Durham University Business School

Department of Economics & Finance

Semiparametric Estimation of the Random Utility Model with Rank-Ordered Choice Data

Jin Yan and Hong Il Yoo

Working Paper No. 02, 2017

Department Economics and Finance
Durham University Business School
Mill Hill Lane
Durham DH1 3LB, UK
Tel: +44 (0)191 3345200
https://www.dur.ac.uk/business/research/economics/

Semiparametric Estimation of the Random Utility Model with Rank-Ordered Choice Data*

Jin Yan† Hong Il Yoo‡

April 27, 2017

Abstract

We propose two semiparametric methods for estimating the random utility model using rank-ordered choice data. The framework is "semiparametric" in that the utility index includes finite dimensional preference parameters but the error term follows an unspecified distribution. Our methods allow for a flexible form of heteroskedasticity across individuals. With complete preference rankings, our methods also allow for heteroskedastic and correlated errors across alternatives, as well as a variety of random coefficients distributions. The baseline method we develop is the generalized maximum score (GMS) estimator, which is strongly consistent but follows a non-standard asymptotic distribution. To facilitate statistical inferences, we make extra regularity assumptions and develop the smoothed GMS estimator, which is asymptotically normal. Monte Carlo experiments show that our estimators perform favorably against popular parametric estimators under a variety of stochastic specifications.

Keywords: Rank-ordered; Random utility; Semiparametric estimation; Smoothing

JEL Classification: C14, C35.

*We thank Xu Cheng, Liran Einav, Jeremy Fox, Bruce Hansen, Han Hong, Arthur Lewbel, Taisuke Otsu, Joris Pinkse, Jack Porter and seminar participants at the 2015 Tsinghua Econometric Conference, the 12th International Symposium on Econometric Theory and Applications, the 2016 Asian and European Meetings of the Econometric Society, Academia Sinica, Newcastle University, Sun Yat-Sen University and the Chinese University of Hong Kong for valuable comments and discussions. We acknowledge funding support provided by the Hong Kong Research Grants Council General Research Fund 2014/2015 (Project No.14413214) and six anonymous referee reports for the project proposal. All errors are ours.

†Department of Economics, The Chinese University of Hong Kong. Email: jyan@cuhk.edu.hk.

‡Durham University Business School, Durham University. Email: h.i.yoo@durham.ac.uk.

1 Introduction

Rank-ordered choices can be elicited using the same type of survey as multinomial choices, specifically one that presents an individual with a finite set of mutually exclusive alternatives. The two elicitation formats may be distinguished by the amount of information that is available to the econometrician. A multinomial choice reports the individual's "choice" or most preferred alternative from the set, whereas a rank-ordered choice reports further about the individual's preference ordering such as her second and third preferences: see for example Hausman and Ruud (1987), Calfee *et al.* (2001), and Train and Winston (2007). One rank-ordered choice observation provides a similar amount of information as several multinomial choice observations, in the sense that it allows inferring what the individual's choices would have been if her more preferred alternatives were not available. This allows fewer individuals to be interviewed to achieve a given level of statistical precision and, as Scarpa *et al.* (2011) point out, the resulting logistic advantages could be substantial for many non-market valuation studies which involve a narrowly defined population of interest.

We develop semiparametric methods for the estimation of random utility models using rank-ordered choice data. Despite the wide availability of parametric counterparts, such semiparametric methods remain almost undeveloped to date. The random utility function of interest has a typical structure: it comprises a systematic component or utility index varying with finite-dimensional explanatory variables, and an additive stochastic component or error term. The objective is to estimate preference parameters, referring to the coefficients on the explanatory variables. The methods are semiparametric in that they maintain the usual parametric form of the systematic component but place only non-parametric restrictions on the stochastic component.

The parametric methods are equally well-established for multinomial choice and rank-ordered choice data. In most cases, an analysis of multinomial choice data involves the maximum (simulated) likelihood estimation of one of four models: multinomial logit (MNL), nested MNL, multinomial probit (MNP), and random coefficients (or "mixed") MNL. Each model assumes a different parametric distribution of the stochastic component, and has its own rank-ordered choice counterpart which shares the same assumption: rank-ordered logit (ROL) of Beggs *et al.* (1981), nested ROL of Dagsvik and Liu (2009), rank-ordered probit (ROP) of Layton and Levine (2003), and mixed ROL of Layton (2000) and Calfee *et al.* (2001). Building on Falmagne (1978) and Barberá and Pattanaik (1986), McFadden (1986) provides a technique that can be applied to translate any parametric multinomial choice model into the corresponding rank-ordered choice model.

The literature on the semiparametric methods is more lopsided. For multinomial choice data, several alternative methods exist including Manski (1975), Ruud (1986), Lee (1995), Lewbel (2000) and Fox (2007). The special case of binomial choice data has attracted even greater attention, and the respectable menagerie includes Ruud (1983), Manski (1985), Han (1987), Horowitz (1992), Klein and Spady (1993) and Sherman (1993) to name a few. When it comes to rank-ordered choice data, we are aware of only one study that aimed at semiparametric estimation of the preference parameters, namely Hausman and Ruud (1987). In that study, the weighted M-estimator (WME) of Ruud (1983, 1986) is generalized for use with rank-ordered choice data, whereas the original WME was intended for use with binomial and multinomial choice data. The generalized WME allows consistent estimation of the ratios of the preference parameters despite stochastic misspecification, but there are two drawbacks affecting its empirical applicability. As the authors acknowledge, the estimator's consistency is confined to the ratios of the coefficients on continuous explanatory variables, and its asymptotic distribution is unknown outside a special case of Newey (1986).

In this paper, we propose a pair of new semiparametric methods for rank-ordered choice data. We call them the generalized maximum score (GMS) estimator and the smoothed generalized maximum score (SGMS) estimator respectively. Both estimators are consistent under more general assumptions concerning explanatory variables than the generalized WME of Hausman and Ruud (1987). Roughly speaking, if one of q explanatory variables is continuous, each estimator allows consistent estimation of all coefficients up to scale regardless of whether the other $q-1$ variables are continuous or discrete. Moreover, the SGMS estimator is asymptotically normal, meaning that it is amenable to the application of usual Wald-type tests. The GMS estimator follows a non-standard asymptotic distribution, but it does not require extra smoothness assumptions.

The GMS estimator generalizes the pairwise maximum score (MS) estimator of Fox (2007), which has been developed for use with multinomial choice data and is a modern extension of the classic MS estimator due to Manski (1975). Suppose that the individual faces J alternatives. A multinomial choice observation allows one to infer the outcomes of $J-1$ pairwise comparisons where each pair comprises her actual choice and an unchosen alternative. A rank-ordered choice observation allows one to infer the outcomes of more pairwise comparisons. For example, in case the individual ranks all J alternatives from best to worst, her rank-ordered choice would allow one to learn the outcomes of all possible $J(J-1)/2$ pairwise comparisons. The GMS estimator extends the MS estimator by incorporating such extra information. The key identification condition comprises an intuitively plausible set of inequalities: in a pairwise comparison, if one alternative's systematic utility exceeds the other's, its chance of being ranked better also exceeds the other's.

The GMS estimator inherits all attractive properties of the MS estimator, two of which are particularly relevant to empirical applications. First, the GMS estimator allows the econometrician to be agnostic about the form of interpersonal heteroskedasticity or "scale heterogeneity" (Hensher *et al.*, 1999; Fiebig *et al.*, 2010), referring to variations in the overall scale of utility across individuals.[1] This property is desirable because in most studies, the exact form of interpersonal heteroskedasticity matters only to the extent that its misspecification leads to inconsistent estimation of the core preference parameters. Second, the GMS estimator is consistent when the data generating process (DGP) comprises an arbitrary mixture of different models, provided that it is consistent for each component model. The empirical evidence from behavioral economics (Harrison and Rutström, 2009; Conte *et al.*, 2011) supports the notion that characterizing observed choices requires more than one model, but the parametric estimation of a mixture model demands the exact knowledge of the number and composition of component models.

In addition, when each individual ranks all alternatives from best to worst, the GMS estimator is substantively more flexible than the MS estimator. As we discuss in details later, the GMS estimator is consistent for all popular parametric models exhibiting flexible substitution patterns, whereas the MS estimator is not.[2] The GMS estimator therefore delivers what empiricists may expect from the use of a semiparametric method, namely the ability to estimate all popular parametric models consistently on top of other types of models. This is an interesting finding because in the parametric framework, the advantage of using rank-ordered choice data instead of multinomial choice data is limited to efficiency gains (Hausman and Ruud, 1987), and a multinomial choice model may be more robust to stochastic misspecification than its rank-ordered choice

[1] This property explains a major difference between the GMS estimator and the maximum rank correlation (MRC) estimator of Han (1987) and Sherman (1993). The GMS method utilizes the observed ranking information and does pairwise comparisons of alternatives *within* each individual, allowing the conditional joint distribution of the error terms to vary across individuals. In comparison, the MCR estimator does pairwise comparisons *between* individuals and requires the error terms to be independent of the explanatory variables, ruling out the possibility of heteroskedasticity across individuals.

[2] The difference arises because the complete ranking information allows us to replace the "exchangeability" assumption (Goeree *et al.*, 2005; Fox, 2007) with a much weaker assumption of zero conditional median.

counterpart (Yan and Yoo, 2014). The efficiency-bias tradeoff does not apply in the semiparametric framework, where the advantage of using rank-ordered choice data also includes robustness to a wider variety of DGPs. We note that in most studies on rank-ordered choices, the complete preference rankings are elicited as required for this result (Hausman and Ruud, 1987; Calfee *et al.*, 2001; Caparrós *et al.*, 2008; Scarpa *et al.*, 2011; Yoo and Doiron, 2013; Oviedo and Yoo, 2016).

The SGMS estimator offers the same types of practical benefits as the GMS estimator, and addresses the latter's major drawbacks in return for requiring extra smoothness assumptions. The GMS estimator's rate of convergence is $N^{-1/3}$, which is slower than the usual rate of $N^{-1/2}$, and it follows a non-standard asymptotic distribution of Kim and Pollard (1990) which is inconvenient for use with conventional hypothesis tests. These properties are inherited from the MS estimator, and arise because the objective function is a sum of step functions. Horowitz (1992) develops the smoothed maximum score (SMS) estimator for binomial choice data which replaces the step functions with smooth functions, and Yan (2012) extends the method to multinomial choice data. Our smoothing technique follows this tradition. We show that the SGMS estimator's convergence rate can be made arbitrarily close to $N^{-1/2}$ under extra smoothness conditions and that its asymptotic distribution is normal, with a covariance matrix which can be consistently estimated.

The remainder of this paper is organized as follows. Section 2 develops the GMS estimator and compares it with popular parametric methods. Section 3 develops the SGMS estimator. Section 4 presents the Monte Carlo evidence on the finite sample properties of the proposed estimators. Section 5 concludes.

2 The Model and the Generalized Maximum Score Estimator

2.1 A Random Utility Framework and Rank-Ordered Choice Data

Consider the standard random utility model. An individual in the population of interest faces a finite collection of alternatives. Let $\mathbb{J} = \{1, \ldots, J\}$ denote the set of alternatives and let $J \geq 2$ be the number of alternatives contained in $\mathbb{J}$. The utility from choosing alternative j, u_j, is assumed as follows:

$$u_j = \boldsymbol{x}_j' \boldsymbol{\beta} + \varepsilon_j \quad \forall\, j \in \mathbb{J}, \tag{1}$$

where $\boldsymbol{x}_j \equiv (x_{j,1}, \ldots, x_{j,q})' \in \mathbb{R}^q$ is an observed q-vector containing the attributes of alternative j and their interactions with the individual's characteristics, $\boldsymbol{\beta} \equiv (\beta_1, \ldots, \beta_q)' \in \mathbb{R}^q$ is the preference parameter vector of interest, and ε_j is the unobserved component of utility to the econometrician. The utility index $\boldsymbol{x}_j'\boldsymbol{\beta}$ is often called systematic (or deterministic) utility, as opposed to the error term ε_j which is called unsystematic (or stochastic) utility. Let $\boldsymbol{X} \equiv (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_J)' \in \mathbb{R}^{J \times q}$ be the matrix of the explanatory variables and $\boldsymbol{\varepsilon} \equiv (\varepsilon_1, \ldots, \varepsilon_J)' \in \mathbb{R}^J$ be the vector of the error terms.

Let $r(j, \boldsymbol{u})$ denote the latent or potentially unobserved ranking of alternative j, based on the vector of underlying alternative-specific utilities $\boldsymbol{u} \equiv (u_1, u_2, \ldots, u_J)' \in \mathbb{R}^J$. We shall follow the notational convention that $r(j, \boldsymbol{u}) = q$ when j is the q^{th} best alternative in the choice set $\mathbb{J}$, meaning that a smaller ranking value indicates a more preferred alternative. For instance, suppose that $J = 4$ and $u_3 > u_4 > u_1 > u_2$. Then, $r(1, \boldsymbol{u}) = 3$, $r(2, \boldsymbol{u}) = 4$, $r(3, \boldsymbol{u}) = 1$, and $r(4, \boldsymbol{u}) = 2$. Purely for technical convenience, our notation handles any utility tie by assigning a better ranking to the alternative that happens to have a smaller numeric label. For instance, suppose instead that $u_3 > u_4 = u_1 > u_2$. Then, $r(1, \boldsymbol{u}) = 2$ and $r(4, \boldsymbol{u}) = 3$ since numeric label "1" is smaller than "4".

A more formal definition of the latent ranking that incorporates our notational convention is as follows. Let $\mathbb{T}(j, \boldsymbol{u})$ be the set of alternatives with the same utility as alternative j. $A(k, \mathbb{T}(j, \boldsymbol{u}))$ maps element $k \in \mathbb{T}(j, \boldsymbol{u})$ one-to-one onto the integers $\{0, \ldots, |\mathbb{T}(j, \boldsymbol{u})| - 1\}$, where $|\mathbb{T}|$ is the number of alternatives in $\mathbb{T}$. For any two alternatives $k, l \in \mathbb{T}(j, \boldsymbol{u})$, $A(k, \mathbb{T}(j, \boldsymbol{u})) < A(l, \mathbb{T}(j, \boldsymbol{u}))$ if and only if $k < l$. For any $j \in \mathbb{J}$, denote its latent ranking as

$$r(j, \boldsymbol{u}) \equiv L(j, \boldsymbol{u}) + 1 + A(j, \mathbb{T}(j, \boldsymbol{u})), \tag{2}$$

where $L(j, \boldsymbol{u})$ denotes the number of alternatives that yield strictly larger utility than alternative j for the individual. Notice that in the absence of utility ties, the last term on the right-hand side of (2) is irrelevant to the latent ranking value since $A(j,\, \mathbb{T}(j, \boldsymbol{u})) = 0$. By definition (2), there is a one-to-one mapping between the set $\{r(j, \boldsymbol{u}) :\, j = 1, \ldots, J\}$ and the set $\{1, \ldots, J\}$.

Next, let r_j denote the reported or actually observed ranking of alternative j, and $\boldsymbol{r} \equiv (r_1, \ldots, r_J)' \in \mathbb{N}^J$ be the vector of the reported rankings of all J alternatives in $\mathbb{J}$. We shall maintain that the reported ranking r_j coincides with the latent ranking $r(j,\boldsymbol{u})$ in case the individual reports the complete ranking of alternatives, and is a censored version of the latent ranking in case she reports a partial ranking. To facilitate further discussion, suppose that the individual reports the ranking of her best M alternatives where $1 \leq M \leq J-1$, and leaves that of the other $J-M$ alternatives unspecified. As before, suppose that $J=4$ and $u_3 > u_4 > u_1 > u_2$. In case $M=3$, the complete ranking is observed since the individual reports her best, second-best, and third-best alternatives, allowing the econometrician to infer that the only remaining alternative is her worst one, $\boldsymbol{r} = (r_1, r_2, r_3, r_4) = (3, 4, 1, 2)$, and that each alternative's reported ranking is identical to its latent ranking. In case $M=2$, only a partial ranking is observed since the individual reports her best and second best alternatives, and the econometrician cannot tell whether alternative 1 is preferable to alternative 2, $\boldsymbol{r} = (3, 3, 1, 2)$, so the reported ranking r_2 is no longer the same as the latent ranking $r(2,\boldsymbol{u})$. Finally, in case $M=1$, the resulting partial ranking observation is identical to a multinomial choice observation since the individual reports only her best alternative, $\boldsymbol{r} = (2, 2, 1, 2)$.

A more formal definition of the reported ranking that incorporates the above discussion is as follows. Let the random set $\mathbb{M}$ ($\mathbb{M} \subset \mathbb{J}$) denote the set of the best M alternatives for the individual, that is, $\mathbb{M} \equiv \{j : r(j,\boldsymbol{u}) \leq M\}$. The reported ranking of alternative j, then, follows the observation rule

$$r_j = \begin{cases} r(j,\boldsymbol{u}) & if \quad r(j,\boldsymbol{u}) \leq M, \, or\, equivalently, \; j \in \mathbb{M}, \\ M+1 & if \quad r(j,\boldsymbol{u}) > M, \, or\, equivalently, \; j \in \mathbb{J} \setminus \mathbb{M}. \end{cases} \tag{3}$$

When $M = J-1$, the complete ranking is observed. When $M=1$, the resulting partial ranking is observationally equivalent to a multinomial choice. The intermediate cases of partial rankings, which occur when $2 \leq M < J-1$ and $J > 3$, are much less common in empirical studies though not unprecedented.[3]

2.2 The Generalized Maximum Score Estimator

This section establishes strong consistency of the Generalized Maximum Score (GMS) estimator, the first of two semiparametric methods that we propose. The GMS estimator is semiparametric in the sense that it allows the econometrician to estimate the preference parameter vector $\boldsymbol{\beta}$ consistently, without commiting to

[3]See for example Layton (2000) and Train and Winston (2007), both of which analyze data on the best and second-best alternatives; their data structures are $M=2$ and $J>3$ according to our notations.

a specific parametric form of the conditional distribution of the error vector given observed attributes $\boldsymbol{\varepsilon}|\boldsymbol{X}$.

Our first assumption pertains to sampling.

Assumption 1. *$\{(\boldsymbol{r}_n, \boldsymbol{X}_n, \boldsymbol{\varepsilon}_n) : n = 1, \ldots, N\}$ is a random sample of $(\boldsymbol{r}, \boldsymbol{X}, \boldsymbol{\varepsilon})$, where $\boldsymbol{r}_n \equiv (r_{n1}, \ldots, r_{nJ})' \in \mathbb{N}^J$, $\boldsymbol{X}_n \equiv (\boldsymbol{x}_{n1}, \ldots, \boldsymbol{x}_{nJ})' \in \mathbb{R}^{J\times q}$, and $\boldsymbol{\varepsilon}_n \equiv (\varepsilon_{n1}, \ldots, \varepsilon_{nJ})' \in \mathbb{R}^J$. For each individual $n = 1, \ldots, N$, $(\boldsymbol{r}_n, \boldsymbol{X}_n)$ is observed.*

Assumption 1 states that we have N observations of $(\boldsymbol{r}, \boldsymbol{X})$, indexed by n, and individuals are independently and identically distributed (*i.i.d.*). For the latter reason, we drop subscript n to avoid notational clutter except when it is needed for clarification.[4]

As usual in discrete choice modeling, identification of the preference vector $\boldsymbol{\beta}$ requires scale normalization since they are unique only up to scale.[5] When a parametric form of the conditional distribution of $\boldsymbol{\varepsilon}|\boldsymbol{X}$ is specified, identification is almost always achieved by normalizing a scale parameter of that distribution.[6] But when no parametric form is specified, no scale parameter is available for normalization. In a semiparametric framework, identification is therefore achieved by normalizing $\boldsymbol{\beta}$ instead.

Subject to the prior knowledge that some element of $\boldsymbol{\beta}$ is non-zero, we can normalize the magnitude of that element.[7] Without loss of generality, we assume that $|\beta_1| = 1$. Let $\tilde{\boldsymbol{\beta}} \equiv (\beta_2, \ldots, \beta_q)' \in \mathbb{R}^{q-1}$ be the vector containing the other elements of $\boldsymbol{\beta}$. The following assumption imposes a requirement on the space of the preference parameters.

Assumption 2. *The preference parameter vector $\boldsymbol{\beta} \in \mathbb{B}$, where $\mathbb{B} \equiv \{-1, 1\} \times \tilde{\mathbb{B}}$, $\tilde{\mathbb{B}}$ is a compact subset of $\mathbb{R}^{q-1}$, and $q \geq 2$.*

Next, we state Assumption 3 which presents a key identification condition pertaining to strong consistency of the GMS estimator. This assumption implicitly places a restriction on the conditional distribution of $\boldsymbol{\varepsilon}|\boldsymbol{X}$, albeit it is a non-parametric restriction that is satisfied by a range of parametric functional forms, some of which we will discuss in the subsequent section. Denote the systematic utility of alternative j as $v_j \equiv \boldsymbol{x}_j'\boldsymbol{\beta}$ for any alternative $j \in \mathbb{J}$.

[4]Throughout this paper, we use n to denote an individual, and j, k, l to denote alternatives.
[5]Multiplying both $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ by any positive constant leads to the same rank-ordered choice data.
[6]For instance, in the binomial probit model, the variance of the conditional distribution is assumed to be one.
[7]For example, economists may agree that the coefficient on the own price variable is negative.

Assumption 3. *For any pair of alternatives* $j, k \in \mathbb{J}$ *and for almost every* $\boldsymbol{X}$,

$$v_j > v_k \;\; if \text{ and only } if$$

$$P(r_j < r_k|\boldsymbol{X}) > P(r_k < r_j|\boldsymbol{X}), \tag{4}$$

i.e., alternative j *generates more systematic utility than alternative* k *if and only if there is a higher chance that* j *is preferable to* k $(r_j < r_k)$ *than the reverse* $(r_k < r_j)$, *conditional on almost all explanatory vectors.*

Assumption 3 immediately implies that $v_j = v_k$ if and only if $P(r_j < r_k|\boldsymbol{X}) = P(r_j > r_k|\boldsymbol{X})$, *i.e.*, alternatives j and k have the same systematic utility if and only if the probability that alternative j is preferable to alternative k is the same as the probability that alternative k is preferable to alternative j.

