. (2004); Bishop et al. (2007); Chen and Xing (2016)), because these studies ignore the substitutions among cigarette tiers. These price elasticities are also high compared to price elasticities of cigarette brands (Liu et al. (2015)). The reason is that the price changes took place at the cigarette tier level in this study, meaning that

Table 6: Comparison of price elasticities

	Foldable Menu					Random Coefficient Logit				
	Tier I	Tier II	Tier III	Tier IV	Tier V	Tier I	Tier II	Tier III	Tier IV	Tier V
Tier I	-2.68	0.926	0.721	0.439	0.121	-0.627	0.017	0.023	0.03	0.03
Tier II	0.107	-3.023	0.15	0.239	0.36	0.005	-0.612	0.013	0.016	0.016
Tier III	0.198	0.401	-2.383	0.754	2.123	0.022	0.043	-0.458	0.057	0.063
Tier IV	0.022	0.078	0.101	-2.138	0.565	0.007	0.013	0.014	-0.318	0.022
Tier V	0.001	0.001	0.013	0.034	-1.446	0.001	0.003	0.003	0.004	-0.185

Note: Compared with the foldable menu model, ignoring varying product availabilities (the random coefficient logit model) leads to underestimated own and cross-price elasticities of cigarette tiers because it falsely attributes limited substitutions of products to lower price elasticities.

the prices of all cigarette brands within the tier increase by the same percentage. Thus, the price changes will have a larger effect on sales than the case when only the price of a single cigarette brand increases.

One study estimating cigarette demand elasticities at the tier level is Li et al. (2016), which uses 2006-2009 individual survey data from China. The authors use the classification criteria before May 2009 to categorize cigarettes into four tiers by combining Tiers I and II together. They estimate the marginal own-price effects for each cigarette tier and find that the share of smokers in each tier decreases by 10.3%, 11.4%, 10%, and 21.1% when increasing the tier price by ¥1 while holding the prices of other tiers constant. A ¥1 increase represents 4.11%, 14.16%, 25.64%, and 75.47% increases in the median prices of cigarette tiers in their data. There are several explanations for why I obtained price elasticities larger than the ones estimated in Li et al. (2016):

1. Their price increases are in nominal terms, while my price increases are in real terms.
2. They estimate the demand for cigarette tiers without incorporating the outside option of not smoking. All individuals are smokers in their study. When the price of one cigarette tier increases, current smokers of the tier can choose to stay with the same tier or substitute to other tiers, but cannot switch to the outside option of not smoking.
3. They ignore the issue of varying product availability and assume all cigarette tiers are equally available to each consumer. As shown in Table 6, when I estimate cigarette

demand using the random coefficient logit model, the estimated elasticities are much smaller than those estimated using the foldable menu model. The elasticities estimated using the random coefficient logit model are within the range of previous estimates of price elasticities of China's tobacco industry, but not consistent with the industry fact, as explained below.

I report the overall price elasticity of cigarette demand, defined by the percent change in total cigarette demand when the prices of all cigarette tiers increase by 1%. The estimate is -1.1983 using the foldable menu model. Although the estimate is larger than the estimated elasticities in the literature of China's tobacco industry, it is close to the actual market changes following China's May 2015 tax change, where the 7% increase in retail price is associated with an 8% drop in retail sales volume between 2014 to 2016 (Goodchild and Zheng (2018)). On the other hand, the estimate is -0.384 using the random coefficient logit model. The latter is in the mid-range of study results about China's cigarette price elasticities spanning 0 - 0.8, but it far from explains the actual market changes following China's May 2015 tax change.

Recovering important unobservable variables from firms' profit maximization conditions allows us to estimate price elasticities more accurately than models that simply ignore the problem of unobservable variables. For example, D'Haultfœuille et al. (2019) recovers unobservable auto sale prices to analyze price discrimination. The authors conclude that assuming all consumers pay the posted prices of autos (the uniform pricing model in their paper) leads to underestimation of price sensitivity parameters and price elasticities, because this assumption overstates the prices paid by consumers in their study. Similarly, in this application, I find that assuming all consumers have the same access to all cigarette tiers also leads to underestimation of price sensitivity parameters and price elasticities, because this assumption falsely attributes the limited substitutions of products due to the foldable menu, to lower price elasticities.

5.3.2 Policy implications

Tobacco taxes have been widely adopted as one of the most effective tobacco control policies in many countries. Following the May 2015 tax adjustment, the weighted tax share as a percentage of retail price increased from 52% in 2014 to 56% in 2015 and 2016, and the weighted excise tax share as a percentage of retail price increased from 31% in 2014 to 34% in 2015. Both are still lower than the WHO recommended standard of 75% and 70%, respectively (Zheng (2018)). A feasible approach to meet the WHO-recommended standard would be to gradually raise the ad valorem excise tax rate. This section thus simulates a cigarette ad valorem excise tax increase of 5%, 10%, 15%, and 20%. For simplicity, I assume the new exercise tax is levied at the retail level and uses the current retail price as the tax base, so it does not affect the current unit after-tax wholesale margins.

Figure 6 depicts the simulation results using the foldable menu model and the random coefficient logit model. For the foldable menu model, a 5% increase in real retail prices leads to a 5.8% drop in cigarette sales, consistent with the actual market change between 2014 and 2016. On the other hand, because of the lower estimated price sensitivities using the random coefficient logit model, the increased taxes and cigarette prices have smaller negative effects on cigarette sales and thus lead to larger increases in tax revenues. However, given that the elasticities are much lower than the level consistent with the actual change in cigarette sales following China's May 2015 tax adjustment, tax policy based on the assumption that all cigarette tiers are equally available to every consumer is misleading.

5.3.3 Market changes without the foldable menu

This paper focuses on demand estimation with the foldable menu, which helps wholesalers to maximize profit and improves the accessibility of low-priced cigarettes to low-income consumers. However, the foldable menu forces less price-sensitive (high-income) consumers to make suboptimal decisions in limited choice sets. It is natural to perform counterfactual experiments to see how the market changes when all consumers have equal access to all

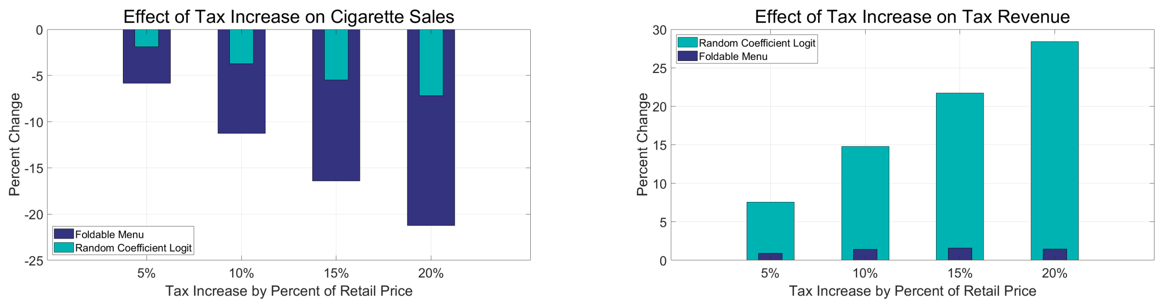

Figure 6: Compared with the foldable menu model, ignoring varying product availabilities (the random coefficient logit model) leads to underestimated effect of increasing tax on cigarette sales and overestimated effect of increasing tax on tax revenue.

cigarette tiers. Using the estimated parameters of the standard foldable menu model (5 cases) and simulated consumer incomes, I calculate the predicted cigarette tier sales, wholesale profits, and consumer utilities in the modified market environment and compare these values with their observed counterparts to quantify the changes. The changes are summarized in Table 7. As expected, when all cigarette tiers are equally available to every consumer, the sales of Tiers I and II cigarettes will decrease, the sales of Tiers III, IV and V cigarettes will increase. Total cigarette sales will increase by 12.11%, but total sale revenue will drop by 2.24% because more consumers choose to buy low-priced cigarettes when available. Cigarette wholesale profits would be approximately 6% lower if the wholesalers did not follow the foldable menu, which may explain why CNTC chose to reduce the availability of low-priced cigarettes as a monopoly strategy in these years. Finally, consumer welfare will increase by 6.75% because all consumers can make their optimal choices from the complete assortment.

5.3.4 Limited assortment distributions

Product availabilities are not observable, but I can use the estimated foldable menu model to simulate the limited assortment distributions to recover unobservable product availabilities. Moreover, consumers suffer more welfare loss when they are given a more limited assortment. I also calculate the welfare loss for consumers given each limited assortment.

