Logit Models for Sets of Ranked Items

Paul D. Allison; Nicholas A. Christakis

Sociological Methodology, Vol. 24. (1994), pp. 199-228.

Stable URL:
http://links.jstor.org/sici?sici=0081-1750%281994%2924%3C199%3ALMFSOR%3E2.0.CO%3B2-A

Sociological Methodology is currently published by American Sociological Association.

6

LOGIT MODELS FOR SETS OF RANKED ITEMS

*Paul D. Allison**
*Nicholas A. Christakis**

Methods are presented for analyzing data generated by asking respondents to rank a set of items. Based on a conditional logit model, these methods allow us to estimate and test for differences among items in respondents' preferences for them; to test for differences in item preferences across subpopulations; and to incorporate predictor variables describing respondents, items, or both. The models can be easily estimated with programs for proportional hazards models, and they can be generalized to allow for ties in the rankings. Detailed examples are given.

1. INTRODUCTION

In this chapter, we describe new methods for analyzing data obtained by asking respondents to rank a set of items according to some criterion. To be more specific, suppose that in January 1992 we asked a sample of potential voters in the New Hampshire primary to rank the five major Democratic presidential candidates according to how likely they would be to vote for each one. Given data of this sort, we might ask the following questions:

Data for the example in this chapter were collected in collaboration with David A. Asch, M.D. For helpful suggestions, we are indebted to Guang Guo, Peter McCullagh, and Herbert Smith.
*University of Pennsylvania

1. Are there systematic differences, overall, among the candidates in voters' preferences for them? That is, do voters tend to rank the candidates in the same way, or is there no particular pattern in rankings? We might also wish to focus on specific pairs of candidates. For example, is there a significant difference between voters' preferences for Bill Clinton and Paul Tsongas? Can we quantify the differences among the candidates?
2. What characteristics of candidates affect their rankings by voters? For example, do candidates from the Northeast do better than those from other regions? How is a candidate's rank determined by the amount of money he spent in the state?
3. What characteristics of voters affect their rank orderings of the candidates? Do men's rankings differ from women's, for example? Do urban voters differ from rural voters? Even if two groups give the same rank order, we might inquire whether one group has more extreme preferences than the other. We might also want to know whether quantitative variables like income or years of schooling have an effect on voters' rankings.
4. What characteristics of voters affect their rankings of individual candidates? Do women favor Paul Tsongas more than men do? Do Vietnam veterans have a special affinity for Bob Kerrey? Do highly educated voters tend to favor Jerry Brown?
5. In addition to the rankings, suppose we also asked voters to rate each candidate on such qualities as likability, honesty, managerial capability, etc. We might then want to know how each of these ratings affects the overall ranking.

We shall show how to answer these and other questions within the framework of a single parametric model. The model is a generalization of the well-known conditional logit regression model introduced by McFadden (1974). In the economics literature, the generalization was proposed by Beggs, Cardell, and Hausman (1981) and further developed by Hausman and Ruud (1987) under the name *rank-ordered logit model.* The model was independently formulated by marketing researchers (Punj and Staelin 1978; Chapman and Staelin 1982) who called it the *exploded logit model.* We adopt the latter terminology because rank-ordered logit model is easily confused with the cumulative logit model[1] that is used for a distinctly

[1]The cumulative logit model is designed for data in which the respondent is presented with a single "item" and asked to rate it on an ordinal scale from 1

different type of ranked data (Agresti 1990). Precedents for the model can also be found in earlier work by Thurstone (1927), Bradley and Terry (1952), Luce (1959), and Plackett (1975). Although the model has seen several applications in the economics and marketing literatures (Lareau and Rae 1989; Moore 1990; Katahira 1990; Kamakura and Mazzon 1991), to the best of our knowledge it has never been employed in sociology.

In addition to explaining and illustrating the method, we generalize the model to accommodate ties in the rankings. We also show how the model can be easily estimated using widely available software for partial likelihood estimation of proportional hazards models for event-history data. Finally, we briefly discuss some alternative approaches to ranked data.

2. THE EXPLODED LOGIT MODEL

We begin with situations in which there are no ties; each respondent assigns a unique rank to each item. Let Y_{ij} be the rank given to item j by respondent i. If there are J items, then Y_{ij} can take on integer values from 1 through J, where 1 is the "best" rank and J is the "worst." A model for such data may be derived from an underlying random utility model—the same model that has been used to justify the standard multinomial logit model. We assume that respondent i has a certain utility U_{ij} for each item j, where j runs from 1 through J, the total number of items. For the moment, we treat J as a constant although, in general, J can differ across respondents. Although the U_{ij}'s are unobserved, we assume that respondent i will give item j a better rank than item k whenever $U_{ij} > U_{ik}$. Furthermore, each U_{ij} is the sum of a systematic component μ_{ij} and a random component ϵ_{ij}:

$$U_{ij} = \mu_{ij} + \epsilon_{ij}, \tag{1}$$

where the ϵ_{ij}'s are independent and identically distributed with an extreme-value distribution.[2] The μ_{ij}'s can be thought of as numerical

to k, where k is some small integer. When there are only two items to be ranked, the exploded logit model is equivalent to both the cumulative logit model and the ordinary binary logit model.

[2]Also known as a Gumbel or double exponential distribution, the standard extreme value distribution has a density function $f(y) = \exp[y - \exp(y)]$. If X and Y both have extreme value distributions, then $X - Y$ has a logistic

quantities that reflect the degree to which respondent i prefers item j over other items. In particular, if the choice is between item j and item k, the odds of choosing j over k is given by $\exp\{\mu_{ij} - \mu_{ik}\}$.

The model may be elaborated by decomposing μ_{ij} into a linear function of a set of explanatory variables

$$\mu_{ij} = \beta_j x_i + \gamma z_j + \theta w_{ij}, \tag{2}$$

where x, z, and w are column vectors of measured variables and β, γ, and θ are row vectors of coefficients to be estimated. The x_i vector contains variables that describe respondents but do not vary over items. The coefficients for such variables must vary over items, and one of the β_j vectors must be set equal to 0 to achieve identification (the choice of the reference item is arbitrary). The z_j vector contains variables that vary across items but are the same for all respondents. In the case of presidential candidates, such variables might include the candidate's age, region of residence, amount of money spent on the campaign, etc. To avoid linear dependence, the number of such variables must be less than or equal to $J - 1$. Moreover, the coefficient vector γ must be constant over items. The w_{ij} vector contains variables that describe a relation between item j and respondent i. For example, we might have a dummy variable indicating whether respondent i contributed any money to candidate j. Also included in this vector would be variables like the respondent i's rating of candidate j's honesty, as well as any interactions between the x_i and the z_j variables.

Some special cases of this model are worth noting. If γ and θ are both 0, we have a model that is equivalent to the usual multinomial logit model. Each of the β_j vectors then describes how characteristics of the respondent affect the log-odds of choosing item j rather than the reference item. If we further specify that the only x variable is a constant of 1, we have a model that allows only for item differences. If θ is 0 and γ and β are nonzero, we have McFadden's conditional logit model. Not all models subsumed by equation (2) are identified. For example, we cannot have J–1 variables in z_j, and at the same time have x_i be a constant of 1.

