The Pennsylvania State University

The Graduate School

ESSAYS ON CONSIDERATION SET MODELS AND DISCRETE CHOICE MODELS

A Dissertation in
Economics
by
Hyung Gyu Rho

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2022

The dissertation of Hyung Gyu Rho was reviewed and approved by the following:

Joris Pinkse
Professor of Economics
Dissertation Co-Adviser
Co-Chair of Committee

Sung Jae Jun
Professor of Economics
Dissertation Co-Adviser
Co-Chair of Committee

Patrik Guggenberger
Professor of Economics

Yizao Liu
Associate Professor of Agricultural Economics

Marc Henry
Professor of Economics
Graduate Program Director

Abstract

Chapter 1 presents two identification results in a consideration set model. The consideration set model relaxes the assumption that a consumer considers all products in a market. The model allows the consumer to endogenously choose products to consider buying. It is assumed that her attention determines this consideration. I discuss identification of model parameters while allowing arbitrary dependencies between preferences and attention. Specifically, I present two identification results. The first identification argument relies on an exclusion restriction and an identification at infinity concept. The exclusion restriction means that there is a variable that affects preferences but not attention. Further, I assume that this variable has unbounded support. The second identification argument relaxes the unbounded support assumption, but it requires two exclusion restrictions. I need an additional variable that influences attention but not preferences. Under the two exclusion restrictions, I show that I can identify the parameters of interest.

Chapter 2 proposes a nonparametric test of the validity of an exclusion restriction in a consideration set model. It is known that point identification in the consideration set model can be achieved with two variables – one variable only controls preferences, and the other only affects attention. Each variable works as an exclusion restriction as they are excluded from either attention or preferences. Conversely, partial identification can be obtained if only one of these variables is available. My test can be used to test the validity of one exclusion restriction when the other exclusion restriction is available. For example, the econometrician can test whether a candidate variable, in fact, only influences

preferences when there is another variable that only shapes attention. The null hypothesis for my test is that the candidate variable does not affect attention. Consequently, under the null hypothesis, the candidate variable only influences preferences. To develop my test, I start by creating a function that depends on the consumer's attention under the null hypothesis. This function is called the test function. I show that if the test function does not vary with the candidate variable, then the candidate variable has no influence on attention. Consequently, the candidate variable only changes preferences. Then, I compute the difference between the test function and the integral of the test function with respect to some weighting function on the space of the candidate variable. This difference must be zero under the null hypothesis. My test is based on measuring the significance of the square of the difference. I then use the kernel method to derive test statistic. I show that the test statistic is consistent.

Chapter 3 studies the welfare cost incurred by information asymmetry between a consumer and an E-commerce platform when the platform endogenously determines its ranking algorithm. In E-commerce, the platform usually has richer information about products than the consumer. The platform provides the consumer with a list of recommended products, called a ranking. The ranking aims to help the consumer, who has limited information, find the product that suits her needs. The platform can rank products according to the relevance score attached to each product based on the consumer's search query. However, the platform can also place more profitable products at the top of the ranking. This practice can happen because the consumer has limited product information, and her attention is drawn to the first few products in the ranking. To quantify the welfare costs of the information asymmetry in this circumstance, I develop a structural model where the platform decides and commits to its ranking algorithm before the consumer enters the market. After the platform's decision, the consumer enters the market and chooses a product. My result shows that the platform exploits the consumer's ignorance. To conduct a counterfactual analysis, I consider a scenario in which the consumer has complete knowledge. Once the consumer has full knowledge, the platform redesigns the ranking algorithm such that it is more beneficial to the consumer than it is under information asymmetry.. As a result, consumer welfare is improved by 2.24%. Surprisingly, the platform's revenue also increases by about 58%. Giving the consumer complete knowledge appears to balance the bargaining power between the consumer and the E-commerce platform.

Contents

List of Figures

List of Tables

Acknowledgments

First, I would like to thank my advisors, Joris Pinkse and Sung Jae Jun, for their abundant patience and invaluable guidance. It would not have been possible to complete this dissertation without their help. I was fortunate to have them as my advisors.

I want to express my gratitude to Patrick Guggenberger, Yizao Liu, Peter Newberry, and Georgy Artemov for their feedback, comments, and suggestions.

I am also deeply grateful to those who responded anonymously to my interview from hotel chains in Korea, Expedia.com, and online travel agencies in Japan. Their interviews gave me valuable information about the online travel agency industry and helped me understand the data.

Next, I would like to show my appreciation to my friends. I was lucky to go through the PhD program with them. I am grateful to Yuta Suzuki, Xiao Lin, Rui Wang, Gouri Khanna, Hyuk Jun Kwon, Michiharu Chewgen Yamashita, Wei Qiao, and Kaida Zhang for valuable discussions and suggestions. Further, I appreciate Tanji Hwang, Nino Doghonadze, Shomak Chakrabarti, Berk Idem, Ece Teoman, Sian Lee, Hokyu Song, Jonathan Lim, Daniel Tiong, Ken Quach for their emotional support.

Last but not least, I would like to thank my parents and wife for their unconditional support and trust in me.

All errors are my own.

Chapter 1

Identification of a Random Consideration Set Model with Arbitrary Dependence

1.1 Introduction

The traditional discrete choice model assumes that a consumer pays full attention to all products in a given choice set. However, in practice, the consumer pays attention to a smaller subset of the choice set (Shocker, Ben-Akiva, Boccara, & Nedungadi, 1991). This subset is called the consideration set. To capture such consumer behavior, econometricians (e.g., Manski, 1977; Goeree, 2008) develop a consideration set model that allows the consumer to choose a consideration set endogenously. They show that the consideration set model estimates consumer behavior significantly differently than the traditional model does. For example, Goeree (2008) shows that the consideration set model estimates 19% higher markup than the traditional model. In addition, Honka, Hortaçsu, and Vitorino (2017) find that advertising does not influence consumer preferences in the consideration set model but that it positively affects consumer preferences in the traditional model. The results suggest that the consideration set model merits more attention from the econometrician.

The consideration set model captures complex consumer behavior but presents the econometrician with an identification challenge when there are no data for the consideration set. When the econometrician observes that the consumer does not choose a product, the econometrician cannot discern whether the consumer does not pay attention to the product or simply dislikes it. To address this identification problem, the applied literature (e.g., Goeree, 2008) assumes that attention and preferences are independent of one another. This independence assumption makes the consideration set model fall under a finite mixture framework. In finite mixture models, identification can be obtained by relying on exclusion restrictions. However, this independence assumption can lead to an erroneous consumer behavior analysis. Several psychology studies (Ajzen & Fishbein, 1977; Fazio & Zanna, 1981; Ajzen & Fishbein, 2000) suggest a relationship between attention and preferences. According to this psychology research, the assumption of independence can lead to inconsistent estimators. Therefore, economic analysis under this assumption may be inaccurate.

In this chapter, I consider a model in which the consumer's attention and preferences are arbitrarily correlated. Using this framework, I identify the joint distribution of preferences and attention and the parameters of interest. Specifically, I present two identification results. The first identification argument relies on an exclusion restriction and an identification at infinity concept. The exclusion restriction means that there is a variable that influences preferences but not attention. I need infinite support for this variable. The second identification argument relaxes the unbounded support assumption but requires two exclusion restrictions. I need an additional variable that impacts attention but not preferences. I show that the parameters of interest can be identified using the two exclusion restrictions.

This chapter is closely related to the work of Barseghyan et al. (as cited in Barseghyan, Molinari, & Thirkettle, 2019a). They consider consumers who are heterogeneous in risk preferences and attention, and they allow for correlations between preferences and attention. They also place no restrictions on how consideration sets are formed. Using this framework, they show that partial identification can be achieved. In this chapter, I do not consider risk attitude, and I specify how the consumer forms her consideration set. Then I show that point identification can be obtained.

To the best of my knowledge, no work other than Barseghyan et al. (as cited in Barseghyan, Molinari, & Thirkettle, 2019a) involves the identification of the consideration set model in which attention and preferences are correlated. However, there is an extensive body of literature on identification in the consideration set model. Kawaguchi, Uetake, and Watanabe (2016) study the effect of time pressure on attention and preferences. They present an identification strategy without exclusion restrictions. Instead, they obtain identification power from sufficient variation in the choice sets and the presence of at least one variable with infinite support. The key to their identification argument is that the characteristics of one product do not alter the consumer's attention to another product. However, this key point may be invalid if attention from one product to another is correlated.

Abaluck and Adams (2017) also show identification without exclusion restrictions. Their identification result requires similar conditions as in Kawaguchi et al. (2016) – that is, the choice sets must vary from consumer to consumer and at least one covariate must have infinite support. They also assume that the utility function is quasi-linear in the covariate with infinite support. Their identification argument stems from the fact that Slutsky symmetry does not hold in the consideration set model. To exploit Slutsky asymmetry, they assume that attention and preferences are independent of one another.

Barseghyan, Molinari, and Thirkettle (2019a) study the consideration set model assuming that consumers are risk-averse. Their identification result requires independence between attention and preferences and an exclusion restriction – that is, that there is at least one regressor that only affects preferences. This regressor needs to have infinite support for the econometrician to arrive at the identification result. The exclusion restriction, the independence assumption, and the unbounded support assumption allow the econometrician to focus on consumers who choose one particular product over all other products regardless of their risk attitude. In this way, their model concerns attention alone. I use a similar argument for my first identification but allow for arbitrary correlation between attention and preferences.

Nchare (2018) studies dynamic consumer behavior in the consideration set model. His model captures consumer inertia (i.e., the tendency of consumers to choose the same product over time). His identification result is based on the independence between attention and preferences. This assumption implies that his model belongs to a finite mixture framework. As such, his identification result comes directly from the partial identification result of a finite mixture model (Henry, Kitamura, & Salanié, 2014). He works on an arbitrary utility function and requires a weaker exclusion restriction than I do in my work. His exclusion restriction is that one regressor impacts only attention but the regressor can be discrete.

The rest of the chapter is organized as follows: Section 1.2 demonstrates the intuition behind my identification strategies using a simple case. Section 1.3 is devoted to the model. In Section 1.4, I discuss identification. Section 1.5 concludes.

1.2 Example

This section uses simple examples to describe two strategies for identifying the joint distribution of preferences and attention and the parameters of interest. This section aims to provide intuition that carries over to the general case. My first identification arguments are based on the identification at infinity concept and an exclusion restriction. My second identification argument depends on two exclusion restrictions, but it does not rely on the identification at infinity concept.

In this example, a consumer selects a product from three possible options $\{0, 1, 2\}$. Option 0 represents the outside option. To choose a product, she has to pay attention to it and prefer it over other products to which she also pays attention. I assume that the consumer always pays attention to the outside option but may or may not pay attention to options 1 and 2. Her attention to options 1 and 2 depends on the level of advertising, denoted by $\boldsymbol{x}_{i1}$ and $\boldsymbol{x}_{i2}$, and her interest shocks for the options, denoted by $\boldsymbol{\eta}_{i1}$ and $\boldsymbol{\eta}_{i2}$. Her attention level to product j thus is

$$A(\boldsymbol{x}_{ij}, \boldsymbol{\eta}_{ij}) = \boldsymbol{x}_{ij} - \boldsymbol{\eta}_{ij},\ j = 1, 2.$$

If her attention to option j is positive, she considers purchasing product j. The set of products that consumer i considers purchasing is called the consideration set, and it is represented by C_i. After choosing consideration set C_i, the consumer perceives the quality of the products in the consideration set. For example, if her consideration set is $\{0, 1\}$, she realizes the quality of product 1. I assume the quality of the outside option is always zero. $\boldsymbol{z}_{ij}$ denotes the quality of product j.

The timeline of consumer behavior has two important implications. First, since consumer i realizes the value of $\boldsymbol{z}_{ij}$ after the consideration set is formed, variable $\boldsymbol{z}_{ij}$ does not play any role in forming it – the exclusion restriction. Second, without loss of generality, if consumer i's consideration set is $\{0, 1\}$, she ignores product 2 completely when choosing between options 0 and 1. Therefore, $\boldsymbol{z}_{i2}$ does not play any role in her choice between options 0 and 1.

After consumer i forms her consideration set and realizes the value of $\boldsymbol{z}_{ij}$ for all $j \in C_i$, she chooses the most preferred item from the consideration set. To order her preferences, consumer i derives utility

$$U(\boldsymbol{z}_{ij}, \boldsymbol{\varepsilon}_{ij}) = \boldsymbol{z}_{ij} + \boldsymbol{\varepsilon}_{ij},\ j = 0, 1, 2,$$

where $(\boldsymbol{x}_{ij}, \boldsymbol{z}_{ij})$ is independent of $(\boldsymbol{\eta}_{ij}, \boldsymbol{\varepsilon}_{ij})$. Note that variable $\boldsymbol{x}_{ij}$ does not enter the utility function. I exploit this exclusion restriction for the second identification strategy.

Now I show the first identification argument. For the first identification argument, I assume that $\boldsymbol{z}_{ij}$ has unbounded support. Since variable $\boldsymbol{x}_{ij}$ does not play any role in this argument, I drop the variable for notation simplicity. Consider product 1 without loss of generality. Let $p(C)$ be the probability that the chosen consideration set is C and $\boldsymbol{d}_{i1}$ be the choice variable, the latter taking the value of one if consumer i chooses product 1. Meanwhile, let $p(1|z_{i1}, z_{i2}, C) = \Pr(\boldsymbol{d}_{i1} = 1|\boldsymbol{z}_{i1} = z_{i1}, \boldsymbol{z}_{i2} = z_{i2}, \boldsymbol{C}_i = C)$ and analogously for other products. Where possible, without creating confusion, I use shorthand notation 012 for $\{0, 1, 2\}$ and likewise for other sets. The choice probability of product 1 is

$$p(1|z_{i1}, z_{i2}) = p(1|z_{i1}, z_{i2}, 012)p(012) + p(1|z_{i1}, 01)p(01).$$

From the displayed equation, the first term on the right-hand side is the probability that consumer i chooses product 1 and considers all products. The second term is the probability that consumer i chooses product 1 and considers product 1 and the outside option. Note that probability $p(1|z_{i1}, 01)$ does not depend on z_{i2} because product 2 is ignored when the consideration set is 01. Note, too, that $p(1|z_{i1}, \infty, 012) = 0$ since the utility from option 2 will always be larger than that from option 1. Therefore,

$$p(1|z_{i1}, z_{i2}, 012)p(012) = p(1|z_{i1}, z_{i2}) - p(1|z_{i1}, \infty). \tag{1.1}$$

The right-hand side of (1.1) depends only on observables; hence, the left-hand side is identified. I repeat the same argument for options 0 and 2 and then sum them so that

$$p(012) = \sum_{j=0}^{2} p(j|z_{i1}, z_{i2}, 012)p(012)$$

is identified, as are $p(j|z_{i1}, z_{i2}, 012)$'s. I can obtain

$$p(1|z_{i1}, 01)p(01) = p(1|z_{i1}, z_{i2}) - p(1|z_{i1}, z_{i2}, 012)p(012)$$

and for other products, too. Following the same argument as before, I can identify $p(j|z_{ij}, 0j)$'s and $p(0j)$'s for all j.

Now I demonstrate the second identification argument. For the second identification argument, I assume that $\boldsymbol{x}_{ij}$, $\boldsymbol{z}_{ij}$, $\boldsymbol{\eta}_{ij}$, and $\boldsymbol{\varepsilon}_{ij}$ are continuous. Additionally, $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ can have bounded supports. Again, I consider product 1 without loss of generality. Since now $\boldsymbol{x}_{ij}$ is not dropped, let $p(C|x_{i1}, x_{i2}) = \Pr(\boldsymbol{C}_i = C|\boldsymbol{x}_{i1} = x_{i1}, \boldsymbol{x}_{i2} = x_{i2})$. Then the choice probability of product 1 is

$$\begin{aligned} p(1|x_{i1}, x_{i2}, z_{i1}, z_{i2}) = {} & p(1|z_{i1}, z_{i2}, x_{i1}, x_{i2}, 012)p(012|x_{i1}, x_{i2}) \\ & + p(1|z_{i1}, x_{i1}, x_{i2}, 01)p(01|x_{i1}, x_{i2}). \end{aligned} \tag{1.2}$$

I first focus on the first term of Equation (1.2) on the right-hand side. The first term captures the event in which options 1 and 2 are considered and option 1 is the most preferred. Based on the simple model described above, this event can be expressed as follows:

$$\{\underbrace{\boldsymbol{\varepsilon}_{i0} - \boldsymbol{\varepsilon}_{i1} < \boldsymbol{z}_{i1},\ \boldsymbol{\varepsilon}_{i2} - \boldsymbol{\varepsilon}_{i1} < \boldsymbol{z}_{i1} - \boldsymbol{z}_{i2}}_{\text{option 1 gives the highest utility}},\ \underbrace{\boldsymbol{\eta}_{i1} < \boldsymbol{x}_{i1},\ \boldsymbol{\eta}_{i2} < \boldsymbol{x}_{i2}}_{\substack{\text{options 1 and 2} \\ \text{are considered}}}\}.$$

Let $F(\cdot)$ be the joint CDF of $(\boldsymbol{\varepsilon}_{i0} - \boldsymbol{\varepsilon}_{i1}, \boldsymbol{\varepsilon}_{i2} - \boldsymbol{\varepsilon}_{i1}, \boldsymbol{\eta}_{i1}, \boldsymbol{\eta}_{i2})$ conditional on $\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \boldsymbol{z}_{i1}, \boldsymbol{z}_{i2}$, and $\partial_\ell F(\cdot)$ be the partial derivative of $F(\cdot)$ with respect to the ℓth argument. Then,

$$p(1|z_{i1}, z_{i2}, x_{i1}, x_{i2}, 012)p(012|x_{i1}, x_{i2}) = F(z_{i1}, z_{i1} - z_{i2}, x_{i1}, x_{i2}). \tag{1.3}$$

Note from Equation (1.3) that

$$\frac{\partial}{\partial z_{i1}} F(z_{i1}, z_{i1} - z_{i2}, x_{i1}, x_{i2}) = \partial_1 F(\cdot) + \partial_2 F(\cdot) \text{ and } \frac{\partial}{\partial z_{i2}} F(z_{i1}, z_{i1} - z_{i2}, x_{i1}, x_{i2}) = -\partial_2 F(\cdot).$$

Next I consider the second term of Equation (1.2) on the right-hand side. Note that this term does not depend on z_{i2}. Again, this is because consumer i ignores product 2 when it is not considered. Therefore,

$$\frac{\partial}{\partial z_{i2}} p(1|z_{i1}, 01)p(01|x_{i1}, x_{i2}) = 0.$$

From the two results, I can obtain

$$\left(\frac{\partial^2}{\partial z_{i1}\partial z_{i2}} + \frac{\partial^2}{\partial z_{i2}^2}\right) p(1|x_{i1}, x_{i2}, z_{i1}, z_{i2}) = -\partial_{12} F(z_{i1}, z_{i1} - z_{i2}, x_{i1}, x_{i2}).$$

Since the left-hand side only depends on data, the right-hand side is identified. Although variables $\boldsymbol{x}_{ij}$'s seem irrelevant to this identification argument, they play a crucial role in the general case.

1.3 Model

This section formally defines the consideration set model. In this model, the consumer makes a choice in two stages: the consideration stage and the decision stage. In the first stage, the consumer decides which products to consider purchasing. The set of considered products is called the consideration set. In the second stage, she chooses the most preferred product from her consideration set.

I consider consumer i who chooses a product from $J+1$ products, where J is the number of products other than the outside option. The set of products is denoted by $\mathcal{J} = \{0, 1, \cdots, J\}$, where 0 denotes the outside option. Product j is described by two bundles of characteristics, $x_{ij} \in \mathcal{X}_j \subset \mathbb{R}^{L_1}$ and $z_{ij} \in \mathcal{Z}_j \subset \mathbb{R}^{L_2}$, where $\mathcal{X}_j$ is support of random vector $\boldsymbol{x}_{ij}$ and $\mathcal{Z}_j$ is support of random vector $\boldsymbol{z}_{ij}$. I use a lowercase letter, x_{ij}, to denote the realization of random vector $\boldsymbol{x}_{ij}$ and likewise for other variables. Furthermore, I define following notation: $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iJ})$, $\mathcal{X} = \prod_{j=1}^{J} \mathcal{X}_j$, and likewise for other variables, vectors, and spaces.

Product characteristics $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ play a role in the consideration stage and the decision stage, respectively. In the consideration stage, only random vector $\boldsymbol{x}_{ij}$ shapes consumer i's attention. In the decision stage, only $\boldsymbol{z}_{ij}$ impacts consumer i's preferences. I allow overlaps between $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$. However, for identification purpose, I do not allow the two random vectors to be identical. A detailed discussion is provided in the next section.

In the consideration stage, consumer i observes the value of $\boldsymbol{x}_i$ and her interest shock for product j, denoted by $\boldsymbol{\eta}_{ij} \in \mathbb{R}$, for all j. Given the realized product characteristics and interests, consumer i derives her attention level to product j as

$$A(\boldsymbol{x}_{ij}, \boldsymbol{\eta}_{ij}) = \boldsymbol{x}_{ij}^{\intercal}\gamma - \boldsymbol{\eta}_{ij}, \tag{1.4}$$

where γ is a conformable vector of parameters. I denote her attention level to product j by $a_{ij} = A(x_{ij}, \eta_{ij})$. If a_{ij} is positive, then she considers buying product j. That is, her consideration set C_i is defined by

$$C_i = \{j \in J : a_{ij} > 0\}.$$

Assumption 1.1: $\boldsymbol{\eta}_{i0} = -\infty$ with probability 1.

Assumption 1.1 says that the outside option is always considered. This assumption is common in the consideration set model literature (Abaluck & Adams, 2017; Goeree, 2008; Kawaguchi et al., 2016). It allows the consumer to walk away without buying a product in the decision stage.

Once consumer i forms her consideration set, she carefully studies the products in that set. For $j \in C_i$, she observes the bundle of product characteristics $\boldsymbol{z}_{ij}$ and her valuation of that product, denoted by $\boldsymbol{\varepsilon}_{ij} \in \mathbb{R}$. Consumer i derives the utility of each product in the consideration set to rank her preferences for the products:

$$U(\boldsymbol{z}_{ij}, \boldsymbol{\varepsilon}_{ij}) = \boldsymbol{z}_{ij}^{\intercal}\beta + \boldsymbol{\varepsilon}_{ij}, \tag{1.5}$$

where β is a conformable vector of parameters. For identification, I normalize the utility from the outside option to be $\boldsymbol{\varepsilon}_{i0}$. Let $\boldsymbol{e}_i = (\boldsymbol{\varepsilon}_i^\intercal, \boldsymbol{\eta}_i^\intercal)^\intercal$.

Assumption 1.2 (Exogeneity): $\boldsymbol{e}_i \perp\!\!\!\perp (\boldsymbol{x}_i, \boldsymbol{z}_i)$.

In the decision stage, consumer i chooses the product that maximizes her utility from her consideration set C_i. Ignoring ties, she chooses product $j \in C_i \subset \mathcal{J}$ if

$$\forall k \in C_i \backslash \{j\},\ u_{ij} > u_{ik},$$

where $u_{ij} = U(z_{ij}, \varepsilon_{ij})$ and $u_{i0} = \varepsilon_{i0}$. Choice variable $\boldsymbol{d}_{ij}$ takes the value of one if her consideration set contains product j and she most prefers product j out of all the products in the consideration set. That is,

$$d_{ij} = \sum_{C \in \mathcal{C}_j} \underbrace{\mathbb{1}(\forall k \in C \backslash \{j\},\ u_{ij} > u_{ik})}_{\substack{\text{take one if } j \text{ gives the highest} \\ \text{utility from set } C}} \underbrace{\mathbb{1}(\forall k \in C,\ a_{ik} > 0)\mathbb{1}(\forall \ell \notin C,\ a_{i\ell} < 0)}_{\text{take one if } C_i = C}, \tag{1.6}$$

where $\mathcal{C}_j$ is a family of consideration sets that contain product j.

Unlike consumer i, the econometrician is not able to observe $\boldsymbol{\eta}_i$ and $\boldsymbol{\varepsilon}_i$. However, suppose the econometrician observes the two bundles of product characteristics, $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$, and consumer choice variable $\boldsymbol{d}_i$. Let

$$p(j|z_i, x_i, C) = \Pr(\boldsymbol{d}_{ij} = 1 | \boldsymbol{z}_i = z_i, \boldsymbol{x}_i = x_i, \boldsymbol{C}_i = C) \text{ and } p(C|x_i) = \Pr(\boldsymbol{C}_i = C | \boldsymbol{x}_i = x_i).$$

From the econometrician's point of view, the choice probability of product j by consumer i conditional on x_i and z_i can be derived by Equations (1.4)–(1.6) and Assumption 1.2 as follows:

$$p(j|x_i, z_i) = \sum_{C \in \mathcal{C}_j} p(j|z_i, x_i, C)p(C|x_i). \tag{1.7}$$

1.4 Identification

In this section, I present two identification arguments when the econometrician only observes $(\boldsymbol{d}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)$ for each consumer i. I first define the notation and parameters of interest; then I discuss identification. All proofs are in Appendix A.

Since consumer choice is determined by the differences between two utilities, I focus on differences between two $\boldsymbol{\varepsilon}_{ij}$'s. In order to express these differences, I define $\boldsymbol{e}_{ij} = (\boldsymbol{\varepsilon}_{i0} - \boldsymbol{\varepsilon}_{ij}, \cdots, \boldsymbol{\varepsilon}_{ij-1} - \boldsymbol{\varepsilon}_{ij}, \boldsymbol{\varepsilon}_{ij+1} - \boldsymbol{\varepsilon}_{ij}, \cdots, \boldsymbol{\varepsilon}_{iJ} - \boldsymbol{\varepsilon}_{ij}, \boldsymbol{\eta}_{i1}, \cdots, \boldsymbol{\eta}_{iJ})^\intercal$. Let $F_{\boldsymbol{e}_{ij}}(\cdot)$ be the joint CDF of $\boldsymbol{e}_{ij}$, and let $f_{\boldsymbol{e}_{ij}}(\cdot)$ be its associated PDF. Note that by Assumption 1.2, I have $F_{\boldsymbol{e}_{ij}}(\cdot) = F_{\boldsymbol{e}_{ij}|\boldsymbol{x}_i, \boldsymbol{z}_i}(\cdot)$. Since the purpose of this chapter is to allow correlation between the consideration set stage and the decision stage, I identify joint distribution $F_{\boldsymbol{e}_{ij}}(\cdot)$ as well as the vectors of parameters β and γ.

1.4.1 Identification with Large Support

In this subsection, I discuss my first identification argument. This identification result is based on an identification at infinity idea and an exclusion restriction. Before presenting the theorem, I first discuss the assumptions needed for identification.

Assumption 1.3 (Exclusion Restriction): *For all $j \in \mathcal{J}$, $\boldsymbol{z}_{ij}$ includes at least one variable that is not contained in $\boldsymbol{x}_{ij}$. Let $\boldsymbol{z}_{ij} = (\boldsymbol{w}_{ij}, \boldsymbol{x}_{ij}^\intercal)^\intercal$.*

Assumption 1.3 indicates that there is at least one variable $\boldsymbol{w}_{ij}$ that enters utility function (1.5) but not attention function (1.4). This assumption can be justified by the model construction. Recall that after consumer i forms her consideration set C_i, she studies the products in her consideration set and learns characteristics $\boldsymbol{z}_{ij}$. While carefully studying the products, she may learn new information about the products; namely, $\boldsymbol{w}_{ij}$. The experiment by Wright and Barbour (1977) shows such consumer behavior.

Assumption 1.4 (Large Support): *Let* $\boldsymbol{x}_{ij} = (\boldsymbol{x}_{ij,1}, \boldsymbol{x}_{ij,-1}^\intercal)^\intercal$. *For any* $x_{i,-1} \in supp(\boldsymbol{x}_{i,-1})$, $supp((\boldsymbol{w}_i^\intercal, \boldsymbol{x}_{i,1}^\intercal)^\intercal | \boldsymbol{x}_{i,-1} = x_{i,-1}) = \mathbb{R}^{2J}$.

Assumption 1.5 (Normalization): *Intercepts in utility function (1.5) and attention function (1.4) are 0. Let* β_1 *be a coefficient for* $\boldsymbol{w}_{ij}$ *and* γ_1 *for* $\boldsymbol{x}_{ij,1}$. $\beta_1 = \gamma_1 = 1$.

Assumption 1.6 (No Multicollinearity): *For any* j, $\mathcal{Z}_j$ *is not contained in a proper linear subspace of* $\mathbb{R}^{L_2}$

Assumption 1.7: $supp(\boldsymbol{e}_{ij}) = \mathbb{R}^{2J}$.

Assumptions 1.4 and 1.5 together imply that $\boldsymbol{w}_{ij}$ and $\boldsymbol{x}_{ij,1}$ are special regressors in the utility and the attention functions, respectively (Lewbel, 2000). I can use these special regressors to identify $F_{\boldsymbol{e}_{ij}}(\cdot)$ and its derivatives everywhere on $\mathbb{R}^{2J}$ (Ichimura & Lee, 1991; Lewbel, 2016).

Assumptions 1.4–1.7 are standard in the class of single-index models (Manski, 1988; Ichimura, 1993; Horowitz, 2012). Ichimura and Lee (1991) show that similar assumptions are needed for multi-index models.

Assumption 1.8: *Let* $\boldsymbol{x}_{ij,-1} = (\boldsymbol{x}_{ij2}, \cdots, \boldsymbol{x}_{ijk}, \cdots, \boldsymbol{x}_{ijL_1})^\intercal$. *For any* j *and* k, *and for any* $i \neq i'$, $\Pr(\boldsymbol{x}_{ijk} \neq \boldsymbol{x}_{i'jk}) > 0$.

Assumption 1.8 states that all product characteristics vary across consumers. This assumption is not needed in the discrete choice model with parametric assumptions (e.g., Berry, Levinsohn, & Pakes, 1995; Goeree, 2008), but this assumption is implicitly imposed in semiparametric models (e.g., Fox, 2007; Lewbel, 2000).

Theorem 1.1: *Suppose Assumptions 1.1–1.8 hold; then for all* j*, the joint CDF* $F_{\boldsymbol{e}_{ij}}(\cdot)$*,* β*, and* γ *are identified.*

1.4.2 Identification with Compact Support

In this subsection, I present the identification without the identification at infinity argument. Instead of drawing on the unbounded support assumption, I assume that $\boldsymbol{x}_{ij}$ includes at least one continuous variable that is not in $\boldsymbol{z}_{ij}$ and vice versa. Under the two exclusion restrictions, I show that parameters β, γ, and PDF $f_{\boldsymbol{e}_{ij}}(\cdot)$ are identified on $\mathcal{XZ}$, the joint support of $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$.

Assumption 1.9 (Exclusion Restriction): *For all* $j \in \mathcal{J}$*,* $\boldsymbol{x}_{ij}$ *includes at least one continuous variable that is not contained in* $\boldsymbol{z}_{ij}$ *and vice versa. Let* $\boldsymbol{x}_{ij} = (\boldsymbol{w}_{ij}^1, \boldsymbol{x}_{ij,-1}^\intercal)^\intercal$ *and* $\boldsymbol{z}_{ij} = (\boldsymbol{w}_{ij}^2, \boldsymbol{x}_{ij,-1}^\intercal)^\intercal$.

Under Assumption 1.9, there is at least one continuous variable that affects utility but not attention and vice versa. Although this assumption is stronger than Assumption 1.3 (the exclusion restriction used in the previous section), it allows me to drop Assumption 1.4 (unbounded support).

Many empirical papers show that there are variables that only influence attention. Ackerberg (2003) shows that the primary role of advertising is to inform consumers. Honka et al. (2017) show that advertising helps consumers become more aware of products and that advertising has a negligible impact on consumer preferences. Online reviews or sponsored reviews work similarly to advertisements in affecting consumer choice (Bailey, 2005;

Lu, Chang, & Chang, 2014; Duan, Gu, & Whinston, 2008). These results justify Assumption 1.9.

Assumption 1.10: *For any j, $\boldsymbol{z}_{ij}$ and $\boldsymbol{x}_{ij}$ are continuous random vectors.*

Assumption 1.10 can be relaxed. I can allow some or all of $\boldsymbol{x}_{ij,-1}$ to be discrete, whereas both $\boldsymbol{w}_{ij}^1$ and $\boldsymbol{w}_{ij}^2$ have to remain continuous. In addition, I need to restrict the parameter space to guarantee identification (Horowitz, 2012; Ichimura & Lee, 1991).

Assumption 1.11: *For all $j \in \mathcal{J}$, $f_{\boldsymbol{e}_{ij}}(\cdot)$ is differentiable with respect to the first through Jth arguments.*

Assumption 1.12: *For all $j \in \mathcal{J}$, $f_{\boldsymbol{e}_{ij}}(\cdot)$ is neither constant nor periodic for any convex 2J-cube with positive Lebesgue measure in $\mathbb{R}^{2J}$.*

Assumptions 1.11 and 1.12 are needed for identification in single-index and multi-index models (Ichimura & Lee, 1991; Ichimura, 1993; Horowitz, 2012).

Theorem 1.2: *Suppose Assumptions 1.1–1.2 and 1.5–1.12 hold; then β and γ are identified, and for all $j \in \mathcal{J}$, PDF $f_{\boldsymbol{e}_{ij}}(\cdot)$ is identified on $\mathcal{XZ}$.*

1.5 Conclusion

In this chapter, I discussed identification of the consideration set model, allowing attention and preferences to be arbitrarily correlated. Specifically, I presented the two identification results. The first identification argument relies on an exclusion restriction and an identification at infinity concept. The exclusion restriction means that there is a variable that changes preferences but not attention. I assume that this variable has unbounded support.

The second identification argument relaxes the unbounded support assumption, instead requiring a stronger exclusion restriction. I need an additional variable that modifies attention but not preferences. I showed that under these two exclusion restrictions, I can identify the parameters of interest.

Chapter 2

Testing an Exclusion Restriction in a Consideration Set Model

2.1 Introduction

The standard discrete choice model assumes that the choice set is exogenously given to a consumer. The consumer pays attention to all products and chooses the product that maximizes her utility from the given choice set. However, in practice, the consumer may not consider all products due to cognitive limitations and may even endogenously choose a subset of products worth buying. The assumption imposed on the standard discrete choice model can lead to inaccurate analysis or prevent the econometrician from addressing interesting questions. For example, Goeree (2008) shows that the standard discrete choice model underestimates price elasticity. Meanwhile, Honka et al. (2017) show that advertising, rather than preferences, influences consumer attention. These results suggest that extending the standard discrete choice model is necessary.

The consideration set model extends the discrete choice model by relaxing the assumption that the consumer considers all products. Instead, the model assumes that the consumer endogenously selects which products to consider buying. The set of considered products is called the consideration set. The consumer's attention to products often drives her decision about the consideration set (Goeree, 2008). Endogenizing the consideration set creates identification problems since the researcher often does not observe the set. Suppose that the consumer is less likely to buy a product when the price rises. The econometrician cannot discern whether the higher price makes the product less favorable or less likely to enter the consideration set. To identify model parameters, the econometrician needs two variables, one of which controls preferences and the other of which affects attention. Each variable works as an exclusion restriction since each is excluded from either attention or preferences. If the researcher has only one of these variables, they can obtain partial identification (Henry et al., 2014). If they have both, they can obtain point identification (Compiani & Kitamura, 2016). Therefore, identification of the consideration set model depends on the available exclusion restrictions.

This chapter proposes a nonparametric test of the validity of an exclusion restriction in the consideration set model. Suppose the econometrician has a variable that only influences the consumer's attention and a candidate variable that may impact either the consumer's attention and preferences or only her preferences. The null hypothesis for my test is that the candidate variable does not affect attention; it only influences preferences. The test relies on three main assumptions. First, I assume that preferences and attention are independent of one another. Second, the characteristics of products outside the consideration set do not shape the consumer's preferences for products in the consideration set. For instance, if the consumer does not consider product j, the characteristics of product j do not modify her preferences for any of the other products in her consideration set. Third, the characteristics of product j do not influence the consumer's attention to the other products.

Based on these assumptions, I devise a test in four steps. First, I derive a function that depends solely on the consumer's attention under the null hypothesis. This derivation is inspired by Henry et al. (2014), and this function is called the test function. I show that if the test function does not vary with the candidate variable, then the candidate variable has no effect on attention. As a result, the candidate variable influences only preferences. Second, my test strategy utilizes this knowledge. I compute the difference between the test function and the integral of the test function with respect to some weighting function on the space of the candidate variable. This difference must be zero under the null hypothesis. Third, my test is based on measuring the significance of the square of the difference. Fourth, I use the kernel method to derive my test statistic, and I show that the test is consistent.

This chapter focuses on a case in which the econometrician wants to test whether the candidate variable changes only preferences when there is a variable that only shapes attention. It is certainly possible to test the opposite case – that is, whether the candidate variable alters only attention when a variable that only impacts preferences is available. I illustrate this opposite case in Appendix C.

The rest of the chapter is structured as follows. Section 2.2 details the identification of the consideration set model and possible scenarios in which my test is useful. Section 2.3 presents a simple model to convey the key idea of my test. Section 2.4 presents the test for a generic model. Section 2.5 reports a simulation study to show the finite sample performance of my test. Section 2.6 concludes the chapter.

2.2 Identification and Potential Applications of the Test

My test can be used to validate a candidate exclusion restriction and evaluate whether a variable of interest governs the consumer's attention, preferences, or both. Since my test is concerned with the identification of the consideration set model, I first review identification and then discuss the possible applications of the test.

2.2.1 Review of Identification

There is a vast body of literature studying the identification of the consideration set model when the econometrician observes only the consumer's final choice. When no data are available for the consideration set, the consideration set model falls under finite mixture model frameworks, assuming that attention and preferences are independent of one another. Therefore, some identification that results from finite mixture models can be applied to the consideration set model. As in finite mixture model frameworks, in the consideration set model, the econometrician needs exclusion restrictions to identify the model parameters. In this chapter the exclusion restrictions mean that there are variables that only impact either attention or preferences, but not both. I refer to the former restriction – variables that shape attention but not preferences – as a *Type 1 exclusion restriction.* These variables are called *Type 1 variables.* Similarly, I refer to the latter exclusion restriction – variables affecting preferences but not attention – as a *Type 2 exclusion restriction*, and I name such variables *Type 2 variables.*

The first identification results depend on whether there are type 1 or type 2 exclusion restrictions. Nchare (2018) shows that partial identification can be obtained with a type 1 exclusion restriction. To the best of my knowledge, there are no identification results achieved solely with a type 2 exclusion restriction unless further assumptions are imposed.[1] To obtain point identification, the econometrician must assume that type 1 or type 2 variables have infinite supports. Kawaguchi, Uetake, and Watanabe (2018) show that parameters of interest can be identified with a type 1 exclusion restriction and an infinite support assumption. They do so by focusing on the consumer who pays attention to every product with probability 1. The identification problem in their study thus becomes identical to the identification problem of the standard discrete choice model. Barseghyan, Molinari, and Thirkettle (2019b) show that identification can be obtained with a type 2 exclusion restriction and an infinite support assumption. They use the infinite support assumption to focus on the consumer who prefers one product over another with probability 1. In this case, the model is concerned only with attention.

The identification at infinity concept is undesirable because the econometrician rarely observes the tail of variables. Compiani and Kitamura (2016) show that the econometri-

[1]Kitamura and Laage (2018) show identification with a type 2 exclusion restriction in a finite mixture framework, but the identification result does not directly apply to the consideration set model.

cian can obtain point identification without the infinite support assumption when both type 1 and 2 exclusion restrictions are imposed. While this is the result of a finite mixture framework, it also applies to the consideration set model. These two exclusion restrictions are often adopted in the applied literature (Goeree, 2008; Heiss, McFadden, Winter, Wuppermann, & Zhou, 2016; Hortaçsu, Madanizadeh, & Puller, 2017).

For completeness, I present different identification results from the consideration set model. These results do not require exclusion restrictions but they do require other assumptions to gain identification power. Abaluck and Adams (2017) and Kawaguchi et al. (2018) assume that each consumer faces a different choice set when entering the market. The econometrician observes each consumer's choice set. Variations in each consumer's choice set allow the econometrician to identify parameters of interest without exclusion restrictions. Barseghyan, Coughlin, Molinari, and Teitelbaum (2019) restrict the size of the consideration set. They assume that the consumer considers at least k products for $k > 1$. This assumption makes partial identification attainable without imposing exclusion restrictions.

