Semiparametric Identification and Estimation of Discrete Choice Models for Bundles

by

Fu Ouyang

and

Thomas Tao Yang

and

Hanghui Zhang

ANU Working Papers in Economics and Econometrics
672

MAY 2020

JEL: C14, C23, C35

ISBN: 0 86831 672 5

Semiparametric Identification and Estimation of Discrete Choice Models for Bundles

Fu Ouyang*
University of Queensland

Thomas Tao Yang†
Australian National University

Hanghui Zhang‡
Shanghai University of Economics and Finance

December 19, 2019

Abstract

We study (point) identification of preference coefficients in semiparametric discrete choice models for bundles. The approach to the identification uses an "identification at infinity" (Chamberlain (1986)) insight in combination with median independence restrictions on unobservables. We propose two-stage maximum score (MS) estimators and show their consistency.

JEL classification Codes: C14, C23, C35.

Keywords: Bundle choices, semiparametric model, median independence, identification at infinity, maximum score estimation.

1 Introduction

In this paper, we study the point identification and estimation of preference coefficients in semiparametric multinomial choice models for bundles. Our paper tightly relates to the growing literatures on the discrete choice models. The most influential early work is Manski (1975, 1985, 1988) which provides a novel identification strategy of index parameters for the binary choice model and leads to a widely used maximum score type estimator. In presence of multiple choices, very recently, among others, Fox (2007), Pakes and Porter (2017), Ahn, Ichimura, Powell and Ruud (2018), Shi, Shum and Shi (2018), Yan (2018), Khan, Ouyang and Tamer (2019) and Lewbel, Yan and Zhou (2019) exploit the semiparametric identification and estimation of the multinomial choice models

*Email: f.ouyang@uq.edu.au.
†Email: tao.yang@anu.edu.au.
‡Email: zhang.hanghui@mail.shufe.edu.cn.

in a static or panel setting. Essentially, all these works could be regarded as extensions of Manski (1975, 1985, 1988) in different perspectives to the multiple choices case. In this sense, our paper fills in the gap in studying the semiparametric bundle choice model in the family of Manski's estimator.

One of the central empirical questions in the industrial organization and marketing research is to investigate the substitutive or complimentary impact between choices of goods and explain the bundle choice behavior of consumers. In empirical studies, Gentzkow (2007) estimates the parametric bundle model in analysis of relationship between the print and online newspapers in demand. Similarly within a Probit framework, Fan (2013) examines the ownership consolidation in newspaper market where they allow households in demand side purchase two copies of newspaper as a bundle choice. If deviating form the parametric specification, the choice and utilities associated with bundles prevent us applying the existing knowledge in the semiparametric multiple choice models directly and it therefore creates special difficulties in identification of model parameters. Sher and Kim (2013) and Fox and Lazzati (2017) are two theory papers rigorously studying the (nonparametric) identification of bundle choice model. Our paper adopts the basic model setup of Fox and Lazzati (2017). However, the identification strategy proposed in our paper is substantially different from Fox and Lazzati (2007) and thus it contributes to the literature in the following perspectives: 1. To our best of knowledge, we are the first paper to discuss the point identification and estimation of bundle choice model in a semiparametric framework; 2. Fox and Lazzati (2007)'s identification heavily relies on the existence of excluded variables. In contrast, we achieve the point identification of preference parameters under median independence restriction thus it allows "full extent of heteroskedasticity" on unobservables.

The remainder of this paper is organized as follows. Section 2 presents a simple bundle choice model and provides sufficient conditions on both observed covariates and unobservables that secure identification. In Section 3, we propose two-stage MS estimators motivated by the identification strategy and show their consistency. Section 4 extends our approach to models with a more complex choice set. Section 5 concludes this paper. The proofs are collected in the Appendix.

For the ease of reference, the notations maintained throughout this paper are listed here.

Notation. All vectors are column vectors. $\mathbb{R}^p$ is a p-dimensional Euclidean space equipped with the Euclidean norm $\|\cdot\|$, and $\mathbb{R}^p_+ \equiv \{x \in \mathbb{R}^p | x \geq 0\}$. We reserve the letter i for indexing agents and j for indexing alternatives. For notational convenience, we might suppress the subscript i in the rest of this paper whenever it is clear from the context that all variables are for each agent. We use $P(\cdot)$ and $\mathbb{E}[\cdot]$ to denote probability and expectation, respectively. $1[\cdot]$ is an indicator function that equals one when the event in the brackets occurs, and zero otherwise. Symbols $\backslash$, $'$, $\Leftrightarrow$, and $\xrightarrow{p}$ represent set difference, matrix transposition, if and only if, and convergence in probability, respectively. For any (random) positive sequences $\{a_n\}$ and $\{b_n\}$, $a_n = O(b_n)$ ($O_p(b_n)$) means that a_n/b_n is bounded (bounded in probability) and $a_n = o(b_n)$ ($o_p(b_n)$) means that $a_n/b_n \to 0$ ($a_n/b_n \xrightarrow{p} 0$).

2 Model and Identification

Consider a choice model[1] where the choice set $\mathcal{J}$ consists of $J = 3$ mutually exclusive alternatives ($j \in \{0, 1, 2\}$) and a bundle of alternatives 1 and 2, i.e., $\mathcal{J} = \{0, 1, 2, (1, 2)\}$. This simple model is sufficient to illustrate our identification method. We discuss in Section 4 how our approach can be modified to be applied to models with more alternatives and their bundles.

For ease of exposition, we re-number alternatives in $\mathcal{J}$ with 2-dimensional vectors of binary indicators $d = (d_1, d_2) \in \{0, 1\}^2$, where d_1 and d_2 indicate if alternative 1 and 2 are chosen, respectively. Then the choice set $\mathcal{J}$ can be one-to-one mapped to the set $\mathcal{D} = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$. An agent chooses the alternative in $\mathcal{D}$ to maximize the latent utility

$$U_d = \sum_{j=1}^{2}(v_j + x_j'\beta + \epsilon_j) \cdot d_j + \eta \cdot (w'\gamma) \cdot d_1 \cdot d_2, \tag{2.1}$$

where $v_j \in \mathbb{R}$ only affects the utility associated to stand-alone alternative j, $x_j \in \mathbb{R}^{k_1}$ and $w \in \mathbb{R}^{k_2}$ collect observed covariates different from $v = (v_1, v_2)'$, of which w is a vector of explanatory variables characterizing the interaction effects of the bundle (e.g., bundle discount)[2], $(\epsilon_1, \epsilon_2, \eta) \in \mathbb{R}^3$ captures unobserved (to econometrician) heterogeneous effects, and $(\beta', \gamma')' \in \mathbb{R}^{k_1+k_2}$ are unknown preference parameters to estimate[3]. Note that expression (2.1) is rather general. By properly re-organizing x_j's and w, expression (2.1) can accommodate both alternative-specific and agent-specific covariates. See Cameron and Trivedi (2005) pp. 498 for a more detailed discussion.

