Accepted Manuscript

Semiparametric estimation of the random utility model with rank-ordered choice data

Jin Yan, Hong Il Yoo

PII: S0304-4076(19)30052-1
DOI: https://doi.org/10.1016/j.jeconom.2019.03.003
Reference: ECONOM 4650

To appear in: *Journal of Econometrics*

Received date : 24 February 2018
Revised date : 1 December 2018
Accepted date : 15 March 2019

Please cite this article as: J. Yan and H.I. Yoo, Semiparametric estimation of the random utility model with rank-ordered choice data. *Journal of Econometrics* (2019), https://doi.org/10.1016/j.jeconom.2019.03.003

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

Jin Yan*
The Chinese University of Hong Kong

Hong Il Yoo†
Durham University

March 22, 2019

Abstract

We propose semiparametric methods for estimating random utility models using rank-ordered choice data. Our primary method is the generalized maximum score (GMS) estimator. With partially rank-ordered data, the GMS estimator allows for arbitrary forms of interpersonal heteroskedasticity. With fully rank-ordered data, the GMS estimator becomes considerably more flexible, allowing for random coefficients and alternative-specific heteroskedasticity and correlations. The GMS estimator has a non-standard asymptotic distribution and a convergence rate of $N^{-1/3}$. We proceed to construct its smoothed version which is asymptotically normal with a faster convergence rate of $N^{-d/(2d+1)}$, where $d \geq 2$ increases in the strength of smoothness assumptions.

Keywords: Random utility, rank-ordered, discrete choice, semiparametric estimation, smoothing.
JEL classification: C14, C35

*Corresponding author. Department of Economics, The Chinese University of Hong Kong, Esther Lee Building, Chung Chi Road, Ma Liu Shui, New Territories, Hong Kong. Email: jyan@cuhk.edu.hk.
†Durham University Business School, Durham University, Mill Hill Lane, Durham DH1 3LB, United Kingdom. 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 on the individual's preference ordering such as her second and third preferences. 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 the resulting logistic advantages could be substantial for non-market valuation studies which typically involve a narrowly defined population of interest (Scarpa *et al.* 2011).

We develop semiparametric methods for 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 (utility index) varying with finite-dimensional explanatory variables, and an additive stochastic component (error term). The objective is to estimate preference parameters, referring to coefficients on the explanatory variables. The methods are semiparametric in that they maintain the usual parametric form of the systematic component but place only nonparametric 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 maximum (simulated) likelihood estimation of one of four models: multinomial logit (MNL), nested MNL, multinomial probit (MNP), and random coefficient or "mixed" MNL. Each model assumes a different parametric distribution of the stochastic component, and has its own rank-ordered choice counterpart that 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), Fox (2007), Bajari, Fox and Ryan (2008), and Yan (2013).[1] 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), Sherman (1993), and Cavanagh and Sherman (1998), to name a few. When it comes to rank-ordered choice data, we are aware of only one study that aims at semiparametric estimation of preference parameters, namely Hausman and Ruud (1987). In their study, the weighted M-estimator (WME) of Ruud (1986) is generalized for use with rank-ordered choice data, whereas the original WME was intended for use with multinomial choice data. The generalized WME imposes independence between the explanatory variables and the error terms, ruling out heteroskedasticity across individuals. Though the generalized WME allows consistent estimation under nonparametric stochastic specification, this consistency is confined to the ratios of the coefficients on continuous explanatory variables and the estimator's 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. The primary method that we develop is the generalized maximum score (GMS) estimator. Unlike the generalized WME, the GMS estimator does not require independence between the explanatory variables and the error terms, and can accommodate flexible forms of interpersonal heteroskedasticity. We also show that the GMS estimator is consistent under more general assumptions concerning the explanatory variables than the generalized WME. Roughly speaking, if one of q explanatory variables is continuous, the GMS estimator allows consistent estimation of the ratios of all coefficients regardless of whether the other $q-1$ variables are continuous or discrete. Like the maximum score (MS) estimator of Manski (1985) that it nests as a special case, the GMS estimator has a slow convergence rate of $N^{-1/3}$ and a non-standard asymptotic distribution. One way to lessen these drawbacks is to introduce extra regular conditions and apply Horowitz's (1992) technique to construct a smoothed version of the GMS estimator. We show that the smoothed GMS (SGMS) estimator achieves a faster convergence rate of $N^{-d/(2d+1)}$, where integer $d \geq 2$ increases in the strength of the smoothness conditions presented in Section 3.1, and possesses a normal limiting distribution with a covariance matrix that can be consistently estimated.

The GMS estimator generalizes the pairwise MS estimator that Fox (2007) has developed for a semiparametric analysis of multinomial choice data. When the individual faces J alternatives, a

[1] Bajari, Fox and Ryan (2008) stands out from other studies in this list, since their objective is to estimate a multinomial choice model in an environment where the econometrician does not observe multinomial choices made by individuals; instead, the econometrician observes aggregated data on sales rankings of alternative products across different markets. This feature poses some challenges for taxonomy. We agree with Fox (2007, p.1004) on classifying their estimator as a multinomial choice method, considering that the behavioral model used in their proofs is a multinomial choice model.

multinomial choice observation allows the econometrician to infer the outcomes of $J-1$ pairwise comparisons where each pair comprises the individual's actual choice and an unchosen alternative. A rank-ordered choice observation provides information that is needed to infer the outcomes of other pairwise comparisons; for example, in case the individual ranks all J alternatives from best to worst, her rank-ordered choice allows the econometrician to infer the outcomes of all possible $J(J-1)/2$ pairwise comparisons. The GMS estimator extends the MS estimator by incorporating this type of extra information, which could come from data on partial rankings (*e.g.*, the individual reports her best and second best out of five alternatives) as well as complete rankings.

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.[2] 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. 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 consistent parametric estimation of a mixture model is extremely difficult, because it demands the exact knowledge of the number and specifications of component models.

The GMS estimator becomes considerably more flexible than the MS estimator when each individual completely ranks all alternatives in her choice set from best to worst. As we discuss in details in Section 2.3, the GMS estimator on complete rankings is consistent for all popular parametric models exhibiting flexible substitution patterns, whereas the MS estimator is not.[3] Thus, the GMS estimator more closely satisfies what an empiricist 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.[4] This is an interesting finding because in the parametric framework, the advantage

[2]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 MRC estimator does pairwise comparisons *between* individuals and requires the error terms to be independent of the explanatory variables, ruling out the possibility of interpersonal heteroskedasticity.

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

[4]When it comes to assumptions on explanatory variables that are needed for the point identification of utility

of using rank-ordered choice data instead of multinomial choice data is limited to efficiency gains (Hausman and Ruud, 1987; Beresteanu and Zincenko, 2018) and a multinomial choice model may be more robust to stochastic misspecification than its rank-ordered choice counterpart (Yan and Yoo, 2014). This kind of efficiency-bias tradeoff does not apply to the comparison of the GMS estimator on complete rankings to the MS estimator on multinomial choices: the GMS estimator is more efficient as indicated by smaller root mean square errors (RMSE) in Monte Carlo simulations (Section 4), and is also robust to a wider variety of DGPs.

As noted earlier, the GMS estimator also inherits less attractive properties of the MS estimator, such as the convergence rate of $N^{-1/3}$ and the non-standard asymptotic distribution of Cavanagh (1987) and Kim and Pollard (1990). Horowitz (1992) develops the smoothed MS (SMS) estimator that addresses these drawbacks in the context of Manski's (1985) MS estimator of binomial choice models. Yan (2013) extends the results to Fox's (2007) MS estimator of multinomial choice models. The SGMS estimator of rank-ordered choice models that we propose builds on this tradition.

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 and states its asymptotic properties. Section 4 presents the Monte Carlo evidence on the finite sample performance of the proposed estimators. Section 5 concludes. Proofs of Theorems 1-3 are provided in Appendices and those of Theorems 4-5 are included in Supplementary Material.

Throughout this paper, we will maintain the following notations. We write scalars in lightface, vectors in lowercase bold, and matrices in uppercase bold. All vectors are column vectors. $\mathbb{R}^q$ is a q-dimensional Euclidean space, $\mathbb{B}$ is a subset of $\mathbb{R}^q$, and other blackboard bold letters such as $\mathbb{J}$ and $\mathbb{M}$ refer to finite sets. We reserve letters j, k and l for indexing alternatives, and letter n for indexing individuals or observations. Vector $\boldsymbol{x}_{jk}$ denotes the difference between two vectors $\boldsymbol{x}_j$ and $\boldsymbol{x}_k$. The first element of $\boldsymbol{x}_j$ ($\boldsymbol{x}_{jk}$) is denoted by $x_{j,1}$ ($x_{jk,1}$), and the subvector comprising its remaining elements is denoted by $\tilde{\boldsymbol{x}}_j$ ($\tilde{\boldsymbol{x}}_{jk}$). Where the distinction needs emphasis, we use $\boldsymbol{x}_{nj}$ ($\boldsymbol{x}_{njk}$) to denote the nth observation of random vector $\boldsymbol{x}_j$ ($\boldsymbol{x}_{jk}$). Letters P and E denote a probability and an expectation, respectively. Function $F(\cdot)$ denotes a cumulative distribution function (CDF), and function $F(\cdot\,|\,\cdot)$ denotes a conditional CDF. The i^{th} derivative of function $K(\cdot)$ is denoted by $K^{(i)}(\cdot)$. Function $1(\cdot)$ is an indicator function that equals one when the event in the brackets is true, and zero otherwise. Symbols $\setminus$, $'$, $\Rightarrow$, and $\xrightarrow{p}$ denote a set difference, matrix transposition, convergence in distribution, and convergence in probability, respectively.

---

coefficients, semiparametric methods are more restrictive than parametric methods and the GMS estimator is no exception. In this respect, the GMS estimator is as restrictive as the MS estimator, and requires the presence of a continuous explanatory variable with large support. See Assumption 3 in Section 2.2.

2 The Model and the Generalized Maximum Score Estimator

2.1 A Random Utility Framework and Rank-Ordered Choice Data

Consider a 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 of covariates, $\boldsymbol{\beta} \equiv (\beta_1, \ldots, \beta_q)' \in \mathbb{R}^q$ is the preference parameter vector of interest, and ε_j is the unobserved component of utility to the econometrician. Let $\boldsymbol{X} \equiv (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_J)' \in \mathbb{R}^{J\times q}$ be the matrix of the covariates and $\boldsymbol{\varepsilon} \equiv (\varepsilon_1, \ldots, \varepsilon_J)' \in \mathbb{R}^J$ be the vector of the error terms. 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.

The random utility function (1) can accommodate both alternative-specific and individual-specific covariates. To see this point, consider a utility function that spells out the distinction the two types of covariates explicitly

$$u_j = \boldsymbol{z}_j'\boldsymbol{\gamma} + \boldsymbol{s}'\boldsymbol{\alpha}_j + \varepsilon_j \quad \forall j \in \mathbb{J}, \tag{2}$$

where q_1-vector $\boldsymbol{z}_j$ includes covariates that vary over alternatives (*e.g.*, product attributes), and q_2-vector $\boldsymbol{s}$ includes a constant term as well as covariates that vary across individuals but not over alternatives (*e.g.*, person's age). Without loss of generality, we set $\boldsymbol{\alpha}_1 = \mathbf{0}_{q_2}$ for location normalization, where $\mathbf{0}_{q_2}$ denotes a q_2-vector of zeros. Following Cameron and Trivedi (2005, p.498), equation (2) can be compactly written in the form of equation (1) as follows. Let $\boldsymbol{\alpha}$ denote a vector that collects alternative-specific parameter vectors, $\boldsymbol{\alpha} \equiv (\mathbf{0}_{q_2}', \boldsymbol{\alpha}_2', \ldots, \boldsymbol{\alpha}_J')' \in \mathbb{R}^{Jq_2}$. Next, let $\boldsymbol{s}_j$ denote a conformable vector that is partitioned into J blocks, where the j^{th} block is $\boldsymbol{s} \in \mathbb{R}^{q_2}$ and each of the remaining $J-1$ blocks is $\mathbf{0}_{q_2}$. For example, $\boldsymbol{s}_1 \equiv (\boldsymbol{s}', \mathbf{0}_{q_2}', \ldots, \mathbf{0}_{q_2}')' \in \mathbb{R}^{Jq_2}$, $\boldsymbol{s}_2 \equiv (\mathbf{0}_{q_2}', \boldsymbol{s}', \ldots, \mathbf{0}_{q_2}')' \in \mathbb{R}^{Jq_2}$ and so on. Then, it follows that $\boldsymbol{s}'\boldsymbol{\alpha}_j = \boldsymbol{s}_j'\boldsymbol{\alpha}$, and equation (1) is obtained by defining $\boldsymbol{x}_j \equiv (\boldsymbol{z}_j', \boldsymbol{s}_j')' \in \mathbb{R}^q$ and $\boldsymbol{\beta} \equiv (\boldsymbol{\gamma}', \boldsymbol{\alpha}')' \in \mathbb{R}^q$, where $q = q_1 + Jq_2$.

Thus, without loss of generality, our subsequent discussion focuses on equation (1). Let $r(j, \boldsymbol{u})$ denote the latent (or potentially unobserved) ranking of alternative j, based on the underlying utility vector $\boldsymbol{u} \equiv (u_1, u_2, \ldots, u_J)' \in \mathbb{R}^J$. We shall follow the notational convention that $r(j, \boldsymbol{u}) = m$ when j is the m^{th} best alternative in the choice set $\mathbb{J}$, *i.e.*, a smaller ranking value indicates a more preferred alternative. A more formal definition of the latent ranking is

$$r(j, \boldsymbol{u}) \equiv 1 + \sum_{k=1}^{J} 1(u_j < u_k) \tag{3}$$

for any $j \in \mathbb{J}$. For instance, suppose that $J = 4$ and $u_3 > u_4 > u_1 > u_2$. Then, $r(3, \boldsymbol{u}) = 1$, $r(4, \boldsymbol{u}) = 2$, $r(1, \boldsymbol{u}) = 3$, and $r(2, \boldsymbol{u}) = 4$. There is a one-to-one mapping between the choice set $\{1, \ldots, J\}$ and the latent ranking set $\{r(j, \boldsymbol{u}) : j = 1, \ldots, J\}$ by definition (3).[5]

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 only a partial ranking. To facilitate further discussion, suppose that the individual reports the ranking of her best M alternatives where $1 \leq M < J-1$, 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, $\mathbb{M} \equiv \{j : r(j, \boldsymbol{u}) \leq M\}$.[6] The reported ranking of alternative j, then, follows the observation

[5] We ignore utility ties here because they happen with probability zero under the assumptions we impose later for point identification.