Two special types of rank-ordered choice data are worth highlighting. First, when $M = 1$, the individual reports only her best alternative and we have multinomial choice data. In this case, alternative j is ranked above alternative k $(r_j < r_k)$ if and only if j is ranked as the best alternative $(r_j = 1)$, so we have

$$P(r_j < r_k|\boldsymbol{X}) = P(r_j = 1|\boldsymbol{X}). \tag{5}$$

If we replace $P(r_j < r_k|\boldsymbol{X})$ with $P(r_j = 1|\boldsymbol{X})$ and replace $P(r_k < r_j|\boldsymbol{X})$ with $P(r_k = 1|\boldsymbol{X})$ in (4), then Assumption 3 becomes the monotonicity property of choice probabilities (Manski, 1975), *i.e.*, the ranking of the choice probability of an alternative is the same as the ranking of the systematic utility of the alternative for any given individual.[8]

Second, when $M = J - 1$, the individual ranks all alternatives from best to worst, and we have fully rank-ordered choice data. With this complete ranking information, we can compare the utilities between any two alternatives. Without loss of generality, let's focus on a pair of alternatives (j, k) such that $j < k$. Alternative j is ranked above alternative k if and only if the utility from choosing alternative j is larger than

[8]See Fox (2007) for a detailed discussion of sufficient conditions for the monotonicity property of choice probabilities.

the utility from choosing alternative k, so we have[9]

$$
\begin{aligned}
P(r_j < r_k|\boldsymbol{X}) &= P(u_j \geq u_k|\boldsymbol{X}) \\
&= P(\varepsilon_k - \varepsilon_j \leq v_j - v_k|\boldsymbol{X}).
\end{aligned} \tag{6}
$$

The "only if" part holds under the definition of ranking $\boldsymbol{r}$, and the "if" part is a direct result of complete ranking. The first equality of (6) may not hold if we only observe a partial ranking, *i.e.*, $M < J-1$. This is because $r_j < r_k$ naturally implies $u_j \geq u_k$, but $u_j \geq u_k$ may not imply $r_j < r_k$. When neither alternative j nor k belongs to the set $\mathbb{M}$, both of them are observed with the same ranking, $M+1$, even if $u_j > u_k$.

For any pair of alternatives, assume that the conditional distribution of $\varepsilon_k - \varepsilon_j$ is a strictly increasing function. Then the well-known (pairwise) zero conditional median (ZCM) restriction, $median(\varepsilon_k - \varepsilon_j|\boldsymbol{X}) = 0$, is a necessary and sufficient condition for Assumption 3 when a complete ranking of J alternatives is available. The proof is straightforward.[10] Notice that $P(r_j < r_k|\boldsymbol{X}) + P(r_k < r_j|\boldsymbol{X}) = 1$ when the choice set is fully rank-ordered. For "necessity", Assumption 3 implies that $v_j - v_k = 0$ if and only if $P(r_j < r_k|\boldsymbol{X}) = 1/2$, or equivalently, $P(\varepsilon_k - \varepsilon_j \leq v_j - v_k|\boldsymbol{X}) = 1/2$ by (6). For "sufficiency", the ZCM assumption implies that $v_j > v_k$ if and only if $P(r_j < r_k|\boldsymbol{X}) > 1/2$ by (6), or equivalently, $P(r_j < r_k|\boldsymbol{X}) > P(r_k < r_j|\boldsymbol{X})$.

Next, we describe the intuition of applying Assumption 3 to construct the GMS estimator for $\boldsymbol{\beta}$. Let $1(\cdot)$ be an indicator function that equals one if the event in the parenthesis is true and zero otherwise, and let $\boldsymbol{b} \equiv (b_1, \tilde{\boldsymbol{b}}')'$ be any vector in the parameters space $\mathbb{B}$. Under Assumption 3, if $\boldsymbol{x}_j'\boldsymbol{\beta} > \boldsymbol{x}_k'\boldsymbol{\beta}$, then event $r_j < r_k$ is more likely to occur than event $r_k < r_j$; if $\boldsymbol{x}_k'\boldsymbol{\beta} > \boldsymbol{x}_j'\boldsymbol{\beta}$, then event $r_k < r_j$ is more likely to be true than event $r_j < r_k$; and if $\boldsymbol{x}_j'\boldsymbol{\beta} = \boldsymbol{x}_k'\boldsymbol{\beta}$, then event $r_j < r_k$ has the same chance to be true as event $r_k < r_j$. Therefore, the expected value of the following match

$$
\begin{aligned}
m_{jk}(\boldsymbol{b}) &= 1(r_j < r_k)\cdot 1(\boldsymbol{x}_j'\boldsymbol{b} > \boldsymbol{x}_k'\boldsymbol{b}) + 1(r_k < r_j)\cdot 1(\boldsymbol{x}_k'\boldsymbol{b} > \boldsymbol{x}_j'\boldsymbol{b}) + 1(r_j < r_k)\cdot 1(\boldsymbol{x}_j'\boldsymbol{b} = \boldsymbol{x}_k'\boldsymbol{b}) \\
&= 1(r_j < r_k)\cdot 1(\boldsymbol{x}_j'\boldsymbol{b} \geq \boldsymbol{x}_k'\boldsymbol{b}) + 1(r_k < r_j)\cdot 1(\boldsymbol{x}_k'\boldsymbol{b} > \boldsymbol{x}_j'\boldsymbol{b})
\end{aligned} \tag{7}
$$

should be maximized at the true preference parameter vector $\boldsymbol{\beta}$ over $\boldsymbol{b} \in \mathbb{B}$. Define $\boldsymbol{x}_{nj}'\boldsymbol{b}$ as the $\boldsymbol{b}$-utility index

[9] If $j > k$, then $P(r_j < r_k|\boldsymbol{X}) = P(u_j > u_k|\boldsymbol{X})$. This is because we break ties using function $A(\cdot, \mathbb{T}(j))$, and rank alternative k above alternative j if $k < j$ when $k \in \mathbb{T}(j)$.

[10] This proof does not apply to partially rank-ordered choice data, *e.g.*, multinomial choice data, because the first equality in (6) does not hold. Goeree *et al.* (2005) give an example showing that the ZCM assumption is not sufficient for the monotonicity property of the choice probabilities.

of alternative j for individual n. Applying the analogy principle, we propose a semiparametric estimator, $\boldsymbol{b}_N \equiv (b_{N,1}, \tilde{\boldsymbol{b}}_N')' \in \mathbb{B}$, for $\boldsymbol{\beta}$ as follows:

$$\boldsymbol{b}_N \in argmax_{\boldsymbol{b}\in\mathbb{B}}\, Q_N(\boldsymbol{b}), \tag{8}$$

where

$$Q_N(\boldsymbol{b}) = N^{-1}\sum_{n=1}^{N}\sum_{1\le j<k\le J}\left[1(r_{nj} < r_{nk}) \cdot 1(\boldsymbol{x}_{nj}'\boldsymbol{b} \ge \boldsymbol{x}_{nk}'\boldsymbol{b}) + 1(r_{nk} < r_{nj}) \cdot 1(\boldsymbol{x}_{nk}'\boldsymbol{b} > \boldsymbol{x}_{nj}'\boldsymbol{b})\right]. \tag{9}$$

In the special case of $M = 1$, *i.e.*, when we have multinomial choice data, the estimator $\boldsymbol{b}_N$ defined by (8) becomes the pairwise maximum score (MS) estimator of Fox (2007). When $J = 2$ or we have binomial choice data, the estimator $\boldsymbol{b}_N$ becomes the MS estimator of Manski (1985). For this reason, the estimator $\boldsymbol{b}_N$ will be called the generalized maximum score (GMS) estimator.

When all the explanatory variables are discrete, we can always find another vector in the neighborhood of $\boldsymbol{\beta}$ that generates the same ranking of utility indexes as $\boldsymbol{\beta}$. To achieve point identification, we need to impose an extra assumption on the explanatory variables, namely, we need a continuous explanatory variable conditional on other explanatory variables. Next, we define a few notations and state the restrictions on explanatory variables formally in Assumption 4.

Since only differences in utilities matter to the observed outcome of random utility maximization, we shall assume $\boldsymbol{x}_J = \boldsymbol{0}$ without any loss of generality.[11] Next, let $\boldsymbol{x}_{jk} \equiv (x_{jk,1}, \ldots, x_{jk,q})' \in \mathbb{R}^q$ denote the difference between the explanatory vectors of alternatives j and k, *i.e.*, $\boldsymbol{x}_{jk} \equiv \boldsymbol{x}_j - \boldsymbol{x}_k$. In Assumption 2, we assumed that the first preference parameter has non-zero value. For each alternative $j \in \mathbb{J}$, partition the vector $\boldsymbol{x}_j$ into $(x_{j,1}, \tilde{\boldsymbol{x}}_j')'$, where $x_{j,1}$ is the first element of $\boldsymbol{x}_j$ and $\tilde{\boldsymbol{x}}_j \equiv (x_{jk,2}, \ldots, x_{jk,q})' \in \mathbb{R}^{q-1}$ refers to the remainder. So the first element of $\boldsymbol{x}_{jk}$ is $x_{jk,1} = x_{j,1} - x_{k,1}$ and its remaining elements are included in vector $\tilde{\boldsymbol{x}}_{jk} = \tilde{\boldsymbol{x}}_j - \tilde{\boldsymbol{x}}_k$. Define $\tilde{\boldsymbol{X}} = (\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_J)' \in \mathbb{R}^{J\times(q-1)}$. Vectors $\boldsymbol{x}_{nj}$, $\boldsymbol{x}_{njk}$, and $\tilde{\boldsymbol{x}}_{njk}$ are the nth observation of vectors $\boldsymbol{x}_j$, $\boldsymbol{x}_{jk}$, and $\tilde{\boldsymbol{x}}_{jk}$, respectively. Matrices $\boldsymbol{X}_n$ and $\tilde{\boldsymbol{X}}_n$ are the nth observation of matrices $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$, respectively.

[11] If $\boldsymbol{x}_J \neq 0$ initially, one can recode $\boldsymbol{x}_j$ as $\boldsymbol{x}_j - \boldsymbol{x}_J$ for all $j \in \mathbb{J}$ including $j = J$.

Assumption 4. *The following statements are true.*

(a) *For any pair of alternatives $j, k \in \mathbb{J}$, the density function of $x_{jk,1}$ conditional on $\tilde{\boldsymbol{x}}_{jk}$, $g_{jk}(x_{jk,1}|\tilde{\boldsymbol{x}}_{jk})$, is positive everywhere on $\mathbb{R}$ for almost every $\tilde{\boldsymbol{x}}_{jk}$.*

(b) *For any constant vector $\boldsymbol{c} \equiv (c_1, \ldots, c_q)' \in \mathbb{R}^q$, $P(\boldsymbol{X}\boldsymbol{c} = \boldsymbol{0}) = 1$ if and only if $\boldsymbol{c} = \boldsymbol{0}$.*

Assumption 4 is sufficient to show that other vectors $\boldsymbol{b} \in \mathbb{B}$ would yield different values for the probability limit of the objective function $Q_N(\boldsymbol{b})$ from the true parameter vector $\boldsymbol{\beta}$. Assumption 4(a) avoids the local failure of identification, which is important for semiparametric setting. Assumption 4(b) is analogous to the full-rank condition for the binomial choice model, which prevents the global failure of identification. The following theorem establishes strong consistency of the GMS estimator. Appendix provides the proofs of all theorems stated in the main text.

Theorem 1. *Let Assumptions* 1-4 *hold. The GMS estimator $\boldsymbol{b}_N$ defined in* (8) *converges almost surely to $\boldsymbol{\beta}$, the true preference parameter vector in the data generating process.*

2.3 Comparisons with Parametric Methods

From empiricists' perspectives, the question of paramount interest would be how flexible the semiparametric model is in comparision with parameteric models that one may consider. Modern desktop computing power makes this question especially relevant. Standard computing resources of today can handle the estimation of models that feature fairly flexible, albeit parametric, error structures.

When applied to data on complete rankings, *i.e.*, $M = J - 1$, the GMS estimator postulates a semiparametric model that nests all popular parametric models and any finite mixture of those models, provided that the explanatory vectors satisfy regularity conditions such as Assumption 4. In most studies on rank-ordered choices, the complete rankings are elicited as required for this result.[12] Such a degree of flexibility is not something to be taken for granted. For instance, the MS estimator (Manski, 1975; Fox, 2007) using multinomial choice data is consistent for a family of parametric models featuring exchangeable errors (*e.g.*, multinomial logit and multinomial probit with equicorrelated errors), but not for those parametric models

[12] See for example, Hausman and Ruud (1987), Calfee *et al.* (2001), Caparrós *et al.* (2008), Scarpa *et al.* (2011), Yoo and Doiron (2013), and Oviedo and Yoo (2016).

that feature more flexible error structures (*e.g.*, nested multinomial logit, multinomial probit with a general error covariance matrix, and mixed logit).

This section elaborates on the semiparametric model that the GMS estimator postulates, and its comparisons with popular parametric models. To clarify the notion of interpersonal heteroskedasticity here (and later, unobserved interpersonal heterogeneity), we reinstate individual subscript n. With a slight abuse of notations, an observationally equivalent form of equation (1) may be specified to express the utility that individual n derives from alternative j as

$$u_{nj} = \sigma_n \times (\boldsymbol{x}'_{nj}\boldsymbol{\beta}) + \varepsilon_{nj} \text{ for } n = 1, 2, ..., N \text{ and } j \in \mathbb{J}, \tag{10}$$

where the new parameter $\sigma_n \in \mathbb{R}^1_+$ captures that portion of the overall scale of utility which varies across individuals.[13] Equivalently, σ_n may be also described as a parameter that is inversely proportional to that portion of error variance which varies across individuals. Consistent estimation of a parametric model requires the correct specification of both the joint density of errors $\boldsymbol{\varepsilon}_n|\boldsymbol{X}_n$ and the functional form of σ_n. The GMS estimator allows both requirements to be relaxed substantially.

Regardless of the depth of rankings observed (*i.e.*, for every M such that $1 \leq M \leq J - 1$), the GMS estimator is consistent for the semiparametric model that accommodates any form of interpersonal heteroskedasticity via σ_n. For verification, note that when $v_{nj} \equiv \boldsymbol{x}'_{nj}\boldsymbol{\beta}$ and $v_{nk} \equiv \boldsymbol{x}'_{nk}\boldsymbol{\beta}$ satisfy the inequality stated in Assumption 3, so does any positive multiple of this pair, $\sigma_n \times v_{nj}$ and $\sigma_n \times v_{nk}$. The GMS estimator, therefore, allows the empiricists to be agnostic about the exact functional form of σ_n. This is a desirable property because in most studies, σ_n demands attention only to the extent that it must be correctly specified for consistent estimation of the preference parameter vector $\boldsymbol{\beta}$.

The remainder of this section assumes the use of complete rankings ($M = J - 1$). This allows the semiparametric model to accommodate any model that satisfies the pairwise zero conditional median (ZCM)

[13]Since an affine transformation of utilities does not alter observed behavior, the random utility specification (10) is observationally equivalent to $u_{nj} = \boldsymbol{x}'_{nj}\boldsymbol{\beta} + \varepsilon_{nj}/\sigma_n$. The slight abuse of notations refers to that ε_j in equation (1) corresponds to $\varepsilon_{nj}/\sigma_n$, rather than ε_{nj} alone. Note that the presence of a parameter like σ_n does not affect any of our earlier results because they do not rely on ε_{nj} having a standardized scale.

restriction, *i.e.*,

$$median(\varepsilon_{nk} - \varepsilon_{nj}|\boldsymbol{X}_n) = 0 \ for \ any \ j,k \in \mathbb{J}, \ where \ j \neq k, \tag{11}$$

which is then a necessary and sufficient condition for Assumption 3 as long as the distribution of $(\varepsilon_{nk} - \varepsilon_{nj})|\boldsymbol{X}_n$ is a strictly increasing function: see Section 2.2. In comparison, any parametric model involves a much stronger set of restrictions affecting other moments too, since the density of $\boldsymbol{\varepsilon}_n|\boldsymbol{X}_n$ is specified in full detail.

The semiparametric model based on (11) offers considerable flexibility not only over possible distributions of idiosyncratic errors, but also over possible distributions of random coefficients. To see this latter aspect, note that one may view $\boldsymbol{\varepsilon}_n$ as composite errors comprising individual-specific coefficients heterogeneity $\boldsymbol{\eta}_n$ (that has the same dimension as $\boldsymbol{\beta}$) and purely idiosyncratic errors $\boldsymbol{\epsilon}_n$ (that has the same dimension as $\boldsymbol{\varepsilon}_n$) such that a typical entry in $\boldsymbol{\varepsilon}_n \equiv \boldsymbol{X}_n \boldsymbol{\eta}_n + \boldsymbol{\epsilon}_n$ is

$$\varepsilon_{nj} \equiv \boldsymbol{x}_{nj}^{'} \boldsymbol{\eta}_n + \epsilon_{nj}. \tag{12}$$

Suppose now that idiosyncratic errors $\boldsymbol{\epsilon}_n$ satisfy the pairwise ZCM restriction, $median(\epsilon_{nk} - \epsilon_{nj}|\boldsymbol{X}_n) = 0$ for any $j,k \in \mathbb{J}$, and the usual random coefficients modeling assumption, $(\boldsymbol{\eta}_n \perp \boldsymbol{\epsilon}_n)|\boldsymbol{X}_n$, holds. Then, as long as individual heterogeneity has ZCM, *i.e.*, $median(\boldsymbol{\eta}_n|\boldsymbol{X}_n) = \boldsymbol{0}$, the composite errors $\boldsymbol{\varepsilon}_n$ satisfy the pairwise ZCM restriction in (11) too: differencing two composite errors results in a linear combination of conditionally independent random variables, $(\boldsymbol{x}_{nk} - \boldsymbol{x}_{nj})'\boldsymbol{\eta}_n$ and $(\epsilon_{nk} - \epsilon_{nj})$, each of which has the conditional median of zero.[14] In comparison, a parametric random coefficients model places more rigid restrictions on the distribution of individual heterogeneity $\boldsymbol{\eta}_n$, because the density of $\boldsymbol{\eta}_n|\boldsymbol{X}_n$ needs be specified in full detail much as that of $\boldsymbol{\epsilon}_n|\boldsymbol{X}_n$.

It is easy to verify that the semiparametric model accommodates the classic troika of parametric random utility models, logit (or ROL), nested logit (or nested ROL), and probit (or ROP). All three models assume away interprsonal heteroskedasticity by setting $\sigma_n = 1 \ \forall \ n = 1, 2, ..., N$, and assume an idiosyncratic error density $\boldsymbol{\varepsilon}_n|\boldsymbol{X}_n$ that implies the pairwise ZCM condition. In case of logit, the idiosyncractic errors are *i.i.d.*

[14]Each element in $\boldsymbol{\beta}$ may be interpreted as the median of a certain random preference coefficient whereas the corresponding element in $\boldsymbol{\eta}_n$ measures the individual-specific deviation around this median.

extreme value type 1 over alternatives and, as the celebrated result of McFadden (1974) shows, differencing two errors results in a standard logistic random variable that is symmetric around 0. The nested logit directly generalizes the logit model by specifying the joint density of $\boldsymbol{\varepsilon}_n|\boldsymbol{X}_n$ as a generalized extreme value (GEV) distribution. This distribution allows for a *positive* correlation between ε_{nj} and ε_{nk} in case alternatives j and k belong to the same "nest" or pre-specified subset of $\mathbb{J}$. Differencing two GEV errors still results in a logistic random variable that is symmetric around 0, though it may not have the unit scale. Finally, in its unrestricted form, the probit model generalizes the nested logit model by specifying the multivariate normal density $\boldsymbol{\varepsilon}_n|\boldsymbol{X}_n \sim N(\mathbf{0}, \boldsymbol{V}_\varepsilon)$ that allows for heteroskedasticity of ε_{nj} over alternatives j, and also for *any sign* of correlation between ε_{nj} and ε_{nk}. Differencing two zero-mean multivariate normal variables results in a zero-mean normal variable which is symmetric around its mean.

Random coefficients or "mixed "logit (or mixed ROL) models have become the workhorse of empirical modeling in the recent decade. The semiparametric model accommodates the most popular variant of mixed logit models, as well as their extensions. In the context of error decomposition (12), a mixed logit model has idiosyncratic errors $\boldsymbol{\epsilon}_n|\boldsymbol{X}_n$ as *i.i.d.* extreme value type 1 over alternatives and incorporates a non-degenerate "mixing" distribution of random heterogeneity $\boldsymbol{\eta}_n|\boldsymbol{X}_n$. While the mixing distribution may take any parametric form, specifying $\boldsymbol{\eta}_n|\boldsymbol{X}_n \sim N(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\eta}})$ is by far the most popular choice, so much so that the generic name "mixed logit" is often associated with this normal-mixture logit model. Differencing the normal-mixture logit model's composite errors results in a linear combination of conditionally independent zero-mean normal and standard logistic random variables, that has the conditional median of zero. Fiebig *et al.* (2010) augment the normal-mixture logit model with a log-normally distributed interpersonal heteroskedasticity parameter σ_n, and find that the resulting Generalized Multinomial Logit model is capable of capturing the multimodality of preferences. Because the semiparametric model allows for any form of σ_n, it nests the Generalized Multinomial Logit model too. Greene *et al.* (2006) extend the normal-mixture model in another direction, by allowing the variance-covariance of random coefficients, $Var(\boldsymbol{\eta}_n|\boldsymbol{X}_n)$ to vary with $\boldsymbol{X}_n$. The semiparametric model nests their heteroskedastic normal-mixture logit model too, since this type of generalization does not affect the conditional median of $\boldsymbol{\eta}_n$.

The semiparametric model also accommodates any finite mixture of the aforementioned parametric models, and more generally that of all parametric models satisfying the pairwise ZCM restriction. In other words, it allows for that the data generating process may comprise different parametric models for different individuals.[15] This flexibility comes from the fact that the GMS estimator does not require the density of $\boldsymbol{\varepsilon}_n | \boldsymbol{X}_n$ to be identical across all individuals $n = 1, 2, ..., N$, as long as each individual's density of the error vector satisfies the pairwise ZCM restriction. While the finite mixture of parametric models approach has not been applied to the analysis of multinomial choice or rank-ordered choice data, it has motivated influential studies in the binomial choice analysis of decision making under risk (Harrison and Rutström, 2009; Conte *et al.*, 2011). The findings from that literature unambiguously suggest that postulating only one parametric model for all individuals may be an unduly restrictive assumption.