The specification of (2.1) parametrizes the nonparametric deterministic utilities in the model considered by Fox and Lazzati (2017). The utility of choosing $(0, 0)$ is normalized to zero, and the utility of choosing $(1, 0)$ ($(0, 1)$) is $v_1 + x_1'\beta + \epsilon_1$ ($v_2 + x_2'\beta + \epsilon_2$). Different from the regular binary choice model, the agent might gain additional payoff if the bundle is selected, i.e., the utility of choosing the bundle is the sum of the two stand-alone utilities plus the interaction term $\eta \cdot (w'\gamma)$ capturing either *complementary* ($\eta \cdot (w'\gamma) \geq 0$) or *substitution* ($\eta \cdot (w'\gamma) \leq 0$) effects[4]. (ϵ_1, ϵ_2) are idiosyncratic shocks associated with each stand-alone alternative, and η reflects unobserved heterogeneity for the bundle. For identification, we restrict $\eta > 0$, i.e., $w'\gamma$ determines the sign of the interaction term, and the presence of η allows agent-specific magnitudes.

Given the latent utility model (2.1), the observed dependent variable y_d is of the form

$$y_d = 1[U_d > U_{d'}, \forall d' \in \mathcal{D} \setminus d]. \tag{2.2}$$

[1] Fox and Lazzati (2017) show that this choice model is mathematically equivalent to a certain class of binary games. The approach developed in this paper can be applied to these game models.

[2] w and x_j's may have common elements.

[3] Note that here we implicitly assume that the coefficient on v_j is non-zero and normalize it to 1 without loss of generality (w.l.o.g.). This serves as usual scale normalization essential for many semi- and non-parametric discrete choice models. See, for example, Lewbel (2000) and Fox and Lazzati (2017), among many others.

[4] Gentzkow (2007) shows that this definition of complements and substitutes is equivalent to the classic definitions based on the sign of cross demand elasticity.

Let $z \equiv (v_1, v_2, x_1', x_2', w')'$. The probabilities of choosing $d = (0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$ conditional on z are expressed respectively as

$$P\left(\max\{v_1 + x_1'\beta + \epsilon_1, v_2 + x_2'\beta + \epsilon_2, v_1 + v_2 + (x_1 + x_2)'\beta + (\epsilon_1 + \epsilon_2) + \eta \cdot (w'\gamma)\} < 0|z\right),$$
$$P\left(\max\{0, v_2 + x_2'\beta + \epsilon_2, v_1 + v_2 + (x_1 + x_2)'\beta + (\epsilon_1 + \epsilon_2) + \eta \cdot (w'\gamma)\} < v_1 + x_1'\beta + \epsilon_1|z\right),$$
$$P\left(\max\{0, v_1 + x_1'\beta + \epsilon_1, v_1 + v_2 + (x_1 + x_2)'\beta + (\epsilon_1 + \epsilon_2) + \eta \cdot (w'\gamma)\} < v_2 + x_2'\beta + \epsilon_2|z\right),$$
$$P\left(\max\{0, v_1 + x_1'\beta + \epsilon_1, v_2 + x_2'\beta + \epsilon_2\} < v_1 + v_2 + (x_1 + x_2)'\beta + (\epsilon_1 + \epsilon_2) + \eta \cdot (w'\gamma)|z\right). \quad (2.3)$$

Denote $u_j = v_j + x_j'\beta$ for $j = 1, 2$. Assuming that $\lim_{v_1 \to -\infty} P(u_1 + \epsilon_1 < 0|z) = \lim_{v_1 \to -\infty} P(u_1 + u_2 + (\epsilon_1 + \epsilon_2) + \eta \cdot (w'\gamma) < 0|z) = 1$[5], we deduce from (2.3) that

$$\lim_{v_1 \to -\infty} P(y_{(0,0)} = 1|z) = P(u_2 + \epsilon_2 < 0|z), \quad (2.4)$$

$$\lim_{v_1 \to -\infty} P(y_{(0,1)} = 1|z) = P(u_2 + \epsilon_2 > 0|z), \quad (2.5)$$

and $\lim_{v_1 \to -\infty} P(y_{(1,0)} = 1|z) = \lim_{v_1 \to -\infty} P(y_{(1,1)} = 1|z) = 0$. The intuition is straightforward. If $d_1 = 0$ with probability approaching 1 as $v_1 \to -\infty$, then the agent chooses $(1,0)$ or $(1,1)$ with probability approaching 0, and hence the bundle choice problem reduces to a standard binary choice model (for alternative 2). If further $\text{Med}(\epsilon_2|z, \eta) = 0$ holds true, we then establish the following (conditional) identification inequality based on (2.4) and (2.5)

$$u_2 = v_2 + x_2'\beta \geq 0 \Leftrightarrow \lim_{v_1 \to -\infty} P(y_{(0,1)} = 1|z) \geq \lim_{v_1 \to -\infty} P(y_{(0,0)} = 1|z). \quad (2.6)$$

The derivation of (2.6) combines the insight of identification at infinity and identification at the median. The former is used for identifying endogenous selection models and censoring models (see e.g., Heckman (1990) and Chamberlain (1986)), and the latter is adopted in a wide range of semiparametric discrete choice models in line with the seminal work of Manski (1975, 1985, 1988).

Similar identification inequality can be obtained for alternative 1 by letting $v_2 \to -\infty$, i.e.,

$$u_1 = v_1 + x_1'\beta \geq 0 \Leftrightarrow \lim_{v_2 \to -\infty} P(y_{(1,0)} = 1|z) \geq \lim_{v_2 \to -\infty} P(y_{(0,0)} = 1|z). \quad (2.7)$$

Collectively, (2.6) and (2.7), in combination with the regularity conditions stated below, establish the identification of β.