[6] Like the popular parametric methods that we will review in Section 2.3, our semiparametric method allows both the choice set $\mathbb{J} = \{1, 2, \ldots, J\}$ and the dimension of the subset $\mathbb{M} \subset \mathbb{J}$, and hence J and M, to vary across individuals. For example, person a may face choice set $\mathbb{J} = \{1, 2, 3, 4, 5\}$, and report his first and second-best alternatives as alternatives 2 and 3 ($\mathbb{M} = \{2, 3\}$), respectively: in his case, $J = 5$ and $M = 2$. Person b, on the other hand, may face $\mathbb{J} = \{1, 2, 3, 4\}$ and report a complete ranking on it, *e.g.*, her first, second-best, and third-best alternatives are

rule

$$r_j = \begin{cases} r(j,\boldsymbol{u}) & if \quad r(j,\boldsymbol{u}) \leq M, \text{ or equivalently, } j \in \mathbb{M}, \\ M+1 & if \quad r(j,\boldsymbol{u}) > M, \text{ or equivalently, } j \in \mathbb{J} \setminus \mathbb{M}. \end{cases} \tag{4}$$

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 $1 < M < J - 1$ and $J > 3$, are much less common in empirical studies though not unprecedented.[7]

2.2 Identification and the Generalized Maximum Score Estimator

This section introduces identification conditions for the preference parameter vector $\boldsymbol{\beta}$ and the primary method that we propose, the Generalized Maximum Score (GMS) estimator. The GMS estimator is semiparametric in the sense that it allows the econometrician to estimate $\boldsymbol{\beta}$ consistently, without committing to a specific parametric form of the conditional distribution of the error vector given observed attributes $\boldsymbol{\varepsilon}|\boldsymbol{X}$.

The first assumption presents a key condition pertaining to our identification strategy. This assumption implicitly places a restriction on the conditional distribution of $\boldsymbol{\varepsilon}|\boldsymbol{X}$, albeit it is a nonparametric restriction 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 $j \in \mathbb{J}$.

Assumption 1. *For any pair of alternatives $j, k \in \mathbb{J}$ and for almost every $\boldsymbol{X} \in \mathbb{R}^{J \times q}$,*

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

$$P(r_j < r_k|\boldsymbol{X}) > P(r_k < r_j|\boldsymbol{X}), \tag{5}$$

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

alternatives 1, 2, and 4 ($\mathbb{M} = \{1, 2, 4\}$), respectively: in her case, $J = 4$ and $M = 3$. Our proofs can be modified to accommodate this generality explicitly, though we do not pursue it to avoid carrying around individual subscripts. Note that when J and M are considered individual-specific, complete rankings data in our subsequent discussion refer to the case where $M = J - 1$ for all individuals, and partial rankings data refer to the case where $M < J - 1$ for at least one individual.

[7] 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.

all covariates.

Assumption 1 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

$$P(r_j < r_k|\boldsymbol{X}) \;=\; P(r_j = 1|\boldsymbol{X}). \tag{6}$$

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 (5), then Assumption 1 becomes the monotonicity property of choice probabilities (Manski, 1975; Fox, 2007), *i.e.*, the ranking of the choice probability of an alternative is the same as the ranking of the systematic utility of that alternative for any given individual.[8]

Second, when $M = J - 1$, the individual reports 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 the utility from choosing alternative k, so we have

$$\begin{aligned} P(r_j < r_k|\boldsymbol{X}) &= P(u_j > u_k|\boldsymbol{X}) \\ &= P(\varepsilon_k - \varepsilon_j < v_j - v_k|\boldsymbol{X}). \end{aligned} \tag{7}$$

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 (7) may not hold if we observe only a partial ranking, *i.e.*, $M < J - 1$. This is because while $r_j < r_k$ naturally implies $u_j > u_k$, $u_j > u_k$ may not imply $r_j < r_k$; when neither alternative j nor alternative k belongs to set $\mathbb{M}$, which includes the best M alternatives, both alternatives j and k are observed with the same ranking, $M + 1$, even if $u_j > u_k$.

For any pair of alternatives, assume that the distribution of $\varepsilon_k - \varepsilon_j$ conditional on the explanatory vectors 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

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

1 when a complete ranking of J alternatives is available. The proof is straightforward.[9] 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 1 implies that $v_j - v_k = 0$ if and only if $P(r_j < r_k|\boldsymbol{X}) = 1/2$, or equivalently, $P(\varepsilon_k - \varepsilon_j < v_j - v_k|\boldsymbol{X}) = 1/2$ by (7). 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 (7), or equivalently, $P(r_j < r_k|\boldsymbol{X}) > P(r_k < r_j|\boldsymbol{X})$.

Our second assumption is about scale normalization and the parameter space. As usual in discrete choice modeling, identification of the preference vector $\boldsymbol{\beta}$ requires scale normalization since they are unique only up to a scale.[10] When a parametric form of the conditional distribution of $\boldsymbol{\varepsilon}|\boldsymbol{X}$ is specified, it is a nearly universal practice to normalize a scale parameter of that distribution to achieve identification.[11] But when no parametric form is specified, no scale parameter is available for normalization. In the semiparametric framework, identification is therefore achieved by normalizing the preference parameter vector $\boldsymbol{\beta}$ instead. Subject to the prior knowledge that some element of vector $\boldsymbol{\beta}$ is non-zero, we can normalize the magnitude of that element.[12] Without loss of generality, we assume that the first element of $\boldsymbol{\beta}$ has absolute value one, *i.e.*, $|\beta_1| = 1$. Let $\tilde{\boldsymbol{\beta}} \equiv (\beta_2, \ldots, \beta_q)' \in \mathbb{R}^{q-1}$ be the vector containing the other elements of $\boldsymbol{\beta}$.

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

Next we formally define the point identification for $\boldsymbol{\beta} \in \mathbb{B}$.

Definition 1. *For any vector $\boldsymbol{b} \in \mathbb{B}$, define function*

$$Q^*(\boldsymbol{b}) = \sum_{1 \leq j < k \leq J} E[1(r_j < r_k) \cdot 1(\boldsymbol{x}_j'\boldsymbol{b} > \boldsymbol{x}_k'\boldsymbol{b}) + 1(r_k < r_j) \cdot 1(\boldsymbol{x}_k'\boldsymbol{b} > \boldsymbol{x}_j'\boldsymbol{b})]. \tag{8}$$

The parameter vector $\boldsymbol{\beta}$ is point identified if $Q^*(\boldsymbol{\beta}) > Q^*(\boldsymbol{b})$ for any $\boldsymbol{b} \in \mathbb{B}$ and $\boldsymbol{b} \neq \boldsymbol{\beta}$.

Identification requires $\boldsymbol{\beta}$ to be the unique maximizer of function $Q^*(\boldsymbol{b})$ for $\boldsymbol{b} \in \mathbb{B}$. Assumption 1 guarantees that $\boldsymbol{\beta}$ maximizes $Q^*(\boldsymbol{b})$ in the parameter space, which will be shown in Theorem 1 later. However, if all the covariates in (8) are discrete, then we can always find another vector $\boldsymbol{b}$ in the

[9] This proof does not apply to partially rank-ordered choice data, of which multinomial choice data is a special case, because the first equality in (7) 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.

[10] Multiplying both the preference parameter vector $\boldsymbol{\beta}$ and the error term $\boldsymbol{\varepsilon}$ by any positive constant leads to the same rank-ordered choice data.

[11] For instance, in the binomial probit model, the variance of the conditional distribution is assumed to be one.

[12] For example, economists may agree that the coefficient on the own price variable is negative.

neighborhood of $\boldsymbol{\beta}$ such that $\boldsymbol{b}$ generates the same ranking of utility indexes as $\boldsymbol{\beta}$ does, and consequently, $Q^*(\boldsymbol{b}) = Q^*(\boldsymbol{\beta})$. To achieve point identification, we need to impose an extra assumption on the covariates, namely, we need a covariate that is continuous conditional on the other covariates. Recall that $\boldsymbol{x}_{jk} \equiv \boldsymbol{x}_j - \boldsymbol{x}_k \in \mathbb{R}^q$, $x_{jk,1}$ is the first element of $\boldsymbol{x}_{jk}$, and $\tilde{\boldsymbol{x}}_{jk} \equiv (x_{jk,2}, \ldots, x_{jk,q})' \in \mathbb{R}^{q-1}$ refers to the remainder. Our third assumption states the continuity requirement on the covariates for point identification.

Assumption 3. *The following statements are true.*

(a) *For any pair of distinct alternatives $j, k \in \mathbb{J}$, the probability 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 3 is essential for the uniqueness of $\boldsymbol{\beta}$ as the maximizer of $Q^*(\boldsymbol{b})$ for $\boldsymbol{b} \in \mathbb{B}$. Assumption 3(a) avoids the local failure of identification, which is not required by parametric models but important in semiparametric settings. In other words, the semiparametric models relax restrictions on the error distribution at the cost of imposing continuity conditions on the covariates. Assumption 3(b) is analogous to the full-rank condition for the binomial choice model, which prevents the global failure of identification.

The following theorem establishes point identification; the proof is available in Appendix A.

Theorem 1. *Let Assumptions* 1-3 *hold. The parameter vector $\boldsymbol{\beta}$ is point identified by Definition* 1.

Next, we describe the intuition behind Theorem 1. Let $\boldsymbol{b} \equiv (b_1, \tilde{\boldsymbol{b}}')'$ be any vector in the parameter space $\mathbb{B}$. Under Assumption 1, 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 of being true as event $r_k < r_j$. Therefore, the expected value of the following match

$$\begin{aligned} m_{jk}(\boldsymbol{b}) &= 1(r_j < r_k) \cdot 1(\boldsymbol{x}_j'\boldsymbol{b} > \boldsymbol{x}_k'\boldsymbol{b}) + 1(r_k < r_j) \cdot 1(\boldsymbol{x}_k'\boldsymbol{b} > \boldsymbol{x}_j'\boldsymbol{b}) + 1(r_j < r_k) \cdot 1(\boldsymbol{x}_j'\boldsymbol{b} = \boldsymbol{x}_k'\boldsymbol{b}) \\ &= 1(r_j < r_k) \cdot 1(\boldsymbol{x}_j'\boldsymbol{b} \geq \boldsymbol{x}_k'\boldsymbol{b}) + 1(r_k < r_j) \cdot 1(\boldsymbol{x}_k'\boldsymbol{b} > \boldsymbol{x}_j'\boldsymbol{b}) \end{aligned} \tag{9}$$

should be maximized at the true preference parameter vector $\boldsymbol{\beta}$ over $\boldsymbol{b} \in \mathbb{B}$. Since

$$Q^*(\boldsymbol{b}) = \sum_{1 \leq j < k \leq J} E[m_{jk}(\boldsymbol{b})] \tag{10}$$

by (8) and (9), function $Q^*(\boldsymbol{b})$ is also maximized at $\boldsymbol{\beta}$. Assumption 2 (scale normalization) and Assumption 3 (regularity conditions on covariates) guarantee that $\boldsymbol{\beta}$ uniquely maximizes $Q^*(\boldsymbol{b})$ over $\boldsymbol{b} \in \mathbb{B}$.

Our fourth assumption pertains to sampling. Matrix $\boldsymbol{X}_n$ and vectors $\boldsymbol{r}_n$ and $\boldsymbol{\varepsilon}_n$ are the n^{th} observation of matrix $\boldsymbol{X}$ and vectors $\boldsymbol{r}$ and $\boldsymbol{\varepsilon}$, respectively.

Assumption 4. *$\{(\boldsymbol{r}_n, \boldsymbol{X}_n, \boldsymbol{\varepsilon}_n) : n = 1, \dots, N\}$ is a random sample of $(\boldsymbol{r}, \boldsymbol{X}, \boldsymbol{\varepsilon})$, where $\boldsymbol{r}_n \equiv (r_{n1}, \dots, r_{nJ})' \in \mathbb{N}^J$, $\boldsymbol{X}_n \equiv (\boldsymbol{x}_{n1}, \dots, \boldsymbol{x}_{nJ})' \in \mathbb{R}^{J \times q}$, and $\boldsymbol{\varepsilon}_n \equiv (\varepsilon_{n1}, \dots, \varepsilon_{nJ})' \in \mathbb{R}^J$. For each individual $n = 1, \dots, N$, $(\boldsymbol{r}_n, \boldsymbol{X}_n)$ is observed.*

Assumption 4 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.

Next, we describe the intuition behind applying Theorem 1 (Identification) and Assumption 4 (Random Sampling) to construct the GMS estimator. Define $\boldsymbol{x}'_{nj}\boldsymbol{b}$ as the $\boldsymbol{b}$-utility index 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:

$$\boldsymbol{b}_N \in \underset{\boldsymbol{b} \in \mathbb{B}}{argmax}\, Q_N(\boldsymbol{b}), \tag{11}$$

where

$$Q_N(\boldsymbol{b}) \equiv N^{-1} \sum_{n=1}^{N} \left\{ \sum_{1 \le j < k \le J} \left[1(r_{nj} < r_{nk}) \cdot 1(\boldsymbol{x}'_{nj}\boldsymbol{b} \ge \boldsymbol{x}'_{nk}\boldsymbol{b}) + 1(r_{nk} < r_{nj}) \cdot 1(\boldsymbol{x}'_{nk}\boldsymbol{b} > \boldsymbol{x}'_{nj}\boldsymbol{b}) \right] \right\} \tag{12}$$

can be viewed as the sample analog of $Q^*(\boldsymbol{b})$ defined by (8). In the special case of $M = 1$, *i.e.*, when we have multinomial choice data, the estimator $\boldsymbol{b}_N$ defined by (11) 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, $\boldsymbol{b}_N$ is called the generalized maximum score (GMS) estimator.

When complete rankings of three or more alternatives are observed ($J \ge 3$ and $M = J - 1$), the inner sum inside the curly brackets in (12) is an increasing function of Kendall's rank correlation between observed rankings and estimated utility indexes across $J(J-1)/2$ alternative pairs within individual n. In this situation, the GMS estimator may be interpreted as an estimator that maximizes the sample mean of within-individual rank correlation.[13] Note that the maximum rank correlation (MRC) estimator of Han (1987) and the rank estimator of Cavanagh and Sherman (1998) are substantively different from ours, both in regards to the models of interest and the maximands. Their semiparametric estimators are for single-equation index models, which include binary choice models ($J = 2$) but not more general types of multinomial choice and rank-ordered choice models ($J \geq 3$).[14] In addition, within-individual rank correlation across alternative pairs is an irrelevant concept for single-equation index models, and what the MRC (rank) estimator maximizes is Kendall's (Spearman's) rank correlation between a dependent variable and an estimated index across $N(N-1)/2$ pairs of individuals in the sample.[15]

Again in the same situation ($J \geq 3$ and $M = J - 1$), $Q_N(\boldsymbol{b})$ is algebraically identical to the objective function of Bajari, Fox and Ryan (2008) at first glance, but the setup of their econometric analysis is quite different from ours. Rankings in their analysis are the aggregate sales rankings of alternative products offered by the same supplier in a specific market, instead of individual-level preference orderings that we consider. Their objective is to estimate a random utility model describing individual-level multinomial choices (that is, $J \geq 3$ and $M = 1$ in our notation), in an environment where the econometrician observes the aggregate sales rankings instead of the individual-level choices. They show that when the error terms are *i.i.d.* over alternatives within individuals, a semiparametric estimator of the multinomial choice model can be constructed using a score function that incorporates all pairwise comparisons of the aggregate sales rankings. In comparison, the GMS estimator with complete rankings ($J \geq 3$ and $M = J - 1$) can accommodate more flexible error structures that satisfy the pairwise ZCM (discussed in Section 2.3), thereby allowing for flexible patterns of heteroskedasticity and correlation over alternatives as well as random coefficients across individuals.

