Department of
Economics

University of Pittsburgh
Kenneth P. Dietrich School of Arts and Sciences

WORKING PAPER SERIES

16/018

Efficiency Gains in Rank-ordered Multinomial Logit Models

Arie Beresteanu and Federico Zincenko

June 13, 2016

University of Pittsburgh • Dietrich School of Arts and Sciences • Department of Economics
4901 Wesley W. Posvar Hall • 230 South Bouquet Street • Pittsburgh, PA 15260

Efficiency Gains in Rank-ordered Multinomial Logit Models*

Arie Beresteanu† and Federico Zincenko‡

June 13, 2016

Abstract

This paper considers estimation of discrete choice models when agents report their ranking of the alternatives (or some of them) rather than just the utility maximizing alternative. We investigate the parametric conditional rank-ordered Logit model. We show that conditions for identification do not change even if we observe ranking. Moreover, we fill a gap in the literature and show analytically and by Monte Carlo simulations that efficiency increases as we use additional information on the ranking.

Keywords: Rank-ordered Logit, Random Utility, Conditional Maximum Likelihood

*We are grateful to JF Richards for useful comments and detailed discussions. We thanks the seminar participants at Vanderbilt, the 10th GNYMA Econometrics Colloquium (Princeton), and the 2015 Pittsburgh Economics Medley.
†Department of Economics, University of Pittsburgh, arie@pitt.edu.
‡Department of Economics, University of Pittsburgh, zincenko@pitt.edu.

1 Introduction

The conditional Logit model (McFadden (1974)) is a widely used estimator for demand estimation in discrete choice models. Typical data sets include the final choice made by decision makers as well as the observed characteristics of the alternatives they faced. The unobserved characteristic of each alternative is assumed to have a type one extreme value distribution independent across alternatives and individuals. In some situations, however, decision makers report their ranking of whole or part of the alternative set. This paper explores the implications of observing more than just the maximal choice on estimation of preferences parameters in the conditional Logit model.

Cases where decision makers report a ranking of several alternatives exist mostly in survey data. For example, in optimal assignment problems it is common to ask respondents to rank their top choices among the set of alternatives from which they can choose. A case that received a lot of attention in the past is assigning medical school graduates to hospitals for internship. This two-sided market and the mechanism design approach taken to make it optimal is described in Roth and Peranson (1999). Another example that received much attention is the literature on school choices by students and parents. The most famous examples are the Boston and NYC public school systems where individuals are asked to report their top two or three choices among the alternatives that they are facing.

Beggs, Cardell, and Hausman (1981) analyze surveys on consumers demand for electric cars. They note that while survey data is usually inferior to real life transactions, surveys allow us to directly ask consumers about hypothetical products. In addition, consumers can be asked to rank several options. Beggs, Cardell, and Hausman (1981) write down the (conditional) log-likelihood function for the case where consumers report a ranking of some top choices under the assumption that the error terms have type one extreme value distribution. They show that this log-likelihood function is globally concave and thus the (conditional) maximum likelihood estimator can be easily found numerically. However, the asymptotic properties of this estimator were not explored. Specifically, there is no comparison between the asymptotic variance of the ranked Logit estimator and that of the regular Logit.

The ranked Logit model is also discussed in Chapman and Staelin (1982). They discuss the impact of the depth of the ranking on the estimator but do not provide analytical results. Our main theorem allows us to fill this gap. Moreover, the literature on rank ordered discrete choice raises the following concern. Suppose individuals are asked to rank a large number of alternatives. It is safe to assume that the top ranked choices receive careful attention from the decision makers, but it is possible that the specific rank at the bottom is uninformative and perhaps even misleading. In other words, having rank data may bias our estimators. Hausman and Ruud (1987) develop a test to see whether the lower ranked alternatives are consistent with the highly ranked one. Hausman and Ruud (1987) compute the standard errors for their estimators based on the outer product of the score, while here we suggest using the Hessian-based estimator. Asymptotically, both methods are equivalent. The analytic results of this paper employs the Hessian-based covariance matrix, so we suggest using it to compute the standard errors. Another example for the concern about ranking quality appears in van Dijk, Fok, and Paap (2012), who develop an estimator in which the ability of a decision maker to evaluate a long list of alternative is its type. They allow for heterogeneous types in the population. The efficiency gains afforded by their method are demonstrated using Monte Carlo experiments.

Our paper makes several contributions to the literature on ranked Logit. The main result of this paper appears in Theorem 2.1. This theorem shows that the ranked Logit has a smaller asymptotic variance than the regular Logit and therefore is a more efficient estimator. Moreover, we show that conditions for identification do not change even if we observe a ranking. In other words, observing ranking contributes to efficiency only (Lemma 2.1). In addition, this lemma also shows that the solution to the log-likelihood of the ranked logit model exist and is unique. Our main theorem has implications for sample design as well. We describe a simple procedure to evaluate the efficiency gains from increasing the depth of the ranking a-priori. We show that based on the current sample and before collecting additional information the researcher can estimate the expected efficiency gain should she decide to collect more information. The Monte Carlo experiment in Section 3 demonstrates the efficiency gains from using a deeper ranking. The results also show that the gains diminish very quickly.

The paper is organized as follows. Section 2 presents the model and main result of our paper. In section 3, we include a Monte Carlo experiment to demonstrate our findings. Section 4 concludes. The proofs of the theorems and lemmas appear in the appendix.

2 Efficiency Gains

In this section we establish the notation for the ranked logit model and present the main results of the paper. Consider a typical single-agent unordered discrete choice model. Let $\mathcal{J} = \{1, 2, \ldots, J\}$ be the choice set that each decision maker faces where $J \geq 2$. Let $x_{i,l} = (x_{i,l}^{(1)}, \ldots, x_{i,l}^{(D)})'$ be a $D \times 1$-vector of observable covariates describing the characteristics of choice l faced by individual i. The vector $x_{i,l}$ differs across alternatives $l = 1, \ldots, J$ and possibly across individuals $i = 1, \ldots, N$ as well. Let $\varepsilon_{i,l}$ be the sum of factors of choice l faced by individual i and unobserved to the econometrician. We assume that individual i's utility from choice l is $u_{i,l} = x_{i,l}'\beta + \varepsilon_{i,l}$ where β is a $D \times 1$ vector of parameters.