Once β is identified, we treat it as a constant vector and move on to identify γ. Let $\mathcal{E} \equiv \{u_1 = u_2 = 0\}$. We deduce from (2.3) that

$$P(y_{(0,0)} = 1|z, \mathcal{E}) = P(\epsilon_1 < 0, \epsilon_2 < 0, \epsilon_1 + \epsilon_2 + \eta \cdot (w'\gamma) < 0|z, \mathcal{E}), \quad (2.8)$$

and

$$P(y_{(1,1)} = 1|z, \mathcal{E}) = P(\epsilon_1 + \eta \cdot (w'\gamma) > 0, \epsilon_2 + \eta \cdot (w'\gamma) > 0, \epsilon_1 + \epsilon_2 + \eta \cdot (w'\gamma) > 0|z, \mathcal{E}). \quad (2.9)$$

[5]This implicitly assumes sufficient variation in v_1 and continuity in its density.

If $P((\epsilon_1,\epsilon_2)<(0,0)|z,\eta)=P((\epsilon_1,\epsilon_2)>(0,0)|z,\eta)$[6] holds, then (2.8) and (2.9) imply the following identification inequality for $w'\gamma$

$$w'\gamma \geq 0 \Leftrightarrow P(y_{(1,1)}=1|z,\mathcal{E}) \geq P(y_{(0,0)}=1|z,\mathcal{E}). \tag{2.10}$$

To identify β and γ based on (2.6), (2.7), and (2.10), the following conditions are sufficient.

B1 The joint distribution of $(\epsilon_1,\epsilon_2,\eta)$ conditional on z is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2\times\mathbb{R}_+$, and $\text{Med}(\epsilon_1|z,\eta)=\text{Med}(\epsilon_2|z,\eta)=0$.

B2 v_1 (v_2) has almost everywhere (a.e.) positive Lebesgue density on $\mathbb{R}$ conditional on x_1 (x_2) and conditional on $\{v_2<-M\}$ ($\{v_1<-M\}$) for all positive constant M.

B3 $\lim_{v_1\to-\infty}P(u_1+\epsilon_1<0|z)=\lim_{v_2\to-\infty}P(u_2+\epsilon_2<0|z)=1$ and $\lim_{v_1\to-\infty}P(u_1+u_2+(\epsilon_1+\epsilon_2)+\eta\cdot(w'\gamma)<0|z)=\lim_{v_2\to-\infty}P(u_1+u_2+(\epsilon_1+\epsilon_2)+\eta\cdot(w'\gamma)<0|z)=1$.

B4 The support $\mathcal{X}_1$ ($\mathcal{X}_2$) of x_1 (x_2) conditional on $\{v_2<-M\}$ ($\{v_1<-M\}$) for all positive constant M is not contained in any proper linear subspace of $\mathbb{R}^{k_1}$.

B5 $\beta\in\mathcal{B}$, where $\mathcal{B}$ is a compact subset of $\mathbb{R}^{k_1}$.

R1 $P((\epsilon_1,\epsilon_2)<(0,0)|z,\eta)=P((\epsilon_1,\epsilon_2)>(0,0)|z,\eta)$.

R2 Let $w^{(1)}$ denote the first element of w and $\tilde{w}$ denote the sub-vector comprising the remaining elements of w. $w^{(1)}$ has a.e. positive Lebesgue density on $\mathbb{R}$ conditional on $\tilde{w}$ and conditional on $(v_1+x_1'\beta,v_2+x_2'\beta)$ in a neighborhood of $(v_1+x_1'\beta,v_2+x_2'\beta)$ near zero.

R3 The support $\tilde{\mathcal{W}}$ of $\tilde{w}$ conditional on $(v_1+x_1'\beta,v_2+x_2'\beta)$ in a neighborhood of $(v_1+x_1'\beta,v_2+x_2'\beta)$ near zero is not contained in any proper linear subspace of $\mathbb{R}^{k_2-1}$.

R4 $\gamma=(\gamma_1,\tilde{\gamma}')'\in\mathcal{R}$, where $\mathcal{R}=\{r=(r_1,...,r_{k_2})'\in\mathbb{R}^{k_2}|\|r\|=1,r_1\neq0\}$.

As discussed above, our approach involves identification at infinity with median independence restriction. Assumptions B1 - B3 are sufficient conditions for establishing the identification inequalities (2.6) and (2.7). Assumption R1 places a "median type restriction" on the joint distribution of (ϵ_1,ϵ_2), from which the identification inequality (2.10) follows. Note that Assumptions B1 and R1 allow general correlation among $(\epsilon_1,\epsilon_2,\eta)$ and flexible dependence of $(\epsilon_1,\epsilon_2,\eta)$ on observed covariates (e.g., conditional heteroskedasticity). This "distribution-free" property is desirable in many applications using micro-level data.

Assumptions B2 and R2 are standard for MS type of estimators, which ensures the point identification, as opposed to a set identification. Assumptions B4 and R3 are familiar full-rank conditions. Assumptions B5 and R4 are about the parameter space. We restrict the search of β in a

[6]Note that $P((\epsilon_1,\epsilon_2)<(0,0)|z,\eta)=P((\epsilon_1,\epsilon_2,\epsilon_1+\epsilon_2)<(0,0,0)|z,\eta)$ and $P((\epsilon_1,\epsilon_2)>(0,0)|z,\eta)=P((\epsilon_1,\epsilon_2,\epsilon_1+\epsilon_2)>(0,0,0)|z,\eta)$.

compact set. As the interaction effect is a multiple of the unobserved heterogeneity η, γ can only be identified up to a scale. Following a substantial literature, we normalize $\|\gamma\|$ to 1 and assume that the coefficient on $w^{(1)}$ is non-zero.

Our identification results are stated in the following theorem, which is proved in Appendix.

Theorem 1. *Suppose Assumptions B1 - B5 hold. Then β is identified. Furthermore, if Assumptions R1 - R4 also hold, γ is identified up to a scale.*

Remark. The identification results do not tell the comparative importance of the stand-alone utilities and the interaction effects in determining the choice of the bundle as the scale of the latter is controlled by η. With additional stochastic restriction(s) on unobservables $(\epsilon_1, \epsilon_2, \eta)$, the scale of η can be identified. For example, assume the joint PDF of $(\epsilon_1, \epsilon_2, \epsilon_1 + \epsilon_2)$ is symmetric around 0 conditional on (z, η). We can show that $u_1 + u_2 + \eta \cdot (w'\gamma) = 0 \Leftrightarrow P(y_{(0,0)} = 1|z, u_1 = u_2) = P(y_{(1,1)} = 1|z, u_1 = u_2)$. If we further assume that η is median independent of z with $\text{Med}(\eta|z) = q_\eta$, then q_η can be identified in a local quantile regression framework with the knowledge of $(x_1'\beta, x_2'\beta, w'\gamma)$. A detailed and thorough discussion on the identification of η is outside the scope of this work. We leave this topic to a separate paper.

