Semiparametric Identification and Estimation of Substitution Patterns *†

Rui Wang‡

June 18, 2022

Abstract

This paper studies semiparametric identification of substitution patterns between two goods using a panel multinomial choice model with bundles. My model allows the two goods to be either substitutes or complements and admits heterogeneous complementarity through observed characteristics. I characterize the sharp identified set for the model parameters and provide sufficient conditions for point identification. My identification analysis accommodates endogenous covariates through flexible dependence structures between observed characteristics and fixed effects while placing no distributional assumptions on unobserved preference shocks. My method is shown to perform more robustly than the parametric method through Monte Carlo simulations. As an empirical illustration, I apply my method to estimate substitution patterns between cigarettes and e-cigarettes using the Nielsen data.

Keywords: Substitution Patterns, Semiparametric Identification, Bundles, Panel Multinomial Choice Model, Endogeneity

---

*I am grateful to Sung Jae Jun, Joris Pinkse, and Keisuke Hirano for their invaluable guidance and support. I thank Andres Aradillas-Lopez, Jeffrey Campbell, Liyu Dou, Paul Grieco, Patrik Guggenberger, Marc Henry, Ming Li, Xiao Lin, Xun Lu, Ewout Verriest, Jiahui Xu, Kaida Zhang, and many seminar participants for their helpful comments.

†Researcher's own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

‡Department of Economics, The Pennsylvania State University, State College, PA 16801. Email: `rxw84@psu.edu`.

1 Introduction

Substitution relationships between goods have been studied in many applications such as online news versus print newspapers, digital books versus traditional books, and cigarettes versus e-cigarettes. The relationship plays a crucial role in consumers' decisions; therefore, understanding substitution patterns is important for predicting demand for a good and analyzing the welfare effects of, for example, a merger of two companies or the introduction of a new good (Petrin, 2002; Goolsbee and Petrin, 2004; Gentzkow, 2007; Liu, Chintagunta, and Zhu, 2010).

The standard multinomial choice models typically assume that consumers can buy only one good at a time, which implies that goods are substitutes. However, many research papers suggest that some goods traditionally perceived as substitutes may in fact be complements. For example, Zhao (2019) suggests that cigarettes and e-cigarettes are complements, and Grzybowski and Pereira (2008) show the complementarity between telephone calls and messages. Motivated by these findings, my paper uses a panel multinomial choice model with fixed effects that allows for bundles to study substitution patterns. This model allows consumers to purchase two goods simultaneously, accommodating the possibility that the two goods are either substitutes or complements. The model also permits heterogeneous complementarity relationships through observed characteristics.

Identifying substitution patterns in a model with bundles presents multiple challenges. First, the demand for one good involves consumers who buy this good alone and those who buy a bundle. Therefore, a large demand for one good could come from consumers' high utility for this good or its complementarity with another good. We need to disentangle the two sources to identify the complementarity relationship. Second, the purchase of two goods together may be due to either the goods' complementarity or the unobserved correlation between consumers' preferences over the two goods. For example, consumers may buy a variety of organic goods because of their preferences over organic goods instead of the complementarity between these goods. Distinguishing the complementarity relationship and the correlation between consumers' tastes for goods is challenging since they are both unobserved and can affect consumers' decisions simultaneously.

To tackle these challenges, my paper uses intertemporal variation in conditional choice probabilities for identification. In doing so, I exploit a conditional stationarity assumption about preference shocks over time, which requires the distribution of the preference shocks to be the same over any pair of two periods conditional on fixed effects and covariates. I derive the sharp identified set for the complementarity parameter and characterize sufficient conditions for point identification. There are two crucial features of my methodology.

First, the model allows for endogenous covariates by accommodating flexible dependence structures between observed characteristics and unobserved fixed effects. One example of endogenous covariates is the price of a product, which is potentially correlated with unobserved heterogeneity such as the quality of the product. Second, the analysis does not impose any parametric assumptions on the distributions of preference shocks, and it allows the shocks to be freely dependent across choices and over time. The simulation results show that misspecifications in distributional assumptions may lead to misleading estimators of substitution patterns.

The main strategy behind my identification analysis is to derive identifying restrictions for the model parameters based on intertemporal comparisons of conditional choice probabilities that can be identified from data. The analysis of substitution patterns consists of two parts. The first part entails identifying the sign of the substitution pattern based on variation in the demand for the two goods. The primary idea is to exploit the relationship between changes in the utility of one good and the demand for the other good. If the two goods are complements, then increasing the utility of one good (by decreasing its price, for example) will encourage consumers to buy the two goods together and thus increase the demand for the other good. When decreasing demand for either of the two goods is observed even as the two goods become more attractive to consumers, the two goods are identified as substitutes. It is because if they were complements, the demand for the two goods would increase.

Additionally, I derive bounds for the complementarity from the sum of conditional probabilities of two different choices over different periods. For example, when the sum of the conditional probabilities of buying two goods together and that of buying neither good is large, then a lower bound for the complementarity can be established. The idea is that when the price of good A increases, consumers will switch from buying two goods together to buying neither good if the two goods are complements, while they will switch to buying only good B if the two goods are substitutes. Therefore, when a high probability of buying two goods together and neither good is observed, we can infer the complementarity between the two goods and provide a lower bound for the complementarity. Similarly, I derive an upper bound for the complementarity parameter when the sum of the conditional probability of buying a single good over different periods is large.

The paper establishes the sharpness of the identification results, suggesting that the results have exhausted all useful information from the data for the model parameters. The strategy for proving sharpness is as follows. I first construct "choice sets," which are the collections of unobserved terms such that a single choice is selected. Then for any parameter satisfying the identifying conditions, I construct a conditional distribution on the choice sets such that the constructed distribution satisfies the model assumptions and matches the observed data, which proves the sharpness. The paper also provides sufficient conditions for point identification of the model parameters under large support conditions of the covariates and a linear specification of the complementarity.

I characterize conditional moment inequalities based on the identification results and apply a similar approach to Shi, Shum, and Song (2018) for estimation. In Monte Carlo simulations, I compare the finite sample performance of the method in this paper to that of a parametric method that assumes a parametric distribution over the error terms and a linear model for fixed effects. The simulation results show that the semiparametric method in this paper has the advantage of performing robustly over different DGP designs, whereas the parametric estimator performs poorly if either the parametric distribution or the linear model is incorrectly specified.

As an empirical illustration of this approach, I estimate the substitution pattern between cigarettes and e-cigarettes using the Nielsen Retail Scanner data. The data contain weekly store-level information about the prices and sales of cigarettes and e-cigarettes. The substitution pattern between the two goods is identified from comparisons of the conditional demand for the two goods over different weeks. The estimation results based on this data show that cigarettes and e-cigarettes are substitutes on average.

As an extension, I study a more general utility function that allows for nonseparability in characteristics, fixed effects, and error terms, where I assume monotonicity in the covariate index. Under this more general utility function, new identification results for the model parameters are established. I also develop a method to test the complementarity relationship between two goods by characterizing conditional moment inequalities under the null hypothesis of the presence of complementarity. Moreover, I allow for unobserved heterogeneity in the complementarity and provide partial identification analyses for the fraction of people for whom the two goods are complements.

1.1 Related Literature

This paper contributes to the literature studying substitution patterns in a discrete choice model with bundles. Gentzkow (2007) uses a model allowing for bundles to study substitution effects between online news and print newspapers. His paper allows for flexible substitution patterns and correlations between preference shocks across choices, but he assumes parametric distributions over unobserved fixed effects and error terms as well as exogeneity of covariates. Dunker, Hoderlein, and Kaido (2015) and Iaria and Wang

(2020) allow for endogeneity and provide identification results of models with bundles by extending the classic BLP approach in Berry, Levinsohn, and Pakes (1995). Their methods rely on demand inversion and parametric distributions over error terms, and they address endogeneity using instrumental variables. Monardo (2021) studies a more flexible model for the inverse demand and uses an instrument to construct moment conditions. My work allows for unknown distributions of both fixed effects and error terms so it does not specify a parametric model for the demand function (or inverse of the demand). Also, my paper mainly exploits intertemporal variation in panel data to derive identification results and this method does not require demand inversion or instrumental variables.

There are some papers that allow for unknown distributions of error terms to study substitution patterns. Fox and Lazzati (2017) study semiparametric identification of a discrete choice model with bundles under a large support assumption and exogenous covariates. I provide sharp identification with bounded support and allow for endogeneity. Allen and Rehbeck (2020) consider unobserved heterogeneous complementarity and provide partial identification for the fraction of people for whom the two goods are complements. My paper focuses on heterogeneous complementarity through observed covariates. My method can identify the sign and bound the value of the complementarity given consumers' characteristics by exploiting panel data. In an extension of this paper, I allow for unobserved heterogeneity in the complementarity relationship while relaxing the exogenous covariates assumption and the exclusion restriction in Allen and Rehbeck (2020).

My paper is also related to a large body of literature on panel multinomial choice models with fixed effects. Chamberlain (1980) provides a conditional fixed effect logit estimator for the panel multinomial choice model under a logistic distribution over disturbances. Manski (1987) as well as Honoré and Lewbel (2002) relax the logistic distribution assumption and study semiparametric identification of a binary choice model. Manski (1987) uses a maximum score approach that relies on a group stationarity assumption, and Honoré and Lewbel (2002) exploit the idea of a special regressor to identify the panel binary choice model.

Pakes and Porter (2019) and Shi, Shum, and Song (2018) extend a binary choice model to a multinomial choice model. Pakes and Porter (2019) derive sharp identification of the model by characterizing conditional moment inequalities, while Shi, Shum, and Song (2018) use cyclic monotonicity for identification and estimation. Gao and Li (2020) relax the separable utility function assumption in the previous papers and study a class of nonseparable utility functions. The above mentioned papers focus on the identification of own price coefficients rather than substitution patterns between different goods. My paper builds on this literature to allow for bundles in the panel multinomial choice model and characterizes sharp identification for substitution patterns.

The rest of this paper is organized as follows. Section 2 introduces the panel multinomial choice model with bundles. Section 3 characterizes the sharp identified set for the model parameters and provides sufficient conditions for point identification. Section 4 develops conditional moment inequalities, and Section 5 examines the finite sample performance of this method via Monte Carlo simulations. Section 6 studies substitution patterns between cigarettes and e-cigarettes as an empirical illustration. Section 7 discusses some extensions of the model, such as nonseparable utility functions and latent complementarity. Section 8 concludes.

2 Panel Multinomial Choice Model

This section presents a panel multinomial choice model allowing for bundles. Consider a short-panel structure: let $i \in \mathcal{I}$ denote consumers and $t \leq T$ denote time periods where the length of the panel $T \geq 2$ is fixed. Since this paper focuses on substitution patterns between two goods, I consider the case of two goods: $\{A, B\}$. Instead of assuming that consumers can buy either only good A or only good B, this model allows consumers to purchase goods A and B simultaneously. The possibility of buying the two goods together allows the two goods to be either substitutes or complements.

The choice set for consumers is $\mathcal{C} = \{A, B, AB, O\}$, where A (or B) denotes purchasing only good A (or B), AB denotes purchasing A and B simultaneously within a single period, and O denotes the outside option. I assume that consumers buy at most one unit of each good, and they select the choice yielding the highest utility in their choice set.

Consumers' utility of a single good has three key components. Let $X_{ijt} \in \mathbb{R}^{d_x}$ denote a vector of observed characteristics, which may include consumer i's characteristics (e.g., income), product j's characteristics (e.g., price), and the interaction terms between them. Let $\alpha_{ij} \in \mathbb{R}$ denote an unobserved individual-specific fixed effect for product j that does not change over time, such as consumers' loyalty to a brand. Let $\epsilon_{ijt} \in \mathbb{R}$ denote an unobserved and time-varying shock that affects consumers' utility over time.

To specify the utility of the choice AB, let Γ_{it} denote the incremental utility from consuming the bundle AB compared to the sum of utilities of consuming goods A and B alone. The sign of Γ_{it} captures the complementarity relationship between the two goods. I discuss later the relationship between the sign of Γ_{it} and an alternative definition of substitution patterns using aggregate demand.

The utility u_{ijt} of consumer i from consuming choice $j \in \mathcal{C}$ at time t is specified as

$$\begin{aligned} u_{iAt} &= X_{iAt}'\beta_0 + \alpha_{iA} + \epsilon_{iAt}, \\ u_{iBt} &= X_{iBt}'\beta_0 + \alpha_{iB} + \epsilon_{iBt}, \\ u_{iABt} &= u_{iAt} + u_{iBt} + \Gamma_{it}, \\ u_{iOt} &= 0, \end{aligned} \tag{1}$$

where $\beta_0 \in \mathbb{R}^{d_x}$ denotes a finite-dimensional unknown parameter vector.

Without loss of generality, the utility of the outside option is normalized to zero so the utility of the remaining choices is defined relative to the utility of the outside option. In this paper, I focus on an additive and separable utility function that is commonly used in the literature. I also study a class of nonseparable utility functions in the extension of this paper and derive the identified set for the model parameters. For simplicity of notation, the coefficient β_0 is assumed to be the same for the two goods. The analysis can generalize to the case in which the coefficients of the two goods are different.

In addition to the covariate X_{ijt}, I assume that consumer i's choice at time t is observed which is denoted as $Y_{it} \in \mathcal{C}$. Consumers select the choice with the highest utility, implying

$$Y_{it} = j \implies u_{ijt} \geq u_{ikt} \text{ for all } k \in \mathcal{C}.$$

When ties between choices happen with nonzero probability, I use a simple selection rule whereby consumers randomly select a choice with a fixed (potentially unknown) probability. The main objective of this paper is to discover the complementarity relationship between goods A and B from consumers' choices Y_{it} and observed covariates X_{ijt}.

The standard discrete choice models assume that consumers can only purchase one good and focus on identifying the coefficient β_0. This is equivalent to imposing a restriction on Γ_{it}: $\Gamma_{it} = -\infty$, which restricts the two goods to be substitutes. In this paper, Γ_{it} can be either positive, negative, or zero, which allows for the possibility that two goods can be either substitutes or complements. My paper derives the sharp identified set for both the coefficient β_0 and the complementarity relationship Γ_{it} between the two goods.

Next, I introduce some assumptions for model (1).

Assumption 1. *The incremental term Γ_{it} in the utility u_{iABt} is specified as*

$$\Gamma_{it} = \Gamma(Z_i, \gamma_0),$$

where the function Γ is known up to a finite-dimensional parameter γ_0, and $Z_i \in \mathbb{R}^{d_z}$

denotes a vector of observed characteristics.

Assumption 1 allows the complementarity Γ_{it} to depend on observed covariates Z_i in a parametric function. The function Γ is flexible and can be nonlinear in the covariate Z_i, which can admit rich complementarity patterns. The covariate Z_i may include consumers' characteristics such as income and age so that Assumption 1 allows for heterogeneous complementarity relationships through observed characteristics. For simplicity of notation, I consider that the covariate Z_i is fixed over time for any t. If the covariate changes over time, then the analysis can be conducted conditional on the same value of the covariate over time: $Z_{is} = Z_{it} = z$.

One restriction of Assumption 1 is that it excludes unobserved heterogeneity in the complementarity Γ_{it}. In an extension, I discuss the case in which Γ_{it} is a random variable such that it incorporates unobserved heterogeneity into the complementarity relationship. In this scenario, the distribution of the sign of Γ_{it} is partially identified. Under Assumption 1, I establish more informative results that not only identify the sign of the complementarity relationship but also bound the magnitude of the complementarity $\Gamma(z, \gamma_0)$ given $Z_i = z$. The more informative results are useful for many analyses, such as the effect of introducing a new good.

Assumption 2 (Exclusion)**.** *There exists at least one characteristic X^*_{it} in $X_{it} = (X_{iAt}, X_{iBt})$ that is not in Z_i, and its coefficient is nonzero.*

The exclusion assumption requires that there exists one variable that only influences the utility for good A or B but not the complementarity between the two goods. One example of this variable is the price of good A or B, which affects the utility of a single good but may not influence the complementarity between the two goods. The sign of the coefficient for X^*_{it} can be still unknown to researchers. Moreover, this assumption does not restrict the covariate Z_i; any variable affecting the complementarity is allowed to influence the utility of a single good.

The last assumption is the stationarity condition for the distribution of the unobserved shocks. Let $X_{it} = (X_{iAt}, X_{iBt}), \alpha_i = (\alpha_{iA}, \alpha_{iB})$, and $\epsilon_{it} = (\epsilon_{iAt}, \epsilon_{iBt})$ collect covariates, fixed effects, and error terms of the two goods.

Assumption 3. *(Stationarity) The distribution of ϵ_{it} conditional on $(X_{is}, X_{it}, Z_i, \alpha_i)$ is stationary over time; that is,*

$$\epsilon_{is} \mid X_{is}, X_{it}, Z_i, \alpha_i \overset{d}{\sim} \epsilon_{it} \mid X_{is}, X_{it}, Z_i, \alpha_i \quad \textit{for any } s, t \leq T.$$

This assumption is a multinomial extension of the conditional homogeneity assumption in Manski (1987). It is commonly used in the literature on panel multinomial choice models, including Pakes and Porter (2019) and Shi, Shum, and Song (2018), which study identification of the coefficient β_0 under this assumption. Assumption 3 restricts the conditional distribution of ϵ_{it} to be stationary over time, but it allows the error term ϵ_{it} to be dependent across choices and over time. In addition, it does not impose any distributional restrictions on the unobserved term ϵ_{it}. Therefore, the standard logit/probit models and i.i.d. assumption of the error term can be nested in Assumption 3.

One crucial feature of Assumption 3 is that it can accommodate endogenous covariates by allowing for arbitrary dependence structures between the fixed effects α_{ij} and the covariates X_{it}. Endogeneity is important in demand estimation since the price of a product potentially depend on the unobserved heterogeneity of the product, such as the quality of the product or consumers' taste for the product. Chesher, Rosen, and Smolinski (2013) and Berry and Haile (2014) provide more detailed discussions about the importance of allowing endogeneity in demand estimation.