3 The Smoothed GMS Estimator

The maximum score (MS) type estimator is $N^{1/3}$-consistent, and its asymptotic distribution is studied in Cavanagh (1987) and Kim and Pollard (1990). Kim and Pollard have shown that $N^{1/3}$ times the centered MS estimator converges in distribution to the random variable that maximizes a certain Gaussian process for the binomial choice data. Their general theorem can be applied to multinomial choice data and rank-ordered choice data too. However, the resulting asymptotic distribution is too complicated to be used for inference in empirical applications. Abrevaya and Huang (2005) prove that the standard bootstrap is not consistent for the MS estimator. Delgado *et al.* (2001) show that subsampling consistently estimates the asymptotic distribution of the test statistic of the MS estimator for the binomial choice data. But subsampling has efficiency loss, and its computational cost is very high for the MS or GMS estimator because a global search method is needed to solve the maximization problem for each subsample.

In this section, we propose an estimator that complements the GMS estimator by addressing these practical limitations, in return for making some additional assumptions. In the context of Manski's (1985) binomial choice MS estimator, Horowitz (1992) develops a smoothed maximum score (SMS) estimator that replaces the step functions with smooth functions. Yan (2012) applies this technique to derive a smoothed version of Fox's (2007) multinomial choice MS estimator. We use the same approach to derive a smoothed GMS (SGMS) estimator, which offers similar benefits as its SMS predecessors. Specifically, we show that the SGMS

[15] For example, the nested logit model may generate 1/3 of the sample while the mixed logit may generate the rest.

estimator has a convergence rate which is faster than $N^{-1/3}$ under extra smoothness conditions, and also that it is asymptotically normal.

3.1 The Smoothed GMS Estimator and its Asymptotic Properties

The objective function in (9) can be rewritten as

$$Q_N(\boldsymbol{b}) = N^{-1}\sum_{n=1}^{N}\sum_{1\leq j<k\leq J}\left\{[1(r_{nj}<r_{nk})-1(r_{nk}<r_{nj})]\cdot 1(\boldsymbol{x}'_{njk}\boldsymbol{b}\geq 0)+1(r_{nk}<r_{nj})\right\} \tag{13}$$

by replacing $1(\boldsymbol{x}'_{nkj}\boldsymbol{b}>0)$ with $[1-1(\boldsymbol{x}'_{njk}\boldsymbol{b}\geq 0)]$.

The indicator function of $\boldsymbol{b}$ in (13) can be replaced by a sufficiently smooth function $K(\cdot)$, where $K(\cdot)$ is analogous to a cumulative distribution function. Application of the smoothing idea in Horowitz (1992) to the right-hand side of (13) yields a smoothed version of GMS (SGMS) estimator

$$\boldsymbol{b}_N^S \in \underset{\boldsymbol{b}\in\mathbb{B}}{argmax} Q_N^S(\boldsymbol{b}, h_N), \tag{14}$$

where

$$Q_N^S(\boldsymbol{b}, h_N) = N^{-1}\sum_{n=1}^{N}\sum_{1\leq j<k\leq J}\left\{[1(r_{nj}<r_{nk})-1(r_{nk}<r_{nj})]\cdot K\left(\boldsymbol{x}'_{njk}\boldsymbol{b}/h_N\right)+1(r_{nk}<r_{nj})\right\} \tag{15}$$

and $\{h_N : N=1,2,\ldots\}$ is a sequence of strictly positive real numbers satisfying $lim_{N\to\infty}h_N=0$.

The next condition states the requirements that the smooth function $K(\cdot)$ should satisfy for the SGMS estimator $\boldsymbol{b}_N^S$ to be consistent.

Condition 1. Let $K(\cdot)$ be a function on $\mathbb{R}$ such that:

(a) $|K(v)|<C$ for some finite C and all $v\in(-\infty,\infty)$; and

(b) $lim_{v\to-\infty}K(v)=0$ and $lim_{v\to\infty}K(v)=1$.

Theorem 2. *Let Assumptions* 1-4 *and Condition* 1 *hold. The SGMS estimator* $\boldsymbol{b}_N^S\in\mathbb{B}$ *defined in* (14) *converges almost surely to the true preference parameter vector* $\boldsymbol{\beta}$.

By Theorem 2, the consistency of the SGMS estimator holds under the same set of assumptions as the GMS estimator, as long as the smooth function is properly chosen. Since any cumulative distribution function (*e.g.*, the standard normal distribution function) satisfies Condition 1, the SGMS does not require more assumptions to achieve strong consistency than the GMS estimator does.

Extra assumptions, however, are required in order to derive the asymptotic distribution of the SGMS estimator. Assume that function $K(\cdot)$ is twice differentiable. Thus the objective function (15) of the SGMS estimator is a smooth function of $\boldsymbol{b}$. Let b_1 denote the first element of $\boldsymbol{b}$, and $\tilde{\boldsymbol{b}}$ denote the vector of its remaining elements. Next, define the first- and second-order derivatives of $Q_N^S(\boldsymbol{b}, h_N)$ with respect to $\tilde{\boldsymbol{b}}$ as $\boldsymbol{t}_N(\boldsymbol{b}, h_N)$ and $\boldsymbol{H}_N(\boldsymbol{b}, h_N)$, respectively, where the vector

$$\boldsymbol{t}_N(\boldsymbol{b}, h_N) = (Nh_N)^{-1}\sum_{n=1}^{N}\sum_{1\le j<k\le J}\left\{[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})] \cdot K'\left(\boldsymbol{x}'_{njk}\boldsymbol{b}/h_N\right)\tilde{\boldsymbol{x}}_{njk}\right\} \tag{16}$$

and the matrix

$$\boldsymbol{H}_N(\boldsymbol{b}, h_N) = (Nh_N^2)^{-1}\sum_{n=1}^{N}\sum_{1\le j<k\le J}\left\{[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})] \cdot K''\left(\boldsymbol{x}'_{njk}\boldsymbol{b}/h_N\right)\tilde{\boldsymbol{x}}_{njk}\tilde{\boldsymbol{x}}'_{njk}\right\}. \tag{17}$$

To derive the first order condition, we make the following assumption:

Assumption 5. *$\tilde{\boldsymbol{\beta}}$ is an interior point of $\tilde{\mathbb{B}}$.*

Let $b_{N,1}^S$ denote the first element of the SGMS estimator $\boldsymbol{b}_N^S \in \mathbb{B}$, and $\tilde{\boldsymbol{b}}_N^S$ denote the vector of the remaining elements. By Theorem 2 and Assumption 5, $b_{N,1}^S = \beta_1$, $\tilde{\boldsymbol{b}}_N^S$ is an interior point of $\tilde{\mathbb{B}}$, and $\boldsymbol{t}_N(\boldsymbol{b}_N^S, h_N) = 0$ with probability approaching one as N approaches ∞. A Taylor series expansion of $\boldsymbol{t}_N(\boldsymbol{b}_N^S, h_N)$ around $\boldsymbol{b}_N^S = \boldsymbol{\beta}$ yields

$$\boldsymbol{t}_N(\boldsymbol{b}_N^S, h_N) = \boldsymbol{t}_N(\boldsymbol{\beta}, h_N) + \boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N)(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}), \tag{18}$$

where $\boldsymbol{b}_N^* \equiv \{b_{N,1}^*, \tilde{\boldsymbol{b}}_N^*\}$, $b_{N,1}^* = b_{N,1}^S = \beta_1$, and $\tilde{\boldsymbol{b}}_N^*$ is a vector between $\tilde{\boldsymbol{b}}_N^S$ and $\tilde{\boldsymbol{\beta}}$. Suppose there is a function $\rho(N)$ such that $\rho(N)\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)$ converges in distribution to a random vector and also that $\boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N)$

converges in probability to a nonsingular, nonstochastic matrix $\boldsymbol{H}$. Then,

$$\rho(N)(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) = -\boldsymbol{H}^{-1}\rho(N)\boldsymbol{t}_N(\boldsymbol{\beta}, h_N) + o_p(1). \tag{19}$$

By (18) and (19), it is essential to derive the limiting distribution of $\rho(N)\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)$ and the probability limit of $\boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N)$ to obtain the asymptotic distribution of the SGMS estimator. Later, we will show that $\rho(N)\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)$ is asymptotically normal if bandwidth h_N is properly chosen according to the smoothness conditions imposed on the distribution of the continuous explanatory variable and error terms. Roughly put, the fastest convergence rate of $\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}$ to zero is $\rho(N)^{-1} \propto N^{-d/(2d+1)}$ when the conditional probability of ranking comparison in (4) is dth ($d \geq 2$) order differentiable with respect to the systematic utility and the conditional density of the continuous explanatory variable is $(d-1)$th order differentiable. Therefore, a higher convergence rate (corresponding to larger d), is achieved at the cost of making stronger smoothness assumptions on the distributions of the continuous explanatory variable and the error terms. By properly choosing the bandwidth $h_N \propto N^{-1/(2d+1)}$ and smooth function $K(\cdot)$ (according to Condition 2 given below), we can conclude that $\rho(N)\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)$ is asymptotically normal. We require the integer d to be no less than 2. If $d = 1$, the random matrix $\boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N)$ does not converge to a non-stochastic matrix $\boldsymbol{H}$, and has an unknown limiting distribution instead; it follows that the limiting distribution of $\rho(N)(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$ is also unknown by (19).

In the binomial choice setting, the SMS estimator is derived from a single latent variable equation, where the conditional choice probability of alternative 1,

$$P(r_1 = 1|\boldsymbol{x}) = P(-\bar{\varepsilon} \leq \boldsymbol{x}'\boldsymbol{\beta}|\boldsymbol{x}), \tag{20}$$

can be expressed as the conditional distribution of the error term $\bar{\varepsilon}$ given a single vector $\boldsymbol{x}$.[16] This conditional distribution function plays an important role in expressing the limiting distribution of the SMS estimator. The SGMS estimator is derived from a model with multiple latent utility equations. Outside the special case of complete rankings, calculating the probability of a ranking comparison, *e.g.*, $P(r_1 < r_2|\boldsymbol{X})$, is even more

[16]Equation (20) uses the common notation adopted in binomial choice analysis. To connect with our notation, $\boldsymbol{x}$ should be interpreted as $\boldsymbol{x}_1 - \boldsymbol{x}_2$ and $\bar{\varepsilon}$ should be interpreted as $\varepsilon_1 - \varepsilon_2$.

complicated than calculating a choice probability. Consider an example where the individual only reports her best and second best alternatives from a set with four alternatives. By the definition of ranking $\boldsymbol{r}$ in (3), we have

$$\begin{aligned} P(r_1 < r_2|\boldsymbol{X}) &= P(u_1 \geq u_2 \geq max\{u_3, u_4\}|\boldsymbol{X}) \\ &+P(u_1 \geq u_3 > max\{u_2, u_4\}|\boldsymbol{X}) + P(u_1 \geq u_4 > max\{u_2, u_3\}|\boldsymbol{X}) \\ &+P(u_3 > u_1 \geq max\{u_2, u_4\}|\boldsymbol{X}) + P(u_4 > u_1 \geq max\{u_2, u_3\}|\boldsymbol{X}). \end{aligned} \tag{21}$$

Calculating $P(r_1 < r_2|\boldsymbol{X})$ by (21) using the joint distribution (or density) function of the error vector $\boldsymbol{\varepsilon}$ is not an easy task. Fortunately, it is not needed for deriving the asymptotic distribution of $\rho(N)\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)$ either. By (16), the convergence rate of $\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)$ to zero depends on the product of the kernel function $K'(\cdot)$ and the pairwise differences between ranking comparisons. For each pair of alternatives (j, k), the difference, $P(r_j < r_k|\boldsymbol{X}) - P(r_k < r_j|\boldsymbol{X})$, is zero if $\boldsymbol{x}'_{jk}\boldsymbol{\beta}$ is zero implied by Assumption 3. The kernel function $K'(\boldsymbol{x}'_{jk}\boldsymbol{\beta}/h_N)$ approaches zero as N goes to ∞ as long as $\boldsymbol{x}'_{jk}\boldsymbol{\beta}$ is non-zero. If the difference, $P(r_j < r_k|\boldsymbol{X}) - P(r_k < r_j|\boldsymbol{X})$, is dth order differentiable with respect to $\boldsymbol{x}'_{jk}\boldsymbol{\beta}$, we choose a dth order kernel $K'(\cdot)$ and an appropriate bandwidth h_N. Analogous to the results on the kernel density estimation, the SGMS estimator's bias is $O(h_N^d)$, variance is $O[(Nh_N)^{-1}]$, and fastest convergence rate is $N^{-d/(2d+1)}$.

To facilitate a formal derivation of the asymptotic distribution of the SGMS estimator, we introduce a series of extra notations first. Recall that $v_j \equiv \boldsymbol{x}'_j\boldsymbol{\beta}$ represents the systematic utility of choosing alternative $j \in \mathbb{J}$. Denote $\boldsymbol{v} \equiv (v_1, \ldots, v_{J-1}, v_J)'$. v_J is zero since $\boldsymbol{x}_J$ is normalized to be a zero vector. There is a one-to-one correspondence between $\boldsymbol{X}$ and $(\boldsymbol{v}, \tilde{\boldsymbol{X}})$ for fixed $\boldsymbol{\beta}$. Define $\boldsymbol{\iota}_J \equiv (1, \ldots, 1) \in \mathbb{R}^J$. For any alternative $j \in \mathbb{J}$, let vector $\boldsymbol{v}_{-j}$ be the difference: $\boldsymbol{v} - \boldsymbol{\iota}_J v_j$. For example, when $1 < j < J$,

$$\boldsymbol{v}_{-j} = (v_1 - v_j, \ldots, v_{j-1} - v_j, 0, v_{j+1} - v_j, \ldots, v_J - v_j)'.$$

In words, $\boldsymbol{v}_{-j}$ is computed by subtracting the systematic utility of alternative j from the raw vector of systematic utilities. For any pair of alternatives $j, k \in \mathbb{J}$, define $v_{-j,k} = v_k - v_j$ and $\tilde{\boldsymbol{v}}_{-j,k}$ as the vector that consists of all elements of $\boldsymbol{v}_{-j}$ excluding $v_{-j,k}$. For example, when $1 < j < k < J$,

$$\tilde{\boldsymbol{v}}_{-j,k} \equiv (v_1 - v_j, \ldots, v_{k-1} - v_j, v_{k+1} - v_j, \ldots, v_J - v_j)'.$$

If $J > 2$, for any three different alternatives $j, k, l \in \mathbb{J}$, define $\tilde{\boldsymbol{v}}_{-j,kl}$ as the vector that consists of all of the elements of $\boldsymbol{v}_{-j}$ excluding $v_{-j,k}$ and $v_{-j,l}$. For example, when $1 < j < k < l < J$,

$$\tilde{\boldsymbol{v}}_{-j,kl} \equiv (v_1 - v_j, \ldots, v_{k-1} - v_j, v_{k+1} - v_j, \ldots, v_{l-1} - v_j, v_{l+1} - v_j, \ldots, v_J - v_j)'.$$

If $J > 3$, for any four different alternatives $j, k, l, m \in \mathbb{J}$, define $\tilde{\boldsymbol{v}}_{-\{k,m\}}$ as the vector that consists of all of the elements of $\boldsymbol{v}$ excluding $\{v_k, v_m\}$. There is a one-to-one correspondence between $\boldsymbol{v}$ and $(v_{jk}, v_{lm}, \tilde{\boldsymbol{v}}_{-\{k,m\}})$.

Let $p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$ denote the conditional density of $v_{-j,k}$ given $(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$. Define the derivatives

$$p^{(i)}_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) = \partial^i p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})/\partial(v_{-j,k})^i$$

and

$$p^{(0)}_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) \equiv p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}).$$

Let $p_{jkl}(v_{-j,k}, v_{-j,l}|\tilde{\boldsymbol{v}}_{-j,kl}, \tilde{\boldsymbol{X}})$ denote the joint density of $(v_{-j,k}, v_{-j,l})$ conditional on $(\tilde{\boldsymbol{v}}_{-j,kl}, \tilde{\boldsymbol{X}})$ and let $p_{jklm}(v_{jk}, v_{lm}|\tilde{\boldsymbol{v}}_{-\{k,m\}}, \tilde{\boldsymbol{X}})$ denote the joint density of (v_{jk}, v_{lm}) conditional on $(\tilde{\boldsymbol{v}}_{-\{k,m\}}, \tilde{\boldsymbol{X}})$.

Given any pair of alternatives $j, k \in \mathbb{J}$, there is a one-to-one correspondence between $\boldsymbol{X}$ and $(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$ for fixed $\boldsymbol{\beta} \in \mathbb{B}$. The probability for each individual to rank alternative j over alternative k depends on her explanatory matrix $\boldsymbol{X}$, or equivalently, $(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$. Define

$$F_{jk}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) = P(r_j < r_k|v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) \tag{22}$$

and

$$\bar{F}_{jk}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) = P(r_j < r_k|v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) - P(r_k < r_j|v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}). \tag{23}$$

Next, for any integer $i > 0$, define the following derivatives:

$$\bar{F}_{jk}^{(i)}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) = \partial^i \bar{F}_{jk}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})/\partial(v_{-j,k})^i,$$

whenever the derivatives exist. Likewise, define the scalar constants k_d and k_Ω by

$$k_d = \int_{-\infty}^{\infty} x^d K'(x) dx$$

and

$$k_\Omega = \int_{-\infty}^{\infty} [K'(x)]^2 dx,$$

whenever these quantities exist. Finally, define the $q-1$ vector $\boldsymbol{a}$, and the $(q-1) \times (q-1)$ matrices $\boldsymbol{\Omega}$ and $\boldsymbol{H}$ as follows:

$$\boldsymbol{a} = \sum_{1 \le j < k \le J} k_d \sum_{i=1}^{d} \frac{1}{i!(d-i)!} E\left[\bar{F}_{jk}^{(i)}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}^{(d-i)}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}\right], \tag{24}$$

$$\boldsymbol{\Omega} = \sum_{1 \le j < k \le J} 2k_\Omega E\left[F_{jk}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk}\right], \tag{25}$$

and

$$\boldsymbol{H} = \sum_{1 \le j < k \le J} E\left[\bar{F}_{jk}^{(1)}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk}\right], \tag{26}$$

whenever these quantities exist.

Now, we turn to the derivation of the asymptotic distribution of the SGMS estimator $\boldsymbol{b}_N^S$. We start off by making the following requirements on the smooth function $K(\cdot)$, in addition to Condition 1.[17]

Condition 2. The following statements are true.

(a) $K(v)$ is twice differentiable for $v \in \mathbb{R}$, $|K'(v)|$ and $|K''(v)|$ are uniformly bounded, and the integrals

[17]These extra requirements, stated in Condition 2, on the smooth function $K(\cdot)$ are similar to those in Assumption 7 of Horowitz (1992).

$\int_{-\infty}^{\infty}[K'(v)]^2dv$, $\int_{-\infty}^{\infty}[K'(v)]^4dv$, $\int_{-\infty}^{\infty}v^2|K''(v)|dv$, and $\int_{-\infty}^{\infty}[K''(v)]^2dv$ are finite.

(b) For some integer $d \geq 2$, $\int_{-\infty}^{\infty}|v^d K'(v)|dv < \infty$ and $k_d \in (0,\infty)$. For any integer i $(1 \leq i < d)$, integrals $\int_{-\infty}^{\infty}|v^i K'(v)|dv < \infty$ and $\int_{-\infty}^{\infty}v^i K'(v)dv = 0$.

(c) For any integer i $(0 \leq i \leq d)$, any $\eta > 0$, and any positive sequence $\{h_N\}$ converging to 0,

$$\underset{N\to\infty}{lim}\, h_N^{i-d}\int_{|h_N v|>\eta}|v^i K'(v)|dv = 0$$

and

$$\underset{N\to\infty}{lim}\, h_N^{-1}\int_{|h_N v|>\eta}|K''(v)|dv = 0.$$

Next, we state extra assumptions needed for deriving the asymptotic distribution of the SGMS estimator, with brief comments on the implications of each assumption.

Assumption 6. *For any pair of alternatives $j<k$ and for $v_{-j,k}$ in a neighborhood of 0, $\bar{F}_{jk}^{(i)}(v_{-j,k},\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})$ exists and is a continuous function of $v_{-j,k}$. Function $|\bar{F}_{jk}^{(i)}(v_{-j,k},\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})|$ is bounded by a constant C for almost every $(\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})$, where C is a finite real number and i is an integer $(1 \leq i \leq d)$.*

By definition (23), function $\bar{F}_{jk}(\cdot)$ can be derived from the conditional distribution of the error terms. Assumption 6 in essence imposes the differentiability requirement on the conditional distribution function of the error vector $\boldsymbol{\varepsilon}$ with respect to systematic utilities.

Assumption 7. *The following statements on the explanatory variables are true.*

(a) *For any pair of different alternatives $j,k \in \mathbb{J}$, $p_{jk}^{(i)}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})$ exists and is a continuous function of $v_{-j,k}$ satisfying $|p_{jk}^{(i)}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})| < C$ for $v_{-j,k}$ in a neighborhood of 0, almost every $(\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})$, some finite constant C, and any integer i $(1 \leq i \leq d-1)$. In addition, $|p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})| < C$ for all $v_{-j,k}$* and almost every $(\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})$.

(b) *For any three different alternatives $j,k,l \in \mathbb{J}$, $p_{jkl}(v_{-j,k},v_{-j,l}|\tilde{\boldsymbol{v}}_{-j,kl},\tilde{\boldsymbol{X}}) < C$ for all $(v_{-j,k},v_{-j,l})$, almost every $(\tilde{\boldsymbol{v}}_{-j,kl},\tilde{\boldsymbol{X}})$, and some finite constant C.*

(c) *For any four different alternatives* $j, k, l, m \in \mathbb{J}$, $p_{jklm}(v_{jk}, v_{lm}|\tilde{\boldsymbol{v}}_{-\{k,m\}}, \tilde{\boldsymbol{X}}) < C$ *for all* (v_{jk}, v_{lm}), *almost every* $(\tilde{\boldsymbol{v}}_{-\{k,m\}}, \tilde{\boldsymbol{X}})$, *and some finite constant* C.

(d) *The components of matrices* $\tilde{\boldsymbol{X}}$, $vec(\tilde{\boldsymbol{X}})vec(\tilde{\boldsymbol{X}})'$, *and* $vec(\tilde{\boldsymbol{X}})vec(\tilde{\boldsymbol{X}})'vec(\tilde{\boldsymbol{X}})vec(\tilde{\boldsymbol{X}})'$ *have finite first absolute moments.*

Assumption 7 imposes regularity conditions on the explanatory variables. In addition to the continuity requirement imposed by Assumption 4, Assumption 7 further requires that the conditional probability density function of the first explanatory variable, $x_{jk,1}$, given other explanatory variables is $(d-1)th$ order differentiable.