The random utility model implies the following likelihood L_i

distribution, which is why the random utility model gives rise to a logistic (logit) regression model.

for a single respondent. Let $\delta_{ijk} = 1$ if $Y_{ik} \geq Y_{ij}$, and 0 otherwise. We then have

$$L_i = \prod_{j=1}^{J} \left[\frac{\exp\{\mu_{ij}\}}{\sum_{k=1}^{J} \delta_{ijk} \exp\{\mu_{ik}\}} \right]. \tag{3}$$

Each of the terms in the product has the form of a conditional logit model. To further develop the connection with logit analysis, we can view the respondent's task as one of the sequential, conditional choice with each choice governed by a logit model. The first step is to choose the most preferred item from among the entire set of J items. McFadden's model for the probability choosing item j from among the entire set is

$$\frac{e^{\mu_j}}{\sum_{k=1}^{J} e^{\mu_k}}. \tag{4}$$

When that choice has been made, the probability that the respondent will choose item m from among the remaining items is

$$\frac{e^{\mu_m}}{\sum_{k=1}^{J} e^{\mu_k} - e^{\mu_j}}; \tag{5}$$

i.e., the term associated with j is removed from the denominator. This continues so that, at each step, the denominator is calculated by subtracting the numerator in the previous step from the denominator in the previous step. If the final choice is between, say, items r and s, the probability of choosing r is

$$\frac{e^{\mu_r}}{e^{\mu_r} + e^{\mu_s}}. \tag{6}$$

Taking the product of all these probabilities, we get equation (3) above.

It must be stressed that this is only a behavioral *interpretation* of the model. The likelihood in equation (3) follows directly from equation (1) and the accompanying assumptions. There is nothing in those assumptions requiring that respondents choose items sequentially from "best" to "worst." On the other hand, this way of conceptualizing the model does help to clarify one of its critical assumptions—that the relative preference for any two items is invariant to all other features of the choice set. Thus the relative preference for item j over item k does not depend on which items are in the current choice set, which other items have already been chosen, the number of items already chosen, the order in which those items were chosen, etc. This invariance is a manifestation of the well-known *independence from irrelevant alternatives* (IIA) assumption that characterizes the usual multinomial logit model. Also known as *Luce's choice axiom,* the IIA assumption corresponds most directly to the assumption that the ϵ_{ij} terms are independent across items, although it also depends, in part, on the assumption of an extreme-value distribution for the ϵ terms.

It is not difficult to devise hypothetical examples in which the IIA assumption is implausible (e.g., Amemiya 1985, p. 298; Hill, Axinn, and Thornton 1993)—that is, where the introduction or elimination of a particular item changes the relative preference for the remaining items. On the other hand, attempts to relax this assumption typically lead to difficult problems of either computation or identification. Given these difficulties, we believe that it is reasonable to employ the exploded logit model as an approximation to what may sometimes be a more complex phenomenon. Moreover, the IIA assumption is no less plausible for ranked data than for data in which respondents choose only one item from among some fixed set of alternatives—the kind of data to which the usual multinomial logit model is applied. On the other hand, violations of the IIA assumption may be more easily detected with ranked data because the respondent provides more information on relative preference. In Section 5.4 we show how some consequences of the IIA assumption can be readily tested with ranked data.

Note also that the exploded logit model is not reversible, in the sense that inverting the rank order does not merely change the signs of the coefficients (as it would in a dichotomous logit model or in the cumulative logit model) but it fundamentally changes the model and its associated likelihood (Yellott 1977). In terms of the sequential choice interpretation, it makes a difference whether the respondent first chooses the most preferred item and then proceeds downward, or chooses the least preferred item and proceeds upward. With respect to the random utility framework, the sensitivity to inversion is a consequence of the asymmetry of the extreme value distribution (Critchlow, Fligner, and Verducci 1991). Symmetrical models have been proposed—for example, models based on the normal distribution rather than the extreme value distribution (Thurstone 1927; Keener and Waldman 1985)—but these tend to be rather more cumbersome computationally, and typically give very similar results (Yellott 1977).

3. ESTIMATION AND COMPUTATION

For a sample of n independent respondents, equation (3) implies a log likelihood of

$$\log L = \sum_{i=1}^{n} \sum_{j=1}^{J_i} \mu_{ij} - \sum_{i=1}^{n} \sum_{j=1}^{J_i} \log \left[\sum_{k=1}^{J_i} \delta_{ijk} \exp(\mu_{ik}) \right]. \quad (7)$$

Note that we now allow J_i to vary across respondents. In practice, the linear model for the μ_{ij}'s in equation (2) is substituted into equation (7), which is then maximized with respect to the coefficient vectors. The likelihood is globally concave, guaranteeing that if a maximum is found, it is a global rather than a local maximum (Beggs et al. 1981). Keener and Waldman (1985) prove the consistency and asymptotic normality of the resulting coefficient estimates.

Maximum likelihood estimation of the model is available with the discrete choice procedure of the LIMDEP package (Greene 1992). It has not been previously recognized that estimation can also be easily accomplished with most partial likelihood procedures for estimating proportional hazards models. Such procedures are designed to estimate regression models for ordered survival (event) times, and are available in SAS, BMDP, SYSTAT, and SPSS (Windows version).

The partial likelihood for a set of ordered times is identical to that in equation (3) above (e.g., Cox and Oakes 1984). To constrain these likelihoods to be calculated *within* respondents and then multiplied across respondents, it is necessary to stratify by respondent, an option available in most current partial likelihood programs. The general procedure is as follows:

1. Create a separate record for each of the J_i items for each respondent. Each record should contain (a) the identifying number of

the item, (b) the rank given to the item, (c) a unique ID number for the respondent, and (d) explanatory variables describing the item and/or the respondent.

2. Specify a model with the item rank as the dependent variable and the various explanatory variables as independent variables.
3. Specify stratification on the ID number.

While conventional partial likelihood applications usually involve censored data, there is no need to specify a censoring indicator for the exploded logit model. More details will be given below.

4. TIES AND INCOMPLETE RANKINGS

So far we have assumed that every respondent gives a unique rank to every item. In practice, especially in self-administered questionnaires, respondents often fail to accomplish this. Presented with a list of six items, some respondents will assign the top ranks (e.g., 1, 2, 3, and 4) and leave the rest blank. Equivalently, respondents may be asked explicitly to rank only, say, the top four items. This creates no difficulty at all for the procedures we have just described. If the last reported rank is a 3, then the unranked items should all be given a value of 4 (or any other constant higher than 3). The last term in the likelihood is then the probability of choosing the third item from among all the remaining items—which is exactly what it should be.[3]

More troublesome is the situation where a respondent assigns the same rank to two or more items, and then proceeds to give less preferred ranks to other items.[4] Thus, if there are six items, we might see ranks (1, 2, 3, 3, 4, 5). Such a pattern cannot be accommodated by the likelihood in equation (3).

Here we propose a generalization of that likelihood for tied data, using the principle of marginal likelihood. Since the likelihood

[3]Although this modification works for any partial likelihood program, it will not work for the discrete choice procedure in LIMDEP, which requires that every respondent give a unique rank to each item.