In most common applications individuals either report their top choice from the set $\mathcal{J}$ or this top choice is observed as their action. In this paper, we look at cases where individuals are asked to report their R top choices from the set $\mathcal{J}$ where $1 \leq R < J$.

Assumption 2.1 *Each individual reports a set of R distinct indexes $(j_1, .., j_R)$ from $\mathcal{J}$ such that $u_{i,j_1} > u_{i,j_2} > \cdots > u_{i,j_R}$ and $u_{i,j_R} > u_{i,l}$ for all $l \in \mathcal{J} \backslash \{j_1, .., j_R\}$.*

With this assumption we rule out strategic behavior in which individuals rank less preferred alternatives higher to influence the choice assigned to them. This type of situation may occur in two-sided market as described in Roth and Peranson (1999). Assumption 2.1 rules out this type of strategic behavior. Moreover, we do not consider ties in the utilities because the assumptions imposed in the rest of the paper imply that the probability of a tie is zero.

We consider the rank-ordered Logit model first studied in Beggs, Cardell, and Hausman (1981). For any integer $1 \leq R < J$, let $\mathcal{J}_R$ denote the set of all ordered R-tuples from $\{1, 2, \ldots, J\}$. In other words, $\mathcal{J}_R$ is the set of sequences of length R of elements taken from $\mathcal{J}$ without repetition.[1] With a slight abuse of notation we let $\mathcal{J}_1 = \mathcal{J} = \{1, 2, \ldots, J\}$, so the case in which $R = 1$ corresponds the standard unordered discrete choice model with J alternatives.

We assume that all individuals report their R top choices for some integer $1 \leq R < J$. Specifi-

[1] E.g., when $J = 3$ and $R = 2$, we have $\mathcal{J}_R = \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\}$.

cally, individual i chooses the R-tuple $j = (j_1, \ldots, j_R) \in \mathcal{J}_R$ if and only if

$$\begin{aligned} x'_{i,j_1}\beta + \varepsilon_{i,j_1} &> x'_{i,j_2}\beta + \varepsilon_{i,j_2}, \\ x'_{i,j_2}\beta + \varepsilon_{i,j_2} &> x'_{i,j_3}\beta + \varepsilon_{i,j_3}, \ldots \\ x'_{i,j_R}\beta + \varepsilon_{i,j_R} &> \max_{l \in \mathcal{J} \setminus \{j_1,\ldots,j_R\}} x'_{i,l}\beta + \varepsilon_{i,l} \end{aligned}$$

or, equivalently, if and only if

$$x'_{i,j_r}\beta + \varepsilon_{i,j_r} > \max_{l \in \mathcal{J} \setminus \{j_1,\ldots,j_r\}} x'_{i,l}\beta + \varepsilon_{i,l} \tag{2.1}$$

for all $r = 1, \ldots, R$. The presentation in equation (2.1) can be thought as a sequence of R classical Logit models where in each one of them the choice set is $\mathcal{J} \setminus \{j_1, \ldots, j_r\}$ and the utility maximizing choice is j_r. This type of presentation is called exploded Logit; see Train (2009), sec. 7.3.1. Denote $\mathbf{x}_i = (x'_{i,1}, \ldots, x'_{i,J})_{[1 \times JD]}$ and $\varepsilon_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,J})_{[1 \times J]}$. We make the following assumption.

Assumption 2.2 $\{(\mathbf{x}_i, \varepsilon_i) : i = 1, \ldots, N\}$ *are independent and identically distributed (i.i.d.) random vectors and the following conditions hold.*

1. $\mathbf{x}_i$ *and* ε_i *are independent.*
2. $\mathbf{x}_i$ *has density* $f(\cdot)$ *with respect to a* σ*-finite measure* $\nu(\cdot)$. $E(\mathbf{x}'_i \mathbf{x}_i)$ *exists and is finite.*
3. $\varepsilon_{i,l}$ *are i.i.d. across* $l = 1, \ldots, J$ *with type 1 extreme value distribution with probability density function* $g(t) = \exp(-t)\exp[-\exp(-t)]$ *for* $t \in \mathbb{R}$.
4. $\beta \in \text{interior}(\mathcal{B})$ *for some compact and convex set* $\mathcal{B} \subset \mathbb{R}^D$.

Let y_i^R denote the R-tuple chosen by individual i from $\mathcal{J}_R$. Several remarks are noteworthy. First, Assumption 2.2 implies that $\{(y_i^R, \mathbf{x}_i) : i = 1, \ldots, N\}$ are i.i.d. Second, part 2 of Assumption 2.2 allows for any type and combination of covariates (discrete, continuous, or some mixture both).

Third, the conditional probability of choosing $(j_1, \ldots, j_R) \in \mathcal{J}_R$ is given by

$$
\begin{aligned}
\mathrm{P}[y_i^R = (j_1, \ldots, j_R) | \mathbf{x}_i; \beta] &= \mathrm{P}\left(x'_{i,j_1}\beta + \varepsilon_{i,j_1} > \max_{l \in \mathcal{J} \setminus \{j_1\}} \{x'_{i,l}\beta + \varepsilon_{i,l}\} \middle| \mathbf{x}_i; \beta \right) \\
&\quad \times \mathrm{P}\left(x'_{i,j_2}\beta + \varepsilon_{i,j_2} > \max_{l \in \mathcal{J} \setminus \{j_1, j_2\}} \{x'_{i,l}\beta + \varepsilon_{i,l}\} \middle| \mathbf{x}_i; \beta \right) \times \cdots \\
&\quad \times \mathrm{P}\left(x'_{i,j_R}\beta + \varepsilon_{i,j_R} > \max_{l \in \mathcal{J} \setminus \{j_1, \ldots, j_R\}} \{x'_{i,l}\beta + \varepsilon_{i,l}\} \middle| \mathbf{x}_i; \beta \right) \\
&= \frac{\exp(x'_{i,j_1}\beta)}{\sum_{l \in \mathcal{J}} \exp(x'_{i,l}\beta)} \times \frac{\exp(x'_{i,j_2}\beta)}{\sum_{l \in \mathcal{J} \setminus \{j_1\}} \exp(x'_{i,l}\beta)} \times \cdots \\
&\quad \times \frac{\exp(x'_{i,j_R}\beta)}{\sum_{l \in \mathcal{J} \setminus \{j_1, \ldots, j_{R-1}\}} \exp(x'_{i,l}\beta)}, \qquad (2.2)
\end{aligned}
$$

This expression is obtained from the properties of the type 1 extreme value distribution. As discussed in McFadden (1984), pp. 1413-1415, the probability of an observed ranking $(j_1, \ldots, j_R)$ is the product of conditional probabilities of choice from successively restricted subsets. This result is an immediate consequence of the Independence from Irrelevant Alternatives (IIA) property.