3 Estimation

The identification inequalities established in (2.6), (2.7), and (2.10) translate into a two-stage MS estimation procedure. Each of the two stages is described in turn below.

Assume a random sample of n observations is generated from the model (2.1)-(2.2). Let δ_{n1} and δ_{n2} denote two positive non-stochastic series such that $\delta_{n1}, \delta_{n2} \to \infty$ as $n \to \infty$. In the first stage, we construct the localized MS estimator $\hat{\beta}$ for β, analogous to the MS estimator proposed in Manski (1975, 1985), defined as the maximizer over the parameter space $\mathcal{B}$, of the following objective function:

$$
\begin{aligned}
Q_{n1}(b) =& \frac{1}{n}\sum_{i=1}^{n}\left\{1[v_{i2} \le -\delta_{n2}]\left(y_{i(1,0)}1[v_{i1} + x_{i1}'b > 0] + y_{i(0,0)}1[v_{i1} + x_{i1}'b \le 0]\right)\right.\\
&\left.+1[v_{i1} \le -\delta_{n1}]\left(y_{i(0,1)}1[v_{i2} + x_{i2}'b > 0] + y_{i(0,0)}1[v_{i2} + x_{i2}'b \le 0]\right)\right\}. \qquad (3.1)
\end{aligned}
$$

Recall that the identification of γ is achieved at $v_1 + x_1'\beta = v_2 + x_2'\beta = 0$. In estimation, we use the $\hat{\beta}$ obtained in the first stage to replace β when constructing the indexes. Note that the probability measure of the conditioned set is zero in the presence of continuous regressors. Following the literature, we use observations whose (v_1, v_2) and (x_1, x_2) make $v_1 + x_1'\beta = v_2 + x_2'\beta = 0$ approximately hold for estimation. To this end, we introduce a standard kernel function $\mathcal{K}(\cdot, \cdot)$ that satisfies Assumption C4 stated below and two smoothing parameters h_{n1}, h_{n2} that satisfy $h_{n1}, h_{n2} \to 0$ as $n \to \infty$. Then we propose the kernel weighted MS estimator $\hat{\gamma}$ of γ maximizing the following objective function over the parameter space $\mathcal{R}$:

$$Q_{n2}(r) = \frac{1}{nh_{n1}h_{n2}} \sum_{i=1}^{n} \mathcal{K}\left(\frac{v_{i1} + x_{i1}'\hat{\beta}}{h_{n1}}, \frac{v_{i2} + x_{i2}'\hat{\beta}}{h_{n2}}\right) \left(y_{i(1,1)} 1[w_i' r > 0] + y_{i(0,0)} 1[w_i' r \leq 0]\right). \quad (3.2)$$

To secure the consistency of $\hat{\beta}$ and $\hat{\gamma}$, we need the following technical conditions.

C1 The data $\{(y_i', z_i')'\}_{i=1}^{n}$ are i.i.d. across i, where $y_i \equiv (y_{i(0,0)}, y_{i(0,1)}, y_{i(1,0)}, y_{i(1,1)})'$.

C2 δ_{n1} and δ_{n2} are sequences of positive numbers such that as $n \to \infty$: (i) $\delta_{n1} \to \infty$ and $\delta_{n2} \to \infty$, and (ii) $nP(v_1 \leq -\delta_{n1}) \to \infty$ and $nP(v_2 \leq -\delta_{n2}) \to \infty$[7].

C3 h_{n1} and h_{n2} are sequences of positive numbers such that as $n \to \infty$: (i) $\|\hat{\beta} - \beta\|/h_{n1} = o_p(1)$ and $\|\hat{\beta} - \beta\|/h_{n2} = o_p(1)$, (ii) $nh_{n1}h_{n2} \to \infty$, and (iii) $nh_{n1}h_{n2}/\log n \to \infty$.

C4 $\mathcal{K} : \mathbb{R}^2 \mapsto \mathbb{R}_+$ is continuously differentiable, takes non-zero values only on $[-1,1]^2$, and has bounded first derivatives. $\int_{\mathbb{R}^2} \mathcal{K}(u)\, du = 1$ and $\int_{\mathbb{R}^2} \|u\|\mathcal{K}(u)\, du < \infty$.

C5 Let $\nu \equiv (v_1 + x_1'\beta, v_2 + x_2'\beta)'$ and $h(r) \equiv y_{(1,1)} 1[w'r > 0] + y_{(0,0)} 1[w'r \leq 0]$. $\mathbb{E}[h(r)|\nu = \cdot]$ is continuously differentiable a.e. with bounded first derivatives.

C6 Let $f_\nu(\cdot)$ denote the joint probability density function of ν. $f_\nu(\cdot)$ is bounded from above on its support, continuously differentiable with bounded first derivatives, and strictly positive in a neighborhood of zero. $\mathbb{E}\|x_1\| < \infty$, $\mathbb{E}\|x_2\| < \infty$.

Assumptions C2 and C3 place mild restrictions on tuning parameters. Assumption C4 collects regularity conditions for kernel function $\mathcal{K}$, all of which are standard in the literature. These assumptions, together with Assumptions C5 and C6, are needed for proving the uniform convergence of the objective functions (3.1) and (3.2) to their population analogues.

The theorem below states that the two-stage procedure described in (3.1) and (3.2) gives consistent estimators of β and γ, whose proof is left to Appendix.

