ON THE USE OF PROBIT BASED MODELS FOR RANKING DATA ANALYSIS

Gopindra S. Nair
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Tel: 512-471-4535; Email: gopindra.s.nair@gmail.com

Chandra R. Bhat (corresponding author)
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Tel: 512-471-4535; Email: bhat@mail.utexas.edu
and
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Ram M. Pendyala
Arizona State University
School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85287-3005
Tel: 480-727-4587; Email: ram.pendyala@asu.edu

Becky P.Y. Loo
The University of Hong Kong
Department of Geography
The Jockey Club Tower, Centennial Campus
Pokfulam Road, Hong Kong
Tel: +852-3917-7024; Email: bpyloo@hku.hk

William H.K. Lam
The Hong Kong Polytechnic University
Department of Civil and Environmental Engineering
9/F, Block Z, 181 Chatham Road South, Hung Hom
Kowloon, Hong Kong
Tel : +852-2766-6045; Email: william.lam@polyu.edu.hk

ABSTRACT

In consumer surveys, more information per response regarding preferences of alternatives may be obtained if individuals are asked to rank alternatives instead of being asked to select only the most preferred alternative. However, the latter method continues to be the common method of preference elicitation. This is because of the belief that ranking of alternatives is cognitively burdensome. In addition, the limited research on modeling ranking data has been based on the rank ordered logit (ROL) model. In this paper, we show that a rank ordered probit (ROP) model can better utilize ranking data information, and that the prevalent view of ranking data as not being reliable (due to the attenuation of model coefficients with rank depth) may be traced to the use of a misspecified ROL model rather than to any cognitive burden considerations.

Keywords: ranking; rank ordered probit (ROP); rank ordered logit (ROL); heteroscedastic ROL (HROL); heteroscedastic ROP (HROP); coefficient attenuation; stated preference.

1. INTRODUCTION

The preferences of individuals regarding market goods and services are typically imputed in choice models using consumer survey data. In the context of transportation planning, surveys have been extensively used to explore individuals' preferences for travel modes, vehicle type choices and route choices, among many other activity-travel choice dimensions. It is common practice to elicit only the most preferred alternative in these surveys. However, it is possible to elicit not just the preferred alternative but a ranking of all the available alternatives. Although it would appear that the additional information available from the ranking of alternatives should prove beneficial in producing more precise estimates of model coefficients, several past studies (*1–3*) have shown that the rankings provided among the less preferred alternatives appear to be less reliable than the rankings provided among the more preferred alternatives. This unreliability was assumed to be a result of an increased cognitive burden placed on respondents when ranking alternatives with lower preference. In other words, individuals are assumed to be more uncertain when ranking alternatives with lower preferences.

The most commonly used model for ranking data has been the Rank Ordered Logit (ROL) model and its variants. The ROL model is a random utility maximization model that assumes a type 1 extreme value distribution for its utility error kernel term. To account for the hypothesis that rankings of less preferred alternatives, i.e., rankings at higher rank depths, are less reliable, "error scaling parameters" have been introduced into the ROL model to capture the varying uncertainty of individuals at each ranking level.[1] The scaling up of the error terms at higher rank depths would represent increasing uncertainty in the rankings among the lesser preferred alternatives.

A study by Yan and Yoo (*4*) has questioned the notion that the increasing scale as one goes down the rankings (that is, at higher ranking depths) is due to cognitive burden or less reliability in the ranking. They show that the perceived unreliability of rankings of less preferred alternatives can be a result of model misspecification in the ROL model. Since the ROL model assumes its error kernel to follow a type 1 extreme value distribution, the model has a special property that the alternative chosen at any rank level among the unranked alternatives is independent of the rank ordering of the higher ranked alternatives. If the true error kernel is any distribution other than the type 1 extreme value distribution, this property would not hold true. In particular, when a generic distribution of the error kernel is incorrectly constrained to be of the type 1 extreme value distribution, the estimated parameters can mimic a situation of coefficient attenuation across rank depths (that is, increasing uncertainty in utility preferences as the rank depth increases).

The findings by Yan and Yoo (*4*) suggest that it may be worthwhile to explore ranking models that do not rely on the IIA property. The Rank Ordered Probit (ROP) model is another random utility maximization model for ranking data which assumes a normal distribution on its error kernel. In this paper, we perform simulation experiments on the ROP model to evaluate how the model performance varies with rank depth and then compare the ROP model results with the ROL model results. We also extend the ROP model similar to the manner in which the ROL model was extended to incorporate scaling of error terms at different conditional rank depths. The performance of the extended ROP model is compared against the performance of other models in the simulated and empirical datasets.

---

[1] To be clear about the terminology of rank depth, consider a set of three alternatives labeled A, B and C, and assume that a person ranks the alternatives (from most preferred to least preferred) as C, A, and B. Then, the alternative selected at a rank depth of 1 is C and the alternative selected at a rank depth of 2 is A.

2. LITERATURE REVIEW

The ROL model was first developed by Beggs et al. (*5*). This model follows a random utility maximization framework and assumes that the error kernels of the utility functions follow an independent and identically distributed (IID) type 1 extreme value (EV) distribution across alternatives. A consequence of this assumption is that the distribution of utility of the most preferred alternative is independent of the ordering of utilities of less preferred alternatives. This property is a manifestation of the independence of irrelevant alternatives (IIA) property that is associated with logit models. Therefore, if the utilities of alternatives are type 1 EV distributed, the probability of any rank ordering of alternatives can be written as the product of the sequence of probabilities of choosing the most preferred alternative among all the unranked alternatives.

Logically, the coefficients estimated using the ROL model should be the same irrespective of the rank depth that is used in estimation. In other words, the coefficients estimated using only the most preferred choice must be around the same as the coefficients obtained when the estimation is undertaken using the first k ranked choices. However, several previous studies have shown that this is not the case. In fact, it is observed that the coefficients of the ROL model tend to attenuate (move closer to zero) when more rank levels are used for estimation (*2*, *3*). This has fostered a belief that progressively higher cognitive demands are placed on individuals when ranking less preferred alternatives, because of which rankings become less reliable and not consistent with the individual's true underlying utilities as the rank depth increases. To accommodate this higher unreliability of rankings, some studies (such as Hausman and Ruud (*3*) and Foster and Mourato (*2*)) have suggested the use of models where the utility function for alternatives changes with each ranking level. These models make the use of scaling parameters to alter the variance of the error kernel with rank levels. The error kernel is assumed to capture an individual's inability to assess utilities reliably. Therefore, scaling up of the error kernel at higher rank depths is considered to be a sign of decreased ability to rank reliably. This extension of the ROL model that makes use of scale parameters to capture coefficient variations across rank levels is called the heteroscedastic ROL (HROL) model.