For every i and for every R-tuple $j \in \mathcal{J}_R$, let $d_{ij}^R = 1\left\{y_i^R = j\right\}$, where $1\{\cdot\}$ is the indicator function. From expression (2.2), the conditional log-likelihood function for observation i is

$$
\begin{aligned}
l_i^R(b) &= \sum_{j \in \mathcal{J}_R} d_{ij}^R \log\left\{P[y_i^R = (j_1, \ldots, j_R) | \mathbf{x}_i; b]\right\} \\
&= \sum_{j \in \mathcal{J}_R} d_{ij}^R \sum_{r=1}^{R} \left\{ x'_{i,j_r} b - \log\left[\sum_{l \in \mathcal{J} \setminus \{j_1, \ldots, j_{r-1}\}} \exp(x'_{i,l} b) \right] \right\}
\end{aligned}
$$

for $b \in \mathbb{R}^{D \times 1}$. When $R = 1$, we denote $\{j_1, \ldots, j_{R-1}\} = \emptyset$ and $\sum_{l \in \mathcal{J} \setminus \{j_1, \ldots, j_{R-1}\}} = \sum_{l \in \mathcal{J}}$, so the above log-likelihood function generalizes the log-likelihood function of the traditional multinomial Logit model. The score of this log-likelihood is

$$
\underset{D \times 1}{s_i^R(b)} \equiv \frac{\partial l_i^R(b)}{\partial b} = \sum_{j \in \mathcal{J}_R} d_{ij}^R \sum_{r=1}^{R} \left\{ x_{i,j_r} - \frac{\sum_{l \in \mathcal{J} \setminus \{j_1, \ldots, j_{r-1}\}} \exp(x'_{i,l} b) x_{i,l}}{\sum_{m \in \mathcal{J} \setminus \{j_1, \ldots, j_{r-1}\}} \exp(x'_{i,l} b)} \right\}.
$$

Define the matrix

$$
\begin{aligned}
\underset{D \times D}{I^R(b)} \equiv -E\left[\frac{\partial l_i^R(b)}{\partial b \partial b'}\right] &= E\left[\sum_{j \in \mathcal{J}_R} d_{ij}^R \sum_{r=1}^{R} H_{\{j_1, \ldots, j_{r-1}\}}(\mathbf{x}_i; b) \right] \\
&= \sum_{j \in \mathcal{J}_R} \sum_{r=1}^{R} E\left[d_{ij}^R H_{\{j_1, \ldots, j_{r-1}\}}(\mathbf{x}_i; b) \right], \qquad (2.3)
\end{aligned}
$$

where

$$H_{\{j_1,\dots,j_{r-1}\}}(\mathbf{x}_i;b) = \frac{\sum_{l\in\mathcal{J}\backslash\{j_1,\dots,j_{r-1}\}}\exp(x'_{i,l}b)x_{i,l}x'_{i,l}}{\sum_{m\in\mathcal{J}\backslash\{j_1,\dots,j_{r-1}\}}\exp(x'_{i,m}b)} - \frac{[\sum_{l\in\mathcal{J}\backslash\{j_1,\dots,j_{r-1}\}}\exp(x'_{i,l}b)x_{i,l}][\sum_{l\in\mathcal{J}\backslash\{j_1,\dots,j_{r-1}\}}\exp(x'_{i,l}b)x'_{i,l}]}{[\sum_{m\in\mathcal{J}\backslash\{j_1,\dots,j_{r-1}\}}\exp(x'_{i,m}b)]^2}.$$

It can be shown that $H_{\{j_1,\dots,j_{r-1}\}}(\mathbf{x}_i;b)$ is positive semi-definite for any value of $\mathbf{x}_i$ and b (Lemma A.1 in Appendix A.1). By the unconditional information matrix equality, we have that $I^R(\beta) = E[s_i^R(\beta)s_i^R(\beta)']$, so $I^R(\beta)$ is the Fisher information matrix. Before proceeding, we make the following identification assumption.

Assumption 2.3 *For every $c\in\mathbb{R}^{D\times 1}\backslash\{0\}$, there are $\bar{l}_1,\bar{l}_2\in\mathcal{J}$ such that*

$$E\left\{[c'(x_{i,\bar{l}_1}-x_{i,\bar{l}_2})]^2\right\} > 0.$$

This assumption provides the key identification condition and is independent of the length of the ranking. It rules out exact collinearity among the regressors and requires variation across alternatives. Alternative-specific constants can be included, with one of them normalized to 0. We remark that the alternatives $\bar{l}_1,\bar{l}_2$ may depend on the value of c. Assumption 2.3 is a necessary and sufficient condition for identification. The next lemma formalizes this result.

Lemma 2.1 *The following statements hold for any $1\le R<J$.*

1. *Under Assumptions 2.1-2.3, $\sum_{j\in\mathcal{J}_R} E\left[d_{ij}^R H_{\{j_1,\dots,j_{R-1}\}}(\mathbf{x}_i;b)\right]$ is positive definite for every $b\in\mathbb{R}^{D\times 1}$. As a consequence, $I^R(\beta)$ is positive definite.*
2. *Under Assumptions 2.1-2.2, there exists a solution to the problem $\max\{E[l_i(b)] : b\in\mathcal{B}\}$. Such a solution is unique if and only if Assumption 2.3 holds.*