2.2.2 Testing the Validity of an Exclusion Restriction

As explained in the previous subsection, for the econometrician to obtain point identification, they must impose both type 1 and type 2 exclusion restrictions. This is because the identification at infinity concept is undesirable, and the composition of each consumer's choice set is often unknown. I do not consider identification strategies that rely on these two restrictions in this chapter. When imposing such restrictions, the econometrician often rests on educated guesses or refers to past literature; unless they run experiments that artificially create the required exclusion restrictions. The econometrician may want to validate the exclusion restrictions obtained from the former two sources since his guesses may be wrong or the results from the past literature may not apply to the context of his research. In this case, the econometrician can use my test to validate a type 2 exclusion restriction if a type 1 exclusion restriction is correctly imposed and vice versa.

2.2.3 Testing the role of variables

My test can be used to study the role of variables of interest in consumer behavior. For a consumer to choose a product, she must pay attention to it and prefer it over other products to which she also pays attention. The econometrician is often interested in whether a variable of interest modifies attention, preferences, or both. This is because the role of variables is closely related to policy implications (Draganska & Klapper, 2011).

One advantage of my test is that the econometrician does not have to make many identification assumptions. Instead of imposing both type 1 and 2 exclusion restrictions, they can impose either type 1 or 2 exclusion restrictions to study the role of variables of interest. For example, suppose the econometrician wants to test whether the consumer is aware of the product quality (which is known to the econometrician) before forming her consideration set. In that case, my test can be used to determine whether the quality variable shapes the consumer's attention. Of course, performing this test requires other variables that control attention but not preferences.

I now review the literature to illustrate cases in which my test can be used. The consideration set model is often used to study the role of advertising. Advertising is considered to play two roles in consumer behavior – influencing attention and influencing preferences (Brierley, 2005). Ackerberg (2003), Honka et al. (2017), and Clark, Doraszelski, and Draganska (2009) show that the main effect of advertising is to inform the consumer about a product. Draganska and Klapper (2011) specifically study the role of advertising using the consideration set model. Their study finds that advertising plays both roles in consumer behavior.

Similar to advertising, reviews can shift the consumer's attention to a product (Vermeulen & Seegers, 2009) and convey information about product quality (Newberry & Zhou, 2016; Lewis & Zervas, 2016). Vermeulen and Seegers (2009) study the effect of hotel reviews on the consumer's attention and preferences. They conclude that positive reviews enhance the consumer's attention to lesser-known hotels. Newberry and Zhou (2016) study the impact of reviews posted on an E-commerce platform on local and national retailers. They show that online reviews have a more significant impact on sales for local sellers than for national sellers. This is because positive online reviews help local retailers differentiate themselves from lower-quality retailers.

In addition to advertising and reviews, Berger, Draganska, and Simonson (2007) study the impact of product diversity on the consumer's attention and preferences. Kawaguchi et al. (2018) study how the consumer's attention or preferences change under time pressure. Ursu (2018) studies the impact of product rankings on consumer behavior. She shows that the ranking of a product shifts the consumer's attention but not her preferences.

2.2.4 Limitation

When my test is used to verify a candidate exclusion restriction, my test can be considered a model selection procedure. For example, the econometrician selects a model between a model with a type 1 exclusion restriction and a model with both type 1 and type 2 exclusion restrictions. Like model selection procedures, my test runs into the uniform consistency issue (Leeb & Pötscher, 2005). Therefore, my test causes post-selection inference problems (Leeb and Pötscher, 2005; 2006; 2008).

This problem can also occur when using my test to study the role of a variable of interest. If the econometrician wants to make additional inferences based on the test result, they will face the same post-selection inference problem. In this case, they can work around the issue by data splitting or carving (Fithian, Sun, & Taylor, 2014), but this procedure will result in a loss of efficiency in testing and inference. Otherwise, the econometrician can derive the exact distribution of parameters of interest conditional on the test result (e.g., Lee, Sun, Sun, & Taylor, 2013). However, the econometrician may not be able to derive the exact distribution.

While it is important to address this type of problem, it is beyond the scope of this chapter to achieve uniform consistency or to derive the exact distribution based on a test result.

2.3 The Simple Model and Test

This section describes the consideration set model with two products and an outside option. I show how to test whether the candidate variable affects only preferences when a type 1 exclusion restriction is imposed. The opposite case is certainly possible; since it is repetitive, however, I present it in Appendix C. I use bold letter $\boldsymbol{x}$ for a random variable, lowercase letter x for its realization, and curly letter $\mathcal{X}$ for its support. I use x_j to represent the jth component of vector x and x_{-j} to represent vector x without the jth component. Although this section focuses on a model with two products, I state general assumptions and a hypothesis to avoid repetition when studying a case with more than two products. This section focuses on a particular consumer; namely, consumer i. For the sake of clarity, I drop from the notation the subscript i representing consumer i. All proofs are in Appendix B.

The consumer chooses a product from three possible options $\mathcal{J} = \{0, 1, \cdots, J\}$, where J represents the number of products other than the option 0, and option 0 represents the outside option. In this chapter, I consider $J = 2$. Product 1 is described by its characteristics, $\boldsymbol{x}_1$. Let $\boldsymbol{x}_1 = (\boldsymbol{w}_1, \boldsymbol{z}_1)$ and $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$. For simplicity, I assume that $\boldsymbol{w}_1$ and $\boldsymbol{z}_1$ are random variables. It is straightforward to extend them to random vectors, and if this is done, all results in this chapter still hold. I define $\boldsymbol{z}$ and $\boldsymbol{w}$ similarly to $\boldsymbol{x}$. In this section, I want to test whether $\boldsymbol{z}$ influences attention provided that $\boldsymbol{w}$ influences attention but not preferences. If $\boldsymbol{z}$ does not impact attention, then I can conclude that $\boldsymbol{z}$ only changes preferences. As a result, point identification is attainable. Before explaining consumer behavior, I make two assumptions:

Assumption 2.1: $\boldsymbol{x}$ is independent of the consumer's idiosyncratic shocks.

Assumption 2.2: Preferences and attention are independent of one another conditional on $\boldsymbol{x}$.

Assumption 2.1 states that observable characteristics are exogenous. Assumption 2.2 is made in most literature that studies the consideration set model (e.g., Abaluck & Adams, 2017; Van Nierop, Bronnenberg, Paap, Wedel, & Franses, 2010; Kawaguchi et al., 2018; Barseghyan, Molinari, & Thirkettle, 2019b).

To choose a product, the consumer has to pay attention to it and prefer it over the other products to which she also pays attention. I assume that the consumer always pays attention to the outside option but may or may not pay attention to options 1 and 2. Her attention to options 1 and 2 depends on their product characteristics, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively.

Assumption 2.3: For all $j, j' \in \mathcal{J}$ with $j \neq j'$, the consumer's attention to product j is independent of $\boldsymbol{x}_{j'}$ given $\boldsymbol{x}_j$.

Assumption 2.3 states that the consumer ignores the characteristics of product 2 when deciding whether to pay attention to product 1. If the consumer pays attention to product 1, she will consider buying it. The set of products she considers purchasing is called the consideration set and is denoted by C.

The attention-based consideration set formation is widely adopted in the literature and is useful for my testing.[2] Where possible, without creating confusion, I use the shorthand 012 for $\{0, 1, 2\}$ and likewise for the other sets. The space of consideration sets when three options are available is

$$\mathcal{C} = \{0, 01, 02, 012\}.$$

[2]See Goeree (2008), Van Nierop et al. (2010), Abaluck and Adams (2017), Nchare (2018), and Barseghyan, Molinari, and Thirkettle (2019b).

In the attention-based consideration set formation, the consideration set (or event) $01 \in \mathcal{C}$ indicates that the consumer pays attention to product 1 but not product 2. In other words, $\Pr(\boldsymbol{C} = 01|\boldsymbol{x})$ depends on both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ because $\boldsymbol{x}_1$ affects how much attention the consumer pays to product 1. Therefore, event 01 or 012 refers to an event in which the consumer is attentive to product 1 regardless of product 2. In other words, whether the consumer chooses consideration set 01 or 012 is independent of $\boldsymbol{x}_2$ by Assumption 2.3 – $\Pr(\boldsymbol{C} = 01 \text{ or } 012|\boldsymbol{x}) = \Pr(\boldsymbol{C} = 01 \text{ or } 012|\boldsymbol{x}_1)$. In addition, an event in $\mathcal{C}$ is mutually exclusive because the consumer can only have one consideration set. I utilize these two facts to test whether $\boldsymbol{z}$ alters the consumer's attention.

After she selects her consideration set C, she chooses the most preferred item from her consideration set. I assume that her preferences for certain products over others depend on $\boldsymbol{x}$.

Let $p(C|x) = \Pr(\boldsymbol{C} = C|\boldsymbol{x} = x)$ be the probability that the consumer's consideration set is C, conditional on product characteristics x. I call this probability *the consideration set probability.* Let $\boldsymbol{d}_1$ be the choice variable, which takes a value of one if the consumer chooses product 1. Meanwhile, let $p(1|x, C) = \Pr(\boldsymbol{d}_1 = 1|\boldsymbol{x} = x, \boldsymbol{C} = C)$ be the choice probability of product 1 given product characteristics x and consideration set C. This choice probability works analogously for the other products. I call this probability *the conditional choice probability.* The choice probability of product 1 is

$$p(1|x) = p(1|x, 012)p(012|x) + p(1|x, 01)p(01|x). \tag{2.1}$$

In this chapter, I am interested in testing whether $\boldsymbol{z}$ is excluded from the consideration set probability when a type 1 exclusion restriction is imposed. If it is excluded, then I can conclude that $\boldsymbol{z}$ indeed only impacts preferences. To construct the test, I assume that the samples from the joint distribution of $(\boldsymbol{d}, \boldsymbol{x})$ are available to the econometrician, where $\boldsymbol{d} = (\boldsymbol{d}_0, \boldsymbol{d}_1, \boldsymbol{d}_2)$. Note that the consumer's consideration set is unavailable to the econometrician.

The hypothesis that I want to test is

$$H_0: \ \Pr\left[p(\mathcal{J}|\boldsymbol{w},\boldsymbol{z}) = p(\mathcal{J}|\boldsymbol{w})\right] = 1.$$
$$H_A: \ \Pr\left[p(\mathcal{J}|\boldsymbol{w},\boldsymbol{z}) = p(\mathcal{J}|\boldsymbol{w})\right] < 1.$$

Notice that my hypothesis only depends on $C = \mathcal{J}$ (012 in this section). This is because I am interested in whether $\boldsymbol{z}$ influences the consumer's attention to products. Recall that $C \in \mathcal{C}$ describes an event in which the consumer is attentive to product $j \in C$ and inattentive to product $j' \notin C$. If the consumer's attention to product 1 depends on z_1, so does any $C \in \mathcal{C}$. Hence, it is sufficient to test $C = \mathcal{J}$ only.

I assume no data is available for the consumer's consideration set. Because of this limitation, $p(012|x)$ cannot be directly identified from the data. Instead, I suggest relying on exclusion restrictions to test the hypothesis.

Assumption 2.4: $\forall j \in \mathcal{J}$ and $\forall C \in \mathcal{C}$, $p(j|\boldsymbol{x}, C) = p(j|\boldsymbol{z}, C)$ with probability 1.

Assumption 2.4 describes a type 1 exclusion restriction – that is, $\boldsymbol{w}$ only affects attention. If $j \notin C$, then $p(j|\boldsymbol{x}, C) = p(j|\boldsymbol{z}, C) = 0$ with probability 1. This is because the consumer is unable to choose a product that she does not consider buying. I can derive a function that depends on $p(012|x)$ from Equation (2.1) using Assumption 2.4. This function, which relies only on $p(012|x)$, is called the test function. Therefore, the hypothesis can be tested using Assumption 2.4. The idea of constructing these test functions using Assumption 2.4 is inspired by Henry et al. (2014).

2.3.1 Testing the Hypothesis

Now I show how Assumptions 2.3 and 2.4 can be helpful when testing the hypothesis. Under Assumptions 2.3 and 2.4, Equation (2.1) can be written as

$$\begin{aligned} p(1|x) &= p(1|z,01)p(01 \text{ or } 012|x) + p(012|x)[p(1|z,012) - p(1|z,01)] \\ &= p(1|z,01)p(01 \text{ or } 012|x_1) + p(012|x)[p(1|z,012) - p(1|z,01)]. \end{aligned} \tag{2.2}$$

In Equation (2.2), the first equality holds true by Assumption 2.4 and because having consideration set 01 and having consideration set 012 are mutually exclusive events. The second equality comes from the fact that the event 01 or 012 represents that the consumer is attentive to product 1 regardless of her attention to product 2. Thus, $\boldsymbol{x}_2$ is independent of her attention to product 1 by Assumption 2.3. Note from Equation (2.2) that when I change w_2 to w_2', the only term that changes is $p(012|x)$. For any function q, I write $\Delta_{x_1'}^{x_1} q(\cdot, x_2)$ to denote $q(x_1, x_2) - q(x_1', x_2)$. Hence,

$$\Delta_{w_2'}^{w_2} p(1|w_1, \cdot, z) = \Delta_{w_2'}^{w_2} p(012|w_1, \cdot, z)[p(1|z,012) - p(1|z,01)]. \tag{2.3}$$

To remove $p(1|z,012) - p(1|z,01)$ from Equation (2.3), I divide two instances of Equation (2.3) at the different values of w_2 and w_2'. I use the resulting function to construct a test. To construct this valid test for the hypothesis, I need the left-hand side of Equation (2.3) to be bounded away from zero. To ensure this, I make the following assumption. Let $\partial_{x_1} q(x_1, \cdots, x_J) = \partial q(x_1, \cdots, x_J)/\partial x_1$ and $\partial_{x_J} \cdots \partial_{x_2} q(x_1, \cdots, x_J) = \partial^{J-1} q(x_1, \cdots, x_J)/(\partial x_2 \cdots \partial x_J)$ for differentiable function $q(\cdot)$.

Assumption 2.5: (i) $\forall j \in \mathcal{J}\ \exists c > 0\ \forall w_j \in \mathcal{W}_j\ \forall z \in \mathcal{Z}\ \forall w_{-j} \in \text{int}(\mathcal{W}_{-j})$, $\partial_{w_J} \cdots \partial_{w_{j+1}} \partial_{w_{j-1}} \cdots \partial_{w_1} p(\mathcal{J}|w_j, w_{-j}, z) \geq c > 0$; (ii) For a given $j \in \mathcal{J}\backslash\{0\}$ and for a given $\mathcal{L} \subset \mathcal{J}$ that contains products 0 and j, define a finite sequence $\{\mathcal{C}_{j,\ell}\}_{\ell=1}^{L(\mathcal{L})}$ where $\mathcal{C}_{j,1} = \{\{0, j\}\}$,

$\mathcal{C}_{j,L(\mathcal{L})} = \{\mathcal{L}\}$ and for $1 < \ell < L(\mathcal{L})$ $\mathcal{C}_{j,\ell} = \{C \cup \{k\} : C \in \mathcal{C}_{j,\ell-1}, k \in \mathcal{L} \backslash C\}$. Then $\exists c > 0$ $\forall j \in \mathcal{J}$ $\forall z \in \mathcal{Z}$ $\forall \mathcal{L} \subset \mathcal{J}$ that contains products 0 and j and for any $j' \notin \mathcal{L}$

$$\sum_{\ell=1}^{L(\mathcal{L})} \sum_{C \in \mathcal{C}_{j,\ell}} (-1)^{|C|} \left\{ p(j|z, C) - p(j|z, C \cup \{j'\}\right\} \geq c > 0.$$

Assumption 2.5–(i) is satisfied if the consumer's attention to product 1 strictly increases in $\boldsymbol{w}_1$ and by Assumption 2.3. For instance, $\boldsymbol{w}_1$ could be advertising product 1. Goeree (2008) and Honka et al. (2017) show that advertising boosts the consumer's attention to a product. Assumption 2.5–(ii) says that products in $\mathcal{J}$ are strictly substitutable goods. For instance, if $\mathcal{L} = 01$, then I have a family of a set, $\mathcal{C}_{1,1} = \{01\}$. Then for any z, for $2 \in \mathcal{J} \backslash 01$, I have $p(1|z, 01) - p(1|z, 012) > 0$. Assumption 2.5–(ii) also states that $p(1|z, C)$ decreases with the size of C, which contains product 1, at a decreasing rate. Moreover, the multinomial logit model satisfies Assumption 2.5–(ii). Under Assumption 2.5, I can ensure that Equation (2.3) is bounded away from zero for any choice of $w_2, w_2' \in \mathcal{W}_2$ with $w_2 \neq w_2'$.

Now for any choice of $w_2, w_2' \in \mathcal{W}_2$ with $w_2 \neq w_2'$, I can construct function $\psi(\cdot)$, which I call the test function:

$$\begin{aligned} \psi(1, w_1, w_2, w_2', z) &= \frac{\Delta_{w_2'}^{w_2} p(1|w_1, \cdot, z)}{\Delta_{w_*(w_1,z)}^{w^*(w_1,z)} p(1|w_1, \cdot, z)} \\ &= \frac{\Delta_{w_2'}^{w_2} p(012|w_1, \cdot, z)}{\Delta_{w_*(w_1,z)}^{w^*(w_1,z)} p(012|w_1, \cdot, z)}, \end{aligned}$$

where $w^*(w_1, z) = \arg\max_{w_2 \in \mathcal{W}_2} p(012|w_1, w_2, z)$ and $w_*(w_1, z) = \arg\min_{w_2 \in \mathcal{W}_2} p(012|w_1, w_2, z)$. In the denominator term of $\psi(\cdot)$, I assume that w_2 takes $w^*(w_1, z)$ and $w_*(w_1, z)$ so that the denominator term is bounded away from zero as much as possible at the given w_1 and z. These choices of $w^*(w_1, z)$ and $w_*(w_1, z)$ and Assumption 2.5–(i) ensure that $\psi(\cdot)$ does not explode to infinity. I suggest testing the hypothesis using $\psi(\cdot)$ by testing whether or not $\psi(\cdot)$ does not vary with z. Even if $\psi(\cdot)$ does not vary with z, $p(012|w,z)$ may still vary across z. In other words, my test, which relies on $\psi(\cdot)$, can yield the incorrect result, which is that z does not affect attention when $p(012|w,z)$ still depends on z. Therefore, I restrict the behavior of $p(012|w,z)$.

Assumption 2.6: $p(\mathcal{J}|w,z)$ is continuously differentiable in w and z.

Assumption 2.7: $\mathcal{X}$ is convex and compact. The joint support $\mathcal{X}$ is a Cartesian product.

Assumptions 2.6 and 2.7 (regarding the convexity of $\mathcal{X}$) are used to invoke the mean value theorem in the proof of Lemma 2.1. The compactness of $\mathcal{X}$ ensures that $w^*(\cdot) \neq \infty$ and $w_*(\cdot) \neq -\infty$. Furthermore, by Assumptions 2.5–(i) and 2.7 (Cartesian product), I have $w^*(w_1,z) = \bar{w}_2$ and $w_*(w_1,z) = \underline{w}_2$, where $\mathcal{W}_2 = [\underline{w}_2, \bar{w}_2]$. I assume the Cartesian product so that I can easily replace $w^*(w_1,z)$ and $w_*(w_1,z)$ with constants. The Cartesian product assumption can be relaxed with convex and compact $\tilde{\mathcal{W}}_{-j} = \cap_{w_j,z}\text{supp}(\boldsymbol{w}_{-j}|w_j,z)$. Then, by Assumption 2.5 and Equation (2.2), the econometrician can pin down constant $w^\star \in \arg\min_{w_2 \in \tilde{\mathcal{W}}_2} p(1|\cdot, w_2, \cdot) = \arg\max_{w_2 \in \tilde{\mathcal{W}}_2} p(012|\cdot, w_2, \cdot)$ and likewise for $w_\star$.

Assumption 2.8: Under H_A, $\forall c \in \mathbb{R}$ for any $j \in \mathcal{J}$, $\forall w'_{-j}, w''_{-j} \in \text{int}(\mathcal{W}_{-j})$, $\exists w_j \in \mathcal{W}_j$ and $z \in \mathcal{Z}$ such that $\partial_{w_J} \cdots \partial_{w_{j+1}} \partial_{w_{j-1}} \cdots \partial_{w_1} p(\mathcal{J}|w_j, w'_{-j}, z) \neq c \partial_{w_J} \cdots \partial_{w_{j+1}} \partial_{w_{j-1}} \cdots \partial_{w_1}$ $p(\mathcal{J}|w_j, w''_{-j}, z)$.

Assumption 2.8 excludes the case in which $p(012|w,z)$ is either additively or multiplicatively separable, i.e., the case in which $p(012|w,z)$ can be represented as $q_1(w) + q_2(z)$ or $q_1(w)q_2(z)$ for some functions q_1 and q_2. Assumption 2.8 is satisfied under another assumption that the applied literature often employs (e.g., Abaluck & Adams, 2017; Van Nierop et al., 2010; Kawaguchi et al., 2018; Barseghyan, Molinari, & Thirkettle, 2019b). That is, the literature assumes that the probability that the consumer considers product 1 follows either the logit or probit form. For instance, if the consumer consider product 1 with probability

$$\frac{\exp(x_1^\intercal \beta)}{1+\exp(x_1^\intercal \beta)},$$

then she has consideration set 012 with probability:

$$p(012|x) = \frac{\exp(x_1^\intercal \beta)}{1+\exp(x_1^\intercal \beta)} \frac{\exp(x_2^\intercal \beta)}{1+\exp(x_2^\intercal \beta)}.$$

Under the logit form, $p(012|w,z)$ cannot be additively or multicatively separable; thus, Assumption 2.8 holds true.

Lemma 2.1: Suppose Assumptions 2.1–2.8 are satisfied. Then $\psi(\cdot)$ does not vary with z if and only if $p(012|w,z)$ does not vary with z.

For the sake of notation, I now write $\omega_j = (w_j, w_{j'}, w'_{j'})$ and $\bar{\omega}_j = (w_j, \bar{w}_{j'}, \underline{w}_{j'})$, where $j' \neq j$. By Lemma 2.1, I can rewrite the hypothesis in the two-product case as

$$H_0: \ \forall j, \ \omega_j, z, z', \ \psi(j, \omega_j, z) = \psi(j, \omega_j, z'),$$

and

$$H_A: \ \exists j, \ \omega_j, z, z', \ \psi(j, \omega_j, z) \neq \psi(j, \omega_j, z').$$

Notice that under the null, I have

$$\forall j \ \forall z \ \forall \omega_j, \ D_a(\psi)(j, \omega_j, z) = \psi(j, \omega_j, z) - \int_{\mathcal{Z}} \psi(j, \omega_j, z') a(z') \, dz' = 0$$

where $a(z')$ is a specified, non-negative weighting function with $\int_{\mathcal{Z}} a(z') \, dz' = 1$. My testing approach is to assess the significance of the square of $D_a(\psi)$.

My test statistic is based on

$$\Gamma(f, F_{\boldsymbol{x}}) = \sum_{j=1,2} \int_{\mathcal{X}} \int_{\underline{w}_{j'}}^{w_{j'}} D_a(\psi(f_j))(j, \omega_j, z)^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x), \tag{2.4}$$

where

$$\psi(j, \omega_j, z; f_j) = \frac{\Delta_{w'_{j'}}^{w_{j'}} p(j|w_j, \cdot, z; f_j)}{\Delta_{\underline{w}_{j'}}^{\bar{w}_{j'}} p(j|w_j, \cdot, z; f_j)}.$$

I use notation $p(j|x; f_j) = f_{\boldsymbol{y}_j}(1, x)/f_{\boldsymbol{x}}(x)$ to explicitly express the dependence of the choice probability of product j on density functions $f_j = (f_{\boldsymbol{y}_j}, f_{\boldsymbol{x}})$, where $\boldsymbol{y}_j = (\boldsymbol{d}_j, \boldsymbol{x})$ and $f_{\boldsymbol{y}_j}(1, x) = \partial \Pr(\boldsymbol{d}_j = 1, \boldsymbol{x} \le x)/\partial x$. In this chapter, I use $\psi(\cdot)$, as in Equation (2.3), to denote $\psi(\cdot; f_j)$ when it is unnecessary to explicitly denote the dependence on f_j. Furthermore, I define $f = (f_{\boldsymbol{y}_1}, f_{\boldsymbol{y}_2}, f_{\boldsymbol{x}})$. $b(x)$ is a specified non-negative weighting function on $\mathcal{X}$ and is of bounded variation, and $F_{\boldsymbol{x}}(\cdot)$ is a CDF of $\boldsymbol{x}$. Under the null, $\Gamma(f, F_{\boldsymbol{x}}) = 0$ for any $F_{\boldsymbol{x}}$ and under the alternative, $\Gamma(f, F_{\boldsymbol{x}}) > 0$ for some $F_{\boldsymbol{x}}$.

2.3.2 Test Statistic and Distribution

Before I define my test statistic, I introduce kernel estimators for unknown densities and distributions. I define

$$K_h(u) = h^{-\delta} K(u/h),$$

where u has dimension δ and h is bandwidth that converges to 0 as $n \to \infty$. I use the standard Nadaraya-Watson density and distribution estimator,

$$\hat{f}(x) = n^{-1} \sum_{i=1}^{n} K_h(x_i - x),$$

and

$$p(j|x; \hat{f}_j) = n^{-1} \sum_{i=1}^{n} \frac{d_{ij} K_h(x_i - x)}{\hat{f}(x)}$$

Let $\hat{F}_{\boldsymbol{x}}$ be the empirical CDF of $\boldsymbol{x}$ and $\hat{f}$ be the Nadaraya-Watson estimator of f, $\omega_{ij} = (w_{ij}, w_{ij'}, w'_{j'})$, $\bar{\omega}_{ij} = (w_{ij}, \bar{w}_{j'}, \underline{w}_{j'})$, where $j' \neq j$, and $b_i = b(x_i)$. My test statistic is

$$\hat{\Gamma} = \Gamma(\hat{f}, \hat{F}_{\boldsymbol{x}}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1,2} \int_{\underline{w}_{j'}}^{w_{ij'}} D_a(\psi(\hat{f}_j))(j, \omega_{ij}, z_i)^2 b_i \, dw'_{j'},$$

where

$$\psi(j, \omega_{ij}, z_i; \hat{f}_j) = \frac{\Delta_{w'_{j'}}^{w_{ij'}} p(j|w_{ij}, \cdot, z_i; \hat{f}_j)}{\Delta_{\underline{w}_{j'}}^{\bar{w}_{j'}} p(j|w_{ij}, \cdot, z_i; \hat{f}_j)}.$$

I make the following assumptions to derive the asymptotic distribution of the test.

Assumption 2.9:

(i) $\{y_i = (d_i, w_i, z_i) \in \{0,1\}^J \times \mathbb{R}^{\delta_1+\delta_2}\}_{i=1}^n$ are i.i.d., and let $\delta = \delta_1 + \delta_2$.

(ii) $\boldsymbol{x}$ has joint distribution $F_{\boldsymbol{x}}$ and joint density $f_{\boldsymbol{x}}$ such that $f_{\boldsymbol{x}}$ has continuous partial derivatives of order $r > 2\delta$, bounded and integrable on $\mathbb{R}^\delta$. $f_{\boldsymbol{x}}$ is bounded away from zero on the compact support $\mathcal{X}$, i.e., $0 < c_1 \leq f_{\boldsymbol{x}}(x) \leq C_1 < \infty$ and $x \in \mathcal{X}$ for some constant c_1 and C_1. Furthermore, $f_{\boldsymbol{x}}$ is of bounded variation in $\mathcal{X}$.

Assumption 2.10: For some integer $r > 2\delta$, kernel K is a product kernel of the bounded symmetric kernel $k : \mathbb{R} \to \mathbb{R}$ satisfying $\int u^i k(u)\, du = \delta_{i0}$ for $i = 0, 1, \cdots, r-1$, $\int u^r k(u)\, du < \infty$, $\int |k(u)|\, du < \infty$, $\int |uk(u)|\, du < \infty$, where δ_{ij} is Kronecker's delta. Furthermore, $k(u)$ is of bounded variation.

Assumption 2.11: As $n \to \infty$, the bandwidth sequence $h = O(n^{-1/\tau})$ for $4\delta < \tau < 2r + \delta/2$.

Under Assumption 2.11, I have (i) $nh^{4\delta}/\ln(n) \to \infty$; (ii) $nh^{\delta/2+2r} \to 0$; and (iii) $nh^{2\delta}/\{\ln(h^{-1})\}^3 \to \infty$.

My main result is that the test statistic is asymptotically normally distributed with an asymptotic bias. I first define the notation necessary to state the result:

$$\kappa(x,\tilde{y}_j) = p(j|x)K_h(\tilde{x}-x)\left\{\frac{\tilde{d}_j}{f_{\boldsymbol{y}_j}(1,x)} - \frac{1}{f_{\boldsymbol{x}}(x)}\right\}, \quad R^*(j,\omega_j,z,\tilde{y}_j) = \frac{\Delta_{w'_{j'}}^{w_{j'}}\kappa(j,w_j,\cdot,z,\tilde{y}_j)}{\Delta_{w'_{j'}}^{w_{j'}}p(j|w_j,\cdot,z)},$$

and

$$\begin{aligned} R(j,\omega_j,z,\tilde{y}_j) =& \psi(j,\omega_j,z)\{R_j^*(j,\omega_j,z,\tilde{y}_j) - R^*(j,\bar{\omega}_j,z,\tilde{y}_j)\} \\ & - \int_{\mathcal{Z}} \psi(j,\omega_j,z')\{R_j^*(j,\omega_j,z',\tilde{y}_j) - R^*(j,\bar{\omega}_j,z',\tilde{y}_j)\}a(z')\,dz'. \end{aligned}$$

I then define

$$\begin{aligned} B =& h^{\delta/2}\sum_{j=1,2}\int_{\mathcal{X}}\int_{\underline{w}_{j'}}^{w_{j'}} \mathbb{E}[R(j,\omega_j,z,\tilde{\boldsymbol{y}}_j)^2]b(x)\,dw'_{j'}\,dF_{\boldsymbol{x}}(x) \\ \sigma^2 =& h^{\delta}\sum_{j=1,2}\sum_{l=1,2}\int_{\mathcal{X}}\int_{\mathcal{X}}\int_{\underline{w}_{j'}}^{w_{j'}}\int_{\underline{w}_{l'}}^{w''_{l'}} \mathbb{E}[R(j,\omega_j,z,\tilde{\boldsymbol{y}}_j)R(l,\omega''_l,z'',\tilde{\boldsymbol{y}}_l)]^2 \\ & \times b(x)b(x'')\,dw'_{j'}\,dw'''_{l'}\,dF_{\boldsymbol{x}}(x)\,dF_{\boldsymbol{x}}(x''). \end{aligned}$$

B and σ^2 converge to some constant as R also depends on h. The details can be found in Appendix B.

Theorem 2.1: Under Assumptions 2.1–2.11 and H_0, I have

$$\frac{nh^{\delta/2}\hat{\Gamma} - B}{\sqrt{2\sigma^2}} \to \mathcal{N}(0,1).$$

The proof relies on the functional expansion introduced by Aït-Sahalia, Bickel, and Stoker (2001).

Note that B and σ^2 are population objects. I write $\hat{R}_{ikj} = \hat{R}(j,\omega_{ij},z_i,y_{kj})$. I estimate

B and σ^2 as follows.

$$\hat{B} = \frac{1}{n^2} \sum_{i,k=1}^{n} h^{\delta/2} \sum_{j=1,2} \int_{\underline{w}_{j'}}^{w_{ij'}} \hat{R}_{ikj}^2 b_i \, dw'_{j'}$$

and

$$\hat{\sigma}^2 = \frac{1}{n^2} h^{\delta} \sum_{i,\ell=1}^{n} \sum_{j,l=1,2} \int_{\underline{w}_{l'}}^{w_{\ell l'}} \int_{\underline{w}_{j'}}^{w_{ij'}} \left[\frac{1}{n} \sum_{k=1}^{n} \hat{R}_{ikj} \hat{R}_{\ell kl} \right]^2 b_i b_\ell \, dw'_{j'} \, dw'_{l'},$$

where $\hat{R}$ is defined as R replacing f_j with $\hat{f}_j$ and $\hat{f}_j$ is the Nadaraya-Watson estimator of f_j. I show in Appendix B that the estimators of B and σ^2 are consistent. To implement the test, I compare

$$T_n = \frac{nh^{\delta/2}\hat{\Gamma} - \hat{B}}{\sqrt{2\hat{\sigma}^2}}$$

to the critical value from the $\mathcal{N}(0,1)$ distribution. Since the test is one-sided, I reject the null if T_n is greater than the critical value.

Now I study the consistency of my test.

Theorem 2.2: If Assumptions 2.1–2.11 hold, the test statistic, T_n, is consistent for $F_{\boldsymbol{x}}$ such that $\Gamma(f, F_{\boldsymbol{x}}) > 0$. Furthermore, $\Pr(T_n > o(nh^{\delta/2})|H_A) \to 1$ as $n \to \infty$.

2.4 General Model and Tests

Now I turn to the case with a general number of products, $J > 2$. I focus on product 1. The choice probability of product 1 under the general case is

$$p(1|x) = \sum_{C \in \mathcal{C}_1^{\mathcal{J}}} p(1|x, C) p(C|x), \tag{2.5}$$

where $\mathcal{C}_1^{\mathcal{J}}$ denotes the family of consideration sets that include product 1 when the choice set is $\mathcal{J}$. Equation (2.5) corresponds to Equation (2.1) in the two-product case.

I discuss how I construct test functions $\psi(\cdot)$ for the general case from Equation (2.5) under Assumptions 2.3–2.5.

2.4.1 Testing the Hypothesis

In this section, I focus on constructing a function analogous to Equation (2.3) from Equation (2.5) under Assumptions 2.3 and 2.4. First, note that the cardinality of $\mathcal{C}_1^{\mathcal{J}}$ is always even for $J \geq 2$, $|C_1^{\mathcal{J}}| = 2^{J-1}$. I can divide $\mathcal{C}_1^{\mathcal{J}}$ into halves, where one half is a family of sets that include product 2 and the other half is a family of sets that does not include product 2. I use notation $\mathcal{A} \sqcup D$ to denote $\{A \cup D : A \in \mathcal{A}\}$, where $\mathcal{A}$ is a family of sets and D is a set. As before, I use the shorthand 012 for $\{0, 1, 2\}$ and likewise for the other sets. I can write $\mathcal{C}_1^{\mathcal{J}} = (\mathcal{C}_1^{\mathcal{J}\setminus 2} \sqcup 2) \cup \mathcal{C}_1^{\mathcal{J}\setminus 2}$; therefore, for any $C' \in \mathcal{C}_1^{\mathcal{J}\setminus 2}$, both C' and $C' \cup 2$ are in $C_1^{\mathcal{J}}$. If I choose any $C' \in \mathcal{C}_1^{\mathcal{J}\setminus 2}$, I have event C' or $C' \cup 2$. This indicates that regardless of her attention to product 2, the consumer is attentive to product $j \in C'$ and inattentive to $j' \notin C'$. Hence, I can rewrite Equation (2.5) as

$$
\begin{aligned}
p(1|w_1, w_{-1}, z) = &\sum_{C' \in \mathcal{C}_1^{\mathcal{J}\setminus 2}} p(1|w_1, w_{-1}, z, C')p(C' \text{ or } C' \cup 2|w_1, w_{-1,-2}, z) \\
&+ \{p(1|x, C' \cup 2) - p(1|w_1, w_{-1}, z, C')\}p(C' \cup 2|w_1, w_{-1}, z) \\
= &\sum_{C' \in \mathcal{C}_1^{\mathcal{J}\setminus 2}} p(1|z, C')p(C' \text{ or } C' \cup 2|w_1, w_{-1,-2}, z) \\
&+ \{p(1|z, C' \cup 2) - p(1|z, C')\}p(C' \cup 2|w_1, w_{-1}, z). \qquad (2.6)
\end{aligned}
$$

From Equation (2.6), the first equality is true by Assumption 2.3 and $p(C'|x) = p(C' \text{ or } C' \cup 2|x) - p(C' \cup 2|x)$. This is possible because C' and $C' \cup 2$ are mutually exclusive events. The second equality is true by Assumption 2.4. Note that Equation (2.6) is analogous to Equation (2.2). When I change w_2 to w_2' with $w_2' < w_2$, the only term that changes in

Equation (2.6) is $p(C' \cup 2 | w_1, w_{-1}, z)$. Then I have

$$\Delta_{w_2'}^{w_2} p(1|w_1, \cdot, w_{-1,-2}, z) = \sum_{C' \in \mathcal{C}_1^{\mathcal{J}\backslash 2}} \{p(1|z, C' \cup 2) - p(1|z, C')\} \Delta_{w_2'}^{w_2} p(C' \cup 2 | w_1, \cdot, w_{-1,-2}, z). \tag{2.7}$$

I repeat the exercise that I performed on Equation (2.6) on (2.7). I can divide $\mathcal{C}_1^{\mathcal{J}\backslash 2}$ into halves, where one half is a family of sets that includes product 3 and the other half is a family of sets that does not include product 3. I can write $\mathcal{C}_1^{\mathcal{J}\backslash 2} = (\mathcal{C}_1^{\mathcal{J}\backslash 23} \sqcup 3) \cup \mathcal{C}_1^{\mathcal{J}\backslash 23}$. For any $C'' \in \mathcal{C}_1^{\mathcal{J}\backslash 23}$, C'' and $C'' \cup 3$ are in $\mathcal{C}_1^{\mathcal{J}\backslash 2}$. As before, I can rewrite Equation (2.7) by setting $p(C'' \cup 2|x) = p(C'' \cup 2 \text{ or } C'' \cup 23|x) - p(C'' \cup 23|x)$:

$$\begin{aligned}\Delta_{w_2'}^{w_2} p(1|w_1, \cdot, w_{-1,-2}, z) = &\sum_{C'' \in \mathcal{C}_1^{\mathcal{J}\backslash 23}} \{p(1|z, C'' \cup 2) - p(1|z, C'')\} \Delta_{w_2'}^{w_2} p(C'' \cup 2 \text{ or } C'' \cup 23 | w_1, \cdot, w_{-1,-2,-3}, z) \\ &+ [p(1|z, C'' \cup 23) - p(1|z, C'' \cup 3) - \{p(1|z, C'' \cup 2) - p(1|z, C'')\}] \\ &\times \Delta_{w_2'}^{w_2} p(C'' \cup 23 | w_1, \cdot, w_{-1,-2}, z).\end{aligned} \tag{2.8}$$

In Equation (2.8), only $\Delta_{w_2'}^{w_2} p(C'' \cup 23|w_1, \cdot, w_{-1,-2}, z)$ varies across w_3. I choose $w_3' < w_3$. Thus, using Equation (2.8), I have

$$\begin{aligned}\Delta_{w_3'}^{w_3}[\Delta_{w_2'}^{w_2} p(1|w_1, \cdot, w_{-1,-2,-3}, z)] = &\sum_{C'' \in \mathcal{C}_1^{\mathcal{J}\backslash 23}} [p(1|z, C'' \cup 23) - p(1|z, C'' \cup 3) - \{p(1|z, C'' \cup 2) - p(1|z, C'')\}] \\ &\times \Delta_{w_3'}^{w_3}[\Delta_{w_2'}^{w_2} p(C'' \cup 23 | w_1, \cdot, w_{-1,-2,-3}, z)].\end{aligned} \tag{2.9}$$

From Equation (2.5) to Equation (2.7) and from Equation (2.7) to Equation (2.9), the number of summands decreases by one each time. I can perform the same exercise repeatedly until only one summand is left. To state the result from the repetition, I define a finite sequence $\{\mathcal{C}_{1,\ell}\}_{\ell=1}^{L(\mathcal{J})}$, where $\mathcal{C}_{1,1} = \{\{0,1\}\}$, $\mathcal{C}_{1,L(\mathcal{J})} = \{\mathcal{J}\}$ and for $1 < \ell < L(\mathcal{J})$ $\mathcal{C}_{1,\ell} = \{C \cup \{k\} : C \in \mathcal{C}_{\ell-1}, k \in \mathcal{J}\backslash C\}$. I further define

$$\psi_{CP}^*(1, z) = (-1)^{J+1} \sum_{\ell}^{L(\mathcal{J})} \sum_{C \in \mathcal{C}_{1,\ell}} (-1)^{|C|} p(1|z, C).$$

The result of the repetition is

$$\Delta^{w_J}_{w'_J}\left[\cdots[\Delta^{w_2}_{w'_2}p(1|w_1,\cdot,z)]\right]=\psi^*_{CP}(1,z)\Delta^{w_J}_{w'_J}\left[\cdots[\Delta^{w_2}_{w'_2}p(\mathcal{J}|w_1,\cdot,z)]\right]. \tag{2.10}$$

Equation (2.10) is analogous to Equation (2.3). As in the case of Equation (2.3), here I need the left-hand side of Equation (2.10) to be bounded away from zero to construct the valid test. By Assumption 2.5–(i), the mean value theorem, and $w_j > w'_j$ for all $j \neq 1$, $\Delta^{w_J}_{w'_J}\left[\cdots[\Delta^{w_2}_{w'_2}p(\mathcal{J}|w_1,\cdot,z)]\right]$ is bounded away from zero. By Assumption 2.5–(ii), $\psi^*_{CP}(1,z)$ is bounded away from zero.