Assumption 3 also imposes some restrictions. For example, it excludes some dependence structures between ϵ_{it} and the covariate X_{it}. Consider that if ϵ_{it} only depends on X_{it} for any period t, then ϵ_{is} may have a different distribution than ϵ_{it} when X_{is} and X_{it} take different values. Some dependence structures between ϵ_{it} and X_{it} are allowed in Assumption 3: for example, if ϵ_{it} depends on covariates in a time-invariant form such as $\frac{1}{T}\sum_{t=1}^{T} X_{it}'\beta_0$, Assumption 3 can still hold.

2.1 Substitution Patterns

Before describing the identification results, I discuss the relationship between two different definitions of substitution patterns. This paper uses the sign of $\Gamma(z, \gamma_0)$ to represent the substitution relationship between two goods, This sign captures the incremental utility from consuming the bundle compared to consuming a single good. Now I introduce an alternative definition of substitution patterns that is widely used in the literature such as Gentzkow (2007). In Lemma 1, an equivalence result between the two different definitions is established.

The alternative definition of substitution patterns centers on how the demand for good A (or B) is affected by an increase in the price of good B (or A). The two goods are substitutes if the demand for good A increases, complements if it decreases, and independent if the demand does not change. Let p_{jt} denote the price of good j whose coefficient is nonzero, and let $\tilde{X}_{it} = X_{it} \setminus \{p_{Bt}\}$ denote the remaining covariates in X_{it}

excluding the price of good B. I fix all other covariates $\tilde{X}_{is} = \tilde{X}_{it} = \tilde{x}$ over time and compare the conditional demand for good A under different prices $p_{Bs} \neq p_{Bt}$ of good B. Under model (1), the demand for good A comes from two sources: individuals who purchase only good A and those who purchase the bundle AB. Let $D_\ell = \{\ell, AB\}$ collect all choices containing good $\ell \in \{A, B\}$. Let $\text{sign}(x) = \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$ denote the sign function.

The substitution pattern $s_{AB}(z)$ conditional on the covariate $Z_i = z$ is defined as

$$s_{AB}(z) \equiv \text{sign}\left\{\frac{\Pr(Y_{is} \in D_A \mid p_{Bs}, p_{Bt}, \tilde{x}, z) - \Pr(Y_{it} \in D_A \mid p_{Bs}, p_{Bt}, \tilde{x}, z)}{p_{Bs} - p_{Bt}}\right\}.$$

The value of $s_{AB}(z) \in \{-1, 0, 1\}$ represents the complementarity relationship between goods A and B. For consumers with the covariate $Z_i = z$, the two goods are substitutes if $s_{AB}(z) = 1$, independent if $s_{AB}(z) = 0$, and complements if $s_{AB}(z) = -1$. Under the aforementioned Assumptions 1-3, the value of $s_{AB}(z)$ is the same defined by any two periods $s \neq t$, and it is independent of other variables except z since the complementarity term $\Gamma(z, \gamma_0)$ depends only on z; therefore, $s_{AB}(z)$ is written as a function of only z.

It is often difficult to study substitution patterns directly from the definition of $s_{AB}(z)$. The term $s_{AB}(z)$ uses only variation in prices and requires the fixing of all other covariates. This may not be feasible since the other covariates may change simultaneously with prices or the covariates may include time-varying variables such as time dummies. In addition, variation in prices may not be available in some scenarios in which the prices of products are constant over time. Moreover, as the definition $s_{AB}(z)$ involves conditional choice probabilities, directly estimating $s_{AB}(z)$ may perform poorly, especially when the dimension of covariates is large.

The next lemma establishes the relationship between $s_{AB}(z)$ and the incremental utility $\Gamma(z, \gamma_0)$.

Lemma 1. *Under Assumptions 1-3, the following holds for any $Z_i = z$:*

$$\Gamma(z, \gamma_0) s_{AB}(z) \leq 0.$$

Lemma 1 shows that $s_{AB}(z)$ always has the opposite sign of the incremental utility term $\Gamma(z, \gamma_0)$. This lemma implies that the sign of $s_{AB}(z)$ can be learned if the sign of the incremental utility $\Gamma(z, \gamma_0)$ is identified. Therefore, identifying the complementarity parameter γ_0 is sufficient for studying substitution patterns defined by $s_{AB}(z)$.

To illustrate the intuition of Lemma 1, I focus on the case in which the incremental utility is positive: $\Gamma(z,\gamma_0) > 0$. If the additional utility from consuming the bundle AB is positive, consumers with a small utility from a single good will still purchase the bundle since they can obtain additional positive utility from consuming the two goods together. When the price of good B increases such that the utility of the bundle decreases, some consumers will switch from buying the bundle to buying the outside option since their utility from a single good is small. Therefore, the demand for good A decreases, which implies $s_{AB}(z) \leq 0$.

A similar result to Lemma 1 is shown in Gentzkow (2007) with cross-sectional data. The difference is that Gentzkow (2007) requires an independence condition between unobserved error terms and observed covariates, so his results do not apply to the case with endogenous covariates. My paper mainly employs the stationarity assumption, which allows for endogenous covariates. Therefore, the result in Lemma 1 shows that even with endogenous covariates, the relationship between the two definitions of substitution patterns ($\Gamma(z,\gamma_0)s_{AB}(z) \leq 0$) still holds by using intertemporal variation in covariates.

3 Identification

This section establishes identification results for the parameter $\theta_0 = (\beta_0, \gamma_0)$, which includes the utility coefficient β_0 and the complementarity parameter γ_0. The observed data are the covariates (X_{it}, Z_i) and consumers' choices $Y_{it} \in \mathcal{C}$ in each period.

Let $P_t(K \mid x_s, x_t, z)$ denote the conditional choice probability (CCP) of $Y_{it} \in K$ for $K \subset \mathcal{C}$ at time t given covariates $(X_{is}, X_{it}) = (x_s, x_t)$ and $Z_i = z$. It is the probability that there exists one choice in the set K generating the highest utility among all choices; that is,

$$
\begin{aligned}
P_t(K \mid x_s, x_t, z) &\equiv \Pr(Y_{it} \in K \mid x_s, x_t, z) \\
&= \Pr\big(\exists j \in K \text{ s.t. } \forall k \in \mathcal{C}\ u_{ijt} \geq u_{ikt} \mid x_s, x_t, z\big).
\end{aligned}
$$

When K is a singleton, this reduces to the conditional probability of selecting one choice. The main idea of my identification analysis is to derive identifying restrictions of the true parameter θ_0 from intertemporal variation in conditional choice probabilities over two different periods. All parameters satisfying those identifying restrictions form an identified set for the true parameter.

Let $\delta_{\ell t} = x'_{\ell t}\beta_0$ denote the covariate index for good $\ell \in \{A, B\}$ given $X_{i\ell t} = x_{\ell t}$. Let $\delta_{ABt} = \delta_{At} + \delta_{Bt}$ and $\delta_{Ot} = 0$ denote the covariate indices for the bundle AB and the outside option, respectively. Let $\Delta_{s,t}\delta_j = \delta_{js} - \delta_{jt}$ denote the change in the covariate index for choice $j \in \mathcal{C}$ between periods s and t.

In models assuming that consumers can buy only one good at a time, two goods can only be substitutes. Since the complementarity relationship is known, the only unknown factor affecting conditional choice probabilities is variation in covariate indices of all choices. My paper allows for the possibility that two goods can be either substitutions ($\Gamma(z,\gamma_0) < 0$) or complements ($\Gamma(z,\gamma_0) > 0$) and the complementarity relationship is unknown. Therefore, there are two unknown sources affecting conditional choice probabilities in my paper: one is changes in covariate indices and the other is the complementarity relationship between the two goods. Distinguishing the two sources and identifying the complementarity makes the identification analysis challenging and different from the literature.

The following proposition characterizes the identifying restrictions for the parameter θ_0 under Assumptions 1-3. Let $C_1 \vee C_2$ mean that either condition C_1 or C_2 holds or both hold, and let $C_1 \wedge C_2$ mean that both C_1 and C_2 hold.

Proposition 1. *Under Assumptions 1-3, the following conditions hold for any (x_s, x_t, z) and $s \neq t \leq T$:*

(1) comparisons of CCP of choice $j \in \mathcal{C}$:

$$P_s(\{j\} \mid x_s, x_t, z) > P_t(\{j\} \mid x_s, x_t, z) \Longrightarrow \exists k \neq j \text{ s.t. } \Delta_{s,t}\delta_j > \Delta_{s,t}\delta_k; \tag{ID1}$$

(2) comparisons of the demand for good $\ell \in \{A, B\}$ and let $\ell_{-1} \neq \ell \in \{A, B\}$,

$$\begin{aligned} &P_s\big(\{\ell, AB\} \mid x_s, x_t, z\big) > P_t\big(\{\ell, AB\} \mid x_s, x_t, z\big) \Longrightarrow \\ &\quad \{\Delta_{s,t}\delta_\ell > 0\} \vee \Big\{\Delta_{s,t}(\delta_\ell + \operatorname{sign}(\Gamma(z,\gamma_0))\delta_{\ell_{-1}}) > 0, |\Gamma(z,\gamma_0)| > -\Delta_{s,t}\delta_\ell\Big\}; \end{aligned} \tag{ID2}$$

(3) comparisons of the sum of CCP of two choices:

$$\begin{aligned} &P_s(\{AB\} \mid x_s, x_t, z) + P_t(\{O\} \mid x_s, x_t, z) > 1 \Longrightarrow \\ &\qquad \Big\{\Gamma(z,\gamma_0) > -\min\{\Delta_{s,t}\delta_A, \Delta_{s,t}\delta_B\}\Big\} \wedge \{\Delta_{s,t}(\delta_A + \delta_B) > 0\}; \\ &P_s(\{A\} \mid x_s, x_t, z) + P_t(\{B\} \mid x_s, x_t, z) > 1 \Longrightarrow \\ &\qquad \Big\{\Gamma(z,\gamma_0) < \min\left\{\Delta_{s,t}\delta_A, -\Delta_{s,t}\delta_B\right\}\Big\} \wedge \{\Delta_{s,t}(\delta_A - \delta_B) > 0\}. \end{aligned} \tag{ID3}$$

Proposition 1 characterizes identification restrictions for the parameter θ_0 from comparisons of conditional choice probabilities over two periods that can be identified from data. The identifying restrictions for θ_0 in Proposition 1 are free from unobserved terms including the fixed effects α_i and the error term ϵ_{it}. Since the above results hold for any fixed length T of panel data, I can use variation in conditional choice probabilities for any two periods to identify θ_0 and take intersections of the identified sets. Later I will formulate conditional moment inequalities based on the identifying restrictions in Proposition 1, which can be used to conduct inference for the parameter θ_0.

Condition (ID1) in Proposition 1 contains the identifying restrictions for the coefficient β_0. The intuition of this result is as follows: if the conditional probability of selecting choice j increases, then it is impossible that choice j becomes worse (in terms of the covariate index) compared to all other choices. Therefore, it can be inferred that the covariate index for choice j should increase relative to at least one other choice.

The remaining two conditions in Proposition 1 provide novel identification results for the complementarity parameter γ_0. Condition (ID2) identifies the sign of the complementarity $\Gamma(z, \gamma_0)$ and bounds its absolute value by comparing the conditional demand of the two goods over time. Condition (ID3) establishes both lower and upper bounds for the complementarity $\Gamma(z, \gamma_0)$ using the sum of probabilities of two different choices over two periods. Next, I describe the intuition for the two conditions.

Condition (ID2) mainly exploits the idea that increasing the utility of one good affects the demand for the other good differently under different complementarity relationships between the two goods. Under model (1), the demand for one good involves individuals who buy a single good and the bundle. Therefore, the demand depends not only on the utility of a single good but also on the complementarity term $\Gamma(z, \gamma_0)$ in the utility of the bundle. When the two goods are complements, increasing the covariate index of good A will encourage consumers to buy the bundle AB so the demand for good B also increases. If the two goods are substitutes, increasing the covariate index of good A will shift consumers from originally buying good B only to buying good A only so that the demand for good B decreases.

Therefore comparisons of the conditional demand over two periods can help identify the sign of the complementarity relationship $\Gamma(z, \gamma_0)$. For example, when decreasing demand for good A or good B is observed even as the covariate indices for the two goods both increase, this implies that the two goods are substitutes ($\Gamma(z, \gamma_0) < 0$). It is because if the two goods were complements, increasing the covariate indices of both goods would lead to an increase in the demand for both goods. Similarly, when the covariate index for good A decreases and for good B increases while the demand for good A increases, the two goods are identified as complements ($\Gamma(z, \gamma_0) > 0$).

Condition (ID3) in Proposition 1 can bound the value of the complementarity $\Gamma(z, \gamma_0)$

by looking at the sum of conditional probabilities of two choices over time. For example, when the sum of the conditional probabilities of buying two goods together and neither good is large, then a lower bound for the complementarity $\Gamma(z, \gamma_0)$ is established. The intuition is that when the price of good A increases, consumers will switch from buying two goods together to buying neither good if the two goods are complements, but they will switch to buying only good B if the two goods are substitutes. Therefore, if we observe a large probability of buying two goods together and neither good, we can infer the complementarity between the two goods and provide a lower bound for $\Gamma(z, \gamma_0)$. Similarly, if the sum of the conditional probabilities of buying a single good is large, then the upper bound for the value of $\Gamma(z, \gamma_0)$ can be obtained.

The identifying restrictions (ID1)-(ID3) in Proposition 1 characterize an identified set Θ_I for θ_0, which is defined as

$$\Theta_I = \{\theta : \text{conditions (ID1)} - \text{(ID3) hold with } \theta \text{ in place of } \theta_0\}.$$

Theorem 1. *Under Assumptions 1-3, the identified set Θ_I is sharp.*

Theorem 1 shows that the identifying restrictions (ID1)-(ID3) have exhausted all possible information from the observed data for the parameter θ_0. The proof of the sharpness is conducted through direct construction. For any parameter in the identified set Θ_I, if I can construct an underlying DGP that satisfies Assumptions 1-3 and matches the observed conditional choice probabilities, it shows the sharpness of the identified set Θ_I. The difficulty is that the unknown DGP involves conditional distributions over the whole space of unobserved error terms that make the construction challenging.

This paper addresses the difficulty by first constructing "choice sets," which are collections of unobserved terms such that a single choice is selected conditional on covariates. It is sufficient to focus on constructing the distributions on the choice sets because their distributions determine the observed choice probabilities. The number of choice sets is finite due to the finite number of choices; accordingly, I need only to assign probabilities on the finite number of sets, which simplifies the construction. Then the paper shows that for any parameter in the identified set Θ_I, there exists a conditional distribution on the choice sets that satisfies the assumptions and generates the observed choice probabilities. The construction of the probabilties on the choice sets depends on the sign of the complementarity $\Gamma(z, \gamma_0)$ as well as the covariate index $\Delta_{s,t}\delta_j$, which is discussed in detail in Section A.3.

There are some interesting facts from Theorem 1. First, the identified set Θ_I employs only marginal choice probabilities at each period yet it is shown to be sharp. Therefore, joint choice probabilities over different periods do not provide any extra information for the parameter θ_0. This is because joint choice probabilities also depend on the unknown dependence structure of the error term ϵ_{it} over different periods, and this dependence structure cannot be distinguished from the effects of variation in covariate indices without further assumptions.

To show sharpness, I need to construct a joint distribution of ϵ_{ijt} across choices and over time to match the observed data. The sharpness result exploits the fact that the unobserved shock ϵ_{ijt} can be freely correlated across choices and over time. If additional restrictions are imposed on the dependence structure of the error terms across choices or over time, then the identified set Θ_I may not be sharp and can be further tightened.

Finally, my identification analysis in Proposition 1 and the sharpness result for Θ_I can be extended to a more general model, $u_{i\ell t} = f(X_{i\ell t}, \beta_0) + g(\alpha_{i\ell}, \epsilon_{i\ell t})$, where f is a known function up to a finite-dimensional parameter β_0 and g is a function that can be unknown to econometrician. This utility function allows for infinite-dimensional fixed effects and idiosyncratic shocks as well as admits arbitrary interactions between them.

3.1 Point Identification

This section studies the conditions under which the parameters β_0 and γ_0 can be point identified up to scale. The analysis depends on the specification of the additional utility term $\Gamma(Z_i, \gamma_0)$. I focus on point identification under a linear specification of the complementarity: $\Gamma(Z_i, \gamma_0) = Z_i'\gamma_0$.

For simplicity of notation, I consider a two-period model ($T = 2$) to illustrate the idea. Let $\Delta X_{i\ell} = X_{i\ell 2} - X_{i\ell 1}$ denote the change in observed covariates for consumer i and good $\ell \in \{A, B\}$ over the two periods, and let $\Delta X_i = (\Delta X_{iA}, \Delta X_{iB})$ collect the changes in covariates for the two goods. I use a superscript k to denote the kth element of a vector, e.g., ΔX_{iA}^k represents the kth element of the vector ΔX_{iA}.

I first introduce sufficient conditions for point identification of the coefficient β_0.

Assumption 4. *The support of the conditional density of ϵ_{it} given $(X_{i1}, X_{i2}, Z_i, \alpha_i)$ is $\mathbb{R}^2$.*

Assumption 5. *For any $\ell \in \{A, B\}$, there exists k_ℓ that satisfies $\beta_0^{k_\ell} \neq 0$. Let $\Delta\tilde{X}_i = \Delta X_i \backslash (\Delta X_{iA}^{k_A}, X_{iB}^{k_B})$ denote the remaining elements in ΔX_i. The support of the conditional density of $(\Delta X_{iA}^{k_A}, X_{iB}^{k_B})$ conditional on $(\Delta\tilde{X}_i, Z_i)$ is $\mathbb{R}^2$. Furthermore, the support of $\Delta X_{i\ell}$ is not contained in any proper linear subspace of $\mathbb{R}^{d_x}$.*

Assumption 4 requires that the conditional density of ϵ_{it} is positive everywhere on $\mathbb{R}^2$. This rules out the uninformative case in which conditional choice probabilities do not vary when the covariate indices change over time. Assumption 5 is a support condition on the covariate ΔX_i. It requires at least one covariate for each good to have large support, while the support of the remaining covariates is unrestricted. The large support condition guarantees that there is sufficient variation in the covariate over time such that the true parameter can be distinguished from any other candidate parameters.