Proof. See Appendix A.1. ■

From the second part of this lemma, β can be characterized as the unique argument that maximizes $E[l_i^R(\cdot)]$ over $\mathcal{B}$, i.e., $\beta = \arg\max\{E[l_i^R(b)] : b\in\mathcal{B}\}$. Then, the conditional maximum likelihood estimator (CMLE) of β is defined as

$$\hat{\beta} = \arg\max_{b\in\mathcal{B}} \frac{1}{N}\sum_{i=1}^{N} l_i^R(b)$$

or, alternatively, as the solution of the nonlinear equation $(1/N)\sum_{i=1}^{N} s_i^R(\hat{\beta}) = 0$. By standard arguments (see e.g. Wooldridge (2010), pp. 476-479), we obtain the asymptotic distribution:

$$\sqrt{N}\left(\hat{\beta}-\beta\right) \xrightarrow{D} N\left(0, I^R(\beta)^{-1}\right), \tag{2.4}$$

where $\xrightarrow{D}$ indicates convergence in distribution. Under additional regularity conditions –such as finite fourth moments–[2], the asymptotic variance can be consistently estimated by its sample analogue and replacing β with $\hat{\beta}$:

$$\hat{I}^R\left(\hat{\beta}\right)^{-1} = \left\{\sum_{j\in\mathcal{J}_R}\sum_{r=1}^{R}\left[\frac{1}{N}\sum_{i=1}^{N} d_{ij}^R H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x}_i;\hat{\beta})\right]\right\}^{-1}.$$

We are now ready to state the main result of this paper. Consider estimating β using a ranking of length $\tilde{R}$ with $1 \le \tilde{R} < R < J$.[3] Let $\tilde{\beta}$ and $I^{\tilde{R}}(\beta)$ denote the corresponding CMLE and Fisher information matrix, respectively. To be specific,

$$I^{\tilde{R}}(\beta) = E\left[\sum_{j\in\mathcal{J}_{\tilde{R}}} d_{ij}^{\tilde{R}}\sum_{r=1}^{\tilde{R}} H_{\{j_1,\ldots,j_r\}}(\mathbf{x}_i;\beta)\right],$$

where $d_{ij}^{\tilde{R}} = 1\{y_i^{\tilde{R}} = j\}$ with $j \in \mathcal{J}_{\tilde{R}}$ and $y_i^{\tilde{R}}$ stands for the $\tilde{R}$-tuple chosen by individual i from $\mathcal{J}_{\tilde{R}}$. By Lemma 2.1 and expression (2.4), $I^{\tilde{R}}(\beta)$ is positive definite and $\sqrt{N}\left(\tilde{\beta}-\beta\right) \xrightarrow{D} N\left(0, I^{\tilde{R}}(\beta)^{-1}\right)$. The next theorem shows analytically that asymptotic efficiency increases with the length of the ranking.

Theorem 2.1 *Under Assumptions 2.1-2.3,*

$$I^{\tilde{R}}(\beta)^{-1} - I^R(\beta)^{-1}$$

is positive definite for every $1 \le \tilde{R} < R < J$.

Proof. See Appendix A.2. ■

Theorem (2.1) states that for efficiency reasons we should use the longest ranking we possibly can to estimate β. Furthermore, this theorem has implications for sample design. Efficiency gain of using a ranking of length R rather than a ranking of length $\tilde{R}$ can be estimated using only

[2] Specifically, $E(|x_{i,l_1}^{(d_1)}x_{i,l_2}^{(d_2)}x_{i,l_3}^{(d_3)}x_{i,l_4}^{(d_4)}|) < +\infty$ for every $(d_1, d_2, d_3, d_4, l_1, l_2, l_3, l_4) \in \{1,\ldots,D\}^4 \times \mathcal{J}^4$.

[3] Note that $\tilde{R} = 1$ corresponds to the traditional multinomial Logit model.

the shortest ranking. Suppose a researcher considers using resources to increase the amount of information collected from individuals such that their R ranking is collected rather than their $\tilde{R}$ ranking ($\tilde{R} < R$). The researcher can estimate the expected efficiency gains from collecting additional information based on the sample she has at hand. For example, the researcher can use a pilot sample to compute the expected gains before conducting the full survey. We suggest here a method for doing that.

Since $E\left[d_{ij}^{R} H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x}_i;\beta)\right] = E\left[P(d_{ij}^{R}=1|\mathbf{x}_i;\beta) H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x}_i;\beta)\right]$ by the law of iterated expectations, the efficiency gain can be estimated in 3 steps:

Step 1 Compute $\tilde{\beta}$ using a ranking of length $\tilde{R}$, $\tilde{\beta} = \arg\max_{b\in\mathcal{B}} \frac{1}{N}\sum_{i=1}^{N} l_i^{\tilde{R}}(b)$, and estimate the asymptotic variance by $I^{\tilde{R}}(\beta)^{-1}$ by

$$\hat{I}^{\tilde{R}}\left(\tilde{\beta}\right)^{-1} = \left\{\sum_{j\in\mathcal{J}_{\tilde{R}}}\sum_{r=1}^{\tilde{R}}\left[\frac{1}{N}\sum_{i=1}^{N} d_{ij}^{\tilde{R}} H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x}_i;\tilde{\beta})\right]\right\}^{-1}.$$

Step 2 For each and $j \in \mathcal{J}_R$, estimate $P(d_{ij}^{R}=1|\mathbf{x}_i;\beta)$ from expression (2.2) and replacing β with $\tilde{\beta}$:

$$\tilde{P}_{ij}^{R} = \prod_{r=1}^{R}\frac{\exp(x'_{i,j_r}\tilde{\beta})}{\sum_{l\in\mathcal{J}\setminus\{j_1,\ldots,j_{r-1}\}}\exp(x'_{i,l}\tilde{\beta})}.$$

Step 3 Estimate $I^{\tilde{R}}(\beta)^{-1}$ by

$$\tilde{I}^{R}\left(\tilde{\beta}\right)^{-1} = \left\{\sum_{j\in\mathcal{J}_{\tilde{R}}}\sum_{r=1}^{\tilde{R}}\left[\frac{1}{N}\sum_{i=1}^{N} \tilde{P}_{ij}^{R} H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x}_i;\tilde{\beta})\right]\right\}^{-1}.$$

Then, the efficiency gain is given by $\hat{I}^{\tilde{R}}(\tilde{\beta})^{-1} - \tilde{I}^{R}(\tilde{\beta})^{-1}$.