In contrast to the prevalent view of cognitive burden considerations with ranking data (based on the observation of coefficient attenuation across conditional rank levels in the ROL), Yan and Yoo (*4*) show, through simulation experiments and computational analyses, that estimates produced by the ROL model can show coefficient attenuation if the true distribution of the utility error term deviates even slightly from the type 1 EV distribution. This is because, if the error term does not follow a type 1 EV distribution, the probability of a ranking pattern can no longer be written as the product (across rank depths) of the probabilities of choosing the most preferred alternative among the unranked alternatives at each rank depth. For a generic distribution of the error term, the probability for selecting an alternative at a rank level must be conditioned on the ordering of alternatives that have already been ranked. Assuming a type 1 EV distribution for the error kernel and using the resulting exploded logit structure ignores this conditioning (except that, when the error kernel is truly a type 1 EV, this conditioning becomes mute). The solution proposed in Yan and Yoo (*4*) to avoid the problem of attenuation of coefficients because of misspecification of the error term distribution is to increase the flexibility of the ROL model by introducing mixing of coefficients and latent classes in addition to scaling parameters at each ranking level. With such a flexible specification, the kernel error will play a lower role in determining the ranking sequence, and therefore the problem of attenuation of coefficients will not be as severe.

The rank ordered probit (ROP) model was developed based on the assumption that the true distribution followed by the utility error terms of alternatives is normal. This model is discussed in Train (*6*). While the ROL model and its variations have been developed upon extensively and used in several empirical contexts (*7–9*), few studies have made use of the ROP model (*10–12*). To the authors' knowledge, Schechter (*11*) is the only paper that applied the ROP model to a stated preference ranking dataset. The reason for the dearth of literature on the ROP model may be because, until the past decade, computing cumulative distribution functions of multivariate normal distributions for evaluation of the ROP model was much more cumbersome than computing logistic distributions for evaluation of the ROL model. However, with recent advancements in analytical (*13*) and simulation methods (*14*) for the approximation of cumulative multivariate normal distribution functions, estimation of a ROP model for the usual modeling contexts encountered in practice should no longer be intractable.

In addition to the dearth of ROP applications, there has been no study that we are aware of that investigates if the problem of unstable coefficients which afflicts the ROL model also prevails for the ROP model. The findings of Yan and Yoo (*4*) suggest that, since the ROP model takes into consideration the dependencies of choices between (the conditioned) rank levels, the effects of unstable coefficients because of misspecification must be much less severe or even non-existent. In this paper, we compare the performance of the ROL and ROP models on simulated datasets in terms of robustness of coefficients to misspecification and goodness of fit across rank depths. Further, we generalize the econometric approach to introduce heteroscedasticity with rank depth through the use of scaling parameters. This approach is used to develop a heteroscedastic version of the rank ordered probit (HROP) model. The tendency for the ROL and ROP models to show coefficient variation across rank levels is studied for the case of two empirical datasets by comparing estimates of the scaling parameters produced by the HROL and HROP models respectively.

The remainder of this paper is organized as follows. Section 3 provides a background of the traditional non-heteroscedastic or homoscedastic ranking models. Section 4 describes the concept behind the development of heteroscedastic ranking models. Further, the ranking probability functions for the HROL and HROP models are derived. Section 5 provides details regarding the simulation experiments conducted to evaluate the different ranking models in terms of coefficient variation and goodness of fit. In Section 6, the ranking models are estimated on two different empirical datasets to investigate whether the insights gained from the simulation studies carry over to the empirical datasets as well. Section 7 concludes the paper.

3. RANKING MODELS

Consider the case of an individual who ranks a set of K different alternatives. Let U_k be the individual's utility for an alternative k $(1,2,...,K)$ expressed as:

$$U_k = \boldsymbol{\beta}'\boldsymbol{x_k} + \xi_k \tag{1}$$

where $\boldsymbol{x_k}$ is a column vector (of dimension $D \times 1$) of individual-level attributes specific to the alternative, $\boldsymbol{\beta}$ is the corresponding column vector of coefficients, and ξ_k is the idiosyncratic random error term. Let $\boldsymbol{r}$ be a specific rank ordering of the alternatives. That is, r^1 is the first alternative, r^2 is the second alternative and so on. $R_{\boldsymbol{r}}$ denotes the event that the alternatives are ranked in the order $\boldsymbol{r}$ by the individual. According to the random utility maximization framework, the probability of $R_{\boldsymbol{r}}$ can be expressed as:

$$P\left(R_{\boldsymbol{r}};\boldsymbol{\beta}\right) = P\left(U_{r^2} - U_{r^1} < 0, U_{r^3} - U_{r^2} < 0, ..., U_{r^K} - U_{r^{K-1}} < 0\right) \tag{2}$$

The above probability can also be considered as the likelihood value for a given ranking observation $\boldsymbol{r}$. When using a likelihood framework for estimating the coefficients of utilities, it is common to consider only a part of the ranking sequence. If only the top d alternatives of an individual are considered during estimation, the resulting probability function of the partial ranking sequence up to a rank depth of d can be expressed as follows:

$$\tilde{P}\left(R_r;\boldsymbol{\beta},d\right)=P\begin{pmatrix}U_{r^2}-U_{r^1}<0,U_{r^3}-U_{r^2}<0,\ldots,U_{r^d}-U_{r^{d-1}}<0,\\ U_{r^{d+1}}-U_{r^d}<0,U_{r^{d+2}}-U_{r^d}<0,\ldots,U_{r^K}-U_{r^d}<0\end{pmatrix} \tag{3}$$

If the rank depth d is set as $K-1$ (because the individual's least preferred alternative is implicitly known once all other alternatives are ranked), Equation (3) is equivalent to Equation (2).

In the ROL model, which assumes the kernel error term ξ_k to follow a type 1 EV distribution, the probability of a ranking sequence up to a rank depth of d reduces to the following equation:

$$\tilde{P}_{ROL}(R_r;\boldsymbol{\beta},d)=\prod_{i=1}^{d}\frac{\exp\left(\boldsymbol{\beta}'\boldsymbol{x}_{r^i}\right)}{\sum_{j=i}^{K}\exp\left(\boldsymbol{\beta}'\boldsymbol{x}_{r^j}\right)} \tag{4}$$

which is the product of probabilities (across ranks up to the rank depth of d) of selecting the most preferred alternative among all the unranked alternatives. The reader is referred to Beggs et al. (*5*) for the complete derivation.

To compute the probability for the ROP model, construct the $K\times D$ matrix $\boldsymbol{x}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_K)'$. Let $\boldsymbol{U}$ ($K\times 1$ vector) be the vector of utility values of alternatives and let $\boldsymbol{\xi}$ ($K\times 1$ vector) be the vector of idiosyncratic error terms associated with the alternatives. Then the vector of utilities $\boldsymbol{U}$ can be expressed as:

$$\boldsymbol{U}=\boldsymbol{x\beta}+\boldsymbol{\xi} \tag{5}$$

Here, $\boldsymbol{\xi}$ is assumed to follow a multivariate normal distribution, with mean **0** and variance **Ω**. Let **M** be the mask matrix of size $\left((K-1)\times K\right)$ which when pre-multiplied with $\boldsymbol{U}$ produces the vector of utility differences that should be less than zero according to Equation (3). To generate the mask matrix **M** corresponding to a ranking sequence $\boldsymbol{r}$ and rank depth d, first generate a matrix of size $\left((K-1)\times K\right)$ filled with zeros. Then, in the first row, place a value of '–1' at the column corresponding to the first ranked alternative and '1' at the column corresponding to the second ranked alternative. Similarly, in the second row, place a value of '–1' at the column corresponding to the second ranked alternative and '1' at the column corresponding to the third ranked alternative. Continue this procedure for d rows. After row d, place '–1' on all rows at the column corresponding to the alternative with rank d. But continue placing '1's at the columns corresponding to the alternative that is ranked one more than the row index. An illustration of the mask matrix for the ranking sequence $(3,5,1,4,2)$ and rank depth 3 is given below.