Now for any choice of $w \in \mathcal{W}$ and $w'_{-1} \in \mathcal{W}_{-1}$ with $w_{-1} > w'_{-1}$, I can construct the test function $\psi(\cdot)$:

$$\begin{aligned}\psi(1,w_1,w_{-1},w'_{-1},z) &= \frac{\Delta^{w_J}_{w'_J}\left[\cdots[\Delta^{w_2}_{w'_2}p(1|w_1,\cdot,z)]\right]}{\Delta^{\bar{w}_J}_{\underline{w}_J}\left[\cdots[\Delta^{\bar{w}_2}_{\underline{w}_2}p(1|w_1,\cdot,z)]\right]}\\ &= \frac{\Delta^{w_J}_{w'_J}\left[\cdots[\Delta^{w_2}_{w'_2}p(\mathcal{J}|w_1,\cdot,z)]\right]}{\Delta^{\bar{w}_J}_{\underline{w}_J}\left[\cdots[\Delta^{\bar{w}_2}_{\underline{w}_2}p(\mathcal{J}|w_1,\cdot,z)]\right]},\end{aligned}$$

where the second equality directly follows from Equation (2.10). As in the previous section, I suggest testing the hypothesis using $\psi(\cdot)$. The following lemma allows me to use $\psi(\cdot)$ for the test:

Lemma 2.2: Suppose Assumptions 2.1–2.8 hold. Then $\psi(\cdot)$ does not vary with z if and only if $p(\mathcal{J}|w,z)$ does not vary with z.

As in Section 2.3.1, I write $\omega_j = (w_j, w_{-j}, w'_{-j})$ and $\bar{\omega}_j = (w_j, \bar{w}_{-j}, \underline{w}_{-j})$. By Lemma 2.2, I can rewrite the hypothesis as

$$H_0: \ \forall j, \ \omega_j, z, z', \ \psi(j,\omega_j,z) = \psi(j,\omega_j,z'),$$

and

$$H_A: \ \exists j, \ \omega_j, z, z', \ \psi(j, \omega_j, z) \neq \psi(j, \omega_j, z').$$

Under the null, I have

$$\forall j \ \forall z \ \forall \omega_j, \ D_a(\psi)(j, \omega_j, z) = \psi(j, \omega_j, z) - \int_{\mathcal{Z}} \psi(j, \omega_j, z') a(z') \, dz' = 0,$$

where $a(z')$ is defined as it is in the previous section. I assess the significance of the square of $D_a(\psi)$. Thus,

$$\Gamma(f, F_{\boldsymbol{x}}) = \sum_{j \in \mathcal{J} \setminus 0} \int_{\mathcal{X}} \int_{\underline{w}_{-j}}^{w_{-j}} D_a(\psi(f_j))(j, \omega_j, z)^2 b(x) \, dw'_{-j} \, dF_{\boldsymbol{x}}(x), \tag{2.11}$$

where

$$\psi(j, \omega_j, z; f_j) = \frac{\Delta_{w'_J}^{w_J} \left[\cdots \Delta_{w'_{j+1}}^{w_{j+1}} \left[\Delta_{w'_{j-1}}^{w_{j-1}} \left[\cdots [\Delta_{w'_1}^{w_1} p(j | w_j, \cdot, z; f_j)] \right] \right] \right]}{\Delta_{\underline{w}_J}^{\bar{w}_J} \left[\cdots \Delta_{\underline{w}_{j+1}}^{\bar{w}_{j+1}} \left[\Delta_{\underline{w}_{j-1}}^{\bar{w}_{j-1}} \left[\cdots [\Delta_{\underline{w}_1}^{\bar{w}_1} p(j | w_j, \cdot, z; f_j)] \right] \right] \right]}$$

and f_j, $b(\cdot)$, and $F_{\boldsymbol{x}}$ are defined as they are in Equation (2.4). I define $f = (f_{\boldsymbol{y}_1}, \cdots, f_{\boldsymbol{y}_J}, f_{\boldsymbol{x}})$.

2.4.2 Test Statistic and Distribution

I derive my test statistic based on the Γ from Equation (2.11). I write $\omega_{ij} = (w_{ij}, w_{i,-j}, w'_{-j})$, $\bar{\omega}_{ij} = (w_{ij}, \bar{w}_{-j}, \underline{w}_{-j})$, and $b_i = b(x_i)$. My test statistic is

$$\hat{\Gamma} = \Gamma(\hat{f}, \hat{F}_{\boldsymbol{x}}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j \in \mathcal{J} \setminus 0} \int_{\underline{w}'_{-j}}^{w_{i,-j}} D_a(\psi(\hat{f}_j))(j, \omega_{ij}, z_i)^2 b_i \, dw'_{-j},$$

where

$$\psi(j, \omega_{ij}, z_i; \hat{f}_j) = \frac{\Delta_{w'_J}^{w_{iJ}} \left[\cdots \Delta_{w'_{j+1}}^{w_{i,j+1}} \left[\Delta_{w'_{j-1}}^{w_{i,j-1}} \left[\cdots [\Delta_{w'_1}^{w_{i1}} p(j | w_{ij}, \cdot, z_i; \hat{f}_j)] \right] \right] \right]}{\Delta_{\underline{w}_J}^{\bar{w}_J} \left[\cdots \Delta_{\underline{w}_{j+1}}^{\bar{w}_{j+1}} \left[\Delta_{\underline{w}_{j-1}}^{\bar{w}_{j-1}} \left[\cdots [\Delta_{\underline{w}_1}^{\bar{w}_1} p(j | w_{ij}, \cdot, z_i; \hat{f}_j)] \right] \right] \right]}.$$

Before I state the asymptotic distribution of my test, I define the notation necessary to state the result:

$$\kappa(j,x,\tilde{y}_j) = p(j|x)K_h(\tilde{x}-x)\left(\frac{\tilde{d}_j}{f_{\boldsymbol{y}_j}(1,x)} - \frac{1}{f_{\boldsymbol{x}}(x)}\right),$$

$$R^*(j,\omega_j,z,\tilde{y}_j) = \frac{\Delta^{w_J}_{w'_J}\left[\cdots\Delta^{w_{j+1}}_{w'_{j+1}}\left[\Delta^{w_{j-1}}_{w'_{j-1}}\left[\cdots[\Delta^{w_1}_{w'_1}\kappa(j,w_j,\cdot,z,\tilde{y}_j)]\right]\right]\right]}{\Delta^{w_J}_{w'_J}\left[\cdots\Delta^{w_{j+1}}_{w'_{j+1}}\left[\Delta^{w_{j-1}}_{w'_{j-1}}\left[\cdots[\Delta^{w_1}_{w'_1}p(j|w_j,\cdot,z)]\right]\right]\right]},$$

and

$$\begin{aligned} R(j,\omega_j,z,\tilde{y}_j) =& \psi(j,\omega_j,z)\{R^*(j,\omega_j,z,\tilde{y}_j) - R^*(j,\bar{\omega}_j,z,\tilde{y}_j)\} \\ &- \int_{\mathcal{Z}} \psi(j,\omega_j,z)\{R^*(j,\omega_j,z,\tilde{y}_j) - R^*(j,\bar{\omega}_j,z,\tilde{y}_j)\}a(z')\,dz'. \end{aligned}$$

I further define

$$\begin{aligned} B =& h^{\delta/2}\sum_{j\in\mathcal{J}\backslash 0}\int_{\mathcal{X}}\int_{\underline{w}_{-j}}^{w_{-j}} \mathbb{E}[R(j,\omega_j,z,\tilde{\boldsymbol{y}}_j)^2]b(x)\,dw'_{-j}\,dF_{\boldsymbol{x}}(x) \\ \sigma^2 =& h^{\delta}\sum_{j\in\mathcal{J}\backslash 0}\sum_{l\in\mathcal{J}\backslash 0}\int_{\mathcal{X}}\int_{\mathcal{X}}\int_{\underline{w}_{-j}}^{w_{-j}}\int_{\underline{w}_{-l}}^{w''_{-l}} \mathbb{E}[R(j,\omega_j,z,\tilde{\boldsymbol{y}}_j)R(l,\omega''_l,z'',\tilde{\boldsymbol{y}}_l)]^2 \\ &\times b(x)b(x'')\,dw'_{-j}\,dw'''_{-l}\,dF_{\boldsymbol{x}}(x)\,dF_{\boldsymbol{x}}(x''). \end{aligned}$$

B and σ^2 converge to some constant since R also depends on h. The details can be found in Appendix B.

Theorem 2.3: Under Assumptions 2.1–2.11 and H_0, I have

$$\frac{nh^{\delta/2}\hat{\Gamma} - B}{\sqrt{2\sigma^2}} \to \mathcal{N}(0,1).$$

The proof relies on the functional expansion introduced by Aït-Sahalia et al. (2001).

Note that B and σ^2 are population objects. I write $\hat{R}_{ikj} = \hat{R}(j, \omega_{ij}, z_i, y_{kj})$. I estimate B and σ^2 by $\hat{B}$ and $\hat{\sigma}^2$, respectively, as follows.

$$\hat{B} = \frac{1}{n^2} \sum_{i,k=1}^{n} h^{\delta/2} \sum_{j \in \mathcal{J} \setminus 0} \int_{\underline{w}_{-j}}^{w_{i,-j}} \hat{R}_{ikj}^2 b_i \, dw'_{-j}$$

and

$$\hat{\sigma}^2 = \frac{1}{n^2} h^{\delta} \sum_{i,\ell=1}^{n} \sum_{j,l \in \mathcal{J} \setminus 0} \int_{\underline{w}_{-l}}^{w_{\ell,-l}} \int_{\underline{w}_{-j}}^{w_{i,-j}} \left[\frac{1}{n} \sum_{k=1}^{n} \hat{R}_{ikj} \hat{R}_{\ell kl} \right]^2 b_i b_\ell \, dw'_{-j} \, dw'_{-l},$$

where $\hat{R}$ is defined as R replacing f_j with $\hat{f}_j$ and $\hat{f}_j$ is the Nadaraya-Watson estimator of f_j. I show in Appendix B that the estimators of B and σ^2 are consistent. To implement the test, I compare

$$T_n = \frac{n h^{\delta/2} \hat{\Gamma} - \hat{B}}{\sqrt{2 \hat{\sigma}^2}}$$

to the critical value from the $\mathcal{N}(0, 1)$ distribution. Since the test is one-sided, I reject the null if T_n is greater than the critical value.

Now I study the consistency of my test.

Theorem 2.4: If Assumptions 2.1–2.11 hold, test statistic T_n is consistent for $F_{\boldsymbol{x}}$ such that $\Gamma(f, F_{\boldsymbol{x}}) > 0$. Furthermore, $\Pr(T_n > o(n h^{\delta/2}) | H_A) \to 1$ as $n \to \infty$.

2.5 A Monte Carlo Study

In this section, I present Monte Carlo study results to provide the finite sample performance of my test. I consider a simple model with two products and an outside option. The consumer's attention $\boldsymbol{a}_{ij}$ to product j is determined as follows:

$$\boldsymbol{a}_{ij} = \begin{cases} 10\boldsymbol{w}_{ij} - \boldsymbol{\eta}_{ij} & \text{if } j = 2 \\ \infty & \text{else} \end{cases}, \tag{2.12}$$

where $\boldsymbol{w}_{i2}$ is distributed normal with zero mean and unit variance, and $\boldsymbol{\eta}_{i2}$ follows the logistic distribution with zero location and unit scale parameters. The consumer considers buying product j if $a_{ij} > 0$. Therefore, equation (2.12) describes that the consumer always considers the outside option and product 1, but may or may not consider buying product 2.

Then, the consumer's utility $\boldsymbol{u}_{ij}$ for product j is determined as follows:

$$\boldsymbol{u}_{ij} = \begin{cases} 0.1\boldsymbol{z}_{ij} + \boldsymbol{\varepsilon}_{ij} & \text{if } j = 1 \\ 5 + \boldsymbol{\varepsilon}_{ij} & \text{if } j = 2 \\ \boldsymbol{\varepsilon}_{ij} & \text{if } j = 0 \end{cases}, \tag{2.13}$$

where $\boldsymbol{z}_{i1}$ is distributed normal with zero mean and unit variance, and $\boldsymbol{\varepsilon}_{ij}$ follows the gumbel distribution with zero location and unit scale parameters. In this simulation study, all random variables are mutually independent. Under this setting, I can only perform my test using the choice probability of product 1, and test function $\psi(\cdot)$ is

$$\psi(1, w_{i2}, w'_{i2}, z_{i1}) = \frac{p(1|w_{i2}, z_{i1}) - p(1|w'_{i2}, z_{i1})}{p(1|\bar{w}_2, z_{i1}) - p(1|\underline{w}_2, z_{i1})}.$$

I choose a univariate normal kernel function for all density estimations and regressions with bandwidth $h = 0.6$. I set $\bar{w}_2 = 2$, $\underline{w}_2 = -2$, and $b(x) = 1$ for all x. To simplify the simulation, I choose $w'_{i2} = -2$ for all i. Instead of choosing the weighting function $a(\cdot)$, I redefine $D_a(\psi)(\cdot)$ as follows:

$$D_a(\psi)(1, w_{i2}, w'_{i2}, z_{i1}) = \psi(1, w_{i2}, w'_{i2}, z_{i1}) - \psi(1, w_{i2}, w'_{i2}, 0).$$

These choices of w'_{i2} and $D_a(\psi)(\cdot)$ do not affect the results of this chapter, but only reduce the power of my test. Under this simulation setting, $\hat{\Gamma}$ is

$$\hat{\Gamma} = \frac{1}{n} \sum_{i=1}^{n} D_a(\psi)(1, w_{i2}, -2, z_{i1})^2.$$

Table 2.1 below reports the estimated rejection rates of my test in the 5% and 10% tests, and the mean and the standard deviation of test statistic T_n. The estimations are obtained from $1,000$ simulations.

Table 2.1: Monte Carlo Simulation Results

	5%	10%	Mean	Std
$n = 500$	3.6	6.7	-0.0962	0.7944
$n = 1000$	4.5	6.8	-0.0628	0.8350
$n = 2000$	5.5	7.8	-0.0120	0.9444
$n = 5000$	6.6	9.2	-0.0252	0.9524

Table 2.1 shows that my test under-rejects the null hypothesis in the 5% test for small sample sizes (n less than 1,000), but it starts to over-reject the null for relatively large sample sizes (n greater than 2,000). In the 10% test, my test tends to under-reject the null, but the rejection rate approaches the population rate as the sample size grows. Moreover, the mean and the standard deviation of my test is getting closer to the population values (0 and 1, respectively) as the sample size increases.

2.6 Conclusion

In this chapter, I studied a nonparametric test of the validity of an exclusion restriction in the consideration set model. The econometrician obtains point identification with two variables – one variable only controls preferences, and the other only affects attention. Each variable works as an exclusion restriction since each is excluded from either attention or preferences. As long as the econometrician has one of the variables, partial identification is possible. My test checks whether the econometrician can use a candidate variable for one type of exclusion restriction if the other type of exclusion restriction is available. To develop this test, I create a test function that does not vary with candidate variable under the null. The difference between the test function and the integral of the test function with respect to some weighting function on the space of the candidate variable is zero if the candidate variable can be used for the exclusion restriction to obtain point identification. My test subsequently measures the significance of the square of the difference. I use the kernel method to construct the test statistic and show that the test statistic is consistent.

Chapter 3

Welfare Cost of Information Asymmetry in E-commerce: Evidence from Expedia

3.1 Introduction

E-commerce sales have grown in recent years, reaching $4.2 trillion in 2020, or 18% of total global retail sales (Statista, 2022). Total sales increased by 24% during the COVID-19 pandemic (Statista, 2022). This growth in E-commerce is due to the vast number of products offered. As a result, E-commerce platforms are considered a major source of product searches. According to a recent survey, approximately 40% of consumers searched for products using E-commerce platforms (Statista, 2021). To provide a better search experience, platforms provide consumers with rankings. A ranking is a list of recommended products that helps consumers with limited information find products that fit their needs.

Crucially, an E-commerce platform has the power to manipulate such rankings to boost its profit. This power comes from two sources. First, the platform may have more product information than the consumer. If the consumer uses the platform as her primary means of product search, it means she has limited knowledge. For example, the consumer may not be aware of product availability or product characteristics. Second, the ranking influences the amount of consumer attention directed toward the various products (De los Santos & Koulayev, 2017; Chen & Yao, 2017; Ursu, 2018). The higher the ranking of a product is (i.e., the more likely the product is visible to the consumer), the more likely the consumer is to be attentive to the product. Therefore, the platform can place highly profitable products in highly visible positions in the rankings to boost its profit. In other words, the consumer may purchase a product advantageous to the platform rather than herself because product information is limited and the ranking influences her attention.

This chapter investigates how information asymmetry between a consumer and an E-commerce platform governs consumer welfare in the presence of an endogenous ranking system. To answer my research question, I take the following two steps. First, I develop a structural model of the E-commerce market featuring an endogenous ranking. Then I use Expedia hotel reservation data to estimate the model parameters. Second, I use the model estimates to quantify the impact of information asymmetry on consumer welfare. To conduct a counterfactual analysis, I consider a scenario in which the consumer has complete knowledge. I then apply my model to propose a policy that can alleviate the problem caused by information asymmetry and improve welfare.

I start by describing my model. My model is based on a two-period sequential game between a platform and a consumer. The platform designs a ranking algorithm to maximize its ex-ante expected profit before any consumer enters the market. The ranking algorithm cannot be modified by the platform once designed. When the consumer enters the market, the consumer submits a search query. Based on the search query, the platform gives her the rankings and characteristics of available products. The consumer then makes a choice.

Using my model and estimates, I find that the platform's optimal ranking algorithm uses both the potential profit from selling a product and the consumer utility from that product to generate a ranking. In other words, the product with greater utility ranks more highly, and the product that generates more profit to the platform ranks more highly. Of course, the relative effects of consumer utility and of potential profit on a ranking are different. An increase in consumer utility by one unit has an effect equivalent to a $28.25 increase in the platform's potential profit.

I perform a counterfactual analysis assuming that the consumer has complete information about every product on the market. When the consumer has complete information, the platform modifies the ranking algorithm such that it is more beneficial to the consumer than it is under information asymmetry. This means that the relative effects of consumer utility and potential profit on a ranking change. Now, a one-unit increase in consumer utility has the same effect as a $37.96 increase in potential profit – that is, the relative significance of consumer utility in the ranking increases. As a result, the consumer is more likely to observe products that match her needs and search query at the top of the ranking. Therefore, the consumer is more likely to purchase a product and experience an average increase in consumer welfare of 2.24%. Surprisingly, just as consumer welfare increases, the platform's ex-ante expected profit rises by about 58%. This gain is due to the inflated probability of purchasing a product. To be more precise, the consumer who would not have bought a product under information asymmetry is now more likely to buy a product. Overall, if the consumer has complete information, total welfare, which is the sum of consumer welfare and platform profit, increases by about 38%. This counterfactual result has an important policy implication. The policy that reduces information asymmetry increases both consumer welfare and the platform's ex-ante expected profit. Such a policy may include mandating the disclosure of certain product information at the search level.

My research relates to the literature on incomplete information in E-commerce. The literature generally considers environments in which the consumer has incomplete product information without considering information asymmetry between agents. Zhu and Zhang (2010), Newberry and Zhou (2016), and Fang (2019) show that online consumer reviews convey unknown product information to the consumer. For example, Newberry and Zhou

(2016) study the impact of online reviews on consumer behavior on an E-commerce platform. They show that online reviews inform the consumer about a retailer's intrinsic quality. As a result, small local retailers can better compete with large national retailers. Bai, Chen, Liu, and Xu (2020) study the consumer's incomplete product information in the presence of a ranking. In their setting, a ranking is determined by past sales. They show that reducing incomplete information improves consumer welfare. They also show that consumer welfare improves when the true quality of each product determines the initial ranking. I build on this literature by showing that consumer welfare improves if the consumer has complete information about every product on the market. This improvement is not only because she has complete information but also because the platform modifies the ranking algorithm such that it is more beneficial to the consumer than it is under information asymmetry.

The rest of the chapter is organized as follows. Section 3.2 illustrates the institutional background information and data. Section 3.3 presents my structural model. Section 3.4 discusses the estimation and identification. Section 3.5 reports the estimation and counterfactual results. Section 3.6 concludes.

3.2 Industry Background and Data

3.2.1 Institutional Detail

This subsection details the online travel agency industry. In 2013, 76% of online hotel reservations were made through online travel agents (OTAs) (Christopher, 2013). In the same year in the United States, the industry was highly concentrated, with five OTAs accounting for 99% of the market share (Statista, 2015). These five OTAs are Expedia, Orbitz, the Priceline Group, Travelocity, and CheapOair. Of them, Expedia had the largest market share with 42%. In 2013, approximately 30% of U.S. hotel room reservations were made through OTAs. In fact, about 12.6% of total room reservations were made on Expedia alone (Max, 2013).

The industry operates on two business models: an agency model and a merchant model. In both the agency and merchant models, an OTA acts as a platform connecting consumers and hoteliers, taking commissions from the hotels. The main difference between the two models is that the OTA collects the payment for the room from the consumer in the merchant model, whereas in the agency model, the hotel collects the payment for the room. Furthermore, the OTA can set prices for packaged products in the merchant model. In particular, Expedia operates according to both models but mainly generates sales from the merchant model (75% of total sales in 2012 and 70% in 2013) (Expedia, 2012, 2013). However, in the first quarter of 2013, 54% of bookings were made through the agency model (Ursu, 2018).

The industry's main feature is a ranking of hotels, which is generated by the OTA's internal ranking algorithm. The OTA uses a machine learning algorithm, *Learning-to-Rank*, to rank hotels.[1] This algorithm aims to find the score function $S(\cdot)$ (Cao, Qin, Liu, Tsai, & Li, 2007). The score function maps relevant consumer-hotel specific information x_{ih} into a scalar: $s_{ih} = S(x_{ih})$. The relevant consumer-hotel specific information x_{ih} may include the consumer's preference for hotel h and potential profits from hotel h. This is because Expedia (2022) states that a ranking is tailored to the consumer's preferences while taking profits into account. A ranking, then, is determined in the descending order of the scores. In addition to using the algorithm, the OTA can sell its ranking positions to hoteliers through auctions. These hotels are tagged as *ads* or *sponsored* on its website. It can also reserve some positions in a ranking for less attractive rooms, called opaque offers. In particular, Expedia reserves the 5th, 11th, 17th, and 23rd positions in a ranking for making less attractive rooms visible to the consumer and thereby clearing its inventory.[2]

The hotel establishes a contractual relationship with the OTA and uploads the number,

[1] The data provider claims that Expedia uses the Learning-to-Rank algorithm. For further comments, visit `https://www.kaggle.com/c/expedia-personalized-sort/discussion/5767#30960`.

[2] The hotelier I interviewed stated that the OTA attempts to bulk-purchase hotel rooms.

type, and price of rooms to the platform. Contractually, prices cannot vary with supply channel. The hotel may update prices at any time based on supply and demand. Updated prices must be reflected simultaneously across all supply channels.

These institutional details allow me to make the following three assumptions for structural modeling. First, the OTA in my model charges the hotel a fixed proportion of the price of each hotel room sold, and the hotel sets the price. It is assumed that this fixed proportion is the same for all hotels.[3] Second, the OTA uses the aforementioned learning-to-rank algorithm to rank hotel rooms. In my model, I assume that the OTA chooses score function $S(\cdot)$ to maximize its ex-ante expected profit. This ex-ante expected profit function is discussed later in this chapter. Furthermore, I assume that $S(\cdot)$ takes two arguments: potential profit and the consumer utility up to the OTA's knowledge. I state the consumer utility up to the OTA's knowledge since the OTA does not know the consumer's idiosyncratic taste shock. Third, I assume in this chapter that prices do not depend on the OTA's ranking algorithm or the consumer's idiosyncratic taste shock. The OTA makes up a small fraction of hotel reservations so the OTA's decision has a negligible impact on the hotel's decisions. Additionally, the hotel cannot set a specific price for a particular consumer. Rather, it must update the price across all supply channels before the consumer enters the market.

3.2.2 Data

The data I use come from Kaggle.com and consist of 10,966 search queries for hotels on Expedia. The data encompass 13 destinations and 1,690 hotels. The data were collected from November 1, 2012, to June 30, 2013.

The data were collected at the consumer's search query level. For each consumer, the data include the consumer's search date, travel dates, travel length, any travel companions, the number of rooms requested, their choices (i.e., clicks and bookings), and their booking

[3] I interviewed a hotelier who has worked for three different major hotel chains in Korea. He attested that the Expedia commission rate was the same for each hotel.

history with Expedia. Along with providing information about each consumer's search query, the data include each hotel's characteristics that each consumer observes.

The data have two unique features. First, consumers are randomly assigned to either a random or Expedia ranking system when entering a search query for hotels. If consumers are assigned to the random ranking algorithm system, they observe a randomly generated order of hotels (e.g., random permutation). Otherwise, consumers observe an order of hotels that is generated by Expedia's internal ranking algorithm. Second, consumers do not know about the experiment so they do not know to which ranking system they are assigned.

Table 3.1 provides summary statistics of hotel characteristics that are identical across all consumers. Hotel characteristics include a brand indicator, a location score, an average review score, and a star rating. The brand indicator tells whether a hotel belongs to one of the major chains. The location score measures the attractiveness of a hotel location, where 7 is the most desirable on a scale of 0 to 7. Expedia calculates this score. The average review score is the average of the scores given by guests who have stayed at the hotel. Finally, the star rating measures the overall quality of a hotel and is rated on a scale of 1 to 5, where 5 indicates the highest-quality hotel. A star rating of 0 indicates that the hotel is not rated. Expedia calculates this rating based on information provided by the hotel.

Table 3.1: Hotel Characteristics Summary Statistics

	Observations	Mean	Median	Std	Min	Max
Brand	1,690	0.6550	1	0.4755	0	1
Location Score	1,690	3.3863	3.6850	1.6039	0	5.97
Review Score	1,690	3.7831	4	0.9916	0	5
Star Rating	1,690	3.0870	3	1.0427	0	5

The data encompass 1,690 hotels across 13 destinations. Most of the hotels belong to major hotel chains and are rated 3 stars or higher. On average, a hotel has a location score of 3.39 and a review score of 3.78.

Table 3.2 shows summary statistics on consumer-hotel specific characteristics and consumer-specific characteristics for individuals assigned to the random ranking system. Consumer-hotel characteristics include the price of a room at each hotel and the promotion flag (i.e., whether the hotel offers a discounted rate). Prices and promotion flags vary from consumer to consumer, as different consumers search for a hotel on different dates and travel on different dates. The data include the number of hotels each consumer observes and which hotels each consumer clicks on or books.

Table 3.2: Consumer-Hotel/Consumer-Specific Characteristics (Random Ranking) Summary Statistics

	Observations	Mean	Median	Std	Min	Max
Price	118,276	195.8034	163.2400	121.1387	10.9200	1,000
Promotion	118,276	0.2832	0	0.4506	0	1
# of Hotels	4,422	26.7472	31	8.3185	5	35
# of Clicks	4,422	1.1192	1	0.5956	1	16
# of Bookings	4,422	0.0692	0	0.2538	0	1

There are 4,422 consumers assigned to the random ranking system. On average, hotels charge about \$196 per night per room, and 28% of these rooms are discounted. On average, consumers observe about 27 hotels. I discuss clicks and bookings later in this subsection.

Table 3.3 illustrates summary statistics on consumer-hotel specific characteristics and consumer-specific characteristics for those individuals assigned to the Expedia ranking system.

Table 3.3: Consumer-Hotel/Consumer-Specific Characteristics (Expedia Ranking) Summary Statistics

	Observations	Mean	Median	Std	Min	Max
Price	181,319	188.8647	160	109.7452	17	999
Promotion	181,319	0.3308	0	0.4705	0	1
# of Hotels	6,544	27.7077	32	8.1502	5	36
# of Clicks	6,544	1.1348	1	0.6550	1	15
# of Bookings	6,544	0.8889	1	0.3143	0	1

There are approximately 6,544 consumers assigned to the Expedia ranking system. On average, hotels charge around \$189 per night per room, and about a third of the rooms are discounted. Consumers observe an average of 28 hotels. The key difference between the two descriptive statistics is that hotels in the Expedia ranking system, on average, are less expensive and more likely to be discounted than those in the random ranking system. Moreover, consumers assigned to the Expedia ranking system click on more hotels than those assigned to the random ranking system. This suggests that the Expedia ranking system indeed reflects consumer utility. I also observe very different figures for the number of reservations. This difference is explained below.

The data have two limitations. First, the data include consumers who click on one or more hotels, as shown in Tables 3.2 and 3.3. I overcome this limitation by defining the market as the set of consumers who decide to travel. Every traveler needs accommodation and searches for hotels on Expedia so it is not unusual that every traveler clicks on at least one hotel. Second, the data are sampled using stratified random sampling. The stratum used is whether a consumer books a hotel. As shown in Tables 3.2 and 3.3, about 0.07% of the consumers assigned to the random ranking system made a reservation, and about 0.89% of the consumers assigned to the Expedia ranking system made a reservation. Expedia determines these two numbers. However, the weighted maximum likelihood estimator can be used to consistently estimate the parameters of interest (Manski & Lerman, 1977; Wooldridge, 2001).

3.3 Model

In this section, I develop a structural model to quantify the welfare costs of information asymmetry. In this model, there are two agents, a consumer and an OTA, and two time periods. The OTA first designs its ranking algorithm and commits to it before the consumer enters the market. The OTA may abide by its commitment to preserve its reputation (Mathevet, Pearce, & Stacchetti, 2019). The consumer then enters the market and makes a choice. In what follows, I introduce notation and describe each agent's problem backward in time.

In this chapter, subscript i refers to consumer i and subscript h refers to hotel h. Let $\mathcal{H}_i$ be a choice set for consumer i. I use x_i to denote a vector of generic variables indexed by both i and h: $x_i = (x_{ih} : h \in \mathcal{H}_i)$. I use ξ to denote a vector of generic variables indexed by h alone: $\xi = (\xi_h : h \in \mathcal{H}_i)$. For generic function $f_h(\cdot)$ indexed by h, I use $f(x_i)$ to denote a vector of the functions' values at x_i: $f(x_i) = (f_h(x_i) : h \in \mathcal{H}_i)$. I use bolded letter $\boldsymbol{x}_{ih}$ to denote a random variable or vector and lowercase letter x_{ih} to denote its realization.

3.3.1 Consumer's Choice

This subsection describes consumer behavior. I start by providing consumer utility and then describe information available to the consumer when entering the market. Finally, I describe consumer choice behavior.

The consumer gains utility by staying in hotel h:

$$\boldsymbol{u}_{ih} = \boldsymbol{\nu}_{ih} + \boldsymbol{\xi}_h + \boldsymbol{\varepsilon}_{ih}, \tag{3.1}$$

$$\boldsymbol{u}_{i0} = \boldsymbol{\varepsilon}_{i0}.$$

where $\boldsymbol{\nu}_{ih}$ is the consumer's valuation of hotel h whose realization is known to both the consumer and the OTA. $\boldsymbol{\varepsilon}_{ih}$ is the consumer's taste shock, and its realization is known only to the consumer. $\boldsymbol{u}_{i0}$ stands for utility from the outside option. I assume that $\boldsymbol{\varepsilon}_{ih}$ is distributed Gumbel with zero location and one scale parameters and independent of $\boldsymbol{\varepsilon}_{ih'}$ for $h \neq h'$ and the other variables introduced later. $\boldsymbol{\xi}_h$ is hotel h's true quality, and its realization is known only to the OTA at the time that the consumer books a hotel. This is because the consumer must stay at hotel h to realize it. However, I assume she knows its distribution. The parametric assumption for $\boldsymbol{\xi}_h$ is discussed in more detail later in this section.

I discuss the information available to the consumer when making a decision. When the consumer enters the market, the OTA provides the consumer with a set of hotels, $\mathcal{H}_i$, and relevant information. This relevant information can be divided into three parts. The first part is ranking $r_i = (h_{i1}, \cdots, h_{iH_i})$. h_{iR} represents the hotel positioned Rth from the top in ranking r_i. I define r_{ih} as the position of hotel h from the top in ranking r_i. For instance, $r_{ik} = 2$ if and only if $h_{i2} = k$. The second part is consumer-hotel specific characteristics, including the price of hotel h, p_{ih}, and promotion indicator x_{ih}. The third part is a set of quality proxies, w_h, containing a location score, a star rating, an average review score, and a brand indicator. These proxies do not directly enter the utility function (3.1). This is because if the consumer knew the true quality, then any additional information about the true quality would not give her more utility. In addition to observing the information provided by the OTA, the consumer is aware of her taste shock, ε_{ih}. Again, the consumer does not observe the realization of $\boldsymbol{\xi}_h$, but she knows its distribution.

The consumer cannot maximize her utility (3.1) for two reasons. The first reason is that, as mentioned earlier, she does not know the realization of true quality, $\boldsymbol{\xi}_h$. However, she knows that $\boldsymbol{\xi}_h$ is normally distributed with mean $w_h^{\intercal}\mu$ and variance σ^2. The second reason why the consumer cannot maximize her utility is that the consumer's attention to hotel h is affected by hotel h's position in ranking r_i (Chen & Yao, 2017; Ursu, 2018). For instance, consider two hotels, A and B. Assume A is at the top of the ranking and B is at the bottom. The consumer will pay more attention to A than B because A is more visible to the consumer than B. Therefore, her choice behavior is governed by her expected utility plus the attention effect from ranking r_i.

Before addressing consumer choice behavior, I discuss how the consumer infers the true quality using the given information. When inferring the true quality, the consumer relies only on the quality proxies and the known distribution. In other words, she does not infer the true quality from p_i and r_i. I assume that $\boldsymbol{\xi}_h$ and $\boldsymbol{p}_{ih}$ are independent for four reasons. First, Ursu (2018), who uses the same dataset, makes a similar argument. She shows that the price variation can be mostly explained by observable characteristics; in particular, the travel date of the consumer explains the majority of it. Furthermore, through an interview with a hotelier and from the past literature (Einav, Kuchler, Levin, & Sundaresan, 2015; Koulayev, 2014), she learns that the remaining variation might be caused by supply and field experiments – that is, the remaining variation is irrelevant to the demand. Second, the consumer is unaware of each hotel's true quality at the time of booking; therefore, her choice probability does not depend on ξ_h. I assume that the same goes for consumers using other booking channels. Assuming that the hotel chooses a price to maximize its expected profit, its price does not depend on ξ_h. Furthermore, while price is updated regularly by the hotel, its intrinsic quality was determined when the hotel was constructed. Third, to avoid overly complicating the model, I assume the consumer is not sophisticated enough to understand the relationship between hotel price and true quality. This also means that hotel h has no incentive to use price as a quality signal. Lastly, De los Santos and Koulayev (2017) presents a supporting result. They show that the estimated price coefficients are similar with and without endogeneity controls. Consequently, I assume that $\boldsymbol{\xi}_h$ and $\boldsymbol{p}_{ih}$ are independent, and the consumer does not infer the true quality from p_{ih}.

Furthermore, based on Chen and Yao (2017) and Ursu (2018), I assume that the consumer is ignorant about how ranking r_i is generated. Therefore, it is impossible for her to infer the true quality from r_i.

Now I discuss consumer choice behavior. Instead of maximizing her utility (3.1), the consumer attempts to maximize her expected utility while her attention is influenced by ranking r_i:

$$\begin{aligned}\mathbb{E}[\boldsymbol{u}_{ih}|p_i, x_i, w, r_i, \varepsilon_{ih}] + \tau r_{ih} =& \mathbb{E}[\boldsymbol{\nu}_{ih}|p_i, x_i, w, r_i, \varepsilon_{ih}] + \mathbb{E}[\boldsymbol{\xi}_h|p_i, x_i, w, r_i, \varepsilon_{ih}] + \varepsilon_{ih} + \tau r_{ih} \\ =& \underbrace{\alpha p_{ih} + \beta x_{ih} + w_h^\intercal \mu + \varepsilon_{ih}}_{\text{Expected Utility}} + \underbrace{\tau r_{ih}}_{\text{Attention Effect}} \qquad (3.2) \\ \mathbb{E}[\boldsymbol{u}_{i0}|p_i, x_i, w, r_i, \varepsilon_{i0}] =& \varepsilon_{i0},\end{aligned}$$

where $\nu_{ih} = \alpha p_{ih} + \beta x_{ih}$, and α, β, μ, and τ are parameters.

The consumer chooses hotel h if

$$\mathbb{E}[\boldsymbol{u}_{ih}|p_i, x_i, w, r_i, \varepsilon_{ih}] + \tau r_{ih}\mathbb{1}(h \neq 0) \geq \max_{k \in \mathcal{H}_i \cup \{0\}} \mathbb{E}[\boldsymbol{u}_{ik}|p_i, x_i, w, r_i, \varepsilon_{ik}] + \tau r_{ik}\mathbb{1}(k \neq 0).$$

The probability that the consumer books hotel h is then

$$\Pr(\boldsymbol{d}_{ih} = 1|p_i, x_i, w, r_i) = \frac{\exp\left(\alpha p_{ih} + \beta x_{ih} + w_h^\intercal \mu + \tau r_{ih}\right)}{1 + \sum_{k \in \mathcal{H}_i} \exp\left(\alpha p_{ik} + \beta x_{ik} + w_k^\intercal \mu + \tau r_{ik}\right)}, \qquad (3.3)$$

where $\boldsymbol{d}_{ih}$ is an indicator of whether the consumer books hotel h – it takes a value of one if the consumer books the hotel, and zero otherwise. I call probability (3.3) the choice probability.

3.3.2 OTA's Ranking Decision

In this section, I describe how the OTA determines its ranking algorithm. The OTA chooses score function $S(\cdot)$ before observing $\boldsymbol{p}_i$, $\boldsymbol{x}_i$, $\boldsymbol{w}$, $\boldsymbol{\xi}$, and $\boldsymbol{\mathcal{H}}_i$. Here, I assume that the OTA does not know the realization of $\boldsymbol{\xi}$ and $\boldsymbol{w}$ because it does not know the consumer's intended destination. It is also possible that hotel h decides to list or remove its hotel rooms from the OTA's platform after the OTA determines its ranking algorithm.

Before discussing the OTA's objective function, I address how ranking r_i is generated from score function $S(\cdot)$. When the consumer enters the market, the OTA observes p_i, x_i, w, ξ, and $\mathcal{H}_i$. As mentioned in Section 3.2.1, I assume that the OTA uses potential profit cp_{ih} and consumer utility up to the OTA's knowledge $u^*_{ih} = \alpha p_{ih} + \beta x_{ih} + \xi_h$ to generate ranking r_i, where $c \in [0, 1]$ is the commission rate that the OTA charges hotels. Again, this assumption is supported by Expedia's statement that consumer preferences and compensation from hotels influence the ranking (Expedia, 2022). As a result, score function $S(\cdot)$ takes potential profit and consumer utility up to the OTA's knowledge as its arguments and maps them into a real scalar:

$$s_{ih} = S(cp_{ih}, u^*_{ih}).$$

Ranking r_i is generated in descending order of s_i. That is, if $s_{ih} > s_{ih'}$, then hotel h is positioned more highly than hotel h' in ranking r_i. I denote the function that maps p_i, x_i, ξ, and $\mathcal{H}_i$ into a permutation of $\mathcal{H}_i$ by $R(\cdot; S)$. That is,

$$r_i = R(p_i, x_i, \xi, \mathcal{H}_i; S)$$

where $R(\cdot; S)$ represents the dependence of ranking on score function S, and likewise for other functions. For instance, for function $f(\cdot)$, I use $f(\cdot; S)$ to denote its dependency on score function S.