Under these assumptions, β_0 can be point identified (up to scale) by using the first identifying restriction (condition (ID1)) in Proposition 1. For any parameter $b \neq k\beta_0$ for any $k > 0$, the large support condition in Assumption 5 implies that there exists one value Δx_ℓ of the covariate such that the covariate index $\Delta x_\ell'\beta$ has different signs under the true parameter $\beta = \beta_0$ and the candidate parameter $\beta = b$. The conditional choice probabilities then change in different directions under β_0 and b so that the parameter β_0 is identified.

For example, suppose that the covariate index satisfies $\Delta x_\ell'\beta_0 > 0$ and $\Delta x_\ell' b < 0$ for any $\ell \in \{A, B\}$. Then under Assumption 4, the conditional choice probability of buying the bundle AB will strictly increase under the true parameter β_0, but strictly decrease under the parameter b. Therefore, β_0 is identified.

Under the conditions for point identification of β_0, the sign of the covariate index $\Delta X_{ij}'\beta_0$ is also identified. Next, I present the conditions for point identification of the complementarity parameter γ_0.

Assumption 6. *There exists k such that $\gamma_0^k \neq 0$, and the support of Z_i^k conditional on (X_{i1}, X_{i2}) is $\mathbb{R}$. Furthermore, the support of Z_i is not contained in any proper linear subspace of $\mathbb{R}^{d_z}$.*

Similar to Assumption 5, this assumption requires a large support restriction on the covariate Z_i. Based on the identifying condition (ID2) in Proposition 1, the sign of the complementarity $Z_i'\gamma_0$ can be identified from intertemporal variation in the conditional demand for the two goods. Then for any candidate parameter $\tilde{\gamma} \neq k\gamma_0$, Assumption 6 implies that there exists some value of the covariate Z_i such that the sign of the complementarity $Z_i'\gamma$ is different under the true parameter γ_0 and the candidate parameter $\tilde{\gamma}$. Thus, the parameter γ_0 can be point identified.

Theorem 2. *Under Assumptions 1-6 and $\Gamma(Z_i, \gamma_0) = Z_i'\gamma_0$, the parameters β_0 and γ_0 are point identified up to scale.*

Theorem 2 establishes the point identification results under the large support assumptions of covariates and the linear specification of the complementarity. Without the large support assumption, I characterize the sharp identified set Θ_I for θ_0 in Theorem 1 without restricting the support of covariates.

4 Conditional Moment Inequalities

The identified set Θ_I characterized by conditions (ID1)-(ID3) in Proposition 1 is abstract and it is a challenging task to check whether every candidate parameter satisfies all of the identifying conditions. This section develops an alternative characterization of the identified set Θ_I by constructing conditional moment inequalities of the parameter. Based on this characterization, the literature has developed many methods to do inference for conditional moment inequalities such as Andrews and Shi (2013), Chernozhukov, Lee, and Rosen (2013), and Armstrong (2015).

The identification conditions (ID1)-(ID3) in Proposition 1 have the same structure of deriving restrictions for the parameter θ_0 from some intertemporal comparisons of conditional choice probabilities. I focus on the first condition (ID1) in Proposition 1 to describe the idea of constructing conditional moment inequalities. Let $W_{ist} = (X_{is}, X_{it}, Z_i)$ collect all of the covariates at the two periods (s, t), and let $w_{st} = (x_s, x_t, z)$ denote one realization of the covariate W_{ist}.

Condition (ID1) exploits comparisons of the conditional probability of a single choice $j \in \mathcal{C}$ to derive restrictions for the parameter. Let $\lambda^j_{s,t}(w_{st}, \theta)$ denote the indicator index of the identifying restriction in condition (ID1), defined as

$$\lambda^j_{s,t}(w_{st}, \theta) = \mathbb{1}\{\exists\ k \neq j \text{ s.t. } \Delta_{s,t} x_j' \beta > \Delta_{s,t} x_k' \beta\}.$$

Condition (ID1) derives the identifying restriction $\lambda^j_{s,t}$ from a positive variation in the conditional probability of selecting choice j over time:

$$P_s(\{j\} \mid w_{st}) - P_t(\{j\} \mid w_{st}) > 0 \Longrightarrow \lambda^j_{s,t}(w_{st}, \theta_0) = 1.$$

The contraposition of the above condition is presented as follows: if the identifying restriction $\lambda^j_{s,t}$ does not hold, then the variation in the conditional probability of selecting choice j is nonpositive.

$$\lambda^j_{s,t}(w_{st}, \theta_0) = 0 \Longrightarrow P_s(\{j\} \mid w_{st}) - P_t(\{j\} \mid w_{st}) \leq 0.$$

Plugging into the definition of the conditional choice probability $P_t(\{j\} \mid w_{st}) = E[\mathbb{1}\{Y_{it} = j\} \mid W_{ist} = w_{st}]$, the above condition leads to the following conditional moment inequality for any w_{st},

$$g^j_{s,t}(w_{st}, \theta_0) = E\big[(1 - \lambda^j_{s,t}(w_{st}, \theta_0))(\mathbb{1}\{Y_{is} = j\} - \mathbb{1}\{Y_{it} = j\}) \mid W_{ist} = w_{st}\big] \leq 0.$$

The above conditional moment inequality holds since either the binary index holds $\lambda^j_{s,t}(w_{st}, \theta_0) = 1$ so that the moment function $g^j_{s,t}$ is zero or the binary index does not hold $\lambda_{s,t}(w_{st}, \theta_0) = 0$ implying that the function $g^j_{s,t}$ is nonpositive. I provide an equivalent characterization to condition (ID1) using conditional moment inequalities. The characterization for the remaining two conditions in Proposition 1 can be constructed similarly.

Next I define the binary indicator of identifying restrictions for the parameter in conditions (ID2)-(ID3). Condition (ID2) derives restrictions of the parameter from comparisons of the demand for good $\ell \in \{A, B\}$. The indicator $\lambda^{D_\ell}_{s,t}(w_{st}, \theta)$ of the identifying restriction in condition (ID2) is defined as follows, let $\ell_{-1} \neq \ell \in \{A, B\}$,

$$\begin{aligned}\lambda^{D_\ell}_{s,t}(w_{st}, \theta) =&\mathbb{1}\Bigg\{\{\Delta_{s,t}x'_\ell\beta > 0\} \\ &\qquad \vee \Big\{\Delta_{s,t}(x_\ell + \text{sign}(\Gamma(z,\gamma))x_{\ell_{-1}})'\beta > 0, |\Gamma(z,\gamma)| > -\Delta_{s,t}x'_\ell\beta\Big\}\Bigg\}.\end{aligned}$$

From comparisons of the demand for good $\ell \in \{A, B\}$, the conditional moment inequality can be constructed as follows:

$$g^{D_\ell}_{s,t}(w_{st}, \theta_0) = E\big[(1 - \lambda^{D_\ell}_{s,t}(w_{st}, \theta_0))(\mathbb{1}\{Y_{is} \in D_\ell\} - \mathbb{1}\{Y_{it} \in D_\ell\}) \mid W_{ist} = w_{st}\big] \leq 0.$$

Condition (ID3) derives lower and upper bounds for the complementarity $\Gamma(z, \gamma_0)$ from the sum of conditional probabilities of two choices over two different periods. The binary indices of the identifying restrictions in condition (ID3) are defined as

$$\begin{aligned}\lambda^L_{s,t}(w_{st}, \theta) &= \mathbb{1}\Big\{\big\{\Gamma(z,\gamma) > -\min\{\Delta_{s,t}x'_A\beta, \Delta_{s,t}x'_B\beta\}\big\} \wedge \{\Delta_{s,t}(x_A + x_B)'\beta > 0\}\Big\},\\ \lambda^U_{s,t}(w_{st}, \theta) &= \mathbb{1}\Big\{\big\{\Gamma(z,\gamma) < \min\{\Delta_{s,t}x'_A\beta, -\Delta_{s,t}x'_B\beta\}\big\} \wedge \{\Delta_{s,t}(x_A - x_B)'\beta > 0\}\Big\}.\end{aligned}$$

Similarly, the conditional moment inequalities are constructed as follows based on condition (ID3) in Proposition 1:

$$g^L_{s,t}(w_{st},\theta_0) = E\big[(1-\lambda^L_{s,t}(w_{st},\theta_0))(\mathbb{1}\{Y_{is}=AB\}+\mathbb{1}\{Y_{it}=O\}-1)\mid W_{ist}=w_{st}\big] \le 0,$$
$$g^U_{s,t}(w_{st},\theta_0) = E\big[(1-\lambda^U_{s,t}(w_{st},\theta_0))(\mathbb{1}\{Y_{is}=A\}+\mathbb{1}\{Y_{it}=B\}-1)\mid W_{ist}=w_{st}\big] \le 0.$$

I have developed conditional moment inequalities that are equivalent to the identifying conditions (ID1)-(ID3) in Proposition 1. Let $g_{s,t} = (\{g^j_{s,t}\}_{j\in\mathcal{C}}, g^{D_A}_{s,t}, g^{D_B}_{s,t}, g^L_{s,t}, g^U_{s,t})'$ denote a vector of all conditional moment functions. The identified set Θ_I is characterized by the set of parameters satisfying the conditional moment inequalities as follows:

Proposition 2. *Under Assumptions 1-3, the following holds:*

$$\Theta_I = \{\theta \in : g_{s,t}(w_{st},\theta) \le 0 \quad \forall w_{st},\ \forall s,t \le T\}.$$

Given the above characterization, the inference for the parameter can be conducted using the methods in the literature developed for general conditional moment inequalities. Since only the relative utility between choices matters for consumers' decisions, the parameter can be only identified up to a constant. Therefore, I normalize the first element θ^1 of the parameter θ to be one for the following analysis: $\Theta = \{\theta : \theta^1 = 1\}$.

5 Simulation Study

This section will compare the method in this paper with a parametric method via Monte Carlo simulation. The parametric method imposes parametric assumptions over ϵ_{ijt} and a linear structure over α_{ij}, which will be described in detail later. The simulation results demonstrate that misspecifications in either parametric distributions or dependence structures between covariates and fixed effects lead to misleading estimation results for the complementarity parameter.

I focus on the case with two periods $T = 2$ and a linear specification of the complementarity: $\Gamma(Z_i,\gamma_0) = Z_i\gamma_0$. Section 3.1 has established sufficient conditions for point identification under the linear specification of the complementarity, so I focus on the case where the parameter is point identified. I implement the criterion function approach in Shi, Shum, and Song (2018) for estimation. The criterion function can be developed as follows based on conditional moment inequalities in Proposition 2:

$$\Omega(\theta) = \sum_{s\neq t\le T} E\Big[\|\max\{(g_{s,t}(W_{ist},\theta),0\}\|_1\Big] \ge \Omega(\theta_0) = 0.$$

Similar to Shi, Shum, and Song (2018), a two-step estimator is developed based on the the above criterion function. The first step estimates the conditional choice probability $P_t(\{j\} \mid w_{st})$ using a nonparametric estimator $\hat{P}_t(\{j\} \mid w_{st})$. I use a single layer artificial neural network estimator and the asymptotic property of this estimator has been established in Chen and White (1999). The neural network estimator is computationally easy to implement and there is a readily used package for the estimator (Bischl et al. (2016)). Let $\hat{g}_{s,t}$ denote the estimated moment function that replaces the conditional choice probability $P_t(\{j\} \mid w_{st})$ with its estimator $\hat{P}_t(\{j\} \mid w_{st})$, then the sample objective function $\hat{\Omega}(\theta)$ is constructed as follows:

$$\hat{\Omega}(\theta) = \frac{1}{N}\sum_{i}^{N}\sum_{s\neq t\leq T} \|\max\{\hat{g}_{s,t}(W_{ist},\theta),0\}\|_1.$$

The second-step estimator for the parameter is obtained by minimizing the sample objective function $\hat{\Omega}$. In the simulation, I also assign a positive weight $G(x) = [2\Phi(x) - 1]\mathbb{1}\{x \geq 0\}$ for intertemporal variation in choice probabilities to increase the performance of the estimator, where Φ is the CDF of the standard normal distribution.

To better evaluate the performance of the two-step estimator (Two-Step Est.) in this paper, I implement a parametric estimator (Parametric Est.) using the method of simulated moments for comparison. For this parametric estimator, the error terms ϵ_{ijt} are assumed to follow a standard Gumbel distribution, independent across choices and time periods, and also independent of all covariates. I allow the fixed effects α_{ij} to depend on covariates through a linear specification: $\alpha_{ij} = \eta_0 + \bar{X}_{ij}'\eta_1 + v_{ij}$, where $\bar{X}_{ij} = \frac{1}{T}\sum_t X_{ijt}$ denote the average covariates over time and $v_{ij} \sim \mathcal{N}(0,1)$ follows standard normal distribution and independent of all covariates. This parametric estimator is $\sqrt{N}$ consistent when its assumptions are all correct, while could be inconsistent if either the parametric distribution or the model of the fixed effects is misspecified.

For the coefficient β_0, I also evaluate the performance of two other estimators for comparison that do not allow for the purchase of bundles $\Gamma_{it} = -\infty$. One estimator is Chamberlain's conditional fixed-effect logit estimator (FE Logit Est.). This estimator assumes ϵ_{ijt} to follow standard Gumbel distribution while leaving the distribution of the fixed effects α_{ij} unrestricted. The other estimator is the semiparametric estimator (Semi. Est.) which is developed under the stationarity assumption but assumes no bundles. Therefore, this estimator only uses conditional choice probabilities of $\{A, B, O\}$ to identify the coefficient β_0.

Now I describe the simulation setup. Let d_x and d_z denote the dimension of the covariates X_{it} and Z_i respectively, and they are set to $d_x = d_z = 2$. In each simulation, $X_{i\ell t}$ is drawn from the normal distribution $\mathcal{N}(0, d_x)$, independently across choices $\ell \in \{A, B\}$ and time $t \leq T$. Let the first element of Z_i be drawn from $\mathcal{N}(1, 1)$ and the second element from $\mathcal{N}(0, 1)$. The true parameters are set as: $\beta_0 = \gamma_0 = (1, 1)$.

I study four different designs of the error terms ϵ_{ijt} and fixed effects α_{ij}. The first design considers the correct specification for the parametric estimator: ϵ_{ij} follows a Gumbel distribution and the fixed effects are specified as $\alpha_{ij} = \bar{X}_{ij}'\beta_0/2 + v_{ij}$. In the second design, the error term ϵ_{it} follows a bivariate normal distribution with the correlation $\rho = -0.7$. So the parametric distribution of the error term ϵ_{it} is misspecified in this design. In the third design, I allow the fixed effects α_{ij} to depend on the covariates of the other good in a non-additive form: $\alpha_{ij} = (\bar{X}_{ij} - \bar{X}_{ik})'\beta_0 * (1 + v_{ij})$ for $j \in \{A, B\}$ and $k \neq j \in \{A, B\}$. In this design, the parametric estimator assumes a wrong model for the fixed effects α_{ij}. The last design combines the second and third design, which considers both a misspecified distribution of ϵ_{it} and a misspecified model of α_{ij} for the parametric estimator. The following summarizes the four designs:

- Design 1: correct specification

$$\begin{aligned}
&\epsilon_{ijt} \sim \text{Gumbel}(0, 1), \\
&\alpha_{ij} = \bar{X}_{ij}'\beta_0/2 + v_{ij}, \quad \text{where } v_{ij} \sim \mathcal{N}(0, 1).
\end{aligned}$$

- Design 2: misspecified distribution

$$\begin{aligned}
&\epsilon_{it} \sim \mathcal{N}_2\big([1.5; -1.5], [1 \ \ -0.7; -0.7 \ \ 1]\big), \\
&\alpha_{ij} = \bar{X}_{ij}'\beta_0/2 + v_{ij}, \quad \text{where } v_{ij} \sim \mathcal{N}(0, 1).
\end{aligned}$$

- Design 3: misspecified fixed effects

$$\begin{aligned}
&\epsilon_{ijt} \sim \text{Gumbel}(0, 1), \\
&\alpha_{ij} = (\bar{X}_{ij}/2 - \bar{X}_{ik})'\beta_0 * (1 + 2v_{ij}), \quad \text{where } v_{ij} \sim \mathcal{N}(0, 1).
\end{aligned}$$

- Design 4: misspecified distribution and misspecified fixed effects

$$\begin{aligned}
&\epsilon_{it} \sim \mathcal{N}_2\big([1.5; -1.5], [1 \ \ -0.7; -0.7 \ \ 1]\big), \\
&\alpha_{ij} = (\bar{X}_{ij}/2 - \bar{X}_{ik})'\beta_0 * (1 + 2v_{ij}), \quad \text{where } v_{ij} \sim \mathcal{N}(0, 1).
\end{aligned}$$

For the above four designs, in terms of the coefficient β_0, I compare the four different estimators by reporting their root mean-squared error (rMSE) and median of absolute deviation (MAD). For the complementarity parameter γ_0, I compare the two-step estimator with the parametric estimator by reporting their standard deviation (SD), root mean-squared error (rMSE), and median of absolute deviation (MAD). I also report the probability of estimating the sign of substitution patterns incorrectly (Err) of the two estimators, defined as

$$\text{Err} = E|\operatorname{sign}(Z_i\gamma_0) - \operatorname{sign}(Z_i\hat{\gamma})|.$$

Let θ^k denote the kth element of the parameter θ. The parameter θ_0 can be only identified up to a constant, so the first element of is normalized to one: $\beta_0^1 = 1$ and the performance of the estimator of β_0^2 is displayed in Table 2. Since only the ratio $\tilde{\gamma}_0 = \gamma_0^2/\gamma_0^1$ matters for the substitution patterns, I focus on the results for the estimator of $\tilde{\gamma}_0$ in Table 1. I study three different sample sizes $N = \{1000, 2000, 4000\}$ and set the simulation repetitions to $B = 1000$.