$$\mathbf{M}=\begin{bmatrix}0 & 0 & -1 & 0 & 1\\ 1 & 0 & 0 & 0 & -1\\ -1 & 0 & 0 & 1 & 0\\ -1 & 1 & 0 & 0 & 0\end{bmatrix}$$

Then the probability of a ranking sequence $\boldsymbol{r}$ up to a rank depth d when using an ROP model is expressed as follows:

$$\tilde{P}_{ROP}(R_r;\boldsymbol{\beta},d)=F_{K-1}\left(\mathbf{0}_{K-1};\mathbf{M}\boldsymbol{x\beta},\mathbf{M\Omega M}'\right) \tag{6}$$

where, $F_{K-1}(\mathbf{0}_{K-1};\boldsymbol{\mu},\mathbf{\Sigma})$ is the K–1 dimensional cumulative multivariate normal distribution function computed at the truncation point vector $\mathbf{0}_{K-1}$ (a vector of zeros of dimension K–1) with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{\Sigma}$.

4. HETEROSCEDASTIC RANKING MODELS

The process of ranking alternatives can be considered as a series of choice decisions in which an individual selects, conditional on earlier choices, the best alternative among all the unranked alternatives (*15*). Let $S_{\boldsymbol{r}}^{r^i}$ denote an event where the individual selects alternative r^i from an ordered set of alternatives $\boldsymbol{r}$.

$$P(S_{\boldsymbol{r}}^{r^i};\boldsymbol{\beta})=P\begin{pmatrix}U_{r^1}-U_{r^i}<0,U_{r^2}-U_{r^i}<0,...,U_{r^{i-1}}-U_{r^i}<0,\\ U_{r^{i+1}}-U_{r^i}<0,U_{r^{i+2}}-U_{r^i}<0,...,U_{r^K}-U_{r^i}<0\end{pmatrix} \tag{7}$$

Let $\boldsymbol{r}^{i:j}$ denote the vector of alternatives between and including the i^{th} and j^{th} alternatives in a ranking $\boldsymbol{r}$ of the K alternatives, with the convention that $\boldsymbol{r}^{1:K}=\boldsymbol{r}$. Equation (3) can be written in the following manner to better reflect the ranking process as a series of conditional single choice decisions.

$$\tilde{P}\left(R_r;\boldsymbol{\beta},d\right)=P\left(S_{\boldsymbol{r}}^{r^1};\boldsymbol{\beta}\right)\prod_{l=2}^{d}P\left(S_{\boldsymbol{r}^{l:K}}^{r^l}\mid S_{\boldsymbol{r}}^{r^1},S_{\boldsymbol{r}^{2:K}}^{r^2},S_{\boldsymbol{r}^{3:K}}^{r^3},...,S_{\boldsymbol{r}^{l-1:K}}^{r^{l-1}};\boldsymbol{\beta}\right) \tag{8}$$

In Equation (8), $P\left(S_{\boldsymbol{r}^{l:K}}^{r^l}\mid S_{\boldsymbol{r}}^{r^1},S_{\boldsymbol{r}^{2:K}}^{r^2},S_{\boldsymbol{r}^{3:K}}^{r^3},...,S_{\boldsymbol{r}^{l-1:K}}^{r^{l-1}};\boldsymbol{\beta}\right)$ denotes the probability that r^l is selected as the best alternative among the unranked alternatives $\boldsymbol{r}^{l:k}$, given that the individual has selected $\boldsymbol{r}^{1:l-1}$ as the first $l-1$ alternatives. The conditioning is required because the unobserved factors that affect the individual's choice of the first $l-1$ alternatives may also affect the choice of the l^{th} alternative.

To account for the phenomenon of attenuation of coefficients with rank depth, it is hypothesized that individuals find it difficult to rank alternatives at higher rank depths reliably. This hypothesis is incorporated into ranking models by multiplying scaling parameters with the attribute coefficients at each rank level. The resulting ranking models are referred to as heteroscedastic ranking models and its general form is as shown below.

$$\tilde{P}_{HR}\left(R_r;\boldsymbol{\beta},d,\boldsymbol{\mu}\right)=P\left(S_{\boldsymbol{r}}^{r^1};\mu_1\boldsymbol{\beta}\right)\prod_{l=2}^{d}P\left(S_{\boldsymbol{r}^{l:K}}^{r^l}\mid S_{\boldsymbol{r}}^{r^1},S_{\boldsymbol{r}^{2:K}}^{r^2},S_{\boldsymbol{r}^{3:K}}^{r^3},...,S_{\boldsymbol{r}^{l-1:K}}^{r^{l-1}};\mu_l\boldsymbol{\beta}\right) \tag{9}$$

The scaling parameter μ_l effectively controls the relative contributions of the exogenous parameters and the idiosyncratic error term in the overall utility function. A lower value for μ_l would introduce higher variability in the utilities of alternatives at rank level l, while a higher value of μ_l would make the utilities more deterministic. Therefore, if the hypothesis that individuals rank alternatives at higher rank depths less reliably is true, the value of μ_l should be lower for larger l. To avoid issues with identification of coefficients, μ_1 is fixed to 1 during estimation. To ensure that the scaling parameter remains positive, it may be reparametrized as $\mu_l = \exp(f_l)$. Note that multiplying the utility coefficients by μ_l has the same effect as dividing the covariance matrix of the error terms by μ_l^2.

The conditional probability of selection of an unranked alternative at any given rank depth can be expressed in terms of the ranking probabilities as follows:

$$\begin{aligned} P\left(S_{\boldsymbol{r}^{l:K}}^{r^l} \mid S_{\boldsymbol{r}}^{r^1}, S_{\boldsymbol{r}^{2:K}}^{r^2}, S_{\boldsymbol{r}^{3:K}}^{r^3}, \ldots, S_{\boldsymbol{r}^{l-1:K}}^{r^{l-1}}; \mu_l \boldsymbol{\beta}\right) &= \frac{P\left(S_{\boldsymbol{r}}^{r^1}, S_{\boldsymbol{r}^{2:K}}^{r^2}, S_{\boldsymbol{r}^{3:K}}^{r^3}, \ldots, S_{\boldsymbol{r}^{l-1:K}}^{r^{l-1}}, S_{\boldsymbol{r}^{l:K}}^{r^l}; \mu_l \boldsymbol{\beta}\right)}{P\left(S_{\boldsymbol{r}}^{r^1}, S_{\boldsymbol{r}^{2:K}}^{r^2}, S_{\boldsymbol{r}^{3:K}}^{r^3}, \ldots, S_{\boldsymbol{r}^{l-1:K}}^{r^{l-1}}; \mu_l \boldsymbol{\beta}\right)} \\ &= \frac{\tilde{P}\left(R_{\boldsymbol{r}}; \mu_l \boldsymbol{\beta}, l\right)}{\tilde{P}\left(R_{\boldsymbol{r}}; \mu_l \boldsymbol{\beta}, l-1\right)} \end{aligned} \tag{10}$$