Now I discuss the objective function that the OTA maximizes to select optimal score function $S(\cdot)$. However, I first introduce some notation. Let $F(\cdot)$ be the joint cumulative distribution function of $\boldsymbol{p}_i$, $\boldsymbol{x}_i$, $\boldsymbol{w}$, $\boldsymbol{\xi}$, and $\boldsymbol{\mathcal{H}}_i$ and

$$\begin{aligned}\Pr(\boldsymbol{d}_{ih} = 1 | p_i, x_i, w, r_i) =& \Pr(\boldsymbol{d}_{ih} = 1 | p_i, x_i, w, R(p_i, x_i, \xi, \mathcal{H}_i; S)) \\ =& \tilde{\Pr}(\boldsymbol{d}_{ih} = 1 | p_i, x_i, w, \xi, \mathcal{H}_i; S).\end{aligned}$$

Then score function $S(\cdot)$ is determined as follows:

$$S = \arg\max_{S^*} c \int \sum_{h\in\mathcal{H}_i} p_{ih} \tilde{\Pr}(\boldsymbol{d}_{ih} = 1 | p_i, x_i, w, \xi, \mathcal{H}_i; S^*) \,\mathrm{d}F(p_i, x_i, w, \xi, \mathcal{H}_i) \tag{3.4}$$

Now I discuss score function $S(\cdot)$. As any monotone transformation of $S(\cdot)$ will result in the same ranking, and thus the same choice probability, I assume that all monotone transformations of $S(\cdot)$ are equivalent. Furthermore, if there are more than two $S(\cdot)'s$ that maximize the ex-ante expected profit, I assume that the OTA chooses the function that is more favorable to the consumer for its future reputation. Suppose that there are two score functions, $S_1(\cdot)$ and $S_2(\cdot)$, that generate the same ex-ante expected profit. Let $S_u(\cdot, u^*_{ih}) = u^*_{ih}$, meaning that $S_u(\cdot)$ only reflects consumer utility. I argue that $S_1(\cdot)$ is more favorable to the consumer than $S_2(\cdot)$ if $\|S_1 - S_u\|_\infty < \|S_2 - S_u\|_\infty$. I define $\|S_1 - S_u\|_\infty$ as follows. Let $\mathcal{S}_1$ be the equivalence class of function $S_1(\cdot)$. Then,

$$\|S_1 - S_u\|_\infty = \inf_{S\in\mathcal{S}_1} \sup_{p_{ih}, u^*_{ih}} |S(cp_{ih}, u^*_{ih}) - u^*_{ih}|.$$

In this example, the OTA chooses score function $S_1(\cdot)$. I assume that score function $S(\cdot)$ is increasing in both arguments. I also assume that if the consumer utilities from two hotels are the same as far as the OTA knows, then the OTA will always assign the more profitable hotel an advantageous position in a ranking. Moreover, if the OTA stands to obtain the same profits from two hotels, the OTA will always put the consumer's preferred hotel in a good position in a ranking, increasing the overall choice probability.

3.4 Identification and Estimation

3.4.1 Identification

In this subsection, I discuss the identification of parameters $\theta = (\alpha, \beta, \mu, \tau)$ and σ. Once θ and σ are identified, $S(\cdot)$ can be directly calculated from Equation (3.4). θ can be easily identified from the variation in p_i, x_i, w, and r_i. Note that, as discussed in the previous sections, I assume that $\boldsymbol{p}_{ih}$ is exogenous in this chapter. $\boldsymbol{r}_i$ is also exogenous as it is a random permutation for those who are assigned to the random ranking system.

Now I discuss the identification of σ. To identify σ, I assume that the econometrician can access an infinite sample from the joint distribution of $\boldsymbol{d}_i$, $\boldsymbol{p}_i$, $\boldsymbol{x}_i$, $\boldsymbol{w}$, and $\boldsymbol{\mathcal{H}}_i$ to uncover $\check{\Pr}(\boldsymbol{d}_{ih} = 1 | p_i, x_i, w, \mathcal{H}_i)$, where

$$\check{\Pr}(\boldsymbol{d}_{ih} = 1 | p_i, x_i, w, \mathcal{H}_i) = \int \tilde{\Pr}(\boldsymbol{d}_{ih} = 1 | p_i, x_i, w, \xi, \mathcal{H}_i)\, \Phi(\xi | w^\intercal \mu, \sigma)$$

and $\Phi(\xi | w^\intercal \mu, \sigma)$ is the normal CDF of $\boldsymbol{\xi}$ with mean $w^\intercal \mu$ and variance σ^2. Let $\tilde{p}_i = (\tilde{p}_{i1}, \cdots, \tilde{p}_{iH_i})$ where $\tilde{p}_{ih} = \tilde{p}_{i1}$ for all h. Then, when I focus on $\tilde{p}_i$, S_σ becomes irrelevant for identification for the reason that I explain below. Recall that I assume $S(\cdot)$ is increasing in profit. If the prices of all the hotels are identical, then ranking r_i will be determined by consumer utility up to the OTA's knowledge $\tilde{u}^*_{ih} = \alpha \tilde{p}_{ih} + \beta x_{ih} + \xi_h$. That is, ranking r_i will be generated in descending order of $\tilde{u}^*_i$. Let $\boldsymbol{\xi}_h = w_h^\intercal \mu + \sigma \tilde{\boldsymbol{\xi}}_h$, where $\tilde{\boldsymbol{\xi}}_h$ is distributed standard normally.

$$\begin{aligned} \tilde{u}^*_{ih} > \tilde{u}^*_{ih'} &\Leftrightarrow \beta x_{ih} + w_h^\intercal \mu + \sigma \tilde{\xi}_h > \beta x_{ih'} + w_{h'}^\intercal \mu + \sigma \tilde{\xi}_{h'} \\ &\Leftrightarrow \frac{\beta}{\sigma} x_{ih} + w_h^\intercal \frac{\mu}{\sigma} + \tilde{\xi}_h > \frac{\beta}{\sigma} x_{ih'} + w_{h'}^\intercal \frac{\mu}{\sigma} + \tilde{\xi}_{h'}. \end{aligned} \tag{3.5}$$

I denote the function that maps x_i, w, $\tilde{\xi}$, and $\mathcal{H}_i$ into a permutation of $\mathcal{H}_i$ by $\tilde{R}(\cdot; \theta, \sigma)$. That is,

$$r_i = \tilde{R}(x_i, w, \tilde{\xi}, \mathcal{H}_i; \theta, \sigma).$$

Then,

$$\check{\Pr}(\boldsymbol{d}_{ih} = 1 | \tilde{p}_i, x_i, w, \mathcal{H}_i; \theta, \sigma) = \int \Pr(\boldsymbol{d}_{ih} = 1 | \tilde{p}_i, x_i, w, \tilde{R}(x_i, w, \tilde{\xi}; \theta, \sigma))\, \mathrm{d}\Phi(\tilde{\xi} | 0, 1).$$

Given θ is identified, it is easy to see that variations in x_i and/or w allow me to identify σ

from Equation (3.5).

3.4.2 Estimation

In this subsection, I describe methods to estimate model parameters θ and σ and calculate $S(\cdot)$. I estimate the parameters in two steps. First, I use a sample of consumers who are assigned to the random ranking system to estimate θ. To estimate θ, I use the maximum likelihood estimator. Second, I use a sample of consumers who are assigned to the Expedia ranking system to estimate σ. To estimate σ, I use the simulated maximum likelihood estimator. Using the estimated parameters, I calculate $S(\cdot)$. For the calculation, I use the simple linear specification for $S(\cdot)$: $S(cp_{ih}, u^*_{ih}) = cp_{ih} + \eta u^*_{ih}$. Here, η captures the weight of consumer utility in the OTA's optimal score function. Thus, the η that maximizes the OTA's ex-ante expected profit function is called the optimal weight. In this subsection, for any function $f(\cdot)$, I use $f(\cdot;\gamma)$ to show its dependence on γ, where γ could be θ, elements of θ, σ, and/or η.

First, I discuss the method to estimate θ. Let I^R denote a set of consumers who are assigned to the random ranking system. The likelihood function is

$$L(\theta^*) = \prod_{i \in I^R} \left[\prod_{h \in \mathcal{H}_i \cup \{0\}} \Pr(\boldsymbol{d}_{ih} = 1 | p_i, x_i, w, r_i; \theta^*)^{d_{ih}} \right]^{W_i^R},$$

where W_i^R is a weight to account for the stratified random sampling method. This weight (Wooldridge, 2001) is calculated by

$$W_i^R = \begin{cases} \frac{\text{The probability that consumer } i \text{ books a hotel}}{\text{The fraction of observations who book a hotel}} & \text{if } i \text{ books a hotel} \\ \frac{\text{The probability that consumer } i \text{ does not books a hotel}}{\text{The fraction of observations who do not book a hotel}} & \text{else} \end{cases}.$$

In this chapter, W_i^R takes the values of 0.03/0.07 when consumer i books a hotel and 0.97/0.93 when she does not book a hotel. In the case that consumer i books a hotel, the numerator is the population probability of consumers i booking a hotel, and the denominator is the sample probability of consumer i booking a hotel in the data. By using this weight, I can consistently estimate the parameters (Wooldridge, 2001). Note that I use 0.03 as the population probability of the consumer booking a hotel as reported by Ursu (2018). Using another dataset from Expedia that is not influenced by stratified sampling,Ursu (2018) computes the population probability conditional on consumers who click on at least one hotel.

Then I obtain the estimator $\hat{\theta}$ of the true parameter θ as follows:

$$\hat{\theta} = \arg\max_{\theta^*} \log L(\theta^*).$$

I now discuss the method to estimate σ. Let I^E denote the set of consumers who are assigned to the Expedia ranking system. To estimate σ, I begin with calculating η as a function of σ.

Given σ^*, I calculate

$$\eta_{\sigma^*} = \arg\max_{\eta^*} c \int \sum_{h \in \mathcal{H}_i} p_{ih} \tilde{\Pr}(\boldsymbol{d}_{ih} = 1 | p_i, x_i, w, \xi, \mathcal{H}_i; \eta^*, \hat{\theta}) \, \mathrm{d}F_n(p_i, x_i, w, \xi, \mathcal{H}_i; \hat{\mu}, \sigma^*),$$

where $F_n(p_i, x_i, w, \xi, \mathcal{H}_i; \hat{\mu}, \sigma^*) = \Phi(\xi | w^\intercal \hat{\mu}, \sigma^*) G_n(p_i, x_i, w, \mathcal{H}_i)$ and $G_n(\cdot)$ is the empirical distribution of $\boldsymbol{p}_i$, $\boldsymbol{x}_i$, $\boldsymbol{w}$, and $\boldsymbol{\mathcal{H}}_i$. I set c to be 0.18.[4] After calculating η_{σ^*}, I calculate

$$\hat{\tilde{\Pr}}(\boldsymbol{d}_{ih} = 1 | p_i, x_i, w, \mathcal{H}_i; \eta_{\sigma^*}, \hat{\theta}, \sigma^*) = \frac{1}{V} \sum_{v} \tilde{\Pr}(\boldsymbol{d}_{ih} = 1 | p_i, x_i, w, \xi^v, \mathcal{H}_i; \eta_{\sigma^*}, \hat{\theta}),$$

where V is the simulation size and v refers to the vth simulation draw. I let V equal 500. Here, ξ_h^v is the vth draw from a normal distribution with mean $w_h^\intercal \hat{\mu}$ and variance ${\sigma^*}^2$.

[4] I thank an anonymous hotelier working at a major hotel chain for revealing this information. He has worked for three different hotel chains, and he states that the commission rates are the same for all three chains.

Then,

$$\hat{\sigma} = \arg\max_{\sigma^*} \log \left[\prod_{i \in I^E} \left\{ \prod_{h \in \mathcal{H}_i \cup \{0\}} \hat{\hat{\Pr}}(\boldsymbol{d}_{ih} = 1 | p_i, x_i, w, \mathcal{H}_i; \eta_{\sigma^*}, \hat{\theta}, \sigma^*)^{d_{ih}} \right\}^{W_i^E} \right],$$

where W_i^E is a weight to account for the stratified random sampling method and is calculated similarly to W_i^R. W_i^E takes the values of 0.03/0.8889 when consumer i books a hotel and 0.97/0.1111 when she does not book a hotel.

Lastly, $\hat{\eta}$, the calculated value of the optimal weight, is obtained by $\eta_{\hat{\sigma}}$.

3.5 Results

3.5.1 Empirical Results

Table 3.4 shows the key results. The main parameter of interest is η. This parameter measures the relative impact of consumer utility and the platform's potential profit on ranking r_i. A one-unit increase in consumer utility has the same effect as a \$28.25 increase in the OTA's potential profit. Furthermore, in the linear specification of $S(\cdot)$, the price coefficient is $c + \eta\alpha \approx -0.015$. This result implies that the lower the room price of a hotel is, the more likely the hotel is to be positioned at the top of ranking r_i. This result is consistent with the descriptive statistics discussed in Section 3.2.2.

Estimates of α and τ help to determine the relative effect of ranking r_i and price p_{ih} on consumer behavior. A one-unit increase in r_{ih} has the same effect on consumer behavior as a price increase of \$6.69. This result somewhat agrees with the existing literature that a price increase between \$0.21 and \$35.15 has the same effect as a hotel dropping one place in ranking r_i (Ghose, Ipeirotis, & Li, 2012; Koulayev, 2014; Chen & Yao, 2017; De los Santos & Koulayev, 2017; Ursu, 2018).

Other parameters are estimated as expected. The coefficient of the promotion indicator β is significant and positive, suggesting that the consumer obtains higher utility by staying at a hotel that offers a discounted price. The coefficients for the quality proxies (i.e., brand indicator, location score, average review score, and star rating) are all positive. The coefficient of the brand indicator is not statistically significant, while the other coefficients are significant at least at the 10% level. The coefficient of the brand indicator shows that a branded hotel does not necessarily mean high quality. Yet the higher the location score, average review score, and star rating, the higher the quality. Finally, σ is estimated to be 0.7175. The result implies that there indeed exists information asymmetry between the OTA and the consumer.

Table 3.4: Estimation Result

	Param (Std. Err.)
α: Price	-0.0069*** (0.0012)
β: Promo	0.2792** (0.1291)
μ_1: Brand	0.0064 (0.1452)
μ_2: Location	0.3126*** (0.0691)
μ_3: Review	0.1567* (0.0939)
μ_4: Star Rating	0.4762*** (0.0988)
τ: Position	-0.0464*** (0.0063)
σ	0.7175*** (0.0014)
η	28.2537

Note: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

3.5.2 Counterfactual Analysis

This subsection quantifies the welfare cost incurred by the information asymmetry between the OTA and the consumer through a counterfactual analysis. To perform the counterfactual analysis, I assume that the consumer has complete information about the true hotel quality, ξ_h. Under this assumption, I first compute the OTA's new optimal weight, η. Second, I simulate consumer choices and measure consumer welfare.

To compute the new optimal score function for the OTA, I first derive the choice probability that fits the counterfactual assumption. Now that the consumer knows the true quality, ξ_h, the choice probability of hotel h changes to

$$\begin{aligned}\Pr{}^*(\boldsymbol{d}_{ih}=1|p_i,x_i,\xi,r_i;\hat{\theta}) =& \Pr{}^*(\boldsymbol{d}_{ih}=1|p_i,x_i,\xi,R(p_i,x_i,\xi,\mathcal{H}_i;\eta);\hat{\theta})\\ =& \tilde{\Pr}^*(\boldsymbol{d}_{ih}=1|p_i,x_i,\xi,\mathcal{H}_i;\eta,\hat{\theta}).\end{aligned}$$

Then I compute the OTA's new optimal weight, η^{cf}. Here the superscript cf represents the counterfactual:

$$\hat{\eta}^{cf} = \arg\max_{\eta^*} 0.18 \int \sum_{h\in\mathcal{H}_i} p_{ih}\tilde{\Pr}^*(\boldsymbol{d}_{ih}=1|p_i,x_i,\xi,\mathcal{H}_i;\eta^*,\hat{\theta})\, F_n(p_i,x_i,w,\xi,\mathcal{H}_i;\hat{\mu},\hat{\sigma}).$$

I take the following four steps to simulate consumer choices. First, I generate for each hotel V normal variables with mean $w_h^{\intercal}\hat{\mu}$ and variance $\hat{\sigma}^2$. Second, I generate V Gumbel variables for each pair of multiple consumers and multiple hotels. Third, I simulate ranking $r_i^v = R(p_i,x_i,\xi^v,\mathcal{H}_i;\hat{\eta}^{cf})$, where superscript v stands for the vth draw. Finally, I simulate the consumer's choice. Let $u_h^v = \alpha p_{ih} + \beta x_{ih} + \xi_h^v + \varepsilon_{ih}^v$ and $u_{i0}^v = \varepsilon_{i0}^v$. The consumer chooses hotel h if

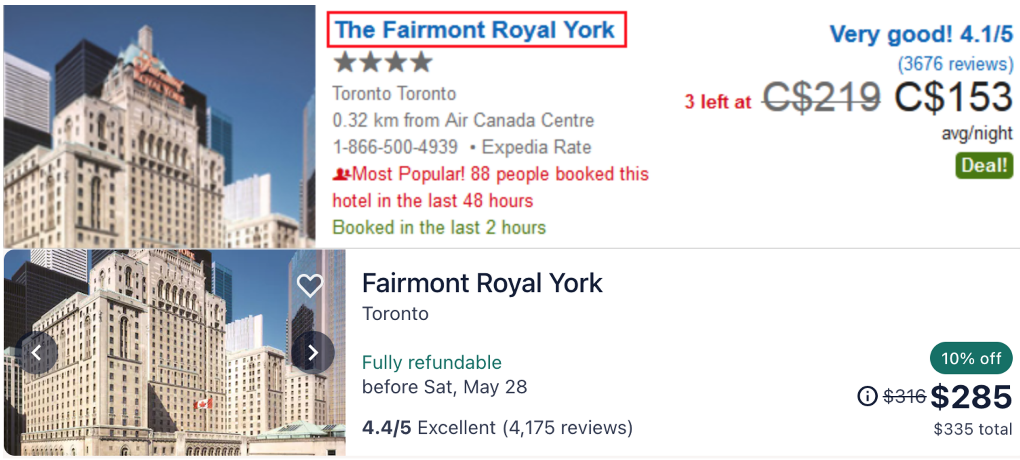
$$u_{ih}^v + \tau r_{ih}^v \mathbb{1}(h\neq 0) \geq \max_{k\in\mathcal{H}_i\cup\{0\}} u_{ik}^v + \tau r_{ik}^v \mathbb{1}(k\neq 0).$$

Consumer welfare is defined as consumer utility from the chosen hotel excluding the attention effect from ranking r_i. For each consumer and each draw, if the consumer chooses hotel h, then the consumer welfare is u_h^v. When reporting the welfare, I use the average value of consumer welfare and set V to 500.

Table 3.5 provides the results of the counterfactual analysis. The first row presents my model's result in the presence of information asymmetry. The second row shows the result in the absence of information asymmetry. The first column shows how the score function changes with the counterfactual assumption. If the consumer has complete information, the OTA redesigns its ranking algorithm such that it is more favorable to the consumer. The η value increases by approximately 34% (by about \$9.71). This result is intuitive and suggests that the OTA loses the power that comes from information asymmetry. In other words, the difference between the current and the initial η values (around \$9.71) quantifies the power of the OTA that arises from the fact that the OTA has richer information than the consumer.

Complete information offers two benefits to the consumer, especially those who have not previously booked. First, she is more likely to observe hotels that suit her needs because of the OTA's new optimal score function. Second, she now knows the true quality. These two benefits allow her to make an accurate decision. Consequently, the choice probability rises and consumer welfare improves by 2.24%. This growth equates to an average price decrease of \$2.1. This number is more substantial than it seems, considering that the probability of booking a hotel is 0.03. The increase in the overall choice probability also boosts the OTA's ex-ante expected profit by approximately 58%. Furthermore, total welfare (ex-ante expected profit + consumer welfare) rises by around 38%.

Figure 3.1: Expedia's Search Page Now & Then



†The first figure is a screenshot of the previous Expedia search page. This screenshot shows the star rating of the hotel and the location information. The second figure is an Expedia screenshot of the same hotel taken in 2022. There is no star rating or location information for the hotel in this screenshot.

Table 3.5: Counterfactual Result

	η	Consumer Welfare	Ex-Ante Expected Profit	Total
Information Asymmetry	28.2537	0.6476	1.1183	1.7636
Full Information	37.9646	0.6621	1.7709	2.5010
% Change	34.37%	2.24%	58.36%	37.78%

The result of this counterfactual analysis has important policy implications. If the OTA provided relevant hotel information (in addition to prices, reviews, and star ratings) so the consumer could infer true quality as much as possible, the welfare of all parties would increase. This implication is especially important because Expedia takes a somewhat different approach. For example, Expedia hides hotel star ratings and other important product information at the search level, as shown in Figure (3.1). Based on my results, this practice will undermine consumer welfare as consumers will have less information. Consequently, the platform will have more power over the consumer, meaning that it can manipulate the ranking algorithm to increase its ex-ante expected profit, which may harm the consumer.

3.6 Conclusion

In this chapter, I studied the welfare cost caused by information asymmetry between a consumer and a platform when the platform endogenously determines the ranking algorithm in E-commerce. To quantify the welfare cost, I developed a structural model in which the platform determines and commits to a ranking algorithm before any consumer enters the market. It is assumed that only the platform has intrinsic product quality information and that the platform determines the amount of information to be conveyed through the ranking. The consumer then enters the market and makes a choice. My results show that the platform exploits consumer ignorance. The counterfactual analysis considers a scenario in which the consumer has complete information. Once the consumer has complete information, the platform redesigns the ranking algorithm in favor of the consumer. As a result, consumer welfare improves by 2.24%. Surprisingly, the platform's profit also increases by about 58%. This increase results from the fact that the consumer is more likely to purchase a product when she has complete information.

References

Abaluck, J., & Adams, A. (2017). *What do consumers consider before they choose? identification from asymmetric demand responses* (Tech. Rep.). National Bureau of Economic Research.

Ackerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review*, *44*(3), 1007–1040.

Aït-Sahalia, Y., Bickel, P. J., & Stoker, T. M. (2001). Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics*, *105*(2), 363–412.

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological bulletin*, *84*(5), 888.

Ajzen, I., & Fishbein, M. (2000). Attitudes and the attitude-behavior relation: Reasoned and automatic processes. *European review of social psychology*, *11*(1), 1–33.

Bai, J., Chen, M., Liu, J., & Xu, D. Y. (2020). *Search and information frictions on global e-commerce platforms: Evidence from aliexpress.* Unpublished.

Bailey, A. A. (2005). Consumer awareness and use of product review websites. *Journal of Interactive Advertising*, *6*(1), 68–81.

Barseghyan, L., Coughlin, M., Molinari, F., & Teitelbaum, J. (2019). *Heterogeneous choice sets and preferences.* Unpublished.

Barseghyan, L., Molinari, F., & Thirkettle, M. (2019a). *Discrete choice under risk with* *limited consideration.* Working Paper.

Barseghyan, L., Molinari, F., & Thirkettle, M. (2019b). Discrete choice under risk with limited consideration. *arXiv preprint arXiv:1902.06629*.

Berger, J., Draganska, M., & Simonson, I. (2007). The influence of product variety on brand perception and choice. *Marketing Science*, *26*(4), 460–472.

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

Brierley, S. (2005). *The advertising handbook.* Routledge.

Cao, Z., Qin, T., Liu, T.-Y., Tsai, M.-F., & Li, H. (2007). Learning to rank: from pairwise approach to listwise approach. In *Proceedings of the 24th international conference on machine learning* (pp. 129–136).

Chen, Y., & Yao, S. (2017). Sequential search with refinement: Model and application with click-stream data. *Management Science*, *63*(12), 4345–4365.

Christopher, W.-G. (2013). *The expedia myth.* Retrieved 2022-02-28, from https://www.forbes.com/sites/christopherwarrengash/2013/08/02/the-expedia-myth-2/?sh=72aad5f41886

Clark, C. R., Doraszelski, U., & Draganska, M. (2009). The effect of advertising on brand awareness and perceived quality: An empirical investigation using panel data. *Qme*, *7*(2), 207–236.

Compiani, G., & Kitamura, Y. (2016). *Using mixtures in econometric models: a brief review and some new results.* Oxford University Press Oxford, UK.

De los Santos, B., & Koulayev, S. (2017). Optimizing click-through in online rankings with endogenous search refinement. *Marketing Science*, *36*(4), 542–564.

Draganska, M., & Klapper, D. (2011). Choice set heterogeneity and the role of advertising: An analysis with micro and macro data. *Journal of Marketing Research*, *48*(4), 653–669.

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter?—an empirical investigation of panel data. *Decision support systems*, *45*(4), 1007–1016.

Einav, L., Kuchler, T., Levin, J., & Sundaresan, N. (2015). Assessing sale strategies in online markets using matched listings. *American Economic Journal: Microeconomics*, *7*(2), 215–47.

Expedia. (2012). *2012 annual report.* Retrieved 2022-02-28, from https://www.expediagroup.com/investors/financial-information/annual-reports/default.aspx

Expedia. (2013). *2013 annual report.* Retrieved 2022-02-28, from https://www.expediagroup.com/investors/financial-information/annual-reports/default.aspx

Expedia. (2022). *How we determine our sort order.* Retrieved 2022-02-28, from https://www.expedia.com/service/?langid=1033&mobileContact=false#/article/1936

Fang, L. (2019). *The effects of online review platforms on restaurant revenue, survival rate, consumer learning and welfare.* Unpublished.

Fazio, R. H., & Zanna, M. P. (1981). Direct experience and attitude-behavior consistency. In *Advances in experimental social psychology* (Vol. 14, pp. 161–202). Elsevier.

Fithian, W., Sun, D., & Taylor, J. (2014). Optimal inference after model selection. *arXiv preprint arXiv:1410.2597*.

Fox, J. T. (2007). Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *The RAND Journal of Economics*, *38*(4), 1002–1019.

Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Surviving social media overload: Predicting consumer footprints on product search engines. In *Workshop on information systems and economics (wise)* (pp. 1–33).

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica*, *76*(5), 1017–1074.

Greene, W. (2004). Fixed effects and bias due to the incidental parameters problem in the tobit model. *Econometric reviews*, *23*(2), 125–147.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of multivariate analysis*, *14*(1), 1–16.

Heiss, F., McFadden, D., Winter, J., Wuppermann, A., & Zhou, B. (2016). *Inattention and switching costs as sources of inertia in medicare part d* (Tech. Rep.). National Bureau of Economic Research.

Henry, M., Kitamura, Y., & Salanié, B. (2014). Partial identification of finite mixtures in econometric models. *Quantitative Economics*, *5*(1), 123–144.

Honka, E., Hortaçsu, A., & Vitorino, M. A. (2017). Advertising, consumer awareness, and choice: Evidence from the us banking industry. *The RAND Journal of Economics*, *48*(3), 611–646.

Horowitz, J. L. (2012). *Semiparametric methods in econometrics* (Vol. 131). Springer Science & Business Media.

Hortaçsu, A., Madanizadeh, S. A., & Puller, S. L. (2017). Power to choose? an analysis of consumer inertia in the residential electricity market. *American Economic Journal: Economic Policy*, *9*(4), 192–226.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics*, *58*(1-2), 71–120.

Ichimura, H., & Lee, L.-F. (1991). Semiparametric least squares estimation of multiple index models: single equation estimation. In *Nonparametric and semiparametric methods in econometrics and statistics: Proceedings of the fifth international symposium in economic theory and econometrics. cambridge* (pp. 3–49).

Kawaguchi, K., Uetake, K., & Watanabe, Y. (2016). Identifying consumer attention: A product-availability approach.

Kawaguchi, K., Uetake, K., & Watanabe, Y. (2018). *Designing context-based marketing: Product recommendations under time pressure.*

Kitamura, Y., & Laage, L. (2018). Nonparametric analysis of finite mixtures. *arXiv* *preprint arXiv:1811.02727*.

Koulayev, S. (2014). Search for differentiated products: identification and estimation. *The RAND Journal of Economics*, *45*(3), 553–575.

Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2013). Exact post-selection inference with the lasso. *arXiv preprint arXiv:1311.6238*, *354*, 355.

Leeb, H., & Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, *21*(1), 21–59.

Leeb, H., & Pötscher, B. M. (2006). Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics*, *34*(5), 2554–2591.

Leeb, H., & Pötscher, B. M. (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory*, *24*(2), 338–376.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics*, *97*(1), 145–177.

Lewbel, A. (2016). The identification zoo-meanings of identification in econometrics. *Forthcoming on Journal of Economic Literature*, *7*.

Lewis, G., & Zervas, G. (2016). The welfare impact of consumer reviews: A case study of the hotel industry. In *Industrial organization workshop (pennsylvania state university, 2016).*

Lu, L.-C., Chang, W.-P., & Chang, H.-H. (2014). Consumer attitudes toward blogger's sponsored recommendations and purchase intention: The effect of sponsorship type, product type, and brand awareness. *Computers in Human Behavior*, *34*, 258–266.

Mack, Y.-p., & Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *61*(3), 405–415.

Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, *8*(3), 229–254.

Manski, C. F. (1988). Identification of binary response models. *Journal of the American statistical Association*, *83*(403), 729–738.

Manski, C. F., & Lerman, S. R. (1977). The estimation of choice probabilities from choice based samples. *Econometrica: Journal of the Econometric Society*, 1977–1988.

Mathevet, L., Pearce, D., & Stacchetti, E. (2019). Reputation and information design. *Unpublished paper, New York University.[451]*.

Max, S. (2013). *A call to arms: How to shift market share from the otas to the hotel website.* Retrieved 2022-02-28, from `https://www.hospitalitynet.org/opinion/4060290.html`

Nchare, K. (2018). *Partial identification of preferences and counterfactuals in a discrete choice model with inattention.* Unpublished.

Newberry, P., & Zhou, X. (2016). Heterogeneous effects of online reputation for local and national retailers. *International Economic Review*.

Shocker, A. D., Ben-Akiva, M., Boccara, B., & Nedungadi, P. (1991). Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing letters*, *2*(3), 181–197.

Statista. (2015). *Online travel agencies (ota) market share in the united states in 2013.* Retrieved 2022-02-28, from `https://www.statista.com/statistics/460012/online-travel-agency-market-share-us/`

Statista. (2021). *Leading online sources where consumers worldwide search for products as of april 2021.* Retrieved 2022-02-28, from `https://www.statista.com/statistics/1034209/global-product-search-online-sources/`

Statista. (2022). *E-commerce worldwide - statistics & facts.* Retrieved 2022-02-28, from `https://www.statista.com/topics/871/online-shopping/#dossier-chapter1`

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, *37*(4), 530–552.

Van Nierop, E., Bronnenberg, B., Paap, R., Wedel, M., & Franses, P. H. (2010). Retrieving unobserved consideration sets from household panel data. *Journal of Marketing Research*, *47*(1), 63–74.

Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism management*, *30*(1), 123–127.

Wooldridge, J. M. (2001). Asymptotic properties of weighted m-estimators for standard stratified samples. *Econometric Theory*, *17*(2), 451–470.

Wright, P., & Barbour, F. (1977). *Phased decision strategies: Sequels to an initial screening.* Graduate School of Business, Stanford University.

Young, W. (1917). On multiple integration by parts and the second theorem of the mean. *Proceedings of the London Mathematical Society*, *2*(1), 273–293.

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing*, *74*(2), 133–148.

Appendix A

Omitted Proofs from Chapter 1

A.1 Proof of Theorem 1.1

Proof: Recall Equation (1.7), where j is arbitrary,

$$p(j|x_i, z_i) = \sum_{C \in \mathcal{C}_j} p(j|z_i, x_i, C)p(C|x_i).$$

Notice that by Assumption 1.3 (exclusion restriction) and construction, for $k \neq j$, $\boldsymbol{w}_{ik}$ only affects the choice probability if $k \in C$. Therefore, if I take the derivative of the choice probability with respect to $\{w_{ik}\}_{k \neq j}$, then all terms will vanish except for the case $C = \mathcal{J}$.

Thus, I take the derivative of the choice probability with respect to $\{w_{ik}\}_{k \neq j}$:

$$\frac{\partial^{J-1} p(j|x_i, z_i)}{\partial w_{i1} \cdots \partial w_{ij-1} \partial w_{ij+1} \cdots \partial w_{iJ}}$$
$$= (-1)^{J-1} \frac{\partial^{J-1} F_{\boldsymbol{e}_{ij}}(z_{ij}^{\intercal}\beta, (z_{ij} - z_{i1})^{\intercal}\beta, \cdots, (z_{ij} - z_{iJ})^{\intercal}\beta, x_{i1}^{\intercal}\gamma, \cdots, x_{iJ}^{\intercal}\gamma)}{\partial w_{i1} \cdots \partial w_{ij-1} \partial w_{ij+1} \cdots \partial w_{iJ}}.$$

The left hand side is well defined since the right hand side is well defined by Assumption

1.7. Let

$$\mathcal{W}_j = \{(t_1, \cdots, t_{j-1}, t_{j+1}, \cdots, t_J) : w_{i1} < t_1, \cdots, w_{ij-1} < t_{j-1}, w_{ij+1} < t_{j+1}, \cdots, w_{iJ} < t_J\}.$$

By Assumptions 1.4 (large support) and 1.5 (parameter normalization), the derivative can be evaluated at any point of w_{ik}. Hence,

$$\left.\frac{\partial^{J-1} p(j|x_i, z_i)}{\partial w_{i1} \cdots \partial w_{ij-1} \partial w_{ij+1} \cdots \partial w_{iJ}}\right|_{w_{i1}=t_1, \cdots w_{ij-1}=t_{j-1}, w_{ij+1}=t_{j+1}, \cdots, w_{iJ}=t_J}$$

can be integrated over $\mathcal{W}_j$. The integration yields

$$- F_{\boldsymbol{e}_{ij}}(z_{ij}^{\intercal}\beta, (z_{ij} - z_{i1})^{\intercal}\beta, \cdots, (z_{ij} - z_{iJ})^{\intercal}\beta, x_{i1}^{\intercal}\gamma, \cdots, x_{iJ}^{\intercal}\gamma). \tag{A.1}$$

For all $k \neq j$, I take integration over t_k from w_{ik} to ∞ because the coefficient for w_{ik} is 1 by Assumption 1.5, and when I increase w_{ik} to infinity, product k becomes more preferable to product j; and hence product j will never be chosen, implying that the choice probability of product j will becomes 0.

Since j is arbitrary, I can obtain the same result (A.1) for all $j \in \mathcal{J}$. Then, I sum Equation (A.1) over all $j \in \mathcal{J}$. That is,

$$\begin{aligned}
&-\sum_{j=0}^{J} F_{\boldsymbol{e}_{ij}}(z_{ij}^{\intercal}\beta, (z_{ij} - z_{i1})^{\intercal}\beta, \cdots, (z_{ij} - z_{iJ})^{\intercal}\beta, x_{i1}^{\intercal}\gamma, \cdots, x_{iJ}^{\intercal}\gamma) \\
&\qquad = -\sum_{j=0}^{J} p(j|z_i, \mathcal{J}) p(\mathcal{J}|x_i)
\end{aligned}$$

Notice that consumer i must choose one item from her consideration set. Therefore, the summation term must equal 1.

$$= -p(\mathcal{J}|x_i) = -F_{\boldsymbol{\eta}_i}(x_{i1}^{\intercal}\gamma, \cdots, x_{iJ}^{\intercal}\gamma),$$

where $F_{\boldsymbol{\eta}_i}(\cdot)$ is the joint CDF of $\boldsymbol{\eta}_i$ conditional on $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$. Suppose I fix x_{ik} for all $k \neq 1$, then $-F_{\boldsymbol{\eta}_i}(\cdot)$ becomes a single-index model with respect to x_{i1}. Hence, I can identify γ (Manski, 1988).

Given that γ is identified, I can identify β and $F_{\boldsymbol{e}_{ij}}(\cdot)$ as follows. Recall Equation (A.1)

$$-F_{\boldsymbol{e}_{ij}}(z_{ij}^{\intercal}\beta, (z_{ij} - z_{i1})^{\intercal}\beta, \cdots, (z_{ij} - z_{iJ})^{\intercal}\beta, x_{i1}^{\intercal}\gamma, \cdots, x_{iJ}^{\intercal}\gamma).$$

Let $x_{ij} = (x_{ij,1}, \cdots, x_{ij,k}, \cdots, x_{ij,L_1})^{\intercal}$. Choose any $k \in (2, \cdots, L_1)$ such that the support of $\boldsymbol{x}_{ij,k}$ contains at least 3 values. I assume that $\boldsymbol{x}_{ij,k}$ can take $x_{ij,k}^1$, $x_{ij,k}^2$ and $x_{ij,k}^3$ and fix them. Now I find $x_{ij,1}^1$ and $x_{ij,1}^2$ so that $x_{ij,1}^1 + \gamma_k x_{ij,k}^1 = x_{ij,1}^2 + \gamma_k x_{ij,k}^2$. I can always find such $x_{ij,1}^1$ and $x_{ij,1}^2$ because I know γ and by Assumption 1.4 (large support). Similarly, I find $x_{ij,1}^3$. Now, I find w_{ij}^1, w_{ij}^2, and w_{ij}^3 so that $w_{ij}^1 + \beta_2 x_{ij,1}^1 + \beta_{k+1} x_{ij,k}^1 = w_{ij}^2 + \beta_2 x_{ij,1}^2 + \beta_{k+1} x_{ij,k}^2$ and $w_{ij}^1 + \beta_2 x_{ij,1}^1 + \beta_{k+1} x_{ij,k}^1 = w_{ij}^3 + \beta_2 x_{ij,1}^3 + \beta_{k+1} x_{ij,k}^3$. I can always find such w_{ij}^1, w_{ij}^2, and w_{ij}^3 since the CDF is strictly increasing in w_{ij} by Assumption 1.7 and by Assumption 1.4 (large support). Now, I have two unknowns β_2 and β_{k+1} and two equations. Hence, I can identify β_2 and β_{k+1}. Since β_2 is identified, I can identify all other β's even if some of random variable $\boldsymbol{x}_{ij,k'}$ takes 2 values by following the same steps. Given that β and γ are identified, I can identify the entire joint CDF $F_{\boldsymbol{e}_{ij}}(\cdot)$ by Assumption 1.4 (large support) (Lewbel, 2016). □

A.2 Proof of Theorem 1.2

Proof: Consider product 1 without loss of generality. Recall Equation (1.7)

$$p(1|x_i, z_i) = \sum_{C \in \mathcal{C}_1} p(1|z_i, x_i, C) p(C|x_i).$$

Recall that by the model construction, if product $k \neq 1$ is not in consideration set C, then $\boldsymbol{z}_{ik}$ does not influence the choice probability of product 1 conditional on C. Let $\partial_\ell f(\cdot)$ be a partial derivative of $f(\cdot)$ with respect to the ℓth argument. Then,

$$\frac{\partial^{J-1} p(1|x_i, z_i)}{\partial w_{i2}^2 \cdots \partial w_{iJ}^2} = (-1)^{J-1} \partial_{2 \cdots J} F_{\boldsymbol{e}_{i1}}(z_{i1}^\intercal \beta, (z_{i1} - z_{i2})^\intercal \beta, \cdots, (z_{i1} - z_{iJ})^\intercal \beta, x_{i1}^\intercal \gamma, \cdots, x_{iJ}^\intercal \gamma), \tag{A.2}$$

The left hand side of Equation (A.2) depends only on data. By Assumptions 1.5–1.12, β and γ can be identified (Ichimura & Lee, 1991).

To identify the joint density $f_{\boldsymbol{e}_{ij}}(\cdot)$. I define

$$\Psi_1(\cdot) = (-1)^{J-1} \partial_{2 \cdots J} F_{\boldsymbol{e}_{i1}}(\cdot)$$

to subdue partial derivative notation. Then,

$$\frac{\partial \Psi_1(z_{i1}^\intercal \beta, (z_{i1} - z_{i2})^\intercal \beta, \cdots, (z_{i1} - z_{iJ})^\intercal \beta, x_{i1}^\intercal \gamma, \cdots, x_{iJ}^\intercal \gamma)}{\partial w_{i1}^2} = \sum_{\ell=1}^{J} \partial_\ell \Psi_1(\cdot)$$

and for all $k \neq 1$

$$\frac{\partial \Psi_1(z_{i1}^\intercal \beta, (z_{i1} - z_{i2})^\intercal \beta, \cdots, (z_{i1} - z_{iJ})^\intercal \beta, x_{i1}^\intercal \gamma, \cdots, x_{iJ}^\intercal \gamma)}{\partial w_{ik}^2} = -\partial_k \Psi_1(\cdot)$$

The sum of the two equations above results in $(-1)^{J-1} \partial_{1 \cdots J} F_{\boldsymbol{e}_{i1}}(\cdot)$. Then, I take the partial derivative of $(-1)^{J-1} \partial_{1 \cdots J} F_{\boldsymbol{e}_{i1}}(\cdot)$ with respect to $\{w_{ik}^1\}_{k=1}^J$. That is,

$$(-1)^{J-1} \frac{\partial^J \partial_{1 \cdots J} F_{\boldsymbol{e}_{i1}}(\cdot)}{\partial w_{i1}^1 \cdots \partial w_{iJ}^1} = (-1)^{J-1} f_{\boldsymbol{e}_{i1}}(\cdot).$$

As a result, given that β and γ are identified, I can identify the joint density function on $\mathcal{XZ}$. $\square$

Appendix B

Omitted Proofs from Chapter 2

B.1 Omitted Proofs from Section 2.3

Proof of Lemma 2.1: I first prove ($\Rightarrow$). Suppose that for some $w_2 \in \mathcal{W}_2$, $p(012|w_1, w_2, z)$ is not constant in z. That is, for some $z \in \mathcal{Z}$, $\nabla_z p(012|w_1, w_2, z) \neq 0$, where ∇_z represents gradient with respect to z. By Assumption 2.8, $p(012|w, z)$ is not additively and multicatively separable. Hence, for some z, $\partial_{w_2} \nabla_z p(012|w_1, w_2, z) \neq 0$.