The following theorem establishes the strong consistency of the GMS estimator.

Theorem 2. *Let Assumptions* 1-4 *hold. The GMS estimator* $\boldsymbol{b}_N$ *defined in* (11) *converges almost surely to the true preference parameter vector,* $\boldsymbol{\beta}$, *in the data generating process.*

[13] Let $m_n(\boldsymbol{b})$ denote inner sum inside the curly brackets in (12) for individual n, then Kendall's rank correlation between observed rankings r_{nj} and utility indexes $\boldsymbol{x}'_{nj}\boldsymbol{b}$, where $j = 1, \ldots, J$, equals $[2m_n(\boldsymbol{b}) - 1] \times [J(J-1)/2]^{-1}$ for this individual. Clearly, $Q_N(\boldsymbol{b})$ is the sample mean of $m_n(\boldsymbol{b})$, where $n = 1, \ldots, N$, and hence is an increasing function of the sample mean of within-individual Kendall's rank correlation.

[14] Single-equation index models include, *inter alia*, Tobit, binary probit, ordered probit, and univariate duration models; the assumed data generating process involves a single latent dependent variable. In comparison, the random utility model for multinomial and rank-ordered choice data can be viewed as a system of $J - 1$ latent dependent variables where each variable is the utility difference between alternative j and alternative J for $j = 1, 2, ..., J - 1$.

[15] More precisely, the rank estimator of Cavanagh and Sherman (1998) is a class of related estimators, of which one that maximizes Spearman's rank correlation is a special case.

2.3 Comparisons with Parametric Methods

From the empiricist's perspective, the question of paramount interest may be how flexible the semiparametric model that the GMS estimator accommodates is relative to parametric models that one may consider. Modern desktop computing power makes this question especially relevant. Standard computing resources of today can handle estimation of models that feature fairly flexible, albeit parametric, error structures. Most semiparametric methods for discrete choice data relax parametric restrictions on error structures at the price of regularity conditions on explanatory variables that parametric methods do not require, and the GMS estimator is no exception. This section maintains that such conditions hold, which have been stated as Assumption 3(a) in the context of the GMS estimator.

When applied to data on complete rankings, *i.e.* $M = J - 1$, the GMS estimator postulates a semiparametric model that can nest all popular parametric models and any finite mixture of those models. In most studies on rank-ordered choices, the complete rankings are elicited as needed for this result.[16] 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 equicorrelated errors (*e.g.*, multinomial logit (MNL) and multinomial probit (MNP) with homoskedastic errors that exhibit the same pairwise correlation), but not for those parametric models that feature more flexible error structures (*e.g.*, nested MNL, MNP with a general error covariance matrix, and mixed MNL).

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 notation, an observationally equivalent form of equation (1) may be specified to express the utility that individual n derives from alternative j as

$$u_{nj} = \sigma_n \times (\boldsymbol{x}'_{nj}\boldsymbol{\beta}) + \varepsilon_{nj} \text{ for } n = 1, 2, ..., N \text{ and } j \in \mathbb{J}, \tag{13}$$

where the new parameter $\sigma_n \in \mathbb{R}^1_+$ captures that portion of the overall scale of utility which varies across individuals.[17] Equivalently, σ_n may be also described as a parameter that is inversely

[16] See for example, Beggs *et al.* (1981), Hausman and Ruud (1987), Calfee and Winston (1998), Calfee *et al.* (2001), McCabe *et al.* (2006), Siikamaki and Layton (2007), Scarpa *et al.* (2011), Yoo and Doiron (2013), and Oviedo and Yoo (201[illegible]).

[17] Since any positive monotonic transformation of the utilities preserves the rank order of the original utilities, the random utility specification (13) is observationally equivalent to $u_{nj} = \boldsymbol{x}'_{nj}\boldsymbol{\beta} + \varepsilon_{nj}/\sigma_n$. The slight abuse of notation 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

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 distribution 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 1, 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 empiricist to be agnostic about the exact distribution 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 the 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) restriction, *i.e.*,

$$median(\varepsilon_{nk} - \varepsilon_{nj}|\boldsymbol{X}_n) = 0 \;\; for\ any\ j, k \in \mathcal{J},\ where\ j \neq k, \tag{14}$$

which is then a necessary and sufficient condition for Assumption 1 as long as the distribution of $(\varepsilon_{nk} - \varepsilon_{nj})|\boldsymbol{X}_n$ is a strictly increasing function (see the proof in 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 (14) 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 vector $\boldsymbol{\varepsilon}_n \equiv \boldsymbol{X}_n\boldsymbol{\eta}_n + \boldsymbol{\epsilon}_n$ is

$$\varepsilon_{nj} \equiv \boldsymbol{x}'_{nj}\boldsymbol{\eta}_n + \epsilon_{nj}. \tag{15}$$

Suppose now that idiosyncratic errors $\boldsymbol{\epsilon}_n$ satisfy the pairwise ZCM restriction, $median(\epsilon_{nk} - \epsilon_{nj}|\boldsymbol{X}_n) = 0$ for any $j, k \in \mathcal{J}$ 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 (14) too: differencing two composite errors

like σ_n does not affect any of our earlier results because they do not rely on ε_{nj} having a standardized scale.

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. 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. 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 fully specified much as that of $\boldsymbol{\varepsilon}_n|\boldsymbol{X}_n$.

It is easy to verify that the semiparametric model accommodates the classic troika of parametric random utility models, MNL (or ROL), nested MNL (or nested ROL), and MNP (or ROP).[18] All three models assume away interpersonal 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 MNL, the idiosyncratic errors are *i.i.d.* 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 zero. The nested MNL directly generalizes the MNL 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 zero, though it may not have the unit scale. Finally, in its unrestricted form, the MNP model generalizes the nested MNL model by specifying the multivariate normal density $\boldsymbol{\varepsilon}_n|\boldsymbol{X}_n \sim N(\boldsymbol{0}, \boldsymbol{V}_\varepsilon)$ 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 that is symmetric around zero.

Mixed MNL (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 (15), a mixed MNL model

[18] A major parametric alternative to these three models is the heteroskedastic rank-ordered logit (HROL) model of Hausman and Ruud (1987). Originally introduced as an *ad hoc* specification to address mounting empirical evidence against the ROL model (Hausman and Ruud, 1987), the HROL model has subsequently inspired several other specifications that share similar motivations (Ben-Akiva *et al.*, 1992; Fok *et al.*, 2012; Yoo and Doiron, 2013). We do not consider the HROL model because it stands on its own behavioral foundation that is not shared by other random utility models. In contrast to the microeconomic interpretation of a ranking as a preference ordering based on a single set of utility draws, the HROL model equates a ranking observation with a collection of observations on stage-by-stage choices that have been made as follows. In stage 1, the individual chooses the best out of J alternatives based on a set of utility draws and excludes it from further consideration; in stage 2, she chooses the best out of the remaining $J-1$ alternatives based on a new set of utility draws and eliminates it from further consideration too; and she repeats this process until stage $J-1$ after which only one alternative is left for further consideration. In her observed ranking r_n, her m^{th} best alternative corresponds to her choice in stage m. The hallmark of this framework is that the individual's preferences for alternatives change from one stage to another even when those alternatives are available in all stages.

has idiosyncratic errors $\epsilon_n|\boldsymbol{X}_n$ 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 $\boldsymbol{\eta}_n|\boldsymbol{X}_n \sim N(\boldsymbol{0}, \boldsymbol{V}_{\boldsymbol{\eta}})$ 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, which 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 is allowed that the data generating process comprises different parametric models for different individuals.[19] This flexibility comes from the fact that the GMS estimator does not require the density of $\varepsilon_n|\boldsymbol{X}_n$ 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 binomial choice data. Their general theorem can be applied to

[19] For example, the nested MNL model may generate 1/3 of the population while the mixed MNL may generate the rest.

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. Delgado *et al.* (2001) show that subsampling consistently estimates the asymptotic distribution of the test statistic of the MS estimator. However, subsampling inference is sensitive to the choice of subsample size.[20] Moreover, the standard bootstrap is inconsistent for the MS estimator for binomial choice data, as shown by Abrevaya and Huang (2005), and also for multinomial and rank-ordered choice data.

In this section, we propose an estimator that complements the GMS estimator by addressing these practical limitations, in return for making some additional smoothness assumptions. In the context of Manski's (1985) MS estimator for binomial choice data, Horowitz (1992) develops a smoothed maximum score (SMS) estimator that replaces indicator functions with smooth functions. Yan (2013) applies this technique to derive a smoothed version of Fox's (2007) MS estimator for multinomial choice data. 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 estimator is consistent under the same set of assumptions as the GMS estimator and has a rate of convergence that is faster than $N^{-1/3}$ under extra smoothness conditions. Its asymptotic distribution is multivariate-normal with a covariance matrix that can be consistently estimated from data.

3.1 The SGMS Estimator and its Asymptotic Properties

In this section, we first derive the SGMS estimator and state its consistency result in Theorem 3. Then we summarize the results on its rate of convergence and asymptotic distribution, and state the formal results on the limiting distribution in Theorem 4. Theorem 5 establishes consistent estimation of the parameters in the limiting distribution of the SGMS estimator.

The objective function in (12) can be rewritten as

$$Q_N(\boldsymbol{b}) = N^{-1}\sum_{n=1}^{N}\sum_{1\leq j<k\leq J}\{[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})]\cdot 1(\boldsymbol{x}'_{njk}\boldsymbol{b} \geq 0) + 1(r_{nk} < r_{nj})\} \quad (16)$$

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 (16) can be replaced by a sufficiently smooth function $K(\cdot)$, where $K(\cdot)$ is analogous to a cumulative distribution function (CDF). Application of this smoothing idea in Horowitz (1992) to the right-hand side of (16) yields the SGMS estimator

[20] The computational cost of subsampling is very high for the MS (or GMS) estimator because a global search method is needed to solve the maximization problem for each subsample.

$$\boldsymbol{b}_N^S \in \underset{\boldsymbol{b}\in\mathbb{B}}{argmax}\, Q_N^S(\boldsymbol{b}, h_N), \tag{17}$$

where

$$Q_N^S(\boldsymbol{b}, h_N) \equiv N^{-1}\sum_{n=1}^{N}\sum_{1\leq j<k\leq J}\left\{[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})] \cdot K\left(x_{njk}\boldsymbol{b}/h_N\right) + 1(r_{nk} < r_{nj})\right\} \tag{18}$$

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 on function $K(\cdot)$ for the consistency of the SGMS estimator.

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

(a) $|K(v)| < C$ *for some finite $C \in \mathbb{R}^1_+$ and all $v \in \mathbb{R}^1$; and*

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

Theorem 3. *Let Assumptions* 1-4 *and Condition* 1 *hold. The SGMS estimator $\boldsymbol{b}_N^S \in \mathbb{B}$ defined in* (17) *converges almost surely to the true preference parameter vector $\boldsymbol{\beta}$.*

By Theorem 3, the consistency of the SGMS estimator holds under the same set of assumptions as the GMS estimator, as long as the smooth function $K(\cdot)$ is properly chosen. Since any CDF (*e.g.*, the standard normal distribution function) satisfies Condition 1, the SGMS estimator does not require more assumptions to achieve strong consistency than the GMS estimator does.

Unlike consistency, extra assumptions on the distributions of the error terms and covariates are required in order to derive the asymptotic distribution of the SGMS estimator. Choosing a smooth function $K(\cdot)$ that is at least twice differentiable. Assume that the true parameter vector is an interior point in the parameter space, that is,

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

Then the objective function (18) of the SGMS estimator is a smooth function of $\boldsymbol{b}$ and we can apply a Taylor series expansion method to derive its asymptotic distribution.[21] Let $b_{N,1}^S$ denote the

[21] Unlike binomial choice data, which are generated by a single latent random utility function, multinomial choice data and rank-ordered choice data are generated by multiple latent random utility functions. Yan (2013) explains the challenge of deriving the asymptotic distribution of the SMS estimator for multinomial choice data based on

first element of $\boldsymbol{b}_N^S$ and $\tilde{\boldsymbol{b}}_N^S$ denote the vector of its remaining elements. Recall that the magnitude of first element of $\boldsymbol{\beta}$ is normalized to be one (Assumption 2). By Theorem 3, $b_{N,1}^S$ is a sign consistent estimator for β_1 and the probability $P(b_{N,1}^S - \beta_1 = 0)$ converges to one as N goes to infinity. Since $b_{N,1}^S$ converges to the true parameter at a faster rate than the remaining elements of the SGMS estimator, we focus on the convergence rate and the asymptotic distribution of $(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$ in the following analysis.

Roughly put, the fastest convergence rate of $(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$ to zero is $N^{-d/(2d+1)}$, where d is the positive integer that indicates the strength of the smoothness conditions in Assumption 6 and Assumption 7(a) discussed later. When $d = 1$, the convergence rate of the SGMS estimator is $N^{-1/3}$ and it has an unknown limiting distribution, thus the SGMS estimator does not offer evident advantages over the GMS estimator. When $d \geq 2$, the convergence rate of the SGMS estimator, by appropriately choosing the smooth function $K(\cdot)$ and bandwidth h_N (Condition 2 and Assumption 8), is $N^{-d/(2d+1)}$, and the asymptotic distribution of the SGMS estimator is multivariate normal, making statistical inference straightforward. In other words, in order to have the asymptotic normality of the SGMS estimator, we require the conditional probability of ranking comparison in (5) to be at least twice differentiable with respect to the systematic utility. A larger integer d corresponds to stronger smoothness conditions. Therefore, a higher rate of convergence of the SGMS estimator is achieved at the cost of making stronger smoothness assumptions on the conditional distributions of the error terms and the continuous explanatory variable. For inferential purposes, we assume $d \geq 2$ and treat it as a given/known parameter.[22]

To facilitate a formal statement of the assumptions required for deriving 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 $\boldsymbol{v} \equiv (v_1, \ldots, v_{J-1}, v_J)' \in \mathbb{R}^J$. There is a one-to-one correspondence between $\boldsymbol{X}$ and $(\boldsymbol{v}, \tilde{\boldsymbol{X}})$ for fixed $\boldsymbol{\beta}$, where $\tilde{\boldsymbol{X}} \equiv (\tilde{\boldsymbol{x}}_1, ..., \tilde{\boldsymbol{x}}_J)' \in \mathbb{R}^{J\times(q-1)}$. Define vector $\boldsymbol{\iota}_J \equiv (1, \ldots, 1)' \in \mathbb{R}^J$. For any alternative $j \in \mathbb{J}$, let vector $\boldsymbol{v}_{-j}$ be the

the properties of Horowitz's (1992) binomial SMS estimator. Since rank-ordered choice data are generated by the same multiple utility function as multinomial choice data, deriving its asymptotic distribution is a straightforward extension of the multinomial SMS estimator in Yan (2013). The sketch for deriving the asymptotic distribution of the SGMS estimator is included in Supplementary Material.