Equation (9) on application of Equation (10) becomes:

$$\tilde{P}_{HR}\left(R_{\boldsymbol{r}}; \boldsymbol{\beta}, d, \boldsymbol{\mu}\right) = \tilde{P}\left(R_{\boldsymbol{r}}; \mu_1 \boldsymbol{\beta}, 1\right) \prod_{l=2}^{d} \frac{\tilde{P}\left(R_{\boldsymbol{r}}; \mu_l \boldsymbol{\beta}, l\right)}{\tilde{P}\left(R_{\boldsymbol{r}}; \mu_l \boldsymbol{\beta}, l-1\right)} \tag{11}$$

Equation (11) can be used to extend the ROL model to produce the heteroscedastic ROL (HROL). Substituting the generic ranking probability function $\tilde{P}\left(R_{\boldsymbol{r}}; \mu_l \boldsymbol{\beta}, l\right)$ in Equation (11) with that of the ROL model, $\tilde{P}_{ROL}$ (from Equation (4)), the probability function for the HROL model is as follows:

$$\tilde{P}_{HROL}(R_{\boldsymbol{r}}; \boldsymbol{\beta}, d, \boldsymbol{\mu}) = \prod_{i=1}^{d} \frac{\exp\left(\mu_i \boldsymbol{\beta}' \boldsymbol{x}_{r^i}\right)}{\sum_{j=i}^{K} \exp\left(\mu_i \boldsymbol{\beta}' \boldsymbol{x}_{r^j}\right)} \tag{12}$$

As mentioned in Section 2, Equation (12) can be viewed as the product of probabilities of choosing the best alternative among the unranked alternatives independent of the alternatives that have already been chosen. This is an artifact of the IIA property of logit based models. In other words, if the errors are assumed to be type 1 EV distributed,

$$P\left(S_{\boldsymbol{r}^{l:K}}^{r^l} \mid S_{\boldsymbol{r}}^{r^1}, S_{\boldsymbol{r}^{2:K}}^{r^2}, S_{\boldsymbol{r}^{3:K}}^{r^3}, \ldots, S_{\boldsymbol{r}^{l-1:K}}^{r^{l-1}}; \mu_l \boldsymbol{\beta}\right) = \tilde{P}_{MNL}\left(S_{\boldsymbol{r}^{l:K}}^{r^l}; \mu_l \boldsymbol{\beta}\right) = \frac{\exp\left(\mu_l \boldsymbol{\beta}' \boldsymbol{x}_{r^l}\right)}{\sum_{j=l}^{K} \exp\left(\mu_l \boldsymbol{\beta}' \boldsymbol{x}_{r^j}\right)} \tag{13}$$

where, $\tilde{P}_{MNL}\left(S_{\boldsymbol{r}^{l:K}}^{r^l}; \mu_l \boldsymbol{\beta}\right)$ denotes the probability of selection of r^l from alternatives $\boldsymbol{r}^{l:K}$ when the utility coefficients are $\mu_l \boldsymbol{\beta}$ and the error terms follow a type 1 EV distribution.

To the authors' knowledge, heteroscedasticity using scale parameters has not been introduced to any ranking model other than ROL. However, the same concept can also be extended to the ROP model to produce the Heteroscedastic ROP (HROP) model. Replacing the generic ranking probability function $\tilde{P}\left(R_{\boldsymbol{r}}; \mu_l \boldsymbol{\beta}, l\right)$ in Equation (11) with that of the ROP model, P_{ROP} (from Equation (6)), the probability function for the HROP model is as follows:

$$\tilde{P}_{HROP}\left(R_{\boldsymbol{r}}; \boldsymbol{\beta}, d, \boldsymbol{\mu}\right) = \tilde{P}_{ROP}\left(R_{\boldsymbol{r}}; \mu_1 \boldsymbol{\beta}, 1\right) \prod_{l=2}^{d} \frac{\tilde{P}_{ROP}\left(R_{\boldsymbol{r}}; \mu_l \boldsymbol{\beta}, l\right)}{\tilde{P}_{ROP}\left(R_{\boldsymbol{r}}; \mu_l \boldsymbol{\beta}, l-1\right)} \tag{14}$$

Unlike the HROL model, the HROP model cannot be simplified to a sequence of independent single choice decisions.

All the heteroscedastic and homoscedastic ranking models discussed in this paper were estimated using the GAUSS matrix programming language, and are available at http://www.caee.utexas.edu/prof/bhat/CodeRepository/CODES/Ranking/Code_Release.zip.

5. TEST FOR COEFFICIENT ATTENUATION IN ROP MODELS

In this section, we compare the extent of coefficient attenuation and goodness of fit for the ROL and ROP models. The comparison is performed on simulated datasets. Details regarding the development of the simulated datasets are provided in Section 5.1, and the results and interpretations from the experiments are presented in Sections 5.2 and 5.3.

5.1 Experimental Setup

The objective of the simulation experiments is to understand the robustness of coefficients obtained from the different ranking models to misspecification of the kernel error term. By robustness of coefficients, we refer to the property of lack of variation in estimated coefficients across rank depths. To focus on the issue of kernel distribution misspecification, we consider strictly IID error terms across alternatives (so that covariance across error terms of alternatives does not add another dimension of misspecification impacting variation in estimated coefficients across rank depths). The data generation process used for the generation of test datasets is the same as Equation (1). In all our experiments we assume 8 alternatives and 2 attributes for each alternative. The alternative specific constant term for the first alternative is set to zero, and that for all other alternatives is set to one. The value of coefficients for the two attributes are set as 1 and –1. We consider 4 different distributions for the IID kernel error terms. They are the normal distribution, type 1 extreme value distribution, uniform distribution and logistic distribution. The parameters of these distributions were set in a way that the mean is zero and the variance is $\pi^2/6$ for all distributions. The values of attributes for each observation are drawn from normal distributions with a variance of one. The mean of normal distributions for the first attribute of the 8 alternatives ranged linearly from 0.5 for the first alternative to –0.5 for the last alternative. This trend was reversed for the distributions of the second attribute. The distributions of attributes were set in this manner to ensure that there are variations in the number of observations that select each alternative as their first choice.

All coefficients are estimated using the maximum likelihood method. The one variate univariate screening method (*13*) is used to evaluate the cumulative multivariate normal distribution function for the probit based models.

5.2 Experimental Results: Robustness of Coefficients

To evaluate the degree of coefficient attenuation, 50 datasets of size 500 were generated for each distribution of the utility error term. For each dataset, coefficients were estimated using the different ranking models at rank depths varying from 1 to 7. In other words, a total of 4 (error distributions) × 50 (datasets) × 4 (ranking models) × 7 (rank depths) = 5600 models were estimated. A box plot of the estimated coefficient of the first attribute (with a coefficient of 1 in the data generation process) is provided in Figure 1. The box plots for the second attribute (with a coefficient of –1 in the data generation process) is not shown here as it is almost the same as that for the first attribute except for a reversal in sign of the estimated coefficients. Note that the variation of coefficient values presented in the box plot in Figure 1 represent the different estimates obtained across the 50 datasets.

For the models generated from datasets having error distributions as normal, uniform or logistic, the extent of coefficient attenuation with the ROP model seems to be much lower than that of the ROL model. For these datasets, the advantage of using a heteroscedastic ranking model over the ROP model seems to be relatively less. The coefficient attenuation shown by the ROL model appears similar to what would be expected if the rankings at higher rank depths were made less reliably. This suggests that it may be possible to rectify the problem of attenuation of coefficients observed in previous literature by using the ROP model instead of the ROL model.

