COMMUN. STATIST.—SIMULA., 27(1), 147–166 (1998)

AN EMPIRICAL COMPARISON OF INFERENCE USING ORDER-RESTRICTED AND LINEAR LOGIT MODELS FOR A BINARY RESPONSE

Alan Agresti and Brent A. Coull

Department of Statistics
University of Florida
Gainesville, Florida 32611-8545

Key Words: contingency table; isotonic regression; logistic regression; monotonicity; trend test.

ABSTRACT

In many applications with a binary response and an ordinal or quantitative predictor, it is natural to expect the response probability to change monotonically. Two possible models are a linear model with some link, such as the linear logit model, and a more general order-restricted model that assumes monotonicity alone. The order-restricted approach is more complex to apply, and we investigate whether it may be worth the extra effort. Specifically, suppose the order restriction truly holds but a simpler linear model does not. For testing the hypothesis of independence, is there the potential of a substantive power gain by performing an order-restricted test? For estimating a set of binomial parameters, how large must the sample size be before the consistency of the order-restricted estimates and inconsistency of the model-based estimates makes a substantive difference to mean square errors?

We conducted a limited simulation study comparing estimators and likelihood-ratio tests for the linear logit model and for the order-restricted model. Results suggest that order-restricted inference is preferable for moderate to large sample sizes when the true probabilities take only a couple of levels, such as in a dose–response experiment when all doses provide a uniform improvement over placebo.

If the true probabilities are strictly monotone but deviate somewhat from the linear logit model, the logit-based inference is usually more powerful unless the sample size is extremely large. When the true probabilities may have slight departures from monotonicity, the order-restricted estimates often perform better, particularly for moderate to large samples.

1. INTRODUCTION

In many applications, it is natural to predict that the relationship between two variables satisfies a rather vague condition such as "Y tends to increase as X increases." A common way for a statistician to handle this is to construct a generalized linear model in which X has a linear effect, on some scale. For a binary response Y with $P(Y = 1 \mid X = x) = \pi(x)$, for instance, one might use the linear logit model

$$\text{logit}[\pi(x)] = \alpha + \beta x \tag{1}$$

with the expectation that $\beta > 0$.

An alternative and more general model assumes solely an order-restricted effect. This model treats the explanatory variable as a factor but imposes a monotonicity constraint on parameters. For instance, suppose the explanatory variable has r levels, $x_1 < x_2 < ... < x_r$, and one expects a positive association. Then, the model takes the more general form

$$\pi(x_i) = \alpha + \beta_i, \tag{2}$$

with constraint $\beta_1 \leq \beta_2 \leq ... \leq \beta_r$.

A substantial literature exists on methods for order-restricted inference, particularly for normal responses. Barlow, Bartholomew, Bremner, and Brunk (1972) and Robertson, Wright, and Dykstra (1988) provided surveys. Despite this extensive literature, order-restricted methods seem to be rarely used for practical application. They do not typically appear in basic texts on applied statistics, in more specialized texts on models for continuous or discrete data, or in the major software packages. There are undoubtedly a variety of reasons for this. First, the methods are not simple to apply, requiring algorithms and/or nonstandard limiting distributions even for simple problems such as the one-way layout with normal responses. This should not be a major hurdle, however, in the modern computing age. Second, much of the survey literature is not easily readable by nonstatisticians or by applied statisticians wanting an introduction to the methods. The area would be well-served by an applied version of the fine theoretical text by Robertson et al. (1988).

A natural question to ask is whether it is worth the extra effort to conduct order-restricted inference. For instance, if the order restriction (2) truly holds but the simpler linear logit model (1) does not, is there the potential of a substantive power gain by performing an order-restricted test instead of the linear logit test or the simple chi-squared test of independence? Do the order-restricted estimates of the true proportions tend to be better than the linear logit estimates and the sample proportions?

This paper studies these questions about order-restricted inference in the context of binary response data. If the linear logit model (1) holds or nearly holds, it is natural to use inference based on it. In practice, though, one would often expect the binomial parameters to increase monotonically without satisfying, even approximately, the linear logit model. Yet, there may be no a priori reason to choose a particular alternative link or structural form for the relationship. Is order-restricted inference superior as one moves sufficiently far away from the linear logit model? One would hope so, but the order-restricted test must counteract the parsimonious benefit the logit-based test has of focusing inference on a single parameter.