Recall ψ,

$$\psi(1, w_1, w_2, w_2', z) = \frac{\Delta_{w_2'}^{w_2} p(012|w_1, \cdot, z)}{\Delta_{\underline{w}_2}^{\bar{w}_2} p(012|w_1, \cdot, z)}.$$

I take the derivative of ψ with respect to z_1. Since ψ is invariant to z by Assumption, I must have

$$\partial_{z_1} \Delta_{w_2'}^{w_2} p(012|w_1, \cdot, z) \Delta_{\underline{w}_2}^{\bar{w}_2} p(012|w_1, \cdot, z) = \partial_{z_1} \Delta_{\underline{w}_2}^{\bar{w}_2} p(012|w_1, \cdot, z) \Delta_{w_2'}^{w_2} p(012|w_1, \cdot, z). \quad \text{(B.1)}$$

Note that Equation (B.1) must hold for all w_1, z and any choice of w_2, w_2'. I take the derivative of Equation (B.1) with respect to w_2 and w_2' separately.

$$w_2 : \partial_{w_2}\partial_{z_1}p(012|w_1,w_2,z)\Delta_{\underline{w}_2}^{\bar{w}_2}p(012|w_1,\cdot,z) = \partial_{z_1}\Delta_{\underline{w}_2}^{\bar{w}_2}p(012|w_1,\cdot,z)\partial_{w_2}p(012|w_1,w_2,z),$$
$$w_2' : \partial_{w_2'}\partial_{z_1}p(012|w_1,w_2',z)\Delta_{\underline{w}_2}^{\bar{w}_2}p(012|w_1,\cdot,z) = \partial_{z_1}\Delta_{\underline{w}_2}^{\bar{w}_2}p(012|w_1,\cdot,z)\partial_{w_2'}p(012|w_1,w_2',z)$$

Rearrange the above two equations, then I obtain

$$\frac{\partial_{w_2}\partial_{z_1}p(012|w_1,w_2,z)}{\partial_{w_2}p(012|w_1,w_2,z)} = \frac{\partial_{w_2'}\partial_{z_1}p(012|w_1,w_2',z)}{\partial_{w_2'}p(012|w_1,w_2',z)}$$

Indefinite integral results in

$$\begin{aligned}\ln[\partial_{w_2'}p(012|w_1,w_2,z)] =& \ln[\partial_{w_2'}p(012|w_1,w_2',z)] \\ &+ C(w_2,w_2').\end{aligned}$$

Take exponential on both sides

$$\partial_{w_2'}p(012|w_1,w_2,z) = c(w_2,w_2')\partial_{w_2'}p(012|w_1,w_2',z), \tag{B.2}$$

where $c(w_2,w_2') = \exp(C(w_2,w_2'))$. Equation (B.2) must hold for all z, w_1 and any value of w_2 and w_2'. This violates Assumption 2.8. Thus the result follows. I can do the same for $j = 2$. Also, the reverse way ($\Leftarrow$) is obvious. □

Proof of Theorem 2.1: Theorem 2.1 will be proved with following Lemmas B1–B7.

Steps of the proof: I expand my test statistic $\Gamma(\hat{f},\hat{F}_{\boldsymbol{x}})$ as

$$\Gamma(\hat{f},\hat{F}_{\boldsymbol{x}}) = \underbrace{\Gamma(\hat{f},\hat{F}_{\boldsymbol{x}}) - \Gamma(\hat{f},F_{\boldsymbol{x}})}_{\text{(i)}} + \underbrace{\Gamma(\hat{f},F_{\boldsymbol{x}}) - \Gamma(f,F_{\boldsymbol{x}})}_{\text{(ii)}}.$$

Note that $\Gamma(f, F_{\boldsymbol{x}}) = 0$ under the null hypothesis. In Lemma B6, I show that term (i) = $o_p(n^{-1}h^{-\delta/2})$. Let $\Gamma(f, F_{\boldsymbol{x}}) = \sum_j \int \phi(f_j)^2\, dF_{\boldsymbol{x}}$ with $\phi = 0$. In Lemmas B1 and B2, I show that term (ii) can be expanded as

$$L(f, \hat{f} - f, F_{\boldsymbol{x}}) + \sum_j \int \underbrace{\phi(f_j)(\hat{f}_j - f_j)\nabla\nabla\phi(f_j)(\hat{f}_j - f_j)^{\intercal}}_{\text{(iii)}} + \underbrace{\{\nabla\phi(f_j)(\hat{f}_j - f_j)^{\intercal}\}^2}_{\text{(iv)}}\, dF_{\boldsymbol{x}}$$
$$+ \underbrace{R(f, \hat{f} - f, F_{\boldsymbol{x}})}_{\text{(v)}}.$$

It is immediate that term (iii)=0 by $\phi = 0$. Likewise, ϕ enters $L(\cdot)$ linearly, thus $L(\cdot) = 0$. In Lemmas B3 and B4, I show that the term with term (iv) is asymptotically normally distributed after multiplying it by $nh^{\delta/2}$. Its bias term is provided in Lemma B5. Lastly, in Lemma B7, I show that term (v) = $o_p(n^{-1}h^{-\delta/2})$.

Lemma B1: Define $\Omega_{\boldsymbol{y}_j} = \{g_{\boldsymbol{y}_j} : \mathbb{R}^{1+\delta} \to \mathbb{R},\ g_{\boldsymbol{y}_j}$ is bounded, $\int g_{\boldsymbol{y}_j} = 0$, and $\|g_{\boldsymbol{y}_j}\|_\infty < c_1/2\}$ and $\Omega_{\boldsymbol{x}} = \{g_{\boldsymbol{x}} : \mathbb{R}^{\delta} \to \mathbb{R},\ g_{\boldsymbol{x}}$ is bounded, $\int g_{\boldsymbol{x}} = 0$, and $\|g_{\boldsymbol{x}}\|_\infty < c_1/2\}$, where $\|g\|_\infty = \sup_x \|g(x)\|$. Let $g = (g_{\boldsymbol{y}_1}, g_{\boldsymbol{y}_2}, g_{\boldsymbol{x}})$ belong to $\Omega_{\boldsymbol{y}_1} \times \Omega_{\boldsymbol{y}_2} \times \Omega_{\boldsymbol{x}}$ and $g_j = (g_{\boldsymbol{y}_j}, g_{\boldsymbol{x}})$ belong to $\Omega_{\boldsymbol{y}_j} \times \Omega_{\boldsymbol{x}}$ for $j = 1, 2$. Then, under Assumptions 2.1–2.8, 2.9–(ii), and H_0, $\Gamma(\cdot, F_{\boldsymbol{x}})$ has the following expansion:

$$\Gamma(f + g, F_{\boldsymbol{x}}) = \sum_{j=1,2} \int_{\mathcal{X}} \int_{\underline{w}_{j'}}^{w_{j'}} D_a(\psi\rho_{g_j})(j, \omega_j, z)^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x) + R(f + t^* g, F_{\boldsymbol{x}}),$$

where $t^* \in [0, 1]$, $\sup\{|R(f + t^* g, F_{\boldsymbol{x}})|/(\|g_{\boldsymbol{y}_1}\|_\infty^3 + \|g_{\boldsymbol{y}_2}\|_\infty^3 + \|g_{\boldsymbol{x}}\|_\infty^3) : (g_{\boldsymbol{y}_1}, g_{\boldsymbol{y}_2}, g_{\boldsymbol{x}}) \in \Omega_{\boldsymbol{y}_1} \times \Omega_{\boldsymbol{y}_2} \times \Omega_{\boldsymbol{x}}\} < \infty$ if $g \neq 0$ and $R(f + t^* g, F_{\boldsymbol{x}}) = 0$ if $g = 0$. I further define

$$\rho_{g_j}(j, \omega_j, z) = \left\{ \frac{\Delta_{w'_{j'}}^{w_{j'}} \kappa_{g_j}(j, w_j, \cdot, z)}{\Delta_{w'_{j'}}^{w_{j'}} p(j|w_j, \cdot, z)} - \frac{\Delta_{\underline{w}_{j'}}^{\bar{w}_{j'}} \kappa_{g_j}(j, w_j, \cdot, z)}{\Delta_{\underline{w}_{j'}}^{\bar{w}_{j'}} p(j|w_j, \cdot, z)} \right\}, \tag{B.3}$$

where

$$\kappa_{g_j}(j,x) = p(j|x)\left\{\frac{g_{\boldsymbol{y}_j}(1,x)}{f_{\boldsymbol{y}_j}(1,x)} - \frac{g_{\boldsymbol{x}}(x)}{f_{\boldsymbol{x}}(x)}\right\}.$$

Proof: From Equation (2.4), I have

$$\Gamma(f+g, F_{\boldsymbol{x}}) = \sum_{j=1,2}\int_{\mathcal{X}}\int_{\underline{w}_{j'}}^{w_{j'}} D_a(\psi(f_j+g_j))(j,\omega_j,z)^2 b(x)\,dw'_{j'}\,dF_{\boldsymbol{x}}(x).$$

I apply the functional expansion introduced by Aït-Sahalia et al. (2001). Define

$$\Psi(t) = \sum_{j=1,2}\int_{\mathcal{X}}\int_{\underline{w}_{j'}}^{w_{j'}} D_a(\psi(f_j+tg_j))(j,\omega_j,z)^2 b(x)\,dw'_{j'}\,dF_{\boldsymbol{x}}(x).$$

I have $(tg_{\boldsymbol{y}_1}, tg_{\boldsymbol{y}_2}, tg_{\boldsymbol{x}}) \in \Omega_{\boldsymbol{y}_1}\times\Omega_{\boldsymbol{y}_2}\times\Omega_{\boldsymbol{x}}$ for all $0\le t\le 1$. Now, Taylor expand the function $\Psi(t)$ about $t=0$:

$$\Psi(t) = \Psi(0) + t\Psi'(0) + t^2\Psi''(0)/2 + t^3\Psi'''(t^*)/6, \tag{B.4}$$

where $0 \le t^* \le t$. Note that $\Psi(0)=0$ under H_0. From now on, I ignore integration area for the sake of notation. Also I define

$$\tilde{p}(j,\omega_j,z;f_j+tg_j) = D_a(\psi(f_j+tg_j))(j,\omega_j,z). \tag{B.5}$$

$$\Psi'(t) = 2\sum\int\int \tilde{p}(j,\omega_j,z;f_j+tg_j)\frac{\partial\tilde{p}(j,\omega_j,z;f_j+tg_j)}{\partial t} b(x)\,dw'_{j'}\,dF_{\boldsymbol{x}}(x).$$

Since for all j, ω_j, z, $\tilde{p}(j, \omega_j, z; f_j) = 0$, under H_0 I have $\Psi'(0) = 0$.

$$\Psi''(t) = 2\sum \int \int \left[\tilde{p}(j, \omega_j, z; f_j + tg_j) \frac{\partial^2 \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t^2} + \left\{ \frac{\partial \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t} \right\}^2 \right] \times b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x).$$

Since $\tilde{p}(j, \omega_j, z; f_j) = 0$ for all j under H_0,

$$\Psi''(0) = 2\sum \int \int \left\{ \left. \frac{\partial \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t} \right|_{t=0} \right\}^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x).$$

I first demonstrate $\left. \frac{\partial \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t} \right|_{t=0}$

$$\left. \frac{\partial \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t} \right|_{t=0} = D_a(\psi \rho_{g_j})(j, \omega_j, z),$$

where

$$\rho_{g_j}(j, \omega_j, z) = \left\{ \frac{\Delta_{w'_{j'}}^{w_{j'}} \kappa_{g_j}(j, w_j, \cdot, z)}{\Delta_{w'_{j'}}^{w_{j'}} p(j|w_j, \cdot, z)} - \frac{\Delta_{\underline{w}_{j'}}^{\bar{w}_{j'}} \kappa_{g_j}(j, w_j, \cdot, z)}{\Delta_{\underline{w}_{j'}}^{\bar{w}_{j'}} p(j|w_j, \cdot, z)} \right\}$$

and

$$\kappa_{g_j}(j, x) = p(j|x) \left\{ \frac{g_{\boldsymbol{y}_j}(1, x)}{f_{\boldsymbol{y}_j}(1, x)} - \frac{g_{\boldsymbol{x}}(x)}{f_{\boldsymbol{x}}(x)} \right\}.$$

To characterize the remainder term,

$$\Psi'''(t) = 2\sum \int \int \left[\tilde{p}(j, \omega_j, z; f_j + tg_j) \frac{\partial^3 \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t^3} + 3 \frac{\partial \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t} \frac{\partial^2 \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t^2} \right] b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x).$$

From Equation (B.5) and the definition of $D_a(\psi)(\cdot)$, $\tilde{p}(\cdot; f_j+tg_j)$ is a sum of two $\psi(j, \omega_j, z; f_j+tg_j)$'s. I only need to focus on one of them to characterize the remainder term instead of $\tilde{p}(\cdot; f_j + tg_j)$. I first define notation

$$\tau_{1,g_j}(j,\omega_j,z,t) = \sum_{w^*_{j'}\in\{w_{j'},w'_{j'}\}} (-1)^{\mathbb{1}(w^*_{j'}=w'_{j'})} \frac{f_{\boldsymbol{y}_j}(1,w_j,w^*_{j'},z)g_{\boldsymbol{x}}(w_j,w^*_{j'},z) - f_{\boldsymbol{x}}(w_j,w^*_{j'},z)g_{\boldsymbol{y}_j}(1,w_j,w^*_{j'},z)}{\{f_{\boldsymbol{x}}(w_j,w^*_{j'},z)+tg_{\boldsymbol{x}}(w_j,w^*_{j'},z)\}^2} \tag{B.6}$$

$$\begin{aligned}\tau_{2,g_j}(j,\omega_j,z,t) =& 2\sum_{w^*_{j'}\in\{w_{j'},w'_{j'}\}} (-1)^{\mathbb{1}(w^*_{j'}=w'_{j'})} \left\{ \frac{g_{\boldsymbol{x}}(w_j,w^*_{j'},z)f_{\boldsymbol{y}_j}(1,w_j,w^*_{j'},z)g_{\boldsymbol{x}}(w_j,w^*_{j'},z)}{\{f_{\boldsymbol{x}}(w_j,w^*_{j'},z)+tg_{\boldsymbol{x}}(w_j,w^*_{j'},z)\}^3}\right.\\ &\left. - \frac{g_{\boldsymbol{x}}(w_j,w^*_{j'},z)f_{\boldsymbol{x}}(w_j,w^*_{j'},z)g_{\boldsymbol{y}_j}(1,w_j,w^*_{j'},z)}{\{f_{\boldsymbol{x}}(w_j,w^*_{j'},z)+tg_{\boldsymbol{x}}(w_j,w^*_{j'},z)\}^3}\right\}\end{aligned} \tag{B.7}$$

$$\begin{aligned}\tau_{3,g_j}(j,\omega_j,z,t) =& 6\sum_{w^*_{j'}\in\{w_{j'},w'_{j'}\}} (-1)^{\mathbb{1}(w^*_{j'}=w'_{j'})} \left\{ \frac{g_{\boldsymbol{x}}(w_j,w^*_{j'},z)^2f_{\boldsymbol{y}_j}(1,w_j,w^*_{j'},z)g_{\boldsymbol{x}}(w_j,w^*_{j'},z)}{\{f_{\boldsymbol{x}}(w_j,w^*_{j'},z)+tg_{\boldsymbol{x}}(w_j,w^*_{j'},z)\}^4}\right.\\ &\left. - \frac{g_{\boldsymbol{x}}(w_j,w^*_{j'},z)^2f_{\boldsymbol{x}}(w_j,w^*_{j'},z)g_{\boldsymbol{y}_j}(1,w_j,w^*_{j'},z)}{\{f_{\boldsymbol{x}}(w_j,w^*_{j'},z)+tg_{\boldsymbol{x}}(w_j,w^*_{j'},z)\}^4}\right\}.\end{aligned} \tag{B.8}$$

To simplify notation, I use $\psi(f_j + tg_j)$ to denote $\psi(j, \omega_j, z; f_j + tg_j)$.

$$\begin{aligned}\partial\psi(f_j+tg_j)/\partial t =& -\frac{\Delta^{w_{j'}}_{w'_{j'}}p(j|w_j,\cdot,z;f_j+tg_j)\tau_{1,g_j}(j,\bar{\omega}_j,z,t)}{\Delta^{\bar{w}_{j'}}_{\underline{w}_{j'}}p(j|w_j,\cdot,z;f_j+tg_j)^2} + \frac{\tau_{1,g_j}(j,\omega_j,z,t)}{\Delta^{\bar{w}_{j'}}_{\underline{w}_{j'}}p(j|w_j,\cdot,z;f_j+tg_j)}\\ =&\Phi_{tg_j}(j,\omega_j,z;\tau_{1,g_j})\end{aligned} \tag{B.9}$$

$$\partial^2\psi(f_j+tg_j)/\partial t^2 = -2\frac{\tau_{1,g_j}(j,\bar{\omega}_j,z,t)}{\Delta^{\bar{w}_{j'}}_{\underline{w}_{j'}}p(j|w_j,\cdot,z;f_j+tg_j)}\Phi_{tg_j}(j,\omega_j,z;\tau_{1,g_j}) + \Phi_{tg_j}(j,\omega_j,z;\tau_{2,g_j}) \tag{B.10}$$

$$\begin{aligned}\partial^3\psi(f_j+tg_j)/\partial t^3 =& 6\frac{\tau_{1,g_j}(j,\bar{\omega}_j,z,t)^2}{\Delta^{w_{j'}}_{w'_{j'}}p(j|w_j,\cdot,z;f_j+tg_j)^2}\Phi_{tg_j}(j,\omega_j,z;\tau_{1,g_j}) - \Phi_{tg_j}(j,\omega_j,z;\tau_{3,g_j})\\ &+3\frac{\tau_{1,g_j}(j,\omega_j,z,t)\tau_{2,g_j}(j,\bar{\omega}_j,z,t) - \tau_{1,g_j}(j,\bar{\omega}_j,z,t)\tau_{2,g_j}(j,\omega_j,z,t)}{\Delta^{\bar{w}_{j'}}_{\underline{w}_{j'}}p(j|w_j,\cdot,z;f_j+tg_j)^2}\\ &-6\frac{\tau_{1,g_j}(j,\bar{\omega}_j,z,t)}{\Delta^{\bar{w}_{j'}}_{\underline{w}_{j'}}p(j|w_j,\cdot,z;f_j+tg_j)}\Phi_{tg_j}(j,\omega_j,z;\tau_{2,g_j}).\end{aligned} \tag{B.11}$$

I argue there exists some constant C_2 such that $\psi(f_j + tg_j)$ is bounded by C_2, Equation (B.9) is bounded by $C_2(\|g_{\boldsymbol{y}_j}\|_\infty + \|g_{\boldsymbol{x}}\|_\infty)$, Equation (B.10) is bounded by $C_2(\|g_{\boldsymbol{y}_j}\|^2_\infty + \|g_{\boldsymbol{x}}\|^2_\infty)$, and Equation (B.11) is bounded by $C_2(\|g_{\boldsymbol{y}_j}\|^3_\infty + \|g_{\boldsymbol{x}}\|^3_\infty)$. It suffices to show that the denominator terms entering Equations (B.6)–(B.11) bounded below for any t

between 0 and 1. First, denominator terms from Equations (B.6)–(B.8) have a form of $f_{\boldsymbol{x}}(x)+tg_{\boldsymbol{x}}(x)$. I show that this bounded below so that Equations (B.6)–(B.8) are bounded above. $|f_{\boldsymbol{x}}+tg_{\boldsymbol{x}}| \geq |f_{\boldsymbol{x}}|-t|g_{\boldsymbol{x}}| \geq |f_{\boldsymbol{x}}|-|g_{\boldsymbol{x}}| \geq c_1/2$ by Assumption 2.9–(ii) and the definition of $\Omega_{\boldsymbol{x}}$. Hence, the denominator terms are bounded below. Second, denominator terms from Equations (B.9)–(B.11) are the difference between two choice probabilities. This difference is bounded below by Assumption 2.5 and H_0 and above by the definition of probability. Thus, I can find constant C_3 such that for all $t^* \in [0,1]$, $\Psi'''(t^*) \leq C_3 \sum(\|g_{\boldsymbol{y}_j}\|_\infty + \|g_{\boldsymbol{x}}\|_\infty)^3$. This shows that I have $\Psi'''(t^*) = O(\|g_{\boldsymbol{y}_1}\|_\infty^3 + \|g_{\boldsymbol{y}_2}\|_\infty^3 + \|g_{\boldsymbol{x}}\|_\infty^3)$. □

Lemma B2: If Assumptions 2.1–2.11 and H_0 are satisfied, then I have for any cdf $F_{\boldsymbol{x}}$,

$$\Gamma(\hat{f}, F_{\boldsymbol{x}}) = \sum_{j=1,2} \int_{\mathcal{X}} \int_{\underline{w}_{j'}}^{w_{j'}} D_a(\psi \rho_{\hat{f}_j})(j, \omega_j, z)^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x)$$
$$+ O_p\left(\|\hat{f}_{\boldsymbol{y}_1}(1,x) - f_{\boldsymbol{y}_1}(1,x)\|_\infty^3 + \|\hat{f}_{\boldsymbol{y}_2}(1,x) - f_{\boldsymbol{y}_2}(1,x)\|_\infty^3 + \|\hat{f}_{\boldsymbol{x}}(x) - f_{\boldsymbol{x}}(x)\|_\infty^3 \right).$$

Proof: To prove Lemma B2, I simply apply Lemma B1. Let

$$g_{\boldsymbol{y}_j} = \hat{f}_{\boldsymbol{y}_j} - f_{\boldsymbol{y}_j} \text{ and } g_{\boldsymbol{x}} = \hat{f}_{\boldsymbol{x}} - f_{\boldsymbol{x}}.$$

Now, I show that $\rho_{\hat{f}_j - f_j}(\cdot) = \rho_{\hat{f}_j}(\cdot)$. From Equation (B.3), the only part that depends on g_j is

$$\frac{g_{\boldsymbol{y}_j}(1,x)}{f_{\boldsymbol{y}_j}(1,x)} - \frac{g_{\boldsymbol{x}}(x)}{f_{\boldsymbol{x}}(x)}.$$

I simply replace $g_{\boldsymbol{y}_j}$ with $\hat{f}_{\boldsymbol{y}_j} - f_{\boldsymbol{y}_j}$ and $g_{\boldsymbol{x}}$ with $\hat{f}_{\boldsymbol{x}} - f_{\boldsymbol{x}}$. Then, the above equation becomes

$$\frac{\hat{f}_{\boldsymbol{y}_j}(1,x)}{f_{\boldsymbol{y}_j}(1,x)} - 1 - \frac{\hat{f}_{\boldsymbol{x}}(x)}{f_{\boldsymbol{x}}(x)} + 1 = \frac{\hat{f}_{\boldsymbol{y}_j}(1,x)}{f_{\boldsymbol{y}_j}(1,x)} - \frac{\hat{f}_{\boldsymbol{x}}(x)}{f_{\boldsymbol{x}}(x)}.$$

As a result, I have $\rho_{\hat{f}_j - f_j}(\cdot) = \rho_{\hat{f}_j}(\cdot)$.

Now, I only need to show that $g = (g_{\boldsymbol{y}_1}, g_{\boldsymbol{y}_2}, g_{\boldsymbol{x}})$ are in $\Omega_{\boldsymbol{y}_1} \times \Omega_{\boldsymbol{y}_2} \times \Omega_{\boldsymbol{x}}$. I can apply uniform consistency results by Mack and Silverman (1982). By Assumptions 2.9–2.11, I have

$$\sup_{w,z} |\hat{f}_{\boldsymbol{x}}(w,z) - f_{\boldsymbol{x}}(w,z)| = O_p\left(\left(\frac{nh^\delta}{\ln h^{-1}}\right)^{-1/2} + h^r\right) = o_p(1).$$

Also, for any $j = 1, 2$

$$\sup_{w,z} |\hat{f}_{\boldsymbol{y}_j}(1,w,z) - f_{\boldsymbol{y}_j}(1,w,z)| = O_p\left(\left(\frac{nh^\delta}{\ln h^{-1}}\right)^{-1/2} + h^r\right) = o_p(1).$$

Then, with probability approaching 1, $g_{\boldsymbol{y}_j}$ $j = 1, 2$ and $g_{\boldsymbol{x}}$ are smaller than $c_1/2$. Hence, g, defined above, will belong to $\Omega_{\boldsymbol{y}_1} \times \Omega_{\boldsymbol{y}_2} \times \Omega_{\boldsymbol{x}}$. □

I define new notation. Let

$$\begin{aligned} I_n &= \sum_{j=1,2} \int_{\mathcal{X}} \int_{\underline{w}_{j'}}^{w_{j'}} D_a(\psi \rho_{\hat{f}_j})(j, \omega_j, z)^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x) \\ &= \sum_{j=1,2} \int_{\mathcal{X}} \int_{\underline{w}_{j'}}^{w_{j'}} r_{\hat{f}_j}(j, \omega_j, z)^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x). \end{aligned}$$

I will suppress the integration area to simplify notation. Throughout the chapter, unless I specify the integration area, it remains fixed as defined above. Then,

$$\begin{aligned} I_n = &\sum \int \int \{r_{\hat{f}_j}(j, \omega_j, z) - \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x) \\ &+ 2 \sum \int \int \{r_{\hat{f}_j}(j, \omega_j, z) - \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\} \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)] b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x) \\ &+ \sum \int \int \{\mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x) \end{aligned}$$

and

$$\mathbb{E}[I_n] = \sum \int \int \mathbb{E}\{r_{\hat{f}_j}(j, \omega_j, z) - \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x)$$
$$+ \sum \int \int \{\mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x).$$

Hence,

$$I_n - \mathbb{E}[I_n] = \sum \int \int \{r_{\hat{f}_j}(j, \omega_j, z) - \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x)$$
$$- \sum \int \int \mathbb{E}\{r_{\hat{f}_j}(j, \omega_j, z) - \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x)$$
$$+ 2 \sum \int \int \{r_{\hat{f}_j}(j, \omega_j, z) - \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}\mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)] b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x).$$

Throughout the chapter, I denote $\boldsymbol{y}_j \in \mathbb{R}^{1+\delta}$, $\boldsymbol{s}_j = \tilde{\boldsymbol{y}}_j \in \mathbb{R}^{1+\delta}$, and $\boldsymbol{t}_j = \tilde{\tilde{\boldsymbol{y}}}_j \in \mathbb{R}^{1+\delta}$ and $\boldsymbol{s} = (\tilde{\boldsymbol{d}}_1, \tilde{\boldsymbol{d}}_2, \tilde{\boldsymbol{x}})$ and $\boldsymbol{t} = (\tilde{\tilde{\boldsymbol{d}}}_1, \tilde{\tilde{\boldsymbol{d}}}_2, \tilde{\tilde{\boldsymbol{x}}})$. I define

$$R^*(j, \omega_j, z, \boldsymbol{s}_j) = \frac{\Delta^{w_{j'}}_{w'_{j'}} \kappa_{\boldsymbol{y}_j}(j, w_j, \cdot, z, \boldsymbol{s}_j) - \Delta^{w_{j'}}_{w'_{j'}} \kappa_{\boldsymbol{x}}(j, w_j, \cdot, z, \tilde{\boldsymbol{x}})}{\Delta^{w_{j'}}_{w'_{j'}} p(j|w_j, \cdot, z)}, \tag{B.12}$$

where

$$\kappa_{\boldsymbol{y}_j}(j, x, \boldsymbol{s}_j) = p(j|x)\frac{\tilde{\boldsymbol{d}}_j K_h(\tilde{\boldsymbol{x}} - x)}{f_{\boldsymbol{y}_j}(1, x)} \quad \text{and} \quad \kappa_{\boldsymbol{x}}(j, x, \tilde{\boldsymbol{x}}) = p(j|x)\frac{K_h(\tilde{\boldsymbol{x}} - x)}{f_{\boldsymbol{x}}(x)}.$$

I further define

$$R(j, \omega_j, z, \boldsymbol{s}_j) = \psi(j, \omega_j, z)\{R^*(j, \omega_j, z, \boldsymbol{s}_j) - R^*(j, \bar{\omega}_j, z, \boldsymbol{s}_j)\}$$
$$- \int_{\mathcal{Z}} \psi(j, \omega_j, z)\{R^*(j, \omega_j, z, \boldsymbol{s}_j) - R^*(j, \bar{\omega}_j, z, \boldsymbol{s}_j)\} a(z')\, dz', \tag{B.13}$$

and

$$\tilde{R}(j, \omega_j, z, \boldsymbol{s}_j) = R(j, \omega_j, z, \boldsymbol{s}_j) - \mathbb{E}[R(j, \omega_j, z, \boldsymbol{s}_j)]. \tag{B.14}$$

Let $J_n(\boldsymbol{s}) = \sum \int \int \tilde{R}(j, \omega_j, z, \boldsymbol{s}_j) h^{-r} \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)] b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x)$ and $H_n(\boldsymbol{s}, \boldsymbol{t}) = h^{\delta/2} \sum \int \int \tilde{R}(j, \omega_j, z, \boldsymbol{s}_j) \tilde{R}(j, \omega_j, z, \boldsymbol{t}_j) b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x)$. Then,

$$\begin{aligned} I_n - \mathbb{E}[I_n] =& 2n^{-1/2} h^r \left\{ n^{-1/2} \sum_{i=1}^{n} J_n(y_i) \right\} \\ &+ 2n^{-1} h^{-\delta/2} \left\{ n^{-1} \sum_{1 \le i < k \le n} [H_n(y_i, y_k) - \mathbb{E}[H_n(\boldsymbol{y}_i, \boldsymbol{y}_k)]] \right\} \\ &+ n^{-1} h^{-\delta/2} \left\{ n^{-1} \sum_{i=1}^{n} [H_n(y_i, y_i) - \mathbb{E}[H_n(\boldsymbol{y}_i, \boldsymbol{y}_i)]] \right\} \\ =& 2n^{-1/2} h^r U_{n,1} + 2n^{-1} h^{-\delta/2} U_{n,2} + n^{-1} h^{-\delta/2} U_{n,3}. \end{aligned}$$

By CLT, it is easy to show $U_{n,3} = O_p(n^{-1/2} h^{-\delta/2}) = o_p(1)$. When I discuss $U_{n,1}$, I drop subscript i by i.i.d. assumption. Thus, I use notation $\boldsymbol{y} = (\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{x})$ and $\boldsymbol{y}_j = (\boldsymbol{d}_j, \boldsymbol{x})$. On the other hand, when I discuss $U_{n,2}$, I keep subscript i and k to make distinction between two random vectors. I use notation $\boldsymbol{y}_i = (\boldsymbol{d}_{i1}, \boldsymbol{d}_{i2}, \boldsymbol{x}_i)$ and $\boldsymbol{y}_{ij} = (\boldsymbol{d}_{ij}, \boldsymbol{x}_i)$, and likewise for k.

I will study asymptotic properties of $U_{n,1}$ and $U_{n,2}$ in Lemmas B3 and B4, respectively. When proving Lemmas B3 and B4, I will focus on only one term of $R(\cdot)$ which is

$$\gamma(j, \omega_j, z, \tilde{\boldsymbol{x}}) = \psi(j, \omega_j, z) \frac{\kappa_{\boldsymbol{x}}(j, w_j, w_{j'}, z, \tilde{\boldsymbol{x}})}{\Delta_{w'_{j'}}^{w_{j'}} p(j | w_j, \cdot, z)}$$

Since I denote $\boldsymbol{x} = (\boldsymbol{w}, \boldsymbol{z})$ and by the definition of $\psi(\cdot)$ and $\kappa_{\boldsymbol{x}}(\cdot)$,

$$= \frac{p(j|x)\left\{\frac{K_h(\tilde{\boldsymbol{x}}-x)}{f_{\boldsymbol{x}}(x)}\right\}}{\Delta_{\underline{w}_{j'}}^{\bar{w}_{j'}} p(j|w_j, \cdot, z)} = \xi(j,x)K_h(x - \tilde{\boldsymbol{x}}).$$

Further, I focus on $j = 1$ since $\xi(j,x)$ is bounded below and above regardless of j by Assumptions 2.5 and 2.9. I focus on $\gamma(\cdot)$ and $j = 1$ because, in the proofs of Lemmas B.3 and B.4, I study orders of $\mathbb{E}[J_n(\boldsymbol{y})^2]$, $\mathbb{E}[H_n(\boldsymbol{y}_i, \boldsymbol{y}_k)^2]$, $\mathbb{E}[H_n(\boldsymbol{y}_i, \boldsymbol{y}_k)^4]$, and $\mathbb{E}[G_n(\boldsymbol{y}_i, \boldsymbol{y}_k)^2]$, where $G_n(y_i, y_k) = \mathbb{E}[H_n(\boldsymbol{y}_i, y_i)H_n(\boldsymbol{y}_i, y_k)]$ as n goes to ∞, and their orders are determined by terms that only concern $\gamma(\cdot)$. Now, I show that my claim is true. First, note that since orders of the expectations are determined by kernels $K_h(\cdot)$ (since this is only function that depends on n), and non kernel functions such as $p(j|\cdot)$, $f_{\boldsymbol{y}_j}$, and $f_{\boldsymbol{x}}$ are bounded above and below, I only focus on kernels only. Then, from Equation (B.13), two terms related to kernels are, ignoring non-kernel functions,

$$\tilde{\boldsymbol{d}}_j K_h(\tilde{\boldsymbol{w}}_j - w_j)K_h(\tilde{\boldsymbol{w}}_{j'} - w_{j'})K_h(\tilde{\boldsymbol{z}} - z) \text{ and } K_h(\tilde{\boldsymbol{w}}_j - w_j)K_h(\tilde{\boldsymbol{w}}_{j'} - w_{j'})K_h(\tilde{\boldsymbol{z}} - z),$$

and other kernel objects can be defined by $w_{j'}$ being replaced with $w'_{j'}$, $\bar{w}_j$, and $\underline{w}_j$ and z with z'. Since $\tilde{\boldsymbol{d}}_j$ takes value 0 or 1, I have

$$|\tilde{\boldsymbol{d}}_j K_h(\tilde{\boldsymbol{w}}_j - w_j)K_h(\tilde{\boldsymbol{w}}_{j'} - w_{j'})K_h(\tilde{\boldsymbol{z}} - z)| \leq |K_h(\tilde{\boldsymbol{w}}_j - w_j)K_h(\tilde{\boldsymbol{w}}_{j'} - w_{j'})K_h(\tilde{\boldsymbol{z}} - z)|.$$

Hence, terms with $\tilde{\boldsymbol{d}}_j$ are dominated by terms without it.

Now, I argue that I only need to consider one $\gamma(1, \omega_j, z, \tilde{\boldsymbol{x}})$ to study the orders, not multiple γ's with different arguments. This is because γ's enter R linearly, and γ's with some constant arguments (other than choice variable 1, such as $\bar{w}$) could be dominated by γ's without constant arguments. This can be shown by comparing (i) $\mathbb{E}[\gamma(1, \bar{\omega}_1, z, \tilde{\boldsymbol{x}})\gamma(1, \bar{\omega}''_1, z'', \tilde{\boldsymbol{x}})]$ and (ii) $\mathbb{E}[\gamma(1, \omega_1, z, \tilde{\boldsymbol{x}})\gamma(1, \omega''_1, z'', \tilde{\boldsymbol{x}})]$. I look at expectation (i). I can find constant

C_4 that satisfies

$$\mathbb{E}[\gamma(1,\bar{\omega}_j,z,\tilde{\boldsymbol{x}})\gamma(1,\bar{\omega}_1'',z'',\tilde{\boldsymbol{x}})] \leq C_4 \int K_h(\tilde{x}-\bar{x})K_h(\tilde{x}-\underline{x})\, f_{\boldsymbol{x}}(\tilde{x})\, d\tilde{x}.$$

Let $u_1 = (\tilde{x}-\bar{x})/h$.

$$= C_4 \int K(u_1)K(u_1-(\underline{x}-\bar{x})/h)/h^{\delta}\, f_{\boldsymbol{x}}(\bar{x}+u_1 h)\, du_1.$$

Since $k(u) = o(1/u)$ by Assumption 2.10 and $f_{\boldsymbol{x}}$ is bounded, I can find constant C_5 so that

$$\leq C_5.$$

On the other hand, for expectation (ii), after a similar change of variables, I can find constant C_6 so that

$$\mathbb{E}[\gamma(1,\omega_1,z,\tilde{\boldsymbol{x}})\gamma(1,\omega_1'',z'',\tilde{\boldsymbol{x}})] \leq C_6 h^{-\delta}.$$

Thus, $\gamma(1,\omega_1,z,\tilde{\boldsymbol{x}})$ dominates other γ's with constant arguments. I focus on only one γ to show the orders of the expectations.

Lemma B3: If Assumptions 2.5, 2.9, and 2.10, and H_0 are satisfied, then $\mathbb{E}[J_n(\boldsymbol{y})] = 0$ and $\mathbb{E}[J_n^2(\boldsymbol{y})] < \infty$.