[4]These kinds of ties can also be produced by design. In a well-known study of parental values, Kohn (1969) presented his respondents with a list of 12 items and then, in effect, asked them to choose the most desirable item, the least desirable item, the two items just below the most desirable, and the two items just above the least desirable.

for ranked items is isomorphic to the partial likelihood for survival data, we draw on previous work on tied survival times (Kalbfleisch and Prentice 1980). For a respondent assigning the ranks (1, 2, 3, 3, 4, 5) to six items, the first two terms in the likelihood have the usual form as in expressions (4) and (5) above. For the third term, we presume that the respondent really has a preference ordering for the two items (which we arbitrarily label 3a and 3b), but we simply don't know what it is.[5] Obviously, there are only two possibilities: (A) item 3a is preferred to item 3b (and both are preferred to 4 and 5); or (B) item 3b is preferred to item 3a (and both are preferred to 4 and 5). Since A and B are mutually exclusive, it follows that Pr(A or B) = Pr(A) + Pr(B). Letting $\mu_{[j]}$ be the systematic component associated with the item given a rank of j, the third term in the likelihood is

$$\left(\frac{e^{\mu_{[3a]}}}{e^{\mu_{[3a]}} + e^{\mu_{[3b]}} + e^{\mu_{[4]}} + e^{\mu_{[5]}}}\right)\left(\frac{e^{\mu_{[3b]}}}{e^{\mu_{[3b]}} + e^{\mu_{[4]}} + e^{\mu_{[5]}}}\right) +$$
$$\left(\frac{e^{\mu_{[3b]}}}{e^{\mu_{[3a]}} + e^{\mu_{[3b]}} + e^{\mu_{[4]}} + e^{\mu_{[5]}}}\right)\left(\frac{e^{\mu_{[3a]}}}{e^{\mu_{[3a]}} + e^{\mu_{[4]}} + e^{\mu_{[5]}}}\right) .$$

The last term in the likelihood is, as before,

$$\frac{e^{\mu_{[4]}}}{e^{\mu_{[4]}} + e^{\mu_{[5]}}} .$$

The general form of the likelihood function for any pattern of ties can be written as follows. Suppose that a respondent assigns K distinct ranks ($K \leq J$) with $k = 1, \ldots, K$. We then let d_k be the number of items tied for rank k. For the d_k items in rank k, assign the arbitrary labels 1 through d_k. Let Q_k be the set of permutations of the numbers $1, \ldots, d_k$ and let $p = (p_1, \ldots, p_{d_k})$ be an element in that set. Let $\mu_{[k]m}$ be the systematic component for item m in rank k. The likelihood for a single respondent is then

[5]An alternative approach is to assume that when respondents assign tied ranks, they really do not have an underlying preference order for the tied items. This leads to a discrete model in which the likelihood of the ranking (1, 2, 3, 3, 4, 5) has a term which is the probability of choosing the pair {3a, 3b} from the set of all possible pairs formed from the set {3a, 3b, 4, 5}. This model can also be estimated with the SAS PHREG procedure by specifying TIES = DISCRETE on the MODEL statement. We have not pursued this model here because it does not follow from the underlying random utility model discussed above.

$$\prod_{k=1}^{K} \sum_{p \in Q_k} \prod_{r=1}^{d_k} \left(\frac{\exp\{\mu_{[k]p_r}\}}{\sum_{s=r}^{d_k} \exp\{\mu_{[k]p_s}\} + \sum_{\ell > k} \sum_{m=1}^{d_\ell} \exp\{\mu_{[\ell]m}\}} \right). \tag{8}$$

When there are no ties, this reduces to equation (3) above. If all the items are given equal rank, the likelihood is 1.0 and, hence, the respondent contributes no information about the parameters. The likelihood for the entire sample is just the product across respondents of the likelihood in equation (8).

Most partial likelihood programs recognize the possibility of tied ranks, but instead of the likelihood in equation (8), which can be very computationally intensive, they use an approximation proposed by Breslow (1974),

$$\prod_{k=1}^{K} \left(\frac{\exp\left(\sum_{m=1}^{d_k} \mu_{[k]m}\right)}{\left[\sum_{\ell \in R} \exp(\mu_\ell)\right]^{d_k}} \right), \tag{9}$$

where R_k is the set of all items in ranks greater than or equal to k. Unfortunately, this approximation has been shown to be inaccurate when the number of items receiving the same rank is a substantial fraction of the number of items at that rank or higher (Farewell and Prentice 1980).

To our knowledge, the only commercial program that calculates the exact likelihood for tied data is the PHREG procedure in SAS (SAS Institute 1992).[6] To implement that likelihood, it is necessary only to specify the TIES = EXACT option on the model statement. Since the exact likelihood can increase the computation time by severalfold, however, it may be advisable to use Breslow's approximation in (9) for exploratory analysis.

There is also the question of how ties should be coded. Standard nonparametric approaches to ranked data typically require that the sum of the ranks must be the same for every respondent regardless of how many ties occur. This requires that appropriate fractional ranks be assigned to tied items. While this sort of coding will certainly work with the partial likelihood method proposed here, it is unnecessarily restrictive. It is sufficient that tied items be assigned the same number, and that the numbers be in the proper order.

[6]The manual for PHREG gives the likelihood function for tied data using an integral representation that appears to be very different from our equation (8) but is, in fact, equivalent.

5. AN EXAMPLE

As part of research on the way physicians make decisions about the withdrawal of life support (Christakis and Asch 1993), a sample of 475 physicians responded to a mail questionnaire that included the following question:

> Some physicians may feel differently about withdrawing life-sustaining therapy depending on what that therapy is. In general, what kinds of life-sustaining therapy are you likely or unlikely to withdraw if the circumstances presented themselves?
>
> Rank your responses from 1 [most likely to withdraw] to 8 [least likely to withdraw]. Use the same number for any therapies you believe deserve the same rank.
>
> __ antibiotics
> __ blood products
> __ intravenous fluids
> __ intravenous vasopressors
> __ mechanical ventilation
> __ renal dialysis
> __ total parenteral nutrition
> __ tube feedings and fluids
>
> *If relevant, please also check* none, one, or both *of these two boxes:*
>
> ☐ I would *not* withdraw any of these medical therapies
> ☐ I consider *all* of these *equally* easy or difficult to withdraw

Nine physicians did not respond to the question and were therefore excluded from the analysis. Fifty-three physicians checked one or both of the boxes, or assigned equal ranks to all eight items. Checking either of the boxes was interpreted as assigning equal ranks to all items.[7] These 53 cases were also excluded from the analysis since they contribute nothing to the likelihood function. Of the remaining 413 cases, 210 had no tied ranks. Of the 203 cases with ties, there was a wide range of tie patterns.

5.1 *Item Differences*

The first task was to estimate a model that allowed for differences among the eight types of therapy but no differences across respondents. This can be formulated as $\mu_{ij} = \beta_j$ for all i and j, with β_j set to 0 for an arbitrarily chosen item, to achieve identification. As in all the analyses for this example, the working data set consisted of a separate record for each therapy for each respondent, for a total of 3,304 records. Each record included the following variables:

1. A unique identification number for the *respondent.*
2. The rank assigned by the respondent to that item.
3. A set of 7 dummy (indicator) variables corresponding to 7 of the 8 different therapies. The reference (omitted) category was antibiotics. For each record, at most one of these variables was coded 1 and the rest were 0. (All the dummy variables were zero for the antibiotics therapy).
4. Explanatory variables describing the respondent (age, sex, etc.).

The model was estimated with the SAS procedure PHREG using the method described in Section 4 for handling ties. Control statements for this analysis are shown in Appendix A.

Estimates of the β_j parameters are shown in rank order in the first column of Table 1. All of the numbers are contrasts with the reference category, antibiotics.[8] These estimates indicate that, on average, physicians are most likely to withdraw blood products and least likely to withdraw intravenous fluids. The numerical values can be interpreted as differences in log odds. Thus, if we exponentiate the coefficient for blood products, $e^{.99} = 2.69$, we may say that the odds of preferring to withdraw blood products are 2.69 times the odds of preferring to withdraw antibiotics. By exponentiating *differ-*

[7]No physicians checked either of the boxes after giving unequal ranks to the items.

[8]Only differences between categories are estimable.