Theorem 2. *Suppose Assumptions B1 - B5, R1 - R4, and C1 - C4 hold. Then we have $\hat{\beta} \xrightarrow{p} \beta$ and $\hat{\gamma} \xrightarrow{p} \gamma$.*

Remark. The process of showing the asymptotic distribution of $\hat{\beta}$ and $\hat{\gamma}$ is rather involved and outside of the scope of this paper. The proof is a combination of the techniques in Andrews and Schafgans (1998), Kim and Pollard (1990), and Seo and Otsu (2018). To utilize the techniques in Andrews and Schafgans (1998), we need the tail thickness restrictions on v_1, v_2, ϵ_1, and ϵ_2. The convergence rates of $\hat{\beta}_1$ and $\hat{\beta}_2$ are slower than cube-root n, and the exactly rate depends on the relative tail thickness of v_1,v_2 to ϵ_1, ϵ_2. If the bias terms are asymptotic negligible and $\delta_{n1} = \delta_{n2} = \delta_n$, we conjecture that $\hat{\beta} - \beta = O_p([n \min\{P(v_{i1} \leq -\delta_{n1}), P(v_{i2} \leq -\delta_{n2})\}]^{-1/3})$. If the generated $\hat{\beta}$

[7] Assumption C2(ii) indicates that the choices of δ_{n1} and δ_{n2} rely on the tail behaviors of v_1 and v_2. For example, if v_1 is sub-Gaussian, Assumption C2(ii) requires that $\delta_{n1} = o(\sqrt{\log n})$, while for sub-exponential v_1, it is $\delta_{n1} = o(\log n)$.

has no effects on the asymptotics on $\hat{\gamma}$, we conjecture that $\hat{\gamma} - \gamma = O_p((nh_1h_2)^{-1/3})$. It is possible that adopting a smoothed MS approach (e.g., Horowitz (1992)) yields faster rates and asymptotically normal estimators. However, the inference based on smoothed estimators involves carefully choosing new bandwidths and sufficiently smooth kernel functions. We leave this topic to future research.

4 Extension to $J > 3$

Our approach can be applied to bundle choice models with $J > 3$. It suffices to illustrate the main idea through the case with $J = 4$ ($j \in \{0, 1, 2, 3\}$) as the same intuition runs through all general cases with $J > 3$.

Consider the model with choice set $\mathcal{J} = \{0, 1, 2, 3, (1, 2), (1, 3), (2, 3), (1, 2, 3)\}$, which can be equivalently expressed as the set $\mathcal{D} \equiv \{d | d = (d_1, d_2, d_3) \in \{0, 1\}^3\}$. An agent chooses $d \in \mathcal{D}$ that maximizes the latent utility

$$\begin{aligned} U_d = & \sum_{j=1}^{3}(v_j + x_j'\beta + \epsilon_j) \cdot d_j + \eta_{110} \cdot (w_1'\gamma_1) \cdot d_1 \cdot d_2 + \eta_{101} \cdot (w_2'\gamma_2) \cdot d_1 \cdot d_3 \\ & + \eta_{011} \cdot (w_3'\gamma_3) \cdot d_2 \cdot d_3 + \eta_{111} \cdot (w_4'\gamma_4) \cdot d_1 \cdot d_2 \cdot d_3 \end{aligned} \tag{4.1}$$

where $w_1, w_2, w_3, w_4 \in \mathbb{R}^{k_2}$ are observed covariates associated to each of the four bundles, and $\eta \equiv (\eta_{110}, \eta_{101}, \eta_{011}, \eta_{111})' \in \mathbb{R}_+^4$ capture bundle-specific unobserved heterogeneities. The specification of (4.1) are analogous to (2.1) but accommodates a more complex choice set. We emphasize that v_j can only affect the stand-alone utility associated to alternative j, while x_j's, $j = 1, 2, 3$, and w_l's, $l = 1, 2, 3, 4$, may have common elements. In the rest of this section, let $y_d \equiv 1[U_d > U_{d'}, \forall d' \in \mathcal{D} \setminus d]$, $z \equiv (v_1, v_2, v_3, x_1', x_2', x_3', w_1', w_2', w_3', w_4')'$, $\epsilon \equiv (\epsilon_1, \epsilon_2, \epsilon_3)'$, $u_j \equiv v_j + x_j'\beta$ for $j = 1, 2, 3$, $\Gamma_{110} \equiv \eta_{110} \cdot (w_1'\gamma_1)$, $\Gamma_{101} \equiv \eta_{101} \cdot (w_2'\gamma_2)$, $\Gamma_{011} \equiv \eta_{011} \cdot (w_3'\gamma_3)$, and $\Gamma_{111} \equiv \eta_{111} \cdot (w_4'\gamma_4)$.

To identify parameters β and γ_l's in model (4.1), the following conditions are sufficient.

- B1′ The joint distribution of (ϵ, η) conditional on z is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^3 \times \mathbb{R}_+^4$, and $\text{Med}(\epsilon_j | z, \eta) = 0$ for all $j = 1, 2, 3$.
- B2′ For all $j = 1, 2, 3$, v_j has a.e. positive Lebesgue density on $\mathbb{R}$ conditional on x_j and conditional on $\{v_m < -M\}$ for all positive constant M and all $m \neq j$.
- B3′ For all $j = 1, 2, 3$, (i) $\lim_{v_j \to -\infty} P(u_j + \epsilon_j < 0 | z) = 1$, (ii) $\lim_{v_j \to -\infty} P(u_j + u_m + (\epsilon_j + \epsilon_m) + \Gamma_{jm} < 0 | z) = 1$ for all $m \neq j$, where Γ_{jm} represents the interaction term associated with bundle $\{j, m\}$, and (iii) $\lim_{v_j \to -\infty} P(u_1 + u_2 + u_3 + (\epsilon_1 + \epsilon_2 + \epsilon_3) + \Gamma_{111} < 0 | z) = 1$.
- B4′ For all $j = 1, 2, 3$, the support $\mathcal{X}_j$ of x_j conditional on $\{v_m < -M\}$ for all positive constant M and all $m \neq j$ is not contained in any proper linear subspace of $\mathbb{R}^{k_1}$.

B5′ $\beta \in \mathcal{B}$, where $\mathcal{B}$ is a compact subset of $\mathbb{R}^{k_1}$.

R1′ (i) $P((\epsilon_1, \epsilon_2) < (0,0)|z,\eta) = P((\epsilon_1, \epsilon_2) > (0,0)|z,\eta)$, (ii) $P((\epsilon_1, \epsilon_3) < (0,0)|z,\eta) = P((\epsilon_1, \epsilon_3) > (0,0)|z,\eta)$, (iii) $P((\epsilon_2, \epsilon_3) < (0,0)|z,\eta) = P((\epsilon_2, \epsilon_3) > (0,0)|z,\eta)$, and (iv) $P((\epsilon_1, \epsilon_2, \epsilon_3) < (0,0,0)|z,\eta) = P((\epsilon_1, \epsilon_2, \epsilon_3) > (0,0,0)|z,\eta)$.