Figure 7 depicts the limited assortment distributions and I summarize the effect of limited

Table 7: Market Changes (%) if all tiers are available

Variables	Percent Change
Tier I sales volume	-19.9
Tier II sales volume	-4.42
Tier III sales volume	4.45
Tier IV sales volume	26.37
Tier V sales volume	109.35
Total Cigarette sales volume	12.11
Wholesale profit	-5.95
Total cigarette sales revenue	-2.24
Consumer welfare	6.75

Note: When all cigarette tiers are available everywhere, Tiers I and II sales decrease, and Tiers III-V sales increase. Total cigarette sales volume increases, but sale revenue drops because more consumers choose to buy low-priced cigarettes. Cigarette wholesale profit drops but consumer welfare increases.

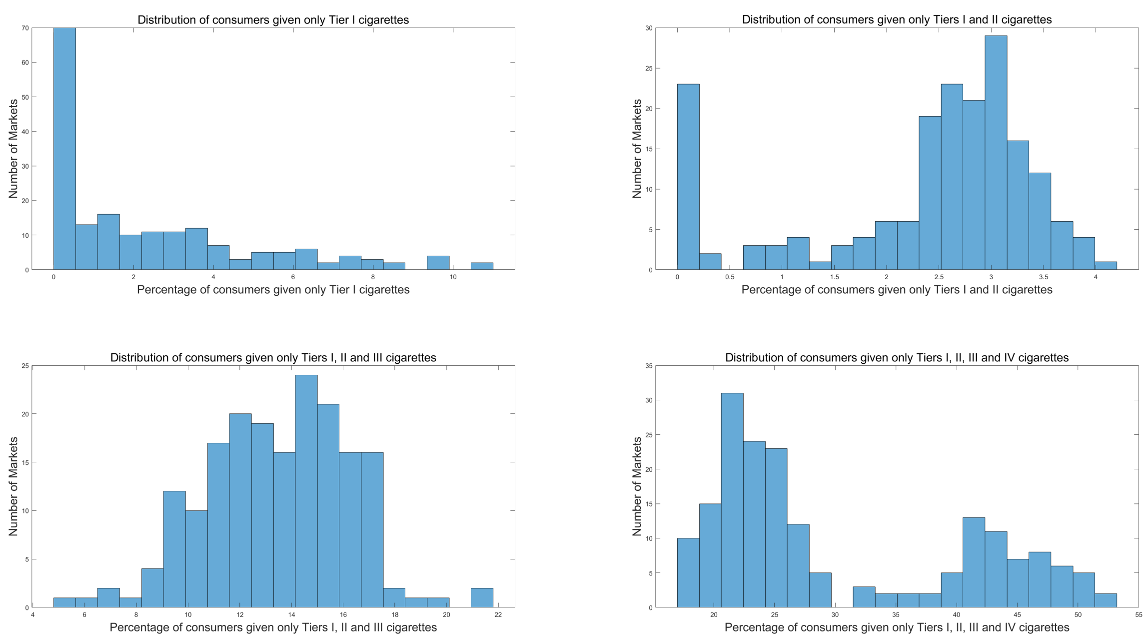

Figure 7: Limited assortment distributions.

assortments on consumer welfare in Table 8. For the 186 markets (31 provinces $\times$ 6 years), on average, 2.36%, 2.35%, 13.47%, and 30.16% consumers are given the limited assortment {I}, {I,II}, {I,II,III}, and {I,II,III,IV}, respectively.

1. 126 out of the 186 markets have consumers only accessing the limited assortment {I}. On average, 47.51% of these consumers make better choices when the complete assortment is available. The mean increase in utility is 1.55, or ¥11.36 in pecuniary terms for purchasing one pack of cigarettes.

2. 164 out of the 186 markets have consumers only accessing the limited assortment {I,II}. On average, 40.93% of these consumers make better choices when the complete assortment is available. The mean increase in utility is 1.45, or ¥7.4 in pecuniary terms for purchasing one pack of cigarettes.

3. All markets have consumers only accessing the limited assortment {I,II,III} or {I,II,III,IV}. On average, 15.25% of consumers in the limited assortment {I,II,III} make better choices when the complete assortment is available. The mean increase in utility is 1.12, or ¥3.87 in pecuniary terms for purchasing one pack of cigarettes.

4. On average, 5.32% of consumers in the limited assortment {I,II,III,IV} make better choices when the complete assortment is available. The mean increase in utility is 1.05, or ¥2.37 in pecuniary terms for purchasing one pack of cigarettes.

The mean increase in pecuniary terms is calculated using the mean increase in utility divided by consumers' mean price sensitivity in the respective limited assortment. These values of consumer welfare loss can serve as a baseline for evaluating the damages of potential collective actions suits of consumers facing limited choice sets.

Table 8: Effect of limited assortments on consumer welfare

	{I}	{I,II}	{I,II,III}	{I,II,III,IV}
Consumers (%) given limited assortments	2.36	2.35	13.47	30.16
	(2.7)	(1.11)	(2.8)	(10.53)
Consumers (%) change choices if given the full assortment	47.51	40.93	15.25	5.32
	(5.82)	(5.28)	(4.9)	(2.46)
Average increase in utility	1.55	1.45	1.12	1.05
	(0.09)	(0.08)	(0.04)	(0.03)
Average increase in pecuniary terms	11.36	7.4	3.87	2.37
	(11.74)	(3.65)	(0.89)	(0.33)
N	126	164	186	186

Note: Values are the mean of the N markets, except for the first row, which contains the mean values of all 186 markets. Standard deviations are in parentheses.

6 Summary and Future Research

This paper provides a new approach for demand estimation of differentiated products with varying availabilities, when product availabilities are unobservable and the proxy variable for the product availability distribution is unavailable.

The theoretical model builds on two assumptions: product prices are exogenous and products have no identical price-cost margins. To maximize profit when the two assumptions hold, firms only need to choose optimal assortments from a foldable menu, which has the same number of assortments as the number of products. Also, each consumer's optimal assortment is uniquely identified, so I can recover unobservable product availabilities from the firms' profit maximization conditions. Furthermore, I prove that the contraction mapping proposed in BLP also extends to my revised framework and employ it in the demand estimation with the foldable menu.

I illustrate the estimation method of the foldable menu model using a simulated dataset and a real dataset on China's tobacco industry. Both estimation results show that assuming all products are available (the random coefficient logit model) leads to underestimated price elasticities because this assumption falsely attributes limited substitutions of products to inelastic demand. Tax policies based on the misspecified model would be misleading and cause unexpected consequences. The foldable menu model can correct the estimation bias of the misspecified model.

One limitation of this study is Simplifying Assumption 2, where I assume the preference parameters do not depend on consumer demographics known to the firm. The benefit is that I only need to calculate one set of price sensitivity cutoffs for each market. Without Simplifying Assumption 2, I need to calculate one set of price sensitivity cutoffs for each group of consumers with the same known demographic, which is a continuous variable (real disposable income) in this application. One way to relax Simplifying Assumption 2 is by dividing consumers into a small number of groups and adding the interactions of group mean income with cigarette tier dummies into the utility function. Estimated coefficients of the interactions can explain whether a cigarette tier is normal or inferior, and I leave this topic for future research.

Appendices

A Proofs of the Lemmas

For simplicity, I write the utility function as

$$U_{ij} = X_j\beta_i + \xi_j - \alpha_i p_j + \epsilon_{ij} = \gamma_{ij} - \alpha_i p_j + \epsilon_{ij}.$$

A.1 Lemma 1

Lemma 1: Any assortment that does not contain top-tier products cannot be optimal.

Proof: Let $\mathcal{J}_i = \{m, m+1, ..., n\}$ and $\mathcal{J}'_i = \{m-1, m, m+1, ..., n\}$, $\forall m > 1$ & $n \leq J$. Let $EP(\cdot)$ denote the expected profit from the corresponding assortment.

$$\begin{aligned}
&EP(\mathcal{J}_i) - EP(\mathcal{J}'_i)\\
&= \int \Big(\frac{\sum_{j=m}^{n} \pi_j \exp(\gamma_{ij} - \alpha_i p_j)}{1 + \sum_{j=m}^{n} \exp(\gamma_{ij} - \alpha_i p_j)} - \frac{\sum_{j=m-1}^{n} \pi_j \exp(\gamma_{ij} - \alpha_i p_j)}{1 + \sum_{j=m-1}^{n} \exp(\gamma_{ij} - \alpha_i p_j)}\Big) dP_v(v_i)\\
&= -\int \frac{\pi_{m-1} \exp(\gamma_{i,m-1} - \alpha_i p_{m-1}) + \sum_{j=m}^{n} (\pi_{m-1} - \pi_j) \exp(\gamma_{i,m-1} - \alpha_i p_{m-1}) \exp(\gamma_{ij} - \alpha_i p_j)}{(1 + \sum_{j=m}^{n} \exp(\gamma_{ij} - \alpha_i p_j))(1 + \sum_{j=m-1}^{n} \exp(\gamma_{ij} - \alpha_i p_j))} dP_v(v_i)\\
&< 0,
\end{aligned}$$

for any consumer i.