In the datasets where the utility follows type 1 EV distribution, the coefficients of the ROL model does not attenuate, but that of the ROP model is amplified. None of the heteroscedastic ranking models show any variation in coefficient parameters with rank depth. This is expected because the variation of coefficients is captured by the scaling parameters in these models. However, the robustness of coefficients in heteroscedastic ranking models comes at the cost of higher variance (determined by the length of the boxes and whiskers) at higher rank depths. The reduction of variance of heteroscedastic models with rank depth seems to be lower than that of homoscedastic models.

Overall, among the homoscedastic models, the ROP model seems to be a better alternative to the ROL model if one is not sure about the distribution of the error term. The probit kernel is much more accommodative and robust to misspecification of the kernel error term. Both the heteroscedastic models do not show significant coefficient attenuation for any distribution of the kernel error term.

5.3 Experimental Results: Goodness of Fit

Coefficient attenuation by itself is not necessarily problematic if it improves the predictive power of a model. The maximum likelihood procedure ensures that the coefficients maximize the probability of observed rankings up to the rank depth used for estimation. However, in most cases where ranking data is used, the true objective is not to be able to predict the probability of an individual's ranking of all alternatives, but to predict the probability of an alternative being selected as the most preferred alternative. The idea behind using the ranking data is to use information of choices at lower rank levels to produce more precise estimates of the probability of selection of the most preferred alternative. Therefore, in this section, we assess the ability of the different models to predict the probability of the most preferred alternative when different rank depths are used.

To evaluate the predictive power of the models estimated in the previous step, test datasets with 1000 observations were generated for each of the utility error distributions. The likelihood of the most preferred alternative in the test dataset was computed for the 5600 models estimated in the previous step. The models were tested on the dataset having the same distribution as the dataset on which the model is estimated. Figure 2 shows a plot of the computed likelihood values. The likelihood value plotted is the average of the likelihoods produced by ranking models estimated using the 50 datasets. The likelihood of the most preferred alternatives in the test dataset is considered to be a metric for goodness of fit or predictive capability.

A comparison between the corresponding plots in Figure 1 and Figure 2 indicates the goodness of fit improves with rank depth for models that showed robust coefficients in Figure 1. The ROL model shows a drop in the goodness of fit when the error terms follow normal, uniform or logistic distribution. The likelihood value of the ROP model deteriorates when the error term is type 1 EV distributed. All heteroscedastic ranking models show improvement in goodness of fit with rank depth. Overall, the plots of goodness of fit reinforce our finding that the ROP model is more robust to the misspecification of distribution of error terms. It should not be surprising that the ROL model performs better than the ROP model when the utilities in the datasets are type 1 EV distributed since this is the same as the utility distribution assumed by the ROL model.

In the broader context, these results further corroborate Yan and Yoo (*4*) that the attenuation of coefficients observed in past studies on the ROL model may be a result of misspecification and not because of inconsistent ranking of alternatives at higher rank depths. The ROP model did not show attenuation of coefficients in three of the four error distributions that were tested. The robustness of the ROP model in our simulated datasets means that, in the context of empirical datasets of stated preference surveys, if the ROP model shows coefficient attenuation with rank depth, or equivalently if estimated values of scale parameters in the HROP model are significantly less than one, this would be a better indication that individuals in fact rank alternatives with lower preference less reliably (while making this same conclusion based on a ROL model is more dubious because of confounding of misspecification in the kernel error distribution with less reliability in the rankings at higher depth).

6. APPLICATION

In this section, we compare the extent of coefficient attenuation and goodness of fit of logit based models and probit based models when the estimation is undertaken on empirical datasets. Coefficient attenuation is captured in the heteroscedastic ranking models through the logarithm of the scale parameter at each rank depth except the last (that is, $f_l = \log(\mu_l)$ $(l = 1, 2, ...K-1)$, where K is the number of alternatives and $f_1 = 0$ as an innocuous normalization) (see Section 4). This term is an indicator of the extent of coefficient variation at each rank level. A negative value for the log scale parameter would indicate coefficient attenuation (that is lower reliability of ranking at higher depths), while a positive value implies that the coefficients are amplified (or higher reliability of ranking at higher depths). Insignificance or a value close to zero for this term at a particular rank level indicates that the coefficient at that rank level is the same as the coefficient at the first rank level where the log scale parameter is fixed to zero. The variance of the error term is fixed to $\pi^2/6$ in all models.

For each of the estimated ranking models, the log-likelihood value at convergence and the adjusted likelihood ratio index (ADLRI) are computed. In computing the ADLRI, we use the likelihood at convergence of the constants only model for the ROL model as the common basis to evaluate the alternative models. The performance of the probit based models are compared against the corresponding logit based models through the use of the non-nested adjusted likelihood ratio test, which determines if the ADLRIs of two non-nested models are significantly different. In other words, this statistic gives the probability that the difference between the ADLRI statistic of two models occurred because of random chance. Within each of the probit-based and logit-based model categories, the performance of the homoscedastic model is compared against its corresponding heteroscedastic counterpart using a nested likelihood ratio test. Additionally, the goodness of fit in predicting the highest and lowest ranked alternatives is computed. That is, using the estimates from each ranked model estimation, the probability of the observed first ranked choice and the probability of the observed last ranked choice for each individual in the sample is computed. Then, the corresponding log-likelihood values and the average probability of correct predictions (across all individuals) are computed. These additional exercises are undertaken to obtain a sense of how the ranked model estimation performs in predicting only the first-ranked choice and only the last-ranked choice (discarding performance at the intermediate ranks). Note, however, that it is not possible to use any rigorous statistical tests to compare performance for the first-ranked choice and last-ranked choice predictions, because the model estimations themselves are undertaken using the full ranking order. But the log-likelihood values and the average probability of correct prediction values serve as informal measures of fit.

6.1 Empirical Example 1

In this section, we analyze the data on ranking of gaming consoles by 91 Dutch students. The students were asked to consider buying a new gaming platform on which to play computer games. The six gaming platforms available were Xbox, PlayStation, Gamecube, PlayStation Portable, Gameboy or a regular personal computer. For more information on this dataset, the reader is referred to van Dijk et al. (*16*). The dataset is available as part of the R package *mlogit* (*17*).

The specification used for modeling is the same as that used by van Dijk et al. (*16*). The exogenous variables are the number of hours spent playing and a binary variable indicating whether the individual owns (or not) the gaming platform under consideration. To conserve on space, and also because the substantive variable effects themselves are not of primary importance in this paper, we do not present the complete model results (suffice it to say though that all the variable effects for all models were as expected and intuitive). Readers interested in the complete model results will find these in an online supplement to this paper available at http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Ranking/OnlineSupplement.pdf. But we should state here that, in terms of the estimated logarithms of the ranking scale parameters, these were statistically significant at the 0.10 level of significance or lower at rank levels two, four, and five for the HROL model, but statistically insignificant at literally any reasonable level for the HROP model. Also, the magnitudes of the estimated logarithm of scale parameters were all negative at each rank level and generally increasing in magnitude with increasing rank depth for both the HROL and HROP models, indicating potentially less reliability for lower ranked alternatives. In addition, the magnitudes at each rank level were higher in the HROL model compared to the HROP model. All of these results, taken together, very strongly imply that the coefficient attenuation observed in the ROL model is a result of model misspecification rather than unreliable rankings.