Assumption 8. $(logN)/(Nh_N^4) \to 0$ *as* $N \to \infty$, *where* $\{h_N\}$ *is a strictly positive sequence converging to* 0.

Assumptions 6-8, together with Condition 2, are analogous to typical assumptions made in the kernel density estimation. A higher convergence rate of the SGMS estimator can be achieved using a higher order kernel $K'(\cdot)$ when the required derivatives of $\bar{F}(\cdot)$ and $p(\cdot)$ exist.

Assumption 9. *The matrix* $\boldsymbol{H}$, *defined by* (26), *is negative definite.*

Note that the matrix $\boldsymbol{H}$ is analogous to the Hessian information matrix in the quasi-MLE. The following theorem presents the main results concerning the asymptotic distribution of the SGMS estimator.

Theorem 3. *Let Assumptions* 1-9 *and Conditions* 1-2 *hold for some integer* $d \geq 2$ *and let* $\{\boldsymbol{b}_N^S\}$ *be a sequence of solutions to problem* (14).

(a) *If* $Nh_N^{2d+1} \to \infty$ *as* $N \to \infty$, *then* $h_N^{-d}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$ *converges in probability to* $-\boldsymbol{H}^{-1}\boldsymbol{a}$.

(b) *If* Nh_N^{2d+1} *has a finite limit* λ *as* $N \to \infty$, *then* $(Nh_N)^{1/2}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$ *converges in distribution to*

$$MVN\left(-\lambda^{1/2}\boldsymbol{H}^{-1}\boldsymbol{a},\, \boldsymbol{H}^{-1}\boldsymbol{\Omega}\boldsymbol{H}^{-1}\right).$$

(c) *Define* $h_N = (\lambda/N)^{1/(2d+1)}$, *where* $\lambda \in (0, \infty)$. *Let* $\boldsymbol{W}$ *be any nonstochastic, positive semidefinite matrix such that* $\boldsymbol{a}'\boldsymbol{H}^{-1}\boldsymbol{W}\boldsymbol{H}^{-1}\boldsymbol{a} \neq 0$. *Denote* E_A *as the expectation with respect to the asymptotic distribution* *of* $N^{d/(2d+1)}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$ *and define* MSE *as* $E_A[(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})'\boldsymbol{W}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})]$. *The* MSE *is then minimized by setting* λ *to be*

$$\lambda^* \equiv \left[trace\left(\boldsymbol{\Omega H}^{-1}\boldsymbol{W H}^{-1}\right)\right] / \left(2d\boldsymbol{a}'\boldsymbol{H}^{-1}\boldsymbol{W H}^{-1}\boldsymbol{a}\right), \tag{27}$$

in that case $N^{d/(2d+1)}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$ *converges in distribution to*

$$MVN\left(-(\lambda^*)^{d/(2d+1)}\boldsymbol{H}^{-1}\boldsymbol{a},\ (\lambda^*)^{-1/(2d+1)}\boldsymbol{H}^{-1}\boldsymbol{\Omega H}^{-1}\right).$$

By Theorem 3, if $Nh_N^{2d+1} \to \infty$ as $N \to \infty$, then $h_N^{-d}/N^{d/(2d+1)} = (Nh_N^{2d+1})^{-d/(2d+1)} \to 0$; if $Nh_N^{2d+1} \to 0$ as $N \to \infty$, then $(Nh_N)^{1/2}/N^{d/(2d+1)} = (Nh_N^{2d+1})^{1/(4d+2)} \to 0$. Therefore, Theorem 3 implies that the fastest rate of convergence of the SGMS estimator is $N^{-d/(2d+1)}$. Choosing bandwidth $h_N = (\lambda/N)^{1/(2d+1)}$ where $\lambda \in (0, \infty)$ can achieve the fastest rate of convergence. Theorem 3(c) shows that λ^*, defined by (27), minimizes the MSE of the SGMS estimator.

To make the results of Theorem 3 useful in applications, it is necessary to be able to estimate the parameters in the limiting distribution, $\boldsymbol{a}$, $\boldsymbol{\Omega}$, and $\boldsymbol{H}$, consistently from observations of $(\boldsymbol{r}, \boldsymbol{X})$. The next theorem shows how this can be done.

Theorem 4. *Let Assumptions* 1-9 *and Conditions* 1-2 *hold for some integer* $d \geq 2$ *and vector* $\boldsymbol{b}_N^S$ *be a consistent estimator based on* $h_N \propto N^{-1/(2d+1)}$. *Let* $h_N^* \propto N^{-\delta/(2d+1)}$, *where* $\delta \in (0,1)$. *Then*

(a) $\hat{\boldsymbol{a}}_N \equiv (h_N^*)^{-d}\boldsymbol{t}_N(\boldsymbol{b}_N^S, h_N^*)$ *converges in probability to* $\boldsymbol{a}$;

(b) *for* $\boldsymbol{b} \in \mathbb{B}$ *and* $n = 1, \ldots, N$, *define*

$$\boldsymbol{t}_{Nn}(\boldsymbol{b}, h_N) = \sum_{1 \leq j < k \leq J} \left[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})\right] K'\left(\boldsymbol{x}_{njk}'\boldsymbol{b}/h_N\right) \tilde{\boldsymbol{x}}_{njk} h_N^{-1},$$

the matrix

$$\hat{\boldsymbol{\Omega}}_N \equiv (h_N/N)\sum_{n=1}^{N}\boldsymbol{t}_{Nn}(\boldsymbol{b}_N^S, h_N)\boldsymbol{t}_{Nn}(\boldsymbol{b}_N^S, h_N)'$$

converges in probability to $\boldsymbol{\Omega}$;

(c) *the matrix* $\boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)$ *converges in probability to* $\boldsymbol{H}$.

By Theorem 3(c), the asymptotic bias of $N^{d/(2d+1)}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$ is $-\lambda^{d/(2d+1)}\boldsymbol{H}^{-1}\boldsymbol{a}$ when the bandwidth $h_N = (\lambda/N)^{1/(2d+1)}$. It follows from Theorem 4 that the bias term $-\lambda^{d/(2d+1)}\boldsymbol{H}^{-1}\boldsymbol{a}$ can be estimated consistently by $-\lambda^{d/(2d+1)}\boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)^{-1}\hat{\boldsymbol{a}}_N$. Therefore, define

$$\tilde{\boldsymbol{b}}_N^u = \tilde{\boldsymbol{b}}_N^S + (\lambda/N)^{d/(2d+1)}\boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)^{-1}\hat{\boldsymbol{a}}_N \tag{28}$$

as the bias-corrected SGMS estimator.

3.2 A Small-Sample Correction

In this subsection, we apply a method proposed by Horowitz (1992) to remove part of the finite sample bias of $\hat{\boldsymbol{a}}_N$. By Theorem 2, $\boldsymbol{b}_{N,1}^S = \beta_1$ with probability approaching one as N goes to ∞. A Taylor expansion of $\hat{\boldsymbol{a}}_N$ around $\tilde{\boldsymbol{\beta}}$ yields

$$\hat{\boldsymbol{a}}_N - \boldsymbol{a} = \left[(h_N^*)^{-d}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N^*) - \boldsymbol{a}\right] + (h_N^*)^{-d}\boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N^*)(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) \tag{29}$$

with probability approaching one as N goes to ∞, where $\boldsymbol{b}_N^*$ is a vector between $\boldsymbol{b}_N^S$ and $\boldsymbol{\beta}$. The right-hand side of (29) shows that the finite sample bias of $\hat{\boldsymbol{a}}_N$ has two components. The first component, $(h_N^*)^{-d}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N^*) - \boldsymbol{a}$, has a non-zero mean due to the use of a non-zero bandwidth h_N^* to estimate $\boldsymbol{a}$. The second component, $(h_N^*)^{-d}\boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N^*)(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$, has a non-zero mean due to the use of an estimate of the true parameter vector $\boldsymbol{\beta}$ in estimating $\boldsymbol{a}$.

The bias correction method described here is aimed at removing the second component of bias by order $N^{-(1-\delta)d/(2d+1)}$. Note that the second component of the right-hand side of (29) can be written as

$$(h_N^*)^{-d}\boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N^*)(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) = \left[Nh_N(h_N^*)^{2d}\right]^{-1/2}\boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N^*)(Nh_N)^{1/2}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}).$$

The probability limit of $\boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N^*)$ is $\boldsymbol{H}$ by Lemmas 8-9 in Appendix B, and $(Nh_N)^{1/2}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$ converges in distribution to $MVN(-\lambda^{1/2}\boldsymbol{H}^{-1}\boldsymbol{a},\, \boldsymbol{H}^{-1}\boldsymbol{\Omega}\boldsymbol{H}^{-1})$ by Theorem 3. Therefore,

$$\left[Nh_N(h_N^*)^{2d}\right]^{1/2}(h_N^*)^{-d}\boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N^*)(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$$

converges in distribution to $MVN(-\lambda^{1/2}\boldsymbol{a},\, \boldsymbol{\Omega})$. By this result, we treat $\hat{\boldsymbol{a}}_N$ as an estimator of

$$\boldsymbol{a} - \left[Nh_N(h_N^*)^{2d}\right]^{-1/2}\lambda^{1/2}\boldsymbol{a}$$

rather than that of $\boldsymbol{a}$. Thus, the bias corrected estimator of $\boldsymbol{a}$ is

$$\hat{\boldsymbol{a}}_N^c = \hat{\boldsymbol{a}}_N / \left\{1 - \left[\lambda^{-1}Nh_N(h_N^*)^{2d}\right]^{-1/2}\right\}. \tag{30}$$

3.3 Bandwidth Selection

Theorem 3(c) provides a way to choose the bandwidth for the SGMS estimator. To achieve the minimum MSE, an optimal λ^* can be consistently estimated by the conclusion of Theorem 4. Therefore, one possible way of choosing bandwidth is to set $h_N = (\hat{\lambda}/N)^{1/(2d+1)}$ given the integer d, where $\hat{\lambda}$ is a consistent estimator for λ^*.

Specifically, the choice of bandwidth can be implemented by taking the following steps.

Step 1. Given d, choose a $h_N \propto N^{-1/(2d+1)}$ and $h_N^* \propto N^{-\delta/(2d+1)}$ for $\delta \in (0,1)$.

Step 2. Compute the SGMS estimator $\boldsymbol{b}_N^S$ using h_N. Use $\boldsymbol{b}_N^S$ and h_N^* to compute $\hat{\boldsymbol{a}}_N^c$. Use $\boldsymbol{b}_N^S$ and h_N to compute $\hat{\boldsymbol{\Omega}}_N$ and $\boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)$.

Step 3. Estimate λ^* by

$$\begin{aligned}\hat{\lambda}_N \quad = \quad & \left\{trace\left[\hat{\boldsymbol{\Omega}}_N\boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)^{-1}\boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)^{-1}\right]\right\} \\ & \cdot\left[2d(\hat{\boldsymbol{a}}_N^c)'\boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)^{-1}\boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)^{-1}\hat{\boldsymbol{a}}_N^c\right]^{-1}.\end{aligned} \tag{31}$$

Step 4. Calculate the estimated bandwidth $h_N^e = (\hat{\lambda}_N/N)^{1/(2d+1)}$.

Step 5. Compute the SGMS estimator using h_N^e.

Note that this approach is analogous to the plug-in method of kernel density estimation. As usual in the application of the plug-in method, the choice of the initial bandwidth h_N and parameter δ would require some exploration, because the estimated bandwidth h_N^e may be sensitive to that choice. In our Monte Carlo experiments in the next section, the bandwidth has been initialized by setting $h_N = N^{-1/5}$ and $\delta = 0.1$.

4 Monte Carlo Experiments

In this section, we provide Monte Carlo simulation results to explore finite-sample properties of the GMS estimator $\boldsymbol{b}_N$ and the SGMS estimator $\boldsymbol{b}_N^S$. We consider six data generating processes (DGPs). In each DGP, individual n's utility from alternative j, u_{nj}, is specified as

$$u_{nj} = x_{nj,1}\beta_1 + x_{nj,2}\beta_{n2} + \varepsilon_{nj} \text{ for } n = 1, 2, ..., N \text{ and } j = 1, 2, ..., 5. \tag{32}$$

Each DGP is used to simulate two sets of 1000 random samples of N individuals, where $N = 100$ in the first set and 500 in the second set.

In all DGPs, the first preference parameter β_1 is a deterministic coefficient and takes the value of one for all individuals: $\beta_1 = 1$. In DGPs 1-4, the second preference parameter β_{n2} is also a deterministic coefficient and takes the value of one for all individuals: $\beta_{n2} = \beta_2 = 1$ for all n. In DGPs 5-6, however, β_{n2} is a random coefficient that varies across individuals, and each individual's coefficient value is a random draw from distribution $N(1,1)$: $\beta_{n2} = \beta_2 + \eta_n$, where $\beta_2 = 1$ and η_n is distributed as $N(0,1)$.[18] Each DGP specifies its own distribution of error terms ε_{nj}: we provide more details below.[19]

The econometrician observes a utility-based ranking $\boldsymbol{r}_n$ of $J = 5$ alternatives in $\mathbb{J}$, as well as attributes $x_{nj,1}$ and $x_{nj,2}$ for $j = 1, 2, ..., 5$ and all n.[20] As usual, the depth of observed rankings would influence the finite sample precision of an estimator; and in the context of our semparametric estimators, it also influences the degree of flexibility that semiparametric models offer. Recall that when the complete rankings

[18] In random coefficients models, we are often interested in discovering a certain central tendency of the random preference parameter, such as its mean or its median. The mixed logit estimator will consistently estimate $E(\beta_{n2})$ under correct parametric specifications and the proposed semiparametric estimators can consistently estimate $median(\beta_{n2})$ under Assumptions 1-4. For the simplicity of demonstration, we choose $\beta_{n2} \sim N(1,1)$ such that $E(\beta_{n2}) = median(\beta_{n2}) = 1$.

[19] In all DGPs, we generate ε_{nj} with variance equal to $\pi^2/6$, subject to rounding errors.

[20] Here we use a relative small choice set mainly because the probit and the mixed logit specifications yield objective functions that require multivariate integration, and consequently a lot of computation time. The computation time of the GMS and SGMS estimators *per se* is affordable even if the choice set is very large.

($M = J - 1 = 4$) are observed, the semiparametric model nests all popular parametric models as special cases; when only partial rankings ($M < 4$) are available, this is not the case because the semiparametric model cannot accommodate alternative-specific heteroekedasticity and flexible correlation patterns. We will therefore explore the finite sample behavior of the estimators at three depth levels: $M = 1$ when only the best alternative is observed, $M = 2$ when the best and second alternatives are observed, and $M = 4$ when the complete ranking is observed. In all DGPs, observed attribute $x_{nj,1}$ is a random draw from $N(0, 2)$ and $x_{nj,2}$ is generated as a ratio of two different uniform draws: specifically, $x_{nj,2} \equiv q_{nj}/z_n$ where q_{nj} is drawn from $U(0, 3)$ and z_n is drawn from $U(\frac{1}{5}, 5)$.[21] Note that $x_{nj,1}$ and q_{nj} vary across both individuals and alternatives, whereas z_n varies only across individuals. All three distributions that generate the observed attributes are independent of one another, and *i.i.d.* across the subscripted dimension(s).

For comparison with our GMS and SGMS estimates, we also compute maximum likelihood estimates using three popular parametric models summarized in Section 2.3, namely rank-ordered logit (ROL), rank-ordered probit (ROP), and mixed ROL (MROL). We do not estimate the nested ROL model, primarily because our analysis already includes the ROP model which is a more flexible parametric method to incorporate correlated errors. In case of ROP and MROL, we opt to place no constraint on the variance-covariance parameters of the underlying multivariate normal densities.[22] This allows us to compare our semiparametric methods with both restrictive (ROL) and very flexible (ROP and MROL) parametric methods.

Our discussion focuses on the ratio of the preference parameters, β_2/β_1, which is identified in both parametric and semiparametric models. In the discrete choice analysis of individual preferences, the main parameter of interest often takes the form of a ratio between coefficients on non-price and price attributes; this type of ratio is known as, *inter alia*, equivalent prices (Hausman and Ruud, 1987), implicit prices (Calfee *et al.*, 2001), and willingness-to-pay (Small *et al.*, 2005). In parametric models, we normalize the scale of the error terms in the usual manner and estimate (β_1, β_2), then we derive the ratio of the relevant coefficient estimates. In semiparametric models, we normalize $|\beta_1| = 1$ and estimate β_2 together with the sign of β_1,

[21] This pair of uniform distributions ensures that the second observed attribute has approximately the same variance as the first attribute, *i.e.*, $Var(q_{nj}/z_n) \simeq 2$.

[22] Our ROP specification requires estimating two slope coefficients (β_1 and β_2) and eight identified variance-covariance parameters of pairwise error differences. Our MROL specification assumes that both slope coefficients are random and bivariate normal: we estimate two mean (β_1 and β_2) and three variance-covariance parameters of the bivariate normal density. The ROP (MROL) model has been estimated in Stata using command -asroprobit- (-mixlogit-); the likelihood function has been simulated by taking 250 pseudo-random draws from Hammersley (Halton) sequences.

then we compute the estimate of the ratio of interest $\beta_2/\beta_1 = \beta_2/sign(\beta_1)$.[23] Since the GMS estimator entails maximizing a sum of step functions, we use a global search method to compute the GMS estimates: specifically the differential evolution algorithm of Storn and Price (1995), which was also Fox's (2007) preferred method for computing his multinomial MS estimates. In this Monte Carlo study, we implement a particular version of the SGMS estimator which uses the standard normal distribution function as the smooth functionl $K(\cdot)$. The resulting objective function is differentiable, and can be maximized by starting any of usual gradient-based algorithms from a set of initial search points. For the SGMS estimator, the bandwidth has been initalized by setting $h_N = N^{-1/5}$ and $\delta = 0.1$, and optimized subsequently by applying the method in Section 3.3.[24]

Table 1 summarizes the true distribution of the error terms in each DGP and whether particular methods can estimate β_2/β_1 consistently. The summary presents a strong case for the importance of considering semiparametric methods for rank-ordered choice data: the GMS/SGMS estimator using complete rankings is the only method that remains consistent throughout all DGPs. The GMS/SGMS estimator using partial rankings is consistent when the error terms are *i.i.d.* (DGP 1-2) or heteroskedastic across individuals (DGP 3), but becomes inconsistent in the presence of alternative-specific heteroskedasticity (DGP 4) and/or random coefficients (DGP 5-6). As usual, a parametric method is consistent only when the DGP happens to coincide with the postulated parametric model itself or its special cases.

Tables 2-7 report each method's bias and RMSE across 1,000 samples of size N simulated from each DGP. In each table, the top and bottom panels summarize the results for sample sizes $N = 100$ and $N = 500$ respectively. Efficiency gains from the use of deeper rankings, alongside the usual play of asymptotics, are apparent from the tables. When a method is consistent for a particular DGP, increasing the depth of rankings M holding the sample size N fixed reduces its bias and RMSE. Increasing the sample size holding the depth of rankings fixed also has the same effects qualitatively.

The GMS estimator using complete rankings is consistent under all DGPs, and displays negligible finite sample bias in most cases. The associated bias is approximately 6% of the true parameter value in DGPs 1 and 2 when $N = 100$, and 2% or less in all other DGPs and/or sample size configurations. These results

[23]The estimator of the sign will converge at a much faster rate than the estimator for β_2 such that there is no need to analyze the finite-sample property of the sign estimator.

[24]When the sample size is small, the parameter λ^* may be estimated with a large standard error due to the slow convergence rate of the bias estimator $\hat{\boldsymbol{a}}_N$, sometimes resulting in a very large estimate of the bandwidth. We apply a trimming procedure to avoid this situation. The estimated λ^* is trimmed at a large constant (1000) for all DGPs.

Table 1: Consistency of estimators by Monte Carlo DGPs

DGP	Distribution of ε_{nj}	ROL	ROP	MROL	(S)GMS
1	ε_{nj} is *i.i.d.* $EV(0,1,0)$	Yes	No	Yes	Yes
2	ε_{nj} is *i.i.d.* $N(0.577, \pi^2/6)$	No	Yes	No	Yes
3	$\varepsilon_{nj} = 0.0055(z_n^4 + 2z_n^2)\epsilon_{nj}$ where ϵ_{nj} is *i.i.d.* $N(0,1)$	No	No	No	Yes
4	$\varepsilon_{nj} = 0.75x_{nj,2}\epsilon_{nj}$ where ϵ_{nj} is *i.i.d.* $N(0,1)$	No	No	No	No when $M < 4$; Yes when $M = 4$
5	ε_{nj} is *i.i.d.* $EV(0,1,0)$	No	No	Yes	No when $M < 4$; Yes when $M = 4$
6	$\varepsilon_{nj} = 0.75x_{nj,2}\epsilon_{nj}$ where ϵ_{nj} is *i.i.d.* $N(0,1)$	No	No	No	No when $M < 4$; Yes when $M = 4$

Note: $EV(0,1,0)$ stands for the extreme value type 1 distribution, assumed by the ROL model, with a mean of 0.577 and a variance of $\pi^2/6$. Where relevant, the error component is *i.i.d.* for $n = 1, \ldots, N$ and $j = 1, \ldots, J$. $M = 4$ ($M < 4$) refers to an estimator that incorporates the complete (partial) rankings. Yes (No) means the estimator of β_2/β_1 is (not) consistent given the DGP.