[22] Following the notations summarized at the end of Introduction, let $K^{(1)}(\cdot)$ denote the first derivative of $K(\cdot)$. As we will point out shortly, $K^{(1)}(\cdot)$ in our analysis is analogous to a d^{th} order kernel in kernel density estimation. If a faster convergence rate is desired, the researcher may assume a larger d and choose $K(\cdot)$ that gives the corresponding higher order kernel $K^{(1)}(\cdot)$, keeping in mind that this gain in the convergence rate is at the cost of making stronger smoothness assumptions. In our Monte Carlo experiments, we find that assuming $d = 2$ allows the SGMS estimator to perform significantly better than the GMS estimator in terms of achieving smaller mean square error under various error distributions.

difference: $\boldsymbol{v} - v_j \boldsymbol{\iota}_J$. For example, when $1 < j < J$,

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

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 $1 < j < k < J$,

$$\tilde{\boldsymbol{v}}_{-j,k} \equiv (v_1 - v_j, \ldots, v_{k-1} - v_j, v_{k+1} - v_j, \ldots, v_J - v_j)' \in \mathbb{R}^{J-1}$$

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 $1 < j < k < l < J$,

$$\tilde{\boldsymbol{v}}_{-j,kl} \equiv (v_1 - v_j, \ldots, v_{k-1} - v_j, v_{k+1} - v_j, \ldots, v_{l-1} - v_j, v_{l+1} - v_j, \ldots, v_J - v_j)' \in \mathbb{R}^{J-2}.$$

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 $(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$. For any integer $i > 0$, define the derivatives

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

whenever they exist. Denote $p_{jk}^{(0)}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) \equiv p_{jk}(v_{-j,k}|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$. 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 $p_{jklm}(v_{jk}, v_{lm}|\tilde{\boldsymbol{v}}_{\{k,m\}}, \tilde{\boldsymbol{X}})$ 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 $\beta \in \mathbb{B}$. The probability for each individual to rank alternative j over alternative k depends on her covariates matrix $\boldsymbol{X}$, or equivalently, $(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$. Define

$$F_{jk}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) = P(r_j < r_k|v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) \tag{19}$$

and

$$\bar{F}_{jk}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) = P(r_j < r_k|v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}) - P(r_k < r_j|v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}}). \tag{20}$$

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, respectively, by

$$k_d = \int_{-\infty}^{\infty} v^d K^{(1)}(v)dv \text{ and } k_\Omega = \int_{-\infty}^{\infty} [K^{(1)}(v)]^2 dv,$$

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:

$$\boldsymbol{a} = \sum_{1\le j<k\le J} k_d \sum_{i=1}^{d} \frac{1}{i!(d-i)!} E\left[\bar{F}_{jk}^{(i)}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}^{(d-i)}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}\right], \tag{21}$$

$$\boldsymbol{\Omega} = \sum_{1\le j<k\le J} 2k_\Omega E\left[F_{jk}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\boldsymbol{x}_{jk}\tilde{\boldsymbol{x}}'_{jk}\right], \tag{22}$$

and

$$\boldsymbol{H} = \sum_{1\le j<k\le J} E\left[\bar{F}_{jk}^{(1)}(0, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\, p_{jk}(0|\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})\tilde{\boldsymbol{x}}_{jk}\tilde{\boldsymbol{x}}'_{jk}\right], \tag{23}$$

whenever these quantities exist.

Now, we turn to the formal description of the smoothness conditions on the distributions of the error terms and the continuous covariate.

Assumption 6. *For any pair of distinct alternatives $j,k \in \mathbb{J}$, any integer i such that $1 \le i \le d$, all $v_{-j,k}$ in a neighborhood of 0, almost every $(\tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$, and some finite constant C, $\bar{F}_{jk}^{(i)}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})$ exists and is a continuous function of $v_{-j,k}$ satisfying $|\bar{F}_{jk}^{(i)}(v_{-j,k}, \tilde{\boldsymbol{v}}_{-j,k}, \tilde{\boldsymbol{X}})| < C$.*

By definition (20), 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. Further elaboration on the latter point using illustrative examples can be downloaded from the corresponding author's website.[23]

[23] https://sites.google.com/site/yanjin2011/research-2

Assumption 7. *The following statements on the covariates are true.*

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

(b) *If $J \geq 3$, then for any three distinct alternatives $j, k, l \in \mathbb{J}$, all $(v_{-j,k}, v_{-j,l})$, almost every $(\tilde{\boldsymbol{v}}_{-j,kl}, \tilde{\boldsymbol{X}})$, and some finite constant C, $p_{jkl}(v_{-j,k}, v_{-j,l}|\tilde{\boldsymbol{v}}_{-j,kl}, \tilde{\boldsymbol{X}}) < C$.*

(c) *If $J \geq 4$, then for any four distinct alternatives $j, k, l, m \in \mathbb{J}$, all (v_{jk}, v_{lm}), almost every $(\tilde{\boldsymbol{v}}_{\{k,m\}}, \tilde{\boldsymbol{X}})$, and some finite constant C, $p_{jklm}(v_{jk}, v_{lm}|\tilde{\boldsymbol{v}}_{\{k,m\}}, \tilde{\boldsymbol{X}}) < C$.*

(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.*

In addition to the continuity requirement imposed by Assumption 3(a), Assumption 7(a) further requires that the conditional probability density function of the first explanatory variable, $x_{jk,1}$, given other explanatory variables is $(d-1)$ times differentiable, or equivalently, the conditional CDF of the first explanatory variable, $x_{jk,1}$, given other explanatory variables is d times differentiable.

Given the smoothness parameter d in Assumption 6 and Assumption 7(a), the smooth function $K(\cdot)$ is chosen in a way such that its first derivative, $K^{(1)}(\cdot)$, is analogous to a d^{th} order *kernel* in kernel density estimation. Condition 2 lists the requirements on the smooth function in addition to Condition 1.[24]

Condition 2. *The following statements about $K(\cdot)$ are true.*

(a) *$K(v)$ is twice differentiable for $v \in \mathbb{R}$, $|K^{(1)}(v)|$ and $|K^{(2)}(v)|$ are uniformly bounded, and the integrals $\int_{-\infty}^{\infty}[K^{(1)}(v)]^2 dv$, $\int_{-\infty}^{\infty}[K^{(1)}(v)]^4 dv$, $\int_{-\infty}^{\infty} v^2|K^{(2)}(v)|dv$, and $\int_{-\infty}^{\infty}[K^{(2)}(v)]^2 dv$ are finite.*

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

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

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

$$\lim_{N\to\infty} h_N^{i-d} \int_{|h_N v|>\eta} |v^i K^{(1)}(v)|dv = 0 \quad \text{and} \quad \lim_{N\to\infty} h_N^{-1} \int_{|h_N v|>\eta} |K^{(2)}(v)|dv = 0.$$

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

Assumption 9. *The matrix* $\boldsymbol{H}$*, defined by* (23)*, is negative definite.*

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 $K^{(1)}(\cdot)$ when the required derivatives of $\bar{F}(\cdot)$ and $p(\cdot)$ exist. The matrix $\boldsymbol{H}$ in Assumption 9 is analogous to the Hessian information matrix in the quasi-MLE.

Theorem 4. *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* (17)*. If* $Nh_N^{2d+1} \to \lambda$ *as* $N \to \infty$*, where* $\lambda \in [0, \infty)$*, then*

$$(Nh_N)^{1/2}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) \Rightarrow MVN\left(-\lambda^{1/2}\boldsymbol{H}^{-1}\boldsymbol{a}, \boldsymbol{H}^{-1}\boldsymbol{\Omega}\boldsymbol{H}^{-1}\right),$$

and if $Nh_N^{2d+1} \to \infty$ *as* $N \to \infty$*, then* $(h_N)^{-d}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) \xrightarrow{p} -\boldsymbol{H}^{-1}\boldsymbol{a}$.

Theorem 4 implies that given a smoothness condition (where the strength of the smoothness condition is governed by integer d), the SGMS estimator centered by the true parameter vector, $\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}$, converges in distribution to a normal vector at the rate of $(Nh_N)^{-1/2}$ by choosing a bandwidth h_N at the rate equal to or faster than $N^{-1/(2d+1)}$. When the bandwidth h_N converges to zero at the rate of $N^{-1/(2d+1)}$ (*i.e.*, Nh_N^{2d+1} converges to a strictly positive real number λ), the convergence rate of the centered SGMS estimator is $(Nh_N)^{-1/2} = N^{-d/(2d+1)}$, which is the fastest rate of convergence as explained below.

In the case of *under-smoothing* (*i.e.*, Nh_N^{2d+1} converges to zero), bandwidth h_N goes to zero at a rate faster than $N^{-1/(2d+1)}$ and the centered SGMS estimator converges in distribution to a zero-mean normal vector (since $\lambda = 0$) at the rate of $(Nh_N)^{-1/2}$, which is slower than the rate of $N^{-d/(2d+1)}$ because $N^{-d/(2d+1)}/(Nh_N)^{-1/2} = (Nh_N^{2d+1})^{1/(4d+2)}$ goes to zero as N goes to infinity.[25] In the case of *over-smoothing* (*i.e.*, Nh_N^{2d+1} diverges to infinity), bandwidth h_N goes to zero at a

[25] In certain applications, under-smoothing may be a more straightforward way to implement statistical inference because it does not require bias-correction, which is discussed in Section 3.2.

rate slower than $N^{-1/(2d+1)}$ and the centered SGMS estimator converges in probability to a bias term at the rate of h_N^d, which is also slower than the rate of $N^{-d/(2d+1)}$ because $N^{-d/(2d+1)}/(h_N)^d = (Nh_N^{2d+1})^{-d/(2d+1)}$ goes to zero as N goes to infinity.

To make the results of Theorem 4 useful in statistical inferences, it is necessary to be able to estimate the parameters, $\boldsymbol{a}$, $\boldsymbol{\Omega}$, and $\boldsymbol{H}$, in the limiting distribution of the SGMS estimator consistently from observations of $(\boldsymbol{r}, \boldsymbol{X})$. The next theorem shows how this can be done.

Theorem 5. *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 real number* $\delta \in (0,1)$. *Then*

(a) $\hat{\boldsymbol{a}}_N \xrightarrow{p} \boldsymbol{a}$, *where vector*

$$\hat{\boldsymbol{a}}_N \equiv (h_N^*)^{-d} N^{-1} \sum_{n=1}^{N} \sum_{1 \leq j < k \leq J} \left[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})\right] K^{(1)}\left(\boldsymbol{x}_{njk}' \boldsymbol{b}_N^S / h_N^*\right) (\tilde{\boldsymbol{x}}_{njk}/h_N^*);$$

(b) $\hat{\boldsymbol{\Omega}}_N \xrightarrow{p} \boldsymbol{\Omega}$, *where matrix* $\hat{\boldsymbol{\Omega}}_N \equiv (h_N/N) \sum_{n=1}^{N} \boldsymbol{t}_{Nn}(\boldsymbol{b}_N^S, h_N) \boldsymbol{t}_{Nn}(\boldsymbol{b}_N^S, h_N)'$ *and vector*

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

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

(c) *and* $\boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N) \xrightarrow{p} \boldsymbol{H}$, *where matrix*

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

3.2 Implementation Suggestions

3.2.1 Asymptotic Bias Correction

Theorem 4 has allowed us to state earlier that the fastest convergence rate of the SGMS estimator centered at the true parameter vector is $N^{-d/(2d+1)}$, which can be achieved by choosing a bandwidth h_N at the rate of $N^{-1/(2d+1)}$ under particular smoothness conditions (indicated by integer d). Neither *under-smoothing* (*i.e.*, $Nh_N^{2d+1} \to 0$) nor *over-smoothing* (*i.e.*, $Nh_N^{2d+1} \to \infty$) can achieve this fastest rate. For any real number $\lambda \in (0, \infty)$, choosing the bandwidth such that $Nh_N^{2d+1} \to \lambda$ allows the centered SGMS estimator to achieve this fastest rate. 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 using bandwidth $h_N = (\lambda/N)^{1/(2d+1)}$. It follows from Theorem 5 that this bias term 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

$$\tilde{\boldsymbol{b}}_N^{bc} = \tilde{\boldsymbol{b}}_N^S + (\lambda/N)^{d/(2d+1)}\boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)^{-1}\hat{\boldsymbol{a}}_N \tag{24}$$

as the bias-corrected SGMS estimator.

3.2.2 Choice of Bandwidth

Using bandwidth $h_N = (\lambda/N)^{1/(2d+1)}$ (where λ is a strictly positive real number) allows the SGMS estimator centered at the true parameter vector to achieve the fastest rate of convergence given certain smoothness conditions. Next we discuss the choice of the positive parameter λ.

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 the mean square error (MSE) as $E_A[(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})'\boldsymbol{W}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})]$. By the cyclic property of trace, $E_A[(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})'\boldsymbol{W}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})] = trace\{\boldsymbol{W}E_A[(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})']\}$. Theorem 4 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, we calculate

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

From the first order condition, we show that MSE is minimized by setting λ to be

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

or equivalently, $\lambda^* = [trace(\boldsymbol{\Omega}\boldsymbol{H}^{-1}\boldsymbol{W}\boldsymbol{H}^{-1})]/(2d\boldsymbol{a}'\boldsymbol{H}^{-1}\boldsymbol{W}\boldsymbol{H}^{-1}\boldsymbol{a})$ by the cyclic property of trace. In this case $N^{d/(2d+1)}(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}})$ converges to $MVN\left(-(\lambda^*)^{d/(2d+1)}\boldsymbol{H}^{-1}\boldsymbol{a},\ (\lambda^*)^{-1/(2d+1)}\boldsymbol{H}^{-1}\boldsymbol{\Omega}\boldsymbol{H}^{-1}\right)$ in distribution.

The optimal λ^* derived here can be consistently estimated by Theorem 5 and the continuous mapping theorem. Therefore, one possible way of choosing bandwidth is to set $h_N = (\hat{\lambda}/N)^{1/(2d+1)}$, where $\hat{\lambda}$ is a consistent estimator for λ^*. Specifically, the choice of bandwidth can be implemented by taking the following steps given integer $d \geq 2$, *i.e.*, the smoothness conditions.

Step 1. Choose a bandwidth $h_N \propto N^{-1/(2d+1)}$ and another bandwidth $h_N^* \propto N^{-\delta/(2d+1)}$ 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 vector $\hat{\boldsymbol{a}}_N$ and use $\boldsymbol{b}_N^S$ and h_N to compute matrices $\hat{\boldsymbol{\Omega}}_N$ and $\boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)$ as Theorem 5 suggests.

Step 3. Estimate λ^* by

$$\hat{\lambda}_N = \frac{trace\left[\hat{\boldsymbol{\Omega}}_N \boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)^{-1} \boldsymbol{W} \boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)^{-1}\right]}{\left[2d\hat{\boldsymbol{a}}_N' \boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)^{-1} \boldsymbol{W} \boldsymbol{H}_N(\boldsymbol{b}_N^S, h_N)^{-1} \hat{\boldsymbol{a}}_N\right]}. \tag{25}$$

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

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

Note that the approach described by steps 1-5 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 $\delta = 0.1$.