The model fit statistics and the statistics for comparison between the different models is provided in Table 1. The non-nested likelihood ratio tests between the ROL and ROP, and the HROL and HROP, show that the superiority in performance of the probit kernel over the logit kernel is significant. The nested likelihood ratio test between the ROL and HROL models indicates that the superior performance of the latter is statistically significant at about the 0.075 level of significance. There is no difference between the ROP and HROP models at any reasonable level of significance. In fact, the ADLRI value for the HROP is worse than that of the ROP, because the HROP adds another four parameters to the ROP with little benefit in prediction.

With regard to the prediction of the observed first choice, once again, the ROP model dominates the ROL model. There is literally no difference in performance between the HROL and HROP models, and also literally no difference in performance between the ROP and HROP models. For the prediction of the last observed choice, the probit-based models perform better than their logit-based counterparts on the log-likelihood measure, though the ROL performs just a little better than the ROP based on the average probability of correct prediction. More important to note is that the heteroscedastic models perform worse here than their homoscedastic counterparts, especially when comparing the two logit-based models. This poor performance of the heteroscedastic models for the last choice prediction is not surprising, given the larger error variance at increasing rank depths.

6.2 Empirical Example 2

In this section, we explore the ranked preferences for buses of public light bus operators obtained from a stated preference (SP) survey conducted in 2002 in Hong Kong. The objective of the SP survey was to gauge the interest of operators in buying LPG powered buses. The survey included an SP game in which each respondent was asked to rank between four hypothetical alternatives for buses. The alternatives were described using the following attributes, fuel type (diesel or LPG powered), fuel price, vehicle price, distance to the nearest refueling station, maximum distance the vehicle can travel between refueling stops (vehicle range), life of the vehicle, number of seats and horsepower of the vehicle. The dataset consisted of 903 valid observations. The reader is referred to Loo et al. (*18*) for further details regarding the survey and Loo et al. (*19*) for a detailed analysis of the dataset. The coefficients estimated by the different ranking models are provided in an online supplement available at http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Ranking/OnlineSupplement.pdf. In this second empirical example (and similar to the first), the estimated logarithms of the ranking scale parameters were negative and increasing in magnitude at higher rank depths. But, unlike the first empirical example, the logarithm of scale parameters were statistically significant at the 0.05 level of significance even in the HROP model (and at a much higher level of significance for the HROL model). But, similar to the first example, the magnitudes of these scale parameters were much higher in the HROL model relative to the HROP model. Overall, the implication is that there may be some degradation in reliability at higher rank depths, but the HROL substantially overestimates any such degradation.

The model fit statistics and the statistics for comparison between the different models are presented in Table 1. The non-nested likelihood ratio tests show again that the probit-based models outperform their logit counterparts. Also, introducing heteroscedasticity helps in both the probit and logit cases, though it improves the ROL model much more than the ROP model. In terms of the ability of the models to predict the most preferred alternative, the probit-based models outperform their logit counterparts, and introducing heteroscedasticity substantially helps in the logit case but relatively less so for the probit case. For predicting the least preferred alternative, the ROP model is superior to all other models in terms of log-likelihood, although the probability of correct prediction is slightly higher for the ROL model.

Across both the datasets, for the full ranking and first choice predictions, our results indicate that the simple ROP model does better than or almost as well as the HROL model. Further, there is not much difference in data fit between the ROP and HROP models. In predicting the least preferred choice, the homoscedastic models fare better, and the ROP model outperforms the ROL model in terms of log-likelihood and the average probability of correct prediction. Taken together, the implication is that the ROP model appears to be far more robust to error term misspecification than the ROL model. Earlier studies that use the ROL model and suggest that ranking data is not reliable due to cognitive burden considerations may be misplaced. This is particularly so because, when using the probit kernel, there was little need to introduce heteroscedasticity, while introducing heteroscedasticity in the ROL model improved fit considerably.

7. CONCLUSION

In this paper, we demonstrate through simulation experiments that the coefficient attenuation with rank depth observed with ROL models in past literature may be a result of model misspecification and not unreliability of the ranking data. Our simulation experiments indicate that the ROP model may be a superior option to the ROL model in terms of robustness of coefficients across rank depths and goodness of fit. The different ranking models are also estimated on two different empirical stated preference survey datasets. In both datasets, on almost all metrics, the probit based models were superior to the logit based models. The ROP model appears to be a particularly robust one to error misspecification, and there was relatively little attenuation observed in the parameters at different conditional rank depths. Our analysis suggests that it would be far superior to use a probit kernel to analyze ranked data, and that ranked data may be quite reliable after all. This is an important result for survey data collection and preference elicitation, and suggests that researchers seriously consider ranking data as a way to elicit preferences. Rank-ordered data are as easy to collect as the most preferred alternative, and also have the distinct advantage of providing the ability to exploit the additional information to achieve a certain desired precision in choice model estimation with a much smaller sample size. Thus, ranked data surveys are more cost-effective for a specified precision level of parameters than purely choice (or first preference) data surveys. Besides, ranked data estimation allows the simultaneous prediction of both first choice and last choice, both of which may be helpful in practical situations. Again, the advantage of the ROP model (compared to all other models) comes through clearly in this context from our empirical results.

The transportation industry has recently witnessed the advent of several new technologies such as autonomous vehicles and affordable electric vehicles. For these technologies, there is a growing need to study and identify early adopters as well as the product features that would be most appealing to individuals. Ranking models applied to stated preference data can be a powerful tool for undertaking such studies because there exists little revealed preference data on these new technologies. We hope that researchers and practitioners will reconsider the use of ranking data in modeling choice behavior, rather than inappropriately and summarily dismissing this type of data collection as being unreliable.