TABLE 1
Physician Preferences Regarding Withdrawal of Different Forms of Life Support

Form of Life Support	β_j (Exact)	β_j (Approx.)	Mean Rank
Blood products	.99	.75	3.27
Hemodialysis	.98	.76	3.28
Intravenous vasopressors	.65	.51	3.75
Total parenteral nutrition	.47	.35	3.93
Antibiotics	.00	.00	4.73
Tube feedings	−.54	−.56	5.42
Mechanical ventilation	−.69	−.46	5.27
Intravenous fluids	−1.14	−.92	6.35

ences in the numbers, we can get the contrast between any two therapies.

The standard errors of the estimates are all about .09. As with most partial likelihood programs, the PHREG procedure also reports the squared ratio of each estimate to its standard error, which is a Wald chi-square statistic for the null hypothesis that the parameter is zero. All had p-values less than .0001. But of course, these tests tell us the significance of the contrasts only with the omitted category, antibiotics. Using the variance-covariance matrix of the estimates, we also calculated Wald chi-squares for all of the other possible contrasts.[9] All but three of the tests were significant, with p-values less than .0001. The differences between blood products and hemodialysis and between mechanical ventilation and tube feedings were not statistically significant at the .05 level. The contrast between intravenous vasopressors and total parenteral nutrition had a p-value of .04.

Most partial likelihood programs also report global chi-square tests for the null hypothesis that all the parameters are 0, which, for this example, is equivalent to saying that there are no differences among the therapies in physicians' preferences for withdrawing them. In this case, the likelihood ratio chi-square is 1011 with 7 degrees of freedom, yielding a p-value much less than .0001; so we reject the null hypothesis.

The second column of Table 1 gives estimates produced using Breslow's approximation to the likelihood for tied data. (While the

[9]With the PHREG procedure, the construction of such tests is greatly simplified by the TEST statement.

exact likelihood required 60 seconds of CPU time on an IBM ES/9121 mainframe, the approximate likelihood used only 5 seconds.) Although the general pattern is similar, there are two reversals in the rank order for adjacent items. These reversals occurred for pairs of items that were not significantly different in the exact analysis. Also, the estimates from the approximate method appear to be somewhat attenuated toward 0: The standard deviation of the approximate estimates is .60 as compared with .76 for the exact method.[10]

For comparison with more traditional, nonparametric methods, the third column of Table 1 gives the average ranks across respondents for each of the eight items. (Ties were normalized so that the ranks summed to 36 for each respondent.) Except for one reversal, the rank ordering is the same as that in the first column. The null hypothesis that all permutations of the items are equally likely can be tested with Friedman's two-way analysis of variance for ranks (Siegel 1956), which in this case yields a p-value less than .001. Comparisons between pairs of items using the Wilcoxon signed rank test (Siegel 1956) found that all differences were significant at the .001 level, except for the three pairs of items that were not significantly different in the parametric analysis. Thus, for this example, the exploded logit model gives results that are qualitatively the same as those produced by standard nonparametric methods.

5.2 *Differences Between Groups*

The preceding model assumed that everyone in the sample had the same probability distribution of item preferences and that the observed differences in people's rankings were due only to random variation. We turn now to models in which there is heterogeneity across respondents that is attributable to measured variables. We begin with the simplest situation in which the sample consists of two groups, with homogeneity within each group.

For the life support example, we wanted to test the null hypothesis that male and female physicians had the same item prefer-

[10]The PHREG procedure can optionally use an alternative approximation proposed by Efron (1977). When applied to these data, Efron's method yielded results that were about midway between the exact likelihood and the Breslow likelihood. This improvement came with only a trivial increase in computation time over Breslow's method.

TABLE 2
Physician Preferences Regarding Withdrawal of Different Forms of Life Support, by Sex

Form of Life Support	Men	Women	Difference
Blood products	.91	1.37	−.46*
Hemodialysis	.85	1.54	−.69**
Intravenous vasopressors	.52	1.24	−.72**
Total parenteral nutrition	.38	.75	−.27
Antibiotics	.00	.00	.00
Tube feedings	−.60	−.46	−.14
Mechanical ventilation	−.75	−.44	−.31
Intravenous fluids	−1.18	−1.06	−.12

*p < .05
**p < .01

ences. Of the 413 cases used in the preceding analysis, 10 did not report their sex and were excluded from the analysis. Of the remainder, 319 were men and 84 were women. The model allowing for sex differences can be written as $\mu_{ij} = \beta_{0j} + \beta_{1j}x_i$, where $x_i = 1$ if male, 0 if female. In practice, we specified a model that included the original 7 dummy variables for item type (whose coefficients are the β_0's), plus the 7 products of each of these variables and a dummy variable for sex. (The coefficients of the 7 product variables are the β_{1j}'s; see Appendix B for SAS code.) To test the hypothesis that there are no sex differences, we computed a Wald chi-square statistic for the hypothesis that all the sex-by-item products have zero coefficients, which had a value of 14.76 with 7 d.f. for a p-value of .039. Thus we can conclude that men and women do differ in their preferences for withdrawing life support therapies.[11]

The parameter estimates are shown in Table 2. The overall rank ordering is approximately the same for men and women, with only two reversals of adjacent items (items that were not significantly different from each other in the combined sample). The estimates for the three most preferred items are significantly higher for women as

[11]The Wald chi-square is the easiest statistic to obtain with SAS PHREG because it can be directly requested without reestimating the model. Alternatively, a likelihood ratio chi-square could be computed by estimating the model with and without the sex interactions, and taking twice the positive difference in the log-likelihoods. For the model without the interactions, it would be necessary to exclude those cases with missing data on sex.

compared with men. But remember that each of these coefficients is a comparison with the omitted category of antibiotics. One possible interpretation is that, instead of preferring these three items *more* than men do, women merely prefer to withdraw antibiotics *less*. It does appear, however, that women have more extreme preferences than men: The standard deviation of the coefficient estimates for women is 0.923, compared with 0.731 with men. (These values are invariant to the choice of omitted category.) Since respondents only assign *ranks* to the items, not quantitative scores, this means that there is more *consensus* among women in their assignment of ranks.

The differences in coefficients for men and women have an odds interpretation. Consider hemodialysis, which has a difference of -0.69. Exponentiating, we get $e^{-0.69} = 0.50$, which means that the odds of preferring hemodialysis to antibiotics are about half as great for men as for women. Contrasts with items other than the reference item can be obtained by taking item differences between the sex differences. Thus, if we want to compare hemodialysis with intravenous fluids, we compute $-0.69 - (-0.12) = -0.57$. Then $e^{-0.57} = 0.56$ tells us that the odds of men preferring hemodialysis to intravenous fluids is about 0.56 times the odds for women.

These methods for comparisons between two groups can be readily extended to three or more groups. For example, to compare Protestants, Catholics, and Jews, we created dummy variables for Jew and Catholic, and then constructed the products of each of these variables with each of the 7 item dummies, for a total of 14. The null hypothesis that all 14 variables had coefficients of 0 is equivalent to the hypothesis of no difference among the three religious groups. With a Wald chi-square of 23.3 with 14 d.f., the p-value for this test was .055.

At least in principle, this sort of test could also be accomplished with nonparametric methods. Schucany and Frawley (1973) introduced a rank test for the hypothesis that respondents agree on the rankings of items within each of two or more groups and between the groups. However, their method appears to be relatively insensitive to small differences in rankings between the groups, and the adjustment to handle ties (Li and Schucany 1975; Shucany and Beckett 1976) is rather cumbersome. Other nonparametric techniques to assess intergroup agreement are also available (Hollander and Sethuraman 1978; Kraemer 1981).