3.2.3 Small-Sample Correction

We describe a method, proposed by Horowitz (1992), to remove part of the finite sample bias of $\hat{\boldsymbol{a}}_N$. A Taylor series expansion of $\hat{\boldsymbol{a}}_N - \boldsymbol{a}$ around $\tilde{\boldsymbol{\beta}}$ yields

$$\hat{\boldsymbol{a}}_N - \boldsymbol{a} = \left[(h_N^*)^{-d} \boldsymbol{t}_N(\boldsymbol{\beta}, h_N^*) - \boldsymbol{a}\right] + (h_N^*)^{-d} \boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N^*)(\tilde{\boldsymbol{b}}_N^S - \tilde{\boldsymbol{\beta}}) \tag{26}$$

with probability approaching one as N goes to infinity, where $\boldsymbol{b}_N^*$ is a vector between $\boldsymbol{b}_N^S$ and $\boldsymbol{\beta}$. The right-hand side of (26) 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 (26) can be written as

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

The probability limit of $\boldsymbol{H}_N(\boldsymbol{b}_N^*, h_N^*)$ is $\boldsymbol{H}$ by Lemmas 7-8 in Supplementary Material, 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 4. 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 $MVN(-\lambda^{1/2}\boldsymbol{a},\ \boldsymbol{\Omega})$.

Based on the above analysis, we treat $\hat{\boldsymbol{a}}_N$ as an estimator of $\{1 - [Nh_N(h_N^*)^{2d}]^{-1/2}\lambda^{1/2}\}\boldsymbol{a}$ rather than that of $\boldsymbol{a}$. Thus, the bias-corrected estimator of vector $\boldsymbol{a}$ is

$$\hat{\boldsymbol{a}}_N^c = \hat{\boldsymbol{a}}_N / \left\{1 - \left[\lambda^{-1}Nh_N(h_N^*)^{2d}\right]^{-1/2}\right\}, \tag{27}$$

which is applied as the estimator of $\boldsymbol{a}$ in our Monte Carlo experiments.

4 Monte Carlo Experiments

In this section, we use Monte Carlo simulation results to study 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

$$u_{nj} = z_{nj,1}\gamma_1 + z_{nj,2}\gamma_{n2} + \alpha_j + \varepsilon_{nj} \text{ for } n = 1, 2, ..., N \text{ and } j = 1, 2, 3, 4. \tag{28}$$

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

In all DGPs, the intercept vector is $\boldsymbol{\alpha} \equiv (\alpha_1, \alpha_2, \alpha_3, \alpha_4)' = (0, 0.25, 0.5, 0.75)'$. The first preference parameter γ_1 is a deterministic coefficient and takes the value of one for all individuals: $\gamma_1 = 1$. In DGPs 1-4, the second preference parameter γ_{n2} is also a deterministic coefficient and takes the value of one for all individuals: $\gamma_{n2} = \gamma_2 = 1$ for all n. In DGPs 5-6, however, γ_{n2} is a random coefficient that varies across individuals, and each individual's coefficient value is a random draw from distribution $N(1, 1)$: $\gamma_{n2} = \gamma_2 + \eta_n$, where $\gamma_2 = 1$ and η_n is distributed as $N(0, 1)$.[26] Each DGP specifies its own distribution of error terms ε_{nj}: we provide more details below.[27]

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

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

The econometrician observes a utility-based ranking $\boldsymbol{r}_n$ of $J = 4$ alternatives in $\mathbb{J}$, as well as attributes $z_{nj,1}$ and $z_{nj,2}$ for $j = 1, ..., 4$ and $n = 1, ..., N$.[28] As usual, the depth of observed rankings influences 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 ($M = J - 1 = 3$) are observed and there is at least one variable satisfying Assumption 3 such as $z_{nj,1}$ and $z_{nj,2}$ in our DGPs, the semiparametric model nests all popular parametric models as special cases; when only partial rankings ($M < 3$) 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 $M = 3$ when the complete ranking is observed. In all DGPs, observed attribute $z_{nj,1}$ is a random draw from $N(0, 2)$ and $z_{nj,2}$ is generated as a ratio of two different uniform draws: specifically, $z_{nj,2} \equiv q_{nj}/w_n$ where q_{nj} is drawn from $U(0, 3)$ and w_n is drawn from $U(\frac{1}{5}, 5)$.[29] Note that $z_{nj,1}$ and q_{nj} vary across both individuals and alternatives, whereas w_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.[30] This allows us to compare our semiparametric methods with both restrictive (ROL) and very flexible (ROP and MROL) parametric methods.

In all estimation runs, we set $\alpha_1 = 0$ for location normalization. Following the notation in

[28]Here we use a relatively small choice set mainly because the probit and the mixed logit specifications yield objective functions that require multivariate integration, and consequently a considerable amount of computation time. The computation time of the GMS and SGMS estimators *per se* is affordable even if the choice set is very large, *e.g.*, $J = 100$ in Yan (2013).

[29]This pair of uniform distributions ensures that the second observed attribute has approximately the same variance as the first attribute, *i.e.*, $Var(q_{nj}/w_n) = 1.9882 \simeq 2$.

[30]Our ROP specification requires estimating five utility index parameters (γ_1, γ_2, α_2, α_3 and α_4) and five 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 coefficients (γ_1 and γ_2), three variance-covariance parameters of their bivariate normal density, and three alternative-specific intercepts (α_2, α_3 and α_4). The ROP (MROL) model has been estimated in Stata using command -asroprobit- (-mixlogit-); the likelihood function has been simulated by taking 200 pseudo-random draws from Hammersley (Halton) sequences.

section 2, let $\beta_1 \equiv \gamma_1$ and $\tilde{\boldsymbol{\beta}} \equiv (\gamma_2, \alpha_2, \alpha_3, \alpha_4)'$. Our discussion focuses on scaled parameter vector, $\tilde{\boldsymbol{\beta}}/\beta_1$, 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 to estimate $(\beta_1, \tilde{\boldsymbol{\beta}}')'$, and then use the results to derive estimated counterparts to $\tilde{\boldsymbol{\beta}}/\beta_1$. In semiparametric models, we normalize $|\beta_1| = 1$ and estimate $\tilde{\boldsymbol{\beta}}$ together with the sign of β_1, and then compute the estimate of the ratio of interest $\tilde{\boldsymbol{\beta}}/sgn(\beta_1)$.[31]

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 (1997), which was also Fox's (2007) preferred method for computing his multinomial MS estimates. As to the SGMS estimator, we assume $d = 2$ (which is the minimum requirement on smoothness conditions for its asymptotic normality) and implement a particular version which uses the standard normal distribution function as the smooth function $K(\cdot)$.[32] 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. The bandwidth has been initialized by setting $h_N = N^{-1/5}$ and $\delta = 0.1$, and optimized subsequently by applying the five steps described in Section 3.2.2 using an identity matrix as the weight matrix $\boldsymbol{W}$.

Table 1 summarizes the true distribution of the error terms in each DGP and whether particular methods can estimate $\tilde{\boldsymbol{\beta}}/\beta_1$ 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 homoskedastic (DGPs 1-2) or heteroskedastic across individuals (DGP 3), but becomes inconsistent in the presence of alternative-specific heteroskedasticity (DGP 4) and/or random coefficients (DGPs 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 through 7 report the bias and root mean square error (RMSE) of each method (in Table 1) using 1,000 samples of size N simulated from DGPs 1 through 6. The last column of each table

[31] The estimator of the sign of β_1 will converge at a much faster rate than the estimators for other parameters such that there is no need to analyze the finite-sample property of the sign estimator.

[32] We follow all the implementation suggestions in Section 3.2 in computing the SGMS estimator, by conducting bias-correction, using the plug-in method to choose the bandwidth, and making a small sample correction.

reports the empirical coverage probability (CP) of the asymptotic 95% confidence interval of the SGMS estimator. While all reported estimation results are for scaled parameters, $\tilde{\boldsymbol{\beta}}/\beta_1$, henceforth we will not stress division by $\beta_1 \equiv \gamma_1$ explicitly for the simplicity of notation and discussion.

We first focus on the slope coefficient γ_2, the results for which vary more widely across DGPs and estimators. The GMS estimator using complete rankings (*i.e.*, $M = 3$) is consistent under all six DGPs, and displays a small finite sample bias, which is less than 3% of the coefficient's true value in DGPs 1 and 2, and 1% in DGPs 3 through 6. In addition, the estimator's RMSE declines noticeably in all DGPs as the sample size grows from $N = 500$ to $N = 1000$, suggesting that its finite sample distribution becomes tighter around the coefficient's true value. The potential benefit of using complete rankings in semiparametric estimation appears considerable. The GMS estimator using partial rankings ($M = 1$ or $M = 2$) is consistent under DGPs 1,2 and 3 but not under DGPs 4, 5 and 6. While the partial rankings estimator still displays a small bias under DGPs 1, 2 and 3, it can be subject to a bias that is about 22% (at $M = 1$) or 8% (at $M = 2$) in DGP 4, and 14% (at $M = 1$) or 5% (at $M = 2$) in DGP 6; the complete rankings estimator's ($M = 3$) bias is practically zero in both DGPs. Comparisons of the SGMS estimator across alternative depth levels and sample sizes lead to similar conclusions, though each SGMS estimator tends to display a larger bias and a smaller RMSE than its GMS counterpart, the expected trade-offs from using a smoothing kernel to construct a surrogate objective function. For DGPs 1, 2, and 5, 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 for γ_2 in DGPs 3, 4, and 6 present particularly interesting examples of the benefit from using our semiparametric methods. Under each of these DGPs, none of the popular parametric methods is consistent but arguably at least one of the parametric methods postulates an approximately correct model. We observe, nevertheless, that even an approximately correct parametric method may display a sizable bias. In DGP 3, for instance, ROP is a correct specification apart from its failure to capture interpersonal heteroskedasticity; yet, the ROP estimator's bias ranges from 37% to 45% 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 $z_{nj,2}$; MROL can readily absorb this component into the normal random coefficient on $z_{nj,2}$, and is therefore a correct specification apart from its inclusion of a redundant extreme value error component. While the MROL estimator's bias is indeed small when only the best alternative indicator is used in estimation ($M = 1$). The bias becomes amplified as deeper ranking information is used and exceeds

13% with the use of complete rankings ($M = 3$).

While the results pertaining to the strong consistency of the GMS and SGMS estimators appear reassuring, the results pertaining to the asymptotic normal distribution of the SGMS estimator sound a cautionary note. The asymptotic 95% confidence intervals for γ_2 have empirical coverage probabilities ranging from 88% to 91% when $N = 500$, and 89% to 92% when $N = 1000$, even when one confines attention to those SGMS estimators that are consistent under a given DGP.[33] It appears that for the asymptotic approximation to work well, one must consider larger sample sizes than what we have examined. For the SMS estimator of binomial choice models, Horowitz (1992) finds an even larger amount of distortion in samples of $N = 500$, which does not improve considerably in larger samples of $N = 1000$, though making head-to-head comparisons with our results is difficult given the use of different DGPs. His subsequent work (Horowitz, 2002) provides a bootstrapping procedure that removes the empirical distortion almost entirely. Our conjecture is that the use of bootstrapping will bring about similarly satisfactory improvement in the present context too. In our view, verification of this conjecture may be best addressed in a dedicated study, for both theoretical and computational reasons. On the theoretical side, one should formally extend Horowitz's (2002) bootstrapping method for the SMS estimator to the SGMS estimator, and verify the validity of the resulting method. On the computational side, we note that obtaining the current set of simulation results for the case of $N = 1000$ and $M = 3$ under one DGP took an average of 10 hours on a powerful workstation; obtaining reliable bootstrapping results involves repeating this type of computing task over several hundred times per each triple of N, M and DGP.[34] Exploring the performance of bootstrapping across alternative DGPs, sample size configurations and preference depths is likely to require several months of computer time, even when one exploits parallel computing.

For the alternative-specific intercepts (α_2, α_3, and α_3), all parametric and semiparametric estimators display practically small biases, even under those DGPs where the estimators in question are inconsistent. We are not aware of any formal explanation for this general robustness, though it appears intuitively plausible that estimating the fixed part of every individual's utility (intercept vector $\boldsymbol{\alpha}$) is an easier task in comparison to estimating the marginal utility weight on an explanatory

[33] As summarized in Table 1, the SGMS estimators using partial rankings are not consistent under DGPs 4-6. In these cases, the coverage probabilities of the asymptotic 95% confidence intervals are not informative about how well asymptotic properties have played out. While the coverage probabilities are sometimes widely off the mark under DGP 4 and DGP 6 when $M < 3$, those results are not alarming considering that the underlying estimators are inconsistent.

[34] We compute the SGMS estimates using Matlab 2018a for 64-bit Windows on a machine with a 3.6 GHz Intel Xeon CPU E3-1271 and a 16GB RAM.

variable that varies across alternatives and individuals (γ_2 on $z_{nj,2}$). The results also suggest that the asymptotic normal distribution of the SGMS estimator provides a better approximation to the finite sample distribution of the intercept estimator than that of the slope coefficient estimator. For each intercept α_j for $j = 2, 3, 4$, the empirical coverage probability of the confidence interval often comes fairly close to the nominal 95% level.

The Monte Carlo experiments were primarily designed to study the properties of our semiparametric methods, but the results provide a fresh perspective on 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 produces different estimates depending on whether data on first preferences ($M = 1$) and deeper rankings ($2 \leq M \leq J-1$) are used, the econometrician should opt for $M = 1$ since deeper ranking information may have been compromised by factors such as decision heuristics: see Chapman and Staelin (1982) and Ben-Akiva *et al.* (1992) for the influential proponents of this view. The results in DGPs 3 through 6, however, caution against testing the reliability of data via comparisons of parametric estimates across alternative levels of M. Since inconsistent parametric estimators may not be equally biased at all levels of M, they may produce estimates that vary across M even when the reliability of data is beyond any doubt as in our simulated samples.

Recall that as Assumption 3(a) in Section 2.2 states, for point identification of parameters, our semiparametric methods requires the presence of a continuous covariate with large support such as $z_{nj,1}$ in the Monte Carlo DGPs. In comparison, parametric methods do not require such a covariate. When Assumption 3(a) fails, there may be a set of parameter vectors that maximize the probability limit of the GMS objective function, instead of a unique parameter vector. Though a detailed theoretical analysis of such partial or set identification is beyond the scope of our paper, we have conducted another Monte Carlo study to develop more insight into the practical consequences of point identification failure using variants of DGP 3 that replace $z_{nj,1}$ and $z_{nj,2}$ with bounded discrete covariates.[35] A summary of the results can be downloaded from the corresponding author's website.[36] We observe that the GMS estimates of the slope parameter γ_2 vary over intervals that are narrow relative to the coefficient's true value as well as RMSEs in Table 4, but those of intercept α_js vary over much wider intervals. Considering the robustness of the parametric estimators of α_js noted earlier, it appears that the complementary use of parametric and semiparametric methods

[35] We use DGP 3 for illustration because it incorporates interpersonal heteroskedasticity (while DGPs 1-2 have homoskedastic errors) and the GMS estimator is consistent across all levels of rankings M under DGP 3 but not DGPs 4-6.