Proof: Recall that

$$\mathbb{E}[J_n(\boldsymbol{y})] = \sum \int \int \mathbb{E}[\tilde{R}(j,\omega_j,z,\boldsymbol{s}_j)]h^{-r}\mathbb{E}[r_{\hat{f}_j}(j,\omega_j,z)]b(x)\, dw_{j'}'\, dF_{\boldsymbol{x}}(x)$$

First, I show

$$h^{-r}\mathbb{E}[r_{\hat{f}_j}(j,\omega_j,z)] = h^{-r}\mathbb{E}[R(j,\omega_j,z,\boldsymbol{s}_j)] < \infty.$$

I focus on the first term of $R^*(\cdot)$ since the two terms in $R^*(\cdot)$ have the same form, and $R(\cdot)$

is sum of four $R^*(\cdot)$'s as shown in Equations (B.12) and (B.13). I denote $R_1^*(\cdot)$ to be the first term of $R^*(\cdot)$:

$$R_1^*(j,\omega_j,z,\boldsymbol{s}_j)=\frac{\kappa_{\boldsymbol{y}_j}(j,w_j,w_{j'},z,\boldsymbol{s}_j)-\kappa_{\boldsymbol{x}}(j,w_j,w_{j'},z,\tilde{\boldsymbol{x}})}{\Delta_{w'_{j'}}^{w_{j'}}p(j|w_j,\cdot,z)}.$$

I change variable to $u=(\tilde{x}-x)/h$ and after multiplying $\psi(\cdot)$, I have

$$\begin{aligned}&h^{-r}\mathbb{E}[\psi(j,\omega_j,z)R_1^*(j,\omega_j,z,\boldsymbol{s}_j)]\\&\quad=\xi(j,x)\sum_{|\alpha|=r}\frac{1}{\alpha!}\int u^\alpha K(u)\left\{\frac{f_{\boldsymbol{x}}(x)\partial^\alpha f_{\boldsymbol{y}_j}(1,x+uh^*)}{f_{\boldsymbol{y}_j}(1,x)}-\partial^\alpha f_{\boldsymbol{x}}(x+uh^*)\right\}du,\end{aligned}$$

where $|\alpha|=\alpha_1+\alpha_2+\cdots+\alpha_m$, $x^\alpha=x_1^{\alpha_1}x_2^{\alpha_2}\cdots x_m^{\alpha_m}$, and $\partial^\alpha f(x)=\partial^{|\alpha|}f(x)/\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}\cdots\partial x_m^{\alpha_m}$, and h^* is some value lying between 0 and h. Then, by Assumption 2.9–(ii), partial derivative of $f_{\boldsymbol{x}}$ is bounded and $\int u^r k(u)\,du<\infty$. Further, it holds true for all other $R^*(\cdot)$'s from $R(\cdot)$. Lastly, $\xi(j,x)$ is bounded below and above by Assumptions 2.5 and Assumption 2.9–(ii). Thus, by Assumptions 2.5, 2.9, and 2.10, I can find constant C_7

$$h^{-r}\mathbb{E}[R(j,\omega_j,z,\boldsymbol{s}_j)]\le C_7. \tag{B.15}$$

Now, $\mathbb{E}[\tilde{R}(j,\omega_j,z,\boldsymbol{s}_j)]=0$ by construction. This shows that $\mathbb{E}[J_n(\boldsymbol{y})]=0$.
Now, I show $\mathbb{E}[J_n(\boldsymbol{y})^2]<\infty$.

$$\begin{aligned}\mathbb{E}[J_n(\boldsymbol{y})^2]=&\sum_j\sum_l\int\int\int\int\mathbb{E}[\tilde{R}(j,\omega_j,z,\boldsymbol{s}_j)\tilde{R}(l,\omega_l'',z'',\boldsymbol{s}_l)]h^{-2r}\mathbb{E}[r_{\hat{f}_j}(j,\omega_j,z)]\mathbb{E}[r_{\hat{f}_l}(l,\omega_l'',z'')]\\&\times b(x)b(x'')\,dw'_{j'}dw'''_{l'}\,dF_{\boldsymbol{x}}(x)\,dF_{\boldsymbol{x}}(x'').\end{aligned}$$

I first need to study $\mathbb{E}[\tilde{R}(j,\omega_j,z,\boldsymbol{s}_j)\tilde{R}(l,\omega_l'',z'',\boldsymbol{s}_l)]$. Note that by Equation (B.15)

$$\mathbb{E}[\tilde{R}(j,\omega_j,z,\boldsymbol{s}_j)\tilde{R}(l,\omega_l'',z'',\boldsymbol{s}_l)]=\mathbb{E}[R(j,\omega_j,z,\boldsymbol{s}_j)R(l,\omega_l'',z'',\boldsymbol{s}_l)]+o(1).$$

Now, I will focus on the term only with $\gamma(\cdot)$ and fix $j = l = 1$ as discussed earlier. So, I will discuss

$$\mathbb{E}[\gamma(1, \omega_1, z, \tilde{\boldsymbol{x}})\gamma(1, \omega_1'', z'', \tilde{\boldsymbol{x}})] = \xi(1, x)\xi(1, x'')\mathbb{E}[K_h(\tilde{\boldsymbol{x}} - x)K_h(\tilde{\boldsymbol{x}} - x'')]$$

since $|\xi(\cdot)| < \infty$ by Assumptions 2.5 and 2.9–(ii), I only look at $\mathbb{E}[K_h(\tilde{\boldsymbol{x}} - x)K_h(\tilde{\boldsymbol{x}} - x''$.

$$\mathbb{E}[K_h(\tilde{\boldsymbol{x}} - x)K_h(\tilde{\boldsymbol{x}} - x'')] = \int K_h(\tilde{x} - x)K_h(\tilde{x} - x'')f_{\boldsymbol{x}}(\tilde{x})\, d\tilde{x}$$

change $u_1 = (\tilde{x} - x)/h$

$$= \int K(u_1)K(u_1 + (x - x'')/h)/h^{\delta} f_{\boldsymbol{x}}(x)\, du_1 + O(1) \tag{B.16}$$

Thus, by Assumptions 2.9 and 2.10, there is constant C_8 such that

$$\mathbb{E}[R(j, \omega_j, z, \boldsymbol{s}_j)R(j, \omega_j'', z'', \boldsymbol{s}_j)] \leq C_8(h^{-\delta}). \tag{B.17}$$

Since from $\mathbb{E}[J_n(\boldsymbol{y})^2]$, the terms only concerning γ's are dominant ones, it is sufficient to show that the following equation is bounded.

$$\int\int\int\int \mathbb{E}[\gamma(1, \omega_1, z, \boldsymbol{s}_1)\gamma(1, \omega_1'', z'', \boldsymbol{s}_1)]h^{-2r}\mathbb{E}[r_{\hat{f}_1}(1, \omega_1, z)]\mathbb{E}[r_{\hat{f}_1}(1, \omega_1'', z'')]$$
$$\times\, b(x)b(x'')\, dw_2' dw_2'''\, dF_{\boldsymbol{x}}(x)\, dF_{\boldsymbol{x}}(x'').$$

By Equation (B.15),

$$\leq C_7^2 \int\int\int\int \mathbb{E}[\gamma(1, \omega_1, z, \boldsymbol{s}_1)\gamma(1, \omega_1'', z'', \boldsymbol{s}_1)]b(x)b(x'')\, dw_2' dw_2'''\, dF_{\boldsymbol{x}}(x)\, dF_{\boldsymbol{x}}(x'').$$

Plugging in Equation (B.16), then

$$=C_7^2 \int\int\int\int \xi(1,\omega_1)\xi(1,\omega_1'') \int K(u_1)K(u_1+(x-x'')/h)/h^\delta f_{\boldsymbol{x}}(x)\, du_1 \\ \times b(x)b(x'')\, dw_2' dw_2''' \, dF_{\boldsymbol{x}}(x)\, dF_{\boldsymbol{x}}(x'') + O(1).$$

Then I change variable $u_2 = (x - x'')/h$, it yields that the above equation is bounded. Since dominating terms from $\mathbb{E}[J_n(\boldsymbol{y})^2]$ is bounded, I have $\mathbb{E}[J_n(\boldsymbol{y})^2] < \infty$. □

By Lemma B3, Assumption 2.9–(i), and CLT, $U_{n,1} = O_p(1)$.

Lemma B4: If Assumptions 2.5, 2.9–2.11, and H_0 hold, then $U_{n,2}/\sqrt{\sigma^2} \to \mathcal{N}(0, 1/2)$, where $\sigma^2 = \mathbb{E}[H_n(\boldsymbol{y}_i, \boldsymbol{y}_k)^2]$.

Proof: I will prove this lemma using Theorem 2.1 by Hall (1984). I need to show that for each n (i) $\mathcal{H}_n(y_i, y_k) = H_n(y_i, y_k) - \mathbb{E}[H_n(\boldsymbol{y}_i, \boldsymbol{y}_k)]$ is symmetric; (ii) $\mathbb{E}[\mathcal{H}_n(\boldsymbol{y}_i, \boldsymbol{y}_k)|\boldsymbol{y}_i] = 0$ a.s.; (iii) $\mathbb{E}[\mathcal{H}_n(\boldsymbol{y}_i, \boldsymbol{y}_k)^2] < \infty$; (iv)

$$\frac{\mathbb{E}[G_n(\boldsymbol{y}_i, \boldsymbol{y}_k)^2] + n^{-1}\mathbb{E}[\mathcal{H}_n(\boldsymbol{y}_i, \boldsymbol{y}_k)^4]}{\mathbb{E}[\mathcal{H}_n(\boldsymbol{y}_i, \boldsymbol{y}_k)^2]} \to 0$$

as $n \to \infty$, where $G_n(y_i, y_k) = \mathbb{E}[\mathcal{H}_n(\boldsymbol{y}_i, y_i)\mathcal{H}_n(\boldsymbol{y}_i, y_k)]$.

By construction, (i) is true. For (ii), it suffices to show $\mathbb{E}[H_n(\boldsymbol{y}_i, \boldsymbol{y}_k)|\boldsymbol{y}_i] = 0$ a.s.

$$\begin{aligned}\mathbb{E}[H_n(\boldsymbol{y}_i, \boldsymbol{y}_k)|\boldsymbol{y}_i] &= h^{\delta/2} \sum \int\int \tilde{R}(j, \omega_j, z, \boldsymbol{y}_{ij})\mathbb{E}[\tilde{R}(j, \omega_j, z, \boldsymbol{y}_{kj})|\boldsymbol{y}_i] b(x)\, dw_{j'}'\, dF_{\boldsymbol{x}}(x) \\ &= h^{\delta/2} \sum \int\int \tilde{R}(j, \omega_j, z, \boldsymbol{y}_{ij})\mathbb{E}[\tilde{R}(j, \omega_j, z, \boldsymbol{y}_{kj})] b(x)\, dw_{j'}'\, dF_{\boldsymbol{x}}(x)\end{aligned}$$

note that $\mathbb{E}[\tilde{R}(j, \omega_j, z, \boldsymbol{y}_{kj})]$ is 0 by construction.

$$= 0 \text{ a.s.} \tag{B.18}$$

(iii): From (ii), $\mathcal{H}_n(y_i, y_k) = H_n(y_i, y_k)$. Now, I will look at $\mathbb{E}[H_n(\boldsymbol{y}_i, \boldsymbol{y}_k)^2] = \mathbb{E}[\mathcal{H}_n(\boldsymbol{y}_i, \boldsymbol{y}_k)^2]$. Since this expectation has 16^4 terms, I will focus on the term only with $\gamma(\cdot)$ and $j = 1$. Further I ignore $\mathbb{E}[R(\cdot)]$ terms as they become asymptotically negligible after multiplying $h^{\delta/2}$ as shown in Equation (B.15). I define

$$H_n^*(\boldsymbol{s}_1, \boldsymbol{t}_1) = h^{\delta/2} \int\int \gamma(1, \omega_1, z, \tilde{\boldsymbol{x}})\gamma(1, \omega_1, z, \tilde{\boldsymbol{x}})b(x)\, dw_2'\, dF_{\boldsymbol{x}}(x).$$

I study $\mathbb{E}[H_n^*(\boldsymbol{y}_{i1}, \boldsymbol{y}_{k1})^2]$.

$$\begin{aligned}
\mathbb{E}[H_n^*(\boldsymbol{y}_{i1}, \boldsymbol{y}_{k1})^2] =& h^{\delta}\mathbb{E}\left[\int\int \gamma(1, \omega_1, z, \boldsymbol{x}_i)\gamma(1, \omega_1, z, \boldsymbol{x}_k)b(x)\, dw_2'\, dF_{\boldsymbol{x}}(x)\right. \\
&\left. \times \int\int \gamma(1, \omega_1'', z'', \boldsymbol{x}_i)\gamma(1, \omega_1'', z'', \boldsymbol{x}_k)b(x'')\, dw_2'''\, dF_{\boldsymbol{x}}(x'')\right] \\
=& h^{\delta}\int\int\int\int \mathbb{E}[\gamma(1, \omega_1, z, \boldsymbol{x}_i)\gamma(1, \omega_1'', z'', \boldsymbol{x}_i)]^2 \\
&\times b(x)b(x'')\, dw_2'\, dw_2'''\, dF_{\boldsymbol{x}}(x)\, dF_{\boldsymbol{x}}(x'')
\end{aligned}$$

By Equation (B.16) and the definition of $\gamma(\cdot)$, I have

$$\begin{aligned}
&\mathbb{E}[\gamma(j, \omega_j, z, \boldsymbol{x}_i)\gamma(j, \omega_j'', z'', \boldsymbol{x}_i)] = \\
&\qquad \xi(j, x)\xi(j, x'')\left[\int K(u_1)K(u_1 + (x - x'')/h)/h^{\delta} f_{\boldsymbol{x}}(x)\, du_1 + O(1)\right].
\end{aligned}$$

Now, I plug this expectation term back into $\mathbb{E}[H_n^*(\boldsymbol{y}_{i1}, \boldsymbol{y}_{k1})^2]$. Then, $O(1)$ part will be negligible after multiplying h^{δ}. And I change variable $u_2 = (x - x'')/h$. This variable change will remove h^{δ} entering $\mathbb{E}[H_n^*(\boldsymbol{y}_{i1}, \boldsymbol{y}_{k1})^2]$ and $1/h^{\delta}$ from $\mathbb{E}[\gamma(1, \omega_1, z, \boldsymbol{x}_i)\gamma(1, \omega_1'', z'', \boldsymbol{x}_i)]$. As a result, $\mathbb{E}[H(\boldsymbol{y}_i, \boldsymbol{y}_k)^2] = O(1)$.

(iv) Following a similar step as (iii), it is easy to show that $\mathbb{E}[H(\boldsymbol{y}_i, \boldsymbol{y}_k)^4] = O(1/h^{\delta})$. Using the results from (ii) and (iii),

$$\frac{\frac{1}{n}\mathbb{E}[\mathcal{H}(\boldsymbol{y}_i, \boldsymbol{y}_k)^4]}{\mathbb{E}[\mathcal{H}(\boldsymbol{y}_i, \boldsymbol{y}_k)^2]} = O(1/nh^{\delta}) = o(1)$$

as $nh^\delta \to \infty$ by Assumption 2.11.

Lastly, I argue that $\mathbb{E}[G_n(\boldsymbol{y}_i, \boldsymbol{y}_k)^2] = O(h^\delta)$. This expectation term will also expand to 16^8 terms. Thus, I will focus on the term only with $\gamma(\cdot)$ and $j = 1$. I define

$$
\begin{aligned}
G_n^*(y_{i1}, y_{k1}) =& \mathbb{E}[H_n^*(\boldsymbol{y}_{i1}, y_{i1}) H_n^*(\boldsymbol{y}_{i1}, y_{k1})] \\
=& h^\delta \mathbb{E}\left[\int\int \gamma(1, \omega_1, z, \boldsymbol{x}_i)\gamma(1, \omega_1, z, x_i) b(x)\, dw_2'\, dF_{\boldsymbol{x}}(x) \right. \\
& \left. \times \int\int \gamma(1, \omega_1'', z'', \boldsymbol{x}_i)\gamma(1, \omega_1'', z'', x_k) b(x'')\, dw_2'''\, dF_{\boldsymbol{x}}(x'')\right] \\
=& h^\delta \int\int\int\int \mathbb{E}\left[\gamma(1, \omega_1, z, \boldsymbol{x}_i)\gamma(1, \omega_1'', z'', \boldsymbol{x}_i)\right] \gamma(1, \omega_1, z, x_i)\gamma(1, \omega_1'', z'', x_k) \\
& \times b(x) b(x'')\, dw_2'\, dw_2'''\, dF_{\boldsymbol{x}}(x)\, dF_{\boldsymbol{x}}(x'') \\
=& \int\int\int\int \xi(j, x)\xi(j, x'') \int K(u_1) K(u_1 + (x - x'')/h) f_{\boldsymbol{x}}(x)\, du_1. \\
& \times \gamma(1, \omega_1, z, x_i)\gamma(1, \omega_1'', z'', x_k) b(x) b(x'')\, dw_2'\, dw_2'''\, dF_{\boldsymbol{x}}(x)\, dF_{\boldsymbol{x}}(x'') + O(h^\delta)
\end{aligned}
$$

Then, the order of $\mathbb{E}[G_n^*(\boldsymbol{y}_{i1}, \boldsymbol{y}_{k1})^2]$ depends on $\mathbb{E}[\gamma(1, \omega_1, z, \boldsymbol{x}_i)\gamma(1, \omega_1^\circ, z^\circ, \boldsymbol{x}_i)]$ and $\mathbb{E}[\gamma(1, \omega_1'', z'', \boldsymbol{x}_k)\gamma(1, \omega_1^\star, z^\star, \boldsymbol{x}_k)]$. By Equation (B.16),

$$
\begin{aligned}
&\mathbb{E}[\gamma(1, \omega_1, z, \boldsymbol{x}_i)\gamma(1, \omega_1^\circ, z^\circ, \boldsymbol{x}_i)] = \\
&\qquad \xi(j, x)\xi(j, x^\circ)\left[\int K(u_2) K(u_2 + (x - x^\circ)/h)/h^\delta f_{\boldsymbol{x}}(x)\, du_2 + O(1)\right] \\
&\mathbb{E}[\gamma(1, \omega_1'', z'', \boldsymbol{x}_k)\gamma(1, \omega_1^\star, z^\star, \boldsymbol{x}_k)] = \\
&\qquad \xi(j, x)\xi(j, x^\star)\left[\int K(u_3) K(u_3 + (x - x^\star)/h)/h^\delta f_{\boldsymbol{x}}(x)\, du_3 + O(1)\right].
\end{aligned}
$$

I change variable $u_4 = (x - x'')/h$, $u_5 = (x - x^\circ)/h$, and $u_6 = (x - x^\star)/h$. Then, I obtain

$$
\mathbb{E}[G_n^*(\boldsymbol{y}_{i1}, \boldsymbol{y}_{k1})^2] = O(h^\delta).
$$

Thus, I have

$$\frac{\mathbb{E}[G_n(\boldsymbol{y}_i,\boldsymbol{y}_k)^2]}{\mathbb{E}[\mathcal{H}_n(\boldsymbol{y}_i,\boldsymbol{y}_k)^2]} = o(1).$$

This proves the lemma. □

Lemma B5: If Assumptions 2.5, 2.9–2.11, and H_0 hold, then

$$nh^{\delta/2}\mathbb{E}[I_n] = B + o(1),$$

where B is defined in the main text.

Proof:

$$\begin{aligned}\mathbb{E}[I_n] =& \sum\int\int \mathbb{E}\{r_{\hat{f}_j}(j,\omega_j,z) - \mathbb{E}[r_{\hat{f}_j}(j,\omega_j,z)]\}^2 b(x)\,dw'_{j'}\,dF_{\boldsymbol{x}}(x)\\ &+\sum\int\int\{\mathbb{E}[r_{\hat{f}_j}(j,\omega_j,z)]\}^2 b(x)\,dw'_{j'}\,dF_{\boldsymbol{x}}(x)\\ =& A_{n,1} + A_{n,2}.\end{aligned}$$

It is easy to show that $nh^{\delta/2}A_{n,2} = o_p(1)$ by $h^{-r}\mathbb{E}[r_{\hat{f}_j}(j,\omega_j,z)] < \infty$ and Assumption 2.11. Now,

$$\begin{aligned}A_{n,1} =& n^{-2}\sum_{i=1}^{n}\mathbb{E}\left\{\sum\int\int\tilde{R}(j,\omega_j,\boldsymbol{y}_{ij})^2 b(x)\,dw'_{j'}\,dF_{\boldsymbol{x}}(x)\right\}\\ &+2n^{-2}\sum_{1\le i<k\le n}\mathbb{E}\left\{\sum\int\int\tilde{R}(j,\omega_j,\boldsymbol{y}_{ij})\tilde{R}(j,\omega_j,z,\boldsymbol{y}_{kj})b(x)\,dw'_{j'}\,dF_{\boldsymbol{x}}(x)\right\}\\ =& n^{-1}h^{-\delta/2}\left\{\mathbb{E}[H_n(\boldsymbol{y}_i,\boldsymbol{y}_i)] + 2n^{-1}\mathbb{E}[H_n(\boldsymbol{y}_i,\boldsymbol{y}_k)]\right\}.\end{aligned}$$

Note that $\mathbb{E}[H_n(\boldsymbol{y}_i,\boldsymbol{y}_k)] = 0$ by Equation (B.18).

$$\begin{aligned}
\mathbb{E}[H_n(\boldsymbol{y}_i, \boldsymbol{y}_i)] &= \mathbb{E}h^{\delta/2} \sum \int \int \tilde{R}(j, \omega_j, \boldsymbol{y}_{ij})^2 b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x) \\
&= h^{\delta/2} \sum \int \int \mathbb{E}[R(j, \omega_j, \boldsymbol{y}_{ij})^2] b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x) + O(h^{\delta/2}) \\
&= B + o(1).
\end{aligned}$$

The result holds. □

Lemma B6: Let $\Delta_n = \Gamma(\hat{f}, \hat{F}_{\boldsymbol{x}}) - \Gamma(\hat{f}, F_{\boldsymbol{x}})$. If Assumptions 2.1–2.11 and H_0 are true, then $nh^{\delta/2}\Delta_n = o_p(1)$.

Proof: Following the same step as Lemmas B1 and B2, I can show that

$$\begin{aligned}
\Gamma(\hat{f}, \hat{F}_{\boldsymbol{x}}) =& \sum \int \int r_{\hat{f}_j}(j, \omega_j, z)^2 b(x)\, dw'_{j'}\, d\hat{F}_{\boldsymbol{x}}(x) \\
&+ O_p(\|\hat{f}_{\boldsymbol{y_1}}(1, x) - f_{\boldsymbol{y_1}}(1, x)\|_\infty^3 + \|\hat{f}_{\boldsymbol{y_2}}(1, x) - f_{\boldsymbol{y_2}}(1, x)\|_\infty^3 + \|\hat{f}_{\boldsymbol{x}}(x) - f_{\boldsymbol{x}}(x)\|_\infty^3).
\end{aligned}$$

So it is sufficient to show that

$$\begin{aligned}
\Delta_n &= \sum \int \int r_{\hat{f}_j}(j, \omega_j, z)^2 b(x)\, dw'_{j'}\, d\{\hat{F}_{\boldsymbol{x}}(x) - F_{\boldsymbol{x}}(x)\} \\
&= o_p(n^{-1}h^{-\delta/2}).
\end{aligned}$$

I write

$$\begin{aligned}
\Delta_n =& \sum \int \int r_{\hat{f}_j}(j, \omega_j, z)^2 b(x)\, dw'_{j'}\, d\{\hat{F}_{\boldsymbol{x}}(x) - F_{\boldsymbol{x}}(x)\} \\
=& n^{-3} \sum_{i,k,\ell=1}^{n} \Big\{ \sum \int R(j, \omega_{\ell j}, z_\ell, y_{ij}) R(j, \omega_{\ell j}, z_\ell, y_{kj}) b(x_\ell)\, dw'_{j'} \\
&- \sum \int \int R(j, \omega_j, z, y_{ij}) R(j, \omega_j, z, y_{kj}) b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x) \Big\} \\
=& \Delta_{n,1} + \Delta_{n,2} + \Delta_{n,3} + \Delta_{n,4},
\end{aligned}$$

where

$$\Delta_{n,1} = n^{-3} \sum_{\ell \neq i,k}^{n} \Big\{ \sum \int R(j, \omega_{\ell j}, z_\ell, y_{ij}) R(j, \omega_{\ell j}, z_\ell, y_{kj}) b(x_\ell)\, dw'_{j'} - \sum \int \int R(j, \omega_j, z, y_{ij}) R(j, \omega_j, z, y_{kj}) b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x) \Big\}$$

is the centered term with $\ell \neq i, k$.

$$\Delta_{n,2} = 2n^{-3} \sum_{i \neq k}^{n} \Big\{ \sum \int R(j, \omega_{ij}, y_{ij}) R(j, \omega_{ij}, z_i, y_{kj}) b(x_i)\, dw'_{j'} \Big\}$$

is the term with $\ell = k$ or $\ell = i$ with $i \neq k$.

$$\Delta_{n,3} = n^{-3} \sum_{i}^{n} \Big\{ \sum \int R(j, \omega_{ij}, z_i, y_{ij})^2 b(x_i)\, dw'_{j'} \Big\}$$

is the term with $\ell = i = k$.

$$\Delta_{n,4} = -\, n^{-3} \sum_{i,k} \sum \int \int R(j, \omega_j, z, y_{ij}) R(j, \omega_j, z, y_{kj}) b(x)\, dw'_{j'}\, dF_{\boldsymbol{x}}(x)$$

is the centering term for $\Delta_{n,2}$ and $\Delta_{n,3}$. I first focus on simpler terms – $\Delta_{n,4}$ and $\Delta_{n,3}$.

$$-\Delta_{n,4} = n^{-1} I_n = n^{-1}(I_n - \mathbb{E} I_n) + n^{-1} \mathbb{E} I_n$$

Using the result from Lemmas B3, B4, B5, and Equation (B.16)

$$\begin{aligned}
&= n^{-2} h^{-\delta/2} \{ n h^{\delta/2} (I_n - \mathbb{E} I_n) \} + n^{-1} O(n^{-1} h^{-\delta} + h^{2r}) \\
&= n^{-1} O_p(n^{-1} h^{-\delta/2}) + n^{-1} h^{-\delta/2} O(n^{-1} h^{-\delta/2}) + h^{2r+\delta/2} O(n^{-1} h^{-\delta/2}) \\
&= o_p(n^{-1} h^{-\delta/2}).
\end{aligned}$$

Now, It is easy to see that

$$\begin{aligned}\mathbb{E}|\Delta_{n,3}| &= n^{-2}\mathbb{E}\left\{\sum \int R(j, \boldsymbol{\omega}_{ij}, \boldsymbol{z}_i, \boldsymbol{y}_{ij})^2 b(\boldsymbol{x}_i)\, dw'_{j'}\right\} \\ &\leq n^{-2}\sum \int_{\mathcal{W}_{j'}} \mathbb{E}[R(j, \boldsymbol{\omega}_{ij}, \boldsymbol{z}_i, \boldsymbol{y}_{ij})^2 b(\boldsymbol{x}_i)]\, dw'_{j'} \\ &=_{(1)} n^{-2}O(h^{-2\delta})\end{aligned}$$

From the above equation, $=_{(1)}$ is true because there exist a term with kernels whose arguments are 0, and this term becomes the dominating term. Then, by Markov inequality and Assumption 2.11, I have $\Delta_{n,3} = o_p(n^{-1}h^{-\delta/2})$.

Now, I look at $\Delta_{n,2}$

$$\begin{aligned}\Delta_{n,2}^2 =& 4n^{-6}\sum_{i\neq k}^{n}\sum_{i'\neq k'}^{n}\left[\sum_j \int R(j, \omega_{ij}, z_i, y_{ij})R(j, \omega_{ij}, z_i, y_{kj})b(x_i)\, dw'_{j'}\right. \\ &\left. \times \sum_l \int R(l, \omega_{i'l}, z_{i'}, y_{i'l})R(l, \omega_{i'l}, z_{i'}, y_{k'l})b(x_{i'})\, dw'_{l'}\right]\end{aligned}$$

I will show $\mathbb{E}[\Delta_{n,2}^2] = O(n^{-2}h^{-\delta})$. $\Delta_{n,2}^2$ can be decomposed to the terms with (i) $i \neq i'$, $k \neq k'$,$i \neq k'$,$i' \neq k$; (ii) $i = i'$, $k = k'$ or $i = k'$, $i' = k$; (iii) only one pair of indexes from (i, i'), (k, k'), (i, k'), and (i', k) is equal. I have $n(n-1)(n-2)(n-3)$ terms for (i), $n(n-1)$ for (ii), and $4n(n-1)(n-2)$ for (iii). For (i), expectation of each term is $O(h^{2r-2\delta})$. $O(h^{2r})$ comes from $\mathbb{E}[R(j, \omega_{ij}, z_i, \boldsymbol{y}_{kj})|x_i] = O(h^r)$ by Equation (B.15), and $O(h^{-2\delta})$ comes from $R(j, \omega_{ij}, z_i, y_{ij})$ – this term has kernels with their argument equal to 0. Thus, I have $O(n^{-2}h^{2r-2\delta}) = o(n^{-2}h^{-\delta})$ by Assumption 2.11. For (ii), expectation of each term is $O(h^{-3\delta})$ or diverge to infinity at a slower rate. In either cases, I only need to show that $\mathbb{E}[R(j, \boldsymbol{\omega}_{ij}, \boldsymbol{z}_i, \boldsymbol{y}_{kj})R(l, \boldsymbol{\omega}_{i'l}, \boldsymbol{z}_{i'}, \boldsymbol{y}_{k'l})] = O(h^{-\delta})$. The order of this expectation is determined by a γ term. Since non-kernel object is bounded below and above, I focus on kernel functions. Also, I focus on $i = i'$ and $k = k'$, but it applies to the other case as well.

Then,

$$\begin{aligned}&\mathbb{E}[\gamma(j,\boldsymbol{\omega}_{ij},\boldsymbol{z}_i,\boldsymbol{y}_{kj})\gamma(l,\boldsymbol{\omega}_{i'l},\boldsymbol{z}_{i'},\boldsymbol{y}_{k'l})]\\ &\quad\leq C_4\int\int K_h(x_i-x_k)^2 f_{\boldsymbol{x}}(x_i)f_{\boldsymbol{x}}(x_i)f_{\boldsymbol{x}}(x_k)\,dx_i\,dx_k.\end{aligned}$$

Let $u_1=(x_i-x_k)/h$.

$$=C_4h^{-\delta}\int\int K(u_1)^2 f_{\boldsymbol{x}}(x_i)f_{\boldsymbol{x}}(x_k+u_1h)f_{\boldsymbol{x}}(x_k)\,du_1\,dx_k$$

I can find constant C_9

$$\leq C_9h^{-\delta}.$$

Further, I can easily find constant C_{10} so that $R(j,\omega_{ij},z_i,y_{ij})<C_{10}h^{-\delta}$. Then, the result follows. Hence, $O(n^{-4}h^{-3\delta})=o(n^{-2}h^{-\delta})$. For (iii), each term is $O(h^{-2\delta})$ or diverge at a slower rate. Hence, $O(n^{-3}h^{-2\delta})=o(n^{-2}h^{-\delta})$. Thus, by Chevyshev inequality, $\Delta_{n,2}=o_p(n^{-1}h^{-\delta/2})$.

Now, I look at $\Delta_{n,1}$.

$$\begin{aligned}\Delta_{n,1}=&n^{-3}\sum_{\ell\neq i,k}^{n}\Big\{\sum\int R(j,\omega_{\ell j},z_\ell,y_{ij})R(j,\omega_{\ell j},z_\ell,y_{kj})b(x_\ell)\,dw'_{j'}\\ &-\sum\int\int R(j,\omega_j,z,y_{ij})R(j,\omega_j,z,y_{kj})b(x)\,dw'_{j'}\,dF_{\boldsymbol{x}}(x)\Big\}\\ =&n^{-3}\sum_{\ell\neq i,k}^{n}\sum\int\big\{R(j,\omega_{\ell j},z_\ell,y_{ij})R(j,\omega_{\ell j},z_\ell,y_{kj})b(x_\ell)dw'_{j'}\\ &-\mathbb{E}\left[\int R(j,\boldsymbol{\omega}_j,\boldsymbol{z},y_{ij})R(j,\boldsymbol{\omega}_j,\boldsymbol{z},y_{kj})b(\boldsymbol{x})\,dw'_{j'}|y_{ij},y_{ik}\right]\Big\}.\end{aligned}$$

I denote $\Delta_{n,11}=\mathbb{E}[\Delta_{n,1}|x_\ell]$ and $\Delta_{n,12}=\Delta_{n,1}-\mathbb{E}[\Delta_{n,1}|x_\ell]$. I will show that $\mathbb{E}[\Delta_{n,11}^2]=$

$o(n^{-2}h^{-\delta})$.

$$\begin{aligned}\mathbb{E}[\Delta_{n,11}^2] =& n^{-6}\sum_{\ell\neq i,k}\sum_{\ell'\neq i',k'}\sum_{j,l}\mathbb{E}\left[\int \mathbb{E}[R(j,\boldsymbol{\omega}_{\ell j},\boldsymbol{z}_\ell,\boldsymbol{y}_{ij})R(j,\boldsymbol{\omega}_{\ell j},\boldsymbol{z}_\ell,\boldsymbol{y}_{kj})b(x_\ell)|\boldsymbol{x}_\ell]\,dw'_{j'}\right.\\ &\times\left.\int \mathbb{E}[R(l,\boldsymbol{\omega}_{\ell' l},\boldsymbol{z}_{\ell'},\boldsymbol{y}_{i'l})R(l,\boldsymbol{\omega}_{\ell' l},\boldsymbol{z}_{\ell'},\boldsymbol{y}_{k'l})b(\boldsymbol{x}_{\ell'})|\boldsymbol{x}_{\ell'}]\,dw'_l\right]\\ &-\mathbb{E}\left[\int R(j,\boldsymbol{\omega}_j,\boldsymbol{z},\boldsymbol{y}_{ij})R(j,\boldsymbol{\omega}_j,\boldsymbol{z},\boldsymbol{y}_{kj})b(\boldsymbol{x})\,dw'_{j'}\right]\\ &\times\mathbb{E}\left[\int R(l,\boldsymbol{\omega}_l,\boldsymbol{z},\boldsymbol{y}_{i'l})R(l,\boldsymbol{\omega}_l,\boldsymbol{z},\boldsymbol{y}_{k'l})b(\boldsymbol{x})\,dw'_{l'}\right]\Bigg\}. \end{aligned}\tag{B.19}$$

Terms from $\mathbb{E}[\Delta_{n,11}^2]$ with index $\ell\neq\ell'$ equals 0 by Assumption 2.9–(i). Therefore, I only consider the case where $\ell=\ell'$. This gives me

$$\begin{aligned}\mathbb{E}[\Delta_{n,11}^2] =& n^{-6}\sum_{\ell\neq i,k,i',k'}\sum_{j,l}\mathbb{E}\left[\int \mathbb{E}[R(j,\boldsymbol{\omega}_{\ell j},\boldsymbol{z}_\ell,\boldsymbol{y}_{ij})R(j,\boldsymbol{\omega}_{\ell j},\boldsymbol{z}_\ell,\boldsymbol{y}_{kj})b(x_\ell)|\boldsymbol{x}_\ell]\,dw'_{j'}\right.\\ &\times\left.\int \mathbb{E}[R(l,\boldsymbol{\omega}_{\ell l},\boldsymbol{z}_{\ell},\boldsymbol{y}_{i'l})R(l,\boldsymbol{\omega}_{\ell l},\boldsymbol{z}_{\ell},\boldsymbol{y}_{k'l})b(\boldsymbol{x}_{\ell})|\boldsymbol{x}_{\ell}]\,dw'_l\right]\\ &-\mathbb{E}\left[\int R(j,\boldsymbol{\omega}_j,\boldsymbol{z},\boldsymbol{y}_{ij})R(j,\boldsymbol{\omega}_j,\boldsymbol{z},\boldsymbol{y}_{kj})b(\boldsymbol{x})\,dw'_{j'}\right]\\ &\times\mathbb{E}\left[\int R(l,\boldsymbol{\omega}_l,\boldsymbol{z},\boldsymbol{y}_{i'l})R(l,\boldsymbol{\omega}_l,\boldsymbol{z},\boldsymbol{y}_{k'l})b(\boldsymbol{x})\,dw'_{l'}\right]. \end{aligned}\tag{B.20}$$

To simplify notation, I use $R(\cdot,x_\ell,y_{ij})$ to denote $R(j,\omega_{\ell j},z_\ell,y_{i'j})$. From Equation (B.20), if $i\neq k$, then $\mathbb{E}[R(\cdot,x_\ell,\boldsymbol{y}_{ij})R(\cdot,x_\ell,\boldsymbol{y}_{kj})b(x_\ell)|x_\ell]=\mathbb{E}[R(\cdot,x_\ell,\boldsymbol{y}_{ij})|x_\ell]\mathbb{E}[R(\cdot,x_\ell,\boldsymbol{y}_{kj})b(x_\ell)|x_\ell]$. Note that $\mathbb{E}[R(\cdot,x_\ell,\boldsymbol{y}_{ij})|x_\ell]<C_7h^r$ from Equation (B.15). Also, when $i\neq k$, $\mathbb{E}[R(\cdot,\boldsymbol{x},\boldsymbol{y}_{ij})$ $R(\cdot,\boldsymbol{x},\boldsymbol{y}_{kj})b(\boldsymbol{x})]=O(h^{2r})$ by the same argument and the law of iterated expectation. On the other hand, if $i=k$, $\mathbb{E}[R(\cdot,x_\ell,\boldsymbol{y}_{ij})R(\cdot,x_\ell,\boldsymbol{y}_{kj})b(x_\ell)|x_\ell]<C_8h^{-\delta}$ by Equation (B.17). This argument and Assumption 2.7 (compactness of $\mathcal{X}$) lead to $\mathbb{E}[\int R(\cdot,\boldsymbol{x},\boldsymbol{y}_{ij})R(\cdot,$ $\boldsymbol{x},\boldsymbol{y}_{kj})b(\boldsymbol{x})]=O(h^{-\delta})$ by law of iterated expectation.

There are 4 possible cases I can consider from Equation (B.20). The first case is $i\neq k$, $i'\neq k'$, $i\neq k'$, $i'\neq k$. The first case has $O(n^5)$ terms and term has an order $O(h^{4r})$. Hence, the first case will be $O(n^{-1}h^{4r})=nh^{\delta+4r}O(n^{-2}h^{-\delta})=o(n^{-2}h^{-\delta})$ by

Assumption 2.11. The second case has $O(n^4)$ terms. Terms in the second case have an order $O(h^{2r-2\delta})$ or converge faster to 0. Terms with $i = k$ and $i' \neq k'$ has order $O(h^{2r-2\delta})$; other possible combinations have the same order or lower order. Hence, I have $O(n^{-2}h^{2r-2\delta}) = h^{2r-\delta}O(n^{-2}h^{-\delta}) = o(n^{-2}h^{-\delta})$. The third case has $O(n^3)$ terms. Some terms have $O(h^{-2\delta})$ or the other terms converge to 0. Terms with $i = k$ and $i' = k'$ have order $O(h^{-2\delta})$; other possible combinations have the same order or lower order. Therefore, The third case has an order $o(n^{-2}h^{-\delta})$. The last term is the case with $i = i' = k = k'$. In the last case, expectation of each term has order $O(h^{-2\delta})$. Therefore, I have $\mathbb{E}[\Delta^2_{n,11}] = o(n^{-2}h^{-\delta})$.

Now, I will show that $\mathbb{E}[\Delta^2_{n,12}] = o(n^{-2}h^{-\delta})$. Firstly, it is easy to see that (i) $\mathbb{E}[\Delta_{n,12}|x_\ell] = 0$; and (ii) $\mathbb{E}[\Delta_{n,12}|y_i, y_k] = 0$. Just like Equation (B.20), $\Delta^2_{n,11}$ has summands with index $\ell, \ell', i, i', k, k'$. Because of (i), terms with $\ell \neq \ell'$ have 0 from $\mathbb{E}[\Delta^2_{n,12}]$. Also, because of (ii), terms with $i \neq i', i' \neq k, k \neq k', k' \neq i$ are 0 from $\mathbb{E}[\Delta^2_{n,12}]$. Since $\mathbb{E}[\Delta^2_{n,12}] = \mathbb{E}[\Delta^2_{n,1}] - \mathbb{E}[\Delta^2_{n,11}]$, and $\mathbb{E}[\Delta^2_{n,11}] = o(n^{-2}h^{-\delta})$, I will show that $\mathbb{E}[\Delta^2_{n,1}] = o(n^{-2}h^{-\delta})$ for other cases than $\ell = \ell'$ and $i \neq i', i' \neq k, k \neq k', k' \neq i$. Again I focus on $j = 1$ as I am only interested in order.

$$
\begin{aligned}
\mathbb{E}[\Delta^2_{n,1}] =& n^{-6} \sum_{\ell \neq i,k,i',k'}^{n} \sum_{j,l} \mathbb{E}\left[\int R(j, \boldsymbol{\omega}_{\ell j}, \boldsymbol{z}_\ell, \boldsymbol{y}_{ij}) R(j, \boldsymbol{\omega}_{\ell j}, \boldsymbol{z}_\ell, \boldsymbol{y}_{kj}) b(x_\ell)\, dw'_{j'}\right. \\
& \left. \times \int R(l, \boldsymbol{\omega}_{\ell l}, \boldsymbol{z}_\ell, \boldsymbol{y}_{i'l}) R(l, \boldsymbol{\omega}_{\ell l}, \boldsymbol{z}_\ell, \boldsymbol{y}_{k'l}) b(\boldsymbol{x}_\ell)\, dw'_l\right] \\
& - \mathbb{E}\left[\mathbb{E}\left[\int R(j, \boldsymbol{\omega}_j, \boldsymbol{z}, \boldsymbol{y}_{ij}) R(j, \boldsymbol{\omega}_j, \boldsymbol{z}, \boldsymbol{y}_{kj}) b(\boldsymbol{x})\, dw'_{j'} | \boldsymbol{y}_{ij}, \boldsymbol{y}_{kj}\right]\right. \\
& \left. \times \mathbb{E}\left[\int R(l, \boldsymbol{\omega}_l, \boldsymbol{z}, \boldsymbol{y}_{i'l}) R(l, \boldsymbol{\omega}_l, \boldsymbol{z}, \boldsymbol{y}_{k'l}) b(\boldsymbol{x})\, dw'_{l'} | \boldsymbol{y}_{i'l}, \boldsymbol{y}_{i'l}\right]\right]. \qquad \text{(B.21)}
\end{aligned}
$$

To simplify notation, I use $R(\cdot, x_\ell, y_{ij})$ to denote $R(1, \omega_{\ell j}, z_\ell, y_{i'j})$. It is sufficient to show that the expectation of the last two terms has $o(n^{-2}h^{-\delta})$ since I can do the same for the first two terms by law of iterated expectation. Further, I ignore $b(x_\ell)$ since it is defined over compact support and thus is bounded above. I have $O(n^4)$ terms when one of the followings is true: $i = i'$, $i = k$, $i = k'$, $i' = k$, $i' = k'$, and $k = k'$. This case considers only one pair of indexes are equal and all else are different. When $i = k$ (equivalently $i' = k'$), the last two terms become (i) $\mathbb{E}[\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{ij})^2 | \boldsymbol{y}_{ij}]\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{i'l}) R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{k'l}) | \boldsymbol{y}_{i'l}, \boldsymbol{y}_{k'l}]]$. The other case is when $i = k'$ (equivalently $i' = k$, $i = i'$, and $k = k'$). In this case, the last two terms become (ii) $\mathbb{E}[\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{ij}) R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{i'j}) | \boldsymbol{y}_{ij}, \boldsymbol{y}_{i'j}]\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{i'l})$ $R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{k'l}) | \boldsymbol{y}_{i'l}, \boldsymbol{y}_{k'l}]]$ Hence, the order of these terms depends on $\mathbb{E}[\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{ij})^2 | \boldsymbol{y}_{ij}]$ and $\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{ij}) R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{i'j}) | \boldsymbol{y}_{ij}, \boldsymbol{y}_{i'j}]$ which will be studied later. Now, I have $O(n^3)$ terms when one of the followings is true: $i \neq i'$, $i \neq k$, $i \neq k'$, $i' \neq k$, $i' \neq k'$, and $k \neq k'$; all else indexes are equal. In this case, the expectation of the last two terms have the same form regardless of pairs of indexes. I choose $i' \neq k' = i = k$ without loss of generality, then $\mathbb{E}[\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{ij})^2 | \boldsymbol{y}_{ij}]\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{i'l}) R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{k'l}) | \boldsymbol{y}_{i'l}, \boldsymbol{y}_{k'l}]]$ – which has the equivalent form as (i). The last case is when $i = k = i' = k'$ which has $O(n^2)$ terms. In this case, the expectation of the last two terms is (iii) $\mathbb{E}[\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{ij})^2 | \boldsymbol{y}_{ij}]\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{il})^2 | \boldsymbol{y}_{il}]]$. The order depends on $\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{ij})^2 | \boldsymbol{y}_{ij}]$.