A.2 Lemma 2

Lemma 2: Let $\vec{\mathbb{J}} = \{\vec{1}, \vec{2}, ..., \vec{J}\}$, and let $\mathcal{J}_i \in \vec{\mathbb{J}}$ be the optimal assortment among $\vec{\mathbb{J}}$ to consumer i. There exist $J-1$ cutoffs of $\bar{\alpha}_i$, $\{c_{i1}, c_{i2}, ..., c_{i,J-1}\}$, such that

$$\mathcal{J}_i = \begin{cases} \vec{1}, & \textbf{if } \bar{\alpha}_i < c_{i1}, \\ \vec{m}, & \textbf{if } c_{i,m-1} \leq \bar{\alpha}_i < c_{im} \textbf{ for m = 2,3,...,J-1}, \\ \vec{J}, & \textbf{if } \bar{\alpha}_i \geq c_{i,J-1}. \end{cases}$$

Proof: 1): For any two consecutive assortments among $\vec{\mathbb{J}}$, i.e., $\vec{m}$, $\vec{n}$ and $n = m+1$, there exists a cutoff c_{im}. When $\bar{\alpha}_i = c_{im}$, the two assortments have the same expected profit from consumer i.

$$\int \frac{\sum_{j=1}^{m} \pi_j \exp(\gamma_{ij} - (c_{im} + \sigma_\alpha v_{i\alpha}) p_j)}{1 + \sum_{j=1}^{m} \exp(\gamma_{ij} - (c_{im} + \sigma_\alpha v_{i\alpha}) p_j)} dP_v(v_i) = \int \frac{\sum_{j=1}^{m+1} \pi_j \exp(\gamma_{ij} - (c_{im} + \sigma_\alpha v_{i\alpha}) p_j)}{1 + \sum_{j=1}^{m+1} \exp(\gamma_{ij} - (c_{im} + \sigma_\alpha v_{i\alpha}) p_j)} dP_v(v_i).$$

Rearranging the equation gives

$$\int \frac{[\sum_{j=1}^{m} (\pi_j - \pi_{m+1}) \exp(\gamma_{ij} - (c_{im} + \sigma_\alpha v_{i\alpha}) p_j) - \pi_{m+1}] \exp(\gamma_{i,m+1} - (c_{im} + \sigma_\alpha v_{i\alpha}) p_{m+1})}{(1 + \sum_{j=1}^{m} \exp(\gamma_{ij} - (c_{im} + \sigma_\alpha v_{i\alpha}) p_j))(1 + \sum_{j=1}^{m+1} \exp(\gamma_{ij} - (c_{im} + \sigma_\alpha v_{i\alpha}) p_j))} dP_v(v_i) = 0. \tag{3}$$

For any v_i, the part

$$A(\cdot | c_{im}, v_i) = \frac{\exp(\gamma_{i,m+1} - (c_{im} + \sigma_\alpha v_{i\alpha}) p_{m+1})}{(1 + \sum_{j=1}^{m} \exp(\gamma_{ij} - (c_{im} + \sigma_\alpha v_{i\alpha}) p_j))(1 + \sum_{j=1}^{m+1} \exp(\gamma_{ij} - (c_{im} + \sigma_\alpha v_{i\alpha}) p_j))} > 0.$$

To make equation (3) = 0, it must be that

$$B(\cdot | c_{im}, v_i) = \sum_{j=1}^{m} (\pi_j - \pi_{m+1}) \exp(\gamma_{ij} - (c_{im} + \sigma_\alpha v_{i\alpha}) p_j) - \pi_{m+1} \geq 0$$

for some v_i and $B(\cdot|c_{im}, v_i) < 0$ for all other v_i. For any $\bar{\alpha}_i \to c_{im}^+$, we have $\lim_{\bar{\alpha}_i \to c_{im}^+} A(\cdot|\bar{\alpha}_i, v_i) = A(\cdot|c_{im}, v_i)$. However, $B(\cdot|\bar{\alpha}_i, v_i)$ is strictly decreasing in $\bar{\alpha}_i$. When we have $\bar{\alpha}_i > c_{im}$, we must have a bigger portion of $B(\cdot|\bar{\alpha}_i, v_i) < 0$ than $B(\cdot|c_{im}, v_i) < 0$. So when $\bar{\alpha}_i > c_{im}$, we have

$$EP(\vec{m}) < EP(\vec{n}).$$

Similarly, when $\bar{\alpha}_i < c_{im}$, we have

$$EP(\vec{m}) > EP(\vec{n}).$$

2): In this step, I show that when $\bar{\alpha}_i < c_{im}$, we also have $EP(\vec{m}) > EP(\vec{n})$, for any $n \in \{m+2, m+3, ..., J\}$.

$$\begin{aligned}
&EP(\vec{m}) - EP(\vec{n}) \\
&= \int \Big(\frac{\sum_{j=1}^m \pi_j \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)}{1 + \sum_{j=1}^m \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)} - \frac{\sum_{j=1}^n \pi_j \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)}{1 + \sum_{j=1}^n \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)}\Big) dP_v(v_i) \\
&= \int \frac{\sum_{r=1}^{n-m} [\sum_{j=1}^m (\pi_j - \pi_{m+r}) \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+r}] \exp(\gamma_{i,m+r} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_{m+r})}{(1 + \sum_{j=1}^m \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1 + \sum_{j=1}^n \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))} dP_v(v_i).
\end{aligned}$$

From the above step, we know when $\bar{\alpha}_i < c_{im}$,

$$\int \frac{[\sum_{j=1}^m (\pi_j - \pi_{m+1}) \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+1}] \exp(\gamma_{i,m+1} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_{m+1})}{(1 + \sum_{j=1}^m \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1 + \sum_{j=1}^{m+1} \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))} dP_v(v_i) > 0.$$

Let $f_i(j|\vec{m}) = \frac{\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)}{1 + \sum_{j=1}^m \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)}$, i.e., the probability of consumer i choosing product j in $\vec{m}$. Then, I can rewrite the above expression as

$$\int [\sum_{j=1}^m (\pi_j - \pi_{m+1}) \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+1}] f_i(0|\vec{m}) f_i(m+1|\overrightarrow{m+1}) dP_v(v_i) > 0.$$

Proposition 1: The sign of

$$\int [\sum_{j=1}^{m}(\pi_j - \pi_{m+1}) \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+1}] f_i(0|\mathcal{J}_i) f_i(k|\mathcal{J}_i') dP_v(v_i)$$

does not change with the choices of any $\mathcal{J}_i \subseteq \vec{J}$ and any $k \in \mathcal{J}_i' \subseteq \vec{J}$.

Proof: Note the part $\sum_{j=1}^{m}(\pi_j - \pi_{m+1}) \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+1}$ increases in $v_{i\beta}$ and decreases in $v_{i\alpha}$; the part $f_i(0|\mathcal{J}_i)$ decreases in $v_{i\beta}$ and increases in $v_{i\alpha}$ for any $\mathcal{J}_i \subseteq \vec{J}$; the part $f_i(k|\mathcal{J}_i')$ increases in $v_{i\beta}$ and decreases in $v_{i\alpha}$ for any $k \in \mathcal{J}_i' \subseteq \vec{J}$. Changing $\mathcal{J}_i$, k, and $\mathcal{J}_i'$ does not affect the rank of $f_i(0|\mathcal{J}_i) f_i(k|\mathcal{J}_i')$ over v_i. So, when we integrate the above expression over an infinite number of v_i, if we have

$$\int [\sum_{j=1}^{m}(\pi_j - \pi_{m+1}) \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+1}] f_i(0|\vec{m}) f_i(m+1|\overrightarrow{m+1}) dP_v(v_i) > 0,$$

we must also have

$$\int [\sum_{j=1}^{m}(\pi_j - \pi_{m+1}) \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+1}] f_i(0|\mathcal{J}_i) f_i(k|\mathcal{J}_i') dP_v(v_i) > 0,$$

for any $\mathcal{J}_i \subseteq \vec{J}$ and $k \in \mathcal{J}_i' \subseteq \vec{J}$. End of proof.