3 Monte Carlo Experiments

In this section we study the efficiency gains in small samples. The data are generated from the random utility model

$$u_{i,l} = \beta_1 x_{i,l}^{(1)} + \beta_2 x_{i,l}^{(2)} + \varepsilon_{i,l}$$

with $i = 1,\ldots,N$ and $l = 1,\ldots,J$. Individual i chooses alternative j_1 over j_2 if and only if $u_{i,j_1} > u_{i,j_2}$, where $j_1, j_2 \in \{1,\ldots,J\}$. The design of the simulations is as follows. The covariates

Table 1: Monte Carlo results –variance–

		Coefficient / Length of ranking (R)									
		$\hat{\beta}_1$					$\hat{\beta}_2$				
N	J	1	2	3	4	5	1	2	3	4	5
100	10	0.173	0.083	0.058	0.044	0.036	0.171	0.081	0.057	0.044	0.034
	15	0.145	0.070	0.046	0.036	0.028	0.138	0.067	0.046	0.034	0.027
	20	0.127	0.065	0.045	0.033	0.027	0.125	0.060	0.038	0.029	0.023
500	10	0.033	0.016	0.011	0.009	0.007	0.031	0.015	0.010	0.008	0.007
	15	0.029	0.015	0.010	0.008	0.006	0.026	0.013	0.009	0.006	0.005
	20	0.025	0.012	0.008	0.006	0.005	0.024	0.012	0.008	0.006	0.005

All figures have been multiplied by 10.

are $x_{i,l}^{(1)} = z_{i,l}^{(1)} + z_{i,l}^{(3)}$ and $x_{i,l}^{(2)} = z_{i,l}^{(2)} + z_{i,l}^{(4)}$, where $z_{i,l}^{(1)} \sim N(0,1)$, $z_{i,l}^{(2)} \sim \text{Uniform}[-2,2]$, and

$$\begin{pmatrix} z_{i,l}^{(3)} \\ z_{i,l}^{(4)} \end{pmatrix} \sim N\left[0, \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}\right].$$

$z_{i,l}^{(1)}$, $z_{i,l}^{(2)}$, and $(z_{i,l}^{(3)}, z_{i,l}^{(4)})'$ are independent between each other and across (i,l). The distribution of the error $\varepsilon_{i,l}$ has a type 1 extreme value distribution. The true values of parameters are $(\beta_1, \beta_2) = (1,1)$. The sample size and number of alternatives are $N \in \{100, 500\}$ and $J \in \{10, 15, 20\}$, respectively. We consider rankings of length $R = 1, \ldots, 5$. The number of replication is 2,000 and, in each replication, (β_1, β_2) is estimated by CMLE for each value $R = 1, \ldots, 5$.

The results are reported in Table 1 and Figure 1. Table 1 shows the variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ obtained in the simulations. As expected, the variance decreases with the length of the ranking.[4] Figure 1 displays the obtained 5th and 95th percentiles of $\hat{\beta}_1$ for the cases $(N, J) = (100, 15)$ and $(N, J) = (500, 15)$. As can be noted, these percentiles approach the true value of the coefficient as the length of the ranking increases.

4 Conclusions

Much attention was devoted by the literature on ranked Logit to the quality of ranking. The main result of this paper shows that the ranked Logit has a smaller asymptotic variance than the regular Logit and therefore is a more efficient estimator. The efficiency gains afforded by using deeper ranking can be computed using the results in Section 2. We show that based on the current sample

[4] The size of the bias is omitted from Table 1 for clarity and is available from the authors upon request. Simulations results suggest that the bias is not affect by the length of the ranking.