Section 2 summarizes order-restricted methods for comparing binomial parameters. Section 3 presents results of a limited simulation study comparing the likelihood-ratio tests of independence applied to $r \times 2$ contingency tables with monotone probabilities for (1) the linear logit model, (2) the more general order restriction, and (3) the most general model. Results suggest that the order-restricted test is preferable to the linear logit test (or the ordinary chi-squared test for the most general model) when the response probability takes a jump but then stays essentially constant; otherwise, the linear logit test performs well even when the model does not hold.

Section 4 studies the potential improvement in estimation from using order-restricted methods. Similar results hold. The main advantage for order-restricted estimation accrues for moderate to large samples when the true probabilities take only a couple of values. Section 5 compares the methods when the probabilities have an irregular monotone trend and may have occasional probabilities that are "out-of-order." Then, order-restricted estimates tend to perform better than the linear logit estimates, and for small samples they also perform better than the sample proportions.

The final section of the article comments about possible reasons for the limited use in practice of order-restricted methods and suggests future research that may be fruitful for increasing their scope and utility.

2. ORDER-RESTRICTED METHODS FOR CONTINGENCY TABLES WITH BINARY RESPONSE

Consider a $r \times 2$ contingency table comparing independent binomial samples at r ordered levels of a predictor. Let n_{i1} denote the number of "successes" out of n_i trials at level x_i of the predictor, and let n_{i2} denote the number of "failures." We assume that $\{n_{i1}, i = 1, ..., r\}$ are independent binomial variates with parameters $\{\pi_i, i = 1, ..., r\}$. Let $p_i = n_{i1}/n_i$, $i = 1, ..., r$, denote the sample proportions.

Bartholomew (1959) presented one of the first order-restricted tests for contingency tables, in testing that $\pi_1 = \pi_2 = ... = \pi_r$ against the alternative $\pi_1 \leq \pi_2 \leq ... \leq \pi_r$. Under the null, the maximum likelihood (ML) estimator of π_i is the overall sample proportion $p = (\sum n_{i1})/(\sum n_i)$. If $p_1 \leq p_2 \leq ... \leq p_r$, then the order-restricted ML estimator of π_i equals $\tilde{\pi}_i = p_i$. Otherwise, one pools "out-of-order" categories until the sample proportions are monotone, and the order-restricted ML estimates of proportions for the original categories are the sample proportions for this partition of categories. Bartholomew's test statistic equals the usual Pearson chi-squared statistic applied to the collapsed table that combines rows having sample proportions falling out of order. The collapsed table and the order-restricted estimates can be calculated using the pool adjacent violators algorithm for isotonic regression (e.g., Robertson et al. 1988, Sec. 1.2).

Alternative large-sample procedures (Barlow et al. 1972, p. 193) include ones based on the approximate normality of the sample proportions, with an inverse sine transformation possibly applied to stabilize the variance. Robertson et al. (1988, p. 167) presented the likelihood-ratio statistic for the order-restricted binomial problem as a special case of a test for parameters in an exponential family distribution (Robertson and Wegman, 1978). The likelihood-ratio test statistic for testing independence in the $r \times 2$ table, assuming the order-restricted model, is

$$G^2(I|O) = \sum n_{i1} \log(\tilde{\pi}_i/p) + \sum n_{i2} \log[(1 - \tilde{\pi}_i)/(1 - p)].$$

The large-sample distribution of this test statistic, like those of most others in order-restricted inference for categorical data, is *chi-bar squared.* This distribution refers to random variables of form $\sum_{d=1}^{r} p_d \chi^2_{d-1}$, where χ^2_d is a chi-squared variate with d degrees of freedom and where the 'weight' p_d is the probability that the order-restricted solution has d distinct sets on which the estimates are level.