Table 1: Performance Comparisons for $\hat{\gamma}$

N	Design	Two-Step Est.				Parametric Est.			
		Err	SD	rMSE	MAD	Err	SD	rMSE	MAD
1000	design 1	0.051	0.379	0.402	0.286	0.013	0.112	0.113	0.075
	design 2	0.051	0.441	0.444	0.274	0.069	0.471	0.742	0.501
	design 3	0.052	0.552	0.559	0.302	0.089	0.430	0.884	0.715
	design 4	0.056	0.807	0.843	0.309	0.222	167.116	167.201	3.568
2000	design 1	0.040	0.300	0.318	0.209	0.009	0.076	0.076	0.050
	design 2	0.041	0.335	0.339	0.222	0.061	0.330	0.588	0.427
	design 3	0.042	0.371	0.372	0.238	0.088	0.261	0.780	0.704
	design 4	0.044	0.427	0.453	0.246	0.210	325.538	325.867	2.995
4000	design 1	0.032	0.218	0.244	0.168	0.006	0.051	0.051	0.034
	design 2	0.028	0.239	0.239	0.149	0.057	0.227	0.488	0.403
	design 3	0.034	0.301	0.301	0.198	0.088	0.194	0.752	0.710
	design 4	0.037	0.318	0.346	0.205	0.204	11.346	11.940	2.959

Table 2: Performance Comparisons for $\hat{\beta}$

N	Design	Estimators with bundles				Estimators assuming no bundles			
		Two-Step Est.		Parametric Est.		FE Logit Est.		Semi. Est.	
		rMSE	MAD	rMSE	MAD	rMSE	MAD	rMSE	MAD
1000	design 1	0.128	0.086	0.079	0.052	0.145	0.092	0.141	0.108
	design 2	0.126	0.089	0.128	0.096	0.138	0.108	0.204	0.133
	design 3	0.128	0.082	0.208	0.200	0.160	0.102	0.145	0.110
	design 4	0.122	0.081	0.257	0.247	0.231	0.151	0.146	0.122
2000	design 1	0.097	0.068	0.048	0.033	0.103	0.070	0.101	0.065
	design 2	0.095	0.065	0.114	0.095	0.166	0.122	0.105	0.072
	design 3	0.090	0.063	0.206	0.202	0.110	0.072	0.103	0.066
	design 4	0.087	0.061	0.252	0.248	0.172	0.131	0.113	0.077
4000	design 1	0.078	0.051	0.033	0.022	0.071	0.047	0.073	0.046
	design 2	0.075	0.051	0.104	0.092	0.143	0.122	0.073	0.045
	design 3	0.070	0.048	0.203	0.198	0.076	0.050	0.075	0.045
	design 4	0.068	0.048	0.245	0.242	0.149	0.125	0.079	0.047

Table 1 compares the performance of the two-step estimator with the parametric estimator for the complementarity parameter γ_0. The parametric estimator performs better only when its assumptions are correctly specified (design 1), but has a worse performance under misspecifications (designs 2-4) especially when the parametric distribution and the model of fixed effects are both misspecified. The two-step estimator has uniform performance in all of the four designs, showing its advantage of performing robustly under different designs of parametric distributions and models of fixed effects. Moreover, as the sample size increases, the deviation and bias of the two-step estimator both shrink significantly. However, the performance of the parametric estimator does not improve as the sample increases in designs 2-4, which shows the inconsistency of this estimator under misspecifications.

Table 2 compares the performance of the two-step estimator with three other estimators described before for the coefficient β_0 under the four designs. Similarly, the two-step estimator performs uniformly better than the other three estimators in designs 2-4, and the difference becomes more significant as the sample size increases. In summary, the results in Table 1 and Table 2 show the advantage of the two-step estimator in performing robustly with respect to parametric distributions or specifications of dependence structures between covariates and fixed effects.

5.1 Longer Panels $T > 2$

This section presents the finite sample performance of different estimators with longer panels. I study the same four DGP designs described before except considering longer panels: $T = \{2, 3, 4\}$. I focus on the sample size $N = 1000$ and the simulation repetition $B = 1000$.

Table 3 and Table 4 compares the performance of the two-step estimator with other estimators for the complementarity parameter γ_0 and the utility coefficient β_0 under longer panels. The performance of the two-step estimator improves significantly when the length of panel increases regardless of DGP designs. The parametric estimator and other estimators perform better with longer panels if the assumptions are correctly specified (design 1), but can become worse if there is any misspecification (designs 2-4). The two-step estimator for γ_0 and β_0 still outperforms other estimators in designs 2-4 when the length of periods increases so the two-step estimator keeps the advantage of performing more robustly with longer panels.

Table 3: Performance Comparisons for $\hat{\gamma}$ (longer panel)

T	Design	Two-Step Est				Parametric Est			
		Err	SD	rMSE	MAD	Err	SD	rMSE	MAD
$T = 2$	design 1	0.051	0.379	0.402	0.286	0.013	0.112	0.113	0.075
	design 2	0.051	0.441	0.444	0.274	0.069	0.471	0.742	0.501
	design 3	0.052	0.552	0.559	0.302	0.089	0.430	0.884	0.715
	design 4	0.056	0.807	0.843	0.309	0.222	167.116	167.201	3.568
T=3	design 1	0.034	0.225	0.259	0.187	0.012	0.094	0.095	0.063
	design 2	0.037	0.335	0.336	0.196	0.111	0.595	1.216	0.934
	design 3	0.036	0.319	0.323	0.196	0.068	0.283	0.609	0.505
	design 4	0.041	0.373	0.410	0.218	0.194	34.356	34.603	2.490
T=4	design 1	0.030	0.168	0.220	0.165	0.010	0.084	0.085	0.054
	design 2	0.031	0.268	0.270	0.164	0.105	0.475	1.077	0.912
	design 3	0.027	0.224	0.224	0.151	0.057	0.208	0.477	0.418
	design 4	0.037	0.309	0.350	0.202	0.175	15.067	15.436	1.961

Table 4: Performance Comparisons for $\hat{\beta}$ (longer panel)

T	Design	Estimators with bundles				Estimators assuming no bundles			
		Two-Step Est.		Parametric Est.		FE Logit Est.		Semi. Est.	
		rMSE	MAD	rMSE	MAD	rMSE	MAD	rMSE	MAD
$T=2$	design 1	0.128	0.086	0.079	0.052	0.145	0.092	0.141	0.108
	design 2	0.126	0.089	0.128	0.096	0.138	0.108	0.204	0.133
	design 3	0.128	0.082	0.208	0.200	0.160	0.102	0.145	0.110
	design 4	0.122	0.081	0.257	0.247	0.231	0.151	0.146	0.122
$T=3$	design 1	0.085	0.058	0.056	0.037	0.091	0.057	0.111	0.079
	design 2	0.082	0.055	0.094	0.067	0.159	0.126	0.107	0.071
	design 3	0.087	0.061	0.148	0.139	0.096	0.063	0.116	0.084
	design 4	0.084	0.059	0.190	0.182	0.165	0.126	0.112	0.075
$T=4$	design 1	0.071	0.048	0.045	0.030	0.067	0.045	0.096	0.064
	design 2	0.069	0.050	0.081	0.058	0.145	0.125	0.095	0.060
	design 3	0.075	0.050	0.110	0.101	0.069	0.047	0.099	0.064
	design 4	0.071	0.050	0.157	0.147	0.147	0.125	0.100	0.068

6 Empirical Illustration

As an empirical illustration of my approach, I estimate the substitution pattern between cigarettes and e-cigarettes as well as the price coefficients. This paper focuses on settings where researchers have access to individual-level panel data of multinomial choices. However, in many scenarios, only aggregate data such as store-level sales data are available. I first present the identification results when only store-level panel data are observed.

6.1 Aggregate Panel Multinomial Choice Model

In the aggregate data, the observed variables include: X_{rjt} which denotes store r's characteristics of good j at time t (e.g., prices and display); q_{rjt} which denotes store r's sales of product j at time t; and Z_r which denotes store r's characteristics related to the complementarity (e.g., demographic information of the store). Under the limitation of only observing store-level data, I assume that agents who purchase products from the same store have the same covariates (X_{rjt}, Z_r) and agents visit the same store over time. The store-specific fixed effects η_r can be allowed in the paper by using a stationarity condition conditional on the fixed effects η_r.

Let $X_{rt} = (X_{rAt}, X_{rBt})$ collect the covariates for goods A and B, and let $E[q_{rjt} \mid x_s, x_t, z]$ denote store r's expected sales of good j at time t conditional on $(X_{rs}, X_{rt}) = (x_s, x_t)$ and $Z_r = z$. Recall that $\Delta_{s,t}\delta_j = x_s'\beta_0 - x_t'\beta_0$ denotes variation in the covariate index over time. The next proposition presents the identification results of the parameter θ_0 with store-level data.

Proposition 3. *Under Assumptions 1-3, the following conditions hold for any (x_s, x_t, z), $\ell \in \{A, B\}$, and $s \neq t \leq T$,*

$$\begin{cases} E\big[q_{r\ell s} \mid x_s, x_t, z\big] > E\big[q_{r\ell t} \mid x_s, x_t, z\big] \Longrightarrow \\ \qquad\qquad \{\Delta_{s,t}\delta_\ell > 0\} \vee \Big\{\Delta_{s,t}\big(\delta_\ell + \text{sign}(\Gamma(z,\gamma_0))\delta_{\ell_{-1}}\big) > 0, |\Gamma(z,\gamma_0)| > -\Delta_{s,t}\delta_\ell\Big\}; \\ E\big[q_{rOs} \mid x_s, x_t, z\big] > E\big[q_{rOt} \mid x_s, x_t, z\big] \Longrightarrow \{\Delta_{s,t}\delta_A < 0\} \vee \{\Delta_{s,t}\delta_B < 0\}. \end{cases}$$

Proposition 3 characterizes identifying restrictions for the parameter θ_0 with store-level sales and covariates. The main difference between store-level data and individual-level data is that in store-level data, only the demand for one good is observed but whether two goods are consumed together or consumed alone is not distinguishable. Therefore, Proposition 3 only exploits information of the demand for the two goods to identify the substitution pattern $\Gamma(z, \gamma_0)$, while the results derived from the probabilities of buying two goods together or buying a single good in Proposition 1 does not apply with store-level data. Based on the identifying restrictions in Proposition 3, the conditional moment inequalities can be developed similarly to Section 4.

6.2 Application

Since the introduction of e-cigarettes to the U.S. market in 2007, their effect on the consumption of traditional cigarettes has been widely discussed. To learn this effect, it is crucial to understand the substitution relationship between cigarettes and e-cigarettes. However, this relationship still keeps ambiguous. Some research papers such as Stoklosa, Drope, and Chaloupka (2016) and Zheng, Zhen, Dench, and Nonnemaker (2017) show that cigarettes and e-cigarettes are substitutes, while other papers including Cotti, Nesson, and Tefft (2018) and Zhao (2019) find the complementarity between the two goods.

I investigate the substitution pattern between cigarettes and e-cigarettes using the approach developed in this paper. The data I use is the Nielsen Retail Scanner Data [1], which contains weekly store-level information of sales and prices of cigarettes, e-cigarettes,

[1] https://www.chicagobooth.edu/research/kilts/datasets/nielsenIQ-nielsen

and other tobacco products in the United States. I focus on the year 2019 which is the most recent period in the Nielsen data. It contains $T = 52$ weeks of store-level sales and prices information. As shown in Proposition 3, comparisons from any two weeks of data can be used to estimate the substitution patterns. In the 52 weeks during 2019, there are around $N = 13000$ stores that sell both cigarettes and e-cigarettes in all panels of weeks.

The Nielsen data records more than 9000 UPCs (universal product code) of cigarettes that are sold in packs and cartons. There are around 500 UPCs of e-cigarettes that mainly include refill cartridges for rechargeable e-cigarettes and disposable e-cigarettes. I aggregate the sales of all UPCs of cigarettes and e-cigarettes to the store level and calculate the weighted average prices of the two goods. In addition, I aggregate the two other main tobacco products, chewing tobacco and cigars, as the outside option for people who have nicotine dependence.

There are three products: cigarettes, e-cigarettes, and the outside option. The covariates X_{rjt} and sales q_{rjt} denote the prices and sales of good j purchased in store r at time t. Table 5 shows the summary statistics of the covariates of cigarettes and e-cigarettes.

Table 5: Summary Statistics

	cigarettes		e-cigarettes	
	price	sales ratio	price	sales ratio
min	3.913	0.187	0.590	0.001
max	25.813	0.992	47.990	0.635
mean	7.105	0.801	13.569	0.045
std	1.338	0.089	3.114	0.039

Since e-cigarettes were introduced to US market, the market share of e-cigarettes has been significantly increased and has reached around 10% of traditional cigarettes in 2019. In the data, there is good variation in prices for both cigarettes and e-cigarettes which will be used to identify the price coefficients and the complementarity between cigarettes and e-cigarettes.

Let $\hat{\beta}_C, \hat{\beta}_E$ denote estimators of price coefficients of cigarettes and e-cigarettes respectively. Let $\hat{\Gamma}$ denote the estimator of the substitution pattern between the two goods, which is assumed to be the same across all stores. I normalize the price coefficient of cigarettes to be one: $|\beta_C| = 1$. Since the parameter is only partially identified, following Chernozhukov, Hong, and Tamer (2007), the set estimator for the identified set Θ_I is proposed as

$$\hat{\Theta}_{\hat{c}_N} = \left\{ \theta \in \Theta : \hat{\Omega}(\theta) \leq \inf_{\theta \in \Theta} \hat{\Omega}(\theta) + \hat{c}_N / a_N \right\},$$

where a_N denotes the uniform convergence rate of the sample objective function $\hat{\Omega}$, which is $a_N \approx N^{1/4}$ when the neural network estimator is used in the first step; $\hat{c}_N$ is chosen as $0.01 \log(N)$ so it satisfies $\hat{c}_N / a_N \to 0$. The following table displays the estimation results by randomly choosing ten and twenty weeks of data out of 52 weeks.

Table 6: Estimation Results

	price coefficient		substitution pattern
Periods	$\hat{\beta}_C$	$\hat{\beta}_E$	$\hat{\Gamma}$
$T = 10$	-1	[-1.453, -1.241]	$[-\infty, -5.832]$
$T = 20$	-1	[-1.517, -1.306]	$[-\infty, -8.331]$

The estimation results in Table 6 with ten weeks and twenty weeks of data are similar, and they both suggest that cigarettes and e-cigarettes are substitutes using the aggregate data. Since the store-level data does not contain information about whether cigarettes and e-cigarettes are purchased together or alone, we cannot rule out the case where consumers always buy one good alone ($\Gamma = -\infty$) . Therefore, the lower bound for the complementarity parameter cannot be provided with the aggregate data. If individual-level panel data were available, the paper would allow for heterogeneous substitution patterns through consumers' covariates and can provide both lower and upper bounds for the complementarity parameter.

7 Extension: Nonseparable Utility Functions

7.1 Identification

In the baseline model, I consider an additive and separable utility function which is commonly used in the literature on discrete choice models. This section studies a more general class of utility functions that can be nonseparable between observed covariates and unobserved heterogeneity. This class of utility functions can allow flexible interactions between observed covariates and unobserved terms.

Similar to model (1), the utility for consumer i from selecting choice j at time t is specified as follows

$$
\begin{aligned}
u_{iAt} &= u(X'_{iAt}\beta_0, \alpha_{iA}, \epsilon_{iAt}), \\
u_{iBt} &= u(X'_{iBt}\beta_0, \alpha_{iB}, \epsilon_{iBt}), \\
u_{iABt} &= u_{iAt} + u_{iBt} + \Gamma_{it}, \\
u_{iOt} &= 0,
\end{aligned} \tag{2}
$$

where the utility function u still depends on the three crucial components: the covariate index $X'_{ijt}\beta_0$; unobserved agent-level fixed effects α_{ij}; and unobserved error terms ϵ_{ijt}. Distinct from model (1), here I do not impose the separability restriction on the utility function and the function u can be potentially unknown to econometrician.

This utility function in model (2) has also been studied in Gao and Li (2020). Their paper does not allow for the purchase of bundles and focuses on identification of the coefficient β_0. My paper allows for the bundle of two goods and focus on identification results for the substitution pattern Γ_{it} between the two goods. Following Gao and Li (2020), I assume a monotonicity assumption on the utility function with respect to the covariate index $X'_{ijt}\beta_0$.

Assumption 7. *(Weak Monotonicity) The utility $u(\delta, \alpha, \epsilon)$ is weakly increasing in the index δ for every realization (α, ϵ), i.e.*

$$
\textit{for any}\ \ (\alpha, \epsilon), \quad u(\tilde{\delta}, \alpha, \epsilon) \geq u(\delta, \alpha, \epsilon) \quad \textit{if}\ \tilde{\delta} \geq \delta.
$$

Assumption 7 only requires monotonicity with respect to the covariate index, but imposes no restrictions on unobserved fixed effects and error terms. The additively separable utility function in model (1) is nested in this assumption. The utility function in model (2) not only admits flexible interactions between observed characteristics and unobserved heterogeneity, but also allows for nonlinear functions of the covariate X_{ijt} such as exponential functions or higher-order polynomial functions.