Given Proposition 1, we have

$$\begin{aligned}
&\int [\sum_{j=1}^{m}(\pi_j - \pi_{m+1}) \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+1}] f_i(0|\vec{m}) f_i(m+1|\overrightarrow{m+1}) dP_v(v_i) > 0 \\
\Rightarrow &\int [\sum_{j=1}^{m}(\pi_j - \pi_{m+1}) \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+1}] f_i(0|\vec{m}) f_i(m+r|\vec{n}) dP_v(v_i) > 0 \\
\Rightarrow &\int [\sum_{j=1}^{m}(\pi_j - \pi_{m+r}) \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+r}] f_i(0|\vec{m}) f_i(m+r|\vec{n}) dP_v(v_i) > 0.
\end{aligned}$$

Because

$$\sum_{j=1}^{m}(\pi_j-\pi_{m+r})\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_{m+r}\geq\sum_{j=1}^{m}(\pi_j-\pi_{m+1})\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_{m+1},$$

then, for any v_i and any $r\in\{1,2,...,n-m\}$, the inequality $EP(\vec{m})>EP(\vec{n})$ holds for any $n\in\{m+2,m+3,...,J\}$.

3): In the last step, I show that when $\bar{\alpha}_i>c_m$, letting $n=m+1$, we also have $EP(\vec{l})<EP(\vec{n})$, for any $l\in\{1,2,...,m-1\}$.

$$\begin{aligned}
&EP(\vec{l})-EP(\vec{n})\\
&=\int(\frac{\sum_{j=1}^{l}\pi_j\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)}{1+\sum_{j=1}^{l}\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)}-\frac{\sum_{j=1}^{n}\pi_j\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)}{1+\sum_{j=1}^{n}\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)})dP_v(v_i)\\
&=\int\frac{\sum_{r=1}^{n-l}[\sum_{j=1}^{l}(\pi_j-\pi_{l+r})\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_{l+r}]\exp(\gamma_{i,l+r}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_{l+r})}{(1+\sum_{j=1}^{l}\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j))(1+\sum_{j=1}^{n}\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j))}dP_v(v_i).
\end{aligned}$$

Because when $\bar{\alpha}_i>c_{im}$,

$$\begin{aligned}
&\int[\frac{[\sum_{j=1}^{m}(\pi_j-\pi_{m+1})\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_{m+1}]\exp(\gamma_{i,m+1}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_{m+1})}{(1+\sum_{j=1}^{m}\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j))(1+\sum_{j=1}^{m+1}\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j))}]dP_v(v_i)<0\\
&\Leftrightarrow\int[\sum_{j=1}^{m}(\pi_j-\pi_{m+1})\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_{m+1}]f_i(0|\vec{m})f_i(n|\vec{n})dP_v(v_i)<0\\
&\Rightarrow\int[\sum_{j=1}^{m}(\pi_j-\pi_{m+1})\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_{m+1}]f_i(0|\vec{l})f_i(l+r|\vec{n})dP_v(v_i)<0\\
&\Rightarrow\int[\sum_{j=1}^{l}(\pi_j-\pi_{m+1})\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_{m+1}]f_i(0|\vec{l})f_i(l+r|\vec{n})dP_v(v_i)<0\\
&\Rightarrow\int[\sum_{j=1}^{l}(\pi_j-\pi_{l+r})\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_{l+r}]f_i(0|\vec{l})f_i(l+r|\vec{n})dP_v(v_i)<0.
\end{aligned}$$

The second last step is correct because

$$\sum_{j=1}^{l}(\pi_j-\pi_{m+1})\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_{m+1}<\sum_{j=1}^{m}(\pi_j-\pi_{m+1})\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_{m+1},$$

for any $l\in\{1,2,...,m-1\}$. And the last step is correct because

$$\sum_{j=1}^{l}(\pi_j-\pi_{l+r})\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_{l+r}\leq\sum_{j=1}^{l}(\pi_j-\pi_{m+1})\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_{m+1},$$

for any $r\in\{1,2,...,n-l\}$. So, the inequality $EP(\vec{l})<EP(\vec{n})$ holds for any $l\in\{1,2,...m-1\}$.

A.3 Lemma 3

Lemma 3: When $\mathcal{J}_i$ is optimal among $\vec{\mathbb{J}}$, $\mathcal{J}_i$ strictly dominates any subset n_s lacking internal products of $\vec{n}\in\vec{\mathbb{J}}$ (not product 1 or n).

Proof: Let $\mathcal{J}_i=\vec{m}$ and start from the case $m\geq 3$.

1): For the case $n=m$, i.e., compare $\vec{m}$ and its own subsets. Let $\mathbf{k}$ be the set of missing internal products of $\vec{m}$, i.e., $\forall k\in\mathbf{k}, 1<k<m$. For the subset $n_s=m_s=\vec{m}\backslash\mathbf{k}$, we have

$$\begin{aligned}
&EP(\vec{m})-EP(n_s)\\
&=\int(\frac{\sum_{j\in\vec{m}}\pi_j\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)}{1+\sum_{j\in\vec{m}}\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)}-\frac{\sum_{j\in n_s}\pi_j\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)}{1+\sum_{j\in n_s}\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)})dP_v(v_i)\\
&=\int(-\frac{\sum_{k\in\mathbf{k}}[\sum_{j=1}^{k-1}(\pi_j-\pi_k)\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)-\pi_k]\exp(\gamma_{ik}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_k)}{(1+\sum_{j\in\vec{m}}\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j))(1+\sum_{j\in n_s}\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j))}\\
&+\frac{\sum_{k\in\mathbf{k}}[\sum_{j=k+1}^{m}(\pi_k-\pi_j)\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j)]\exp(\gamma_{ik}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_k)}{(1+\sum_{j\in\vec{m}}\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j))(1+\sum_{j\in n_s}\exp(\gamma_{ij}-(\bar{\alpha}_i+\sigma_\alpha v_{i\alpha})p_j))})dP_v(v_i).
\end{aligned}$$

When $\bar{\alpha}_i \geq c_{i,m-1} > c_{i,k-1}$, we have

$$\int \frac{\sum_{j=1}^{k-1}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k]\exp(\gamma_{ik} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_k)dP_0(v)}{(1+\sum_{j=1}^{k-1}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1+\sum_{j=1}^{k}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))} dP_v(v_i)$$
$$= \int [\sum_{j=1}^{k-1}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k] f_i(0|\overrightarrow{k-1}) f_i(k|\vec{k}) dP_v(v_i) < 0$$
$$\Rightarrow \int [\sum_{j=1}^{k-1}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k] f_i(0|n_s) f_i(k|\vec{m}) dP_v(v_i) < 0,$$

$\forall k \in \mathbf{k}, 1 < k < m$. Then, we have $EP(\vec{m}) > EP(n_s)$.

2): For the case $n > m$, let $\mathbf{k}$ be the set of missing internal products of $\vec{n}$, i.e., $\forall k \in \mathbf{k}, 1 < k < n$. For the subset $n_s = \vec{n} \backslash \mathbf{k}$, we have

$$EP(\vec{m}) - EP(n_s)$$
$$= \int (\frac{\sum_{j=1}^{m}\pi_j \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)}{1+\sum_{j=1}^{m}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)} - \frac{\sum_{j\in n_s}\pi_j \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)}{1+\sum_{j\in n_s}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)}) dP_v(v_i)$$
$$= \int (-\frac{\sum_{k\in\mathbf{k}}[\sum_{j=1}^{m}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k]\exp(\gamma_{ik} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_k)}{(1+\sum_{j=1}^{m}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1+\sum_{j\in n_s}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))}$$
$$+ \frac{\sum_{r=1}^{n-m}[\sum_{j=1}^{m}(\pi_j - \pi_{m+r})\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+r}]\exp(\gamma_{i,m+r} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_{m+r})}{(1+\sum_{j=1}^{m}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1+\sum_{j\in n_s}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))}) dP_v(v_i).$$

For any $k > m$, the part

$$\int -\frac{[\sum_{j=1}^{m}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k]\exp(\gamma_{ik} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_k)}{(1+\sum_{j=1}^{m}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1+\sum_{j\in n_s}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))} dP_v(v)$$

will cancel out with one part of

$$\int \frac{\sum_{r=1}^{n-m}[\sum_{j=1}^{m}(\pi_j - \pi_{m+r})\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+r}]\exp(\gamma_{i,m+r} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_{m+r})}{(1+\sum_{j=1}^{m}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1+\sum_{j\in n_s}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))}] dP_v(v).$$

After the cancellation, the rest $m + r \in n_s$. When $\bar{\alpha}_i < c_{im}$, we have

$$\int \frac{[\sum_{j=1}^{m}(\pi_j - \pi_{m+1})\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+1}]\exp(\gamma_{i,m+1} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_{m+1})}{(1 + \sum_{j=1}^{m}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1 + \sum_{j=1}^{m+1}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))} dP_v(v_i)$$
$$= \int [\sum_{j=1}^{m}(\pi_j - \pi_{m+1})\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+1}] f_i(0|\vec{m}) f_i(m+1|\overrightarrow{m+1}) dP_v(v_i) > 0$$
$$\Rightarrow \int [\sum_{j=1}^{m}(\pi_j - \pi_{m+1})\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+1}] f_i(0|\vec{m}) f_i(m+r|n_s) dP_v(v_i) > 0$$
$$\Rightarrow \int [\sum_{j=1}^{m}(\pi_j - \pi_{m+r})\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+r}] f_i(0|\vec{m}) f_i(m+r|n_s) dP_v(v_i) > 0.$$

The last step is correct because

$$\sum_{j=1}^{m}(\pi_j - \pi_{m+r})\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+r} \geq \sum_{j=1}^{m}(\pi_j - \pi_{m+1})\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{m+1}$$

$\forall r \geq 1$.