For the order-restricted binomial problem, the weights for the chi-bar-squared statistic depend on the $\{n_i\}$. Robertson et al. (1988) provided tables of the distribution for $r = 3$ and 4 and provided critical values for larger r values for the equal sample size case. The large-sample chi-bar-squared approximation for the distribution of the likelihood-ratio statistic may be inadequate if some null expected cell counts are small. Agresti and Coull (1996) presented small-sample tests using the likelihood-ratio statistic and provided extensions for stratified tables, and Eddy et al. (1995) discussed a model having monotone partial effects for multiple predictors in a model.

3. A POWER COMPARISON OF ORDER-RESTRICTED AND LINEAR LOGIT TESTS

This section compares likelihood-ratio tests of $\pi_1 = \pi_2 = ... = \pi_r$. Suppose that the rows are ordered and the binomial parameter increases in the row index. Are we better off using the order-restricted test for the alternative of ordered binomial parameters, or the test based on the linear trend model (1)?

3.1 Design of power study

Let $G^2(I)$ denote the ordinary likelihood-ratio statistic for testing independence against the general alternative that the $\{\pi_i\}$ are not all equal. Then we have that

$$G^2(I) = 2\sum_{i,j} n_{ij}\log[n_{ij}/\hat{\mu}_{ij}],$$

where $\hat{\mu}_{ij} = n_{i+}n_{+j}/n$. Let $G^2(I|L)$ denote the likelihood-ratio statistic for testing H_0: $\beta = 0$ against H_a: $\beta > 0$ for the linear logit model. Because of the nesting of the alternative parameter spaces, it follows that $G^2(I|L) \leq G^2(I|O) \leq G^2(I)$.

The null asymptotic distribution for $G^2(I)$ is chi-squared with $df = r - 1$. The null asymptotic distribution for the likelihood-ratio test of H_0: $\beta = 0$ against H_a : $\beta \neq 0$ in model (1) is chi-squared with $df = 1$. For the one-sided alternative, H_a: $\beta > 0$, the null asymptotic distribution of $G^2(I|L)$ is $(1/2)\chi_0^2 + (1/2)\chi_1^2$, where χ_0^2 is degenerate at 0. By contrast, the null asymptotic distribution of $G^2(I|O)$ is a weighting of chi-squared variates with df ranging from 0 to $r - 1$.

If the linear logit model truly holds with $\beta > 0$, then the logit statistic $G^2(I|L)$ is approximately noncentral chi-squared with $df = 1$. In that case, $P(G^2(I|O) =$

$G^2(I)$) converges to 1 as n increases; thus, $G^2(I|O)$ is approximately noncentral chi-squared with $df = (r-1)$ and the same noncentrality parameter as the logit statistic. From standard results (e.g., Das Gupta and Perlman, 1974), when the linear logit model holds with $\beta > 0$, the logit test is asymptotically more powerful than the order-restricted test, which is itself asymptotically more powerful than the test for the general alternative.

To make power comparisons of the tests based on $G^2(I|L)$, $G^2(I|O)$, and $G^2(I)$, we conducted a small simulation study. Using 100,000 simulations at each combination of study factors, we estimated the probability that the test statistic falls in the rejection region for a test of fixed size. The simulation study used the following factors in a factorial design, with the levels indicated.

1. Number of rows r: 3, 5, or 7.
2. Total sample size n: 100 or 250, allocated equally or as equally as possible among rows (For instance, 100 trials were allocated to three rows as 34, 33, 33).
3. True model:

   a. $\text{logit}(\pi_i) = \alpha + \beta(i-1)/(r-1)$

   b. $\text{logit}(\pi_i) = \alpha + \beta(i-1)^2/(r-1)^2$

   c. $\text{logit}(\pi_i) = \alpha + \beta(\log i)/(\log r)$

   d. $\text{logit}(\pi_i) = \alpha + \beta I(i \geq 2)$

   e. $\text{logit}(\pi_i) = \alpha + \beta I(i \geq r/2)$

   f. $\text{logit}(\pi_i) = \alpha + (\beta/2)I(i \geq 2) + (\beta/2)I(i \geq 3)$

   The I functions in the last three cases are indicator functions. For comparability, we parameterized the six cases so that $\text{logit}(\pi_1) = \alpha$ and $\text{logit}(\pi_r) = \alpha + \beta$.