The next proposition characterizes the sharp identified set for the parameter θ_0 under model (2).

Proposition 4. *Under model* (2) *and Assumptions 1-3* & *7, the sharp identified set for θ_0 is the set of parameters that satisfy the following conditions: $\forall (x_s, x_t, z)$ and $\forall s \neq t \leq T$,*

(1) comparisons of CCP of the choice $j \in \mathcal{C}$,

$$
P_s(\{j\} \mid x_s, x_t, z) > P_t(\{j\} \mid x_s, x_t, z) \Longrightarrow \\
\left\{(-1)^{\mathbb{1}[j \in D_A]} \Delta_{s,t}\delta_A < 0\right\} \vee \left\{(-1)^{\mathbb{1}[j \in D_B]} \Delta_{s,t}\delta_B > 0\right\};
$$

(2) comparisons of the demand for good $\ell \in \{A, B\}$, and let $\ell_{-1} \neq \ell \in D_\ell$,

$$P_s(D_\ell \mid x_s, x_t, z) > P_t(D_\ell \mid x_s, x_t, z) \Longrightarrow \{\Delta_{s,t}\delta_\ell > 0\} \vee \{\operatorname{sign}(\Gamma(z, \gamma_0))\Delta_{s,t}\delta_{\ell_{-1}} > 0\}.$$

Similar to Proposition 1, Proposition 4 derives identifying conditions for the parameter θ_0 using intertemporal variation in conditional choice probabilities over any two periods. It is not surprising that the identified set in Proposition 4 is wider compared to the result in Proposition 1 since model (2) allows for a larger class of utility functions. The main difference between the two models is: model (1) imposes restrictions on both the direction and the degree of how covariate indices affect agents' utility u_{ijt}; model (2) only assumes the monotonicity assumption but is flexible about the degree of how covariate indices affect the utility. So the results in Proposition 4 are robust to a more general class of utility functions.

7.2 Testing Complementarity

The previous sections focused on identification results for substitution patterns. This section develops a method to test the complementarity (or substitutability) relationship between the two goods. To convey the idea, I assume that the complementarity $\Gamma_{it} = \Gamma_0$ is constant across agents; the analysis can be extended to the case where Γ_{it} depends on observed covariates in Assumption 1.

I first consider testing the nonnegative complementarity $\Gamma_0 \geq 0$ between the two goods, so the first pair of hypotheses is given as

$$H_0 : \Gamma_0 \geq 0 \qquad H_1 : \Gamma_0 < 0.$$

The main idea of testing H_0 is that under the null hypothesis H_0 and the alternative hypothesis H_1, increasing the utility of one good affects the demand for the other good in different directions. When the two goods are complements (H_0), increasing the covariate index of good A will motivate agents to purchase the bundle and thus increase the demand for good B. When the two goods are substitutes (H_1), increasing the covariate index of of good A will encourage consumers to buy good A only and decrease the demand for good B.

Therefore, the first step to test the complementarity is to infer the sign of covariate indices of two goods. Let $\xi^1_{s,t}(x_s, x_t)$ denote an indicator for increasing probabilities for all choices $j \in \{A, B, AB\}$ conditional on $(X_{is}, X_{it}) = (x_s, x_t)$, which is defined as

$$\xi^1_{s,t}(x_s, x_t) = \mathbb{1}\Big\{P_s(\{j\} \mid x_s, x_t) - P_t(\{j\} \mid x_s, x_t) \geq 0,\ \forall j \in \{A, B, AB\}\Big\}.$$

As shown in Proposition 1, increasing conditional probabilities for all choices $j \in \{A, B, AB\}$ imply that the covariate indices for the two goods both increase:

$$\xi^1_{s,t}(x_s, x_t) = 1 \Longrightarrow \Delta_{s,t}\delta_A \geq 0,\ \Delta_{s,t}\delta_B \geq 0.$$

Under the null hypothesis of the two goods being complements, increasing the covariate indices of both goods would imply increasing demand for both goods which generates testable implications for the null hypothesis. When a decreasing demand for either good is observed, it can be inferred that the two goods are substitutes and the null hypothesis is rejected.

Proposition 5. *Under model (2) and Assumptions 3 & 7, the null hypothesis H_0 implies the following conditional moment inequality:* $\forall(x_s, x_t), \forall \ell \in \{A, B\}, \forall(s, t) \leq T$,

$$E\Big[\xi^1_{s,t}(x_s, x_t)\big(\mathbb{1}\{Y_{is} \in D_\ell\} - \mathbb{1}\{Y_{it} \in D_\ell\}\big) \;\big|\; x_s, x_t\Big] \geq 0.$$

Proposition 5 has provided testable implications for the null hypothesis H_0 by characterizing conditional moment restrictions of the parameter that only depend on observed variables. Therefore, the null hypothesis can be tested by directly testing the above conditional moment inequalities. Moreover, the results in Proposition 5 are derived under the utility functions in model (2), so they are robust to a class of nonseparable utility functions.

Tests for the substitutability between the two goods can be conducted similarly. The pair of hypotheses is described as follows:

$$H_0' : \Gamma_0 \leq 0 \qquad H_1' : \Gamma_0 > 0.$$

The distinction from testing H_0 is that the relationship between the demand for one good and the covariate index of the other good is different under the new null hypothesis H_0'. To characterize the testable implications of the new null hypothesis, I consider the covariates such that the covariate index of good A increases and of good B decreases. Let $\xi^2_{s,t}(x_s, x_t) = \mathbb{1}\big\{P_s(\{j\} \mid x_s, x_t) - P_t(\{j\} \mid x_s, x_t) \geq 0,\ \forall j \in \{A, AB, O\}\big\}$ indicate increasing conditional probabilities for all choices $j \in \{A, AB, O\}$, implying

$$\xi^2_{s,t}(x_s, x_t) = 1 \Longrightarrow \Delta_{s,t}\delta_A \geq 0,\ \Delta_{s,t}\delta_B \leq 0.$$

Given the above sign of covariate indices for the two goods, the conditional demand for good A should increase and the demand for good B should decrease under the null hypothesis of the two goods being substitutes. The next proposition characterizes the testable implications of the null hypothesis H_0'.

Proposition 6. *Under model (2) and Assumptions 3 & 7, the null hypothesis H_0' implies the following conditional moment inequalities: $\forall(x_s, x_t)$ and $\forall(s,t) \leq T$,*

$$\begin{cases} E\Big[\xi^2_{s,t}(x_s, x_t)\big(\mathbb{1}\{Y_{is} \in D_A\} - \mathbb{1}\{Y_{it} \in D_A\}\big) \mid x_s, x_t\Big] \geq 0; \\ E\Big[\xi^2_{s,t}(x_s, x_t)\big(\mathbb{1}\{Y_{is} \in D_B\} - \mathbb{1}\{Y_{it} \in D_B\}\big) \mid x_s, x_t\Big] \leq 0. \end{cases}$$

7.3 Latent Complementarity

The previous sections focused on the case where heterogeneity in the complementarity Γ_{it} only comes from observed covariates: $\Gamma_{it} = \Gamma(Z_i, \gamma_0)$. This section allows for unobserved heterogeneity in the complementarity across individuals. The latent complementarity term Γ_{it} can be a random variable with an unknown distribution.

In this section, the sign of Γ_{it} can be different for each individual regardless of their covariates and it captures the heterogeneous complementarity relationship among the two goods for each individual. Therefore, I focus on identifying the distribution of the sign of Γ_{it} which represents the fraction of people for whom the two goods are complements or substitutes.

Next, I introduce some assumptions for the complementarity Γ_{it} and the error terms.

Assumption 8. *The joint distribution of $(\epsilon_{it}, \Gamma_{it})$ conditional on $(\alpha_i, X_{is}, X_{it})$ is stationary over time:*

$$(\epsilon_{is}, \Gamma_{is}) \mid X_{is}, X_{it}, \alpha_i \overset{d}{\sim} (\epsilon_{it}, \Gamma_{it}) \mid X_{is}, X_{it}, \alpha_i \quad \textit{for any } s, t \leq T.$$

Assumption 8 is similar to Assumption 3, except it also assumes a stationarity condition for the complementarity Γ_{it}. Under Assumption 1 which assumes that the complementarity Γ_{it} only depends on covariates, Assumption 8 degenerates to the stationarity condition in Assumption 3 since Γ_{it} is a constant conditional on the covariate.

Assumption 8 only requires that the distribution of Γ_{it} is stationary over time, but it still allows the complementarity Γ_{it} for each individual to vary over time. Moreover, this assumption does not restrict the correlation relationship between the complementarity Γ_{it} with other unobserved terms, including the fixed effects α_i and the error terms ϵ_{it}.

Let $X_i = (X_{it})_{t=1}^T$ collect the covariate of all time periods.

Assumption 9. *The complementarity Γ_{it} is independent of the covariate X_i conditional on the fixed effects α_i: $\Gamma_{it} \perp\!\!\!\perp X_i \mid \alpha_i$.*

Assumption 9 assumes the independence between the complementarity Γ_{it} and the vector of covariates for all periods. Variation in all covariates can be used to identify the distribution of the complementarity $\Pr(\Gamma_{it} \geq 0)$. This assumption can be relaxed to the scenario in which there is a subset of covariates that are independent of Γ_{it} while other covariates can be correlated with Γ_{it}. In this case, the analysis is conducted conditional on the covariates that are potentially correlated with the complementarity.

Under the above assumptions, I establish identification of the fraction of individuals for whom the two goods are complements, denoted as $\eta = \Pr(\Gamma_{it} \geq 0)$. According to Assumption 8, the distribution of Γ_{it} is stationary over time so that η does not depend on t. The identification result for $\Pr(\Gamma_{it} < 0)$ can be directly derived using the formula $\Pr(\Gamma_{it} < 0) = 1 - \Pr(\Gamma_{it} \geq 0)$ so it is skipped here.

The idea of the identification strategy for η is described as follows. The conditional demand for one good can be expressed as a mixture of two groups: people for whom the two goods are complements ($\Gamma_{it} > 0$) and people for whom the two goods are substitutes ($\Gamma_{it} < 0$). Variation in covariate indices of one good affects the demand for the other good for the two groups of people in different directions, which can help identify the fraction of people for whom the two goods are complements.

Similar to Section 7.2, the first step is to derive the sign of covariate indices $(\Delta_{s,t}\delta_A, \Delta_{s,t}\delta_B)$. Let $\mathcal{X}^1_{s,t} = \{(x_s, x_t) \mid \xi^1_{s,t}(x_s, x_t) = 1\}$ and $\mathcal{X}^2_{s,t} = \{(x_s, x_t) \mid \xi^2_{s,t}(x_s, x_t) = 1\}$ denote the collection of covariates such that $\xi^1_{s,t}(x_s, x_t) = 1$ and $\xi^2_{s,t}(x_s, x_t) = 1$ respectively, implying

$$
\begin{aligned}
(x_s, x_t) \in \mathcal{X}^1_{s,t} &\implies \Delta_{s,t}\delta_A \geq 0,\ \Delta_{s,t}\delta_B \geq 0, \\
(x_s, x_t) \in \mathcal{X}^2_{s,t} &\implies \Delta_{s,t}\delta_A \geq 0,\ \Delta_{s,t}\delta_B \leq 0.
\end{aligned}
$$

I first consider that the covariate indices for goods A and B both increase: $(x_s, x_t) \in \mathcal{X}^1_{s,t}$. In this case, the demand for the two goods increases for people for whom the two goods are complements, but may decline for people for whom the two goods are substitutes. Therefore, a decreased demand for either of the two goods in data can only come from people with $\Gamma_{it} < 0$, which can help establish a lower bound for the fraction of people with $\Gamma_{it} < 0$ and thus an upper bound for the fraction of people with $\Gamma_{it} \geq 0$. Similarly, a lower bound for the fraction of people for whom the two goods are complements can be provided when covariates satisfy $(x_s, x_t) \in \mathcal{X}^2_{s,t}$.

The next proposition characterizes identification results for $\eta = \Pr(\Gamma_{it} \geq 0)$.

Proposition 7. *Under model (2) and Assumptions 7-9, η can be bounded as $\eta \in [L_\eta, U_\eta]$, where*

$$
\begin{aligned}
L_\eta &= \sup_{(x_s,x_t)\in\mathcal{X}^2_{st},\ell\in\{A,B\},s,t\leq T} \Big\{(-1)^{\mathbb{1}\{\ell=A\}}[P_s(D_\ell \mid x_s,x_t) - P_t(D_\ell \mid x_s,x_t)]\Big\},\\
U_\eta &= \inf_{(x_s,x_t)\in\mathcal{X}^1_{st},\ell\in\{A,B\},s,t\leq T} \Big\{P_s(D_\ell \mid x_s,x_t) - P_t(D_\ell \mid x_s,x_t)\Big\} + 1.
\end{aligned}
$$

Proposition 7 establishes both lower and upper bounds for η by exploiting variation in the demand for the two goods under different sets of covariate indices. From the formulas for the lower and upper bounds, we can see that they have used variation in the demand over any two periods and all values of covariates. The results in Proposition 7 also provide testable implications for Assumptions 7-9 since the results imply that the upper bound should be no smaller than the lower bound: $U_\eta \geq L_\eta$.

Allen and Rehbeck (2020) also discuss latent complementarity and provide bounds for the fraction of the population for whom the two goods are complements with cross-sectional data. Their paper mainly relies on an exclusion restriction: there exists one covariate that only affects the utility of good A but not good B. Also, their paper requires an independence assumption between the covariates and all unobserved terms. As a complement to this paper, my paper considers panel data setting and mainly exploits intertemporal variation over time. My method allows covariates to be arbitrarily dependent with unobserved fixed effects. In addition, the analysis in this paper does not require an exclusion restriction and can still (partially) identify η when covariates of both goods change simultaneously.

8 Conclusion

This paper characterizes the sharp identification of a panel multinomial choice model allowing for bundles and provides novel identification results for substitution patterns between two goods. The model in this paper allows for the possibility that two goods are either substitutes or complements and admits heterogeneous complementarity relationships through observed covariates. The identification analysis does not assume parametric distributions over idiosyncratic error terms and allows for endogeneity by admitting flexible dependence structures between observed covariates and unobserved fixed effects.

The primary identification strategy is to derive identifying restrictions on unknown parameters through intertemporal variation in conditional choice probabilities that are identified from data. I develop conditional moment inequalities to characterize the identified set. The method in the paper is shown via Monte Carlo simulations to perform more robustly than the parametric method. As an empirical illustration, I estimate the substitution pattern between cigarettes and e-cigarettes. The estimation result suggests that they are substitutes. In the extension, I also study identification under a nonseparable utility function, provide methods to test complementarity, and develop identification under latent complementarity.

My work focuses on substitution patterns between two goods. The idea can apply to the case of more than two goods when consumers only purchase bundles of two goods and the complementarity between any pair of two goods is the same. Determining whether and how the identification strategy in this paper can be extended to a more general case requires additional research. When there are more than two goods, the variation in the covariate index of one good can affect the demand for another good directly through their complementarity and also indirectly through the complementarity with other goods. Therefore, there are multiple channels and interactions affecting the demand for any single good which makes the identification analysis challenging. In addition, this paper considers consumers' choice sets to be homogeneous, it would be interesting to investigate how to identify the complementarity and utility coefficients with heterogeneous and unknown choice sets.

References

R. Allen and J. Rehbeck. Latent complementarity in bundles models. *Available at SSRN 3257028*, 2020.

D. W. Andrews and X. Shi. Inference based on conditional moment inequalities. *Econometrica*, 81(2):609–666, 2013.

T. B. Armstrong. Asymptotically exact inference in conditional moment inequality models. *Journal of Econometrics*, 186(1):51–65, 2015.

S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica*, pages 841–890, 1995.

S. T. Berry and P. A. Haile. Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797, 2014.

B. Bischl, M. Lang, L. Kotthoff, J. Schiffner, J. Richter, E. Studerus, G. Casalicchio, and Z. M. Jones. mlr: Machine learning in r. *The Journal of Machine Learning Research*, 17(1):5938–5942, 2016.

G. Chamberlain. Analysis of covariance with qualitative data. *Review of Economic Studies*, 47:225–238, 1980.

X. Chen and H. White. Improved rates and asymptotic normality for nonparametric neural network estimators. *IEEE Transactions on Information Theory*, 45(2):682–691, 1999.

V. Chernozhukov, H. Hong, and E. Tamer. Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284, 2007.

V. Chernozhukov, S. Lee, and A. M. Rosen. Intersection bounds: estimation and inference. *Econometrica*, 81(2):667–737, 2013.

A. Chesher, A. M. Rosen, and K. Smolinski. An instrumental variable model of multiple discrete choice. *Quantitative Economics*, 4(2):157–196, 2013.

C. Cotti, E. Nesson, and N. Tefft. The relationship between cigarettes and electronic cigarettes: Evidence from household panel data. *Journal of Health Economics*, 61: 205–219, 2018.

F. Dunker, S. Hoderlein, and H. Kaido. Nonparametric identification of endogenous and heterogeneous aggregate demand models: complements, bundles and the market level. *cemmap working paper*, 2015.

J. T. Fox and N. Lazzati. A note on identification of discrete choice models for bundles and binary games. *Quantitative Economics*, 8(3):1021–1036, 2017.

W. Y. Gao and M. Li. Robust semiparametric estimation in panel multinomial choice models. *Available at SSRN 3282293*, 2020.

M. Gentzkow. Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*, 97(3):713–744, 2007.

A. Goolsbee and A. Petrin. The consumer gains from direct broadcast satellites and the competition with cable tv. *Econometrica*, 72(2):351–381, 2004.

L. Grzybowski and P. Pereira. The complementarity between calls and messages in mobile telephony. *Information Economics and Policy*, 20(3):279–287, 2008.

B. E. Honoré and A. Lewbel. Semiparametric binary choice panel data models without strictly exogeneous regressors. *Econometrica*, 70(5):2053–2063, 2002.