For any $k \leq m$, when $\bar{\alpha}_i \geq c_{i,m-1} \geq c_{i,k-1}$, we have

$$\int \frac{\sum_{j=1}^{k-1}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k]\exp(\gamma_{ik} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_k)}{(1 + \sum_{j=1}^{k-1}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1 + \sum_{j=1}^{k}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))} dP_v(v)$$
$$= \int [\sum_{j=1}^{k-1}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k] f_i(0|\overrightarrow{k-1}) f_i(k|\vec{k}) dP_v(v_i) \leq 0$$
$$\Rightarrow \int [\sum_{j=1}^{k-1}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k] f_i(0|n_s) f_i(k|\vec{m}) dP_v(v_i) \leq 0$$
$$\Rightarrow \int [\sum_{j=1}^{m}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k] f_i(0|n_s) f_i(k|\vec{m}) dP_v(v_i) \leq 0.$$

The last step is correct because

$$\sum_{j=1}^{m}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k \leq \sum_{j=1}^{k-1}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k$$

for any $k \leq m$. So we have $EP(\vec{m}) > EP(n_s)$ for all $n > m$.

3): For the case $n < m$, let $\mathbf{k}$ be the set of missing internal products of $\vec{n}$, i.e., $\forall k \in \mathbf{k}, 1 < k < n$. For the subset $n_s = \vec{n}\backslash\mathbf{k}$, we have

$$\begin{aligned}
&EP(\vec{m}) - EP(n_s)\\
&= \int \Big(\frac{\sum_{j=1}^{m}\pi_j \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)}{1 + \sum_{j=1}^{m}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)} - \frac{\sum_{j\in n_s}\pi_j \exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)}{1 + \sum_{j\in n_s}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j)}\Big)dP_v(v_i)\\
&= \int \Big(-\frac{\sum_{k\in\mathbf{k}}[\sum_{j=1}^{m}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k]\exp(\gamma_{ik} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_k)}{(1 + \sum_{j=1}^{m}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1 + \sum_{j\in n_s}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))}\\
&- \frac{\sum_{r=1}^{m-n}[\sum_{j=1}^{n}(\pi_j - \pi_{n+r})\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{n+r}]\exp(\gamma_{i,n+r} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_{n+r})}{(1 + \sum_{j=1}^{m}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1 + \sum_{j\in n_s}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))}\Big)dP_v(v_i).
\end{aligned}$$

For any $k < n < m$ and when $\bar{\alpha}_i \geq c_{i,m-1} > c_{i,k-1}$,

$$\begin{aligned}
&\int \frac{\sum_{j=1}^{k-1}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k]\exp(\gamma_{ik} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_k)dP_0(v)}{(1 + \sum_{j=1}^{k-1}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1 + \sum_{j=1}^{k}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))}dP_v(v_i)\\
&= \int [\sum_{j=1}^{k-1}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k]f_i(0|\overrightarrow{k-1})f_i(k|\vec{k})dP_v(v_i) < 0\\
&\Rightarrow \int [\sum_{j=1}^{k-1}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k]f_i(0|n_s)f_i(k|\vec{m})dP_v(v_i) < 0\\
&\Rightarrow \int [\sum_{j=1}^{m}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k]f_i(0|n_s)f_i(k|\vec{m})dP_v(v_i) < 0.
\end{aligned}$$

The last step is correct because

$$\sum_{j=1}^{m}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k < \sum_{j=1}^{k-1}(\pi_j - \pi_k)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_k$$

for any $k < m$. Similarly, when $\bar{\alpha}_i \geq c_{i,m-1}$, we have

$$\begin{aligned}
&\int \frac{\sum_{j=1}^{m-1}(\pi_j - \pi_m)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_m]\exp(\gamma_{im} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_m)}{(1+\sum_{j=1}^{m-1}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))(1+\sum_{j=1}^{m}\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j))} dP_v(v_i) \\
&= \int [\sum_{j=1}^{m-1}(\pi_j - \pi_m)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_m] f_i(0|\overrightarrow{m-1}) f_i(m|\vec{m}) dP_v(v_i) \leq 0 \\
&\Rightarrow \int [\sum_{j=1}^{m-1}(\pi_j - \pi_m)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_m] f_i(0|n_s) f_i(n+r|\vec{m}) dP_v(v_i) \leq 0 \\
&\Rightarrow \int [\sum_{j=1}^{n}(\pi_j - \pi_m)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_m] f_i(0|n_s) f_i(n+r|\vec{m}) dP_v(v_i) \leq 0 \\
&\Rightarrow \int [\sum_{j=1}^{n}(\pi_j - \pi_{n+r})\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{n+r}] f_i(0|n_s) f_i(n+r|\vec{m}) dP_v(v_i) \leq 0.
\end{aligned}$$

The second-last step holds because

$$\sum_{j=1}^{n}(\pi_j - \pi_m)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_m \leq \sum_{j=1}^{m-1}(\pi_j - \pi_m)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_m$$

for any $n < m$. And the last step holds because

$$\sum_{j=1}^{n}(\pi_j - \pi_{n+r})\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_{n+r} \leq \sum_{j=1}^{n}(\pi_j - \pi_m)\exp(\gamma_{ij} - (\bar{\alpha}_i + \sigma_\alpha v_{i\alpha})p_j) - \pi_m$$

for any $r \in \{1, 2, ..., m-n\}$. So we have $EP(\vec{m}) > EP(n_s)$ for all $n < m$.

B Proof of Theorem 2

Theorem 2 Let $h(\delta) = \delta + ln(s^{obs}) - ln(s(p, \pi, X, \delta; \theta_2, P_v))$, $h(\delta)$ is a contraction of modulus less than one.

I prove by showing that $h(\delta)$ is a function that falls under the assumptions of the theorem stated in Appendix 1 of Berry et al. (1995). For simplicity, I suppress the subscript for market (l) in this proof.

(1) $\forall \delta \in R^J$, $h(\delta)$ is continuously differentiable with, $\forall j$ and k,

$$\partial h(\delta_j)/\partial \delta_k \geq 0$$

and

$$\sum_{k=1}^{J} \partial h(\delta_j)/\partial \delta_k < 1.$$

Proof: Given equation (1),

$$s_j(p, \pi, X, \delta; \theta_2, P_v) = \int f_{ij}(v_i, D_i, p, X, \delta; \theta_2, \mathcal{J}_i^*) dP_v(v_i),$$

where

$$f_{ij}(v_i, D_i, p, X, \delta; \theta_2, \mathcal{J}_i^*) = \begin{cases} \frac{\exp(\delta_j + \mu_{ij})}{1 + \sum_{k \in \mathcal{J}_i^*} \exp(\delta_k + \mu_{ik})}, & \text{if } j \in \mathcal{J}_i^* \\ 0, & \text{otherwise} \end{cases}.$$

Then, for $k = j$, I have

$$\begin{aligned}
&\frac{\partial s_j(p, \pi, X, \delta; \theta_2, P_v)}{\partial \delta_j} \\
&= \int f_{ij}(v_i, D_i, p, X, \delta; \theta_2, \mathcal{J}_i^*)[1 - f_{ij}(v_i, D_i, p, X, \delta; \theta_2, \mathcal{J}_i^*)] dP_v(v_i) \\
&= s_j(p, \pi, X, \delta; \theta_2, P_v) - \int f_{ij}^2(v_i, D_i, p, X, \delta; \theta_2, \mathcal{J}_i^*) dP_v(v_i) \\
&< s_j(p, \pi, X, \delta; \theta_2, P_v).
\end{aligned}$$