5.3 *Effects of Quantitative Variables*

Since men and women differ in many ways, the sex differences we detected in the preceding analysis could well be spurious. In particular, there is a substantial age difference between the two groups: The average age of the men is 43, while the average age of the women is 36, corresponding to a correlation of .23 between age and sex. In this section we show how to estimate and test the effect of sex controlling for age, and the effect of age controlling for sex. In contrast to the previous section, there is no nonparametric method that will accommodate covariates, either quantitative or qualitative.

The setup is a simple extension of what has already been done. The model can be written as $\mu_{ij} = \beta_{0j} + \beta_{1j}x_{1i} + \beta_{2j}x_{2i}$ where $x_{1i} = 1$ if male, 0 if female, and x_{2i} = age in years. In addition to the 7 sex-by-item products, we also created 7 age-by-item products. The model includes all 14 of these products, plus the original 7 item dummies (SAS code is in Appendix C). We then calculate two Wald chi-square tests: one for the null hypothesis that all 7 sex-by-item coefficients are 0, and the other for the null hypothesis that all 7 age-by-item coefficients are 0.

The chi-square for the age effects was 34.5 with 7 d.f., yielding a p-value less than .0001. Thus there is strong evidence that preferences vary with age. On the other hand, the chi-square for sex was only 8.30 with 7 d.f., for a p-value of .30. Controlling for age, therefore, sex differences in preferences are no longer statistically significant. Table 3 reports the coefficients for the product terms. The sex coefficients can be directly compared to the "difference" column in Table 2. The coefficients that were significant in Table 2 are now smaller in magnitude, and the significance levels are greatly attenuated.

The age coefficients can be interpreted as follows. For each coefficient β, compute $100(e^{\beta} - 1)$. This is the percent change in the odds of preferring that item over antibiotics for each 1-year increase in age. Thus, for intravenous vasopressors, we have $100(e^{-.02} - 1) = -2\%$, which implies that with each 1-year increase in age, the odds of preferring to withdraw intravenous vasopressors over antibiotics goes down by 2%.[12]

[12] For small values of β, one may use the approximation $e^{\beta} - 1 \cong \beta$.

TABLE 3
Effects of Sex and Age on Preferences Regarding Withdrawal of Different Forms of Life Support

Form of Life Support	Sex[a]	Age
Blood products	−.39	−.008
Hemodialysis	−.54*	−.021**
Intravenous vasopressors	−.57*	−.020**
Total parenteral nutrition	−.36	−.002
Antibiotics	.00	.00
Tube feedings	−.23	.012
Mechanical ventilation	−.16	−.023**
Intravenous fluids	−.18	.007

*$p < .05$
**$p < .01$
[a]Coded 1 for men and 0 for women.

5.4 *Stability of Preferences*

As noted earlier, the exploded logit model assumes that a respondent's relative preferences for different items are invariant to the stage of the ranking process. In comparing two items, A and B, out of a set of six items, it should make no difference whether we are choosing the most preferred item from the set of six, or we have already chosen items for the top four ranks and only A and B are left. Despite this assumption, there may be reason to suspect that respondents are diligent in assigning the first few ranks but become careless in assigning ranks to the remaining items.

Chapman and Staelin (1982) and Hausman and Ruud (1987) describe some methods for testing this possibility. Here we present a similar method that can be easily implemented with partial likelihood programs that allow for "time dependent covariates." Basically, the strategy is to define a dummy variable that is equal to 0 for earlier ranks and 1 for later ranks. We then create product terms for this variable and the item indicators. The product terms are included in the model along with the item dummies, and the product terms are tested for significance as a group.

We applied this method to the life support data, with a division between early and late ranks at 4.5. (SAS code is in Appendix D). The Wald chi-square for the hypothesis that all the product terms

TABLE 4
Preferences Regarding Withdrawal of Different Forms of Life Support, by Ranking Stage

Form of Life Support	Early	Late
Blood products	1.16	.67*
Hemodialysis	1.17	.37**
Intravenous vasopressors	.82	.17**
Total parenteral nutrition	.59	.38
Antibiotics	.00	.00
Tube feedings	−.44	−.71
Mechanical ventilation	−.11	−1.51**
Intravenous fluids	−1.32	−1.27

*Differs from first column at $p < .05$
**Differs from first column at $p < .01$

are 0 was 74.4 with 7 d.f. for a *p*-value less than .0001. Clearly the later rankings differ in some way from the early rankings.

Table 4 shows the differences. The numbers in the first column are the coefficients for the item dummies. They correspond exactly to the results one would get if all ranks greater than or equal to 4.5 were treated as ties. The numbers in the second column are obtained by adding the coefficient for each product term to the corresponding item dummy. The rank orderings are approximately the same, although, as usual, there are some reversals of adjacent or nearby items. On the other hand, the preferences for the higher ranked items have become smaller at the later stages, while the preferences for the lower ranked items have either stayed the same or become more negative.

This pattern could be parsimoniously explained by an upward valuation of the reference item, antibiotics, at later stages. On the other hand, it cannot be so easily explained by the notion of "carelessness" or "noise" at later ranking stages. As Chapman and Staelin (1982) observe, greater randomness in people's choices should be manifest in an attenuation of the item values toward zero. To check this possibility, we computed the standard deviation of the eight coefficient estimates for early and late choices, yielding .810 for early choices and .767 for later choices. Although the difference is in the predicted direction, it does not seem large enough to draw any substantive conclusions.

There are several possible variations on this testing strategy. One could, of course, choose the division between early and late stages at some point other than 4.5. One could also divide the rank stages into three groups—early, middle, and late—and define two dummy variables (and their resulting products) rather than just one. Finally, if we were willing to postulate a linear change in the values with rank stage, we could simply include the products of the item dummies with rank itself.

What should be done if the test indicates rejection of the stability assumption? If the results suggest that the quality of the rankings deteriorates with later ranks, it may be advisable to use only the earlier ranks. As already noted, this is accomplished by treating all ranks beyond a certain level as ties. In the extreme case where only the top rank is used, the method reduces to standard conditional logit analysis. If, as in the example here, the results do not suggest diminished ranking performance at later stages, and the rank orderings are reasonably consistent, there is probably not too much danger in ignoring the discrepancy. The alternative—treating the later rankings as ties—can result in a substantial loss of efficiency.[13]

5.5 *Effects of Item Characteristics*

In economics and marketing research, the exploded logit model has been primarily used to investigate the effects of item characteristics on the rankings that people give those items. Punj and Staelin (1978), for example, wanted to show how characteristics of colleges—size, cost, and "quality," for example—affected students' rankings of the schools to which they had been accepted. Beggs et al. (1981) sought to determine how characteristics of automobiles affected people's preferences for them. We now illustrate how item characteristics can be incorporated into the study of withdrawal of life support.

Since we had no direct measures of the characteristics of life support therapies, we relied on a panel of experts to assign scores on several dimensions. First, a Delphi method was used to identify several underlying dimensions that differentiated the eight forms of life

[13]We have also shown that apparent differences between early and later stages can result from heterogeneity across individuals rather than changes in the individual's valuation of performance. Proof (by counterexample) is available on request.

support.[14] These dimensions included scarcity, expense, continuity, invasiveness, pain on withdrawal, and rapidity of death after withdrawal. Next, a sample of 23 pulmonologists practicing in a tertiary intensive care unit was asked to rate each of the eight forms of life support on a 10-point scale for each of the several dimensions. To each item, the mean score (across the 23 experts) for each of the six dimensions was assigned. To each of the 3,304 person-item records, the six scores for that particular item are assigned. Thus, although the measurement of the item dimensions was "subjective," the measurements were completely independent of our principal sample of physicians, and they represented a consensus by expert informants.