Table 2: Monte Carlo results on $\hat{\beta}_2/\hat{\beta}_1$ of DGP 1

		ROL		ROP		MROL		GMS		SGMS	
N	M	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
100	1	0.0300	0.2698	0.0490	0.2972	0.0654	0.3100	0.1453	0.5777	0.1403	0.4759
	2	0.0119	0.1883	0.0142	0.1966	0.0297	0.2001	0.0843	0.4077	0.0927	0.3122
	4	-0.0016	0.1382	-0.0075	0.1466	0.0100	0.1450	0.0653	0.3355	0.0632	0.2422
500	1	0.0034	0.1124	0.0091	0.1170	0.0151	0.1206	0.0363	0.2858	0.0528	0.2029
	2	-0.0007	0.0805	-0.0026	0.0834	0.0076	0.0841	0.0200	0.2157	0.0338	0.1439
	4	-0.0006	0.0601	-0.0070	0.0630	0.0047	0.0626	0.0045	0.1739	0.0224	0.1044

Table 3: Monte Carlo results on $\hat{\beta}_2/\hat{\beta}_1$ of DGP 2

		ROL		ROP		MROL		GMS		SGMS	
N	M	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
100	1	0.0243	0.2491	0.0379	0.2781	0.0543	0.2801	0.1301	0.5560	0.1280	0.4260
	2	0.0113	0.1817	0.0125	0.1845	0.0284	0.1942	0.1106	0.4572	0.1002	0.3434
	4	0.0154	0.1488	0.0058	0.1429	0.0278	0.1592	0.0597	0.3781	0.0749	0.2805
500	1	0.0052	0.1079	0.0097	0.1086	0.0201	0.1147	0.0363	0.2756	0.0463	0.1823
	2	0.0081	0.0811	0.0056	0.0788	0.0164	0.0847	0.0315	0.2262	0.0383	0.1430
	4	0.0134	0.0679	0.0038	0.0629	0.0185	0.0719	0.0191	0.2072	0.0305	0.1205

Table 4: Monte Carlo results on $\hat{\beta}_2/\hat{\beta}_1$ of DGP 3

		ROL		ROP		MROL		GMS		SGMS	
N	M	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
100	1	0.1760	0.2517	0.1892	0.2848	0.1039	0.1810	0.0307	0.1873	0.0532	0.1446
	2	0.1609	0.1980	0.1806	0.2224	0.0844	0.1222	0.0055	0.0940	0.0342	0.0864
	4	0.1661	0.1904	0.1928	0.2179	0.0770	0.1013	0.0029	0.0561	0.0329	0.0644
500	1	0.1569	0.1752	0.2009	0.2205	0.0830	0.1033	0.0021	0.0603	0.0266	0.0590
	2	0.1570	0.1655	0.2038	0.2122	0.0780	0.0873	0.0005	0.0309	0.0214	0.0381
	4	0.1593	0.1645	0.2032	0.2082	0.0697	0.0754	-0.0002	0.0193	0.0196	0.0294

Table 5: Monte Carlo results on $\hat{\beta}_2/\hat{\beta}_1$ of DGP 4

		ROL		ROP		MROL		GMS		SGMS	
N	M	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
100	1	-0.1276	0.2794	-0.0674	0.2881	0.0073	0.2721	0.3087	0.5129	0.3674	0.5121
	2	-0.3021	0.3529	-0.2494	0.3187	-0.1520	0.2457	0.1593	0.3600	0.2065	0.3252
	4	-0.4880	0.5123	-0.4191	0.4485	-0.2963	0.3324	-0.0063	0.2591	0.0457	0.2099
500	1	-0.1502	0.1913	-0.1541	0.2037	-0.0008	0.1181	0.2872	0.3687	0.3221	0.3559
	2	-0.3021	0.3145	-0.3055	0.3182	-0.1401	0.1634	0.1500	0.2356	0.1785	0.2093
	4	-0.4945	0.4998	-0.4658	0.4711	-0.2828	0.2906	-0.0032	0.1537	0.0277	0.0904

Table 6: Monte Carlo results on $\hat{\beta}_2/\hat{\beta}_1$ of DGP 5

		ROL		ROP		MROL		GMS		SGMS	
N	M	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
100	1	-0.2618	0.4159	-0.2739	0.4408	-0.0157	0.3681	0.0196	0.5917	0.0390	0.4891
	2	-0.2553	0.3728	-0.2763	0.3900	-0.0105	0.2877	0.0093	0.4857	0.0469	0.3968
	4	-0.2531	0.3538	-0.2763	0.3713	-0.0037	0.2466	0.0161	0.4255	0.0633	0.3398
500	1	-0.2970	0.3268	-0.3488	0.3766	-0.0117	0.1614	-0.0442	0.3193	-0.0220	0.2348
	2	-0.2752	0.2983	-0.3329	0.3532	0.0124	0.1277	-0.0020	0.2670	0.0193	0.1823
	4	-0.2692	0.2905	-0.3259	0.3438	0.0194	0.1132	0.0141	0.2280	0.0412	0.1660

Table 7: Monte Carlo results on $\hat{\beta}_2/\hat{\beta}_1$ of DGP 6

		ROL		ROP		MROL		GMS		SGMS	
N	M	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
100	1	-0.2859	0.4019	-0.2519	0.4001	-0.0375	0.3176	0.2058	0.5294	0.2816	0.5007
	2	-0.4141	0.4707	-0.3838	0.4493	-0.1543	0.2891	0.0988	0.4181	0.1716	0.3763
	4	-0.5540	0.5848	-0.5042	0.5371	-0.2750	0.3478	0.0012	0.3607	0.0622	0.2960
500	1	-0.3179	0.3411	-0.3488	0.3766	-0.0451	0.1442	0.1926	0.3225	0.2355	0.3008
	2	-0.4322	0.4434	-0.3329	0.2892	-0.1575	0.1907	0.1058	0.2370	0.1368	0.2012
	4	-0.5700	0.5764	-0.3259	0.3478	-0.2768	0.2888	0.0006	0.1977	0.0358	0.1356

illustrate a considerable benefit that the use of deeper rankings offers for semiparametric estimation: the partial rankings GMS estimator is consistent for only first three DGPs (DGP 1-3), and even under those DGPs, the estimator exhibits larger bias which sometimes exceeds 10% of the true value when the sample size $N = 100$ (though bias stays below 4% when $N = 500$). Across all depth levels and sample size configurations, the SGMS estimator behaves similarly as its GMS counterpart but tends to display a small increase in bias and a reduction in RMSE, the expected trade-offs from using a smoothing kernel to construct a surrogate objective function. For DGPs 1, 2 and 4, at least one parametric method allows consistent maximum likelihood estimation. The results suggest that the efficiency gains (as measured by the reduction in RMSE) that a consistent SGMS estimator offers over a consistent GMS estimator are comparable to what a consistent parametric estimator offers over the SGMS estimator itself.

The results pertaining to DGPs 3, 4 and 6 present a particularly strong case for the considering the use of the semiparametric methods in empirical practice. While none of the popular parametric methods is consistent under these DGPs, at least one parametric method arguably comes close to getting each DGP approximately right; yet, even in the larger sample configuration ($N = 500$), an approximately correct parametric method may still exhibit a substantial amount of bias. In the context of DGP 3, for instance, the ROP model is a correct specification apart from its failure to capture interpersonal heteroskedasticity; yet, the ROP method's bias stays around 20% of the true parameter value. In DGP 4 and DGP 6, there is alternative-specific heteroskedasticity induced via a normal error component which multiplies the second attribute $x_{nj,2}$; this error component can be absorbed into the normal random coefficient on $x_{nj,2}$, and the MROL model is therefore a correct specification apart from that it postulates the presence of a redundant extreme value error component. While the MROL method's bias is indeed negligible when only information on the most preferred alternative is used ($M = 1$), it becomes amplified as deeper ranking information is used and may reach 28% with complete rankings ($M = 4$).

While our experiments were designed to illustrate the properties of the semiparametric methods, the results also add some cautionary notes to the debate over the reliability of rank-ordered choice data. Based on the intuitively convincing premise that ranking is a more cognitively demanding task than making a choice, some researchers contend that in case a parametric method using the first preference ($M = 1$) and complete rankings ($M = J - 1$) yield systematically different estimates, the econometrician should not make use of complete rankings: see Chapman and Staelin (1982) and Ben-Akiva *et al.* (1992) for the influential and earliest proponents of this view. The results pertaining to DGPs 3-6, however, caution against basing data and model selection on the comparisons of the first preference estimates and the complete rankings estimates. Inconsistent parametric methods may or may not be equally biased at all depth levels, and it is not always the case that the first preference estimates are subject to smaller misspecification bias than the complete rankings estimates.

5 Conclusions

To collect more preference information from a given sample of individuals, multinomial choice surveys can be readily modified to elicit rank-ordered choices. All parametric methods for multinomial choices have their rank-ordered choice counterparts that exploit the extra information to estimate the underlying random utility model more efficiently. But semiparametric methods for rank-ordered choices remain undeveloped, apart from the seminal work of Hausman and Ruud (1987) that is only applicable to continuous regressors. We develop two semiparametric methods for rank-ordered choices: the generalized maximum score (GMS) estimator and the smoothed generalized maximum score (SGMS) estimator. The GMS estimator builds on the maximum score (MS) estimator (Manski, 1975; Fox, 2007) for multinomial choices. Like its predecessor, the GMS estimator allows consistent estimation of coefficients on both continuous and discrete regressors when there is a suitable continuous regressor for normalizing the scale of utility. We establish conditions for strong consistency of the GMS estimator, which follows a non-standard asymptotic distribution and displays a slow convergence rate. The SGMS estimator complements the GMS estimator, much as Horowitz's (1992) smoothed MS estimator complements Manski's (1985) MS estimator in the context of binomial choices. By adding mild regularity conditions, we show that the SGMS estimator is also strongly consistent, and that it is asymptotically normal with a convergence rate approaching $N^{-1/2}$ as the strength of the smoothness conditions increases. Our results are fairly general and cover data on complete rankings as well as partial rankings.

Our study finds that rank-ordered choices provide an interesting data environment which can facilitate and benefit from the development of semiparametric methods. Most interestingly, our results show that using the extra information from rank-ordered choices is not just a matter of efficiency gains, to the contrary of what parametric analyses might lead one to anticipate. For our semiparametric estimators, it is also a matter of consistency in the sense that using complete rankings instead of partial rankings allows the semiparametric estimators to become robust to wider classes of stochastic specifications. More specifically, the MS estimator for multinomial choices and the GMS/SGMS estimators for partial rankings are robust to any form of interpersonal heteroskedasticity. But they are not robust to any error variance-covariance structure that varies across alternatives, meaning that they cannot consistently estimate flexible parametric models including nested logit, unrestricted probit and random coefficients logit. By contrast, the GMS/SGMS estimators for complete rankings (*i.e.*, fully rank-ordered choices) can accommodate error structures as such, fulfilling the usual expectations for a semiparametric method to be more flexible than popular parametric methods. The main intuition behind this contrast is that the use of complete rankings allows one to infer which alternative is more preferred in every possible pair of alternatives in a choice set. The strong consistency of the GMS/SGMS estimators for fully rank-ordered choices can therefore be shown under almost the same assumptions as that of the MS estimator for binomial choices, without invoking stronger assumptions needed to address more analytically complex cases of multinomial choices or partial rank-ordered choices.

Together with our Monte Carlo evidence on the bias of parametric methods under misspecification, this finding calls for a reconsideration of the conventional wisdom prevailing in the empirical literature. Since Chapman and Staelin (1982), several studies have contended that in case the estimates using complete rankings diverge from the estimates using information on the best alternative alone (or other types of partial rankings), one should have more faith in the latter set of estimates and question the reliability of data on deeper preference rankings. But with our semiparametric methods, it is the former set of estimates that is consistent under a wider variety of true models. And with parametric methods, the discrepancy may arise even when the reliability of data is beyond any doubt as in our simulated samples, because the amount of misspecification bias may vary (non-monotonically) in the depth of rankings used. While the premise that an individual finds it easier to tell her best alternative than, say third- or fourth-best alternative, is intuitively appealing, testing the validity of the conventional wisdom would require the use of a semiparametric method which offers the same degree of robustness regardless of the depth of rankings used in estimation. In our view, the development of a method as such is a promising avenue for future research.

A Proof of Theorem 1

In Appendix A, we provide the proofs of Theorem 1 and of Lemmas 1-3. Lemma 1 establishes that the true preference parameter vector is the unique maximizer of the probability limit of $Q_N(\boldsymbol{b})$. Lemma 2 verifies the continuity property of the probability limit of $Q_N(\boldsymbol{b})$. Lemma 3 shows the uniform convergence of $Q_N(\boldsymbol{b})$ to its probability limit.

Throughout, for $\boldsymbol{b} \in \mathbb{R}^q$, let

$$Q^*(\boldsymbol{b}) \equiv E\left\{\sum_{1\le j<k\le J}\left[1(r_j < r_k)\cdot 1(\boldsymbol{x}'_{jk}\boldsymbol{b} \ge 0) + 1(r_k < r_j)\cdot 1(\boldsymbol{x}'_{kj}\boldsymbol{b} > 0)\right]\right\} \tag{A1}$$

denote the probability limit of $Q_N(\boldsymbol{b})$ in (9).

Lemma 1. *Under Assumptions* 3-4*, the true preference parameter vector $\boldsymbol{\beta}$ uniquely maximizes $Q^*(\boldsymbol{b})$ for $\boldsymbol{b} \in \mathbb{B}$.*

Proof. Applying the law of iterated expectations to the right-hand side of (A1) yields

$$\begin{aligned} Q^*(\boldsymbol{b}) &= E\left\{\sum_{1\le j<k\le J}\left[P(r_j < r_k|\boldsymbol{X})\cdot 1(\boldsymbol{x}'_{jk}\boldsymbol{b} \ge 0) + P(r_k < r_j|\boldsymbol{X})\cdot 1(\boldsymbol{x}'_{kj}\boldsymbol{b} > 0)\right]\right\} \\ &= \sum_{1\le j<k\le J} E\left\{[P(r_j < r_k|\boldsymbol{X}) - P(r_k < r_j|\boldsymbol{X})]\cdot 1(\boldsymbol{x}'_{jk}\boldsymbol{b} \ge 0) + P(r_k < r_j|\boldsymbol{X})\right\}. \end{aligned}$$

By Assumption 3, $\boldsymbol{\beta}$ globally maximizes $Q^*(\boldsymbol{b})$ for $\boldsymbol{b} \in \mathbb{B}$ because the sign of $[P(r_j < r_k|\boldsymbol{X}) - P(r_k < r_j|\boldsymbol{X})]$ is the same as the sign of $\boldsymbol{x}'_{jk}\boldsymbol{\beta}$. Next, we show that $\boldsymbol{\beta}$ is a unique global maximizer of $Q^*(\boldsymbol{b})$. Consider a different parameter vector $\boldsymbol{\beta}^- \in \mathbb{B}$. If, for values of $\boldsymbol{X}$ with positive probability, $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^-$ yield different rankings of systematic utilities, then $\boldsymbol{\beta}^-$ will not maximize $Q^*(\boldsymbol{b})$. In other words, for any $\boldsymbol{X}$ with positive probability, if we observe that $\boldsymbol{x}'_{jk}\boldsymbol{\beta}$ and $\boldsymbol{x}'_{jk}\boldsymbol{\beta}^-$ have opposite signs for some pair of alternatives $j,k \in \mathbb{J}$, then we can conclude $Q^*(\boldsymbol{\beta}) > Q^*(\boldsymbol{\beta}^-)$. We will show this argument for $\beta_1 = 1$; the argument for $\beta_1 = -1$ is similar. If the first element of $\boldsymbol{\beta}^-$, β_1^-, is also 1, then the set of points where $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^-$ yield different rankings of systematic utilities is [25]

$$
\begin{aligned}
D(\boldsymbol{\beta},\boldsymbol{\beta}^-) &= \{\boldsymbol{X}|\boldsymbol{x}'_{jk}\boldsymbol{\beta} < 0 < \boldsymbol{x}'_{jk}\boldsymbol{\beta}^- \; for\, some\, j,k \in \mathbb{J},\; where\; j \neq k\} \\
&= \{\boldsymbol{X}|\tilde{\boldsymbol{x}}'_{jk}\tilde{\boldsymbol{\beta}} < -x_{jk,1} < \tilde{\boldsymbol{x}}'_{jk}\tilde{\boldsymbol{\beta}}^- \; for\, some\, j,k \in \mathbb{J},\; where\; j \neq k\}.
\end{aligned}
$$

By Assumption 4(a), the set $D(\boldsymbol{\beta},\boldsymbol{\beta}^-)$ has probability zero if and only if $\tilde{\boldsymbol{x}}'_{jk}\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{x}}'_{jk}\tilde{\boldsymbol{\beta}}^-$ with probability one for any pair of alternatives $j,k \in \mathbb{J}$, that is, $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta}^-$ with probability one. This contradicts Assumption 4(b). If $\beta_1^- = -1$, the set of points where $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^-$ give different predictions is

$$
D(\boldsymbol{\beta},\boldsymbol{\beta}^-) = \{\boldsymbol{X}|x_{jk,1} < min(\tilde{\boldsymbol{x}}'_{jk}\tilde{\boldsymbol{\beta}}^-, -\tilde{\boldsymbol{x}}'_{jk}\tilde{\boldsymbol{\beta}})\; for\, some\, j,k \in \mathbb{J}\; where\; j \neq k\}.
$$

The probability of $D(\boldsymbol{\beta},\boldsymbol{\beta}^-)$ is positive by Assumption 4(a). Thus, we have proved that the true preference parameter vector $\boldsymbol{\beta}$ uniquely maximizes $Q^*(\boldsymbol{b})$ *for* $\boldsymbol{b} \in \mathbb{B}$. □

Lemma 2. *Under Assumption* 4*,* $Q^*(\boldsymbol{b})$ *is continuous in* $\boldsymbol{b} \in \mathbb{B}$*.*

Proof. For any pair of alternatives $j < k$, define

$$
Q^*_{jk}(\boldsymbol{b}) = E\left\{[1(r_j < r_k) - 1(r_k < r_j)] \cdot 1(\boldsymbol{x}'_{jk}\boldsymbol{b} \geq 0) + 1(r_k < r_j)\right\}. \tag{A2}
$$

Consider the case $b_1 = 1$. The argument for $b_1 = -1$ is symmetric. By the law of iterated expectations,

$$
\begin{aligned}
Q^*_{jk}(\boldsymbol{b}) &= E\left\{[P(r_j < r_k|\boldsymbol{x}_{jk}) - P(r_k < r_j|\boldsymbol{x}_{jk})] \cdot 1(\boldsymbol{x}'_{jk}\boldsymbol{b} \geq 0) + P(r_k < r_j|\boldsymbol{x}_{jk})\right\} \\
&= \int \left\{\int_{-\tilde{\boldsymbol{x}}'_{jk}\tilde{\boldsymbol{b}}}^{\infty} [P(r_j < r_k|\boldsymbol{x}_{jk}) - P(r_k < r_j|\boldsymbol{x}_{jk})]\, g_{jk}(x_{jk,1}|\tilde{\boldsymbol{x}}_{jk})dx_{jk,1}\right\} dP(\tilde{\boldsymbol{x}}_{jk}) + P(r_k < r_j),
\end{aligned} \tag{A3}
$$

where $P(\tilde{\boldsymbol{x}}_{jk})$ denotes the cumulative distribution function of $\tilde{\boldsymbol{x}}_{jk}$. The curly brackets inner integral of the

[25] Recall that $\boldsymbol{x}_{jk} = \boldsymbol{x}_j - \boldsymbol{x}_k$ for any $j,k \in \mathbb{J}$, so we have $\boldsymbol{x}_{jk} = -\boldsymbol{x}_{kj}$. The set $\{\boldsymbol{X}|\boldsymbol{x}'_{jk}\boldsymbol{\beta}^- < 0 < \boldsymbol{x}'_{jk}\boldsymbol{\beta}\; for\, some\, j,k \in \mathbb{J}\}$ is the same as the set $\{\boldsymbol{X}|\boldsymbol{x}'_{kj}\boldsymbol{\beta} < 0 < \boldsymbol{x}'_{kj}\boldsymbol{\beta}^-\; for\, some\, k,j \in \mathbb{J}\}$.

right-hand side of (A3) is a function of $\tilde{\boldsymbol{x}}_{jk}$ and $\tilde{\boldsymbol{b}}$ that is continuous in $\tilde{\boldsymbol{b}} \in \tilde{\mathbb{B}}$. Therefore, by (A1) and (A2),

$$Q^*(\boldsymbol{b}) = \sum_{1 \leq j < k \leq J} Q^*_{jk}(\boldsymbol{b})$$

is also continuous in $\boldsymbol{b} \in \mathbb{B}$. □

Lemma 3. *Under Assumption* 1, $Q_N(\boldsymbol{b})$ *converges almost surely to* $Q^*(\boldsymbol{b})$ *uniformly over* $\boldsymbol{b} \in \mathbb{B}$.

Proof. For any pair of alternatives $j, k \in \mathbb{J}$, define

$$Q_{Njk}(\boldsymbol{b}) = N^{-1} \sum_{n=1}^{N} \left\{ [1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})] \cdot 1(\boldsymbol{x}'_{njk}\boldsymbol{b} \geq 0) + 1(r_{nk} < r_{nj}) \right\}.$$

By Assumption 1 and (A2), we have $Q^*_{jk}(\boldsymbol{b}) = E[Q_{Njk}(\boldsymbol{b})]$. By definition (9),

$$Q_N(\boldsymbol{b}) = \sum_{1 \leq j < k \leq J} Q_{Njk}(\boldsymbol{b}).$$

Recall that $Q^*(\boldsymbol{b}) = \sum_{1 \leq j < k \leq J} Q^*_{jk}(\boldsymbol{b})$. Lemma 4 of Manski (1985) implies that with probability one

$$lim_{N \to \infty} sup_{\boldsymbol{b} \in \mathbb{R}^q} \left| Q_{Njk}(\boldsymbol{b}) - Q^*_{jk}(\boldsymbol{b}) \right| = 0$$

by replacing "$\boldsymbol{x}$" in Manski (1985)'s proof with $\boldsymbol{x}_{jk}$ and "y" in that proof with $[1(r_j < r_k) - 1(r_k < r_j)]$, respectively, for any pair of alternatives $j, k \in \mathbb{J}$. Because $Q_N(\boldsymbol{b})$ is the sum of a finite number of term $Q_{Njk}(\boldsymbol{b})$, $Q_N(\boldsymbol{b})$ converges almost surely to $Q^*(\boldsymbol{b})$ uniformly over $\boldsymbol{b} \in \mathbb{B}$. □

Proof. (Theorem 1) The proof of strong consistency involves verifying the conditions of Theorem 2.1 in Newey and McFadden (1994):

(1) $Q^*(\boldsymbol{b})$ is uniquely maximized at $\boldsymbol{\beta}$;

(2) The parameter space $\mathbb{B}$ is compact;

(3) $Q^*(\boldsymbol{b})$ is continuous in $\boldsymbol{b}$; and

(4) $Q_N(\boldsymbol{b})$ converges almost surely to its probability limit, $Q^*(\boldsymbol{b})$, uniformly over $\boldsymbol{b} \in \mathbb{B}$.