I have, $\forall k \neq j$,

$$\begin{aligned}
&\frac{\partial s_j(p, \pi, X, \delta; \theta_2, P_v)}{\partial \delta_k} \\
&= -\int f_{ij}(v_i, D_i, p, X, \delta; \theta_2, \mathcal{J}_i^*) f_{ik}(v_i, D_i, p, X, \delta; \theta_2, \mathcal{J}_i^*) dP_v(v_i) \\
&\leq 0.
\end{aligned}$$

Then,

$$\sum_{k=1}^{J} \frac{\partial s_j(p, \pi, X, \delta; \theta_2, P_v)}{\partial \delta_k} < s_j(p, \pi, X, \delta; \theta_2, P_v).$$

Given $h(\delta_j) = \delta_j + ln(s_j^{obs}) - ln(s_j(p, \pi, X, \delta; \theta_2, P_v))$, taking the derivative gives

$$\frac{\partial h(\delta_j)}{\partial \delta_j} = 1 - \frac{1}{s_j(p, \pi, X, \delta; \theta_2, P_v)} \frac{\partial s_j(p, \pi, X, \delta; \theta_2, P_v)}{\partial \delta_j} > 0;$$

and $\forall k \neq j$,

$$\frac{\partial h(\delta_j)}{\partial \delta_k} = -\frac{1}{s_j(p, \pi, X, \delta; \theta_2, P_v)} \frac{\partial s_j(p, \pi, X, \delta; \theta_2, P_v)}{\partial \delta_k} \geq 0;$$

and

$$
\begin{aligned}
&\sum_{k=1}^{J} \frac{\partial h(\delta_j)}{\partial \delta_k} \\
&= 1 - \frac{1}{s_j(p,\pi,X,\delta;\theta_2,P_v)} \sum_{k=1}^{J} \frac{\partial s_j(p,\pi,X,\delta;\theta_2,P_v)}{\partial \delta_k} \\
&= 1 - \frac{1}{s_j(p,\pi,X,\delta;\theta_2,P_v)} \{ \frac{\partial s_j(p,\pi,X,\delta;\theta_2,P_v)}{\partial \delta_j} + \sum_{k \in \vec{J} \backslash j} \frac{\partial s_j(p,\pi,X,\delta;\theta_2,P_v)}{\partial \delta_k} \} \\
&= 1 - \frac{1}{s_j(p,\pi,X,\delta;\theta_2,P_v)} \{ s_j(p,\pi,X,\delta;\theta_2,P_v) - \int f_{ij}^2(v_i,D_i,p,X,\delta;\theta_2,\mathcal{J}_i^*) dP_v(v_i) \\
&- \sum_{k \in \vec{J} \backslash j} \int f_{ij}(v_i,D_i,p,X,\delta;\theta_2,\mathcal{J}_i^*) f_{ik}(v_i,D_i,p,X,\delta;\theta_2,\mathcal{J}_i^*)) dP_v(v_i) \} \\
&= \frac{1}{s_j(p,\pi,X,\delta;\theta_2,P_v)} \{ \int f_{ij}^2(v_i,D_i,p,X,\delta;\theta_2,\mathcal{J}_i^*) dP_v(v_i) \\
&+ \sum_{k \in \vec{J} \backslash j} \int f_{ij}(v_i,D_i,p,X,\delta;\theta_2,S_i^*) f_{ik}(v_i,D_i,p,X,\delta;\theta_2,\mathcal{J}_i^*)) dP_v(v_i) \} \\
&= \frac{1}{s_j(p,\pi,X,\delta;\theta_2,P_v)} \{ \int f_{ij}(v_i,D_i,p,X,\delta;\theta_2,\mathcal{J}_i^*) [\sum_{k=1}^{J} f_{ik}(v_i,D_i,p,X,\delta;\theta_2,\mathcal{J}_i^*)] dP_v(v_i) \} \\
&< \frac{1}{s_j(p,\pi,X,\delta;\theta_2,P_v)} \{ \int f_{ij}(v_i,D_i,p,X,\delta;\theta_2,\mathcal{J}_i^*) dP_v(v_i) \} = 1.
\end{aligned}
$$

Because $\sum_{k=1}^{J} f_{ik}(v_i, D_i, p, X, \delta; \theta_2, \mathcal{J}_i^*) < 1$ for any assortment $\mathcal{J}_i^*$, since the sum does not include the outside good.

(2) The function h has a finite lower bound.

Keeping every consumer i with $j \in \mathcal{J}_i^*$, then

$$\begin{aligned} h(\delta_j) &= \delta_j + ln(s_j^{obs}) - ln(s_j(p, \pi, X, \delta; \theta_2, P_v)) \\ &= \delta_j + ln(s_j^{obs}) - ln(\int \frac{\exp(\delta_j + \mu_{ij})}{1 + \sum_{k \in \mathcal{J}_i^*} \exp(\delta_k + \mu_{ik})} dP_v(v_i)) \\ &= \delta_j + ln(s_j^{obs}) - ln(\exp(\delta_j) \int \frac{\exp(\mu_{ij})}{1 + \sum_{k \in \mathcal{J}_i^*} \exp(\delta_k + \mu_{ik})} dP_v(v_i)) \\ &= ln(s_j^{obs}) - ln(\int \frac{\exp(\mu_{ij})}{1 + \sum_{k \in \mathcal{J}_i^*} \exp(\delta_k + \mu_{ik})} dP_v(v_i)). \end{aligned}$$

If δ_j goes to $-\infty$, $h(\delta_j)$ converges to

$$ln(s_j^{obs}) - ln(\int \exp(\mu_{ij}) dP_v(v_i)).$$

C Summary Statistics of Cigarette Sales

Table 9: Summary Statistics of Cigarette Sales

Tier \ Year	2011	2012	2013	2014	2015	2016
I	99.37 (84.1)	120.26 (97.87)	138.55 (111.01)	161.8 (127.89)	168.21 (136.13)	156.07 (132.12)
II	48.02 (43.14)	60.19 (50.98)	69.88 (57.84)	83.53 (66.89)	91.88 (67.72)	95.71 (70.56)
III	297.44 (216.02)	356.2 (239.92)	369.41 (248.82)	362.23 (236.06)	347.95 (228.35)	326.51 (219.95)
IV	219.48 (138.67)	169.19 (100.73)	155.18 (92.32)	149 (91.59)	135.76 (87.19)	125.93 (84.19)
V	102.04 (74.06)	76.81 (53.88)	58.64 (40.68)	51.25 (38.97)	45.82 (36.78)	40.9 (33.56)
N	31	31	31	31	31	31

Note: This table is based on the information provided by the China Tobacco Yearbooks. Data are in units of 100 million cigarettes. Standard deviations are in parentheses.

D Cigarette Tax Increases

There were two exogenous tax increases in the industry: one occurring in May 2009 and the other in May 2015. Table 10 summarizes the details of the tax increases. The allocation price in Table 10 is the price negotiated between STMA and the tax authority in China (i.e., the State Administration of Taxation). Under China's monopoly system, STMA makes cigarette 'allocation plans' so that the price in the process is known as allocation price, which is also the price tobacco producers offer to wholesalers. Since June 2001, the allocation price replaced producer price and became the tax base for cigarette ad valorem excise tax. The tax authority uses the allocation price for tax collection purposes, and STMA uses it as a criterion to categorize cigarettes into five tiers (Tier I - Tier V), where Tier I represents top-price cigarettes (Gao et al. (2012)).

The tax adjustment in 2009 mainly includes: (1) Grade A (Tier I and II) cigarettes' excise tax rates increasing from 45 to 56%, and Grade B (Tier III, IV and V) cigarettes excise tax rates increasing from 30 to 36%; (2) 5% of ad valorem excise tax applied additionally at the wholesale level; (3) specific tax remains at ¥0.06 per pack; (4) adjusted standards for Grade A and Grade B cigarettes where products with allocation prices greater than or equal to ¥7 per pack (20 cigarettes) were considered Grade A, while those with allocation prices less than ¥7 were considered Grade B. In contrast, the price threshold was ¥5 before the tax adjustment. Accordingly, the classification standards for cigarette Tiers II and III also changed. The Chinese government raised cigarette excise tax again at the wholesale level in 2015, which increased the ad valorem excise tax rate from 5% to 11% and added a ¥0.10/pack specific tax for all tiers (Zheng (2018)).