A. Iaria and A. Wang. Identification and estimation of demand for bundles. *Working Paper*, 2020.

H. Liu, P. K. Chintagunta, and T. Zhu. Complementarities and the demand for home broadband internet services. *Marketing Science*, 29(4):701–720, 2010.

C. F. Manski. Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, pages 357–362, 1987.

J. Monardo. Measuring substitution patterns with a flexible demand model. *Available at SSRN 3921601*, 2021.

A. Pakes and J. Porter. Moment inequalities for multinomial choice with fixed effects. *Working Paper*, 2019.

A. Petrin. Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy*, 110(4):705–729, 2002.

X. Shi, M. Shum, and W. Song. Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. *Econometrica*, 86(2):737–761, 2018.

M. Stoklosa, J. Drope, and F. J. Chaloupka. Prices and e-cigarette demand: evidence from the european union. *Nicotine & Tobacco Research*, 18(10):1973–1980, 2016.

J. Zhao. E-cigarettes and cigarettes: Complements or substitutes? *Working Paper*, 2019.

Y. Zheng, C. Zhen, D. Dench, and J. M. Nonnemaker. Us demand for tobacco products in a system framework. *Health Economics*, 26(8):1067–1086, 2017.

A Appendix

In the following proofs, I will suppress the covariate Z_i and use Γ_0 to denote the incremental utility $\Gamma(Z_i, \gamma_0)$.

A.1 Proof of Lemma 1

Proof. Lemma 1 contains two results to be shown: $\Gamma_0 \geq 0$ implies $s_{AB} \leq 0$, and $\Gamma_0 \leq 0$ implies $s_{AB} \geq 0$. I will show the proof for the first result, and the same idea can be applied to the second case.

Suppose that the complementarity term is positive $\Gamma_0 \geq 0$, and I need to show $s_{AB} \leq 0$. From the definition of s_{AB}, proving $s_{AB} \leq 0$ is equivalent to showing that if $p_{Bs} > p_{Bt}$, then $\Pr(Y_{is} \in D_A \mid p_{Bs}, p_{Bt}, \tilde{x}) \leq \Pr(Y_{it} \in D_A \mid p_{Bs}, p_{Bt}, \tilde{x})$. Since the covariate $\tilde{x}$ is fixed over time which does not affect variation in conditional choice probabilities, so it is suppressed in this proof.

Let $\beta_{0,p} \leq 0$ denote the coefficient for price $p_{\ell t}$. Let $v_{i\ell t} = \alpha_{i\ell} + \epsilon_{i\ell t}$ for $\ell \in \{A, B\}$. The utility for good B can be expressed as $u_{iBt} = p_{Bt}\beta_{0,p} + v_{iBt}$ and for good A is $u_{iAt} = v_{iAt}$ since all other covariates for good A are suppressed.

Let $\mathcal{V}_{D_A}(p_{Bt})$ denote the collection of $v = (v_A, v_B)$ such that there exists one choice in $D_A = \{A, AB\}$ being chosen conditional on price p_{Bt}. The set $\mathcal{V}_{D_A}(p_{Bt})$ includes two cases: either choice A or choice AB has higher utility than all other options not in D_A. So $\mathcal{V}_{D_A}(p_{Bt})$ can be expressed as follows:

$$\begin{aligned}\mathcal{V}_{D_A}(p_{Bt}) = &\big\{v \mid v_A \geq p_{Bt}\beta_{0,p} + v_B,\ v_A \geq 0\big\} \equiv \mathcal{V}_1(p_{Bt})\\ &\cup \big\{v \mid v_A + \Gamma_0 \geq 0,\ v_A + p_{Bt}\beta_{0,p} + v_B + \Gamma_0 \geq 0\big\} \equiv \mathcal{V}_2(p_{Bt}).\end{aligned}$$

The demand for good A conditional on fixed effects and prices can be expressed as follows:

$$\Pr(Y_{it} \in D_A \mid \alpha_i, p_{Bs}, p_{Bt}) = \Pr(v_{it} \in \mathcal{V}_{D_A}(p_{Bt}) \mid \alpha_i, p_{Bs}, p_{Bt}).$$

Under Assumption 3 (stationarity), the conditional distribution of v_{it} is stationarity over time since the conditional distribution of ϵ_{it} is the same over time and the fixed effects α_i are constant. Therefore, a larger set would imply a higher conditional probability as follows:

$$\mathcal{V}_{D_A}(p_{Bs}) \subseteq \mathcal{V}_{D_A}(p_{Bt}) \Longrightarrow \Pr(Y_{is} \in D_A \mid \alpha_i, p_{Bs}, p_{Bt}) \leq \Pr(Y_{it} \in D_A \mid \alpha_i, p_{Bs}, p_{Bt}).$$

By taking expectations with respect to the fixed effects α_i conditional on covariates, the above condition leads to

$$\mathcal{V}_{D_A}(p_{Bs}) \subseteq \mathcal{V}_{D_A}(p_{Bt}) \Longrightarrow \Pr(Y_{is} \in D_A \mid p_{Bs}, p_{Bt}) \leq \Pr(Y_{it} \in D_A \mid p_{Bs}, p_{Bt}).$$

Now Lemma 1 is proved if I can show that when $p_{Bs} > p_{Bt}$, it implies $\mathcal{V}_{D_A}(p_{Bs}) \subseteq \mathcal{V}_{D_A}(p_{Bt})$. I will prove $\mathcal{V}_{D_A}(p_{Bs}) \subseteq \mathcal{V}_{D_A}(p_{Bt})$ by showing that for any element $v \in \mathcal{V}_{D_A}(p_{Bs})$, it satisfies $v \in \mathcal{V}_{D_A}(p_{Bt})$. The proof will proceed by discussing two cases: $v \in \mathcal{V}_1(p_{Bs})$ and $v \in \mathcal{V}_2(p_{Bs})$.

Case 1: $v \in \mathcal{V}_1(p_{Bs})$. If v satisfies $v_A \geq p_{Bt}\beta_{0,p} + v_B$ then $v \in \mathcal{V}_1(p_{Bt})$. Otherwise v should satisfy

$$v_A < p_{Bt}\beta_{0,p} + v_B,\ v_A \geq 0.$$

Since $\Gamma_0 \geq 0$, it has the following implication:

$$v_A + \Gamma_0 \geq 0,\ (v_A + \Gamma_0) + p_{Bt}\beta_{0,p} + v_B > (v_A + \Gamma_0) + v_A \geq 0.$$

Therefore we know that $v \in \mathcal{V}_2(p_{Bt}) \subseteq \mathcal{V}_{D_A}(p_{Bt})$.

Case 2: $v \in \mathcal{V}_2(p_{Bs})$. According to the definition of the set $\mathcal{V}_2(p_{Bs})$, it decreases with p_{Bs} given $\beta_{0,p} \leq 0$. Since $p_{Bs} > p_{Bt}$, it implies $v \in \mathcal{V}_2(p_{Bs}) \subseteq \mathcal{V}_2(p_{Bt})$.

I have shown that for any element $v \in \mathcal{V}_{D_A}(p_{Bs})$, it satisfies $v \in \mathcal{V}_{D_A}(p_{Bt})$ when $p_{Bs} > p_{Bt}$. Therefore, we can conclude that $\Gamma_0 \geq 0$ implies $s_{AB} \leq 0$.

□

A.2 Proof of Proposition 1

Proof. Let $v_{i\ell t} = \alpha_{i\ell} + \epsilon_{i\ell t}$ for $\ell \in \{A, B\}$ denote the sum of fixed effects and error term. For any set $K \subset \mathcal{C}$, let $\mathcal{V}_K(x_t)$ denote the collection of $v = (v_A, v_B)$ such that there exists one choice in $K \subset \mathcal{C}$ being chosen given $X_{it} = x_t$. Let $v_{AB} = v_A + v_B + \Gamma_0$ and $v_O = 0$ denote the error term for bundle AB and the outside option respectively. The set $\mathcal{V}_K(x_t)$ can be expressed as

$$\mathcal{V}_K(x_t) = \big\{ v \mid \exists j \in K \text{ s.t. } \delta_{jt} + v_j \geq \delta_{kt} + v_k \text{ for all } k \in K^c \big\},$$

where $\delta_{\ell t} = x_{\ell t}'\beta_0$ for $\ell \in \{A, B\}$, $\delta_{ABt} = \delta_{At} + \delta_{Bt}$, and $\delta_{Ot} = 0$.

The conditional probability that there exists one choice in set K being chosen can be expressed as follows:

$$\Pr(Y_{it} \in K \mid \alpha_i, x_s, x_t) = \Pr\big(v_{it} \in \mathcal{V}_K(x_t) \mid \alpha_i, x_s, x_t\big).$$

Under Assumption 3 (stationarity), the conditional distribution of v_{it} is stationarity over time for any $s \neq t$ since fixed effects α_i are constant over time. Therefore, a larger set implies a larger conditional choice probability over time:

$$\mathcal{V}_K(x_s) \subseteq \mathcal{V}_K(x_t) \implies \Pr(Y_{is} \in K \mid \alpha_i, x_s, x_t) \leq \Pr(Y_{it} \in K \mid \alpha_i, x_s, x_t). \tag{3}$$

Next I will provide sufficient conditions on the parameter θ_0 for the set relationship $\mathcal{V}_K(x_s) \subseteq \mathcal{V}_K(x_t)$, which would imply a decreasing conditional probability of the set K over time. Then by contraposition, identifying restrictions for θ_0 can be derived from increasing choice probabilities over time. Proposition 1 includes three parts of identifying restrictions, I will show the proof for each part one by one.

Part 1: comparisons of the conditional probability of a single good $j \in \mathcal{C}$ over time. According to the definition of the set $\mathcal{V}_j(x_t)$, it increases with respect to $\delta_{jt} - \delta_{kt}$ for $k \neq j$. So when the covariate index of choice j compared to all other choices decreases over time, it implies the following set relationship:

$$\delta_{js} - \delta_{ks} \leq \delta_{jt} - \delta_{kt}\ \forall k \neq j \implies \mathcal{V}_j(x_s) \subseteq \mathcal{V}_j(x_t).$$

Plugging into the notation $\Delta_{s,t}\delta_j = \delta_{js} - \delta_{jt}$ and condition (3), it has the following implication:

$$\Delta_{s,t}\delta_j - \Delta_{s,t}\delta_k \leq 0\ \forall k \neq j \implies \Pr(Y_{is} = \{j\} \mid \alpha_i, x_s, x_t) \leq \Pr(Y_{it} = \{j\} \mid \alpha_i, x_s, x_t).$$

By contraposition and taking expectation over α_i conditional on (x_s, x_t), it yields the first identifying restriction in Proposition 1:

$$P_s(\{j\} \mid x_s, x_t) > P_t(\{j\} \mid x_s, x_t) \Longrightarrow\ \exists\ k \neq j \text{ s.t. } \Delta_{s,t}\delta_j - \Delta_{s,t}\delta_k > 0.$$

Part 2: comparisons of the demand for good $\ell \in \{A, B\}$ over time. I take good A as an example to show the proof. The set $\mathcal{V}_{D_A}(x_t)$ can be expressed as the union of two sets as follows: the set of choice A and the set of choice AB generating higher utility than other choices not in D_A,

$$\mathcal{V}_{D_A}(x_t) = \left\{v \mid \delta_{At} + v_A \geq \delta_{Bt} + v_B,\ \delta_{At} + v_A \geq 0\right\} \equiv \mathcal{V}_1(x_t)$$
$$\cup \left\{v \mid \delta_{At} + v_A + \Gamma_0 \geq 0, \delta_{At} + \delta_{Bt} + v_{AB} \geq 0\right\} \equiv \mathcal{V}_2(x_t).$$

To prove condition (ID2) in Proposition 1, I look at the contrapositive statement of (ID2) given as follows:

$$\left\{\Delta_{s,t}\delta_A \leq 0,\ \Delta_{s,t}(\delta_A + \text{sign}(\Gamma_0)\delta_B) \leq 0\right\} \vee \left\{|\Gamma_0| \leq -\Delta_{s,t}\delta_A\right\} \Longrightarrow \mathcal{V}_{D_A}(x_s) \subseteq \mathcal{V}_{D_A}(x_t). \tag{4}$$

If the above condition is shown, then similarly condition (ID2) is proved by contraposition and taking conditional expectation over α_i. The conditions on the parameter θ_0 in (4) also depends on the sign of the complementarity Γ_0. I focus on the case $\Gamma_0 > 0$ and the idea applies to the case $\Gamma_0 \leq 0$.

When $\Gamma_0 > 0$, the restriction on the parameter θ_0 in (4) includes two parts: $C_1 = \left\{\Delta_{s,t}\delta_A \leq 0, \Delta_{s,t}(\delta_A + \delta_B) \leq 0\right\}$ and $C_2 = \{\Gamma_0 \leq -\Delta_{s,t}\delta_A\}$. Now I need to show that one of the two conditions $C_1 \vee C_2$ implies $\mathcal{V}_{D_A}(x_s) \subseteq \mathcal{V}_{D_A}(x_t)$. It can be proved by showing that any element v belonging to $\mathcal{V}_{D_A}(x_s)$ also belongs to $\mathcal{V}_{D_A}(x_t)$ under either condition C_1 or C_2. For any element $v \in \mathcal{V}_{D_A}(x_s)$, I discuss two cases: $v \in \mathcal{V}_1(x_s)$ and $v \in \mathcal{V}_2(x_s)$.

Case 1: $v \in \mathcal{V}_1(x_s)$. If v satisfies $\delta_{At} + v_A \geq \delta_{Bt} + v_B$, then $v \in \mathcal{V}_1(x_t)$ since either condition C_1 or C_2 both implies $\delta_{As} \leq \delta_{At}$. Otherwise v should satisfy the following inequality:

$$\delta_{Bt} + v_B > \delta_{At} + v_A \geq \delta_{As} + v_A \geq 0.$$

Since the complementarity is nonnegative $\Gamma_0 \geq 0$, the following conditions hold:

$$\delta_{At} + v_A + \Gamma_0 \geq 0,\ (\delta_{At} + v_A) + (\delta_{Bt} + v_B) + \Gamma_0 \geq 0.$$

Therefore, $v \in \mathcal{V}_2(x_t) \subseteq \mathcal{V}_{D_A}(x_t)$.

Case 2: $v \in \mathcal{V}_2(x_s)$. I first consider that condition C_1 holds. According to the definition of the set $\mathcal{V}_2(x_s)$, the set increases when the indices δ_{As} and $\delta_{As} + \delta_{Bs}$ both increase. Condition C_1 implies increasing covariate indices $\delta_{At} \geq \delta_{As}$ and $\delta_{At} + \delta_{Bt} \geq \delta_{As} + \delta_{Bs}$, so $v \in \mathcal{V}_2(x_t)$.

Now consider that condition C_2 holds. For any element $v \in \mathcal{V}_2(x_s)$, condition C_2 implies the following condition:

$$\delta_{At} + v_A \geq \delta_{As} + \Gamma_0 + v_A \geq 0.$$

If v also satisfies the second condition in $\mathcal{V}_2(x_t)$ which is $\delta_{At} + \delta_{Bt} + v_A + v_B + \Gamma_0 \geq 0$, then v belongs to the set $\mathcal{V}_2(x_t)$: $v \in \mathcal{V}_2(x_t)$. Otherwise v should satisfy

$$\delta_{Bt} + v_B < -(\delta_{At} + v_A + \Gamma_0) \leq \delta_{At} + v_A.$$

It implies that $v \in \mathcal{V}_1(x_t)$. I have shown whenever $v \in \mathcal{V}_{D_A}(x_s)$, it satisfies $v \in \mathcal{V}_{D_A}(x_t)$ under either condition C_1 or C_2.

Part 3: comparisons of the sum of conditional probabilities of two choices over time. Condition (ID3) includes two parts of identifying restrictions: one is the sum of the conditional probabilities of buying a single good and the other is the conditional probabilities of buying the bundle and the outside option. I focus on the condition using the sum of the conditional probabilities of buying a single good.

Similarly, I look at the contrapositive statement of condition (ID3) in Proposition 1. Let $C_3 = \{\min\left\{\Delta_{s,t}\delta_A, -\Delta_{s,t}\delta_B\right\} \leq \Gamma_0\}$ and $C_4 = \{\Delta_{s,t}(\delta_A - \delta_B) \leq 0\}$, the contraposition of condition (ID3) is given as

$$C_3 \vee C_4 \implies \mathcal{V}_A(x_s) \subseteq \mathcal{V}_{\{A,AB,O\}}(x_t),$$

where the set $\mathcal{V}_A(x_t)$ and $\mathcal{V}_{A,AB,O}(x_t)$ is given as

$$\mathcal{V}_A(x_t) = \{v \mid \delta_{At} + v_A \geq 0,\ \delta_{At} + v_A \geq \delta_{Bt} + v_B,\ \delta_{Bt} + v_B + \Gamma_0 \leq 0\},$$
$$\mathcal{V}_{A,AB,O}(x_t) = \{v \mid \delta_{At} + v_A \geq \delta_{Bt} + v_B \text{ or } \delta_{At} + v_A + \Gamma_0 \geq 0 \text{ or } 0 \geq \delta_{Bt} + v_B\}.$$

First consider that condition C_3 holds, which has the following implications:

$$\delta_{As} \leq \delta_{At} + \Gamma_0 \quad \text{or} \quad \delta_{Bs} + \Gamma_0 \geq \delta_{Bt}.$$

For any element $v \in \mathcal{V}_A(x_s)$, condition C_3 implies

$$\delta_{At} + v_A + \Gamma_0 \geq 0 \quad \text{or} \quad 0 \geq \delta_{Bt} + v_B.$$

Therefore it is concluded that $v \in \mathcal{V}_{A,AB,O}(x_t)$.