ACKNOWLEDGEMENT

This research was partially supported by the Center for Teaching Old Models New Tricks (TOMNET) (Grant No. 69A3551747116) as well as the Data-Supported Transportation Operations and Planning (D-STOP) Center (Grant No. DTRT13GUTC58), both of which are Tier 1 University Transportation Centers sponsored by the US Department of Transportation. The authors are grateful to Lisa Macias for her help in formatting this document, and to three anonymous reviewers for excellent suggestions on improving the paper.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: G.S. Nair, C.R. Bhat, B.P.Y. Loo, W.H.K. Lam, R.M. Pendyala; data collection: D. Fok, R. Paap, B. van Dijk (first dataset), B.P.Y. Loo (second dataset); analysis and interpretation of results: G.S. Nair, C.R. Bhat, B.P.Y. Loo, W.H.K. Lam, R.M. Pendyala; draft manuscript preparation: G.S. Nair, C.R. Bhat, B.P.Y. Loo, W.H.K. Lam, R.M. Pendyala. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Chapman, R.G., and R. Staelin. Exploiting Rank Ordered Choice Set Data within the Stochastic Utility Model. *Journal of Marketing Research*, 1982. 19: 288–301. https://doi.org/10.2307/3151563
2. Foster, V., and S. Mourato. Testing for Consistency in Contingent Ranking Experiments. *Journal of Environmental Economics and Management*, 2002. 44: 309–328.
3. Hausman, J.A., and P.A. Ruud. Specifying and Testing Econometric Models for Rank-Ordered Data. *Journal of Econometrics*, 1987. 34: 83–104. https://doi.org/10.1016/0304-4076(87)90068-6
4. Yan, J., and H.I. Yoo. The Seeming Unreliability of Rank-Ordered Data as a Consequence of Model Misspecification. MPRA Paper No. 56285, 2014. https://mpra.ub.uni-muenchen.de/56285/. Accessed June 18, 2018.
5. Beggs, S., S. Cardell, and J. Hausman. Assessing the Potential Demand for Electric Cars. *Journal of Econometrics*, 1981. 17: 1–19. https://doi.org/10.1016/0304-4076(81)90056-7
6. Train, K.E. *Discrete Choice Methods with Simulation*, 2nd edition. Cambridge University Press, Cambridge, 2009.
7. Beaumais, O., A. Casabianca, X. Pieri, and D. Prunetti. Why Not Allow Individuals to Rank Freely? A Scaled Rank-Ordered Logit Approach Applied to Waste Management in Corsica. *Annals of Economics and Statistics*, 2016. 121/122: 187–212. https://doi.org/10.15609/annaeconstat2009.121-122.187
8. Bogue, S., R. Paleti, and L. Balan. A Modified Rank Ordered Logit Model to Analyze Injury Severity of Occupants in Multivehicle Crashes. *Analytic Methods in Accident Research*, 2017. 14: 22–40. https://doi.org/10.1016/j.amar.2017.03.001
9. Oviedo, J.L., and H.I. Yoo. A Latent Class Nested Logit Model for Rank-Ordered Data with Application to Cork Oak Reforestation. *Environmental and Resource Economics*, 2017. 68: 1021–1051. https://doi.org/10.1007/s10640-016-0058-7
10. Nair, G.S., S. Astroza, C.R. Bhat, S. Khoeini and R.M. Pendyala. An Application of a Rank Ordered Probit Modeling Approach to Understanding Level of Interest in Autonomous Vehicles. *Transportation*, 2018. 45(6): 1623–1637.
11. Schechter, L. The Apple and your Eye: Visual and taste rank-ordered probit analysis with correlated errors. *Food Quality and Preference*, 2010. 21: 112–120. https://doi.org/10.1016/j.foodqual.2009.08.009
12. Tamiya, R., S.Y. Lee, and F. Ohtake. Second to Fourth Digit Ratio and the Sporting Success of Sumo Wrestlers. *Evolution and Human Behavior*, 2012. 33: 130–136. https://doi.org/10.1016/j.evolhumbehav.2011.07.003
13. Bhat, C.R. New Matrix-based Methods for the Analytic Evaluation of the Multivariate Cumulative Normal Distribution Function. *Transportation Research Part B*, 2018. 109: 238–256.
14. Pace, R.K., and J.P. LeSage. Fast Simulated Maximum Likelihood Estimation of the Spatial Probit Model Capable of Handling Large Samples. In: *Spatial Econometrics: Qualitative and Limited Dependent Variables, Advances in Econometrics*. Emerald Group Publishing Limited, 2016, pp. 3–34. https://doi.org/10.1108/S0731-905320160000037008
15. Luce, R.D., and P. Suppes. Preference, Utility and Subjective Probability. In: *Handbook of Mathematical Psychology* (R.D. Luce, R.R. Bush, E.H. Galanter, eds.), Wiley, New York, 1965, Vol. 3, pp. 249–410.

16. van Dijk, B., D. Fok, and R. Paap. A Rank-ordered Logit Model with Unobserved Heterogeneity in Ranking Capabilities. Econometric Institute Research Paper No. EI 2007-07, Erasmus University Rotterdam, Erasmus School of Economics (ESE), Econometric Institute, 2007.
17. Croissant, Y. mlogit: Multinomial Logit Models. R package version 0.3-0, 2018.
18. Loo, B.P.Y., S.C. Wong, and T.D. Hau. Introducing Alternative Fuel Vehicles in Hong Kong: Views from the public light bus industry. *Transportation*, 2006. 33: 605–619. https://doi.org/10.1007/s11116-006-7947-5
19. Loo, B., S. Wang, and T. Hau. Choice or Rank Data in Stated Preference Surveys? *The Open Transportation Journal*, 2008. 2: 74–79. https://doi.org/10.2174/1874447800802010074

LIST OF FIGURES

LIST OF TABLES

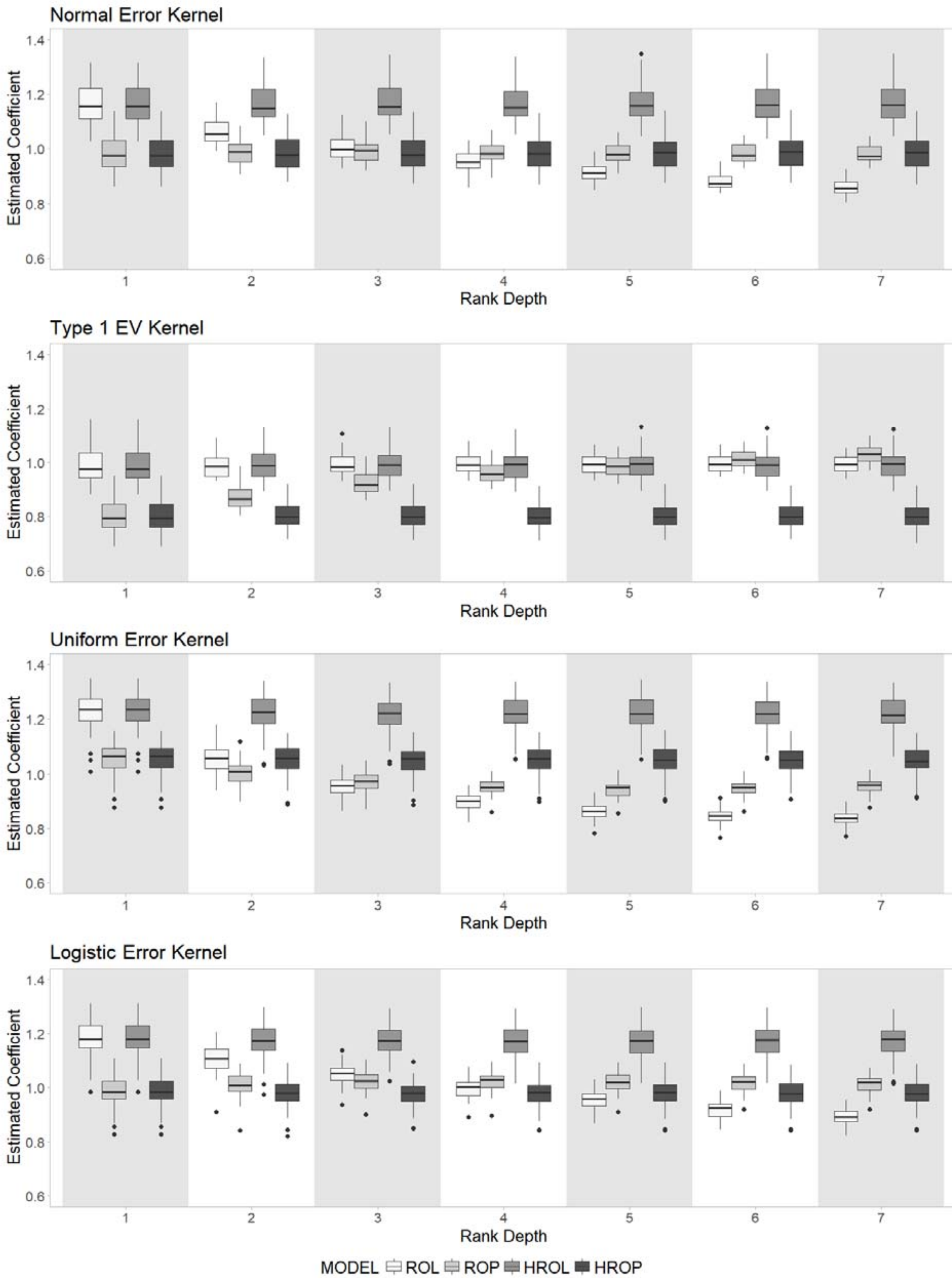

FIGURE 1 Coefficient estimated by ROL, ROP, HROL, and HROP models for different kernel error distributions.