4. Parameter values for true models: α = -0.7 and β = 0, .5, 1.0, 1.5. These give success probabilities varying from .332 at $i = 1$ to .332, .450, .574, .690 (for the four beta choices) at $i = r$.

The six choices for the model form for the true probabilities all satisfy the order-restricted model (2). They correspond respectively to (a) linear logit model with positive trend truly holding; (b) logit model with quadratic increase; (c) logit model with log increase; (d) first level differs from others, which are identical; (e) shift in probability at midpoint of levels, with constant value below and above it; (f) linear increase for three levels, then plateau. The study used relatively large sample sizes ($n = 100$ and 250) to ensure that the actual size of the large-sample chi-squared and chi-bar-squared tests did not deviate greatly from the nominal size. This was checked by the estimated power when $\beta = 0$.

3.2 Results of power study

Table I shows the estimated powers for the linear logit and order-restricted tests. The standard error of the estimates equals .0007 when the true power is .05 (or .95) and .0016 when the true power is .50. When the linear logit model truly holds, the power for the order-restricted test was as much as .04 less than the power of the model-based test, this difference occuring when both powers were in the middle of the range. The quadratic and log increases exemplify cases in which the true relationship is strictly monotone but not linear. The model-based test maintains its superiority in these cases, but not as markedly. For the plateau cases, the two tests are not much different except in case (d), in which a single step follows the first level. This case exhibits some large differences in power, with the advantage going to the order-restricted analysis. These latter results are qualitatively similar to those obtained by Shi (1991, Table III) in comparing analogous score tests (e.g., Mantel's trend test). The differences become more noticeable as r increases. When $n = 100$ with $r = 5$ or 7, however, this result is tempered by the fact that the order-restricted test shows evidence of having true size somewhat larger than the nominal value. A referee pointed out to us that similar behavior occurs for order-restricted methods in other contexts: for instance, Silvapulle (1992) obtained better approximations for the null distribution by replacing weighted chi-squared terms by weighted F terms.

For contrast, the $n = 250$ part of the table also displays the power of the ordinary chi-squared likelihood-ratio test, $G^2(I)$. Not surprisingly, this test fares poorly compared to the others, particularly as r (and hence df) increases. The only exception is case (d), in which the test based on the linear logit model fails to uniformly dominate it. Since this chi-squared test refers to a general alternative, it is weak for directed alternatives such as monotone trends.

TABLE I

Estimated Powers for Likelihood-Ratio Tests Based on Logit Model with Linear (first entry), Order-Restricted (second entry), and Saturated (third entry, $n = 250$) Alternatives.