Now, I first study $\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{i1}) R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{i'1}) | \boldsymbol{y}_{i1}, \boldsymbol{y}_{i'1}]$. I argue that I only need to focus on $\gamma(\cdot)$ as before since terms concerning γ is dominant terms. I use $\gamma(\cdot, x_\ell, x_i)$ to denote $\gamma(1, w_{\ell 1}, w_{\ell 2}, w_2', z_\ell, x_i)$

$$\mathbb{E}[\gamma(\cdot, \boldsymbol{x}_\ell, x_i)\gamma(\cdot, \boldsymbol{x}_\ell, x_{i'}) | x_i, x_i'] \leq C_4 \int K_h(x_\ell - x_i) K_h(x_\ell - x_{i'}) f_{\boldsymbol{x}}(x_\ell)\, dx_\ell$$

Let $u_1 = (x_\ell - x_i)/h$

$$= C_4 \int K(u_1) K_h(u_1 + (x_i - x_{i'})/h)/h^\delta f_{\boldsymbol{x}}(x_i)\, du_1 + O(1) \tag{B.22}$$

This could also imply

$$\leq C_6 h^{-\delta}. \tag{B.23}$$

Now, I study $\mathbb{E}[\int R(\cdot, \boldsymbol{x}_\ell, \boldsymbol{y}_{i1})^2 | \boldsymbol{y}_{i1}]$. Again, I focus on a single $\gamma(\cdot)$ as a square of $\gamma(\cdot)$ is dominant term when I study a square of $R(\cdot)$.

$$\mathbb{E}[\gamma(\cdot, \boldsymbol{x}_\ell, x_i)^2 | x_i] \leq C_4^2 \int K_h(x_\ell - x_i)^2 f(x_\ell)\, dx_\ell$$

Let $u_1 = (x_\ell - x_i)/h$

$$\begin{aligned} &= C_4^2 h^{-\delta} \int K(u_1)^2 f_{\boldsymbol{x}}(x_i)\, du_1 \\ &\leq C_9 h^{-\delta}. \end{aligned} \tag{B.24}$$

Using Equations (B.22–B.24), I show that the sums of the terms that have the form of (i), (ii), and (iii) are $o(n^{-1}h^{-\delta/2})$ after multiplying n^{-6}. I focus on γ terms to study their orders. First, I have $O(n^3 + n^4)$ terms that have the form of expectation (i). Expectation (i) is bounded by $C_6 C_9 h^{-2\delta}$ by Equations (B.23) and (B.24). Consequently, the sum of terms that have the form of expectation (i), after multiplying n^{-6}, has order $O(n^{-2}h^{-2\delta})$. By Assumption 2.11, I have $nh^{3/2\delta} \to \infty$, and thus $O(n^{-2}h^{-2\delta}) = o(n^{-1}h^{-\delta/2})$.

Second, I have $O(n^4)$ terms that have the form of expectation (ii). Expectation (ii) is bounded by $C_6^2 h^{-2\delta}$ by Equation (B.23). As a result, the sum of terms that have the form of expectation (ii), after multiplying n^{-6}, has order $O(n^{-2}h^{-2\delta}) = o(n^{-1}h^{-\delta/2})$ for the same reason as the first case.

Third, I have $O(n^2)$ terms that have the form of expectation (iii). Expectation (iii) is bounded by $C_9^2 h^{-2\delta}$ by Equation (B.24). Therefore, the sum of terms that have the form of expectation (iii), after multiplying n^{-6} has order $O(n^{-4}h^{-2\delta}) = n^{-2}o(n^{-1}h^{-\delta/2}) = o(n^{-1}h^{-\delta/2})$.

Putting all together, I show that $\mathbb{E}[\Delta_{n,12}^2] = o(n^{-2}h^{-\delta})$. Since $\Delta_{n,1} = \Delta_{n,11} + \Delta_{n,12}$, I have $\Delta_{n,1} = o_p(n^{-1}h^{-\delta/2})$. Hence, I have the desired result. □

Lemma B7: If Assumptions 2.9–2.11 are satisfied, then $nh^{\delta/2} \|\hat{f}_{\boldsymbol{y}_j}(1, x) - f_{\boldsymbol{y}_j}(1, x)\|_\infty^3 =$

$o_p(1)$ for $j = 1, 2$ and $nh^{\delta/2}\|\hat{f}_{\boldsymbol{x}}(x) - f_{\boldsymbol{x}}(x)\|_\infty^3 = o_p(1)$.

Proof: Under Assumption 2.11,

$$nh^{\delta/2+3r} \to 0 \text{ and } \frac{\{\ln(h^{-1})\}^3}{nh^{2\delta}} \to 0.$$

Then, the result follows. □

Putting Lemmas B1–B7 together proves Theorem 2.1.

Proof of consistency of estimator: I first discuss the consistency of $\hat{B}$. First I show that $\|\hat{R} - R\|_\infty = o_p(1)$. This can be shown with series of triangle inequalities. for some functions q_1 and q_2, I have

$$\begin{aligned}
\|\hat{q}_1 - \hat{q}_2 - (q_1 - q_2)\|_\infty \leq& \|\hat{q}_1 - q_1\|_\infty + \|\hat{q}_2 - q_2\|_\infty \\
\|\hat{q}_1/\hat{q}_2 - q_1/q_2\|_\infty \leq& \|\hat{q}_1/\hat{q}_2 - q_1/\hat{q}_2 + q_1/q_2 \cdot q_2/\hat{q}_2 - q_1/q_2 \cdot \hat{q}_2/\hat{q}_2\|_\infty \\
\leq& \|1/\hat{q}_2\|_\infty \|\hat{q}_1 - q_1\|_\infty + \|q_1/(q_2 \cdot \hat{q}_2)\|_\infty \|\hat{q}_2 - q_2\|_\infty \\
\|\hat{q}_1\hat{q}_2 - q_1 q_2\|_\infty \leq& \|\hat{q}_1\|_\infty \|\hat{q}_2 - q_2\|_\infty + \|q_2\|_\infty \|\hat{q}_1 - q_1\|_\infty \\
\|\hat{q}_1{}^2 - q_1^2\|_\infty \leq& \|\hat{q}_1 + q_1\|_\infty \|\hat{q}_1 - q_1\|_\infty.
\end{aligned}$$

After replacing $\hat{q}_1$ and q_1 with adequate functions, applying above inequality multiple times, I can show that $\|\hat{R} - R\|_\infty = o_p(1)$ by Assumptions 2.9–2.11. Recall that $\gamma(\cdot)$ is one term from $R(\cdot)$. I show that $\|\hat{\gamma} - \gamma\|_\infty = o_p(1)$, where $\hat{\gamma}$ is defined by γ replacing population densities with their counterpart estimates. Just to illustrate the example, I write $p_{-3} = p(j|\omega_{j,-3}, z)$, $p_{-2} = p(j|\omega_{j,-2}, z)$, $\hat{p}_{-3} = p(j|\omega_{j,-3}, z)$, $\hat{p}_{-2} = p(j|\omega_{j,-2}, z)$.

$f = f_{\boldsymbol{x}}(x)$ and $K_h = K_h(\tilde{x} - x)$. Note that $\|K_h\|_\infty = K(0)h^{-\delta}$.

$$\begin{aligned}
K(0)^{-1}h^{\delta}\|\hat{\gamma} - \gamma\|_\infty =& \|\hat{\xi} - \xi\|_\infty = \left\| \frac{\hat{p}_{-3}/\hat{f}}{\hat{p}_{-3} - \hat{p}_{-2}} - \frac{p_{-3}/f}{p_{-3} - p_{-2}} \right\|_\infty \\
\leq& \left\| \frac{1}{\hat{p}_{-3} - \hat{p}_{-2}} \right\|_\infty \cdot \|\hat{p}_{-3}/\hat{f} - p_{-3}/f\|_\infty \\
&+ \left\| \frac{p_{-3}/f}{(\hat{p}_{-3} - \hat{p}_{-2})(p_{-3} - p_{-2})} \right\|_\infty \cdot \|\hat{p}_{-3} - \hat{p}_{-2} - (p_{-3} - p_{-2})\|_\infty \\
\leq& \left\| \frac{1}{\hat{p}_{-3} - \hat{p}_{-2}} \right\|_\infty \cdot \left(\|1/\hat{f}\|_\infty \cdot \|\hat{p}_{-3} - p_{-3}\|_\infty + \left\| \frac{p_{-3}}{\hat{f}f} \right\|_\infty \cdot \|\hat{f} - f\|_\infty \right) \\
&+ 2 \left\| \frac{p_{-3}/f}{(\hat{p}_{-3} - \hat{p}_{-2})(p_{-3} - p_{-2})} \right\|_\infty \cdot \|\hat{p}_{-3} - p_{-3}\|_\infty \\
=& O_p\left(\left(\frac{nh^{\delta}}{\ln h^{-1}} \right)^{-1/2} + h^r \right).
\end{aligned}$$

This result can be easily generalized to R, meaning that

$$\|\hat{R} - R\|_\infty = h^{-\delta} O_p\left(\left(\frac{nh^{\delta}}{\ln h^{-1}} \right)^{-1/2} + h^r \right).$$

For $\hat{B}$, I need to show that $h^{\delta/2}\|\hat{R}^2 - R^2\|_\infty = o_p(1)$.

$$\begin{aligned}
h^{\delta/2}\|\hat{R}^2 - R^2\|_\infty \leq& h^{\delta/2}\|\hat{R} + R\|_\infty \cdot \|\hat{R} - R\|_\infty \\
=& h^{-3/2\delta} O_p\left(\left(\frac{nh^{\delta}}{\ln h^{-1}} \right)^{-1/2} + h^r \right) = o_p(1).
\end{aligned}$$

I write $R_{ikj} = R(j, w_{ij}, w_{ij'}, w'_{j'}, z_i, y_{kj})$. Hence, I have

$$\begin{aligned}
\hat{B} =& \frac{1}{n^2} \sum_{i,k=1}^{n} h^{\delta/2} \sum_{j=1,2} \int_{\underline{w}_{j'}}^{w_{ij'}} R_{ikj}^2 b(x_i)\, dw'_{j'} + o_p(1) \\
=& h^{\delta/2} \sum_{j=1,2} \int_{\mathcal{X}} \int_{\underline{w}_{j'}}^{w_{j'}} \mathbb{E}[R(j, \omega_j, z, \boldsymbol{y}_j)^2 b(x)]\, dw'_{j'}\, d\{\hat{F}_{\boldsymbol{x}}(x) - F_{\boldsymbol{x}}(x)\} \\
&+ \frac{1}{n} \sum_{i=1}^{n} h^{\delta/2} \sum_{j=1,2} \int_{\mathcal{Y}_j} \int_{\underline{w}_{j'}}^{w_{ij'}} R(j, \omega_{ij}, z_i, y_j)^2 b(x_i)\, dw'_{j'}\, d\{\hat{F}_{\boldsymbol{y}_j}(y_j) - F_{\boldsymbol{y}_j}(y_j)\} \\
&+ B + o_p(1) \\
=& B_{n,1} + B_{n,2} + B + o_p(1).
\end{aligned}$$

I will show that $B_{n,1}$ and $B_{n,2}$ are $o_p(1)$. To show $B_{n,1} = o_p(1)$, I state the following lemma by Young (1917).

Lemma B8 (Young, 1917, equation (17)): Let q_1 and q_2 be of bounded variation δ dimensional real functions. If one of q_1 and q_2 is continuous, then

$$\begin{aligned}
\int_{\alpha}^{\beta} q_1\, dq_2 = q_1 q_2|_{\alpha}^{\beta} - \sum_i \int_{\alpha_i}^{\beta_i} q_1\, dq_2|_{\alpha_{-i}}^{\beta_{-i}} + \sum_{i,j} \int_{\alpha_i,\alpha_j}^{\beta_i,\beta_j} q_1\, dq_2|_{\alpha_{-i,-j}}^{\beta_{-i,-j}} \\
- \sum_{i,j,k} + \sum_{i,j,k,l} + \cdots + (-1)^{\delta} \int_{\alpha}^{\beta} q_1\, dq_2. \qquad \text{(B.25)}
\end{aligned}$$

Suppose that the total variation of q_1 is μ_{q_1} and q_2 is bounded above. Each term from Equation (B.25) is bounded by $\|q_2\|_\infty \mu_{q_1}$. Thus, Equation (B.25) is bounded by $2^\delta \|q_2\|_\infty \mu_{q_1}$. I will use this to show that $B_{n,1}$ is $o_p(1)$.

$$B_{n,1} = h^{-\delta/2} \sum_{j=1,2} \int_{\mathcal{X}} \int_{\underline{w}_{j'}}^{w_{j'}} \mathbb{E}[h^{\delta} R(j, \omega_j, z, \boldsymbol{y}_j)^2 b(x)]\, dw'_{j'}\, d\{\hat{F}_{\boldsymbol{x}}(x) - F_{\boldsymbol{x}}(x)\}$$

Let μ_1 be total variation of $\int_{\underline{w}_{j'}}^{w_{j'}} \mathbb{E}\left[h^{\delta} R(j, w_j, w_{j'}, w'_{j'}, z, \boldsymbol{y}_j)^2 b(x)\right] dw'_{j'}$, then since $f_{\boldsymbol{x}}$ is bounded below and above and of bounded variation and $k(u)$ is of bounded variation, $\mu_1 < \infty$. Then by Lemma B8, I have

$$\leq 2^{\delta+1}\mu_1 h^{-\delta/2} \sup_{x\in\mathcal{X}} |\hat{F}_{\boldsymbol{x}}(x) - F_{\boldsymbol{x}}(x)|.$$

Then, by Bonferroni inequality and Bernstein's inequality, it is easy to show that $B_{n,1} = o_p(1)$ as $nh^{\delta}/\log(n) \to \infty$.

Now, I show that $B_{n,2}$ is $o_p(1)$.

$$\begin{aligned}
B_{n,2} =& \frac{1}{n}\sum_i h^{\delta/2} \sum_{j=1,2} \int_{\underline{w}_{j'}}^{w_{ij'}} \int_{\mathcal{Y}_j} R(j, \omega_{ij}, z_i, y_j)^2 b(x_i)\, d\{\hat{F}_{\boldsymbol{y}_j}(y_j) - F_{\boldsymbol{y}_j}(y_j)\}\, dw'_{j'} \\
\leq & \sup_{x, w'_{j'} \in \mathcal{X}\times\mathcal{W}_{j'}} h^{\delta/2} \sum_{j=1,2} \left| \int_{\mathcal{Y}_j} R(j, \omega_j, z, y_j)^2 b(x)\, d\{\hat{F}_{\boldsymbol{y}_j}(y_j) - F_{\boldsymbol{y}_j}(y_j)\} \right| \\
=_{(1)} & \sup_{m=1,\cdots,M_n} \sup_{\substack{|x - x_m| \\ \times |w'_{j'} - w'_{mj'}| < \eta_n}} h^{\delta/2} \sum_{j=1,2} \left| \int_{\mathcal{Y}_j} R(j, \omega_j, z, y_j)^2 b(x)\, d\{\hat{F}_{\boldsymbol{y}_j}(y_j) - F_{\boldsymbol{y}_j}(y_j)\} \right| \\
\leq_{(2)} & \sup_{m=1,\cdots,M_n} h^{\delta/2} \sum_{j=1,2} \left| \int_{\mathcal{Y}_j} R(j, \omega^*_{mj}, z_m, y_j)^2 b(x_m)\, d\{\hat{F}_{\boldsymbol{y}_j}(y_j) - F_{\boldsymbol{y}_j}(y_j)\} \right| \\
& + \sup_{m=1,\cdots,M_n} \sup_{\substack{|x - x_m| \\ \times |w'_{j'} - w'_{mj'}| < \eta_n}} h^{\delta/2} \sum_{j=1,2} \left| \int_{\mathcal{Y}_j} R(j, \omega_j, z, y_j)^2 b(x) - R(j, \omega^*_{mj}, z_m, y_j)^2 \right. \\
& \left. \times\, b(x_m)\, d\{\hat{F}_{\boldsymbol{y}_j}(y_j) - F_{\boldsymbol{y}_j}(y_j)\} \right|, \qquad \text{(B.26)}
\end{aligned}$$

where $\omega^*_{mj} = (w_{mj}, w_{mj'}, w'_{mj'})$. $=_{(1)}$ holds by the compactness of $\mathcal{X}$ and $\leq_{(2)}$ holds by triangle inequality. I can make the second term of Equation (B.26) is negligible compared to the first term by choosing smaller η_n. Thus, I focus on the first term. I denote the first term of Equation (B.26) by $B_{n,21}$. I have

$$
\begin{aligned}
&\Pr(B_{n,21} > \varepsilon) \\
&\leq M_n \\
&\quad \times \sup_{m=1,\cdots,M_n} \Pr\left(h^{\delta/2} \sum_{j=1,2} \left| \int_{\mathcal{Y}_j} R(j,\omega_{mj},z_m,y_j)^2 b(x_m)\, d\{\hat{F}_{\boldsymbol{y}_j}(y_j) - F_{\boldsymbol{y}_j}(y_j)\} \right| > \varepsilon \right) \\
&\leq \sum_{j=1,2} M_n \\
&\quad \times \sup_{m=1,\cdots,M_n} \Pr\left(\left| \int_{\mathcal{Y}_j} R(j,\omega_{mj},z_m,y_j)^2 b(x_m)\, d\{\hat{F}_{\boldsymbol{y}_j}(y_j) - F_{\boldsymbol{y}_j}(y_j)\} \right| > h^{-\delta/2}\varepsilon \right).
\end{aligned} \tag{B.27}
$$

By Berstein's inequality, the probability object from Equation (B.27) is bounded by

$$
2\exp\left(\frac{-n(\varepsilon h^{-\delta/2}/4)^2}{2\tilde{\sigma}_n^2 + 2/3\mu_1^2 \varepsilon h^{-\delta/2}} \right),
$$

where

$$
\tilde{\sigma}_n^2 = Var\left(R(j,\omega_{mj},z_m,\boldsymbol{y}_j)^2 b(x_m)^2 \right) = O(h^{-3\delta}).
$$

For $M_n = O(n)$, I have

$$
Pr(B_{n,21} > \varepsilon) \leq M_n 2\exp\left(\frac{-n(\varepsilon h^{-\delta/2}/4)^2}{2\tilde{\sigma}^2 + 2/3\mu_1^2 \varepsilon h^{-\delta/2}} \right).
$$

By $nh^{2\delta}/\log(n) \to \infty$, I have $B_{n,2} = o_p(1)$.

For $\hat{\sigma}^2$, I need to show that $h^{\delta}\|\hat{R}_j\hat{R}_l - R_jR_l\| = o_p(1)$. I write $R_j = R(j,\omega_{ij},z_i,y_{kj})$

and $R_l = R(l, \omega_{\ell l}, z_\ell, y_{kl})$.

$$h^\delta \|\hat{R}_j \hat{R}_l - R_j R_l\|_\infty \leq h^\delta \|\hat{R}_j\|_\infty \cdot \|\hat{R}_l - R_l\| + h^\delta \|R_l\|_\infty \cdot \|\hat{R}_j - R_j\|_\infty$$
$$= h^{-\delta} O_p \left(\left(\frac{nh^\delta}{\ln h^{-1}} \right)^{-1/2} + h^r \right) = o_p(1).$$

Following the same argument as $\hat{B}$, I can show that

$$\begin{aligned}
\hat{\sigma}^2 =& \frac{1}{n^2} h^\delta \sum_{i,\ell=1}^{n} \sum_{j,l=1,2} \int_{\underline{w}_{j'}}^{w_{ij'}} \int_{\underline{w}_{l'}}^{w_{\ell l'}} \left[\frac{1}{n} \sum_{k=1}^{n} R(j, \omega_{ij}, z_i, y_{kj}) R(l, \omega_{\ell l}, z_\ell, y_{kl}) \right]^2 \\
& \times b(x_i) b(x_\ell)\, dw'_{j'}\, dw'_{l'} + o_p(1) \\
=& \frac{2}{n^2} h^\delta \sum_{i,\ell=1}^{n} \sum_{j,l=1,2} \int_{\underline{w}_{j'}}^{w_{ij'}} \int_{\underline{w}_{l'}}^{w_{\ell l'}} \int_{\mathcal{Y}} R(j, \omega_{ij}, z_i, y_j) R(l, \omega_{\ell l}, z_\ell, y_l) d\{\hat{F}_{\boldsymbol{y}}(y) - F_{\boldsymbol{y}}(y)\} \\
& \times \mathbb{E}[R(j, \omega_{ij}, z_i, \boldsymbol{y}_j) R(l, \omega_{\ell l}, z_\ell, \boldsymbol{y}_l)] b(x_i) b(x_\ell)\, dw'_{j'}\, dw'_{l'} \\
& + \frac{1}{n^2} h^\delta \sum_{i,\ell=1}^{n} \sum_{j,l=1,2} \int_{\underline{w}_{j'}}^{w_{ij'}} \int_{\underline{w}_{l'}}^{w_{\ell l'}} \left[\int_{\mathcal{Y}} R(j, \omega_{ij}, z_i, y_j) R(l, \omega_{\ell l}, z_\ell, y_l) d\{\hat{F}_{\boldsymbol{y}}(y) - F_{\boldsymbol{y}}(y)\} \right]^2 \\
& \times b(x_i) b(x_\ell)\, dw'_{j'}\, dw'_{l'} \\
& + \frac{1}{n} \sum_{i=1}^{n} h^\delta \sum_{j,l=1,2} \int_{\mathcal{X}} \int_{\underline{w}_{j'}}^{w_{ij'}} \int_{\underline{w}_{l'}}^{w_{\ell l'}} \mathbb{E}[R(j, \omega_{ij}, z_i, \boldsymbol{y}_j) R(l, \omega_{\ell l}, z_\ell, \boldsymbol{y}_l)]^2 \\
& \times b(x_i) b(x_\ell)\, dw'_{j'}\, dw'_{l'} d\{\hat{F}_{\boldsymbol{x}}(x_\ell) - F_{\boldsymbol{x}}(x_\ell)\} \\
& + h^\delta \sum_{j,l=1,2} \int_{\mathcal{X}} \int_{\mathcal{X}} \int_{\underline{w}_{j'}}^{w_{ij'}} \int_{\underline{w}_{l'}}^{w_{\ell l'}} \mathbb{E}[R(j, \omega_{ij}, z_i, \boldsymbol{y}_j) R(l, \omega_{\ell l}, z_\ell, \boldsymbol{y}_l)]^2 \\
& \times b(x_i) b(x_\ell)\, dw'_{j'}\, dw'_{l'}\, dF_{\boldsymbol{x}}(x_\ell) d\{\hat{F}_{\boldsymbol{x}}(x_i) - F_{\boldsymbol{x}}(x_i)\} + \sigma^2 + o_p(1) \\
=& \sigma_{n,1}^2 + \sigma_{n,2}^2 + \sigma_{n,3}^2 + \sigma_{n,4}^2 + \sigma^2 + o_p(1).
\end{aligned}$$

I will show that $\sigma_{n,1}^2, \cdots, \sigma_{n,4}^2$ are $o_p(1)$.

$$\sigma_{n,1}^2 = \frac{2}{n^2} h^{\delta} \sum_{i,\ell=1}^{n} \sum_{j,l=1,2} \int_{\underline{w}_{j'}}^{w_{ij'}} \int_{\underline{w}_{l'}}^{w_{\ell l'}} \int_{\mathcal{Y}} R(j, \omega_{ij}, z_i, y_j) R(l, \omega_{\ell l}, z_\ell, y_l)$$
$$\times\, d\{\hat{F}_{\boldsymbol{y}}(y) - F_{\boldsymbol{y}}(y)\} \mathbb{E}[R(j, \omega_{ij}, z_i, \boldsymbol{y}_j) R(l, \omega_{\ell l}, z_\ell, \boldsymbol{y}_l)] b(x_i) b(x_\ell)\, dw'_{j'}\, dw'_{l'}$$

From Equation (B.17), I have $\mathbb{E}[R(j, \omega_{ij}, z_i, \boldsymbol{y}_j) R(l, \omega_{\ell l}, z_\ell, \boldsymbol{y}_l)] \leq C_8 h^{-\delta}$. Also, since $b(\cdot)$ is bounded as $\mathcal{X}$ is compact, I can find constant C_{12} so that

$$\begin{aligned}
&\leq C_{12} \sum_{j,l=1,2} \sup_{(x_i,x_\ell)\in\mathcal{X}^2} \int_{\mathcal{W}_{j'}} \int_{\mathcal{W}_{l'}} \left| \frac{1}{n} \sum_{k=1}^{n} R(j, \omega_{ij}, z_i, y_{kj}) R(l, \omega_{\ell l}, z_\ell, y_{kl}) \right. \\
&\quad \left. - \mathbb{E}\left[R(j, \omega_{ij}, z_i, \boldsymbol{y}_j) R(l, \omega_{\ell l}, z_\ell, \boldsymbol{y}_l)\right] \right| dw'_{j'}\, dw'_{l'} \\
&= C_{12} \sum_{j,l=1,2} \sup_{m=1,\cdots,M_n} \sup_{\substack{|x_i - x_{mi}| \\ \times |x_\ell - x_{m\ell}| < \eta_n}} \int_{\mathcal{W}_{j'}} \int_{\mathcal{W}_{l'}} \left| \frac{1}{n} \sum_{k=1}^{n} R(j, \omega_{ij}, z_i, y_{kj}) R(l, \omega_{\ell l}, z_\ell, y_{kl}) \right. \\
&\quad \left. - \mathbb{E}\left[R(j, \omega_{ij}, z_i, \boldsymbol{y}_j) R(l, \omega_{\ell l}, z_\ell, \boldsymbol{y}_l)\right] \right| dw'_{j'}\, dw'_{l'} \\
&\leq C_{12} \sum_{j,l=1,2} \sup_{m=1,\cdots,M_n} \int_{\mathcal{W}_{j'}} \int_{\mathcal{W}_{l'}} \left| \frac{1}{n} \sum_{k=1}^{n} R(j, \omega_{mij}, z_{mi}, y_{kj}) R(l, \omega_{m\ell l}, z_{m\ell}, y_{kl}) \right. \\
&\quad \left. - \mathbb{E}\left[R(j, \omega_{mij}, z_{mi}, \boldsymbol{y}_j) R(l, \omega_{,\ell l}, z_{m\ell}, \boldsymbol{y}_l)\right] \right| dw'_{j'}\, dw'_{l'} \\
&+ C_{12} \sum_{j,l=1,2} \sup_{m=1,\cdots,M_n} \sup_{\substack{|x_i - x_{mi}| \\ \times |x_\ell - x_{m\ell}| < \eta_n}} \int_{\mathcal{W}_{j'}} \int_{\mathcal{W}_{l'}} \left| \frac{1}{n} \sum_{k=1}^{n} R(j, \omega_{ij}, z_i, y_{kj}) R(l, \omega_{\ell l}, z_\ell, y_{kl}) \right. \\
&\quad - \mathbb{E}\left[R(j, \omega_{ij}, z_i, \boldsymbol{y}_j) R(l, \omega_{\ell l}, z_\ell, \boldsymbol{y}_l)\right] - \frac{1}{n} \sum_{k=1}^{n} R(j, \omega_{mij}, z_i, y_{kj}) R(l, \omega_{m\ell l}, z_\ell, y_{kl}) \\
&\quad \left. + \mathbb{E}\left[R(j, \omega_{mij}, z_{mi}, \boldsymbol{y}_j) R(l, \omega_{m\ell l}, z_{m\ell}, \boldsymbol{y}_l)\right] \right| dw'_{j'}\, dw'_{l'},
\end{aligned} \tag{B.28}$$

where $\omega_{mij} = (w_{mij}, w_{mij'}, w'_{j'})$ and M_n intervals $\{(x_i, x_\ell) : |x_i - x_{mi}| \times |x_\ell - x_{m\ell}| < \eta_n\}$ cover $\mathcal{X}^2$. From Equation (B.28), I can always choose η_n small enough that the second term is negligible compared to the first term. Now, I will show that Equation (B.28) is bounded in probability, and so is $\sigma_{n,1}^2$. When showing Equation (B.28) is bounded in probability I focus on $\gamma(\cdot)\gamma(\cdot)$ since $R(\cdot)R(\cdot)$ is a sum of $\gamma(\cdot)\gamma(\cdot)$'s. By Benferroni inequality, I consider

$$M_n \sup_{m=1,\cdots,M_n} \Pr\left(\left|\frac{1}{n}\sum_{k=1}^{n}\gamma(j,\omega_{mij},z_{mi},x_k)\gamma(l,\omega_{m\ell\ell},z_{m\ell},x_k) - \mathbb{E}\left[\gamma(j,\omega_{mij},z_{mi},\boldsymbol{x})\gamma(l,\omega_{m\ell l},z_{m\ell},\boldsymbol{x})\right]\right| > \varepsilon\right). \quad \text{(B.29)}$$

By Bernstein's inequality, Equation (B.29) is bounded by, assuming $|k(u)| < C_{14}$,

$$M_n 2\exp\left(\frac{-n(\varepsilon/4)^2}{2\tilde{\sigma}_n^2 + 2/3C_{14}^{\delta}\varepsilon}\right),$$

where $\tilde{\sigma}_n^2 = Var\left(\gamma(j,\omega_{mij},z_{mi},\boldsymbol{x}_k)\gamma(l,\omega_{m\ell l},z_{m\ell},\boldsymbol{x}_k)\right) = O(h^{-\delta})$. Let $M_n = O(n)$, then Equation (B.29) is $o_p(1)$. This leads that $\sigma_{n,1}^2 = o_p(1)$.

I can easily show that $\sigma_{n,2}^2$ by Equations (B.28) and (B.29) and $\sup|f^2| = (\sup|f|)^2$.

Also, it is easy to show that $\sigma_{n,3}^2$ and $\sigma_{n,4}^2$ are $O_p(n^{-1/2}h^{-\delta})$. By Equation (B.17), I have $\mathbb{E}\left[R(j,\omega_{mij},z_{mi},\boldsymbol{y}_j)R(l,\omega_{m\ell l},z_{m\ell},\boldsymbol{y}_l)\right] \le C_8^2 h^{-\delta}$. then, the result is obtained by Chebyshev's inequality. □

Proof of Theorem 2.2: From Lemma B1 by Equation (B.4),

$$\begin{aligned}
\Gamma(\hat{f}, F_{\boldsymbol{x}}) &= \Gamma(f, F_{\boldsymbol{x}}) + \Psi'(0) + \Psi''(0) + \Psi'''(t^*) \\
&= \Gamma(f, F_{\boldsymbol{x}}) + O_p(\|\hat{f}_{\boldsymbol{y}_1}(1,x) - f_{\boldsymbol{y}_1}(1,x)\|_\infty + \|\hat{f}_{\boldsymbol{y}_2}(1,x) - f_{\boldsymbol{y}_2}(1,x)\|_\infty \\
&\quad + \|\hat{f}_{\boldsymbol{x}}(x) - f_{\boldsymbol{x}}(x)\|_\infty).
\end{aligned}$$

Notice that Lemma B6 holds true under the alternative, too. Then, by Lemma B6,

$$\Gamma(\hat{f}, \hat{F}_{\boldsymbol{x}}) = \Gamma(\hat{f}, F_{\boldsymbol{x}}) + o_p(n^{-1}h^{-\delta/2}).$$

Therefore,

$$\Gamma(\hat{f},\hat{F}_{\boldsymbol{x}}) = \Gamma(f,F_{\boldsymbol{x}}) + O_p(\|\hat{f}_{\boldsymbol{y}_1}(1,x) - f_{\boldsymbol{y}_1}(1,x)\|_\infty + \|\hat{f}_{\boldsymbol{y}_2}(1,x) - f_{\boldsymbol{y}_2}(1,x)\|_\infty + \|\hat{f}_{\boldsymbol{x}}(x) - f_{\boldsymbol{x}}(x)\|_\infty) + o_p(n^{-1}h^{-\delta/2}).$$

This shows that

$$T_n = nh^{\delta/2}\Gamma(f,F_{\boldsymbol{x}}) + nh^{\delta/2}O_p(\|\hat{f}_{\boldsymbol{y}_1}(1,x) - f_{\boldsymbol{y}_1}(1,x)\|_\infty + \|\hat{f}_{\boldsymbol{y}_2}(1,x) - f_{\boldsymbol{y}_2}(1,x)\|_\infty + \|\hat{f}_{\boldsymbol{x}}(x) - f_{\boldsymbol{x}}(x)\|_\infty) + o_p(1) \to \infty.$$

if $\Gamma(f,F_{\boldsymbol{x}}) > 0$. This proves Theorem 2.2. □

B.2 Omitted Proofs from Section 2.4

Proof of Lemma 2.2: I first prove (⇒) and focus on $j = 1$. Cases with other j's can be proved similarly. Suppose that for some w_{-1}, $p(\mathcal{J}|w_1, w_{-1}, z)$ is not constant in z. That is, for some $z \in \mathcal{Z}$, $\nabla_z p(\mathcal{J}|w_1, w_{-1}, z) \neq 0$, where ∇_z is gradient. By Assumption 2.8, $p(\mathcal{J}|w, z)$ is not additively separable. Hence for some z, I have $\partial_{w_j}[\nabla_z p(\mathcal{J}|w, z)] \neq 0$.

Recall ψ,

$$\psi(1, w_1, w_{-1}, w'_{-1}, z) = \frac{\Delta^{w_J}_{w'_J}\left[\cdots[\Delta^{w_2}_{w'_2} p(\mathcal{J}|w_1, \cdot, z)]\right]}{\Delta^{\bar{w}_J}_{\underline{w}_J}\left[\cdots[\Delta^{\bar{w}_2}_{\underline{w}_2} p(\mathcal{J}|w_1, \cdot, z)]\right]}.$$

With the assumption that ψ does not vary with z, I must have $\nabla_z \psi(1, w_1, w_{-1}, w'_{-1}, z) = 0$. Choose any element from $\nabla_z \psi$. Without loss of generality, I choose the first element, and it will be $\partial_{z_1}\psi$. For all w, z and w'_{-1}, I have

$$\partial_{z_1}\psi(1, w_1, w_{-1}, w'_{-1}, z) = 0. \tag{B.30}$$

I take the derivative of Equation (B.30) with respect to $w_2, \cdots, w_J$ and $w'_2, \cdots, w'_J$, separately.

$$\begin{aligned}
w_{-1}:\ & \partial_{w_2}\cdots\partial_{w_J}\partial_{z_1}p(\mathcal{J}|w_1, w_{-1}, z)\Delta_{\underline{w}_J}^{\bar{w}_J}\left[\cdots[\Delta_{\underline{w}_2}^{\bar{w}_2}p(\mathcal{J}|w_1, \cdot, z)]\right] \\
&= \partial_{w_2}\cdots\partial_{w_J}p(\mathcal{J}|w_1, w'_{-1}, z)\partial_{z_1}\Delta_{\underline{w}_J}^{\bar{w}_J}\left[\cdots[\Delta_{\underline{w}_2}^{\bar{w}_2}p(\mathcal{J}|w_1, \cdot, z)]\right] \\
w'_{-1}:\ & \partial_{w'_2}\cdots\partial_{w'_J}\partial_{z_1}p(\mathcal{J}|w_1, w_{-1}, z)\Delta_{\underline{w}_J}^{\bar{w}_J}\left[\cdots[\Delta_{\underline{w}_2}^{\bar{w}_2}p(\mathcal{J}|w_1, \cdot, z)]\right] \\
&= \partial_{w'_2}\cdots\partial_{w'_J}p(\mathcal{J}|w_1, w'_{-1}, z)\partial_{z_1}\Delta_{\underline{w}_J}^{\bar{w}_J}\left[\cdots[\Delta_{\underline{w}_2}^{\bar{w}_2}p(\mathcal{J}|w_1, \cdot, z)]\right].
\end{aligned}$$

Rearrange the two equations above,

$$\frac{\partial_{w_2}\cdots\partial_{w_J}\partial_{z_1}p(\mathcal{J}|w_1, w_{-1}, z)}{\partial_{w_2}\cdots\partial_{w_J}p(\mathcal{J}|w_1, w_{-1}, z)} = \frac{\partial_{w'_2}\cdots\partial_{w'_J}\partial_{z_1}p(\mathcal{J}|w_1, w_{-1}, z)}{\partial_{w'_2}\cdots\partial_{w'_J}p(\mathcal{J}|w_1, w_{-1}, z)}.$$

The rest of argument is identical to the proof of Lemma 2.1. Therefore, the result follows. □

Proof of Theorem 2.3:

Lemma B1: Define $\Omega_{\boldsymbol{y}_j} = \{g_{\boldsymbol{y}_j} : \mathbb{R}^{1+\delta} \to \mathbb{R},\ g_{\boldsymbol{y}_j}$ is bounded, $\int g_{\boldsymbol{y}_j} = 0$, and $\|g_{\boldsymbol{y}_j}\|_\infty < c_1/2\}$ and $\Omega_{\boldsymbol{x}} = \{g_{\boldsymbol{x}} : \mathbb{R}^\delta \to \mathbb{R},\ g_{\boldsymbol{x}}$ is bounded, $\int g_{\boldsymbol{x}} = 0$, and $\|g_{\boldsymbol{x}}\|_\infty < c_1/2\}$, where $\|g\|_\infty = \sup_x \|g(x)\|$. Let $g = (g_{\boldsymbol{y}_1}, \cdots, g_{\boldsymbol{y}_J}, g_{\boldsymbol{x}})$ belong to $\prod_{j\in\mathcal{J}\backslash 0}\Omega_{\boldsymbol{y}_j} \times \Omega_{\boldsymbol{x}}$ and $g_j = (g_{\boldsymbol{y}_j}, g_{\boldsymbol{x}})$ belong to $\Omega_{\boldsymbol{y}_j} \times \Omega_{\boldsymbol{x}}$ for $j \in \mathcal{J}\backslash 0$. Then, under Assumptions 2.1–2.8, 2.9–(ii), and

H_0, $\Gamma(\cdot, F_{\boldsymbol{x}})$ has the following expansion:

$$\Gamma(f+g, F_{\boldsymbol{x}}) = \sum_{j \in \mathcal{J}\backslash 0} \int_{\mathcal{X}} \int_{\underline{w}_{-j}}^{w_{-j}} D_a(\psi \rho_{g_j})(j, \omega_j, z)^2 b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x) + R(f + t^* g, F_{\boldsymbol{x}}),$$

where $t^* \in [0,1]$, $\sup\{|R(f+t^*g, F_{\boldsymbol{x}})|/(\sum_{j\in\mathcal{J}\backslash 0} \|g_{\boldsymbol{y}_j}\|_\infty^3 + \|g_{\boldsymbol{x}}\|_\infty^3) : (g_{\boldsymbol{y}_1}, \cdots, g_{\boldsymbol{y}_J}, g_{\boldsymbol{x}}) \in \prod_{j\in\mathcal{J}\backslash 0} \Omega_{\boldsymbol{y}_j} \times \Omega_{\boldsymbol{x}}\} < \infty$ if $g \neq 0$ and $R(f+t^*g, F_{\boldsymbol{x}}) = 0$ if $g = 0$. I further define

$$\begin{aligned} \rho_{g_j}(j, \omega_j, z) =& \frac{\Delta_{w'_J}^{w_J}\left[\cdots \Delta_{w'_{j+1}}^{w_{j+1}}\left[\Delta_{w'_{j-1}}^{w_{j-1}}\left[\cdots [\Delta_{w'_1}^{w_1} \kappa_{g_j}(j, w_j, \cdot, z)]\right]\right]\right]}{\Delta_{w'_J}^{w_J}\left[\cdots \Delta_{w'_{j+1}}^{w_{j+1}}\left[\Delta_{w'_{j-1}}^{w_{j-1}}\left[\cdots [\Delta_{w'_1}^{w_1} p(j|w_j, \cdot, z)]\right]\right]\right]} \\ &- \frac{\Delta_{\underline{w}_J}^{\bar{w}_J}\left[\cdots \Delta_{\underline{w}_{j+1}}^{\bar{w}_{j+1}}\left[\Delta_{\underline{w}_{j-1}}^{\bar{w}_{j-1}}\left[\cdots [\Delta_{\underline{w}_1}^{\bar{w}_1} \kappa_{g_j}(j, w_j, \cdot, z)]\right]\right]\right]}{\Delta_{\underline{w}_J}^{\bar{w}_J}\left[\cdots \Delta_{\underline{w}_{j+1}}^{\bar{w}_{j+1}}\left[\Delta_{\underline{w}_{j-1}}^{\bar{w}_{j-1}}\left[\cdots [\Delta_{\underline{w}_1}^{\bar{w}_1} p(j|w_j, \cdot, z)]\right]\right]\right]}, \end{aligned} \tag{B.31}$$

with

$$\kappa_{g_j}(j, w_j, w_{-j}, z) = p(j|w_j, w_{-j}, z)\left(\frac{g_{\boldsymbol{y}_j}(1, w_j, w_{-j}, z)}{f_{\boldsymbol{y}_j}(1, w_j, w_{-j}, z)} - \frac{g_{\boldsymbol{x}}(w_j, w_{-j}, z)}{f_{\boldsymbol{x}}(w_j, w_{-j}, z)}\right).$$

Proof: From Equation (2.11), I have

$$\Gamma(f+g, F_{\boldsymbol{x}}) = \sum_{j\in\mathcal{J}\backslash 0} \int_{\mathcal{X}} \int_{\underline{w}_{-j}}^{w_{-j}} D_a(\psi(f_j + g_j))(j, \omega_j, z)^2 b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x).$$

I apply the functional expansion introduced by Aït-Sahalia et al. (2001). Define

$$\Psi(t) = \sum_{j\in\mathcal{J}\backslash 0} \int_{\mathcal{X}} \int_{\underline{w}_{-j}}^{w_{-j}} D_a(\psi(f_j + tg_j))(j, \omega_j, z)^2 b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x).$$

and $(tg_{\boldsymbol{y}_1}, \cdots, tg_{\boldsymbol{y}_J}, tg_{\boldsymbol{x}}) \in \prod_{j\in\mathcal{J}\backslash 0} \Omega_{\boldsymbol{y}_j} \times \Omega_{\boldsymbol{x}}$ for all $0 \le t \le 1$. Now, Taylor expand the function $\Psi(t)$ about $t = 0$:

$$\Psi(t) = \Psi(0) + t\Psi'(0) + t^2\Psi''(0)/2 + t^3\Psi'''(t^*)/6, \tag{B.32}$$

where $0 \leq t^* \leq t$. Note that $\Psi(0) = 0$ under H_0. From now on, I ignore integration area for the sake of notation. Also I define

$$\tilde{p}(j, \omega_j, z; f_j + tg_j) = D_a(\psi(f_j + tg_j))(j, \omega_j, z). \tag{B.33}$$

I first study $\Psi'(t)$.

$$\Psi'(t) = 2\sum \int\int \tilde{p}(j, \omega_j, z; f_j + tg_j)\frac{\partial \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t} b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x).$$

Since for all j, ω_j, z, $\tilde{p}(j, \omega_j, z; f_j) = 0$, under H_0 I have $\Psi'(0) = 0$.

$$\begin{aligned}\Psi''(t) =& 2\sum \int\int \left[\tilde{p}(j, \omega_j, z; f_j + tg_j)\frac{\partial^2 \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t^2} + \left\{\frac{\partial \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t}\right\}^2\right] \\ & \times b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x).\end{aligned}$$

Thus,

$$\Psi''(0) = 2\sum \int\int \left\{\frac{\partial \tilde{p}(j, \omega_j, z; f_j)}{\partial t}\right\}^2 b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x)$$

under H_0. To characterize the remainder term,

$$\begin{aligned}\Psi'''(t) =& 2\sum \int\int \left[\tilde{p}(j, \omega_j, z; f_j + tg_j)\frac{\partial^3 \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t^3}\right. \\ & \left. + 3\frac{\partial \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t}\frac{\partial^2 \tilde{p}(j, \omega_j, z; f_j + tg_j)}{\partial t^2}\right] \\ & \times b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x).\end{aligned}$$

From Equation (B.33) and the definition of $D_a(\psi)(\cdot)$, $\tilde{p}(\cdot)$ is a sum of two $\psi(\cdot; f_j + tg_j)$'s.