E Calculation of Wholesale Margins

Equation (4) (first introduced in Gao et al. (2012)) demonstrates China's cigarette pricing mechanism where A represents allocation price, a indicates the allocation-wholesale profit

Table 10: China Cigarette Excise Tax

	Before May 2009	After May 2009	After May 2015
Producer level			
Specific excise tax	¥0.06 /pack	¥0.06 /pack	¥0.06 /pack
Ad valorem excise tax rate			
Class A cigarettes	Allocation price $\geq$ ¥5 /pack 45%	Allocation price $\geq$ ¥7 /pack 56%	Allocation price $\geq$ ¥7 /pack 56%
Class B cigarettes	Allocation price $<$ ¥5 /pack 30%	Allocation price $<$ ¥7 /pack 36%	Allocation price $<$ ¥7 /pack 36%
Wholesale level			
Specific excise tax	¥0 /pack	¥0 /pack	¥0.10 /pack
Ad valorem excise tax rate	0%	5%	11%

Note: Grade A cigarettes include Tiers I and II. Grade B cigarettes include Tiers III, IV and V. The standard changed in May 2009.

margin, b indicates the wholesale-retail profit margin, and $Rtvat$ indicates the VAT (value-added tax) rate. P_r is the retail price at which retailers sell cigarettes to consumers, and it is the final price of tobacco products. All cigarette retailers are required to obtain cigarette sales permission from STMA and CNTC.

$$P_r = A \times (1 + a) \times (1 + b) \times (1 + Rtvat). \tag{4}$$

As addressed, retail price is equal to the allocation price plus the allocation-wholesale margin, wholesale-retail margin, and VAT tax. A includes excise tax but excludes VAT, so no excise tax appears in equation (4). VAT is collected at all circulation segments, including produce, wholesale, and retail, and uses the added value at each level as the tax base. Therefore, putting the factor $(1 + Rtvat)$ at the end of the equation includes all collected VAT (Zheng (2018)).

Under China's cigarette monopoly system, both (a) allocation-wholesale profit margin and (b) wholesale-retail profit margin are set by STMA. Given the allocation price, STMA can easily control wholesale price and retail price by adjusting (a) allocation-wholesale margin or (b) wholesale-retail margin. Table 11 summarizes the allocation-wholesale margin adjustment in 2009, and we can see high-priced cigarettes have margin rate no lower than low-priced cigarettes.

Calculation procedure for wholesale after-tax price-cost margins.

Table 11: Cigarette allocation prices and profit margins

Tier	Allocation Price	Allocation-wholesale Margin (%) After May 2009	Allocation-wholesale Margin (%) Before May 2009	Wholesale-retail Margin (%)
I	[10, ∞)	31.5	47	15
II	[7,10)	25	43	15
III	[5,7)	25	43	15
III	[3,5)	25	38	10
IV	[1.65, 3)	20	28	10
V	(0, 1.65)	15	18	10

Note: Tier III allocation price falls into either [3,5) and [5,7) ranges after May 2009. Before May 2009, price range [3,5) belonged to Tier III while price range [5,7) belonged to Tier II. Margin rates are collected from Zheng (2018).

1. Rearrange equation (4)

$$Pr = A \times (1 + a) \times (1 + b) \times (1 + Rtvat),$$

   to get the allocation price

$$A = \frac{B}{(1 + a) \times (1 + Rtvat)},$$

   where B represents the wholesale price in Table 2, a indicates the allocation-wholesale profit margin in Table 10, and $Rtvat$ indicates the 17% VAT rate.

2. Calculate after-tax wholesale margin using

$$\pi_w = A \times a - A \times t_a - t_s,$$

   where π_w represents the after-tax wholesale margin, t_a indicates the wholesale ad valorem excise tax rate, and t_s indicates the wholesale specific excise tax listed in Table 12.

F Monte Carlo Experiment

F.1 Data Generation Process

Use the case of five products. I set the nominal price vector as $p = \{3.6, 2.4, 1.6, 1.2, 1\}$, and the nominal profit vector as $\pi = \{3, 2, 1.2, 0.9, 0.75\}$, so that product 1 is the most expensive one and has the highest unit profit. I simulate 100 markets. All markets have the same nominal prices and profits but have different inflation rates. I generate random market inflation rates using $e \sim U(0, 0.1)$ with size 100×1, and divide p and π by $1 + e$ to get real prices and profits. Products differ in qualities where the more expensive product has better quality. Consumers' mean preferences for product qualities are fully captured by the mean fixed utilities ξ. I decompose the mean fixed utilities linearly into a constant which is the same in all markets, and a random deviation from the constant. The mean fixed utility of product j in market l satisfies $\xi_{jl} = \bar{\xi}_j + \Delta\xi_{jl}$, where $\bar{\xi} = \{\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3, \bar{\xi}_4, \bar{\xi}_5\} = \{1, 0.8, 0.6, 0.4, 0.2\}$, and $\Delta\xi \sim N(0, 0.01)$. Individual deviation from the mean fixed utilities v_{ij} follows independent standard normal distributions, and the scale matrix for v_{ij} is $\Sigma = 0.5 \times I_5$, i.e., $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5 = 0.5$.

I assume consumer i's deterministic price sensitivity is a function of real disposable income with the form $\bar{\alpha}_i = \alpha - \alpha_{inc} \times inc_i$, where inc_i is the real disposable income of consumer i, and α and α_{inc} are parameters of the function. With the above setup, the price sensitivity cutoffs range from 0.0373 to 1.0681. To ensure the simulated price sensitivities cover all product assortments, I set $\alpha_{inc} = 0.5$ and generate $inc_i \sim U(0, 3)$ with size 10000×1 for each market l. I set $\alpha = \alpha_{inc} \times max(inc) + e^{-6}$ (1.5025 in the simulated dataset), so that all consumers have positive deterministic price sensitivities and the wealthiest consumer in all markets has deterministic price sensitivity close to zero ($\bar{\alpha}_{min} = e^{-6}$). Individual deviation from deterministic price sensitivities $v_{i\alpha}$ also follows a standard normal distribution with a scaling scalar of $\sigma_\alpha = 0.1$. Then, for each market l, I generate the observed market shares s_{jl}^{obs} using the following procedure.

1. Calculate the cutoffs of $\bar{\alpha}_i$ using the set of equations

$$\sum_{j=1}^{m}(\pi_{jl} - \pi_{m+1,l}) \exp(\xi_{jl} - c_{ml} p_{jl}) - \pi_{m+1,l} = 0,$$

   for $m \in \{1, 2, 3, 4\}$.

2. Randomly draw the vector of consumer characteristics $v_i = \{v_{i1}, v_{i2}, ..., v_{i5}, v_{i\alpha}\}$ from independent standard normal distributions, and $inc_i \sim U(0, 3)$ for 10,000 consumers, and calculate the values $\bar{\alpha}_i = \alpha - \alpha_{inc} \times inc_i$, $\alpha_i = \bar{\alpha}_i + \sigma_\alpha v_{i\alpha}$ and $\xi_{ijl} = \xi_{jl} + \sigma_j v_{ij}$.

3. Classify each consumer i into the optimal product assortment $\mathcal{J}_i^*$ according to Theorem 1, and find the optimal choice j of consumer i using

$$U_{ijl} \geq U_{ikl}, \text{ for } j, k \in \{0, \mathcal{J}_i^*\},$$

   where $U_{ijl} = \xi_{ijl} - \alpha_i p_{jl} + \epsilon_{ijl}$.

4. Sum up the number of consumers who choose product j, then divide by the sample size of 10,000 to obtain the observed market share of product j in market l, s_{jl}^{obs}.

F.2 Estimation Procedure and Results

The detailed estimation procedure for the foldable menu model is as follows.

1. For each market l, use random starting values of $\xi_l = (\xi_{1l}, \xi_{2l}, ..., \xi_{5l})$ to calculate the cutoffs of $\bar{\alpha}_i$ using

$$\sum_{j=1}^{m}(\pi_{jl} - \pi_{m+1,l}) \exp(\xi_{jl} - c_{ml} p_{jl}) - \pi_{m+1,l} = 0,$$

   for $m \in \{1, 2, 3, 4\}$.

2. Randomly draw a different vector of consumer characteristics $v_i' = \{v_{i1}', v_{i2}', ..., v_{i5}', v_{i\alpha}'\}$ from independent standard normal distributions for 10,000 consumers, and obtain the vector of $\bar{\alpha}_i = \alpha - \alpha_{inc} \times inc_i$, $\alpha_i = \bar{\alpha}_i + \sigma_\alpha v_{i\alpha}'$, and $\xi_{ijl} = \xi_{jl} + \sigma_j v_{ij}'$ for any given values of $\theta_2 = (\alpha, \alpha_{inc}, \sigma_\alpha, \sigma_j$ for $j = 1, 2, ..., 5)$.