R2′ Let $w^{(1)}$ denote the first element of a vector w and $\tilde{w}$ denote the sub-vector comprising the remaining elements of w. Then, (i) $w^{(1)}$ ($w^{(2)}$, $w^{(3)}$) has a.e. positive Lebesgue density on $\mathbb{R}$ conditional on $\tilde{w}_1$ ($\tilde{w}_2$, $\tilde{w}_3$) and conditional on (u_1, u_2) ((u_1, u_3), (u_2, u_3)) in a neighborhood of (u_1, u_2) ((u_1, u_3), (u_2, u_3)) near zero, (ii) $w^{(4)}$ has a.e. positive Lebesgue density on $\mathbb{R}$ conditional on $\tilde{w}_4$ and conditional on $(u_1, u_2, u_3, w_1'\gamma_1, w_2'\gamma_2, w_3'\gamma_3)$ in a neighborhood of $(u_1, u_2, u_3, w_1'\gamma_1, w_2'\gamma_2, w_3'\gamma_3)$ near zero.

R3′ (i) The support $\tilde{\mathcal{W}}_1$ ($\tilde{\mathcal{W}}_2$, $\tilde{\mathcal{W}}_3$) of $\tilde{w}_1$ ($\tilde{w}_2$, $\tilde{w}_3$) conditional on $\{v_3 < -M\}$ ($\{v_2 < -M\}$, $\{v_1 < -M\}$) for all positive constant M and conditional on (u_1, u_2) ((u_1, u_3), (u_2, u_3)) in a neighborhood of (u_1, u_2) ((u_1, u_3), (u_2, u_3)) near zero is not contained in any proper linear subspace of $\mathbb{R}^{k_2-1}$, and (ii) The support $\tilde{\mathcal{W}}_4$ of $\tilde{w}_4$ conditional on $(u_1, u_2, u_3, w_1'\gamma_1, w_2'\gamma_2, w_3'\gamma_3)$ in a neighborhood of $(u_1, u_2, u_3, w_1'\gamma_1, w_2'\gamma_2, w_3'\gamma_3)$ near zero is not contained in any proper linear subspace of $\mathbb{R}^{k_2-1}$.

R4′ $\gamma_l \in \mathcal{R}$ for all $l = 1,2,3,4$, where $\mathcal{R} = \{r = (r_1, ..., r_{k_2})' \in \mathbb{R}^{k_2} | \|r\| = 1, r_1 \neq 0\}$.

Assumptions B1′ - B5′ and R1′ - R4′ are parallel to Assumptions B1 - B5 and R1 - R4. Next we outline how these assumptions can help achieve the identification of β and γ_l's.

First, note that Assumption B3′ leads to

$$\lim_{v_2, v_3 \to -\infty} P(y_{(1,0,0)} = 1|z) = P(u_1 + \epsilon_1 > 0|z) \text{ and } \lim_{v_2, v_3 \to -\infty} P(y_{(0,0,0)} = 1|z) = P(u_1 + \epsilon_1 < 0|z).$$

Together with Assumption B1′, they imply that

$$u_1 = v_1 + x_1'\beta \geq 0 \Leftrightarrow \lim_{v_2, v_3 \to -\infty} P(y_{(1,0,0)} = 1|z) \geq \lim_{v_2, v_3 \to -\infty} P(y_{(0,0,0)} = 1|z). \tag{4.2}$$

Similarly, we can show

$$u_2 = v_2 + x_2'\beta \geq 0 \Leftrightarrow \lim_{v_1, v_3 \to -\infty} P(y_{(0,1,0)} = 1|z) \geq \lim_{v_1, v_3 \to -\infty} P(y_{(0,0,0)} = 1|z), \tag{4.3}$$

and

$$u_3 = v_3 + x_3'\beta \geq 0 \Leftrightarrow \lim_{v_1, v_2 \to -\infty} P(y_{(0,0,1)} = 1|z) \geq \lim_{v_1, v_2 \to -\infty} P(y_{(0,0,0)} = 1|z). \tag{4.4}$$

Then using identification inequalities (4.2) - (4.4) in combination of regularity conditions B2′ and B4′, β can be identified using similar arguments for proving Theorem 1.

To identify γ_1, we first let $v_3 \to -\infty$ so that the model reduces to the $J = 3$ case by Assumption B3′. Then conditioning on $\mathcal{E}_1 \equiv \{u_1 = u_2 = 0\}$, we write

$$\lim_{v_3 \to -\infty} P(y_{(1,1,0)} = 1|z, \mathcal{E}_1) = P(\epsilon_1 + \Gamma_{110} > 0, \epsilon_2 + \Gamma_{110} > 0, \epsilon_1 + \epsilon_2 + \Gamma_{110} > 0|z, \mathcal{E}_1),$$

and

$$\lim_{v_3 \to -\infty} P(y_{(0,0,0)} = 1|z, \mathcal{E}_1) = P(\epsilon_1 < 0, \epsilon_2 < 0, \epsilon_1 + \epsilon_2 + \Gamma_{110} < 0|z, \mathcal{E}_1).$$

It then follows from Assumption R1′ that

$$w_1'\gamma_1 \geq 0 \Leftrightarrow \lim_{v_3 \to -\infty} P(y_{(1,1,0)} = 1|z, \mathcal{E}_1) \geq \lim_{v_3 \to -\infty} P(y_{(0,0,0)} = 1|z, \mathcal{E}_1), \tag{4.5}$$

The identification inequality (4.5), along with regularity conditions R2′ and R3′, can be used to identify γ_1 (up to a scale). Similar ideas apply to the identification of γ_2 and γ_3.

With β, γ_1, γ_2, and γ_3 identified, γ_4 can be identified on the set $\mathcal{E}_4 \equiv \{u_1 = u_2 = u_3 = 0, w_1'\gamma_1 = w_2'\gamma_2 = w_3'\gamma_3 = 0\}$. Specifically, under Assumption R1′, we have

$$w_4'\gamma_4 \geq 0 \Leftrightarrow P(y_{(1,1,1)} = 1|z, \mathcal{E}_4) \geq P(y_{(0,0,0)} = 1|z, \mathcal{E}_4) \tag{4.6}$$

Then using similar arguments as those used for identifying γ in Section 2, we can establish the identification of γ_4 by (4.6) and Assumptions R2′ and R3′.