I only need to focus on one of them to characterize the remainder term. I use notation $\psi(f_j + tg_j)$ to denote $\psi(j, \omega_j, z; f_j + tg_j)$. By quotient rule, I can derive the first, second, and third derivative of $\psi(f_j + tg_j)$ with respect to t. Further, I use $den(\psi(f_j + tg_j))$ to denote the denominator of $\psi(f_j + tg_j)$ and $num(\psi(f_j + tg_j))$ to denote the numerator of $\psi(f_j + tg_j)$.

$$\frac{\partial \psi(f_j + tg_j)}{\partial t} = \frac{den(\psi(f_j + tg_j))\frac{\partial num(\psi(f_j+tg_j))}{\partial t} - num(\psi(f_j + tg_j))\frac{\partial den(\psi(f_j+tg_j))}{\partial t}}{den(\psi(f_j + tg_j))^2} \tag{B.34}$$

$$\begin{aligned}\frac{\partial^2 \psi(f_j + tg_j)}{\partial t^2} =& \frac{\frac{\partial^2 num(\psi(f_j+tg_j))}{\partial t^2}}{den(\psi(f_j + tg_j))} + 2\frac{num(\psi(f_j + tg_j))\left(\frac{\partial den(\psi(f_j+tg_j))}{\partial t}\right)^2}{den(\psi(f_j + tg_j))^3} \\ &- \frac{num(\psi(f_j + tg_j))\frac{\partial^2 den(\psi(f_j+tg_j))}{\partial t^2}}{den(\psi(f_j + tg_j))^2}\end{aligned} \tag{B.35}$$

$$\begin{aligned}\frac{\partial^3 \psi(f_j + tg_j)}{\partial t^3} =& \frac{\frac{\partial^3 num(\psi(f_j+tg_j))}{\partial t^3}}{den(\psi(f_j + tg_j))} - 6\frac{num(\psi(f_j + tg_j))\left(\frac{\partial den(\psi(f_j+tg_j))}{\partial t}\right)^3}{den(\psi(f_j + tg_j))^4} \\ &+ 2\frac{\frac{\partial num(\psi(f_j+tg_j))}{\partial t}\left(\frac{\partial den(\psi(f_j+tg_j))}{\partial t}\right)^2}{den(\psi(f_j + tg_j))^3} - \frac{\frac{\partial^2 num(\psi(f_j+tg_j))}{\partial t^2}\frac{\partial den(\psi(f_j+tg_j))}{\partial t}}{den(\psi(f_j + tg_j))^2} \\ &+ 6\frac{num(\psi(f_j + tg_j))\frac{\partial den(\psi(f_j+tg_j))}{\partial t}\frac{\frac{\partial^2 den(\psi(f_j+tg_j))}{\partial t^2}}{\partial t}}{den(\psi(f_j + tg_j))^3} \\ &- \frac{\frac{\partial num(\psi(f_j+tg_j))}{\partial t}\frac{\partial^2 den(\psi(f_j+tg_j))}{\partial t^2}}{den(\psi(f_j + tg_j))^2} - \frac{num(\psi(f_j + tg_j))\frac{\partial^3 den(\psi(f_j+tg_j))}{\partial t^3}}{den(\psi(f_j + tg_j))^2}.\end{aligned} \tag{B.36}$$

I argue that there is constant C_{13} such that $\psi(f_j+tg_j)$ is bounded by C_{13}; Equation (B.34) is bounded by $C_{13}(\|g_{\boldsymbol{y}_j}\|_\infty + \|g_{\boldsymbol{x}}\|_\infty)$; Equation (B.35) is bounded by $C_{13}(\|g_{\boldsymbol{y}_j}\|_\infty^2 + \|g_{\boldsymbol{x}}\|_\infty^2)$; and Equation (B.36) is bounded by $C_{13}(\|g_{\boldsymbol{y}_j}\|_\infty^3 + \|g_{\boldsymbol{x}}\|_\infty^3)$. First, by Assumption 2.5, I can show that $num(\psi(f_j + tg_j))$ and $den(\psi(f_j + tg_j))$ are bounded above and below. Then, I need to study the first, second, and third order derivatives of $num(\psi(f_j + tg_j))$ (and $den(\psi(f_j + tg_j))$) with respect to t. By construction, $den(\psi(f_j + tg_j))$ is linear in

$(f_{\boldsymbol{y}_j}+tg_{\boldsymbol{y}_j})/(f_{\boldsymbol{x}}+tg_{\boldsymbol{x}})$'s. Hence, I only need to study the derivatives of $(f_{\boldsymbol{y}_j}+tg_{\boldsymbol{y}_j})/(f_{\boldsymbol{x}}+tg_{\boldsymbol{x}})$ with respect to t, which is shown below.

$$\begin{aligned}
\partial \frac{f_{\boldsymbol{y}_j}+tg_{\boldsymbol{y}_j}}{f_{\boldsymbol{x}}+tg_{\boldsymbol{x}}}/\partial t =& \frac{f_{\boldsymbol{x}}(x)g_{\boldsymbol{y}_j}(1,x)-f_{\boldsymbol{y}_j}(1,x)g_{\boldsymbol{x}}(x)}{\{f_{\boldsymbol{x}}(x)+tg_{\boldsymbol{x}}(x)\}^2} \qquad \text{(B.37)}\\
\partial^2 \frac{f_{\boldsymbol{y}_j}+tg_{\boldsymbol{y}_j}}{f_{\boldsymbol{x}}+tg_{\boldsymbol{x}}}/\partial t^2 =& 2\frac{g_{\boldsymbol{x}}(x)\{f_{\boldsymbol{y}_j}(1,x)g_{\boldsymbol{x}}(x)-f_{\boldsymbol{x}}(x)g_{\boldsymbol{y}_j}(1,x)\}}{\{f_{\boldsymbol{x}}(x)+tg_{\boldsymbol{x}}(x)\}^3} \qquad \text{(B.38)}\\
\partial^3 \frac{f_{\boldsymbol{y}_j}+tg_{\boldsymbol{y}_j}}{f_{\boldsymbol{x}}+tg_{\boldsymbol{x}}}/\partial t^3 =& 6\frac{g_{\boldsymbol{x}}(x)^2\{f_{\boldsymbol{x}}(x)g_{\boldsymbol{y}_j}(1,x)-f_{\boldsymbol{y}_j}(1,x)g_{\boldsymbol{x}}(x)\}}{\{f_{\boldsymbol{x}}(x)+tg_{\boldsymbol{x}}(x)\}^4}. \qquad \text{(B.39)}
\end{aligned}$$

Equations (B.37)–(B.39) are equivalent to Equations (B.6)–(B.8). Thus, denominators from Equations (B.37)–(B.39) are bounded below and $f_{\boldsymbol{y}_j}$ and $f_{\boldsymbol{x}}$ is bounded below and above. Putting Equations (B.34)–(B.36) and (B.37)–(B.39) yield the result.

Thus, I can find constant C_{14} such that for all $t^* \in [0,1]$, $\Psi'''(t^*) \leq C_{14}\sum_{j\in\mathcal{J}\setminus 0}(\|g_{\boldsymbol{y}_j}\|_\infty+\|g_{\boldsymbol{x}}\|_\infty)^3$. This shows that I have $\Psi'''(t^*) = O(\sum_{j\in\mathcal{J}\setminus 0}\|g_{\boldsymbol{y}_j}\|_\infty^3+\|g_{\boldsymbol{x}}\|_\infty^3)$.

Now, I derive the functional form of $\frac{\partial\tilde{p}(j,\omega_j,z;f_j+tg_j)}{\partial t}\Big|_{t=0}$, this can be derived from Equations (B.34) and (B.37) by letting $t=0$. I can rewrite Equation (B.34) at $t=0$

$$\partial\psi(f_j+tg_j)/\partial t\Big|_{t=0} = \psi(f_j)\left(\frac{\frac{\partial num(\psi(f_j+tg_j))}{\partial t}\Big|_{t=0}}{num(\psi(f_j))} - \frac{\frac{\partial den(\psi(f_j+tg_j))}{\partial t}\Big|_{t=0}}{den(\psi(f_j))}\right).$$

Also, I can rewrite Equation (B.37) at $t=0$ as

$$\begin{aligned}
\partial \frac{f_{\boldsymbol{y}_j}+tg_{\boldsymbol{y}_j}}{f_{\boldsymbol{x}}+tg_{\boldsymbol{x}}}/\partial t\Big|_{t=0} =& \frac{f_{\boldsymbol{x}}(x)g_{\boldsymbol{y}_j}(1,x)-f_{\boldsymbol{y}_j}(1,x)g_{\boldsymbol{x}}(x)}{f_{\boldsymbol{x}}(x)^2}\\
=& p(j|x)\left(\frac{g_{\boldsymbol{y}_j}(1,x)}{f_{\boldsymbol{y}_j}(1,x)} - \frac{g_{\boldsymbol{x}}(x)}{f_{\boldsymbol{x}}(x)}\right).
\end{aligned}$$

I define

$$\rho_{g_j}(j,\omega_j,z) = \frac{\Delta_{w'_J}^{w_J}\left[\cdots\Delta_{w'_{j+1}}^{w_{j+1}}\left[\Delta_{w'_{j-1}}^{w_{j-1}}\left[\cdots[\Delta_{w'_1}^{w_1}\kappa_{g_j}(j,w_j,\cdot,z)]\right]\right]\right]}{\Delta_{w'_J}^{w_J}\left[\cdots\Delta_{w'_{j+1}}^{w_{j+1}}\left[\Delta_{w'_{j-1}}^{w_{j-1}}\left[\cdots[\Delta_{w'_1}^{w_1}p(j|w_j,\cdot,z)]\right]\right]\right]} - \frac{\Delta_{\underline{w}_J}^{\bar{w}_J}\left[\cdots\Delta_{\underline{w}_{j+1}}^{\bar{w}_{j+1}}\left[\Delta_{\underline{w}_{j-1}}^{\bar{w}_{j-1}}\left[\cdots[\Delta_{\underline{w}_1}^{\bar{w}_1}\kappa_{g_j}(j,w_j,\cdot,z)]\right]\right]\right]}{\Delta_{\underline{w}_J}^{\bar{w}_J}\left[\cdots\Delta_{\underline{w}_{j+1}}^{\bar{w}_{j+1}}\left[\Delta_{\underline{w}_{j-1}}^{\bar{w}_{j-1}}\left[\cdots[\Delta_{\underline{w}_1}^{\bar{w}_1}p(j|w_j,\cdot,z)]\right]\right]\right]}$$

with

$$\kappa_{g_j}(j,w_j,w_{-j},z) = p(j|w_j,w_{-j},z)\left(\frac{g_{\boldsymbol{y}_j}(1,w_j,w_{-j},z)}{f_{\boldsymbol{y}_j}(1,w_j,w_{-j},z)} - \frac{g_{\boldsymbol{x}}(w_j,w_{-j},z)}{f_{\boldsymbol{x}}(w_j,w_{-j},z)}\right).$$

Then, I have

$$\left.\frac{\partial\psi(j,\omega_j,z;f_j+tg_j)}{\partial t}\right|_{t=0} = \psi(j,\omega_j,z)\rho_{g_j}(j,\omega_j,z). \tag{B.40}$$

Plug Equation (B.40) back into $\partial\tilde{p}(\cdot;f_j+tg_j)/\partial t\big|_{t=0}$, then I have the result. □

Lemma B2: If Assumptions 2.1–2.11 and H_0 are satisfied, then I have for any cdf $F_{\boldsymbol{x}}$,

$$\Gamma(\hat{f},F_{\boldsymbol{x}}) = \sum_{j\in\mathcal{J}\backslash 0}\int_{\mathcal{X}}\int_{\underline{w}_{-j}}^{w_{-j}} D_a(\psi\rho_{\hat{f}_j})(j,\omega_j,z)^2 b(x)\,dw'_{-j}\,dF_{\boldsymbol{x}}(x) + O_p\left(\sum_{j\in\mathcal{J}\backslash 0}\|\hat{f}_{\boldsymbol{y}_j}(1,x)-f_{\boldsymbol{y}_j}(1,x)\|_\infty^3 + \|\hat{f}_{\boldsymbol{x}}(x)-f_{\boldsymbol{x}}(x)\|_\infty^3\right).$$

Proof: As shown in the proof of Lemma B3, let

$$g_{\boldsymbol{y}_j} = \hat{f}_{\boldsymbol{y}_j} - f_{\boldsymbol{y}_j} \text{ and } g_{\boldsymbol{x}} = \hat{f}_{\boldsymbol{x}} - f_{\boldsymbol{x}}.$$

Then, by uniform consistency of Nadaraya-Watson estimators, the result follows. □

I define new notation. Let

$$\begin{aligned} I_n &= \sum_{j \in \mathcal{J} \setminus 0} \int_{\mathcal{X}} \int_{\underline{w}_{-j}}^{w_{-j}} D_a(\psi \rho_{\hat{f}_j})(j, \omega_j, z)^2 b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x) \\ &= \sum_{j \in \mathcal{J} \setminus 0} \int_{\mathcal{X}} \int_{\underline{w}_{-j}}^{w_{-j}} r_{\hat{f}_j}(j, \omega_j, z)^2 b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x). \end{aligned}$$

I will suppress the integration area to simplify notation. Throughout the chapter, unless I specify the integration area, it remains fixed as defined above. Then,

$$\begin{aligned} I_n =& \sum \int \int \{r_{\hat{f}_j}(j, \omega_j, z) - \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}^2 b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x) \\ &+ 2 \sum \int \int \{r_{\hat{f}_j}(j, \omega_j, z) - \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\} \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)] b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x) \\ &+ \sum \int \int \{\mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}^2 b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x) \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[I_n] =& \sum \int \int \mathbb{E}\{r_{\hat{f}_j}(j, \omega_j, z) - \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}^2 b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x) \\ &+ \sum \int \int \{\mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}^2 b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x). \end{aligned}$$

Hence,

$$\begin{aligned} I_n - \mathbb{E}[I_n] =& \sum \int \int \{r_{\hat{f}_j}(j, \omega_j, z) - \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}^2 b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x) \\ &- \sum \int \int \mathbb{E}\{r_{\hat{f}_j}(j, \omega_j, z) - \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\}^2 b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x) \\ &+ 2 \sum \int \int \{r_{\hat{f}_j}(j, \omega_j, z) - \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)]\} \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)] b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x). \end{aligned}$$

Throughout the chapter, I denote $\boldsymbol{y}_j \in \mathbb{R}^{1+\delta}$, $\boldsymbol{s}_j = \tilde{\boldsymbol{y}}_j \in \mathbb{R}^{1+\delta}$, and $\boldsymbol{t}_j = \tilde{\tilde{\boldsymbol{y}}}_j \in \mathbb{R}^{1+\delta}$ and

$\boldsymbol{s} = (\tilde{\boldsymbol{d}}_1, \cdots, \tilde{\boldsymbol{d}}_J, \tilde{\boldsymbol{x}})$ and $\boldsymbol{t} = (\tilde{\tilde{\boldsymbol{d}}}_1, \cdots, \tilde{\tilde{\boldsymbol{d}}}_J, \tilde{\tilde{\boldsymbol{x}}})$.

$$\kappa(j, x, \boldsymbol{s}_j) = p(j|x) K_h(\tilde{\boldsymbol{x}} - x) \left(\frac{\tilde{\boldsymbol{d}}_j}{f_{\boldsymbol{y}_j}(1, x)} - \frac{1}{f_{\boldsymbol{x}}(x)} \right),$$

and

$$R^*(j, \omega_j, z, \boldsymbol{s}_j) = \frac{\Delta_{w'_J}^{w_J}\left[\cdots\left[\Delta_{w'_{j+1}}^{w_{j+1}}\left[\Delta_{w'_{j-1}}^{w_{j-1}}\left[\cdots[\Delta_{w'_1}^{w_1}\kappa(j, w_j, \cdot, z, \boldsymbol{s}_j)]\right]\right]\right]\right]}{\Delta_{w'_J}^{w_J}\left[\cdots\left[\Delta_{w'_{j+1}}^{w_{j+1}}\left[\Delta_{w'_{j-1}}^{w_{j-1}}\left[\cdots[\Delta_{w'_1}^{w_1}p(j|w_j, \cdot, z)]\right]\right]\right]\right]}. \tag{B.41}$$

I further define

$$\begin{aligned} R(j, \omega_j, z, \boldsymbol{s}_j) =& \psi(j, \omega_j, z)\{R^*(j, \omega_j, z, \boldsymbol{s}_j) - R^*(j, \bar{\omega}_j, z, \boldsymbol{s}_j)\} \\ & - \int_{\mathcal{Z}} \psi(j, \omega_j, z)\{R^*(j, \omega_j, z, \boldsymbol{s}_j) - R^*(j, \bar{\omega}_j, z, \boldsymbol{s}_j)\} a(z')\, dz', \end{aligned} \tag{B.42}$$

and

$$\tilde{R}(j, \omega_j, z, \boldsymbol{s}_j) = R(j, \omega_j, z, \boldsymbol{s}_j) - \mathbb{E}[R(j, \omega_j, z, \boldsymbol{s}_j)]. \tag{B.43}$$

Let $J_n(\boldsymbol{s}) = \sum \int \int \tilde{R}(j, \omega_j, z, \boldsymbol{s}_j) h^{-r} \mathbb{E}[r_{\hat{f}_j}(j, \omega_j, z)] b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x)$ and $H_n(\boldsymbol{s}, \boldsymbol{t}) = h^{\delta/2} \sum \int \int \tilde{R}(j, \omega_j, z, \boldsymbol{s}_j) \tilde{R}(j, \omega_j, z, \boldsymbol{t}_j) b(x)\, dw'_{-j}\, dF_{\boldsymbol{x}}(x)$. Then,

$$\begin{aligned} I_n - \mathbb{E}[I_n] =& 2n^{-1/2} h^r \left\{ n^{-1/2} \sum_{i=1}^n J_n(y_i) \right\} \\ & + 2n^{-1} h^{-\delta/2} \left\{ n^{-1} \sum_{1 \le i < k \le n} [H_n(y_i, y_k) - \mathbb{E} H_n(\boldsymbol{y}_i, \boldsymbol{y}_k)] \right\} \\ & + n^{-1} h^{-\delta/2} \left\{ n^{-1} \sum_{i=1}^n [H_n(y_i, y_i) - \mathbb{E} H_n(\boldsymbol{y}_i, \boldsymbol{y}_i)] \right\} \\ =& 2n^{-1/2} h^r U_{n,1} + 2n^{-1} h^{-\delta/2} U_{n,2} + n^{-1} h^{-\delta/2} U_{n,3}. \end{aligned}$$

Notice that the above equation is identical to the two-product case. Now, I argue that the proofs of Lemmas B3–B6 are directly applicable to the general case. As before, I need to study asymptotic properties of $U_{n,1}$ and $U_{n,2}$. I first argue that

$$h^{-r}\mathbb{E}[R(j,\omega_j,z,\boldsymbol{s}_j)] < \infty.$$

From Equations (B.41) and (B.42) and by construction of $\kappa(\cdot)$, it is easy to see that $\kappa(\cdot)$ enters linearly in $R(\cdot)$. Since non-kernel functions from $R(\cdot)$ is bounded above and below by Assumptions 2.5 and 2.9, I only need to show that

$$h^{-r}\mathbb{E}[\kappa(j,x,\boldsymbol{s}_j)] < \infty$$

which is shown in the proof of Lemma B3.

Second, I only need to focus on $\gamma(\cdot)$ and $j=1$, where

$$\gamma(j,\omega_j,z,\tilde{\boldsymbol{x}}) = \psi(j,\omega_j,z)\frac{\kappa_{\boldsymbol{x}}(j,x,\tilde{\boldsymbol{x}})}{\Delta_{w_J'}^{w_J}\left[\cdots\left[\Delta_{w_{j+1}'}^{w_{j+1}}\left[\Delta_{w_{j-1}'}^{w_{j-1}}\left[\cdots[\Delta_{w_1'}^{w_1}p(j|w_j,\cdot,z)]\right]\right]\right]\right]},$$

and

$$\kappa_{\boldsymbol{x}}(j,x,\tilde{\boldsymbol{x}}) = p(j|x)\frac{K_h(\tilde{\boldsymbol{x}}-x)}{f_{\boldsymbol{x}}(x)}$$

because studying asymptotic properties of $U_{n,1}$ and $U_{n,2}$ concerns orders of expectation as n goes to ∞ as shown in the proofs of Lemmas B3–B6. I argue that $\gamma(\cdot)$ is a dominating term in those expectations. As discussed before Lemma B3, $\gamma(\cdot)$ dominates terms with $\tilde{\boldsymbol{d}}_j$ and it enters $R(\cdot)$ linearly. These allow me to to focus on one $\gamma(\cdot)$. Furthermore, I focus on $j=1$ since non-kernel functions such as $f_{\boldsymbol{y}_j}$, $f_{\boldsymbol{x}}$, and $p(j|\cdot)$ are bounded above and below regardless of j. Now, $\gamma(\cdot)$ defined here and in the two-product case look identical other than non-kernel functions, which are bounded above and below. Since Lemmas B3–B6 are proved using $\gamma(\cdot)$ only, I can directly apply the lemmas to the general case.

Then, I can show that $U_{n,1} = O_p(1)$, $U_{n,2}/\sigma \xrightarrow{d} \mathcal{N}(0,1/2)$, where $\sigma^2 = \mathbb{E}[H_n(\boldsymbol{y}_i,\boldsymbol{y}_k)^2]$, and $U_{n,3} = O_p(n^{-1/2}h^{-\delta/2}) = o_p(1)$. Furthermore, $\mathbb{E}[I_n] = B + o(1)$, where $B = \mathbb{E}[H_n(\boldsymbol{y}_i, \boldsymbol{y}_i)]$. Last, define $\Delta_n = \Gamma(\hat{f}, \hat{F}_{\boldsymbol{x}}) - \Gamma(\hat{f}, F_{\boldsymbol{x}})$ and $nh^{\delta/2}\Delta_n = o_p(1)$.

Furthermore, Lemma B7 holds in the general case for any $j \in \mathcal{J}\backslash 0$ by Mack and Silverman (1982) under Assumptions 2.9–2.11. This proves Theorem 2.3. □

Proof of consistency of estimator: Similar to the proof in the two-product case. I only need to replace $R(j, w_j, w_{j'}, w'_{j'}, z, \boldsymbol{y}_j)$ with $R(j, w_j, w_{-j}, w'_{-j}, z, \boldsymbol{y}_j)$ and likewise for $\hat{R}$. □

Proof of Theorem 2.4: Similar to the proof of Theorem 2.2. I only need to replace

$$\|\hat{f}_{\boldsymbol{y}_1}(1,x) - f_{\boldsymbol{y}_1}(1,x)\|_\infty + \|\hat{f}_{\boldsymbol{y}_2}(1,x) - f_{\boldsymbol{y}_2}(1,x)\|_\infty + \|\hat{f}_{\boldsymbol{x}}(x) - f_{\boldsymbol{x}}(x)\|_\infty$$

with

$$\sum_{j\in\mathcal{J}\backslash 0} \|\hat{f}_{\boldsymbol{y}_j}(1,x) - f_{\boldsymbol{y}_j}(1,x)\|_\infty + \|\hat{f}_{\boldsymbol{x}}(x) - f_{\boldsymbol{x}}(x)\|_\infty.$$

Then, the result follows. □

Appendix C

Omitted Case from Chapter 2

In this appendix, I present the omitted case from chapter 2. That is, I consider the case where the econometrician wants to test if a candidate variable indeed only affects attention when a variable that governs preferences only is available. In other words, I test if $\boldsymbol{w}$ affects preferences when it is known that $\boldsymbol{z}$ only affects preferences. In section C.1, I describe the hypothesis for the omitted case. Section C.2 illustrates required assumptions to construct the test. Section C.3 shows how to construct the test function for the omitted case. Once constructing the test function, the remainder parts (the lemma, the test statistic, and the theorems) are identical as the main text. Thus, I omit them here.

C.1 Hypothesis

In this section, I illustrate the hypothesis for my test.

$$H_0: \ \forall j \in \mathcal{J}, \ \Pr\left[p(j|\boldsymbol{w}, \boldsymbol{z}, \mathcal{J}) = p(j|\boldsymbol{z}, \mathcal{J})\right] = 1.$$

$$H_A: \ \exists j \in \mathcal{J}, \ \Pr\left[p(j|\boldsymbol{w}, \boldsymbol{z}, \mathcal{J}) = p(j|\boldsymbol{z}, \mathcal{J})\right] < 1.$$

Also, my hypothesis only relies on $C = \mathcal{J}$ (012 in this section). The hypothesis tests whether $\boldsymbol{w}$ influences preferences when the consumer considers all products. Assume that $\boldsymbol{w}_j$ for any $C \in \mathcal{C}$ changes the consumer's preference for product j in C. According to Assumption 2.2, her preference for product j must be independent of consideration set C. This means that even if the consideration set is $\mathcal{J}$, $\boldsymbol{w}_j$ still changes the preference for product j. Therefore, testing in $C = \mathcal{J}$ is sufficient.

C.2 Assumptions

In this section, I state the assumptions. Some assumptions in the main text are still required for the omitted case, some have to be replaced. I only state the assumptions that need replacing in this section. When I replace assumptions with new ones, I use the the asterisk. For instance, if I replace Assumption 2.3 with new one, I state it as Assumption 2.3*.

Assumption 2.3*: If $j' \notin C$, then the consumer's preferences for products in C does not depend on $\boldsymbol{x}_{j'}$.

Assumption 2.3* says that if her consideration set is 01, then $\boldsymbol{x}_2$ has nothing to do with the preference order between products 0 and 1. The idea behind Assumption 2.3* is that if the consumer does not pay attention to product 2, she will completely ignore product 2. Therefore, the characteristics of product 2 do not play any role in preferences between products 0 and 1. As far as I know, Assumption 2.3* is often imposed in the literature. Assumption 2.3* is used to test whether $\boldsymbol{w}$ affects preferences.

Assumption 2.4*: $\forall C \in \mathcal{C}$, $p(C|\boldsymbol{x}) = p(C|\boldsymbol{w})$ with probability 1.

Assumption 2.4* describes a type 2 exclusion restriction – that is, $\boldsymbol{z}$ only affects preferences. Using Assumption 2.4*, I can construct the test function that depends on $p(1|x, 012)$ from Equation (2.1) in this section. Therefore, the hypothesis can be tested using Assumption

2.4*.

Assumption 2.5* (i) $\forall j \in \mathcal{J}\ \exists c < 0\ \forall w \in \mathcal{W}\ \forall z_j \in \mathcal{Z}_j\ \forall z_{-j} \in \text{int}(\mathcal{Z}_{-j})$, $\partial_{z_J} \cdots \partial_{z_{j+1}} \partial_{z_{j-1}} \cdots \partial_{z_1} p(j|w, z_j, z_{-j}, \mathcal{J}) \leq c < 0$; (ii)$\exists c > 0\ \forall w \in \mathcal{W}$, $p(\mathcal{J}|w) \geq c > 0$.

Assumption 2.5*–(i) is similar to Assumption 2.5–(i), stating that the consumer is less likely to choose product j when z_{-j} increases. This assumption is reasonable when, for example, z_j represents the quality of product j. Assumption 2.5*–(ii) states that the probability of considering all available products is positive.

Assumption 2.6* $p(j|w, z, \mathcal{J})$ is continuously differentiable in w and z.

Assumption 2.8* Under H_A, $\forall j \in \mathcal{J}$, $\forall z'_{-j}, z''_{-j} \in \text{int}(\mathcal{Z}_{-j})$ and $\forall c \in \mathbb{R}$, $\exists z_j \in \mathcal{Z}_j$ and $w \in \mathcal{W}$ such that $\partial_{z_J} \cdots \partial_{z_{j+1}} \partial_{z_{j-1}} \cdots \partial_{z_1} p(j|w, z_j, z'_{-j}, \mathcal{J}) \neq c \partial_{z_J} \cdots \partial_{z_{j+1}} \partial_{z_{j-1}} \cdots \partial_{z_1} p(j|w, z_j, z''_{-j}, \mathcal{J})$.

Just like Assumption 2.8, Assumption 2.8* excludes the case that $p(j|w, z, \mathcal{J})$ is additively and multiplicatively separable. The random utility model satisfies Assumption 2.8*. Assumptions 2.1, 2.2, 2.3*–2.5* are required to construct the test function in this appendix. Assumptions 2.6*, 2.7, and 2.8* are required for constructing the valid test – required for a lemma analogous to Lemma 2.2.

C.3 Test Function

Now, I show how to construct the test function from Assumptions 2.1, 2.2, 2.3*–2.5*. Under Assumptions 2.3* and 2.4*, Equation (2.5) can be written as

$$\begin{aligned} p(1|w, z_1, z_{-1}) = & \sum_{C' \in \mathcal{C}_1^{\mathcal{J}\setminus 2}} p(1|w, z_1, z_{-1}, C')p(C' \text{ or } C' \cup 2|w, z_1, z_{-1}) \\ & + \{p(1|w, z_1, z_{-1}, C' \cup 2) - p(1|w, z_1, z_{-1}, C')\}p(C' \cup 2|w, z_1, z_{-1}) \\ = & \sum_{C' \in \mathcal{C}_1^{\mathcal{J}\setminus 2}} p(1|w, z_1, z_{-1,-2}, C')p(C' \text{ or } C' \cup 2|w) \\ & + \{p(1|w, z_1, z_{-1}, C' \cup 2) - p(1|w, z_1, z_{-1,-2}, C')\}p(C' \cup 2|w). \end{aligned} \tag{C.1}$$

From Equation (C.1), the first equality is true as shown from Equation (2.6). The second equality is true by Assumptions 2.3* and 2.4*. From Equation (C.1), when I change z_2 to z_2' with $z_2' < z_2$, the only term that changes is $p(1|x, C' \cup 2)$. Then, I have

$$\Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot, z_{-1,-2}) = \sum_{C' \in \mathcal{C}_1^{\mathcal{J}\setminus 2}} p(C' \cup 2|w)\Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot, z_{-1,-2}, C' \cup 2). \tag{C.2}$$

I repeat the same process for Equation (C.2) as I performed on Equation (2.8). I can rewrite Equation (C.2) by putting $p(C'' \cup 2|w) = p(C'' \cup 2 \text{ or } C'' \cup 23|w) - p(C'' \cup 23|w)$ for $C'' \in \mathcal{C}_1^{\mathcal{J}\setminus 23}$ as

$$\begin{aligned} \Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot, z_{-1,-2}) = & \sum_{C'' \in \mathcal{C}_1^{\mathcal{J}\setminus 23}} p(C'' \cup 2 \text{ or } C'' \cup 23|w)\Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot, z_{-1,-2,-3}, C'' \cup 2) \\ & + [\Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot, z_{-1,-2}, C'' \cup 23) - \Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot, z_{-1,-2,-3}, C'' \cup 2)] \\ & \times p(C'' \cup 23|w). \end{aligned} \tag{C.3}$$

When I change z_3 to z_3' with $z_3' < z_3$ from Equation (C.3), the term that changes is $\Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot, z_{-1,-2}, C'' \cup 23)$. Then, I have

$$\Delta_{z_3'}^{z_3}[\Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot, z_{-1,-2,-3})] = \sum_{C'' \in \mathcal{C}_1^{\mathcal{J} \setminus 23}} p(C'' \cup 23|w)[\Delta_{z_3'}^{z_3}[\Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot, z_{-1,-2,-3}, C'' \cup 23)]]. \quad \text{(C.4)}$$

I do the same exercise on (C.4) recursively until only one summand is left as I perform on Equations (C.2). The resulting outcome is

$$\Delta_{z_J'}^{z_J}\left[\cdots[\Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot)]\right] = \Delta_{z_J'}^{z_J}\left[\cdots[\Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot, \mathcal{J})]\right] p(\mathcal{J}|w). \quad \text{(C.5)}$$

Again, to construct the valid test from Equation (C.5), I need the left hand side to be bounded away from zero. I can easily show that $\Delta_{z_J'}^{z_J}[\cdots[\Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot, \mathcal{J})]]$ is bounded away from zero by the mean value theorem, Assumption 2.5*–(i) and the choice of z_j' for all $j \neq 1$. By Assumption 2.5*–(ii), I have $p(\mathcal{J}|w) > c$ for some positive constant c.

Now, for any choice of $z \in \mathcal{Z}$ and $z_{-1}' \in \mathcal{Z}_{-1}$ with $z_{-1} > z_{-1}'$, I can construct test function $\lambda(\cdot)$

$$\begin{aligned}\lambda(1, w, z_1, z_{-1}, z_{-1}') &= \frac{\Delta_{z_J'}^{z_J}\left[\cdots[\Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot)]\right]}{\Delta_{\underline{z}_J}^{\bar{z}_J}\left[\cdots[\Delta_{\underline{z}_2}^{\bar{z}_2} p(1|w, z_1, \cdot)]\right]} \\ &= \frac{\Delta_{z_J'}^{z_J}\left[\cdots[\Delta_{z_2'}^{z_2} p(1|w, z_1, \cdot, \mathcal{J})]\right]}{\Delta_{\underline{z}_J}^{\bar{z}_J}\left[\cdots[\Delta_{\underline{z}_2}^{\bar{z}_2} p(1|w, z_1, \cdot, \mathcal{J})]\right]},\end{aligned}$$

where the second equality is directly from Equation (C.5).

Appendix D

Data Cleaning for Chapter 3

The raw data includes 9,917,530 observations. I follow Ursu (2018) and use five criteria to filter the data. The resulting data includes 299,595 observations.

First, I remove travel destinations that have at least one hotel with an average review score of NaN. NaN means review information is unavailable or missing. If there are no reviews left for a hotel, the data is padded with zeros.

Second, the data suffers from incorrect price information. For instance, one hotel charges more than 19 million USD per night while another hotel charges 0.01 USD per night. Following Ursu (2018), I focus on consumers who observe hotels whose prices are bounded between 10 USD and 1,000 USD.

Third, The data contains some consumers who have paid more than they should have paid. To mitigate this issue, I eliminate any consumers whose total payment for a hotel room exceeds the displayed price times the sum of the length of the trip and the number of requested rooms plus taxes. Following Ursu (2018), I set tax rate equals 30%. This filtration also mitigate the second problem.

Fourth, the raw data includes more than 20,000 travel destinations, with a median of two consumers per destination. I will suffer from the incidental parameter problem (Greene, 2004) when I estimate parameters of interest using MLE while allowing destination fixed effects. Thus, I focus on the destinations searched by more than 200 consumers assigned to the random ranking system.

Fifth, I eliminate all consumers who observe opaque offers. Some hotels are placed 5th, 11th, 17th, or 23rd on a ranking. If hotels are positioned at one of these four slots, then its characteristics are available to the consumer, except for price and star rating. Because of this nature, I remove consumers who observe one or more opaque offers. This filtering will not impact my analysis as opaque offers appear to be exogenous (Ursu, 2018).

Vita

Hyung Gyu Rho

Education

Ph.D. Economics (2022), Pennsylvania State University

B.Com.Hons Economics (2015), The University of Melbourne

B.Com Economics and Finance (2014), The University of Melbourne

Teaching

Teaching Assistant Pennsylvania State University (2016–2021)

Econometrics I (Econ 501, Ph.D.; Econ 510, Master)

Urban Economics (Econ 432)

Tutor The University of Melbourne (2015-2016)

Microeconomics (Econ 30019)

Microeconomics 2 (Econ 90045)

Economics of Financial Markets (Econ 30024)