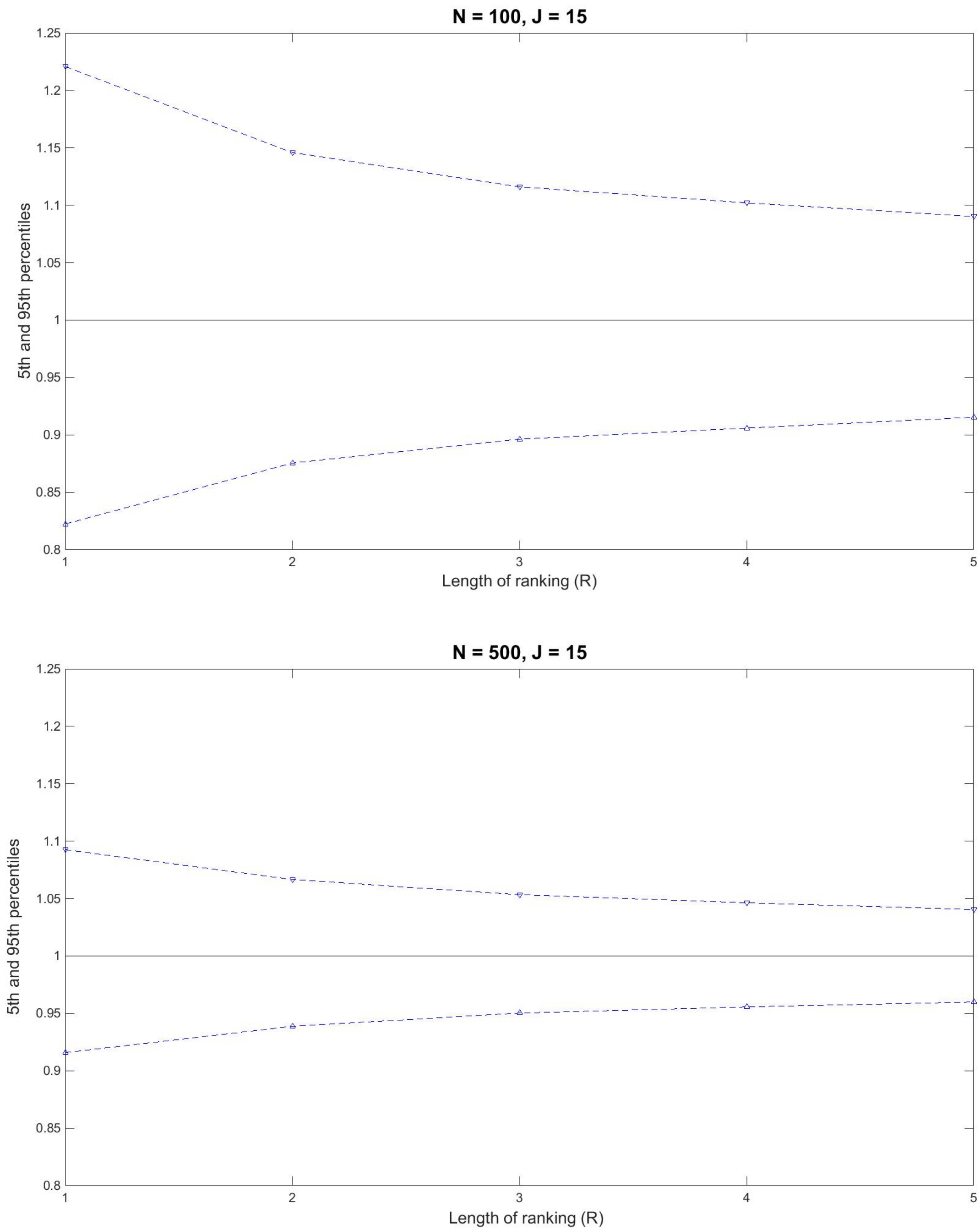

Figure 1: 5th and 95th percentiles of $\hat{\beta}_1$

and before collecting additional information the researcher can estimate the expected efficiency gain should she decide to collect more information. Our Monte Carlo experiment in Section 3, however, demonstrates that these efficiency gains diminish very quickly as we use deeper ranking. We conclude that using the top three choices gives the lion share of efficiency gains while avoiding the concerns raised by previous literature that using lower ranked options may reduce the quality of the estimator.

A Appendix: Proofs

A.1 Proof of Lemma 2.1

1. Pick any $b \in \mathbb{R}^{D\times 1}$. We start with an auxiliary lemma.

Lemma A.1 *Under Assumption 2.2, $H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x};b)$ is positive semi-definite for any value of $\mathbf{x} = (x_1',\ldots,x_J') \in \mathbb{R}^{1\times JD}$, $r \in \{1,\ldots,R\}$, and $\{j_1,\ldots,j_{r-1}\} \subseteq \mathcal{J}$.*

Proof. Pick any $\mathbf{x} = (x_1',\ldots,x_J')$, r, and $\{j_1,\ldots,j_{r-1}\}$. Define the weights

$$w_l(\mathbf{x}) = \frac{\exp(x_l'b)}{\sum_{m\in\mathcal{J}\setminus\{j_1,\ldots,j_{r-1}\}}\exp(x_m'b)}$$

for $l \in \mathcal{J}\setminus\{j_1,\ldots,j_{r-1}\}$ and observe that $w_l(\mathbf{x}) > 0$, as well as, $\sum_{l\in\mathcal{J}\setminus\{j_1,\ldots,j_{r-1}\}} w_l(\mathbf{x}) = 1$. We show next that $c'H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x};b)c \geq 0$ for any $c \in \mathbb{R}^{D\times 1}$: write

$$\begin{aligned}
c'H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x};\beta)c &= \frac{\sum_{l\in\mathcal{J}\setminus\{j_1,\ldots,j_{r-1}\}}\exp(x_l'\beta)(c'x_l)^2}{\sum_{m\in\mathcal{J}\setminus\{j_1,\ldots,j_{r-1}\}}\exp(x_m'\beta)} - \frac{[\sum_{l\in\mathcal{J}\setminus\{j_1,\ldots,j_{r-1}\}}\exp(x_l'\beta)(c'x_l)]^2}{[\sum_{m\in\mathcal{J}\setminus\{j_1,\ldots,j_{r-1}\}}\exp(x_m'\beta)]^2} \\
&= \left[\sum_{l\in\mathcal{J}\setminus\{j_1,\ldots,j_{r-1}\}}(c'x_l)^2 w_l(\mathbf{x})\right] - \left[\sum_{l\in\mathcal{J}\setminus\{j_1,\ldots,j_{r-1}\}}(c'x_l)w_l(\mathbf{x})\right]^2 \\
&= \sum_{l\in\mathcal{J}\setminus\{j_1,\ldots,j_{r-1}\}}\left[(c'x_l) - \mu_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x};c)\right]^2 w_l(\mathbf{x}), \qquad \text{(A.1)}
\end{aligned}$$

where $\mu_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x};c) = \sum_{m\in\mathcal{J}\setminus\{j_1,\ldots,j_{r-1}\}}(c'x_m)w_m(\mathbf{x})$. Note that the last expression is clearly nonnegative. ∎

Before proceeding, we remark that $E\left[d_{ij}^R H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x}_i;b)\right]$ is finite (Assumption 2.2.2) and positive semi-definite (Lemma A.1) for every $r = 1,\ldots,R$. Consequently, $I^R(b)$ is positive semi-definite. We show next that

$$E\left[\sum_{j\in\mathcal{J}_R} d_{ij}^R H_{\{j_1,\ldots,j_{R-1}\}}(\mathbf{x};b)\right]$$

is positive definite. Pick any $c \in \mathbb{R}^{D\times 1}$ and, using Assumption 2.3, choose $j^* = (j_1^*,\ldots,j_{R-1}^*,j_R^*) \in \mathcal{J}_R$ so that $\bar{l}_1,\bar{l}_2 \in \mathcal{J}\setminus\{j_1^*,\ldots,j_{R-1}^*\}$; this is possible because $R-1 \leq J-2$. Note that

$$0 < E\left\{\left[c'\left(x_{i,\bar{l}_1} - x_{i,\bar{l}_2}\right)\right]^2\right\} < +\infty$$

by Assumption 2.3 and write

$$E\left\{\left[c'\left(x_{i,\bar{l}_1} - x_{i,\bar{l}_2}\right)\right]^2\right\} = \int_{\mathcal{D}}\left[c'(x_{\bar{l}_1} - x_{\bar{l}_2})\right]^2 f(\mathbf{x})\nu(d\mathbf{x}),$$

where $\mathcal{D} = \{\mathbf{x} = (x_1', \ldots, x_J') \in \mathbb{R}^{1\times JD} : |c'(x_{\bar{l}_1} - x_{\bar{l}_2})| > 0\}$. Denote

$$\mathcal{D}_s = \left\{\mathbf{x} \in \mathbb{R}^{1\times JD} : |c'(x_{\bar{l}_1} - x_{\bar{l}_2})| \geq \frac{1}{s}, \|(x_1', \ldots, x_J')\|_\infty \leq s\right\}$$