The theorem below summarizes our identification results for model (4.1). The proof is omitted as it uses similar arguments for proving Theorem 1.

Theorem 3. *Suppose Assumptions B1′ - B5′ hold. Then β is identified. Furthermore, if Assumptions R1′ - R4′ also hold, γ_l, $l = 1, 2, 3, 4$, are identified up to a scale.*

5 Conclusions

This paper studies (point) identification and estimation of preference coefficients in semiparametric discrete choice models for bundles, by means of a combination of "identification at infinity" and median independence restrictions. We provide estimators and show their consistency. This paper leaves some open questions, e.g., the asymptotics and small sample properties of the estimators. We leave these for future research.

References

ANDREWS, D. W. AND M. M. SCHAFGANS (1998): "Semiparametric estimation of the intercept of a sample selection model," *The Review of Economic Studies*, 65, 497–517.

CAMERON, A. C. AND P. K. TRIVEDI (2005): *Microeconometrics: methods and applications*, Cambridge university press.

CHAMBERLAIN, G. (1986): "Asymptotic efficiency in semi-parametric models with censoring," *Journal of Econometrics*, 32, 189–218.

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

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

HECKMAN, J. (1990): "Varieties of selection bias," *American Economic Review*, 80, 313–318.

HOROWITZ, J. L. (1992): "A smoothed maximum score estimator for the binary response model," *Econometrica*, 60, 505–531.

KIM, J. AND D. POLLARD (1990): "Cube root asymptotics," *The Annals of Statistics*, 18, 191–219.

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

MANSKI, C. F. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics*, 3, 205–228.

——— (1985): "Semiparametric analysis of discrete response," *Journal of Econometrics*, 27, 313–333.

——— (1988): "Identification of binary response models," *Journal of the American Statistical Association*, 83, 729–738.

NEWEY, W. K. AND D. MCFADDEN (1994): "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 4, 2111–2245.

SEO, M. H. AND T. OTSU (2018): "Local M-estimation with discontinuous criterion for dependent and limited observations," *The Annals of Statistics*, 46, 344–369.

A Appendix

Proof of Theorem 1. It suffices to show the identification of β based on identification inequality (2.6). By the dominated convergence theorem (DCT), Assumptions B3 and B1, (2.6) implies that β maximizes objective function $Q_1(b) \equiv \lim_{v_1 \to -\infty} \mathbb{E}[(P(y_{(0,1)} = 1|z) - P(y_{(0,0)} = 1|z)) \cdot sgn(v_2 + x_2'b)|v_1]$ where $sgn(\cdot)$ denote the sign function. To show that β attains a unique maximum, let $b \in \mathcal{B}$ such that $Q_1(b) = Q_1(\beta)$. We want to show that $b = \beta$ must hold. To see this, first note that if $\lim_{v_1 \to -\infty} P[(x_2'\beta < -v_2 < x_2'b) \cup (x_2'b < -v_2 < x_2'\beta)|v_1] > 0$, β and b yield different values of the $sgn(\cdot)$ function in $Q_1(\cdot)$ with strictly positive probability, and thus $Q_1(b) < Q_1(\beta)$. This implies that for all b satisfying $Q_1(b) = Q_1(\beta)$, $\lim_{v_1 \to -\infty} P[(x_2'\beta < -v_2 < x_2'b) \cup (x_2'b < -v_2 < x_2'\beta)|v_1] = 0$ must hold, which is equivalent to $\lim_{v_1 \to -\infty} P(x_2'\beta = x_2'b|v_1) = 1$ under Assumption B2. Then the desired result follows from Assumption B4.

Next, we prove the identification of γ. Note that by Assumption R1, (2.10) implies that γ maximizes objective function $Q_2(r) \equiv \mathbb{E}[(P(y_{(1,1)} = 1|z,\mathcal{E}) - P(y_{(0,0)} = 1|z,\mathcal{E})) \cdot sgn(w'\gamma)|\mathcal{E}]$. The remaining task is to show that γ is unique in $\mathcal{R}$, i.e., $Q_2(r) = Q_2(\gamma)$ implies $r = \gamma$. Here we assume $\gamma_1 > 0$ w.l.o.g. as the case $\gamma_1 < 0$ is symmetric.

First note that for any $r \in \mathcal{R}$ such that $Q_2(r) = Q_2(\gamma)$, $r_1 > 0$ must hold, for otherwise by Assumption R2 we have $P(w_1 r_1 + \tilde{w}'\tilde{r} < 0 < w_1\gamma_1 + \tilde{w}'\tilde{\gamma}|\mathcal{E}) = P(w_1 > -\tilde{w}'\tilde{r}/r_1, w_1 > -\tilde{w}'\tilde{\gamma}/\gamma_1|\mathcal{E}) > 0$. Then γ and r yield different realized values of the sign function in $Q_2(\cdot)$ with strictly positive probability, and thus $Q_2(\gamma) > Q_2(r)$.

Focus on the case with $r_1 > 0$ to write

$$\begin{aligned}&P([w_1 r_1 + \tilde{w}'\tilde{r} < 0 < w_1\gamma_1 + \tilde{w}'\tilde{\gamma}] \cup [w_1\gamma_1 + \tilde{w}'\tilde{\gamma} < 0 < w_1 r_1 + \tilde{w}'\tilde{r}]|\mathcal{E})\\ =&P([-\tilde{w}'\tilde{\gamma}/\gamma_1 < w_1 < -\tilde{w}'\tilde{r}/r_1] \cup [-\tilde{w}'\tilde{r}/r_1 < w_1 < -\tilde{w}'\tilde{\gamma}/\gamma_1]|\mathcal{E}),\end{aligned}$$

which implies that to make $Q_2(r) = Q_2(\gamma)$ hold we must have $P(\tilde{w}'\tilde{r}/r_1 = \tilde{w}'\tilde{\gamma}/\gamma_1|\mathcal{E}) = 1$ by Assumption R2. However, whenever r is not a scalar multiple of γ, $P(\tilde{w}'\tilde{r}/r_1 = \tilde{w}'\tilde{\gamma}/\gamma_1|\mathcal{E}) = 1$ implies that $\tilde{\mathcal{W}}$ is contained in a proper linear subspace of $\mathbb{R}^{k_2-1}$ a.e., violating Assumption R3. As a result, we must have r being a scalar multiple of γ, which leads to the desired result $r = \gamma$ as $\|r\| = \|\gamma\| = 1$ by the construction of the parameter space $\mathcal{R}$ in Assumption R4. □

Proof of Theorem 2. We use the idea of Theorem 2.1 in Newey and McFadden (1994) to show the consistency. Note that this theorem requires two conditions. The first one is that the objective function uniformly converges to the population of the objective function. The second one is that the true value of parameters uniquely maximizes the population of the objective function.