The model may be written as $\mu_{ij} = \gamma z_j$ where z_j is a vector containing the six dimensions describing the items. The model does *not* contain the seven dummy variables for the items themselves. Including them would induce a linear dependency (perfect multicollinearity), because the seven item dummies capture *all* of the variation among the items. Similarly, a model with seven independent dimensions (instead of six) would be exactly equivalent to the model with seven item dummies (it would have an identical log-likelihood), because there are only 7 d.f. to differentiate the items. Any more than seven dimensions would, again, induce linear dependence. In short, models of this sort can be seen as attempts to (a) show how the item differences can be more meaningfully represented in terms of underlying dimensions, and/or (b) determine whether the item differences can be adequately represented by fewer dimensions than the number of item dummies.

Table 5 shows the coefficient estimates and test statistics for the model with six dimensions. All but one of the coefficients is significantly different from zero at the .05 level, and the one exception is nearly significant. A positive coefficient indicates that items high on the dimension are more preferred for withdrawal. Thus physicians prefer to withdraw therapies that are scarce, expensive, invasive, and that cause rapid death when withdrawn. They are

[14]A modified Delphi method was used with a panel of ten internists to identify a consensus list of 11 dimensions of forms of life support that might be relevant to physicians considering withdrawal. In addition to the six used here, the dimensions included painfulness in application, emotional toll, level of technology, artificiality, and need for an ICU. These five were excluded from the present report.

TABLE 5
Effects of Item Dimensions on Preferences Regarding Withdrawal of Life Support

Dimension	Coefficient	*t*
Continuity	−.039	−1.92
Scarcity	.141	8.42
Expense	.112	3.61
Rapidity of death	.097	5.40
Invasiveness	.064	2.45
Pain on withdrawal	−.384	−19.31

reluctant to withdraw therapies that are continuously applied and that cause pain on withdrawal. The magnitudes of the coefficients can be interpreted as follows. Blood products has a scarcity score of 8.0 while antibiotics has a scarcity score of 2.0, for a difference of 6 points. Mutiplying this difference by the coefficient for scarcity of .141 yields .846. This indicates how much the log-odds of preferring to withdraw blood products over antibiotics is increased over what it would be if their scarcity scores were equal, controlling for other factors. Equivalently, applying the transformation $100(e^{\beta} - 1)$, we can say that the difference in scarcity scores has produced a 133 percent increase in the odds of preferring blood products over antibiotics.

Because the six dimensions are all measured on scales with the same range, we can directly compare the coefficient magnitudes to determine the relative importance of the variables. The largest by far is pain on withdrawal. Physicians appear to be extremely sensitive to this dimension when deciding whether to withdraw a therapy. Importance of the items can also be judged by the size of the *t*-statistics. With one reversal, the rank ordering of these statistics is the same as the coefficients themselves.

We can also compare this model directly with the model containing the seven item dummies in Table 1. The likelihood ratio chi-square (obtained by taking twice the positive difference in the log-likelihoods) is only .67 with 1 d.f., indicating that these six dimensions do a quite satisfactory job in accounting for the differences among the items. It would also be possible to investigate whether the six dimensions have different effects for different subgroups (e.g., men and women) by including interaction terms between subgroup dummies and the dimension measures.

6. EXTENSIONS AND DESIGN ISSUES

The exploded logit model offers several design and analysis possibilities that are not encompassed by the life support withdrawal study. Here we briefly mention some of these.

1. *Rank only "top" choices.* As previously noted, it is possible to analyze only the first few ranks and treat later ranks as ties. This might be done if there is reason to suspect that lower rankings are "noisier." But if this is anticipated in advance, we might as well reduce the burden on respondents by asking them, for example, to "rank your top three choices out of the following list of 10 items."
2. *Different choice sets for different respondents.* In the life support withdrawal study, every respondent was asked to rank the same set of eight items. As in the standard conditional logit model, the exploded logit model can easily incorporate different choice sets for different respondents. In Punj and Staelin's (1978) study of college choice, for example, respondents were asked to rank only those colleges to which they had been accepted. An extreme version of this approach is the well-known "matched-pair" design in which respondents are given only two items to rank (Bradley and Terry 1952). If the aim is to estimate models with dummy variables for each item, some care must be taken to see that there is sufficient overlap in the choice sets to identify the parameters. This problem does not occur, however, for models in which the independent variables are item characteristics.
3. *Presentation of bundles of characteristics.* The items that people are asked to rank may be sets of characteristics rather than named items. Beggs et al. (1981), for example, presented people with sets of cards describing hypothetical automobiles and asked them to sort the cards in order of preference. The descriptions included such characteristics as price, driving range, energy consumption, and size. The aim was to determine how these characteristics affected automobile preference. In designs of this sort, the trick is

to vary the characteristics across items in such a way as to minimize collinearity and maximize variability of the characteristics.

4. *Subjective ratings of item dimensions.* In the life support withdrawal study, characteristics of the items were measured by getting ratings from an expert panel of 23 pulmonologists. As an alternative design, we could have asked our sample of 475 physicians to perform the same rating task. The we could have modeled the effect of these ratings on the rank ordering of the eight items. The form of the model would be identical to that in Table 5, but the measured effects would be for each respondent's perception of the item characteristics rather than some consensus rating by experts. In many contexts, perceptions of items may mediate any effects of their objective characteristics. On the other hand, a danger in this design is that the ranking of items may contaminate the ratings of characteristics, or vice versa. A technical advantage of the design is that, unlike the objective rating design, the number of characteristics that can be included in the model is not limited by the number of items ranked.
5. *Item-respondent relationships.* Some variables may describe an objective relationship between a respondent and an item. In Punj and Staelin's (1978) college choice study, for example, one of the explanatory variables was the distance between the student and the college. In a study of voter's preference for candidates, one might include a dummy variable indicating whether the candidate came from the same region as the voter.
6. *Instructions for ranking.* Since the exploded logit model can be conceptualized as a sequence of choices from most preferred to least preferred, it may be desirable to instruct the ranker to follow that procedure explicitly. Instructions could say "First choose the item you prefer most and give it a rank of 1. Then, from the remaining items, choose the one you prefer most and give it a rank of 2. Continue downward in this way until you have assigned ranks to all the items."
7. *Mixture models.* In our examples, we have allowed for population heterogeneity by introducing measured variables like sex and age. An alternative approach is to allow for heterogeneity in unmeasured variables by formulating mixture models. Kamakura and Mazzon (1991), for example, estimated a model for the ranking of human values that allowed for six latent classes hav-

ing different item parameters in each class. For other examples, see Croon and Luijkx (1993) and Stern (1993). Estimation of such models requires specialized software.

7. ALTERNATIVE METHODS

Before concluding, we briefly consider some alternative approaches. As mentioned in Sections 5.1 and 5.2, nonparametric methods are available for testing some of the simpler hypotheses (Siegel 1956). Within the realm of parametric methods, many different stochastic models have been proposed for sets of ranked items (Fligner and Verducci 1988; Critchlow et al. 1991), but only a few of these incorporate explanatory variables. Most similar to the exploded logit model is the Thurstone (random utility) model that is equivalent to our equations (1) and (2) but assumes an underlying normal distribution rather than an extreme value distribution (Keener and Waldman 1985). Although this model has some attractive properties, it is much more computationally demanding than the exploded logit model, making it impractical for routine applications. Less well known is the Mallows-Bradley-Terry model, which is an alternative generalization of paired-comparison models. Critchlow and Fligner (1993) have shown how explanatory variables can be incorporated into this model, and how maximum likelihood estimates can be obtained with standard algorithms for generalized linear models. For either of these two models, all the techniques we have illustrated for the exploded logit model could, in principle, be applied.