Conditions (1), (3), and (4) are verified by Lemmas 1, 2, and 3, respectively. Condition (2) is guaranteed by Assumption 2. Therefore, the GMS estimator that maximizes $Q_N(\boldsymbol{b})$ converges to $\boldsymbol{\beta}$ almost surely under Assumptions 1-4. $\square$

B Proofs of Theorems 2-4

In Appendix B, we provide the proofs of Theorems 2-4 and of Lemmas 4-9. Lemma 4 shows the uniform convergence of the SGMS objective function to its probability limit. Lemmas 5-6 establish the asymptotic distribution of the normalized form of $\boldsymbol{t}_N(\boldsymbol{\beta},\, h_N)$. Lemmas 7-9 justify that $\boldsymbol{H}_N(\boldsymbol{b}_N^*,\, h_N)$ converges to a nonstochastic matrix in probability. By applying a Taylor series expansion, Lemmas 5-9 can be used to derive the asymptotic distribution of the centered, properly normalized SGMS estimator for the random utility model.

Lemma 4. *Under Assumptions* 1, 4, *and Condition* 1, $Q_N^S(\boldsymbol{b}, h_N)$ *converges almost surely to* $Q^*(\boldsymbol{b})$ *uniformly over* $\boldsymbol{b} \in \mathbb{B}$.

Proof. First, we show that $Q_N^S(\boldsymbol{b}, h_N)$ converges almost surely to $Q_N(\boldsymbol{b})$ uniformly over $\boldsymbol{b} \in \mathbb{B}$ following the method in Lemma 4 of Horowitz (1992). By (9) and (15), we calculate

$$\left|Q_N^S(\boldsymbol{b}, h_N) - Q_N(\boldsymbol{b})\right| \leq \frac{1}{N}\sum_{n=1}^{N}\sum_{1\leq j<k\leq J}\left|1\left(\boldsymbol{x}_{njk}'\boldsymbol{b} > 0\right) - K\left(\boldsymbol{x}_{njk}'\boldsymbol{b}/h_N\right)\right|. \tag{B1}$$

The right-hand side of (B1) is the sum of $c_{N1}(\eta)$ and $c_{N2}(\eta)$, where

$$c_{N1}(\eta) \equiv \frac{1}{N}\sum_{n=1}^{N}\sum_{1\leq j<k\leq J}\left|1\left(\boldsymbol{x}_{njk}'\boldsymbol{b} > 0\right) - K\left(\boldsymbol{x}_{njk}'\boldsymbol{b}/h_N\right)\right| \cdot 1\left(\left|\boldsymbol{x}_{njk}'\boldsymbol{b}\right| \geq \eta\right),$$

$$c_{N2}(\eta) \equiv \frac{1}{N}\sum_{n=1}^{N}\sum_{1\leq j<k\leq J}\left|1\left(\boldsymbol{x}_{njk}'\boldsymbol{b} > 0\right) - K\left(\boldsymbol{x}_{njk}'\boldsymbol{b}/h_N\right)\right| \cdot 1\left(\left|\boldsymbol{x}_{njk}'\boldsymbol{b}\right| < \eta\right),$$

and $\eta \in \mathbb{R}$ is a positive number. Condition 1(b) implies that for any $\delta > 0$, there exists $c > 0$ such that $|K(v) - 1| < \delta \cdot J^{-2}$ and $|K(-v)| < \delta \cdot J^{-2}$ for any $v > c$. As $h_N \to 0$, there exists $N_0 \in \mathbb{N}$ such that $\eta/h_N > c$ for any $N > N_0$. Therefore, $c_{N1}(\eta) < \delta$ for any $N > N_0$. We have shown that for each $\eta > 0$, $c_{N1}(\eta)$ goes to zero uniformly over $\boldsymbol{b} \in \mathbb{B}$ as N goes to ∞. Next consider $c_{N2}(\eta)$. By Condition 1(a), there is a finite C such that

$$c_{N2}(\eta) \leq \sum_{1 \leq j < k \leq J} \left[\frac{C}{N} \sum_{n=1}^{N} 1\left(\left| \boldsymbol{x}'_{njk} \boldsymbol{b} \right| < \eta \right) \right]. \tag{B2}$$

Lemma 4 of Horowitz (1992) implies that the inner-bracket part of the right-hand side of (B2) converges almost surely to $CP(|\boldsymbol{x}'_{jk}\boldsymbol{b}| < \eta)$ uniformly over $\boldsymbol{b} \in \mathbb{B}$ as N approaches ∞ under Assumption 1 and that $P(|\boldsymbol{x}'_{jk}\boldsymbol{b}| < \eta)$ converges to zero uniformly over $\boldsymbol{b} \in \mathbb{B}$ as η goes to zero under Assumption 4 by replacing "$\boldsymbol{x}$" in Horowitz's proof with $\boldsymbol{x}_{jk}$ for any pair of alternatives $1 \leq j < k \leq J$. Because J is finite, the right-hand side of (B2) also converges almost surely to zero uniformly over $\boldsymbol{b} \in \mathbb{B}$ as N goes to ∞ and η goes to 0. Since the right-hand side of (B1) is the sum of $c_{N1}(\eta)$ and $c_{N2}(\eta)$ for any $\eta > 0$, $\left|Q_N^S(\boldsymbol{b}, h_N) - Q_N(\boldsymbol{b})\right|$ converges almost surely to zero uniformly over $\boldsymbol{b} \in \mathbb{B}$ as $N \to \infty$. The absolute difference between $Q_N^S(\boldsymbol{b}, h_N)$ and $Q^*(\boldsymbol{b})$ converges almost surely to zero uniformly over $\boldsymbol{b} \in \mathbb{B}$ because

$$\begin{aligned} \underset{\boldsymbol{b} \in \mathbb{B}}{sup} \left| Q_N^S(\boldsymbol{b}, h_N) - Q^*(\boldsymbol{b}) \right| &\leq \underset{\boldsymbol{b} \in \mathbb{B}}{sup} \left\{ \left| Q_N^S(\boldsymbol{b}, h_N) - Q_N(\boldsymbol{b}) \right| + \left| Q_N(\boldsymbol{b}) - Q^*(\boldsymbol{b}) \right| \right\} \\ &\leq \underset{\boldsymbol{b} \in \mathbb{B}}{sup} \left| Q_N^S(\boldsymbol{b}, h_N) - Q_N(\boldsymbol{b}) \right| + \underset{\boldsymbol{b} \in \mathbb{B}}{sup} \left| Q_N(\boldsymbol{b}) - Q^*(\boldsymbol{b}) \right|, \end{aligned} \tag{B3}$$

and we have proved that the right-hand side of (B3) converges to zero almost surely.[26] □

Proof. (Theorem 2) The proof of strong consistency involves verifying the conditions of Theorem 2.1 in Newey and McFadden (1994):

(1) $Q^*(\boldsymbol{b})$ is uniquely maximized at $\boldsymbol{\beta}$;

(2) The parameter space $\mathbb{B}$ is compact;

(3) $Q^*(\boldsymbol{b})$ is continuous in $\boldsymbol{b}$; and

(4) $Q_N^S(\boldsymbol{b}, h_N)$ converges almost surely to its probability limit, $Q^*(\boldsymbol{b})$, uniformly over $\boldsymbol{b} \in \mathbb{B}$.

Conditions (1), (3), and (4) are verified by Lemmas 1, 2, and 4, respectively. Condition (2) is guaranteed by Assumption 2. Therefore, the SGMS estimator that maximizes $Q_N^S(\boldsymbol{b}, h_N)$ converges to $\boldsymbol{\beta}$ almost surely under Assumptions 1-4 and Condition 1. □

[26] The second term on the right-hand side of (B3) converges almost surely to zero by Lemma 3.

Lemma 5. *Let Assumptions* 1, 3, 6-7 *and Condition* 2 *hold. Then*

(a) $lim_{N\to\infty} E\left[h_N^{-d}\boldsymbol{t}_N(\boldsymbol{\beta},h_N)\right] = \boldsymbol{a}$;

(b) $lim_{N\to\infty} Var\left[(Nh_N)^{1/2}\boldsymbol{t}_N(\boldsymbol{\beta},h_N)\right] = \boldsymbol{\Omega}$.

Proof. First, under Assumption 1 we calculate that

$$\begin{aligned} E\left[h_N^{-d}\boldsymbol{t}_N(\boldsymbol{\beta},h_N)\right] &= \sum_{1\le j<k\le J} E\left\{\left[1(r_j<r_k)-1(r_k<r_j)\right]K'(\boldsymbol{x}'_{jk}\boldsymbol{\beta}/h_N)\tilde{\boldsymbol{x}}_{jk}h_N^{-d-1}\right\} \\ &= \sum_{1\le j<k\le J} \boldsymbol{d}_{jk}, \end{aligned} \tag{B4}$$

where

$$\boldsymbol{d}_{jk} \equiv E\left\{\left[1(r_j<r_k)-1(r_k<r_j)\right]K'(\boldsymbol{x}'_{jk}\boldsymbol{\beta}/h_N)\tilde{\boldsymbol{x}}_{jk}h_N^{-d-1}\right\}. \tag{B5}$$

By the law of iterated expectations,

$$\begin{aligned} \boldsymbol{d}_{jk} &= E\left\{\left[P(r_j<r_k|v_{-j,k},\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}}) - P(r_k<r_j|v_{-j,k},\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})\right]\cdot K'(-v_{-j,k}/h_N)\tilde{\boldsymbol{x}}_{jk}h_N^{-d-1}\right\} \\ &= E\left[\bar{F}_{jk}(v_{-j,k},\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})\cdot K'(-v_{-j,k}/h_N)\tilde{\boldsymbol{x}}_{jk}h_N^{-d-1}\right]. \end{aligned} \tag{B6}$$

By Assumption 3, $\bar{F}_{jk}(0,\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})=0$ for almost every $(\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})$. Under Assumption 6, applying a Taylor series expansion of $\bar{F}_{jk}(v_{-j,k},\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})$ around $v_{-j,k}=0$ yields

$$\bar{F}_{jk}(v_{-j,k},\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}}) = \sum_{i=1}^{d-1}\frac{1}{i!}\bar{F}_{jk}^{(i)}(0,\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})(v_{-j,k})^i + \frac{1}{d!}\bar{F}_{jk}^{(d)}(\xi,\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})(v_{-j,k})^d, \tag{B7}$$

where ξ is between 0 and $v_{-j,k}$. Under Assumption 7(a), applying a Taylor series expansion of the density function $p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})$ around $v_{-j,k}=0$ yields

$$p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}}) = \sum_{c=0}^{d-i-1}\frac{1}{c!}p_{jk}^{(c)}(0|\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})(v_{-j,k})^c + \frac{1}{(d-i)!}p_{jk}^{(d-i)}(\xi_i|\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})(v_{-j,k})^{d-i}, \tag{B8}$$

where ξ_i is between 0 and $v_{-j,k}$, and $1\le i\le d-1$. Combining (B7) and (B8) yields

$$\bar{F}_{jk}(v_{-j,k}, -\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) = \sum_{i=1}^{d-1} \frac{1}{i!(d-i)!} \bar{F}_{jk}^{(i)}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}^{(d-i)}(\xi_i|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})(v_{-j,k})^d$$
$$+\tfrac{1}{d!}\bar{F}_{jk}^{(d)}(\xi, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})(v_{-j,k})^d$$
$$+\sum_{i=1}^{d-1}\sum_{c=0}^{d-i-1} \frac{1}{i!c!} \bar{F}_{jk}^{(i)}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}^{(c)}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})(v_{-j,k})^{i+c}. \tag{B9}$$

Assumptions 6 and 7(a) imply that all of the derivatives on the right-hand side of (B9) are uniformly bounded for almost every $(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$ if $|v_{-j,k}| \leq \eta$ for some $\eta > 0$. Let $\zeta_{jk} \equiv -v_{-j,k}/h_N$. Decompose $\boldsymbol{d}_{jk}$ into two parts, $\boldsymbol{d}_{jk} = \boldsymbol{d}_{jk1} + \boldsymbol{d}_{jk2}$, where

$$\boldsymbol{d}_{jk1} \equiv h_N^{-d} \int_{|h_N\zeta_{jk}|>\eta} \bar{F}_{jk}(-\zeta_{jk}h_N, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{j,k}(-\zeta_{jk}h_N|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}K'(\zeta_{jk})d\zeta_{jk}dP(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) \tag{B10}$$

and

$$\boldsymbol{d}_{jk2} \equiv h_N^{-d} \int_{|h_N\zeta_{jk}|\leq\eta} \bar{F}_{jk}(-\zeta_{jk}h_N, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{j,k}(-\zeta_{jk}h_N|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}K'(\zeta_{jk})d\zeta_{jk}dP(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}). \tag{B11}$$

Under Assumptions 7(a), 7(d), and Condition 2(c),

$$|\boldsymbol{d}_{jk1}| \leq Ch_N^{-d} \int_{|h_N\zeta_{jk}|>\eta} |\tilde{\boldsymbol{x}}_{jk}| \cdot |K'(\zeta_{jk})|\, d\zeta_{jk}dP(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) \to \boldsymbol{0} \; as\, N \to \infty,$$

where $|\boldsymbol{d}_{jk1}|$ denotes the vector of the absolute value of each element in $\boldsymbol{d}_{jk1}$. Plugging (B9) into (B11) and use of Condition 2(b) yield the result that

$$\boldsymbol{d}_{jk2} \to k_d \sum_{i=1}^{d} \frac{1}{i!(d-i)!} E\left[\bar{F}_{jk}^{(i)}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}^{(d-i)}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}\right] \tag{B12}$$

as N goes to ∞ by Lebesgue's dominated convergence theorem. Part (a) is established by (B4), (B12), and (24).

Next consider $Var[(Nh_N)^{1/2}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)]$. By Assumption 1,

$$Var[(Nh_N)^{1/2}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)] = h_N Var\left\{\sum_{1\le j<k\le J}\left[1(r_j<r_k)-1(r_k<r_j)\right]K'(\boldsymbol{x}'_{jk}\boldsymbol{\beta}/h_N)\tilde{\boldsymbol{x}}_{jk}h_N^{-1}\right\}.$$

Let

$$\boldsymbol{e}_N \equiv \sum_{1\le j<k\le J}\left[1(r_j<r_k)-1(r_k<r_j)\right]K'(\boldsymbol{x}'_{jk}\boldsymbol{\beta}/h_N)\tilde{\boldsymbol{x}}_{jk}h_N^{-1}, \tag{B13}$$

then

$$Var[(Nh_N)^{1/2}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)] = h_N E(\boldsymbol{e}_N\boldsymbol{e}'_N) - h_N E(\boldsymbol{e}_N)E(\boldsymbol{e}'_N). \tag{B14}$$

In part (a), we show that $E[h_N^{-d}\boldsymbol{e}_N]$, which equals $E[h_N^{-d}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)]$, converges to $\boldsymbol{a}$, implying that $h_N E(\boldsymbol{e}_N)E(\boldsymbol{e}'_N)$ converges to zero as N goes to ∞. Since the binomial choice setting where $J=2$ has been discussed in Horowitz (1992), the following discussion focuses on the case where $J\ge 3$. Define

$$h_N E(\boldsymbol{e}_N\boldsymbol{e}'_N) = \boldsymbol{L}_{N1}+\boldsymbol{L}_{N2}+\boldsymbol{L}_{N3}, \text{ where} \tag{B15}$$

$$\boldsymbol{L}_{N1} \equiv \sum_{1\le j<k\le J} h_N^{-1}E\left\{\left[1(r_j<r_k)-1(r_k<r_j)\right]^2\left[K'(\boldsymbol{x}'_{jk}\boldsymbol{\beta}/h_N)\right]^2\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk}\right\}, \tag{B16}$$

$$\begin{aligned}\boldsymbol{L}_{N2} \equiv \sum_{1\le j<k<l\le J} 2h_N^{-1}E\{ & \\ &\left[1(r_j<r_k)-1(r_k<r_j)\right]\left[1(r_j<r_l)-1(r_l<r_j)\right]K'(\boldsymbol{x}'_{jk}\boldsymbol{\beta}/h_N)K'(\boldsymbol{x}'_{jl}\boldsymbol{\beta}/h_N)\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jl}\\ &+\left[1(r_j<r_k)-1(r_k<r_j)\right]\left[1(r_k<r_l)-1(r_l<r_k)\right]K'(\boldsymbol{x}'_{jk}\boldsymbol{\beta}/h_N)K'(\boldsymbol{x}'_{kl}\boldsymbol{\beta}/h_N)\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{kl}\\ &+\left[1(r_j<r_l)-1(r_l<r_j)\right]\left[1(r_k<r_l)-1(r_l<r_k)\right]K'(\boldsymbol{x}'_{jl}\boldsymbol{\beta}/h_N)K'(\boldsymbol{x}'_{kl}\boldsymbol{\beta}/h_N)\tilde{\boldsymbol{x}}_{jl}\tilde{\boldsymbol{x}}'_{kl}\Big\},\end{aligned} \tag{B17}$$

$$\begin{aligned}\boldsymbol{L}_{N3} \equiv \sum_{j<k,l<m,j<l,k\ne l,k\ne m}^{j,k,l,m\in\mathbb{J}} & 2h_N^{-1}E\left\{\left[1(r_j<r_k)-1(r_k<r_j)\right]\left[1(r_l<r_m)-1(r_m<r_l)\right]\right.\\ &\left.\cdot K'(\boldsymbol{x}'_{jk}\boldsymbol{\beta}/h_N)K'(\boldsymbol{x}'_{lm}\boldsymbol{\beta}/h_N)\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{lm}\right\}\end{aligned} \tag{B18}$$

when $J>3$, and $\boldsymbol{L}_{N3}\equiv 0$ when $J=3$. Define $\zeta_{jk} = -v_{-j,k}/h_N$ for any pair of alternatives $j,k\in\mathbb{J}$. By the law of iterated expectations,

$$\boldsymbol{L}_{N1} = \sum_{1\le j<k\le J} \int [2F_{jk}(-h_N\zeta_{jk}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) - \bar{F}_{jk}(-h_N\zeta_{jk}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})] \cdot p_{jk}(-h_N\zeta_{jk}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk}[K'(\zeta_{jk})]^2 d\zeta_{jk} dP(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}). \tag{B19}$$

By Assumptions 3, 6, 7, Condition 2(a), and Lebesgue's dominated convergence theorem, the right-hand side of (B19) converges to $\boldsymbol{\Omega}$ when N goes to ∞. By Assumption 7(b),

$$|\boldsymbol{L}_{N2}| \le \sum_{1\le j<k<l\le J} 2Ch_N[\int |K'(\zeta_{jk})K'(\zeta_{jl})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jl}| d\zeta_{jk}d\zeta_{jl}dP(\tilde{\boldsymbol{v}}_{-j,kl}, \tilde{\boldsymbol{X}}) + \int |K'(\zeta_{jk})K'(\zeta_{kl})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{kl}| d\zeta_{kj}d\zeta_{kl}dP(\tilde{\boldsymbol{v}}_{-k,jl}, \tilde{\boldsymbol{X}}) + \int |K'(\zeta_{jl})K'(\zeta_{kl})\tilde{\boldsymbol{x}}_{jl}\tilde{\boldsymbol{x}}'_{kl}|\, d\zeta_{lj}d\zeta_{lk}dP(\tilde{\boldsymbol{v}}_{-l,jk}, \tilde{\boldsymbol{X}})]. \tag{B20}$$

Therefore, the right-hand side of (B20) converges to zero under Assumption 7(d) and Condition 2(a). If $J > 3$ and by Assumption 7(c),

$$|\boldsymbol{L}_{N3}| \le \sum_{j<k,l<m,j<l,k\ne l,k\ne m}^{j,k,l,m\in\mathbb{J}} 2Ch_N \int |K'(\zeta_{jk})K'(\zeta_{lm})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{lm}| d\zeta_{jk}d\zeta_{lm}dP(\tilde{\boldsymbol{v}}_{-\{k,m\}}, \tilde{\boldsymbol{X}}). \tag{B21}$$

By Assumption 7(d) and Condition 2(a), the right-hand side of (B21) converges to zero as N goes to ∞. We have proved part (b) by (B14) and (B15). □

Lemma 6. *Let Assumptions* 1, 3, 6-7 *and Condition* 2 *hold. Then*

(a) *If* $Nh_N^{2d+1} \to \infty$ *as* $N \to \infty$, *then* $h_N^{-d}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N) \xrightarrow{p} \boldsymbol{a}$.

(b) *If* $Nh_N^{2d+1} \to \lambda$, *where* $\lambda \in (0,\infty)$, *as* $N \to \infty$, *then* $(Nh_N)^{1/2}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N) \xrightarrow{d} MVN(\lambda^{1/2}\boldsymbol{a}, \boldsymbol{\Omega})$.