To ease technical proof, we assume that $P(v_1 = -\infty), P(v_2 = -\infty) > 0$. In the case when $P(v_j = -\infty) = 0$, we need $nP(v_{ij} \le -\delta_{nj}) \to \infty$ so that the number of the observations used for estimation tends to infinity and the objective function is not degenerated. By Theorem 1, β uniquely maximizes

$$\begin{aligned}&E\left(1\left[v_1 + x_1'b > 0\right]1\left[y = (1,0)\right] + 1\left[v_1 + x_1'b \le 0\right]1\left[y_i = (0,0)\right]|v_2 = -\infty\right)\\ &+ E\left(1\left[v_2 + x_2'b > 0\right]1\left[y = (0,1)\right] + 1\left[v_2 + x_2'b \le 0\right]1\left[y_i = (0,0)\right]|v_1 = -\infty\right).\end{aligned}$$

That is equivalent to that β uniquely maximizes

$$\begin{aligned}&E\left(1\left[v_1 + x_1'b > 0\right]1\left[y = (1,0)\right]1\left[v_2 = -\infty\right] + 1\left[v_1 + x_1'b \le 0\right]1\left[y_i = (0,0)\right]1\left[v_2 = -\infty\right]\right)\\ &+ E\left(1\left[v_2 + x_2'b > 0\right]1\left[y = (0,1)\right]1\left[v_1 = -\infty\right] + 1\left[v_2 + x_2'b \le 0\right]1\left[y_i = (0,0)\right]1\left[v_1 = -\infty\right]\right).\end{aligned}$$

The sample objective function obviously satisfies the technical conditions in Kim and Pollard (1990) for the uniform convergence, by Assumptions B5, C1, and C2. Using the same arguments in Kim and Pollard (1990), we have

$$\begin{aligned}&n^{-1}\sum_{i=1}^{n}\left\{1\left[v_{i1} + x_{i1}'b > 0\right]1\left[y_i = (1,0)\right]1\left[v_{i2} \le -\delta_{n2}\right] + 1\left[v_{i1} + x_{i1}'b \le 0\right]1\left[y_i = (0,0)\right]1\left[v_{i2} \le -\delta_{n2}\right]\right\}\\ &\xrightarrow{P} E\left(1\left[v_1 + x_1'b > 0\right]1\left[y = (1,0)\right]1\left[v_2 = -\infty\right] + 1\left[v_1 + x_1'b \le 0\right]1\left[y_i = (0,0)\right]1\left[v_2 = -\infty\right]\right),\end{aligned}$$

and

$$n^{-1}\sum_{i=1}^{n}\left\{1\left[v_{i2}+x_{i2}'b>0\right]1\left[y_i=(0,1)\right]1\left[v_{i1}\leq-\delta_{n1}\right]+1\left[v_{i2}+x_{i2}'b\leq 0\right]1\left[y_i=(0,0)\right]1\left[v_{i1}\leq-\delta_{n1}\right]\right\}$$
$$\xrightarrow{P} E\left(1\left[v_2+x_2'b>0\right]1\left[y=(0,1)\right]1\left[v_1=-\infty\right]+1\left[v_2+x_2'b\leq 0\right]1\left[y_i=(0,0)\right]1\left[v_1=-\infty\right]\right),$$

uniformly over the parameter space of β. Then by Theorem 2.1 in Newey and McFadden (1994),

$$\hat{\beta}\xrightarrow{P}\beta.$$

If we know the true values of β,then $\hat{\gamma}$ may be obtained from maximizing

$$n^{-1}\sum_{i=1}^{n}\frac{1}{h_1h_2}\mathcal{K}\left(\frac{v_{i1}+x_{i1}'\beta}{h_1},\frac{v_{i2}+x_{i2}'\beta}{h_2}\right)\left\{1\left[y_i=(1,1)\right]1\left[\omega_i'r>0\right]+1\left[y_i=(0,0)\right]1\left[\omega_i'r\leq 0\right]\right\}.$$

This infeasible sample objective function is a special case in Seo and Otsu (2018). Using the arguments in Seo and Otsu (2018) and by Assumptions B5, C1, C3 and C4, the above uniformly converges to

$$E\left[\left\{1\left[y_i=(1,1)\right]1\left[\omega_i'r>0\right]+1\left[y_i=(0,0)\right]1\left[\omega_i'r\leq 0\right]\right\}|u_1=0,u_2=0\right]f_{u_1u_2}(0,0). \tag{A.1}$$

Assumptions C3 and C4 imply that

$$n^{-1}\sum_{i=1}^{n}\frac{1}{h_1h_2}\left[\mathcal{K}\left(\frac{v_{i1}+x_{i1}'\hat{\beta}}{h_1},\frac{v_{i2}+x_{i2}'\hat{\beta}}{h_2}\right)-\mathcal{K}\left(\frac{v_{i1}+x_{i1}'\beta}{h_1},\frac{v_{i2}+x_{i2}'\beta}{h_2}\right)\right]=o_P(1),$$

and it is not a function of r. Combing the results so far yields that

$$n^{-1}\sum_{i=1}^{n}\frac{1}{h_1h_2}\mathcal{K}\left(\frac{v_{i1}+x_{i1}'\hat{\beta}}{h_1},\frac{v_{i2}+x_{i2}'\hat{\beta}}{h_2}\right)\left\{1\left[y_i=(1,1)\right]1\left[\omega_i'r>0\right]+1\left[y_i=(0,0)\right]1\left[\omega_i'r\leq 0\right]\right\}$$

converges to equation (A.1) uniformly for r in $\mathbb{B}^{k_2}$.

By Theorem 2, γ uniquely maximizes(A.1) in $\mathbb{B}^{k_2}$. Applying Theorem 2.1 in Newey and McFadden (1994) again delivers

$$\hat{\gamma}\xrightarrow{P}\gamma.$$

□