Taking a very different approach, Jackson and Alwin (1981) showed how a confirmatory factor model could be fit to sets of ranked items by deleting one item and imposing appropriate constraints on the covariance matrix for the error terms. Imbedding such a model in a more general structural equation model would allow one to test many hypotheses similar to those considered here. On the other hand, this approach makes the rather unrealistic assumption that the observed ranks are *linear* functions of some set of latent variables. Duncan and Brody (1982) considered a variety of log-linear and related models for rankings of three items when the resulting data are arrayed in a contingency table. Their approach is unlikely to be practical, however, when the number of items is six or greater since the contingency table would have at least 6! cells.

8. CONCLUSION

Social researchers have tended to avoid asking respondents to rank items because the resulting data did not fit into any standard analytic scheme. The methods we have presented here should make that option much more attractive. The methods are extremely flexible, allow for a wide range of research designs, and make it possible to answer many different questions about the process governing the rankings. They are computationally practical and can be implemented with most standard statistical packages. And, finally, the models are closely related to standard multinomial logit models, which means that the resulting parameters have a familiar and appealing interpretation.

APPENDIX A

SAS code for estimating the model reported in Table 1:

```
PROG PHREG NOSUMMARY;
MODEL RANK = BP IF IP MV HD TN TF /
  TIES=EXACT;
STRATA=ID;
TEST BP - IF;
.
.
.
TEST TN - TF;
```

Notes:

1. BP through TF are dummy variables corresponding to seven of the eight items, excluding antibiotics.
2. The variable ID contains an identification number for each respondent.
3. The TEST statements are optional. Each produces a Wald chi-square test for the null hypotheses that coefficients for the two named variables are equal.

4. The NOSUMMARY option suppresses output that is both voluminous and irrelevant for this application.

APPENDIX B

SAS code for estimating the model reported in Table 2:

```
PROC PHREG NOSUMMARY;
MODEL RANK = BP IF IP MV HD TN TF
  SEXBP SEXIF SEXIP SEXMV SEXHD SEXTN
  SEXTF / TIES = EXACT;
STRATA = ID;
TEST SEXBP, SEXIF, SEXIP, SEXMV, SEXHD,
  SEXTN, SEXTF;
```

Notes:

1. The DATA step (not shown) must contain a series of statements defining the interaction terms, e.g., SEXBP = SEX*BP where SEX is a dummy variable.
2. The TEST statement produces a Wald chi-square test for the null hypothesis that all the named variables have coefficients of 0.

APPENDIX C

SAS code for estimating the model reported in Table 3.

```
PROC PHREG NOSUMMARY;
MODEL RANK = BP IF IP MV HD TN TF
  AGEBP AGEIF AGEIP AGEMV AGEHD
  AGETN AGETF SEXBP SEXIF SEXIP SEXMV
  SEXHD SEXTN SEXTF / TIES = EXACT;
STRATA ID;
TEST AGEBP, AGEIF, AGEIP, AGEMV,
  AGEHD, AGETN, AGETF;
TEST SEXBP, SEXIF, SEXIP, SEXMV, SEXHD,
  SEXTN, SEXTF;
```

APPENDIX D

SAS code for estimating the model reported in Table 4.

```
PROC PHREG NOSUMMARY;
MODEL RANK = BP IF IP MV HD TN TF
  BPTIM IFTIM IPTIM MVTIM HDTIM TNTIM
  TFTIM / TIES = EXACT;
IF RANK >= 4.5 THEN TDUM = 1; ELSE
TDUM = 0;
BPTIM = TDUM*BP;
IFTIM = IF*TDUM;
IPTIM = IP*TDUM;
MVTIM = MV*TDUM;
HDTIM = HD*TDUM;
TNTIM = TN*TDUM;
TFTIM = TF*TDUM;
STRATA ID;
TEST BPTIM, IFTIM, IPTIM, MVTIM, HDTIM,
  TNTIM, TFTIM;
```

Note: The "time dependent" variables are defined *after* they are first listed in the MODEL statement.

REFERENCES

Agresti, Alan. 1990. *Categorical Data Analysis.* New York: Wiley.

Amemiya, Takeshi. 1985. *Advanced Econometrics.* Cambridge: Harvard University Press.

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

Bradley, Ralph Allen, and Milton E. Terry. 1952. "Rank Analysis of Incomplete Block Designs." *Biometrika* 39:324–35.

Breslow, Norman E. 1974. "Covariance Analysis of Censored Survival Data." *Biometrics* 30:89–99.

Chapman, Randall G., and Richard Staelin. 1982. "Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model." *Journal of Marketing Research* 14:288–301.

Christakis, Nicholas A., and David A. Asch. 1993. "Biases in How Physicians Choose to Withdraw Life Support." *The Lancet* 342:642–46.

Cox, David R., and D. Oakes. 1984. *Analysis of Survival Data.* London: Chapman and Hall.

Critchlow, Douglas E., and Michael A. Fligner. 1993. "Ranking Models with Item Covariates." In *Probability Models and Statistical Analyses for Ranking Data,* edited by Michael A. Fligner and Joseph S. Verducci, 1–19. New York: Springer-Verlag.

Critchlow, Douglas E., Michael A. Fligner, and Joseph Verducci. 1991. "Probability Models on Rankings." *Journal of Mathematical Psychology* 35:294–318.

Croon, M. A., and R. Luijkx. 1993. "Latent Structure Models for Ranking Data." In *Probability Models and Statistical Analyses for Ranking Data,* edited by Michael A. Fligner and Joseph S. Verducci, 53–74. New York: Springer-Verlag.

Duncan, Otis Dudley, and Charles Brody. 1982. "Analyzing *n* Rankings of Three Items." In *Social Structure and Behavior,* edited by Robert M. Hauser, David Mechanic, Archibald O. Haller, and Taissa S. Hauser, 269–310. New York: Academic Press.

Efron, Bradley. 1977. "The Efficiency of Cox's Likelihood Function for Censored Data." *Journal of the American Statistical Association* 72:557–65.

Farewell, V. T., and Ross L. Prentice. 1980. "The Approximation of Partial Likelihood with Emphasis on Case-Control Studies." *Biometrics* 67:273–78.

Fligner, Michael A., and Joseph S. Verducci. 1988. "Multistage Ranking Models." *Journal of the American Statistical Association* 83:892–901.

Greene, William H. 1992. *LIMDEP: User's Manual and Reference Guide Version 6.0.* Bellport, N.Y.: Econometric Software.

Hausman, Jerry A., and Paul A. Ruud. 1987. "Specifying and Testing Econometric Models for Rank-Ordered Data." *Journal of Econometrics* 34: 83–104.

Hill, Daniel H., William G. Axinn, and Arland Thornton. 1993. "Competing Hazards with Shared Unmeasured Risk Factors." In *Sociological Methodology 1993,* edited by Peter V. Marsden, 245–77. Oxford: Basil Blackwell.

Hollander, M., and Sethuraman, J. 1978. "Testing for Agreement Between Two Groups of Judges." *Biometrika* 65:403–41.