True	r = 3				r = 5				r = 7			
	Size of Effect (Beta)											
Relation	0	.5	1.0	1.5	0	.5	1.0	1.5	0	.5	1.0	1.5
					n=100							
a	.051	.241	.621	.905	.052	.216	.543	.836	.051	.202	.498	.801
	.052	.244	.616	.893	.054	.205	.507	.801	.054	.195	.466	.757
b	.051	.249	.624	.903	.051	.221	.541	.832	.051	.203	.493	.790
	.051	.248	.628	.909	.055	.211	.519	.816	.055	.194	.476	.765
c	.052	.241	.622	.904	.051	.209	.525	.817	.052	.186	.467	.761
	.051	.247	.621	.900	.054	.206	.501	.791	.055	.188	.447	.732
d	.052	.237	.623	.914	.052	.164	.402	.680	.052	.126	.288	.508
	.053	.269	.697	.949	.055	.193	.503	.822	.056	.159	.400	.704
e	.051	.235	.623	.913	.052	.272	.676	.934	.052	.273	.680	.939
	.051	.269	.700	.949	.054	.262	.679	.944	.055	.256	.671	.941
f	.052	.242	.623	.904	.053	.215	.543	.833	.050	.169	.424	.706
	.052	.243	.612	.893	.054	.219	.553	.850	.055	.187	.471	.769
					n=250							
a	.053	.477	.933	.999	.050	.393	.865	.993	.051	.357	.829	.988
	.051	.448	.922	.999	.052	.356	.832	.989	.053	.325	.784	.978
	.052	.274	.820	.993	.054	.168	.583	.929	.055	.133	.459	.846
b	.052	.488	.934	.999	.050	.391	.868	.993	.052	.358	.828	.987
	.051	.462	.937	.999	.052	.365	.849	.992	.052	.327	.800	.983
	.052	.297	.851	.996	.052	.177	.626	.950	.055	.140	.493	.874
c	.053	.474	.932	.999	.051	.376	.848	.991	.050	.328	.789	.979
	.052	.448	.927	.999	.051	.352	.825	.988	.051	.309	.760	.972
	.052	.277	.828	.994	.054	.164	.580	.930	.054	.129	.437	.828
d	.053	.464	.932	.999	.051	.276	.703	.958	.052	.203	.504	.822
	.050	.504	.967	1.000	.051	.346	.858	.995	.052	.261	.727	.971
	.050	.343	.919	.999	.053	.201	.703	.977	.054	.140	.502	.889
e	.052	.462	.932	.999	.051	.492	.951	1.000	.050	.499	.956	1.000
	.052	.502	.967	1.000	.051	.486	.963	1.000	.054	.474	.959	1.000
	.052	.352	.918	.999	.052	.291	.881	.999	.055	.250	.842	.997
f	.052	.476	.931	.999	.050	.385	.865	.993	.051	.290	.721	.958
	.049	.448	.923	.999	.050	.387	.881	.996	.052	.317	.800	.986
	.051	.276	.819	.993	.054	.201	.704	.976	.055	.151	.541	.915

NOTE: True relation (a) = linear logit, (b) = quadratic, (c) = log, (d) = step at level 2, (e) = change in level midway, (f) = linear then plateau

In summary, the good news for the order-restricted approach is that it has weaker assumptions than the linear logit model but never fares much worse than the logit test. The bad news is that it often has weaker power even if reality departs quite strongly from the linear logit model. However, these results do suggest an applica-

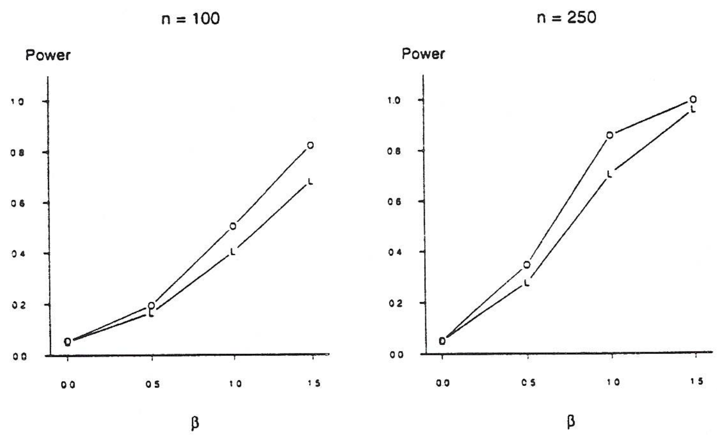

FIG. 1: Estimated Powers for Likelihood-Ratio Tests Based on Linear Logit Model (L) and Order-Restricted Model (O), When the Data are Generated from Model (d) with r=5 (Based on 100,000 Simulations)

tion in which the order-restricted test is preferable to the logit test or the ordinary chi-squared test. Consider a dose–response experiment comparing zero dosage (e.g., placebo) to several doses of a drug. Suppose one expects the drug to have an effect, but such that successively increasing dose levels may not provide much if any improvement over the lowest dose administered. Then, based on the power results shown for case (d), the order-restricted test would be much preferred. Figure 1 illustrates this case, plotting the powers for the order-restricted and linear logit tests for each n when $r = 5$.