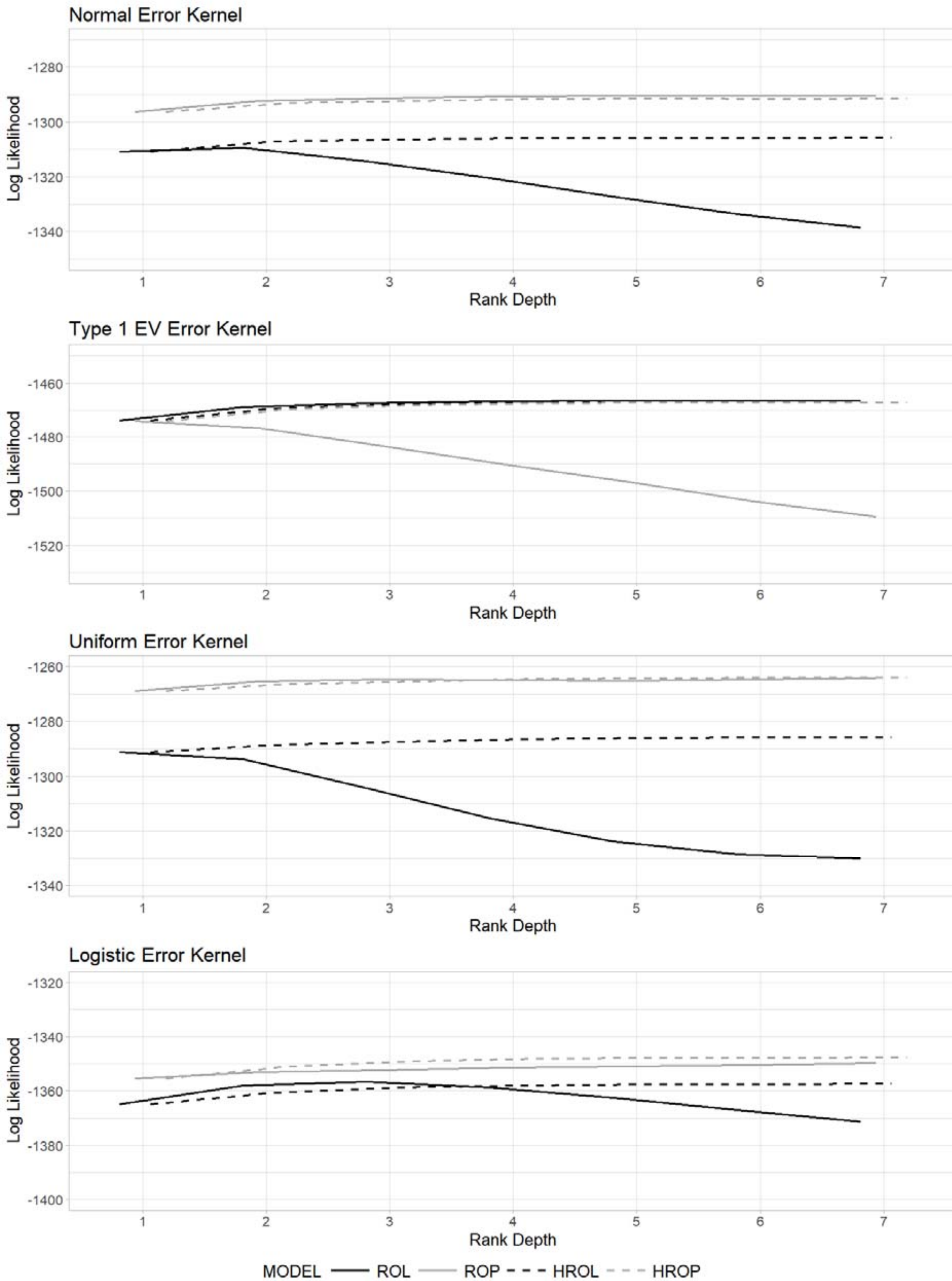

FIGURE 2 Likelihood of test dataset for the different error distributions when using ROL, ROP, HROL, and HROP models.

TABLE 1 Goodness of fit and comparison statistics for models estimated using ROL, ROP, HROL and HROP models

Summary Statistics	Model			
	ROL	**ROP**	**HROL**	**HROP**
Measures of fit for empirical example 1 – Gaming console dataset				
No. of variables in full specification	11	11	15	15
No. of variables in constants model	5	5	5	5
For complete rankings				
Log-likelihood at constants for the ROL model	-546.82			
Log-Likelihood at convergence	-517.37	-507.34	-513.13	-506.41
ADLRI	0.0429	0.0612	0.0433	0.0556
Non-nested comparison between models	**ROL and ROP**		**HROL and HROP**	
p value	$\Phi(-4.48) < 0.0001$		$\Phi(-3.67) = 0.0001$	
Nested comparison between models	**ROL and HROL**		**ROP and HROP**	
p value	$1 - F_{\chi^2_4}(8.48) = 0.0754$		$1 - F_{\chi^2_4}(1.87) = 0.7604$	
For observed first choice				
Log-Likelihood value	-131.30	-124.77	-124.53	-124.05
Average probability of correct prediction	0.28	0.32	0.32	0.32
For observed last choice				
Log-Likelihood value	-133.65	-130.38	-138.02	-131.26
Average probability of correct prediction	0.30	0.29	0.26	0.28
Measures of fit for empirical example 2 – LPG bus utility dataset				
No. of variables in full specification	8	8	10	10
No. of variables in constants model	0	0	0	0
For complete rankings				
Log-likelihood at constants for the ROL model	-2869.78			
Log-Likelihood at convergence	-2489.54	-2416.67	-2419.88	-2402.49
ADLRI	0.1297	0.1551	0.1533	0.1593
Non-nested comparison between models	**ROL and ROP**		**HROL and HROP**	
p value	$\Phi(-12.07) < 0.0001$		$\Phi(-5.90) < 0.0001$	
Nested comparison between models	**ROL and HROL**		**ROP and HROP**	
p value	$1 - F_{\chi^2_2}(139.33) < 0.0001$		$1 - F_{\chi^2_2}(28.37) = 0.0004$	
For observed first choice				
Log-Likelihood of full specification	-928.69	-878.79	-869.37	-866.50
Average probability of correct prediction	0.41	0.46	0.48	0.48
For observed last choice				
Log-Likelihood of full specification	-1141.847	-1135.127	-1161.127	-1147.029
Average probability of correct prediction	0.37	0.36	0.32	0.34