[36] https://sites.google.com/site/yanjin2011/research-2

could be a useful strategy when Assumption 3(a) is violated. The results also point to another potential benefit of using complete rankings in semiparametric estimation, as intervals become tighter as the depth of ranking increases from $M = 1$ through $M = 3$.

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), which rules out interpersonal heteroskedasticity and is only applicable to continuous regressors. Building on Fox's (2007) maximum score (MS) estimator of semiparametric multinomial choice models, we develop the generalized maximum score (GMS) estimator of semiparametric rank-ordered choice models. We show that the GMS estimator allows for arbitrary forms of interpersonal heteroskedasticity and consistent estimation of coefficients on all types of regressors, as long as there is one continuous regressor with large support that can be used to normalize the scale of utility. Like other MS-type estimators, the GMS estimator has a slow convergence rate of $N^{-1/3}$ and a non-standard asymptotic distribution. In the context of binomial choice models, Horowitz (1992) develops the smoothed MS estimator that addresses similar drawbacks of Manski's (1985) MS estimator in return for making stronger assumptions. Yan (2013) extends the results to Fox's (2007) MS estimator of multinomial choice models. Following this tradition, we propose the smoothed GMS (SGMS) estimator which achieves a faster convergence rate and has an asymptotic normal distribution.

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 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 estimators to become robust to wider classes of stochastic specifications. More specifically, the MS estimator using multinomial choices and the GMS estimator using partial rankings do not allow for an error variance-covariance structure that varies across alternatives, meaning that they cannot consistently estimate flexible parametric models including nested logit, unrestricted probit, and mixed logit. By contrast, the GMS estimator using complete rankings (*i.e.*, fully rank-ordered choices) can accommodate error structures as such, fulfilling the usual expectations that a semiparametric model should nest competing parametric models. 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 estimator (and hence that of the SGMS estimator) using fully rank-ordered choices can be therefore shown under almost the same assumptions as the strong consistency of the MS estimator using binomial choices, without invoking stronger assumptions needed to address more analytically complex cases of multinomial choices or partially 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 calls for 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.

Acknowledgements

We thank the editor, associate editor, and anonymous reviewers for their constructive comments on our initial draft. We also thank Arie Beresteanu, Xu Cheng, Liran Einav, Jeremy Fox, Bryan Graham, Bruce Hansen, Han Hong, Arthur Lewbel, Taisuke Otsu, Joris Pinkse, Jack Porter, Xiaoxia Shi, and seminar participants at the 2015 Tsinghua Econometric Conference, Academia Sinica, Sun Yat-Sen University, the 12th International Symposium on Econometric Theory and Applications, the 2016 Asian and European Meetings of the Econometric Society, Newcastle University, the Chinese University of Hong Kong, the 2nd Guangzhou Econometrics Workshop, and National University of Singapore for valuable comments and discussions. Jin Yan acknowledges 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.

Appendices

We provide the proofs of Theorems 1-3 and those of relevant lemmas in Appendices A and B. Specifically, Appendix A provides the proof of identification (Theorem 1) and Appendix B includes the proofs of the strong consistency of the proposed estimators (Theorems 2-3 and Lemmas 1-3). The derivation of the asymptotic distribution of the SGMS estimator and the results for statistical inference (Theorems 4-5 and Lemmas 4-8) require a relatively long list of technical conditions; we present these conditions and associated proofs in Supplementary Material.

Throughout, we use acronyms, LIE, SLLN, and DCT, for Law of Iterated Expectations, Strong Law of Large Numbers, and Dominated Convergence Theorem, respectively. Set $\mathbb{Z}_+$ denotes the collection of positive integers. Symbol $\|\boldsymbol{v}\|$ denotes the L^2 norm of vector $\boldsymbol{v}$ and $|\boldsymbol{v}|$ denotes the vector of the absolute value of each element in $\boldsymbol{v}$. Symbol $O(1)$ ($O_p(1)$) denotes a sequence that is bounded (bounded in probability) and symbol $o(1)$ ($o_p(1)$) denotes a sequence that converges to zero (converges to zero in probability). For any summation indexed by an alternative (alternatives), we suppress the statement that the alternative (alternatives) is (are) in the choice set $\mathbb{J}$. For example, $\sum_{j<k}$ means $\sum_{j<k,\, j\in\mathbb{J},\, k\in\mathbb{J}}$, or equivalently, $\sum_{1\le j<k\le J}$.

A Identification

Proof. (**Theorem 1**) Recall that in Definition 1

$$
\begin{aligned}
Q^*(\boldsymbol{b}) &\equiv \sum_{j<k} E\left[1(r_j<r_k)\cdot 1(\boldsymbol{x}_j'\boldsymbol{b}\ge \boldsymbol{x}_k'\boldsymbol{b}) + 1(r_k<r_j)\cdot 1(\boldsymbol{x}_k'\boldsymbol{b}>\boldsymbol{x}_j'\boldsymbol{b})\right] \\
&= \sum_{j<k} E\left\{[1(r_j<r_k)-1(r_k<r_j)]\cdot 1(\boldsymbol{x}_{jk}'\boldsymbol{b}\ge 0)+1(r_k<r_j)\right\},
\end{aligned} \tag{A1}
$$

where $\boldsymbol{x}_{jk}'\boldsymbol{\beta}\equiv \boldsymbol{x}_j'\boldsymbol{\beta}-\boldsymbol{x}_k'\boldsymbol{\beta}$. Applying the LIE to the right-hand side (RHS) of (A1) yields

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

By Assumption 1, the true parameter vector $\boldsymbol{\beta}$ globally maximizes $Q^*(\boldsymbol{b})$ in (A2) for $\boldsymbol{b}\in\mathbb{B}$ because the sign of the difference, $[P(r_j<r_k|\boldsymbol{X})-P(r_k<r_j|\boldsymbol{X})]$, 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 distinct alternatives $j,k \in \mathbb{J}$, then we can conclude $Q^*(\boldsymbol{\beta}) > Q^*(\boldsymbol{\beta}^-)$. By scale normalization in Assumption 2, 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 covariates where $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^-$ yield different rankings of systematic utilities is [37]

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

By Assumption 3(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 distinct alternatives $j,k \in \mathbb{J}$, that is, $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{\beta}^-$ with probability one. This contradicts Assumption 3(b). If $\beta_1^- = -1$, the set of points where $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^-$ give different predictions is

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

The $D(\boldsymbol{\beta},\boldsymbol{\beta}^-)$ has positive probability by Assumption 3(a). Thus, we have proved that the true preference parameter vector $\boldsymbol{\beta}$ uniquely maximizes $Q^*(\boldsymbol{b})$ *for* $\boldsymbol{b} \in \mathbb{B}$ under Assumptions 1-3. □

B Strong Consistency of the GMS and the SGMS Estimators

We prove Lemmas 1-3 to establish the strong consistency of the GMS and SGMS estimators (Theorem 2-3). Lemma 1 verifies the continuity property of function $Q^*(\boldsymbol{b})$, which is the probability limit of the objective functions of the GMS and SGMS estimators. Lemmas 2 and 3 show the uniform convergence of the GMS objective function, $Q_N(\boldsymbol{b})$, and the SGMS objective function, $Q_N^S(\boldsymbol{b}, h_N)$, to this probability limit function $Q^*(\boldsymbol{b})$, respectively.

Lemma 1. *Under Assumptions 2-3, $Q^*(\boldsymbol{b})$ is continuous in $\boldsymbol{b} \in \mathbb{B}$.*

Proof. Denote each term in the summation on the RHS of (A1) as

$$Q^*_{jk}(\boldsymbol{b}) \equiv E\left\{[1(r_j < r_k) - 1(r_k < r_j)] \cdot 1(\boldsymbol{x}'_{jk}\boldsymbol{b} \geq 0) + 1(r_k < r_j)\right\}. \tag{B1}$$

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

Then,

$$Q^*(\boldsymbol{b}) \;=\; \sum_{j<k} Q^*_{jk}(\boldsymbol{b}). \tag{B2}$$

Therefore, it is sufficient to prove that $Q^*_{jk}(\boldsymbol{b})$ is continuous in $\boldsymbol{b} \in \mathbb{B}$ for any pair of alternatives $j < k$. Consider the case $b_1 = 1$ by the scale normalization in Assumption 2. The argument for $b_1 = -1$ is symmetric. Applying the LIE to the RHS of (B1) yields

$$\begin{aligned} Q^*_{jk}(\boldsymbol{b}) \;&=\; E\left\{[P(r_j < r_k|\boldsymbol{x}_{jk}) - P(r_k < r_j|\boldsymbol{x}_{jk})] \cdot 1(\boldsymbol{x}'_{jk}\boldsymbol{b} \geq 0)\right\} + P(r_k < r_j) \\ &=\; \int \left\{ \int_{-\tilde{\boldsymbol{x}}'_{jk}\tilde{\boldsymbol{b}}}^{\infty} [P(r_j < r_k|\boldsymbol{x}_{jk}) - P(r_k < r_j|\boldsymbol{x}_{jk})] \cdot g_{jk}(x_{jk,1}|\tilde{\boldsymbol{x}}_{jk}) dx_{jk,1} \right\} dF(\tilde{\boldsymbol{x}}_{jk}) \\ &\qquad + P(r_k < r_j), \end{aligned} \tag{B3}$$

where the second equality in (B3) holds by Assumption 3(a) and $F(\tilde{\boldsymbol{x}}_{jk})$ denotes the CDF of $\tilde{\boldsymbol{x}}_{jk}$. The curly brackets inner integral on the RHS of (B3) is a function of $\tilde{\boldsymbol{x}}_{jk}$ and $\tilde{\boldsymbol{b}}$ that is continuous in $\tilde{\boldsymbol{b}} \in \tilde{\mathbb{B}}$. ☐

Lemma 2. *Under Assumption 4, $Q_N(\boldsymbol{b})$ converges almost surely to $Q^*(\boldsymbol{b})$ uniformly over $\boldsymbol{b} \in \mathbb{B}$.*

Proof. Denote the sample analog of (B1) as

$$Q_{Njk}(\boldsymbol{b}) \;\equiv\; N^{-1}\sum_{n=1}^{N} \left\{[1(r_{nj} < r_{nk}) - 1(r_{nk} < r_{nj})] \cdot 1(\boldsymbol{x}'_{njk}\boldsymbol{b} \geq 0) + 1(r_{nk} < r_{nj})\right\}. \tag{B4}$$

By (B1), (B4), and Assumption 4, we have $E[Q_{Njk}(\boldsymbol{b})] = Q^*_{jk}(\boldsymbol{b})$ for any pair of alternatives $j < k$. By (12),

$$Q_N(\boldsymbol{b}) \;=\; \sum_{j<k} Q_{Njk}(\boldsymbol{b}). \tag{B5}$$

Combination of (B2) and (B5) implies that it is sufficient to show that $Q_{Njk}(\boldsymbol{b})$ converges almost surely to $Q^*_{jk}(\boldsymbol{b})$ uniformly over $\boldsymbol{b} \in \mathbb{B}$ for any pair of alternatives $j < k$. Assumption 4 and the uniform SLLN (Theorem 6.2 of Rao (1962) or Lemma 4 of Manski (1985)) imply that

$$P\left[\lim_{N\to\infty} \sup_{\boldsymbol{b}\in\mathbb{B}} \left|Q_{Njk}(\boldsymbol{b}) - Q^*_{jk}(\boldsymbol{b})\right| = 0\right] = 1 \tag{B6}$$

for each pair of alternatives. □

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

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

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

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

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

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

Lemma 3. *Under Assumptions* 2-4 *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 the SGMS objective function $Q_N^S(\boldsymbol{b}, h_N)$ 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 definitions (12) and (18), we calculate

$$\left|Q_N^S(\boldsymbol{b}, h_N) - Q_N(\boldsymbol{b})\right| \leq \frac{1}{N}\sum_{n=1}^{N}\sum_{j<k}\left|1\left(\boldsymbol{x}'_{njk}\boldsymbol{b} > 0\right) - K\left(\boldsymbol{x}'_{njk}\boldsymbol{b}/h_N\right)\right|. \tag{B7}$$

The RHS of (B7)) is the sum of $c_{N1}(\eta)$ and $c_{N2}(\eta)$, where

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

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

and $\eta \in \mathbb{R}_+^1$ 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 an integer $N_0 \in \mathbb{Z}_+$ 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 $\eta > 0$, $c_{N1}(\eta) \to 0$ uniformly over $\boldsymbol{b} \in \mathbb{B}$ as $N \to \infty$. Next consider $c_{N2}(\eta)$. By Condition 1(a), there is a finite C such that

$$c_{N2}(\eta) \leq \sum_{j<k} C \cdot \left[N^{-1} \sum_{n=1}^{N} 1\left(|\boldsymbol{x}'_{njk}\boldsymbol{b}| < \eta \right) \right]. \tag{B8}$$

Assumption 4 and the uniform SLLN (Theorem 7.2 of Rao, 1962) imply that

$$P\left\{ \lim_{N\to\infty} \sup_{\boldsymbol{b}\in\mathbb{B}} \left| C \cdot \left[N^{-1} \sum_{n=1}^{N} 1\left(|\boldsymbol{x}'_{njk}\boldsymbol{b}| < \eta \right) \right] - C \cdot P(|\boldsymbol{x}'_{jk}\boldsymbol{b}| < \eta) \right| = 0 \right\} = 1 \tag{B9}$$

for any pair of alternatives $j < k$. Next, we prove that $P(|\boldsymbol{x}'_{jk}\boldsymbol{b}| < \eta) \to 0$ uniformly over $\boldsymbol{b} \in \mathbb{B}$ as $\eta \to 0$ by verifying the three conditions (*i.e.*, continuity, monotonicity, and pointwise convergence) of Dini's theorem (Theorem 7.13 of Rudin, 1976). We consider $b_1 = 1$; case $b_1 = -1$ is similar. By Assumption 3(a),

$$P(|\boldsymbol{x}'_{jk}\boldsymbol{b}| < \eta) = \int_{-\eta - \tilde{\boldsymbol{x}}'_{jk}\tilde{\boldsymbol{b}}}^{\eta - \tilde{\boldsymbol{x}}'_{jk}\tilde{\boldsymbol{b}}} g_{jk}(x_{jk,1}|\tilde{\boldsymbol{x}}_{jk}) dx_{jk,1} dF(\tilde{\boldsymbol{x}}_{jk}). \tag{B10}$$