When condition C_4 holds, it implies that $\delta_{As} - \delta_{Bs} \leq \delta_{At} - \delta_{Bt}$. Then the element v satisfying $v \in \mathcal{V}_A(x_s)$ also satisfies $\delta_{At} + v_A \geq \delta_{Bt} + v_B$, we can conclude that $v \in \mathcal{V}_{A,AB,O}(x_t)$.

The analysis for the sum of the conditional probabilities of purchasing the bundle and the outside option is similar, so I omit the analysis here.

□

A.3 Proof of Theorem 1

Proof. To prove sharpness, I need to show that for any parameter θ in the identified set Θ_I, I can construct a data generating process such that it matches observed choice probabilities and satisfies assumptions.

Let $X_i = (X_{it})_{t=1}^T$ and $Y_i = (Y_{it})_{t=1}^T$ collect covariates and choice variables at all periods. Let $F_{Y|X}(j_1, j_2, ..., j_T \mid x)$ denote joint choice probabilities of choosing $j_t \in \mathcal{C}$ at all periods $t \leq T$ given $X_i = x$, which are identified from data. I set fixed effects to be zero $\alpha_i = 0$ and focus on constructing the conditional distribution of the error term $\epsilon_i \mid x$.

The first requirement of sharpness is that the constructed distribution of error terms can match the observed choice probabilities $F_{Y|X}(j_1, j_2, ..., j_T \mid x)$ according to model (1):

$$F_{Y|X}(j_1, j_2, ..., j_T \mid x) = \Pr(u_{ij_tt} \geq u_{ik_tt} \quad \forall k_t \neq j_t, \forall t \leq T \mid x). \tag{5}$$

The left hand term represents observed choice probabilities in data, and the right hand term represents choice probabilities generated from model (1) which depend on the constructed distribution of the error term.

The second requirement is that the constructed distribution of the error term satisfies Assumption 3 (stationarity), which is equivalent to the following condition given $(X_{is}, X_{it}) = (x_s, x_t)$:

$$\Pr(\epsilon_{is} \in K \mid x_s, x_t) = \Pr(\epsilon_{it} \in K \mid x_s, x_t) \quad \text{for any set } K. \tag{6}$$

To construct a conditional distribution of the error term to satisfy the above two requirements, the first step is to construct choice sets. Since only consumers' choices are observed in data, I will define choice sets and construct the conditional distribution of $\epsilon_i \mid x$ over those sets. Let $\mathcal{E}_K(x_t)$ denote the collection of $\epsilon = (\epsilon_A, \epsilon_B)$ such that there exists one choice in the set K being selected given $X_{it} = x_t$, defined as

$$\mathcal{E}_K(x_t) = \{\epsilon \mid \exists j \in K \text{ s.t. } \delta_{jt} + \epsilon_j \geq \delta_{kt} + \epsilon_k \quad \forall k \in K^c \mid x_t\},$$

where $\epsilon_{AB} = \epsilon_A + \epsilon_B + \Gamma_0$ and $\epsilon_O = 0$. When $K = \{j\}$ is a singleton, $\mathcal{E}_j(x_t)$ is the set of error terms such that choice j is selected given x_t.

The four choice sets $\{\mathcal{E}_j(x_t)\}_{j \in \mathcal{C}}$ form partitions of the space of ϵ_{it} conditional on x_t.

The conditional probability of selecting choice j can be represented as follows:

$$\Pr(Y_{it} = j \mid x_t) = \Pr(\epsilon_{it} \in \mathcal{E}_j(x_t) \mid x_t).$$

For any $j_t \in \mathcal{C}$, condition (5) is satisfied when the conditional distribution of $\epsilon_i \mid x$ on the set $\mathcal{E}_j(x_t)$ is constructed as follows:

$$F_{Y|X}(j_1, j_2, ..., j_T \mid x) = \Pr(\epsilon_{i1} \in \mathcal{E}_{j_1}(x_1), ..., \epsilon_{iT} \in \mathcal{E}_{j_T}(x_T) \mid x). \tag{7}$$

The joint distribution of $\epsilon_i \mid x$ over choice sets $\mathcal{E}_j(x_t)$ is pinned down to match observed choice probabilities. Now I only need to verify that the stationarity assumption in condition (6) can be satisfied. To show it, I will construct a marginal distribution of $\epsilon_{it} \mid (x_s, x_t)$ over smaller sets such that it is stationary over any two periods $s \neq t$ and it is consistent with the distribution over choice sets derived from equation (7).

Equation (7) restricts the distribution of $\epsilon_{it} \mid (x_s, x_t)$ on the choice set $\mathcal{E}_j(x_t)$ for any $t \leq T$. The choice set $\mathcal{E}_j(x_t)$ depends on x_t so it changes over time when the covariate x_t changes. It is difficult to compare the two distributions defined over different sets at different periods and verify the stationarity assumption. To tackle this difficulty, I construct the marginal distribution of $\epsilon_{is} \mid (x_s, x_t)$ and $\epsilon_{it} \mid (x_s, x_t)$ on the intersection of the two choice sets $\mathcal{E}_j(x_s)$ and $\mathcal{E}_j(x_t)$, then their distributions are defined over the same set. Let $J_{j,k}(x_s, x_t)$ denote the intersection of the two sets $\mathcal{E}_j(x_s)$ and $\mathcal{E}_j(x_t)$, defined as follows:

$$J_{j,k}(x_s, x_t) = \mathcal{E}_j(x_s) \cap \mathcal{E}_k(x_t).$$

Let $P_t(j \mid x_s, x_t) = \Pr(Y_{it} = j \mid x_s, x_t)$ denote the marginal probability of choosing j at time t which is identified from data. The requirements for the conditional distribution of $\epsilon_{it} \mid (x_s, x_t)$ over the set $J_{j,k}(x_s, x_t)$ are equivalent to the following conditions: for any $j, k \in \mathcal{C}$ and any $s \neq t$,

$$\begin{aligned}
\Pr(\epsilon_{is} \in J_{j,k}(x_s, x_t) \mid x_s, x_t) &= \Pr(\epsilon_{it} \in J_{j,k}(x_s, x_t) \mid x_s, x_t), \\
\sum_k \Pr(\epsilon_{is} \in J_{j,k}(x_s, x_t) \mid x_s, x_t) &= P_s(j \mid x_s, x_t), \\
\sum_j \Pr(\epsilon_{it} \in J_{j,k}(x_s, x_t) \mid x_s, x_t) &= P_t(k \mid x_s, x_t).
\end{aligned} \tag{8}$$

The first equation guarantees the conditional stationarity assumption (Assumption 3), and the other two equations ensure that the constructed marginal distribution of $\epsilon_{it} \mid (x_s, x_t)$ is consistent with observed choice probabilities.

Now I need to show that there exists nonnegative probabilities of $\epsilon_{it} \mid (x_s, x_t)$ over the set $J_{j,k}(x_s, x_t)$ such that condition (8) holds. Let $r_{j,k}(x_s, x_t) \geq 0$ denote the conditional probability over the set $J_{j,k}$:

$$r_{j,k}(x_s, x_t) = \Pr(\epsilon_{is} \in J_{j,k}(x_s, x_t) \mid x_s, x_t) = \Pr(\epsilon_{it} \in J_{j,k}(x_s, x_t) \mid x_s, x_t).$$

The stationarity assumption is satisfied since the probability $r_{j,k}(x_s, x_t)$ is time invariant. For the following analysis, I will suppress the covariate (x_s, x_t) for the conditional probabilities $r_{j,k}(x_s, x_t)$ and $P_t(j \mid x_s, x_t)$ to simplify notation. I only need to construct nonnegative probabilities $r_{j,k} \geq 0$ such that the last two conditions in (8) hold for all $j, k \in \mathcal{C}$:

$$\begin{aligned} \sum_k r_{j,k} &= P_s(j), \\ \sum_j r_{j,k} &= P_t(k). \end{aligned} \tag{9}$$

I focus on the case $\Gamma_0 \geq 0$ and the idea applies to the case $\Gamma_0 < 0$. The construction of $r_{j,k}$ depends on the relationship between covariate indices and the complementarity term $\{\Delta_{s,t}\delta_A, \Delta_{s,t}\delta_B, \Delta_{s,t}\delta_{AB}, \Gamma_0\}$, since their relationship determines the relationship between choice sets. I discuss the following cases to show the construction of $r_{j,k}$:

Case 1: $\Delta_{s,t}\delta_A \geq \Delta_{s,t}\delta_{AB} \geq 0 \geq \Delta_{s,t}\delta_B$, and $\Gamma_0 \geq \min\{\Delta_{s,t}\delta_A, -\Delta_{s,t}\delta_B\}$. From the proof for Proposition 1 in A.2, the relationship between covariate indices and the complementarity term implies the following set inclusion relationship:

$$\begin{aligned} &\mathcal{E}_J(x_t) \subseteq \mathcal{E}_J(x_s) \quad \text{for any } J \in \{\{A\}, \{A, AB\}, \{A, AB, O\}\}, \\ &\mathcal{E}_B(x_t) \subseteq \mathcal{E}_{\{B,AB,O\}}(x_s). \end{aligned}$$

According to the definition of $J_{j,k}$, the above set inclusion relationship implies that the following sets are empty:

$$J_{k_1,A} = J_{k_2,AB} = J_{B,O} = J_{A,B} = \emptyset \quad \text{for } k_1 \neq A, k_2 = \{B, O\}.$$

Given the relationship between covariate indices and the complementarity term, the contraposition of conditions (ID1)-(ID3) in Proposition 1 are equivalent to the following inequalities:

$$\begin{aligned} P_t(A) &\leq P_s(A), \\ P_t(B) &\geq P_s(B), \\ P_t(A) + P_t(AB) &\leq P_s(A) + P_s(AB), \\ P_t(B) + P_s(A) &\leq 1. \end{aligned} \tag{10}$$

Now I need to show that when the above restrictions (10) hold, there exists nonnegative probabilities $r_{j,k} \geq 0$ on nonempty sets $J_{j,k}$ such that (9) holds. The following displays all probabilities $r_{j,k}$ which are not determined:

$$\begin{array}{llll} r_{B,B} & & & \\ r_{O,B} & r_{O,O} & & \\ r_{AB,B} & r_{AB,O} & r_{AB,AB} & \\ & r_{A,O} & r_{A,AB} & r_{A,A} \end{array}$$

Condition (9) requires that the sum of each row of $r_{j,k}$ equals to $P_s(j)$ and the sum of each column equals to $P_t(j)$. Then the following two probabilities can be determined:

$$r_{B,B} = P_s(B), \qquad r_{A,A} = P_t(A).$$

I look at the sum of probabilities in the first column and second row:

$$\begin{aligned} &r_{O,B} + r_{AB,B} = P_t(B) - P_s(B), \\ &r_{O,B} + r_{O,O} = P_s(O). \end{aligned}$$

Based on the above conditions, I construct nonnegative probabilities as follows:

$$\begin{aligned} &r_{O,B} = \min\{P_t(B) - P_s(B), P_s(O)\}, \\ &r_{AB,B} = P_t(B) - P_s(B) - r_{O,B}, \\ &r_{O,O} = P_s(O) - P_{O,B}. \end{aligned}$$

Similarly, I look at the sum of probabilities in the last row and third column, and the corresponding probabilities can be constructed as

$$\begin{aligned} &r_{A,AB} = \min\{P_s(A) - P_t(A), P_t(AB)\}, \\ &r_{A,O} = P_s(A) - P_t(A) - r_{AB,A}, \\ &r_{AB,AB} = P_t(AB) - r_{AB,A}. \end{aligned}$$

The last term $r_{AB,O}$ can be determined by the third row or the second column:

$$r_{AB,O} = \begin{cases} 1 - P_t(B) - P_s(A) & \text{if } P_s(A) \geq P_t(\{A, AB\}), P_t(B) \geq P_s(\{B, O\}), \\ P_s(AB) & \text{if } P_s(A) \geq P_t(\{A, AB\}), P_t(B) \leq P_s(\{B, O\}), \\ P_t(O) & \text{if } P_s(A) \leq P_t(\{A, AB\}), P_t(B) \geq P_s(\{B, O\}), \\ P_s(\{A, AB\}) - P_t(\{A, AB\}) & \text{if } P_s(A) \leq P_t(\{A, AB\}), P_t(B) \leq P_s(\{B, O\}). \end{cases}$$

The constructed probabilities $r_{j,k}$ satisfy condition (9) by construction. Also, the probability $r_{AB,O}$ is nonnegative when condition (10) holds and all other probabilities $r_{j,k}$ are nonnegative according to their definitions. The idea of constructing nonnegative probabilities $r_{j,k}$ for the following cases is similar.

Case 2: $\Delta_{s,t}\delta_A \geq \Delta_{s,t}\delta_{AB} \geq 0 \geq \Delta_{s,t}\delta_B$ and $\Gamma_0 < \min\{\Delta_{s,t}\delta_A, -\Delta_{s,t}\delta_B\}$. It implies following set inclusion relationship:

$$J_{k_1,A} = J_{k_2,AB} = J_{k_3,O} = \emptyset \quad \text{for } k_1 \neq A, k_2 = \{B, O\}, k_3 = \{B, AB\}.$$

Given the relationship between covariate indices and the complementarity term, the contraposition of conditions in Proposition 1 implies the following inequalities:

$$\begin{aligned} P_t(A) &\leq P_s(A), \\ P_t(B) &\geq P_s(B), \\ P_t(A) + P_t(AB) &\leq P_s(A) + P_s(AB), \\ P_t(B) + P_t(AB) &\geq P_s(B) + P_s(AB). \end{aligned} \tag{11}$$

The probability $r_{j,k}$ can be constructed as follows:

$$\begin{aligned} & r_{B,B} = P_s(B), && r_{A,A} = P_t(A), \\ & r_{O,O} = \min\{P_t(O), P_s(O)\}, && r_{O,B} = P_s(O) - r_{O,O}, && r_{A,O} = P_t(O) - r_{O,O}, \\ & r_{AB,AB} = \min\{P_t(AB), P_s(AB)\}, && r_{AB,B} = P_s(AB) - r_{AB,AB}, && r_{A,AB} = P_t(AB) - r_{AB,AB}. \end{aligned}$$

By construction, the above probabilities are all nonnegative. The last probability is determined as $r_{A,B} = P_s(A) + P_t(B) - 1 + r_{AB,AB} + r_{O,O}$, and it can be shown to be nonnegative $r_{A,B} \geq 0$ when condition (11) holds.

Case 3: $\Delta_{s,t}\delta_A \geq 0 \geq \Delta_{s,t}\delta_{AB} \geq \Delta_{s,t}\delta_B$. In this case, I also need to discuss two cases $\Gamma_0 \geq \min\{\Delta_{s,t}\delta_A, -\Delta_{s,t}\delta_B\}$ and $\Gamma_0 < \min\{\Delta_{s,t}\delta_A, -\Delta_{s,t}\delta_B\}$. The two cases are similar to case 1 and case 2 by exchanging the place of bundle AB and outside option O.

Case 4: $\Delta_{s,t}\delta_{AB} \geq \Delta_{s,t}\delta_A \geq \Delta_{s,t}\delta_B \geq 0$. The construction is the same with case 2, except changing the place of AB with A and changing B with O.

Case 5: $0 \geq \Delta_{s,t}\delta_{AB} \geq \Delta_{s,t}\delta_A \geq \Delta_{s,t}\delta_B$. The construction is the same with case 3 by exchanging the order of the bundle AB and the outside option O.

Case 6: all other cases are the same as above cases, except exchanging the place of choice A and choice B.

□

A.4 Proof of Theorem 2

Proof. I will show the proof for point identification of the coefficient β_0 and the idea applies to the parameter γ_0. The first step is to show that for any candidate $b \neq k\beta_0$, there exists some value of the covariate such that the sign of the covariate index $\Delta X_{i\ell}'\beta$ for good $\ell \in \{A, B\}$ is different under the true parameter β_0 and the candidate b. I take good A as an example to illustrate the idea.

From Assumption 5, the conditional density of $\Delta X_{iA}^{k_A}$ is positive everywhere. Let $\Delta \tilde{X}_{iA} = \Delta X_{iA} \setminus \Delta X_{iA}^{k_A}$ denote the remaining covariates in ΔX_{iA} and $\tilde{\beta}_0$ denote its coefficient. Consider that the coefficient of $\Delta X_{iA}^{k_A}$ is positive $\beta_0^{k_A} > 0$, and the analysis applies to the case $\beta_0^{k_A} < 0$. For any candidate b, I discuss three cases: $b^{k_A} < 0$, $b^{k_A} = 0$, and $b^{k_A} > 0$.

Case 1: $b^{k_A} < 0$. When the covariate $\Delta X_{iA}^{k_A}$ takes a large positive value $\Delta X_{iA}^{k_A} = \Delta x_A^{k_A} \to +\infty$ and all other covariates take a bounded number in their support, it implies that $\Delta x_A'\beta_0 > 0$ and $\Delta x_A' b < 0$ since the true coefficient and the candidate coefficient have different signs $\beta_0^{k_A} > 0$ and $b^{k_A} < 0$.

Case 2: $b^{k_A} = 0$. For any value $\Delta X_{iA} = \Delta x_A$, the value of $\Delta x_A' b$ is either positive or nonpositive. First consider that $\Delta x_A' b > 0$ is positive. When $\Delta x_A^{k_A}$ takes a large negative number $\Delta x_A^{k_A} \to -\infty$ such that $\Delta x_A'\beta_0 < 0$, which has a different sign from $\Delta x_A' b$. Similarly, if $\Delta x_A' b \leq 0$, there exists $\Delta x_A^{k_A} \to +\infty$ such that $\Delta x_A'\beta_0 > 0$.