3. Classify each consumer i into the optimal product assortment $\mathcal{J}_i^*$ according to Theorem 1, then calculate the probability of consumer i choosing product $j \in \mathcal{J}_i^*$ using

$$f_j(v_i', inc_i, p_l, \xi_l; \theta_2, \mathcal{J}_i^*) = \begin{cases} \frac{\exp(\xi_{ijl} - \alpha_i p_{jl})}{1 + \sum_{k \in \mathcal{J}_i^*} \exp(\xi_{ikl} - \alpha_i p_{kl})}, & \text{if } j \in \mathcal{J}_i^* \\ 0, & \text{otherwise} \end{cases}.$$

4. Calculate the predicted market share of product j in market l using

$$s_{jl}(p_l, \pi_l, \xi_l; \theta_2, P_{ns}) = \frac{1}{ns} \sum_{i=1}^{ns} f_j(v_i', inc_i, p_l, \xi_l; \theta_2, \mathcal{J}_i^*).$$

5. Calculate converged ξ_{jl}^* iteratively using the contraction mapping $h(\xi_{jl}) = \xi_{jl} + ln(s_{jl}^{obs}) - ln(s_{jl}(p_l, \pi_l, \xi_l; \theta_2, P_{ns}))$.

6. Regress ξ^* (vector of all ξ_{jl}^*) on the five product dummies to obtain the estimates of $\bar{\xi}$, and calculate the predicted residuals of the regression, which are the random demand shocks $\Delta\xi_{jl}$.

7. Search for values of θ_2 to minimize the absolute value of the covariance between real product prices and the random demand shocks, based on the assumption that they are independent.

On the other hand, if we ignore the issue of varying product availabilities and assume all products are equally available to every consumer, we would calculate the predicted market share for product j in market l using

$$s_{jl}(p_l, \xi_l; \theta_2, P_{ns}) = \frac{1}{ns} \sum_{i=1}^{ns} \frac{\exp(\xi_{ijl} - \alpha_i p_{jl})}{1 + \sum_{k=1}^{J} \exp(\xi_{ikl} - \alpha_i p_{kl})}.$$

Now, the model is just the random coefficient logit model, which can be estimated using steps 2, 5, 6, and 7 of the above procedure. To obtain bootstrap standard errors, I repeat the above estimation procedures 20 times for each model with a different consumer characteristics vector v_i for each estimation. Table 12 summarizes the estimation results of the foldable menu model and the random coefficient logit model using the same simulated dataset.

As shown in Table 12, the estimates are unbiased when using the estimation procedure for the foldable menu model. On the other hand, if we ignore varying product availabilities, and thus use the random coefficient logit model for the demand estimation, most estimates are biased as expected. The parameters for deterministic price sensitivities and product mean utilities are underestimated when using the random coefficient logit model.

Table 12: Estimation results using the simulated dataset

Coefficient	True value	Foldable menu	Random coefficient logit
α	1.5025	1.4769	0.549
		(0.0289)	(0.0147)
α_{inc}	0.5	0.4915	0.1827
		(0.0097)	(0.0049)
σ_α	0.1	0.1001	0.1028
		(0.0022)	(0.0063)
σ_1	0.5	0.5123	0.5454
		(0.0156)	(0.0213)
σ_2	0.5	0.5064	0.5352
		(0.0121)	(0.0198)
σ_3	0.5	0.5026	0.5131
		(0.0088)	(0.01)
σ_4	0.5	0.5048	0.5776
		(0.007)	(0.0264)
σ_5	0.5	0.4981	0.519
		(0.0094)	(0.0206)
$\bar{\xi}_1$	1	0.9706	-0.0259
		(0.0303)	(0.0234)
$\bar{\xi}_2$	0.8	0.7724	-0.1977
		(0.0438)	(0.0142)
$\bar{\xi}_3$	0.6	0.5808	-0.8858
		(0.0286)	(0.0137)
$\bar{\xi}_4$	0.4	0.376	-1.327
		(0.0172)	(0.016)
$\bar{\xi}_5$	0.2	0.2082	-1.6914
		(0.0045)	(0.0146)

Note: Coefficient values are the mean estimates of 20 estimations. Bootstrap standard errors are in parentheses. All coefficients are statistically significance at 1% significance level.

References

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Bishop, J. A., L. Haiyong, and M. Qi (2007). Are chinese smokers sensitive to price? *China Economic Review 18*(2), 113–121.

Bruno, H. A. and N. J. Vilcassim (2008). Research note—structural demand estimation with varying product availability. *Marketing Science 27*(6), 1126–1131.

Chen, Y. and W. Xing (2016). Quantity, quality, and regional price variation of cigarettes: demand analysis based on a household survey in china. In *Economics of tobacco control in China: from policy research to practice*, pp. 61–87. World Scientific.

Ciliberto, F., C. Murry, and E. Tamer (2018). Market structure and competition in airline markets.

D'Haultfœuille, X., I. Durrmeyer, and P. Février (2019). Automobile prices in market equilibrium with unobserved price discrimination. *The Review of Economic Studies 86*(5), 1973–1998.

Draganska, M., M. Mazzeo, and K. Seim (2009). Beyond plain vanilla: Modeling joint product assortment and pricing decisions. *QME 7*(2), 105–146.

Dubé, J.-P., A. Hortaçsu, and J. Joo (2021). Random-coefficients logit demand estimation with zero-valued market shares. *Marketing Science*.

Gandhi, A., Z. Lu, and X. Shi (2020). Estimating demand for differentiated products with zeroes in market share data. *Available at SSRN 3503565*.

Gao, S., R. Zheng, and T.-w. Hu (2012). Can increases in the cigarette tax rate be linked to cigarette retail prices? solving mysteries related to the cigarette pricing mechanism in china. *Tobacco control 21*(6), 560–562.

Goodchild, M. and R. Zheng (2018). Early assessment of china's 2015 tobacco tax increase. *Bulletin of the World Health Organization 96*(7), 506.

Gruen, T. W., D. Corsten, and S. Bharadwaj (2002). Retail out-of-stocks: A worldwide examination of causes, rates, and consumer responses. washington. *DC: Grocery Manufacturers of America*.

Hu, T.-W. and Z. Mao (2002). Effects of cigarette tax on cigarette consumption and the chinese economy. *Tobacco control 11*(2), 105–108.

Iaria, A. (2014). Consumer benefit of big-box supermarkets: The importance of controlling for endogenous entry. Technical report, Working paper, University of Warwick.

Lance, P. M., J. S. Akin, W. H. Dow, and C.-P. Loh (2004). Is cigarette smoking in poorer nations highly sensitive to price?: Evidence from russia and china. *Journal of health economics 23*(1), 173–189.

Li, J., J. S. White, T.-w. Hu, G. T. Fong, and J. Yuan (2016). The heterogeneous effects of cigarette prices on brand choice in china: implications for tobacco control policy. In *Economics of Tobacco Control in China: From Policy Research to Practice*, pp. 109–127. World Scientific.

Li, S., J. Mazur, Y. Park, J. Roberts, A. Sweeting, and J. Zhang (2018). Repositioning and market power after airline mergers. *Unpublished Manuscript, University of Maryland*.

Liu, H., J. A. Rizzo, Q. Sun, and F. Wu (2015). How do smokers respond to cigarette taxes? evidence from china's cigarette industry. *Health economics 24*(10), 1314–1330.

Luo, Y., I. Perrigne, and Q. Vuong (2018). Structural analysis of nonlinear pricing. *Journal of Political Economy 126*(6), 2523–2568.

Misra, K. (2008). Understanding retail assortments in competitive markets. *Northwestern University*.

Musalem, A. (2015). When demand projections are too optimistic: A structural model of product line and pricing decisions. *Available at SSRN 2560662*.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307–342.

Quan, T. W. and K. R. Williams (2018). Product variety, across-market demand heterogeneity, and the value of online retail. *The RAND Journal of Economics 49*(4), 877–913.

Shah, D., V. Kumar, and Y. Zhao (2015). Diagnosing brand performance: Accounting for the dynamic impact of product availability with aggregate data. *Journal of Marketing Research 52*(2), 147–165.

Tenn, S. and J. M. Yun (2008). Biases in demand analysis due to variation in retail distribution. *International Journal of Industrial Organization 26*(4), 984–997.

Zheng, R. (2018). Tobacco tax and tobacco economy in china. In *Tobacco Control in China*, pp. 285–309. Springer.