for $s \in \mathbb{N}$, being $\|\cdot\|_\infty$ the sup-norm of a vector, and observe that $\mathcal{D} = \cup_{s\in\mathbb{N}}\mathcal{D}_s$. Since

$$\lim_{s\to+\infty} \int_{\mathcal{D}_s} \left[c'(x_{\bar{l}_1} - x_{\bar{l}_2})\right]^2 f(\mathbf{x})\nu(d\mathbf{x}) = \int_{\mathcal{D}} \left[c'(x_{\bar{l}_1} - x_{\bar{l}_2})\right]^2 f(\mathbf{x})\nu(d\mathbf{x}) > 0$$

by the Lebesgue's dominated convergence theorem (see e.g., Theorem 16.4 in Billingsley (1995)), there is $s^* \in \mathbb{N}$ such that

$$\int_{\mathcal{D}_{s^*}} \left[c'(x_{\bar{l}_1} - x_{\bar{l}_2})\right]^2 f(\mathbf{x})\nu(d\mathbf{x}) > 0,$$

which implies $\int_{\mathcal{D}_{s^*}} f(\mathbf{x})\nu(d\mathbf{x}) > 0$. Now define the sets

$$\mathcal{D}_{s^*,l} = \left\{\mathbf{x} \in \mathbb{R}^{1\times JD} : \left|(c'x_{\bar{l}_l}) - \mu_{\{j_1^*,\ldots,j_{R-1}^*\}}(\mathbf{x}; c)\right| \geq \frac{1}{2s^*}, \|(x_1', \ldots, x_J')\|_\infty \leq s^*\right\}$$

for $l = 1, 2$ that satisfy $\mathcal{D}_{s^*} \subseteq \mathcal{D}_{s^*,1} \cup \mathcal{D}_{s^*,2}$ because

$$|c'(x_{\bar{l}_1} - x_{\bar{l}_2})| \leq \left|(c'x_{\bar{l}_1}) - \mu_{\{j_1^*,\ldots,j_{R-1}^*\}}(\mathbf{x}; c)\right| + \left|(c'x_{\bar{l}_2}) - \mu_{\{j_1^*,\ldots,j_{R-1}^*\}}(\mathbf{x}; c)\right|.$$

As a consequence,

$$0 < \int_{\mathcal{D}_{s^*}} f(\mathbf{x})\nu(d\mathbf{x}) \leq \int_{\mathcal{D}_{s^*,1}} f(\mathbf{x})\nu(d\mathbf{x}) + \int_{\mathcal{D}_{s^*,2}} f(\mathbf{x})\nu(d\mathbf{x}),$$

and without loss of generality we assume $\int_{\mathcal{D}_{s^*,1}} f(\mathbf{x})\nu(d\mathbf{x}) > 0$. To obtain the desired result, write

$$E\left[\sum_{j\in\mathcal{J}_R} d_{ij}^R H_{\{j_1,\ldots,j_{R-1}\}}(\mathbf{x}; b)\right] = E\left[\sum_{j\in\mathcal{J}_R\setminus\{j^*\}} d_{ij}^R H_{\{j_1,\ldots,j_{R-1}\}}(\mathbf{x}; b)\right] + E\left[d_{ij^*}^R H_{\{j_1^*,\ldots,j_{R-1}^*\}}(\mathbf{x}; b)\right]$$

and observe that the second term satisfies

$$\begin{aligned}
&c'E\left[d_{ij^*}^R H_{\{j_1^*,\ldots,j_{R-1}^*\}}(\mathbf{x}_i; b)\right] c \\
&\quad= \int \mathrm{P}(d_{ij^*}^R = 1|\mathbf{x}; \beta)[c'H_{\{j_1^*,\ldots,j_{R-1}^*\}}(\mathbf{x}; b)c] f(\mathbf{x})\nu(d\mathbf{x}) \\
&\quad= \int \mathrm{P}(d_{ij^*}^R = 1|\mathbf{x}; \beta)\left\{\sum_{l\in\mathcal{J}\setminus\{j_1^*,\ldots,j_{R-1}^*\}} \left[(c'x_l) - \mu_{\{j_1^*,\ldots,j_{R-1}^*\}}(\mathbf{x}; c)\right]^2 w_l(\mathbf{x})\right\} f(\mathbf{x})\nu(d\mathbf{x}) \\
&\quad\geq \int_{\mathcal{D}_{s^*,1}} \mathrm{P}(d_{ij^*}^R = 1|\mathbf{x}; \beta)\left[(c'x_{\bar{l}_1}) - \mu_{\{j_1^*,\ldots,j_{R-1}^*\}}(\mathbf{x})\right]^2 w_{\bar{l}_1}(\mathbf{x}) f(\mathbf{x})\nu(d\mathbf{x}) \\
&\quad\geq \left(\frac{1}{2s^*}\right)^2 \int_{\mathcal{D}_{s^*,1}} \mathrm{P}(d_{ij^*}^R = 1|\mathbf{x}; \beta) w_{\bar{l}_1}(\mathbf{x}) f(\mathbf{x})\nu(d\mathbf{x}) \\
&\quad\geq \left(\frac{1}{2s^*}\right)^2 \left[\min_{\mathbf{x}\in\mathcal{D}_{s^*,1}} \mathrm{P}(d_{ij^*}^R = 1|\mathbf{x}; \beta) w_{\bar{l}_1}(\mathbf{x})\right] \int_{\mathcal{D}_{s^*,1}} f(\mathbf{x})\nu(d\mathbf{x}) > 0.
\end{aligned}$$

The first equality follows by the law of iterated expectation and the second by expression (A.1). We highlight that $\min_{\mathbf{x}\in\mathcal{D}_{s^*,1}}[\mathrm{P}(d_{ij^*}^R = 1|\mathbf{x};\beta)w_{\bar{l}_1}(\mathbf{x})]$ is positive because $\mathrm{P}[d_{ij^*}^R = 1|\cdot;\beta]w_{\bar{l}_1}(\cdot)$ is continuous and strictly positive, while $\mathcal{D}_{s^*,1}$ is compact.

To show that $I^R(\beta)$ is positive definite, using expression (2.3), just write

$$I^R(\beta) = \sum_{j\in\mathcal{J}_R}\sum_{r=1}^{R-1} E\left[d_{ij}^R H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x};\beta)\right] + \sum_{j\in\mathcal{J}_R} E\left[d_{ij}^R H_{\{j_1,\ldots,j_{R-1}\}}(\mathbf{x};\beta)\right]. \tag{A.2}$$

Note that the first term is positive semi-definite by Lemma A.1, while the second one is positive definite by the previous result.