Proof. If $Nh_N^{2d+1} \to \infty$ as $N \to \infty$, then

$$Var[h_N^{-d}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)] = \frac{1}{Nh_N^{2d+1}} Var[(Nh_N)^{1/2}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)]$$

converges to zero by Lemma 5(b). Therefore, Lemma 5 and Chebyshev's Theorem imply Lemma 6(a). Next consider part (b). Define $\mathbf{w}_N = (Nh_N)^{1/2} \{\boldsymbol{t}_N(\boldsymbol{\beta}, h_N) - E[\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)]\}$. Lemma 5(a) implies that

$$(Nh_N)^{1/2}E[\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)] = (Nh_N^{2d+1})^{1/2}E[h_N^{-d}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N)] \to \lambda^{1/2}\boldsymbol{a},$$

so it suffices to prove that $\boldsymbol{\gamma}'\mathbf{w}_N$ is asymptotically distributed as $MVN(\mathbf{0}, \boldsymbol{\gamma}'\boldsymbol{\Omega}\boldsymbol{\gamma})$ for any nonstochastic $q-1$ dimensional vector $\boldsymbol{\gamma}$ such that $\boldsymbol{\gamma}'\boldsymbol{\gamma} = 1$. Let

$$\boldsymbol{t}_N(\boldsymbol{\beta}, h_N) = N^{-1}\sum_{n=1}^{N}\boldsymbol{t}_{Nn}(\boldsymbol{\beta}, h_N),$$

where

$$\boldsymbol{t}_{Nn}(\boldsymbol{\beta}, h_N) \equiv \sum_{1\leq j<k\leq J}[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})]\, K'(\boldsymbol{x}'_{njk}\boldsymbol{\beta}/h_N)\tilde{\boldsymbol{x}}_{njk}h_N^{-1}.$$

By definition, we have

$$\boldsymbol{\gamma}'\mathbf{w}_N = (h_N/N)^{1/2}\boldsymbol{\gamma}'\sum_{n=1}^{N}\{\boldsymbol{t}_{Nn}(\boldsymbol{\beta}, h_N) - E[\boldsymbol{t}_{Nn}(\boldsymbol{\beta}, h_N)]\}.$$

Let $CF_N(\tau)$ denote the characteristic function of $\boldsymbol{\gamma}'\mathbf{w}_N$. Applying the proof of Lemma 6 in Horowitz (1992) yields the result that

$$\lim_{N\to\infty} CF_N(\tau) = exp(-\boldsymbol{\gamma}'\boldsymbol{\Omega}\boldsymbol{\gamma}\tau^2/2),$$

which is the same as the characteristic function of $MVN(\mathbf{0}, \boldsymbol{\gamma}'\boldsymbol{\Omega}\boldsymbol{\gamma})$. □

Lemma 7. *Let Assumptions* 1, 3-4, 6-8 *and Conditions* 1-2 *hold. For any pair of alternatives* $j, k \in \mathbb{J}$, *assume that* $||\tilde{\boldsymbol{x}}_{jk}|| \leq c$ *for some* $c > 0$. *Let* η *be some positive real number such that* $p_{jk}^{(1)}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$, $\bar{F}_{jk}^{(1)}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$, *and* $\bar{F}_{jk}^{(2)}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$ *exist and are uniformly bounded for almost every* $(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$

if $|v_{-j,k}| \leq \eta$*. For* $\boldsymbol{\theta} \in \mathbb{R}^{q-1}$*, define*

$$\boldsymbol{t}_N^*(\boldsymbol{\theta}) = (Nh_N^2)^{-1} \sum_{n=1}^{N} \sum_{1 \leq j < k \leq J} \left[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})\right] K'(\boldsymbol{x}_{njk}'\boldsymbol{\beta}/h_N + \tilde{\boldsymbol{x}}_{njk}'\boldsymbol{\theta})\tilde{\boldsymbol{x}}_{njk}.$$

Define the sets $\boldsymbol{\Theta}_N$*, where* $N = 1, 2, \ldots,$ *by* $\boldsymbol{\Theta}_N = \left\{\boldsymbol{\theta} : \boldsymbol{\theta} \in \mathbb{R}^{q-1}, h_N \|\boldsymbol{\theta}\| \leq \eta/2c\right\}$.

(a) *Then*

$$\underset{N\to\infty}{plim} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_N} \left\|\boldsymbol{t}_N^*(\boldsymbol{\theta}) - E\left[\boldsymbol{t}_N^*(\boldsymbol{\theta})\right]\right\| = 0. \tag{B22}$$

(b) *There are finite numbers* α_1 *and* α_2 *such that for all* $\boldsymbol{\theta} \in \Theta_N$

$$\left\|E[\boldsymbol{t}_N^*(\boldsymbol{\theta})] - \boldsymbol{H}\boldsymbol{\theta}\right\| \leq o(1) + \alpha_1 h_N \|\boldsymbol{\theta}\| + \alpha_2 h_N \|\boldsymbol{\theta}\|^2 \tag{B23}$$

uniformly over $\boldsymbol{\theta} \in \boldsymbol{\Theta}_N$.

Proof. Define

$$\begin{aligned} \boldsymbol{g}_{Nn}(\boldsymbol{\theta}) \quad = \quad & \sum_{1 \leq j < k \leq J} \Big\{\left[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})\right] K'(\boldsymbol{x}_{njk}'\boldsymbol{\beta}/h_N + \tilde{\boldsymbol{x}}_{njk}'\boldsymbol{\theta})\tilde{\boldsymbol{x}}_{njk} \\ & -E\Big[\left[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})\right] K'(\boldsymbol{x}_{njk}'\boldsymbol{\beta}/h_N + \tilde{\boldsymbol{x}}_{njk}'\boldsymbol{\theta})\tilde{\boldsymbol{x}}_{njk}\Big]\Big\}. \end{aligned} \tag{B24}$$

The remaining part of the proof of (B22) follows the proof of (A15) in Lemma 7 of Horowitz (1992). Next, we prove (B23). Define

$$E\left[\boldsymbol{t}_N^*(\boldsymbol{\theta})\right] = \sum_{1 \leq j < k \leq J} \boldsymbol{t}_{Njk}^*(\boldsymbol{\theta}),$$

where

$$\begin{aligned} \boldsymbol{t}_{Njk}^*(\boldsymbol{\theta}) \quad & \equiv \quad h_N^{-2} E\left\{\left[1(r_j < r_k) - 1(r_k < r_j)\right] K'(\boldsymbol{x}_{jk}'\boldsymbol{\beta}/h_N + \tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\theta})\tilde{\boldsymbol{x}}_{jk}\right\} \\ & = \quad h_N^{-2} E\left[\bar{F}_{jk}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) K'(-v_{-j,k}/h_N + \tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\theta})\tilde{\boldsymbol{x}}_{jk}\right] \end{aligned} \tag{B25}$$

and the second equality in (B25) is implied by the law of iterated expectations.

Decompose the right-hand side of (B25) into two parts: $\boldsymbol{t}^*_{Njk}(\boldsymbol{\theta}) = \boldsymbol{t}^*_{Njk1} + \boldsymbol{t}^*_{Njk2}$, where

$$\boldsymbol{t}^*_{Njk1} \equiv h_N^{-2} \int_{|v_{-j,k}|>\eta} \bar{F}_{jk}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) K'(-v_{-j,k}/h_N + \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}) \cdot \tilde{\boldsymbol{x}}_{jk} p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) dv_{-j,k} dP(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$$

and

$$\boldsymbol{t}^*_{Njk2} \equiv h_N^{-2} \int_{|v_{-j,k}|\leq\eta} \bar{F}_{jk}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) K'(-v_{-j,k}/h_N + \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}) \cdot \tilde{\boldsymbol{x}}_{jk} p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) dv_{-j,k} dP(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}).$$

For some finite constant $C > 0$, by Assumption 7(a) and $||\tilde{\boldsymbol{x}}_{jk}|| \leq c$,

$$\left\|\boldsymbol{t}^*_{Njk1}\right\| \leq Cch_N^{-2} \int_{|v_{-j,k}|>\eta} \left|K'(-v_{-j,k}/h_N + \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta})\right| dv_{-j,k} dP(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}).$$

Let $\zeta_{jk} \equiv -v_{-j,k}/h_N + \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}$. Since $h_N||\boldsymbol{\theta}|| \leq \eta/2c$ and $||\tilde{\boldsymbol{x}}_{jk}|| \leq c$, $|v_{-j,k}| > \eta$ implies that

$$|\zeta_{jk}| > | - v_{-j,k}/h_N| - |\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}| > \eta/2h_N$$

and

$$\left\|\boldsymbol{t}^*_{Njk1}\right\| \leq Cch_N^{-1} \int_{|\zeta_{jk}|>\eta/2h_N} |K'(\zeta_{jk})| \, d\zeta_{jk}. \tag{B26}$$

We conclude

$$\lim_{N\to\infty} \sup_{\boldsymbol{\theta}\in\Theta_N} \left\|\boldsymbol{t}^*_{Njk1}\right\| = 0 \tag{B27}$$

because the term on the right-hand side of (B26) converges to zero by Condition 2(c). Next, we consider

$\boldsymbol{t}^*_{Njk2}$. If $|v_{-j,k}| \leq \eta$, substitution of $d = 2$ into the right-hand side of (B9) yields

$$\begin{aligned}\bar{F}_{jk}(v_{-j,k}, -\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) = {} & \bar{F}^{(1)}_{jk}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})v_{-j,k} \\ & + \bar{F}^{(1)}_{jk}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p^{(1)}_{jk}(\xi_1|\xi_1, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})(v_{-j,k})^2 \\ & + (1/2)\bar{F}^{(2)}_{jk}(\xi, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})(v_{-j,k})^2,\end{aligned} \tag{B28}$$

where ξ and ξ_1 are between zero and $v_{-j,k}$.

Decompose $\boldsymbol{t}^*_{Njk2}$ into two parts, $\boldsymbol{t}^*_{Njk2} = \boldsymbol{s}_{Njk1} + \boldsymbol{s}_{Njk2}$, where

$$\begin{aligned}\boldsymbol{s}_{Njk1} \equiv h_N^{-2} \int_{|v_{-j,k}| \leq \eta} & \bar{F}^{(1)}_{jk}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) \\ & \cdot \tilde{\boldsymbol{x}}_{jk} v_{-j,k} K'(-v_{-j,k}/h_N + \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}) dv_{-j,k} dP(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{s}_{Njk2} \equiv h_N^{-2} \int_{|v_{-j,k}| \leq \eta} & \Big[\bar{F}^{(1)}_{jk}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p^{(1)}_{jk}(\xi_1|\xi_1, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) \\ & + (1/2)\bar{F}^{(2)}_{jk}(\xi, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\Big] \\ & \cdot \tilde{\boldsymbol{x}}_{jk}(v_{-j,k})^2 K'(-v_{-j,k}/h_N + \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}) dv_{-j,k} dP(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}).\end{aligned}$$

Define $\zeta_{jk} = -v_{-j,k}/h_N + \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}$, then

$$\begin{aligned}\boldsymbol{s}_{Njk1} = \int_{|\zeta_{jk} - \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}| \leq \eta/h_N} & \bar{F}^{(1)}_{jk}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) \\ & \cdot \tilde{\boldsymbol{x}}_{jk}(\zeta_{jk} - \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}) K'(\zeta_{jk}) d\zeta_{jk} dP(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}).\end{aligned}$$

Because $\int_{-\infty}^{\infty} \zeta K'(\zeta) d\zeta = 0$ and $|\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta} h_N| \leq \eta/2$,

$$\begin{aligned}\left|\int_{|\zeta_{jk} - \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}| \leq \eta/h_N} \zeta_{jk} K'(\zeta_{jk}) d\zeta_{jk}\right| &= \left|\int_{|\zeta_{jk} - \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}| > \eta/h_N} \zeta_{jk} K'(\zeta_{jk}) d\zeta_{jk}\right| \\ &\leq \int_{|\zeta_{jk}| > \eta/2h_N} |\zeta_{jk} K'(\zeta_{jk})| d\zeta_{jk}.\end{aligned} \tag{B29}$$

By Condition 2(c), the right-hand term of (B29) is bounded uniformly over $\boldsymbol{\theta} \in \boldsymbol{\Theta}_N$ and it converges to zero.

Therefore, by Lebesgue's dominated convergence theorem,

$$\underset{N\to\infty}{lim}\ \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}_N}{sup}\left|\int_{|\zeta_{jk}-\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}|\leq\eta/h_N}\bar{F}^{(1)}_{jk}(0,\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})p_{jk}(0|\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})\right. \\ \left.\cdot\tilde{\boldsymbol{x}}_{jk}\zeta_{jk}K'(\zeta_{jk})d\zeta_{jk}dP(\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})\right|=0. \tag{B30}$$

In addition,

$$||\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}\int_{|\zeta_{jk}-\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}|\leq\eta/h_N}K'(\zeta_{jk})d\zeta_{jk}-\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}|| \\ \leq|\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}h_N|h_N^{-1}\int_{|\zeta_{jk}-\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}|>\eta/h_N}|K'(\zeta_{jk})|d\zeta_{jk}\leq(\eta/2)h_N^{-1}\int_{|\zeta_{jk}-\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}|>\eta/h_N}|K'(\zeta_{jk})|d\zeta_{jk}. \tag{B31}$$

By Condition 2(c), the right-hand side of (B31) is bounded uniformly over $\boldsymbol{\theta}\in\boldsymbol{\Theta}_N$ and it converges to zero. Next, by Lebesgue's dominated convergence theorem, Condition 1, and the definition of $\boldsymbol{H}$,

$$\underset{N\to\infty}{lim}||\underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}_N}{sup}\sum_{1\leq j<k\leq J}\int_{|\zeta_{jk}-\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}|\leq\eta/h_N}\bar{F}^{(1)}_{jk}(0,\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})p_{jk}(0|\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}}) \\ \cdot\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}K'(\zeta_{jk})d\zeta_{jk}dP(\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}})-\boldsymbol{H}\boldsymbol{\theta}||=0. \tag{B32}$$

For some finite $C>0$,

$$||\boldsymbol{s}_{Njk2}||\leq Ch_N\int_{|\zeta_{jk}-\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}|\leq\eta/h_N}(\zeta_{jk}-\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta})^2|K'(\zeta_{jk})|d\zeta_{jk}dP(\tilde{\boldsymbol{v}}_{-j,k},\tilde{\boldsymbol{X}}) \\ \leq o(1)+\alpha_{jk1}h_N||\boldsymbol{\theta}||+\alpha_{jk2}h_N||\boldsymbol{\theta}||^2 \tag{B33}$$

for some finite α_{jk1} and α_{jk2}. Since J is finite, part (b) is established by combining (B27), (B30), (B32), and (B33). □

Lemma 8. *Let Assumptions* 1-9 *and Conditions* 1-2 *hold and define* $\boldsymbol{\theta}_N=(\tilde{\boldsymbol{b}}^S_N-\tilde{\boldsymbol{\beta}})/h_N$, *where* $\boldsymbol{b}^S_N$ *is a SGMS estimator. Then the probability limit of* $\boldsymbol{\theta}_N$ *is* $\mathbf{0}$.

Proof. Pick γ to be a finite number such that $P(||\tilde{\boldsymbol{x}}_{jk}||\leq\gamma,\forall 1\leq j<k\leq J)\geq 1-\delta$ for any $\delta>0$. Let P_δ be the probability distribution of $\boldsymbol{X}$ conditional on the event $C_\gamma\equiv\{\boldsymbol{X}:||\tilde{\boldsymbol{x}}_{jk}||\leq\gamma,\forall 1\leq j<k\leq J\}$. Define C'_γ as the complement of C_γ. The remaining part of the proof follows the proof of Lemma 8 of Horowitz (1992). □

Lemma 9. *Let Assumptions* 1-8 *and Conditions* 1-2 *hold. Let* $\{\boldsymbol{\beta}_N \equiv (\beta_{N1}, \tilde{\boldsymbol{\beta}}_N')'\}$ *be any sequence in* $\mathbb{B}$ *such that* $(\boldsymbol{\beta}_N - \boldsymbol{\beta})/h_N \to \mathbf{0}$ *as* $N \to \infty$*. Then the probability limit of* $\boldsymbol{H}_N(\boldsymbol{\beta}_N, h_N)$ *is* $\boldsymbol{H}$*.*

Proof. Assume that $\beta_{N1} = \beta_1 \in \{-1, 1\}$ because this is true for sufficiently large N if $(\beta_{N1} - \beta_1)$ converges to zero as N goes to ∞. Let $\boldsymbol{\theta}_N \equiv (\tilde{\boldsymbol{\beta}}_N - \tilde{\boldsymbol{\beta}})/h_N$. Let $\{a_N\}$ be a positive sequence such that a_N goes to ∞ and $a_N\boldsymbol{\theta_N}$ goes to $\mathbf{0}$. Define $\boldsymbol{X}_N = \{\boldsymbol{X} : ||\tilde{\boldsymbol{x}}_{jk}|| \le a_N, \forall\, 1 \le j < k \le J\}$. Given any $\epsilon > 0$,

$$\underset{N\to\infty}{lim} P\left[|\boldsymbol{H}_N(\boldsymbol{\beta}_N, h_N) - \boldsymbol{H}| > \epsilon\right] = \underset{N\to\infty}{lim} P\left[|\boldsymbol{H}_N(\boldsymbol{\beta}_N, h_N) - \boldsymbol{H}| > \epsilon | \boldsymbol{X}_N\right].$$

Therefore, it suffices to show that $E[\boldsymbol{H}_N(\boldsymbol{\beta}_N, h_N)|\boldsymbol{X}_N]$ converges to $\boldsymbol{H}$ and $Var[\boldsymbol{H}_N(\boldsymbol{\beta}_N, h_N)|\boldsymbol{X}_N]$ converges to $\mathbf{0}$ as N goes to ∞ by Chebyshev's Theorem. Consider $E[\boldsymbol{H}_N(\boldsymbol{\beta}_N, h_N)|\boldsymbol{X}_N]$ first.

Define $\boldsymbol{E}_N = E[\boldsymbol{H}_N(\boldsymbol{\beta}_N, h_N)|\boldsymbol{X}_N]$, then $\boldsymbol{E}_N = \sum_{1\le j<k\le J} \boldsymbol{E}_{Njk}$, where

$$\begin{aligned}\boldsymbol{E}_{Njk} \equiv h_N^{-2} \textstyle\int & \bar{F}_{jk}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}_{jk}' \\ & \cdot K''(-v_{-j,k}/h_N + \tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\theta}_N)dv_{-j,k}dP_{Njk}(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}),\end{aligned} \tag{B34}$$

and P_{Njk} denote the distribution of $(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$ conditional on $\boldsymbol{X}_N$. By Assumptions 6 and 7(a), there exists an η such that $\bar{F}_{jk}^{(1)}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$, $\bar{F}_{jk}^{(2)}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$, and $p_{jk}^{(1)}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$ exist and are almost surely uniformly bounded if $|v_{-j,k}| \le \eta$. Therefore, substitution of (B28) into (B34) yields

$$\boldsymbol{E}_{Njk} = \boldsymbol{I}_{Njk1} + \boldsymbol{I}_{Njk2} + \boldsymbol{I}_{Njk3},$$

where

$$\begin{aligned}\boldsymbol{I}_{Njk1} \equiv h_N^{-2} \textstyle\int_{|v_{-j,k}|\le\eta} & \bar{F}_{jk}^{(1)}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}_{jk}' \\ & \cdot v_{-j,k}K''(-v_{-j,k}/h_N + \tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\theta}_N)dv_{-j,k}dP_{Njk}(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}),\end{aligned}$$

$$\begin{aligned}\boldsymbol{I}_{Njk2} \equiv h_N^{-2} \textstyle\int_{|v_{-j,k}|\le\eta} & \Big[\bar{F}_{jk}^{(1)}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}^{(1)}(\xi_1|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) \\ & + (1/2)\bar{F}_{jk}^{(2)}(\xi, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\Big]\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}_{jk}' \\ & \cdot (v_{-j,k})^2 K''(-v_{-j,k}/h_N + \tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\theta}_N)dv_{-j,k}dP_{Njk}(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}),\end{aligned}$$

and

$$\boldsymbol{I}_{Njk3} \equiv h_N^{-2} \int_{|v_{-j,k}|>\eta} \bar{F}_{jk}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk} \cdot K''(-v_{-j,k}/h_N + \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}_N)dv_{-j,k}dP_{Njk}(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}). \tag{B35}$$

Define $\zeta_{jk} = -v_{-j,k}/h_N + \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}_N$. Then

$$\boldsymbol{I}_{Njk1} = -\int_{|\zeta_{jk}-\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}_N|\le\eta/h_N} \bar{F}_{jk}^{(1)}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk} \cdot(\zeta_{jk} - \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}_N)K''(\zeta_{jk})d\zeta_{jk}dP_{Njk}(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}).$$

Because $|\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}_N| \le a_N||\boldsymbol{\theta}_N|| \to 0$, by Conditions 1-2,

$$\boldsymbol{I}_{Njk1} \to E\left[\bar{F}_{jk}^{(1)}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk}\right]. \tag{B36}$$

For some finite $C > 0$, by Assumptions 6-7 and Condition 2(a),

$$|\boldsymbol{I}_{Njk2}| \le Ch_N \int_{|\zeta_{jk}-\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}_N|\le\eta/h_N} |\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk}| \cdot(\zeta_{jk} - \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}_N)^2|K''(\zeta_{jk})|d\zeta_{jk}dP_{Njk}(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) \to \boldsymbol{0} \; as\, N \to \infty. \tag{B37}$$

Finally by Assumption 7(a) we calculate (B35):

$$|\boldsymbol{I}_{Njk3}| \le Ch_N^{-1} \int_{|\zeta_{jk}-\tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}_N|>\eta/h_N} |\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk}| \cdot |K''(\zeta_{jk})|d\zeta_{jk}dP_{Njk}(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}). \tag{B38}$$

Under Assumption 7(d) and Condition 2(c), the right-hand side of (B38) converges to zero as N goes to ∞. Since J is finite, combination of (B36), (B37), and (B38) establishes that

$$\boldsymbol{E}_N = \sum_{1\le j<k\le J} \boldsymbol{E}_{Njk} = \sum_{1\le j<k\le J} (\boldsymbol{I}_{Njk1} + \boldsymbol{I}_{Njk2} + \boldsymbol{I}_{Njk3}) \to \boldsymbol{H} \; as\; N \to \infty.$$

Next, based on $\boldsymbol{E}_N$, calculate $Var[\boldsymbol{H}(\boldsymbol{\beta}_N, h_N)|\boldsymbol{X}_N]$:

$$\begin{aligned} & Var[\boldsymbol{H}(\boldsymbol{\beta}_N, h_N)|\boldsymbol{X}_N] \qquad \text{(B39)} \\ = \; & N^{-1} Var\left\{ \sum_{1\le j<k\le J} \left[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})\right] K''(-v_{-j,k}/h_N + \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}_N)\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk}h_N^{-2} \,|\boldsymbol{X}_N \right\} \\ = \; & N^{-1}E[\boldsymbol{r}_N(\boldsymbol{\theta}_N)\boldsymbol{r}_N(\boldsymbol{\theta}_N)'|\boldsymbol{X}_N] + O(N^{-1}), \end{aligned}$$

where $\boldsymbol{r}_N(\boldsymbol{\theta}_N) \equiv \sum_{1\le j<k\le J} \left[1(r_j < r_k) - 1(r_k < r_j)\right] K''(-v_{-j,k}/h_N + \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}_N) vec(\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk})h_N^{-2}$.