Define a sequence of functions $\{f_i^{jk}(\boldsymbol{b}) \equiv P(|\boldsymbol{x}'_{jk}\boldsymbol{b}| < i^{-1}) : i \in \mathbb{Z}_+\}$ for each pair of alternatives $j < k$. By Assumption 3(a) and (B10), it is straightforward to verify that $f_i^{jk}(\boldsymbol{b})$ is continuous in $\boldsymbol{b}$ and $f_i^{jk}(\boldsymbol{b}) > f_{i+1}^{jk}(\boldsymbol{b})$ for any $i \in \mathbb{Z}_+$ and $\boldsymbol{b} \in \mathbb{B}$. As $i \to \infty$, $f_i^{jk}(\boldsymbol{b})$ converges to zero at each $\boldsymbol{b} \in \mathbb{B}$ by Assumption 3(a). Since $\mathbb{B}$ is a compact space (Assumption 2), this pointwise convergence of $f_i^{jk}(\boldsymbol{b})$ to zero implies the uniform convergence of $f_i^{jk}(\boldsymbol{b})$ to zero over $\boldsymbol{b} \in \mathbb{B}$ by Dini's theorem. By (B9), the RHS of (B8) also converges almost surely to zero uniformly over $\boldsymbol{b} \in \mathbb{B}$ as $N \to \infty$ and $\eta \to 0$. The absolute difference $\left|Q_N^S(\boldsymbol{b}, h_N) - Q_N(\boldsymbol{b})\right|$ converges almost surely to zero uniformly over $\boldsymbol{b} \in \mathbb{B}$ as $N \to \infty$ because the RHS of (B7) is the sum of $c_{N1}(\eta)$ and $c_{N2}(\eta)$ for any $\eta > 0$. Since

$$\sup_{\boldsymbol{b}\in\mathbb{B}} \left|Q_N^S(\boldsymbol{b}, h_N) - Q^*(\boldsymbol{b})\right| \leq \sup_{\boldsymbol{b}\in\mathbb{B}} \left|Q_N^S(\boldsymbol{b}, h_N) - Q_N(\boldsymbol{b})\right| + \sup_{\boldsymbol{b}\in\mathbb{B}} \left|Q_N(\boldsymbol{b}) - Q^*(\boldsymbol{b})\right| \tag{B11}$$

and each term on the RHS of (B11) converges to zero almost surely, we have shown that $Q_N^S(\boldsymbol{b}, h_N)$ converges to its probability limit $Q^*(\boldsymbol{b})$ almost surely uniformly over $\boldsymbol{b} \in \mathbb{B}$. □

Proof. (**Theorem 3**) The proof of strong consistency of the SGMS estimator is similar to that of the GMS estimator, which involves verifying the four conditions of Theorem 2.1 in Newey and McFadden (1994). As shown in Theorem 2, the first three conditions are verified by Theorem 1, Assumption 2, and Lemma 1, respectively. The last condition is proved by Lemma 3. Therefore, the SGMS estimator that maximizes its objective function $Q_N^S(\boldsymbol{b}, h_N)$ converges to $\boldsymbol{\beta}$ almost surely under Assumptions 1-4 and Condition 1. $\square$

References

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

[2] Bajari P, Fox JT and Ryan SP. 2008. Evaluating wireless carrier consolidation using semiparametric demand estimation. Quantitative Marketing and Economics 6: 299-338.

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

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

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

[6] Beresteanu A, Zincenko F. 2018. Efficiency Gains in Rank-ordered Multinomial Logit Models. Oxford Bulletin of Economics and Statistics: 80: 122-134.

[7] Calfee J and Winston C. 1998. The value of automobile travel time: implications for congestion policy. Journal of Public Economics 69: 83-102.

[8] 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.

[9] Cameron AC and Trivedi PK. 2005. Microeconometrics: Methods and Applications. Cambridge University Press.

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

[11] Cavanagh C. and Sherman, R. 1998. Rank estimators for monotonic index models. Journal of Econometrics 84: 351-381.

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

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

[14] 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.

[15] 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.

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

[17] 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.

[18] Fok D, Paap R, Van Dijk B. 2012. A Rank-Ordered Logit Model With Unobserved Heterogeneity in Ranking Capabilities. Journal of Applied Econometrics 27: 831-846.

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

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

[21] 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.

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

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

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

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

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

[27] Horowitz JL. 2002. Bootstrap critical values for tests based on the smoothed maximum score estimator. Journal of Econometrics 111: 141-167.

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

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

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

[31] 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.

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

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

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

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

[36] McCabe C, Brazier J, Gilks P, Tsuchiya A, Roberts J, O'Hagan A, Stevens K. 2006. Using rank data to estimate health state utility models. Journal of Health Economics 25: 418-431.

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

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

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

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

[41] Oviedo JL and Yoo H. 2017. A latent class nested logit model for rank-ordered data with application to cork oak reforestation. Environmental and Resource Economics 68: 1021-1051.

[42] Rao R.R. 1962. Relations between weak and uniform convergence of measures with applications. The Annals of Mathematical Statistics 33: 659-680.

[43] Rudin WR. 1976. Principles of Mathematical Analysis. Third Edition. McGraw-Hill.

[44] 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.

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

[46] 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.

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

[48] Siikamaki J, Layton DF. 2007. Discrete choice survey experiments: A comparison using flexible methods. Journal of Environmental Economics and Management 53: 122-139.

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

[50] 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.

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

[52] Yan J. 2013. A smoothed maximum score estimator for multinomial discrete choice models. Working Paper.

[53] 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/

[54] 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.

Table 1: Consistency of estimators by Monte Carlo DGPs

DGP	Distribution of ε_{nj}	ROL	ROP	MRCL	GMS & SGMS
	(a) True parameters: $\gamma_1 = 1$, $\gamma_{n2} = 1$ for all n, and $\alpha_j = (j-1)/4$				
1	ε_{nj} is *i.i.d.* $EV(0,1,0)$	Yes	No	Yes	Yes
2	ε_{nj} is *i.i.d.* $N(0,\pi^2/6)$	No	Yes	No	Yes
3	$\varepsilon_{nj} = 0.82\overline{z}_{n,2}\epsilon_{nj}$ where ϵ_{nj} is *i.i.d.* $N(0,1)$	No	No	No	Yes
4	$\varepsilon_{nj} = 0.75z_{nj,2}\epsilon_{nj}$ where ϵ_{nj} is *i.i.d.* $N(0,1)$	No	No	No	No when $M < 3$; Yes when $M = 3$
	(b) True parameters: $\gamma_1 = 1$, $\gamma_{n2} \overset{i.i.d.}{\sim} N(1,1)$, and $\alpha_j = (j-1)/4$				
5	ε_{nj} is *i.i.d.* $EV(0,1,0)$	No	No	Yes	No when $M < 3$; Yes when $M = 3$
6	$\varepsilon_{nj} = 0.75z_{nj,2}\epsilon_{nj}$ where ϵ_{nj} is *i.i.d.* $N(0,1)$	No	No	No	No when $M < 3$; Yes when $M = 3$

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 $j = 1,\ldots,J$. $M = 3$ ($M < 3$) refers to an estimator that incorporates the complete (partial) rankings. Yes (No) means the estimator of $\tilde{\boldsymbol{\beta}}/\beta_1$ is (not) consistent given the DGP. $\overline{z}_{n,2}$ is the within-individual average of the second covariate, *i.e.*, $\overline{z}_{n,2} = J^{-1}\sum_{j=1}^{J} z_{nj,2}$.

Table 2: Monte Carlo results of DGP 1 (extreme value type 1 errors)

			ROL		ROP		MROL		GMS		SGMS		
M	N		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	CP
1	[illegible]	γ_2	[illegible]	0.1200	0.0110	0.1249	0.0215	0.1271	0.0490	0.2650	0.0925	0.2067	0.920
		α_2	[illegible]	0.1709	0.0162	0.2415	-0.0005	0.1683	0.0213	0.3362	0.0235	0.2325	0.944
		α_3	[illegible]	[illegible]	0.0210	0.2259	-0.0124	0.1647	0.0215	0.3291	0.0366	0.2328	0.947
		α_4	[illegible]	[illegible]	0.0386	0.2221	-0.0154	0.1619	0.0057	0.3280	0.0565	0.2259	0.940
	1000	γ_2	0.0001	[illegible]	[illegible]	0.0881	0.0103	0.0915	0.0248	0.2075	0.0690	0.1555	0.901
		α_2	0.0053	0.1174	[illegible]	[illegible]	0.0016	0.1159	0.0064	0.2666	0.0166	0.1605	0.947
		α_3	0.0041	0.1142	[illegible]	[illegible]	-0.0039	0.1131	0.0150	0.2491	0.0401	0.1597	0.951
		α_4	0.0008	0.1150	[illegible]	[illegible]	-0.0120	0.1151	0.0045	0.2534	0.0519	0.1685	0.940
2	500	γ_2	0.0042	0.0869	0.0034	0.0898	[illegible]	[illegible]	0.0365	0.2097	0.0744	0.1572	0.905
		α_2	0.0020	0.1215	0.0096	0.1280	[illegible]	[illegible]	0.0170	0.2627	0.0173	0.1669	0.937
		α_3	-0.0032	0.1174	0.0074	0.1226	-0.0095	[illegible]	0.0070	0.2587	0.0296	0.1669	0.949
		α_4	-0.0014	0.1169	0.0136	0.1234	-0.0109	[illegible]	[illegible]	0.2620	0.0491	0.1751	0.937
	1000	γ_2	0.0004	0.0597	-0.0043	0.0611	0.0073	0.0623	[illegible]	[illegible]	0.0558	0.1125	0.910
		α_2	0.0004	0.0816	0.0064	0.0856	-0.0018	0.0810	[illegible]	[illegible]	0.0099	0.1136	0.950
		α_3	0.0016	0.0832	0.0118	0.0878	-0.0031	0.0828	0.0034	[illegible]	[illegible]	0.1194	0.943
		α_4	-0.0009	0.0838	0.0114	0.0869	-0.0080	0.0842	0.0038	[illegible]	0.0397	[illegible]	0.927
3	500	γ_2	0.0021	0.0730	-0.0041	0.0759	0.0099	0.0751	0.0184	0.1864	[illegible]	[illegible]	[illegible]
		α_2	-0.0027	0.0998	-0.0008	0.1032	-0.0055	0.0987	0.0073	0.2387	0.0134	[illegible]	[illegible]
		α_3	-0.0059	0.0997	-0.0046	0.1037	-0.0114	0.0992	0.0043	0.2353	0.0271	0.1417	0.952
		α_4	-0.0056	0.1010	-0.0044	0.1054	-0.0136	0.1015	0.0091	0.2431	0.0453	0.1509	0.935
	1000	γ_2	0.0013	0.0519	-0.0062	0.0533	0.0073	0.0542	0.0135	0.1442	0.0498	0.0976	0.905
		α_2	-0.0004	0.0670	0.0026	0.0699	-0.0024	0.0665	-0.0014	0.1838	0.0071	0.0954	0.945
		α_3	0.0029	0.0680	0.0042	0.0708	-0.0012	0.0677	0.0053	0.1848	0.0271	0.1023	0.953
		α_4	-0.0005	0.0714	-0.0008	0.0741	-0.0065	0.0717	0.0080	0.1862	0.0357	0.1080	0.927

Table 3: Monte Carlo results of DGP 2 (normal errors)

			ROL		ROP		MROL		GMS		SGMS		
M	N		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	CP
1	500	γ_2	[illegible]	0.1140	0.0074	0.1157	0.0192	0.1220	0.0367	0.2559	0.0928	0.2070	0.912
		α_2	[illegible]	0.1611	-0.0033	0.2276	0.0011	0.1604	0.0020	0.3316	0.0105	0.2203	0.953
		α_3	[illegible]	[illegible]	-0.0053	0.2102	-0.0052	0.1549	-0.0098	0.3265	0.0280	0.2211	0.945
		α_4	[illegible]	[illegible]	0.0003	0.2085	-0.0064	0.1544	0.0017	0.3140	0.0538	0.2208	0.945
	1000	γ_2	0.0051	0.0828	[illegible]	0.0826	0.0170	0.0893	0.0329	0.2009	0.0726	0.1532	0.904
		α_2	0.0015	0.1131	[illegible]	[illegible]	0.0010	0.1130	-0.0050	0.2659	0.0140	0.1536	0.952
		α_3	0.0069	0.1123	[illegible]	[illegible]	0.0056	0.1120	0.0079	0.2571	0.0378	0.1619	0.944
		α_4	0.0012	0.1097	[illegible]	[illegible]	-0.0009	0.1096	0.0077	0.2517	0.0546	0.1680	0.933
2	500	γ_2	0.0077	0.0874	0.0048	0.0855	[illegible]	[illegible]	0.0371	0.2140	0.0761	0.1610	0.911
		α_2	0.0000	0.1203	-0.0022	0.1252	[illegible]	[illegible]	0.0081	0.2808	0.0143	0.1640	0.958
		α_3	-0.0032	0.1156	-0.0042	0.1169	-0.0084	[illegible]	0.0021	0.2771	0.0281	0.1670	0.948
		α_4	-0.0053	0.1215	-0.0048	0.1243	-0.0133	[illegible]	[illegible]	0.2724	0.0463	0.1755	0.946
	1000	γ_2	0.0063	0.0616	0.0022	0.0595	0.0130	0.0671	[illegible]	[illegible]	0.0600	0.1201	0.892
		α_2	0.0025	0.0853	0.0028	0.0870	0.0006	0.0847	[illegible]	[illegible]	0.0109	0.1196	0.957
		α_3	0.0017	0.0882	0.0015	0.0899	-0.0021	0.0875	0.0027	[illegible]	[illegible]	0.1288	0.946
		α_4	0.0012	0.0859	0.0016	0.0878	-0.0048	0.0858	0.0010	[illegible]	0.0418	[illegible]	0.934
3	500	γ_2	0.0115	0.0803	0.0039	0.0767	0.0184	0.0842	0.0230	0.1991	[illegible]	[illegible]	[illegible]
		α_2	0.0002	0.1087	0.0009	0.1044	-0.0036	0.1072	0.0010	0.2560	0.0139	[illegible]	[illegible]
		α_3	-0.0037	0.1067	-0.0015	0.1048	-0.0114	0.1062	-0.0051	0.2620	0.0239	0.1582	0.943
		α_4	-0.0034	0.1121	-0.0022	0.1093	-0.0145	0.1121	-0.0035	0.2648	0.0434	0.1668	0.937
	1000	γ_2	0.0107	0.0586	0.0029	0.0546	0.0141	0.0619	0.0226	0.1550	0.0574	0.1133	0.895
		α_2	-0.0018	0.0747	-0.0009	0.0732	-0.0051	0.0740	0.0024	0.2017	0.0092	0.1084	0.956
		α_3	-0.0015	0.0789	0.0000	0.0775	-0.0080	0.0784	0.0095	0.2113	0.0241	0.1170	0.937
		α_4	0.0021	0.0775	0.0021	0.0750	-0.0077	0.0771	0.0057	0.2135	0.0416	0.1220	0.931

Table 4: Monte Carlo results of DGP 3 (heteroskedastic errors across individuals)