2. Since $E[s_i^R(\beta)] = 0$ and $E\left[\partial l_i^R(b)/\partial b\partial b'\right] = -I^R(b)$ is negative semi-definite for every $b \in R^{D\times 1}$, it follows that $\beta \in \arg\max_{b\in\mathcal{B}} E[l_i^R(b)]$.

On the one hand, if Assumption 2.3 holds, $E\left[\partial l_i^R(b)/\partial b\partial b'\right]$ is negative definite for every $b \in R^{D\times 1}$, so $\arg\max_{b\in\mathcal{B}} E[l_i^R(b)]$ is a singleton. On the other hand, if Assumption 2.3 does not hold, there is $c^* \in \mathbb{R}^{D\times 1}\backslash\{0\}$ such that $E\left\{[c^{*\prime}(x_{i,l_1} - x_{i,l_2})]^2\right\} = 0$ for every $l_1, l_2 \in \mathcal{J}$. This implies that $c^{*\prime}x_{i,l_1} = c^{*\prime}x_{i,l_2}$ for every l_1, l_2 with probability 1 (w.p.1), so there is $\bar{c} \in \mathbb{R}$ such that $\bar{c} = c^{*\prime}x_{i,l}$ for every $l \in \mathcal{J}$ w.p.1. Consider the vector $\beta + \lambda c^*$ with $\lambda > 0$ sufficiently small so that $\beta + \lambda c^* \in \text{interior}(\mathcal{B})$; recall that $\beta \in \text{interior}(\mathcal{B})$ by Assumption 2.2.4. To complete the proof, observe that

$$\begin{aligned} l_i^R(\beta + \lambda c^*) &= \sum_{j\in\mathcal{J}_R} d_{ij}^R \sum_{r=1}^R \left\{ x_{i,j_r}'(\beta + \lambda c^*) - \log\left[\sum_{l\in\mathcal{J}\backslash\{j_1,\ldots,j_{r-1}\}} \exp(x_{i,l}'(\beta + \lambda c^*))\right]\right\} \\ &= \sum_{j\in\mathcal{J}_R} d_{ij}^R \sum_{r=1}^R \left\{ x_{i,j_r}'\beta + \lambda\bar{c} - \log\left[\exp(\lambda\bar{c}) \sum_{l\in\mathcal{J}\backslash\{j_1,\ldots,j_{r-1}\}} \exp(x_{i,l}'\beta)\right]\right\} \\ &= l_i^R(\beta) \end{aligned}$$

w.p.1. Then, $E[l_i^R(\beta)] = E[l_i^R(\beta + \lambda c^*)]$ and therefore $\beta + \lambda c^* \in \arg\max_{b\in\mathcal{B}} E[l_i^R(b)]$.

A.2 Proof of Theorem 2.1

We consider $\tilde{R} = R - 1$ as the result for the general case $1 \leq \tilde{R} < R$ follows by induction. Since

$$d_{i(j_1,\ldots,j_{R-1})}^{R-1} = \sum_{l\in\mathcal{J}\backslash\{j_1,\ldots,j_{R-1}\}} d_{i(j_1,\ldots,j_{R-1},l)}^R,$$

we have that

$$
\begin{aligned}
I^{R-1}(\beta) &= \sum_{j \in \mathcal{J}_{R-1}} \sum_{r=1}^{R-1} E\left[d_{ij}^{R-1} H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x};\beta)\right] \\
&= \sum_{r=1}^{R-1} \sum_{j \in \mathcal{J}_{R-1}} E\left[\left(\sum_{l \in \mathcal{J} \setminus \{j_1,\ldots,j_{R-1}\}} d_{i(j_1,\ldots,j_{R-1},l)}^{R}\right) H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x};\beta)\right] \\
&= \sum_{r=1}^{R-1} \sum_{j \in \mathcal{J}_R} E\left[d_{ij}^{R} H_{\{j_1,\ldots,j_{r-1}\}}(\mathbf{x};\beta)\right]
\end{aligned}
$$

because $\sum_{j \in \mathcal{J}_{R-1}} \sum_{l \in \mathcal{J} \setminus \{j_1,\ldots,j_{R-1}\}} = \sum_{j \in \mathcal{J}_R}$. From expression (A.2), it follows that

$$
I^{R-1}(\beta) - I^{R}(\beta) = -\sum_{j \in \mathcal{J}_R} E\left[d_{ij}^{R} H_{\{j_1,\ldots,j_{R-1}\}}(\mathbf{x};\beta)\right].
$$

Since right-hand side is negative definite (Lemma 2.1.1), we have that $I^{R-1}(\beta)^{-1} - I^{R}(\beta)^{-1}$ is positive definite.

References

Beggs, S., S. Cardell, and J. Hausman (1981): "Assessing the Potential demand for Electric Cars," *Journal of Econometrics*, 16, 1–19.

Billingsley, P. (1995): *Probability and Measure*. Wiley-Interscience, New York, 3rd edn.

Chapman, R. G., and R. Staelin (1982): "Exploiting Rank Ordered Choice Set Data within the Stochastic Utility Model," *Journal of Marketing Research*, 19(3), 288–301.

Hausman, J. A., and P. A. Ruud (1987): "Specifying and testing econometric models for rank-ordered data," *Journal of Econometrics*, 34, 83–104.

McFadden, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, pp. 105–142. Academic Press, New York.

McFadden, D. (1984): "Econometric Analysis of Qualitative Response Models," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 2, chap. 24, pp. 1395–1457. Elsevier.

Roth, A. E., and E. Peranson (1999): "The Redesign of the Matching Market for American Physicians: some Engineering Aspect of Economic Design," *The American Economic Review*, 89(4), 748–780.

Train, K. E. (2009): *Discrete Choice Methods with Simulation*. Cambridge University Press, 2nd edn.

van Dijk, B., D. Fok, and R. Paap (2012): "A Rank-Ordered Logit Model with Unobserved Heterogeneity in Ranking Capabilities," *Journal of Applied Econometrics*, 27, 831–846.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, 2nd edn.