Case 3: $b^{k_A} > 0$. Assumption 5 says that the support of the covariate ΔX_{iA} is not contained in any proper linear subspace, so there exists $\Delta \tilde{X}_{iA} = \Delta \tilde{x}_A$ such that $\Delta \tilde{x}_A' \tilde{\beta}_0 / \beta_0^{k_A} \neq \Delta \tilde{x}_A' \tilde{b} / b^{k_A}$. Suppose that $\Delta \tilde{x}_A' \tilde{\beta}_0 / \beta_0^{k_A} - \Delta \tilde{x}_A' \tilde{b} / b^{k_A} = k > 0$, then when the covariate takes the value $\Delta X_{iA}^{k_A} = -\Delta \tilde{x}_A' \tilde{b} / b^{k_A} - \epsilon$ with $0 < \epsilon < k$. The sign of the covariate index satisfies: $\Delta x_A'\beta_0 = \beta_0^{k_A}(k - \epsilon) > 0$ and $\Delta x_A' b = -b^{k_A}\epsilon < 0$. The construction is similar when $k < 0$.

I have shown that there exists some value of the covariate such that the sign of the index $\Delta X_{i\ell}'\beta$ is different under the true parameter β_0 and the candidate parameter b. From

Assumption 4, there exists at least one choice such that the conditional probability of this choice changes in strictly different directions under β_0 and b so β_0 is identified. Consider that $\Delta x_\ell'\beta_0 > 0$ and $\Delta x_\ell' b \leq 0$ for $\ell \in \{A, B\}$, then the conditional probability of choosing AB strictly increases under β_0 and decreases under b. When $\Delta x_A'\beta_0 > 0, \Delta x_B'\beta_0 < 0$ and $\Delta x_A' b \leq 0, \Delta x_B' b \geq 0$, then the conditional probability of choosing A strictly increases under β_0 and decreases under b. It is similar for the other cases.

□

A.5 Proof of Proposition 4

Proof. The proof for the identifying restrictions in Proposition 4 is similar to Proof A.2 for Proposition 1, so I skip the proof here and only show the sharpness result. The idea to show sharpness is similar to A.3 for Theorem 1, I need to construct nonnegative probabilities $r_{j,k}$ for the set $J_{j,k}$ such that the following conditions hold:

$$\begin{aligned} \sum_k r_{j,k} &= P_s(j), \\ \sum_j r_{j,k} &= P_t(k). \end{aligned} \tag{12}$$

I focus on the case $\Gamma_0 > 0$, and discuss the following cases of the sign of covariate indices.

Case 1: $\Delta_{s,t}\delta_A \geq 0, \Delta_{s,t}\delta_B \geq 0$. The analysis for the case $\Delta_{s,t}\delta_A \leq 0, \Delta_{s,t}\delta_B \leq 0$ is the same except changing the place of choice AB and choice O. The sign of covariate indices implies the following set relationship:

$$\begin{aligned} \mathcal{E}_{AB}(x_t) &\subseteq \mathcal{E}_{AB}(x_s), \\ \mathcal{E}_{\{\ell,AB\}}(x_t) &\subseteq \mathcal{E}_{\{\ell,AB\}}(x_s) \quad \text{for } \ell \in \{A, B\}. \end{aligned}$$

By the definition of the set $J_{j,k}$, the above set relationship implies:

$$J_{k_1,AB} = J_{k_2,A} = J_{k_3,B} = \emptyset \quad \text{for } k_1 \neq AB, k_2 \notin \{A, AB\}, k_3 \notin \{B, AB\}.$$

Given the sign of covariates index, conditions in Proposition 4 are given as follows:

$$\begin{aligned} P_t(AB) &\leq P_s(AB), \\ P_t(O) &\geq P_s(O), \\ P_t(\ell) + P_t(AB) &\leq P_s(\ell) + P_s(AB) \quad \text{for } \ell \in \{A, B\}. \end{aligned} \tag{13}$$

Now we want to show that as long as condition (13) holds, we can construct nonnegative probabilities $r_{j,k} \geq 0$ on nonempty sets $J_{j,k}$ such that condition (12) holds. For $\ell \in \{A, B\}$, the probability $r_{j,k}$ is constructed as follows:

$$
\begin{aligned}
& r_{O,O} = P_s(O), && r_{AB,AB} = P_t(AB), \\
& r_{\ell,\ell} = \min\{P_t(\ell), P_s(\ell)\}, && r_{AB,\ell} = P_t(\ell) - r_{\ell,\ell}, \quad r_{\ell,O} = P_s(\ell) - r_{\ell,\ell}, \\
& r_{AB,O} = P_s(AB) - P_t(AB) - r_{AB,A} - r_{AB,B}.
\end{aligned}
$$

It can be verified that the above probabilities satisfy condition (12). All probabilities are nonnegative by construction except $r_{AB,O}$. So I only need to verify $r_{AB,O}$ is nonnegative, which is simplified as follows:

$$
r_{AB,O} = \begin{cases} P_s(AB) - P_t(AB) & \text{if } P_s(A) \geq P_t(A), P_s(B) \geq P_t(B), \\ P_s(\{B, AB\}) - P_t(\{B, AB\}) & \text{if } P_s(A) \geq P_t(A), P_s(B) \leq P_t(B), \\ P_s(\{A, AB\}) - P_t(\{A, AB\}) & \text{if } P_s(A) \leq P_t(A), P_s(B) \geq P_t(B), \\ P_t(O) - P_s(O) & \text{if } P_s(A) \leq P_t(A), P_s(B) \leq P_t(B). \end{cases}
$$

Therefore, condition (13) implies that $r_{AB,O} \geq 0$.

Case 2: $\Delta_{s,t}\delta_A \geq 0, \Delta_{s,t}\delta_B \leq 0$. The analysis is symmetric for the case $\Delta_{s,t}\delta_A \leq 0, \Delta_{s,t}\delta_B \geq 0$ so it is skipped. Similarly, the sign of covariate indices implies the following set relationship:

$$
\mathcal{E}_A(x_t) \subseteq \mathcal{E}_A(x_s), \qquad \mathcal{E}_B(x_t) \supseteq \mathcal{E}_B(x_s).
$$

The above relationship implies that the following sets $J_{j,k}$ are empty:

$$
J_{k_1,A} = J_{B,k_2} = \emptyset \quad \text{for } k_1 \neq A, \ k_2 \neq B.
$$

Given the sign of covariate indices, the identifying restrictions in Proposition 4 are given as follows:

$$
P_t(A) \leq P_s(A), \qquad P_t(B) \geq P_s(B). \tag{14}
$$

Now I need to show if (14) holds, I can construct $r_{j,k} \geq 0$ on nonempty sets $J_{j,k}$ such that condition (12) is satisfied. The probabilities $r_{j,k}$ on nonempty sets are constructed as follows in different scenarios:

- when $P_s(A) \geq P_t(\{A, AB, O\})$,

$$\begin{aligned}
&r_{A,A} = P_t(A), && r_{B,B} = P_s(B), && \\
&r_{A,AB} = P_t(AB), && r_{AB,AB} = 0, && r_{O,AB} = 0, \\
&r_{A,O} = P_t(O), && r_{AB,O} = 0, && r_{O,O} = 0, \\
&r_{A,B} = P_s(A) - P_t(\{A, AB, O\}), && r_{AB,O} = P_s(AB), && r_{O,B} = P_s(O).
\end{aligned}$$

- when $P_t(\{A, AB\}) \leq P_s(A) < P_t(\{A, AB, O\})$, let $q_{s,t} = P_t(\{A, AB, O\}) - P_s(A)$,

$$\begin{aligned}
&r_{A,A} = P_t(A), && r_{B,B} = P_s(B), && \\
&r_{A,AB} = P_t(AB), && r_{AB,AB} = 0, && r_{O,AB} = 0, \\
&r_{A,O} = P_s(A) - P_t(\{A, AB\}), && r_{AB,O} = \min\{q_{s,t}, P_s(AB)\}, && r_{O,O} = q_{s,t} - r_{AB,O}, \\
&r_{A,B} = 0, && r_{AB,B} = P_s(AB) - r_{AB,O}, && r_{O,B} = P_s(O) - r_{O,O}.
\end{aligned}$$

- when $P_t(\{A, AB\}) - P_s(AB) \leq P_s(A) < P_t(\{A, AB\})$,

$$\begin{aligned}
&r_{A,A} = P_t(A), && r_{B,B} = P_s(B), && \\
&r_{A,AB} = P_s(A) - P_t(A), && r_{AB,AB} = P_t(\{A, AB\}) - P_s(A), && r_{O,AB} = 0, \\
&r_{A,O} = 0, && r_{AB,O} = P_t(O) - r_{O,O}, && r_{O,O} = \min\{P_s(O), P_t(O)\}, \\
&r_{A,B} = 0, && r_{AB,B} = P_t(B) - P_s(\{B, O\}) + r_{O,O}, && r_{O,B} = P_s(O) - r_{O,O}.
\end{aligned}$$

- when $P_s(A) < P_t(\{A, AB\}) - P_s(AB)$,

$$\begin{aligned}
&r_{A,A} = P_t(A), && r_{B,B} = P_s(B) && \\
&r_{A,AB} = P_s(A) - P_t(A), && r_{AB,AB} = P_s(AB), && r_{O,AB} = P_t(\{A, AB\}) - P_s(\{A, AB\}) \\
&r_{A,O} = 0, && r_{AB,O} = 0, && r_{O,O} = P_t(O), \\
&r_{A,B} = 0, && r_{AB,B} = 0, && r_{O,B} = P_t(B) - P_s(B).
\end{aligned}$$

It can be verified that all probabilities $r_{j,k} \geq 0$ are nonnegative under condition (14), and condition (12) is satisfied by construction.

$\square$

A.6 Proof of Proposition 7

Proof. The conditional demand for one good can be expressed as a mixture of two groups: one group is people to whom the two goods are complements $\Gamma_{it} \geq 0$ and the other is people to whom the two goods are substitutes $\Gamma_{it} < 0$. Therefore, the demand for good A (or B) conditional on the covariate and the fixed effects is given as follows:

$$\begin{aligned}\Pr(Y_{it} \in D_A \mid \alpha_i, x_s, x_t) = \Pr(Y_{it} \in D_A \mid \alpha_i, x_s, x_t, \Gamma_{it} \geq 0) \Pr(\Gamma_{it} \geq 0 \mid \alpha_i) + \\ \Pr(Y_{it} \in D_A \mid \alpha_i, x_s, x_t, \Gamma_{it} < 0)[1 - \Pr(\Gamma_{it} \geq 0 \mid \alpha_i)].\end{aligned}$$

The above equation holds since Assumption 9 (conditional independence) implies that $\Pr(\Gamma_{it} \geq 0 \mid \alpha_i, x_s, x_t) = \Pr(\Gamma_{it} \geq 0 \mid \alpha_i)$.

The main strategy is to use intertemporal variation in the conditional demand for good A to bound the probability $\Pr(\Gamma_{it} \geq 0 \mid \alpha_i)$. It can be shown that when $(x_s, x_t) \in \mathcal{X}^1_{s,t}$, the conditional demand $\Pr(Y_{it} \in D_A \mid \alpha_i, x_s, x_t, \Gamma_{it} \geq 0)$ increases at time s compared to time t, which will be proven later. The variation in the conditional demand given $\Gamma_{it} < 0$ can be bounded as $[-1, 1]$ since it is the difference of two probabilities. Therefore, the variation in the aggregate demand for good A can be bounded as: for any $(x_s, x_t) \in \mathcal{X}^1_{s,t}$,

$$\Pr(Y_{is} \in D_A \mid \alpha_i, x_s, x_t) - \Pr(Y_{it} \in D_A \mid \alpha_i, x_s, x_t) \geq 0 + (-1) * [1 - \Pr(\Gamma_{it} \geq 0 \mid \alpha_i)].$$

By taking expectation over the fixed effect α_i conditional on the covariate, the probability $\eta = \Pr(\Gamma_{it} \geq 0)$ can be bounded above as follows: for any $(x_s, x_t) \in \mathcal{X}^1_{s,t}$,

$$\eta \leq P_s(D_A \mid x_s, x_t) - P_t(D_A \mid x_s, x_t) + 1.$$

Since the probability η does not depend on covariates and is stationarity over time under Assumption 8, it can be bounded by taking infimum over all values of the covariates and any two periods. Moreover, the variation in the demand for good B can also be exploited to bound the probability η similarly. Therefore, the upper bound for η can be established as follows:

$$\eta \leq \inf_{(x_s,x_t) \in \mathcal{X}^1_{s,t}, \ell \in \{A,B\}, (s,t) \leq T} \Big\{ P_s(D_\ell \mid x_s, x_t) - P_t(D_\ell \mid x_s, x_t) \Big\} + 1 = U_\eta.$$

Now I need to show that the conditional probability $\Pr(Y_{it} \in D_A \mid \alpha_i, x_s, x_t, \Gamma_{it} \geq 0)$ increases at time s compared to time t when the covariate satisfies $(x_s, x_t) \in \mathcal{X}^1_{s,t}$. Let $v_{it} = \epsilon_{it} + \alpha_i$, and let $\mathcal{V}^{\Gamma}_{D_A}(x_t)$ denote the collection of $(v, \Gamma \geq 0)$ such that either choice A

or AB is chosen:

$$\begin{aligned}\mathcal{V}^{\Gamma}_{D_A}(x_t) = &\{(v,\Gamma \geq 0) \mid \delta_{At} + v_A \geq \delta_{Bt} + v_B,\ \delta_{At} + v_A \geq 0\} \equiv \mathcal{V}^{\Gamma}_1(x_t), \\ &\cup\{(v,\Gamma \geq 0) \mid \delta_{At} + v_A + \Gamma \geq 0,\ \delta_{At} + v_A + \delta_{Bt} + v_B + \Gamma \geq 0\} \equiv \mathcal{V}^{\Gamma}_2(x_t).\end{aligned}$$

The conditional demand $\Pr(Y_{it} \in D_A \mid \alpha_i, x_s, x_t, \Gamma_{it} \geq 0)$ can be expressed as the conditional probability of the set $\mathcal{V}^{\Gamma}_{D_A}(x_t)$:

$$\Pr(Y_{it} \in D_A \mid \alpha_i, x_s, x_t, \Gamma_{it} \geq 0) = \Pr((v_{it}, \Gamma_{it}) \in \mathcal{V}^{\Gamma}_{D_A}(x_t) \mid \alpha_i, x_s, x_t, \Gamma_{it} \geq 0).$$

Assumption 8 implies that the distribution (v_{it}, Γ_{it}) conditional on $(\alpha_i, X_{is}, X_{it}, \Gamma_{it})$ is stationarity over time. Then I only need to show $\mathcal{V}^{\Gamma}_{D_A}(x_t) \subseteq \mathcal{V}^{\Gamma}_{D_A}(x_s)$ when the covariate satisfies $(x_s, x_t) \in \mathcal{X}^1_{s,t}$, which has the following implication:

$$\begin{aligned}&\mathcal{V}^{\Gamma}_{D_A}(x_t) \subseteq \mathcal{V}^{\Gamma}_{D_A}(x_s) \Longrightarrow \\ &\qquad\qquad \Pr(Y_{it} \in D_A \mid \alpha_i, x_s, x_t, \Gamma_{it} \geq 0) \leq \Pr(Y_{is} \in D_A \mid \alpha_i, x_s, x_t, \Gamma_{it} \geq 0).\end{aligned}$$

To prove $\mathcal{V}^{\Gamma}_{D_A}(x_t) \subseteq \mathcal{V}^{\Gamma}_{D_A}(x_s)$, I will show that for any element $(v, \Gamma) \in \mathcal{V}^{\Gamma}_{D_A}(x_t)$, it also satisfies $(v, \Gamma) \in \mathcal{V}^{\Gamma}_{D_A}(x_s)$ when $(x_s, x_t) \in \mathcal{X}^1_{s,t}$. As shown before, $(x_s, x_t) \in \mathcal{X}^1_{s,t}$ satisfies $\delta_{At} \geq 0, \delta_{Bt} \geq 0$. I discuss two cases to prove the statement: $(v, \Gamma) \in \mathcal{V}^{\Gamma}_1(x_t)$ and $(v, \Gamma) \in \mathcal{V}^{\Gamma}_2(x_t)$.

Case 1: $(v, \Gamma) \in \mathcal{V}^{\Gamma}_2(x_t)$. According to the definition of the set $\mathcal{V}^{\Gamma}_2(x_t)$, it increases with respect to the covariate index for good A and good B. Therefore it is shown that $(v, \Gamma) \in \mathcal{V}^{\Gamma}_2(x_s)$ since the covariate indices for goods A and B both increase when $(x_s, x_t) \in \mathcal{X}^1_{s,t}$.

Case 2: $(v, \Gamma) \in \mathcal{V}^{\Gamma}_1(x_t)$. If v satisfies $\delta_{As} + v_A \geq \delta_{Bs} + v_B$, it implies $(v, \Gamma) \in \mathcal{V}^{\Gamma}_1(x_s)$ since the covariate index for good A increases at time s relative to time t. Otherwise v should satisfy $\delta_{As} + v_A < \delta_{Bs} + v_B$. Also the complementarity is nonnegative $\Gamma \geq 0$, which implies the following conditions:

$$\delta_{As} + v_A + \Gamma \geq \delta_{At} + v_A \geq 0,\ \delta_{As} + v_A + \Gamma + \delta_{Bs} + v_B \geq 2(\delta_{As} + v_A) \geq 0.$$

The above condition implies $(v, \Gamma) \in \mathcal{V}^{\Gamma}_2(x_s) \subseteq \mathcal{V}^{\Gamma}_{D_A}(x_s)$.

In summary, I have shown that $\mathcal{V}^{\Gamma}_{D_A}(x_t) \subseteq \mathcal{V}^{\Gamma}_{D_A}(x_s)$ for any $(x_s, x_t) \in \mathcal{X}^1_{s,t}$, implying that the conditional probability $\Pr(Y_{it} \in D_A \mid \alpha_i, x_s, x_t, \Gamma_{it} \geq 0)$ increases at time s compared to time t.

The proof of the lower bound for η can be established similarly, so it is omitted here.

□