			ROL		ROP		MROL		GMS		SGMS		
M	N		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	CP
1	500	γ_2	-0.3192	0.3378	-0.3686	0.3873	-0.0902	0.1515	0.0054	0.2117	0.0576	0.1669	0.892
		α_2	-0.0[illegible]41	0.1102	0.0193	0.1841	-0.0244	0.0993	-0.0043	0.1822	0.0112	0.1188	0.962
		α_3	-0.006[illegible]	0.[illegible]132	0.0264	0.1741	-0.0527	0.1122	-0.0102	0.1807	0.0258	0.1232	0.950
		α_4	-0.00[illegible]6	0.111[illegible]	0.0349	0.1714	-0.0844	0.1319	-0.0075	0.1776	0.0437	0.1269	0.938
	1000	γ_2	-0.3221	0.3320	-0.389[illegible]	0.3988	-0.0870	0.1224	0.0022	0.1601	0.0450	0.1156	0.903
		α_2	-0.0030	0.0770	0.0205	0.[illegible]229	-0.0249	0.0700	0.0031	0.1367	0.0066	0.0812	0.967
		α_3	-0.0014	0.0776	0.03[illegible]6	0.12[illegible]2	-0.0504	0.0824	0.0014	0.1356	0.0236	0.0838	0.956
		α_4	0.0002	0.0732	0.038[illegible]	0.[illegible]165	-0.0823	0.1042	0.0020	0.1306	0.0384	0.0878	0.941
2	500	γ_2	-0.3432	0.3546	-0.4182	0.4284	-0.1[illegible]27	0.1590	0.0092	0.1663	0.0490	0.1263	0.903
		α_2	0.0018	0.0773	0.0067	0.0910	[illegible]	0.068[illegible]	0.0006	0.1358	0.0167	0.0778	0.972
		α_3	-0.0020	0.0795	0.0019	0.0918	-0.0526	0.[illegible]845	-0.0034	0.1362	0.0265	0.0849	0.947
		α_4	-0.0036	0.0783	0.0020	0.0865	-0.0833	0.106[illegible]	[illegible]	0.1366	0.0402	0.0917	0.921
	1000	γ_2	-0.3439	0.3495	-0.4267	0.4316	-0.1332	0.1474	[illegible]	[illegible]	0.0364	0.0891	0.900
		α_2	-0.0001	0.0562	0.0035	0.0653	-0.0243	0.0535	0.00[illegible]	[illegible]	0.0084	0.0585	0.957
		α_3	0.0014	0.0554	0.0063	0.0634	-0.0516	0.0692	0.0009	0.[illegible]074	0.020[illegible]	0.0599	0.937
		α_4	0.0009	0.0553	0.0066	0.0625	-0.0815	0.0938	0.0013	0.1041	0.0313	0.0646	0.912
3	500	γ_2	-0.3546	0.3644	-0.4372	0.4449	-0.1378	0.1589	0.0054	0.1554	0.045[illegible]	0.1[illegible]47	[illegible]
		α_2	0.0008	0.0661	0.0013	0.0731	-0.0254	0.0601	-0.0003	0.1191	0.0136	0.[illegible]68[illegible]	0.958
		α_3	-0.0034	0.0690	-0.0025	0.0770	-0.0557	0.0792	-0.0014	0.1196	0.0235	0.0736	0.942
		α_4	-0.0040	0.0691	-0.0022	0.0760	-0.0833	0.1015	-0.0026	0.1194	0.0368	0.0811	0.921
	1000	γ_2	-0.3563	0.3609	-0.4455	0.4492	-0.1396	0.1506	0.0002	0.1247	0.0342	0.0837	0.897
		α_2	0.0008	0.0466	-0.0005	0.0505	-0.0256	0.0462	0.0050	0.0954	0.0094	0.0502	0.947
		α_3	0.0017	0.0471	0.0010	0.0515	-0.0515	0.0643	0.0051	0.0928	0.0206	0.0506	0.946
		α_4	0.0021	0.0485	0.0026	0.0526	-0.0782	0.0878	0.0045	0.0983	0.0312	0.0580	0.911

Table 5: Monte Carlo results of DGP 4 (heteroskedastic errors across alternatives)

			ROL		ROP		MROL		GMS		SGMS		
M	N		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	CP
1	500	γ_2	[illegible]	0.2324	-0.2150	0.2503	-0.0137	0.1284	0.2192	0.3093	0.2844	0.3266	0.540
		α_2	[illegible]	0.1204	0.0208	0.1995	-0.0302	0.1067	-0.0065	0.1830	0.0093	0.1242	0.957
		α_3	[illegible]	[illegible]	0.0352	0.1995	-0.0597	0.1183	-0.0081	0.1795	0.0247	0.1241	0.953
		α_4	[illegible]	[illegible]	0.0486	0.1901	-0.0888	0.1366	-0.0018	0.1703	0.0459	0.1288	0.939
	1000	γ_2	-0.2017	0.2199	-0.2442	0.2620	-0.0153	0.0921	0.2201	0.2786	0.2645	0.2881	0.359
		α_2	0.0003	0.0821	[illegible]	[illegible]	-0.0251	0.0736	-0.0002	0.1332	0.0095	0.0843	0.954
		α_3	-0.0006	0.0825	[illegible]	[illegible]	-0.0563	0.0889	0.0011	0.1334	0.0241	0.0881	0.948
		α_4	0.0016	0.0776	[illegible]	[illegible]	-0.0902	0.1123	0.0016	0.1312	0.0378	0.0901	0.938
2	500	γ_2	-0.3540	0.3651	-0.3691	0.3802	[illegible]	0.1414	0.0794	0.1890	0.1246	0.1735	0.772
		α_2	-0.0021	0.0769	0.0024	0.0908	[illegible]	[illegible]	0.0050	0.1292	0.0134	0.0777	0.960
		α_3	-0.0049	0.0791	-0.0002	0.0928	-0.0542	[illegible]	0.0016	0.1304	0.0237	0.0843	0.945
		α_4	-0.0038	0.0796	0.0032	0.0903	-0.0820	[illegible]	[illegible]	0.1290	0.0396	0.0902	0.922
	1000	γ_2	-0.3560	0.3619	-0.3811	0.3869	-0.1132	0.1290	[illegible]	[illegible]	0.1143	0.1423	0.710
		α_2	0.0003	0.0593	0.0047	0.0709	-0.0236	0.0549	[illegible]	[illegible]	0.0091	0.0596	0.941
		α_3	-0.0010	0.0569	0.0058	0.0670	-0.0533	0.0711	0.0000	[illegible]	[illegible]	0.0595	0.942
		α_4	-0.0002	0.0567	0.0075	0.0664	-0.0817	0.0946	-0.0019	[illegible]	0.0308	[illegible]	0.906
3	500	γ_2	-0.4579	0.4654	-0.4568	0.4641	-0.1588	0.1771	-0.0040	0.1600	[illegible]	[illegible]	[illegible]
		α_2	-0.0017	0.0645	-0.0010	0.0709	-0.0254	0.0585	-0.0009	0.1083	0.0122	[illegible]	0.949
		α_3	-0.0038	0.0663	-0.0033	0.0747	-0.0527	0.0760	-0.0001	0.1063	0.0231	0.0725	0.942
		α_4	-0.0039	0.0688	-0.0023	0.0754	-0.0776	0.0970	-0.0024	0.1047	0.0381	0.0797	0.924
	1000	γ_2	-0.4598	0.4638	-0.4656	0.4693	-0.1591	0.1688	-0.0007	0.1274	0.0328	0.0847	0.901
		α_2	0.0016	0.0469	0.0016	0.0516	-0.0235	0.0448	0.0035	0.0854	0.0103	0.0500	0.936
		α_3	0.0002	0.0461	0.0012	0.0511	-0.0503	0.0626	0.0052	0.0863	0.0201	0.0508	0.931
		α_4	0.0002	0.0477	0.0020	0.0526	-0.0759	0.0854	0.0021	0.0855	0.0308	0.0575	0.906

Table 6: Monte Carlo results of DGP 5 (random coefficient with extreme value errors)

			ROL		ROP		MROL		GMS		SGMS		
M	N		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	CP
1	500	γ_2	-0.2998	0.3328	-0.3508	0.3808	-0.0116	0.1741	-0.0346	0.2975	0.0304	0.2326	0.904
		α_2	[illegible]	0.1781	0.0147	0.2795	-0.0029	0.1725	0.0205	0.3495	0.0265	0.2372	0.950
		α_3	[illegible]	[illegible]	0.0358	0.2771	-0.0000	0.1744	0.0250	0.3331	0.0572	0.2428	0.932
		α_4	[illegible]	[illegible]	0.0520	0.2653	-0.0060	0.1661	0.0244	0.3457	0.0852	0.2476	0.936
	1000	γ_2	-0.2920	0.3099	[illegible]	0.3715	-0.0011	0.1230	-0.0391	0.2358	0.0182	0.1737	0.914
		α_2	-0.0006	0.1264	[illegible]	[illegible]	-0.0041	0.1217	-0.0036	0.2653	0.0111	0.1690	0.951
		α_3	0.0014	0.1208	[illegible]	[illegible]	-0.0068	0.1179	0.0006	0.2642	0.0360	0.1715	0.945
		α_4	0.0003	0.1193	[illegible]	[illegible]	-0.0116	0.1181	0.0065	0.2633	0.0543	0.1768	0.932
2	500	γ_2	-0.2848	0.3116	-0.3407	0.3630	[illegible]	[illegible]	-0.0226	0.2447	0.0432	0.1880	0.910
		α_2	0.0030	0.1242	0.0091	0.1348	[illegible]	[illegible]	0.0076	0.2707	0.0208	0.1681	0.958
		α_3	0.0042	0.1315	0.0144	0.1402	-0.0003	[illegible]	[illegible]	0.2741	0.0421	0.1819	0.935
		α_4	0.0037	0.1234	0.0179	0.1308	-0.0040	[illegible]	[illegible]	0.2740	0.0620	0.1840	0.924
	1000	γ_2	-0.2766	0.2898	-0.3401	0.3515	-0.0008	0.0866	[illegible]	[illegible]	0.0408	0.1392	0.921
		α_2	-0.0025	0.0897	0.0034	0.0952	-0.0037	0.0874	[illegible]	[illegible]	0.0073	0.1269	0.936
		α_3	-0.0017	0.0877	0.0088	0.0917	-0.0049	0.0858	-0.0006	[illegible]	[illegible]	0.1291	0.934
		α_4	-0.0014	0.0862	0.0120	0.0916	-0.0070	0.0853	-0.0107	[illegible]	0.0433	[illegible]	0.923
3	500	γ_2	-0.2817	0.3060	-0.3368	0.3567	-0.0079	0.1143	0.0024	0.2357	0.0581	[illegible]	[illegible]
		α_2	0.0007	0.1064	0.0035	0.1105	-0.0008	0.1022	0.0088	0.2484	0.0178	[illegible]	[illegible]
		α_3	0.0016	0.1126	0.0059	0.1178	-0.0017	0.1085	0.0111	0.2503	0.0390	0.1615	0.927
		α_4	0.0022	0.1099	0.0035	0.1145	-0.0039	0.1065	0.0188	0.2575	0.0609	0.1662	0.925
	1000	γ_2	-0.2721	0.2839	-0.3332	0.3434	0.0016	0.0786	-0.0015	0.1867	0.0510	0.1339	0.907
		α_2	-0.0043	0.0756	-0.0019	0.0781	-0.0046	0.0732	-0.0104	0.1972	0.0074	0.1073	0.949
		α_3	-0.0014	0.0742	0.0021	0.0773	-0.0034	0.0728	-0.0058	0.1982	0.0263	0.1104	0.945
		α_4	-0.0011	0.0750	-0.0003	0.0776	-0.0045	0.0736	-0.0069	0.2049	0.0410	0.1192	0.921

Table 7: Monte Carlo results of DGP 6 (random coefficients with heteroskedastic errors)

M	N		ROL		ROP		MROL		GMS		SGMS		
			Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	CP
1	500	γ_2	-0.3556	0.3781	-0.4000	0.4216	-0.0435	0.1601	0.1450	0.3046	0.2061	0.2931	0.761
		α_2	[illegible]	0.1265	0.0252	0.2047	-0.0191	0.1094	0.0085	0.2018	0.0183	0.1338	0.942
		α_3	[illegible]	[illegible]	0.0388	0.1977	-0.0376	0.1143	0.0043	0.1933	0.0379	0.1389	0.936
		α_4	[illegible]	[illegible]	0.0498	0.1937	-0.0626	0.1251	0.0101	0.1939	0.0539	0.1398	0.938
	1000	γ_2	-0.3532	0.3650	[illegible]	0.4243	-0.0448	0.1186	0.1399	0.2483	0.1984	0.2430	0.684
		α_2	-0.0048	0.0914	0.0190	[illegible]	-0.0225	0.0787	-0.0001	0.1482	0.0072	0.0934	0.955
		α_3	-0.0007	0.0849	[illegible]	[illegible]	-0.0435	0.0840	-0.0025	0.1468	0.0263	0.0921	0.948
		α_4	-0.0006	0.0827	[illegible]	[illegible]	-0.0703	0.1023	0.0009	0.1454	0.0410	0.0983	0.942
2	500	γ_2	-0.4686	0.4800	-0.4957	0.5056	[illegible]	0.1564	0.0597	0.2352	0.1041	0.1913	0.847
		α_2	-0.0003	0.0903	0.0047	0.1050	[illegible]	[illegible]	0.0034	0.1456	0.0150	0.0913	0.956
		α_3	-0.0017	0.0901	0.0048	0.1022	-0.0418	[illegible]	0.0029	0.1451	0.0311	0.0958	0.937
		α_4	0.0006	0.0893	0.0084	0.1021	-0.0636	[illegible]	[illegible]	0.1414	0.0486	0.1035	0.926
	1000	γ_2	-0.4687	0.4746	-0.5035	0.5084	-0.1129	0.1383	[illegible]	[illegible]	0.0905	0.1469	0.825
		α_2	-0.0019	0.0641	0.0030	0.0733	-0.0221	0.0554	[illegible]	[illegible]	0.0075	0.0649	0.948
		α_3	0.0028	0.0603	0.0107	0.0696	-0.0434	0.0647	0.0007	[illegible]	0.0241	0.0641	0.954
		α_4	-0.0014	0.0616	0.0070	0.0706	-0.0712	0.0881	-0.0025	[illegible]	0.0336	[illegible]	0.921
3	500	γ_2	-0.5453	0.5536	-0.5522	0.5596	-0.1334	0.1688	0.0050	0.2087	0.0471	[illegible]	[illegible]
		α_2	-0.0002	0.0763	-0.0007	0.0825	-0.0203	0.0617	-0.0009	0.1223	0.0143	[illegible]	0.948
		α_3	-0.0020	0.0752	-0.0016	0.0822	-0.0429	0.0734	0.0032	0.1278	0.0315	0.0834	0.929
		α_4	-0.0012	0.0782	0.0005	0.0833	-0.0648	0.0904	0.0001	0.1249	0.0465	0.0911	0.917
	1000	γ_2	-0.5431	0.5474	-0.5563	0.5599	-0.1379	0.1551	-0.0115	0.1579	0.0331	0.1111	0.887
		α_2	-0.0021	0.0541	-0.0024	0.0585	-0.0229	0.0477	-0.0019	0.0948	0.0082	0.0544	0.950
		α_3	0.0033	0.0534	0.0025	0.0566	-0.0426	0.0594	-0.0025	0.0949	0.0240	0.0565	0.940
		α_4	-0.0015	0.0554	-0.0010	0.0593	-0.0684	0.0813	-0.0013	0.0969	0.0331	0.0635	0.914