Define $\zeta_{jk} = -v_{-j,k}/h_N + \tilde{\boldsymbol{x}}'_{jk}\boldsymbol{\theta}_N$. For some finite constant C, by Assumption 7,

$$\begin{aligned} N^{-1}E[\boldsymbol{r}_N(\boldsymbol{\theta}_N)\boldsymbol{r}_N(\boldsymbol{\theta}_N)'|\boldsymbol{X}_N] \le \; & Ch_N(Nh_N^4)^{-1} \sum_{1\le j<k\le J} \int vec(\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk})vec(\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk})' \qquad \text{(B40)} \\ & \cdot [K''(\zeta_{jk})]^2 d\zeta_{jk} dP_{Njk}(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) \\ & + Ch_N^2(Nh_N^4)^{-1} \sum_{\substack{j,k,l\in\mathbb{J} \\ j\ne k, j\ne l, k\ne l}} \int vec(\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk})vec(\tilde{\boldsymbol{x}}_{jl}\tilde{\boldsymbol{x}}'_{jl})' \\ & \cdot |K''(\zeta_{jk})K''(\zeta_{jl})| d\zeta_{jk} d\zeta_{jl} dP_{Njkl}(\tilde{\boldsymbol{v}}_{-j,kl}, \tilde{\boldsymbol{X}}) \\ & + 2Ch_N^2(Nh_N^4)^{-1} \sum_{\substack{j,k,l,m\in\mathbb{J} \\ j<k, l<m, j<l, k\ne l, k\ne m}} \int vec(\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk})vec(\tilde{\boldsymbol{x}}_{lm}\tilde{\boldsymbol{x}}'_{lm})' \\ & \cdot |K''(\zeta_{jk})K''(\zeta_{lm})| d\zeta_{jk} d\zeta_{lm} dP_{Njklm}(\tilde{\boldsymbol{v}}_{-\{k,m\}}, \tilde{\boldsymbol{X}}), \end{aligned}$$

where P_{Njkl} denotes the distribution of $(\tilde{\boldsymbol{v}}_{-j,kl}, \tilde{\boldsymbol{X}})$ conditional on $\boldsymbol{X}_N$ and P_{Njklm} denotes the distribution of $(\tilde{\boldsymbol{v}}_{-\{k,m\}}, \tilde{\boldsymbol{X}})$ conditional on $\boldsymbol{X}_N$. Notice that the second term on the right-hand side of (B40) is relevant only if $J > 2$, and the third term on the right-hand side of (B40) is relevant only if $J > 3$. We can show that each term on the right-hand side of (B40) converges to zero by Assumptions 7-8 and Condition 2(a). Therefore, it follows from (B39) that $Var[\boldsymbol{H}(\boldsymbol{\beta}_N, h_N)|\boldsymbol{X}_N]$ converges to zero as N goes to ∞. □

Proof. (Theorem 3) By Theorem 2 and Assumption 5, $b^S_{N,1} = \beta_1$ and $\tilde{\boldsymbol{b}}^S_N$ is an interior point of $\tilde{\mathbb{B}}$ with probability approaching one as N goes to ∞. Consequently, the first order condition $\boldsymbol{t}_N(\boldsymbol{b}^S_N, h_N) = 0$ holds with probability approaching one. A Taylor series expansion of $\boldsymbol{t}_N(\boldsymbol{b}_N^S, h_N)$ around $\boldsymbol{b}_N^S = \boldsymbol{\beta}$ yields

$$\boldsymbol{t}_N(\boldsymbol{\beta}, h_N) + \boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N)(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) = 0, \tag{B41}$$

where $\boldsymbol{b}_N^*$ is between $\boldsymbol{\beta}$ and $\boldsymbol{b}_N^S$.

Part (a): By (B41),

$$h_N^{-d}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N) + \boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N)h_N^{-d}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) = 0$$

with probability approaching one as N goes to ∞. Lemmas 8-9 imply that the probability limit of $\boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N)$ is $\boldsymbol{H}$. Because $\boldsymbol{H}$ is nonsingular by Assumption 9, we have

$$h_N^{-d}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) = -\boldsymbol{H}^{-1}h_N^{-d}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N) + o_p(1).$$

Part (a) is a direct result of Lemma 6(a).

Part (b): By (B41),

$$(Nh_N)^{1/2}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N) + \boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N)(Nh_N)^{1/2}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) = 0$$

with probability approaching one as N goes to ∞. Application of Lemmas 8-9 and Assumption 9 yields

$$(Nh_N)^{1/2}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) = -\boldsymbol{H}^{-1}(Nh_N)^{1/2}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N) + o_p(1).$$

Part (b) is a direct result of Lemma 6(b).

Part (c): By the cyclic property of trace,

$$E_A[(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})'\boldsymbol{W}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})] = trace\{\boldsymbol{W}E_A[(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})']\}.$$

Part (b) implies that

$$E_A[(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})']$$
$$= N^{-2d/(2d+1)} \left[\lambda^{-1/(2d+1)} \boldsymbol{H}^{-1} \boldsymbol{\Omega} \boldsymbol{H}^{-1} + \lambda^{2d/(2d+1)} \boldsymbol{H}^{-1} \boldsymbol{a}\boldsymbol{a}' \boldsymbol{H}^{-1}\right].$$

Therefore, by definition,

$$MSE = N^{-2d/(2d+1)} Tr \left[\boldsymbol{W}\boldsymbol{H}^{-1} \left(\lambda^{-1/(2d+1)} \boldsymbol{\Omega} + \lambda^{2d/(2d+1)} \boldsymbol{a}\boldsymbol{a}'\right) \boldsymbol{H}^{-1}\right]. \tag{B42}$$

To minimize the MSE, take the differentiation of the right-hand side of (B42) with respect to λ. From the first order condition, we show that MSE is minimized by setting λ to be

$$\lambda^* = [trace(\boldsymbol{W}\boldsymbol{H}^{-1}\boldsymbol{\Omega}\boldsymbol{H}^{-1})]/[trace(2d\boldsymbol{W}\boldsymbol{H}^{-1}\boldsymbol{a}\boldsymbol{a}'\boldsymbol{H}^{-1})]. \tag{B43}$$

By the cyclic property of trace, $\lambda^* = [trace(\boldsymbol{\Omega}\boldsymbol{H}^{-1}\boldsymbol{W}\boldsymbol{H}^{-1})]/(2d\boldsymbol{a}'\boldsymbol{H}^{-1}\boldsymbol{W}\boldsymbol{H}^{-1}\boldsymbol{a})$. Part (b) implies that $N^{d/(2d+1)}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$ has the asymptotic distribution $MVN(-(\lambda^*)^{d/(2d+1)}\boldsymbol{H}^{-1}\boldsymbol{a}, (\lambda^*)^{-1/(2d+1)}\boldsymbol{H}^{-1}\boldsymbol{\Omega}\boldsymbol{H}^{-1})$. □

Proof. (Theorem 4)

Part (a): Applying a Taylor series expansion to $\boldsymbol{t}_N(\boldsymbol{b}_N^S, h_N^*)$ around $\boldsymbol{b}_N^S = \boldsymbol{\beta}$ yields

$$(h_N^*)^{-d}\boldsymbol{t}_N(\boldsymbol{b}_N^S, h_N^*) = (h_N^*)^{-d}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N^*) + [\partial \boldsymbol{t}_N(\boldsymbol{b}_N^*, h_N^*)/\partial \tilde{\boldsymbol{b}}']\,(h_N^*)^{-d}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) \tag{B44}$$

with probability approaching one as N goes to ∞, where $\boldsymbol{b}_N^*$ is between $\boldsymbol{b}_N^S$ and $\boldsymbol{\beta}$. Lemma 8 implies that $(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})/h_N$ converges to zero in probability, which indicates that $(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})/h_N^*$ also converges to zero in probability because h_N^* goes to zero at a slower rate than h_N. By Lemma 9, the probability limit of $[\partial \boldsymbol{t}_N(\boldsymbol{b}_N^*, h_N^*)/\partial \tilde{\boldsymbol{b}}']$ equals $\boldsymbol{H}$. Theorem 3 implies that $(h_N)^{-d}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) = O_p(1)$. We have $(h_N^*)^{-d}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) = o_p(1)$ because h_N^* goes to zero at a slower rate than h_N. Lastly, the probability limit of $[(h_N^*)^{-d}\boldsymbol{t}_N(\boldsymbol{\beta}, h_N^*)]$ is $\boldsymbol{a}$ by Lemma 6(a). We can prove part (a) by taking probability limits of each side of (B44).

Part (b): By Chebyshev's Theorem, it suffices to show that $E(\hat{\boldsymbol{\Omega}}_N) \to \boldsymbol{\Omega}$ and $Var(\hat{\boldsymbol{\Omega}}_N) \to \boldsymbol{0}$. First consider $E(\hat{\boldsymbol{\Omega}}_N)$:

$$E(\hat{\boldsymbol{\Omega}}_N) = h_N E[\boldsymbol{t}_{Nn}(\boldsymbol{b}_N^S, h_N)\boldsymbol{t}_{Nn}(\boldsymbol{b}_N^S, h_N)'] = \boldsymbol{L}_{N1}^* + \boldsymbol{L}_{N2}^* + \boldsymbol{L}_{N3}^*, \tag{B45}$$

where

$$\boldsymbol{L}_{N1}^* \equiv (Nh_N)^{-1}\sum_{n=1}^{N}\sum_{1\le j<k\le J} E\left\{[1(r_{nj} < r_{nk}) + 1(r_{nk} < r_{nj})]\left[K'(\boldsymbol{x}_{njk}'\boldsymbol{b}_N^S/h_N)\right]^2 \tilde{\boldsymbol{x}}_{njk}\tilde{\boldsymbol{x}}_{njk}'\right\},$$

$$\begin{aligned}\boldsymbol{L}_{N2}^* \equiv (Nh_N)^{-1}\sum_{n=1}^{N}\sum_{j<k<l}^{j,k,l\in\mathbb{J}} 2E\{&[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})]\,[1(r_{nj} < r_{nl}) - 1(r_{nl} < r_{nj})] \\ &\cdot K'(\boldsymbol{x}_{njk}'\boldsymbol{b}_N^S/h_N)K'(\boldsymbol{x}_{njl}'\boldsymbol{b}_N^S/h_N)\tilde{\boldsymbol{x}}_{njk}\tilde{\boldsymbol{x}}_{njl}' \\ &+ [1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})]\,[1(r_{nk} < r_{nl}) - 1(r_{nl} < r_{nk})] \\ &\cdot K'(\boldsymbol{x}_{njk}'\boldsymbol{b}_N^S/h_N)K'(\boldsymbol{x}_{nkl}'\boldsymbol{b}_N^S/h_N)\tilde{\boldsymbol{x}}_{njk}\tilde{\boldsymbol{x}}_{nkl}' \\ &+ [1(r_{nj} < r_{nl}) - 1(r_{nl} < r_{nj})]\,[1(r_{nk} < r_{nl}) - 1(r_{nl} < r_{nk})] \\ &\cdot K'(\boldsymbol{x}_{njl}'\boldsymbol{b}_N^S/h_N)K'(\boldsymbol{x}_{nkl}'\boldsymbol{b}_N^S/h_N)\tilde{\boldsymbol{x}}_{njl}\tilde{\boldsymbol{x}}_{nkl}'\Big\},\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{L}_{N3}^* \equiv (Nh_N)^{-1}\sum_{n=1}^{N}\sum_{j<k,l<m,j<l,k\ne l,k\ne m}^{j,k,l,m\in\mathbb{J}} 2E\{&[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})] \\ &\cdot [1(r_{nl} < r_{nm}) - 1(r_{nm} < r_{nl})]\,K'(\boldsymbol{x}_{njk}'\boldsymbol{b}_N^S/h_N)K'(\boldsymbol{x}_{nlm}'\boldsymbol{b}_N^S/h_N)\tilde{\boldsymbol{x}}_{njk}\tilde{\boldsymbol{x}}_{nlm}'\Big\}.\end{aligned}$$

Let $\boldsymbol{\theta}_N \equiv (\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})/h_N$ and $\zeta_{jk} \equiv -v_{-j,k}/h_N + \tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\theta}_N$. By Assumption 1 and the law of iterated expectations, we have

$$\begin{aligned}\boldsymbol{L}_{N1}^* = \sum_{1\le j<k\le J}\int &\left\{2F_{jk}[h_N(-\zeta_{jk} + \tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\theta}_N), \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}] - \bar{F}_{jk}[h_N(-\zeta_{jk} + \tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\theta}_N), \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}]\right\} \\ &\cdot p_{jk}[h_N(-\zeta_{jk} + \tilde{\boldsymbol{x}}_{jk}'\boldsymbol{\theta}_N)|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}]\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}_{jk}'[K'(\zeta_{jk})]^2 d\zeta_{jk} dP(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}).\end{aligned} \tag{B46}$$

By Assumptions 3, 7(a), Condition 2, and Lebesgue's dominated convergence theorem, the right-hand side of (B46) converges to $\boldsymbol{\Omega}$ when $N \to \infty$. Under Assumption 7(b),

$$|\boldsymbol{L}_{N2}^{*}| \leq \sum_{j<k<l}^{j,k,l\in\mathbb{J}} 2Ch_N \Big\{ \int |K'(\zeta_{jk})K'(\zeta_{jl})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}_{jl}'| d\zeta_{jk}d\zeta_{jl}dP(\tilde{\boldsymbol{v}}_{-j,kl},\tilde{\boldsymbol{X}}) + \int |K'(\zeta_{jk})K'(\zeta_{kl})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}_{kl}'| d\zeta_{kj}d\zeta_{kl}dP(\tilde{\boldsymbol{v}}_{-k,jl},\tilde{\boldsymbol{X}}) + \int |K'(\zeta_{jl})K'(\zeta_{kl})\tilde{\boldsymbol{x}}_{jl}\tilde{\boldsymbol{x}}_{kl}'| \, d\zeta_{lj}d\zeta_{lk}dP(\tilde{\boldsymbol{v}}_{-l,jk},\tilde{\boldsymbol{X}}) \Big\}. \tag{B47}$$

Thus, the right-hand side of (B47) converges to zero when N goes to ∞ by Assumption 7 and Condition 2.

$$|\boldsymbol{L}_{N3}^{*}| \leq \sum_{j<k,l<m,j<l,k\neq l,k\neq m}^{j,k,l,m\in\mathbb{J}} 2Ch_N [\int |K'(\zeta_{jk})K'(\zeta_{lm})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}_{lm}'| d\zeta_{jk}d\zeta_{lm}dP(\tilde{\boldsymbol{v}}_{-\{k,m\}},\tilde{\boldsymbol{X}})]$$

under Assumption 7(c). $\boldsymbol{L}_{N3}^{*}$ converges to zero by Assumption 7 and Condition 2. So $E(\hat{\boldsymbol{\Omega}}_N) \to \boldsymbol{\Omega}$ by (B45). Next consider $Var(\hat{\boldsymbol{\Omega}}_N)$. By Assumption 1, we can calculate

$$\begin{aligned} Var(\hat{\boldsymbol{\Omega}}_N) &= \left(h_N^2/N\right) Var\left[\boldsymbol{t}_{Nn}\left(\boldsymbol{b}_N^S, h_N\right)\boldsymbol{t}_{Nn}\left(\boldsymbol{b}_N^S, h_N\right)'\right] \\ &= \left(h_N^2/N\right) E\left\{vec\left[\boldsymbol{t}_{Nn}\left(\boldsymbol{b}_N^S, h_N\right)\boldsymbol{t}_{Nn}\left(\boldsymbol{b}_N^S, h_N\right)'\right]\right. \\ &\quad \left.\cdot\, vec\left[\boldsymbol{t}_{Nn}\left(\boldsymbol{b}_N^S, h_N\right)\boldsymbol{t}_{Nn}\left(\boldsymbol{b}_N^S, h_N\right)'\right]'\right\} + o(1) = (Nh_N^2)^{-1}E\left[\boldsymbol{c}\boldsymbol{c}'\right] + o(1), \end{aligned} \tag{B48}$$

where $\boldsymbol{c} \equiv \sum_{j<k,l<m}^{j,k,l,m\in\mathbb{J}} \boldsymbol{c}_{jklm}$, and

$$\boldsymbol{c}_{jklm} \equiv \left[1(r_j < r_k) - 1(r_k < r_j)\right]\left[1(r_l < r_m) - 1(r_m < r_l)\right] K'(\boldsymbol{x}_{jk}'\boldsymbol{b}_N^S/h_N)K'(\boldsymbol{x}_{lm}'\boldsymbol{b}_N^S/h_N)vec(\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}_{lm}').$$

Following the method of proving $E(\hat{\boldsymbol{\Omega}}_N) \to \boldsymbol{\Omega}$, the right-hand side of (B48) converges to zero under Assumption 7 and Condition 2. Therefore we have proved that $Var(\hat{\boldsymbol{\Omega}}_N) = \boldsymbol{0}$.

Part (c): Lemmas 8 and 9 imply the result of this part. □

References

[1] Abrevaya J and Huang J. 2005. On the bootstrap of the maximum score estimator. *Econometrica* **73**: 1175-1204.

[2] Barberá S and Pattanaik PK. 1986. Falmagne and the rationalizability of stochastic choices in terms of random orderings. *Econometrica* **54**: 707-715.

[3] Beggs S, Cardell S, and Hausman J. 1981. Assessing the potential demand for electric cars. *Journal of Econometrics* **16**: 1-19.

[4] Ben-Akiva M, Morikawa T, and Shiroishi F. 1992. Analysis of the reliability of preference ranking data. *Journal of Business Research* **24**: 149-164.

[5] Calfee J, Winston C, and Stempski R. 2001. Econometric issues in estimating consumer preferences from stated preference data: A case study of the value of automobile travel time. *Review of Economics and Statistics* **83**: 699-707.

[6] Caparrós A, Oviedo JL, and Campos P. 2008. Would you choose your preferred option? Comparing choice and recoded ranking experiments. *American Journal of Agricultural Economics* **90**: 843-855.

[7] Cavanagh CL. 1987. Limiting behavior of estimators defined by optimization. *Unpublished manuscript*, Department of Economics, Harvard University, Cambridge, MA.

[8] Chapman RG and Staelin R. 1982. Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research* **19**: 288-301.

[9] Conte A, Hey JD, and Moffatt PG. 2011. Mixture models of choice under risk. *Journal of Econometrics* **162**: 79-88.

[10] Dagsvik JK and Liu G. 2009. A framework for analyzing rank ordered data with application to automobile demand. *Transportation Research Part A* **43**: 1-12.

[11] Delgado MA, Rodríguez-Poo JM, and Wolf M. 2001. Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator. *Economics Letters* **73**: 241-250.

[12] Falmagne JC. 1978. A representation theorem for finite random scale systems. *Journal of Mathematical Psychology* **18**: 52-72.

[13] Fiebig DG, Keane MP, Louviere J, and Wasi N. 2010. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Market Science* **29**: 393-421.

[14] Fox JT. 2007. Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *RAND Journal of Economics* **38**: 1002-1019.

[15] Goeree JK, Holt CA, and Palfrey TR. 2005. Regular quantal response equilibrium. *Experimental Economics* **8**: 347-367.

[16] Greene WH, Hensher DA, and Rose J. 2006. Accounting for heterogeneity in the variance of unobserved effects in mixed logit models. *Transportation Research Part B* **40**: 75-92.

[17] Han AK. 1987. Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics* **35**: 303-316.

[18] Harrison GW and Rutström EE. 2009. Expected utility theory and prospect theory: one wedding and a decent funeral. *Experimental Economics* **12**: 133-158.

[19] Hausman JA and Ruud PA. 1987. Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics* **34**: 83-104.

[20] Hensher D, Louviere J, and Swait J. 1999. Combining sources of preference data. *Journal of Econometrics* **89**: 197-221.

[21] Horowitz JL. 1992. A smoothed maximum score estimator for the binary response model. *Econometrica* **60**: 505-531.

[22] Kim J and Pollard D. 1990. Cube root asymptotics. *Annals of Statistics* **18**: 191-219.

[23] Klein RW and Spady RH. 1993. An efficient semiparametric estimator for binary response models. *Econometrica* **61**: 387-421.

[24] Layton DF. 2000. Random coefficient models for stated preference surveys. *Journal of Environmental Economics and Management* **40**: 21-36.

[25] Layton DF and Levine RA. 2003. How much does the far future matter? A hierarchical Bayesian analysis of the public's willingness to mitigate ecological impacts of climate change. *Journal of the American Statistical Association* **98**: 533-544.

[26] Lee L. 1995. Semiparametric maximum likelihood estimation of polychotomous and sequential choice models. *Journal of Econometrics* **65**: 381-428.

[27] Lewbel A. 2000. Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics* **97**: 145-177.

[28] Manski CF. 1975. Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* **3**: 205-228.

[29] Manski CF. 1985. Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics* **27**: 313-333.

[30] McFadden D. 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka P. (Ed.), *Frontiers in Econometrics*. Academic Press: New York, pp.105-142.

[31] McFadden D. 1986. The choice theory approach to market research. *Marketing Science* **5**: 275-297.

[32] Newey WK. 1986. Linear instrumental variable estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* **32**: 127-141.

[33] Newey WK and McFadden D. 1994. Large sample estimation and hypothesis testing. *Handbook of Econometrics* Vol **4**: 2111-2245.

[34] Oviedo JL and Yoo H. 2016. A latent class nested logit model for rank-ordered data with application to cork oak reforestation. *Environmental and Resource Economics*. DOI: `10.1007/s10640-016-0058-7`.

[35] Ruud PA. 1983. Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models. *Econometrica* **51**: 225-228.

[36] Ruud PA. 1986. Consistent estimation of limited dependent variable models despite misspecification of distribution. *Journal of Econometrics* **32**: 157-187.

[37] Scarpa R, Notaro S, Louviere J, and Raffaelli R. 2011. Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in alpine grazing commons. *American Journal of Agricultural Economics* **93**: 813-828.

[38] Sherman RP. 1993. The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61**: 123-137.

[39] Small KA, Winston C, and Yan J. 2005. Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica* **73**: 1367-1382.

[40] Storn R and Price K. 1997. Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* **11**: 341-359.

[41] Train KE and Winston C. 2007. Vehicle choice behavior and the declining market share of U.S. automakers. *International Economic Review* **48**: 1469-1496.

[42] Yan J. 2012. A smoothed maximum score estimator for multinomial discrete choice models. *Working paper*.

[43] Yan J and Yoo H. 2014. The seeming unreliability of rank-ordered data as a consequence of model misspecification. MPRA Paper No. 56285. http://mpra.ub.uni-muenchen.de/56285/

[44] Yoo H and Doiron D. 2013. The use of alternative preference elicitation methods in complex discrete choice experiments. *Journal of Health Economics* **32**: 1166-1179.