Jackson, D. J., and Alwin, D. F. 1981. "The Factor Analysis of Ipsative Measures." *Sociological Methods and Research* 9:218–38.

Kalbfleisch, John D., and Ross L. Prentice. 1980. *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Kamakura, Wagner A., and José Alfonso Mazzon. 1991. "Value Segmentation: A Model for the Measurement of Values and Value Systems." *Journal of Consumer Research* 18:208–18.

Katahira, Hotaka. 1990. "Perceptual Mapping Using Ordered Logit Analysis." *Marketing Science* 9:1–17.

Keener, Robert W., and Donald M. Waldman. 1985. "Maximum Likelihood Regression of Rank-Censored Data." *Journal of the American Statistical Association* 80:385–92.

Kohn, Melvin. 1969. *Class and Conformity: A Study in Values*. Homewood, Ill. Irwin.

Kraemer, H. C. 1981. "Intergroup Concordance: Definition and Estimation." *Biometrika* 68:641–46.

Lareau, Thomas J., and Douglas A. Rae. 1989. "Valuing WTP for Diesel Odor Reductions: An Application of Contingent Ranking Technique." *Southern Economic Journal* 55:728–42.

Li, L., and W. R. Schucany. 1975. "Some Properties of a Test for Concordance of Two Groups of Rankings." *Biometrika* 62:417–23.

Luce, R. D. 1959. *Individual Choice Behavior*. New York: Wiley.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by Paul Zarembka, 105–42. New York: Academic Press.

Moore, Laurence. 1990. "Segmentation of Store Choice Models Using Stated Preferences." *Papers of the Regional Science Association* 69:121–31.

Plackett, R. L. 1975. "The Analysis of Permutations." *Applied Statistics* 24:193–201.

Punj, Girish N., and Richard Staelin. 1978. "The Choice Process for Graduate Business Schools." *Journal of Marketing Research* 15:588–98.

SAS Institute. 1992. *SAS/STAT Software: The PHREG Procedure, Version 6*. SAS Technical Report P-217, Cary, N.C.

Schucany, W. R., and J. Beckett. 1976. "Analysis of Multiple Sets of Incomplete Rankings." *Communications in Statistics and Theoretical Mathematics*, A5: 1327–34.

Schucany, W. R., and W. H. Frawley. 1973. "A Rank Test for Two Group Concordance." *Psychometrika* 38:249–58.

Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

Stern, Hal. 1993. "Probability Models on Rankings and the Electoral Process." In *Probability Models and Statistical Analyses for Ranking Data*, edited by Michael A. Fligner and Joseph S. Verducci, 173–95. New York: Springer-Verlag.

Thurstone, L. L. 1927. "A Law of Comparative Judgment." *Psychological Review* 34:273–86.

Yellott, John I., Jr. 1977. "The Relationship Between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the Double Exponential Distribution." *Journal of Mathematical Psychology* 15:109–44.

LINKED CITATIONS

Logit Models for Sets of Ranked Items
Paul D. Allison; Nicholas A. Christakis
Sociological Methodology, Vol. 24. (1994), pp. 199-228.
Stable URL:
http://links.jstor.org/sici?sici=0081-1750%281994%2924%3C199%3ALMFSOR%3E2.0.CO%3B2-A

References

Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons
Ralph Allan Bradley; Milton E. Terry
Biometrika, Vol. 39, No. 3/4. (Dec., 1952), pp. 324-345.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28195212%2939%3A3%2F4%3C324%3ARAOIBD%3E2.0.CO%3B2-C

Covariance Analysis of Censored Survival Data
N. Breslow
Biometrics, Vol. 30, No. 1. (Mar., 1974), pp. 89-99.
Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28197403%2930%3A1%3C89%3ACAOCSD%3E2.0.CO%3B2-S

Exploiting Rank Ordered Choice Set Data within the Stochastic Utility Model
Randall G. Chapman; Richard Staelin
Journal of Marketing Research, Vol. 19, No. 3. (Aug., 1982), pp. 288-301.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28198208%2919%3A3%3C288%3AEROCSD%3E2.0.CO%3B2-S

The Efficiency of Cox's Likelihood Function for Censored Data
Bradley Efron
Journal of the American Statistical Association, Vol. 72, No. 359. (Sep., 1977), pp. 557-565.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28197709%2972%3A359%3C557%3ATEOCLF%3E2.0.CO%3B2-X

LINKED CITATIONS

Multistage Ranking Models
Michael A. Fligner; Joseph S. Verducci
Journal of the American Statistical Association, Vol. 83, No. 403. (Sep., 1988), pp. 892-901.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198809%2983%3A403%3C892%3AMRM%3E2.0.CO%3B2-3

Testing for Agreement Between Two Groups of Judges
Myles Hollander; Jayaram Sethuraman
Biometrika, Vol. 65, No. 2. (Aug., 1978), pp. 403-410.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28197808%2965%3A2%3C403%3ATFABTG%3E2.0.CO%3B2-6

Value Segmentation: A Model for the Measurement of Values and Value Systems
Wagner A. Kamakura; José Afonso Mazzon
The Journal of Consumer Research, Vol. 18, No. 2. (Sep., 1991), pp. 208-218.
Stable URL:
http://links.jstor.org/sici?sici=0093-5301%28199109%2918%3A2%3C208%3AVSAMFT%3E2.0.CO%3B2-G

Perceptual Mapping Using Ordered Logit Analysis
Hotaka Katahira
Marketing Science, Vol. 9, No. 1. (Winter, 1990), pp. 1-17.
Stable URL:
http://links.jstor.org/sici?sici=0732-2399%28199024%299%3A1%3C1%3APMUOLA%3E2.0.CO%3B2-W

Maximum Likelihood Regression of Rank-Censored Data
Robert W. Keener; Donald M. Waldman
Journal of the American Statistical Association, Vol. 80, No. 390. (Jun., 1985), pp. 385-392.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198506%2980%3A390%3C385%3AMLRORD%3E2.0.CO%3B2-M

Intergroup Concordance: Definition and Estimation
Helena Chmura Kraemer
Biometrika, Vol. 68, No. 3. (Dec., 1981), pp. 641-646.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28198112%2968%3A3%3C641%3AICDAE%3E2.0.CO%3B2-L

LINKED CITATIONS

Valuing WTP for Diesel Odor Reductions: An Application of Contingent Ranking Technique
Thomas J. Lareau; Douglas A. Rae
Southern Economic Journal, Vol. 55, No. 3. (Jan., 1989), pp. 728-742.
Stable URL:
http://links.jstor.org/sici?sici=0038-4038%28198901%2955%3A3%3C728%3AVWFDOR%3E2.0.CO%3B2-L

Some Properties of a Test for Concordance of Two Groups of Rankings
Loretta Li; William R. Schucany
Biometrika, Vol. 62, No. 2. (Aug., 1975), pp. 417-423.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28197508%2962%3A2%3C417%3ASPOATF%3E2.0.CO%3B2-B

The Analysis of Permutations
R. L. Plackett
Applied Statistics, Vol. 24, No. 2. (1975), pp. 193-202.
Stable URL:
http://links.jstor.org/sici?sici=0035-9254%281975%2924%3A2%3C193%3ATAOP%3E2.0.CO%3B2-8

The Choice Process for Graduate Business Schools
Girish N. Punj; Richard Staelin
Journal of Marketing Research, Vol. 15, No. 4. (Nov., 1978), pp. 588-598.
Stable URL:
http://links.jstor.org/sici?sici=0022-2437%28197811%2915%3A4%3C588%3ATCPFGB%3E2.0.CO%3B2-W