4. A COMPARISION OF ORDER-RESTRICTED, LINEAR LOGIT, AND SAMPLE PROPORTION ESTIMATORS

An order-restricted approach may be more useful for estimation than testing. When one expects an increasing trend in binomial probabilities, the order-restricted fit smooths the sample data toward that expectation, though the smoothing is less severe than with a parametric model. Suppose the true probabilities satisfy the order restriction, but not the linear logit model. Then, one would expect the order-restricted damping of the sampling error for small samples to result in estimated cell probabilities with considerably smaller mean squared errors than the sample proportions have; yet, for large samples, those estimates would maintain consistency, like the sample proportions but unlike the linear logit estimates. If an order restriction truly holds but the linear logit model does not, how large must the sample size be before the consistency of the order-restricted estimates and inconsistency of the model-based estimates makes a substantive difference to mean square errors?

To describe the extent of potential improved estimation provided by order-restricted approaches, we compared the sample proportions, order-restricted estimates, and linear logit model-based estimates of the proportions, when the true relationships are the ones mentioned in the previous section. For any estimator $\tilde{\pi}_i$ of π_i, we evaluated the average mean square error of the estimators of the r probabilities,

$$[\sum_i E(\tilde{\pi}_i - \pi_i)^2]/r.$$

For the sample proportions $\{p_i\}$ with sample size n allocated equally among the r rows, this equals $\sum_i[\pi_i(1-\pi_i)]/n$. For the other cases, we estimated the average mean squared errors by simulating 100,000 sample tables. When the true probabilities are strictly monotone increasing, the probability that the order-restricted estimates are identical to the sample proportions converges to 1 as $n \to \infty$, since asymptotically the sample proportions satisfy the ordering by the Law of Large Numbers. Thus, asymptotically their average mean squared errors are identical in this case.

We performed the comparison for sample sizes of $n = 25$, 100, 250, 500 and 1000. Table II shows the sample-size normalized MSE values (i.e., multiplied by n, to make results comparable for differing n) for $\beta = 1.5$. For contrast, the bottom panel contains results for $\beta = 0$, in which case all the models hold and all estimators are consistent. The estimated mean squared errors have standard error values no greater than .006.

When there is no effect ($\beta = 0$), both the linear logit and order-restricted estimators perform much better than the sample proportion. The relative performance of the three estimators is the same at each sample size, for given r. As r increases, the relative performance of the linear logit model improves. In fact, Table II shows that for every case considered, the sample proportions deteriorate as r increases, since there are additional parameters to estimate with the same total amount of data; by contrast, when the linear logit model holds, the average mean square error for that model is relatively stable in r.

TABLE II

Estimated Average Mean Squared Errors for Estimating Binomial Probabilities Using Linear Logit Model (first entry), Order-Restricted Estimates (second entry), and Sample Proportions (third entry), for Models with $\beta = 1.5$.

True Relation	r = 3					r = 5					r = 7				
	Sample Size														
	25	100	250	500	1000	25	100	250	500	1000	25	100	250	500	1000
Linear logit	0.45	0.45	0.45	0.45	0.45	0.46	0.46	0.46	0.46	0.46	0.48	0.46	0.46	0.46	0.46
	0.52	0.63	0.68	0.69	0.69	0.62	0.76	0.90	1.02	1.11	0.69	0.82	0.98	1.13	1.31
	0.69	0.69	0.69	0.69	0.69	1.17	1.17	1.17	1.17	1.17	1.68	1.65	1.65	1.65	1.65
Quadratic	0.49	0.62	0.88	1.31	2.19	0.49	0.58	0.77	1.09	1.74	0.50	0.56	0.73	1.00	1.54
	0.52	0.60	0.64	0.66	0.68	0.62	0.76	0.88	0.96	1.01	0.69	0.82	0.98	1.11	1.24
	0.68	0.68	0.68	0.68	0.68	1.16	1.16	1.16	1.16	1.16	1.67	1.63	1.63	1.63	1.63
Log	0.46	0.49	0.56	0.68	0.93	0.48	0.53	0.65	0.85	1.25	0.49	0.55	0.69	0.92	1.39
	0.52	0.62	0.67	0.68	0.68	0.62	0.75	0.87	0.97	1.05	0.68	0.80	0.96	1.09	1.23
	0.68	0.68	0.68	0.68	0.68	1.16	1.16	1.16	1.16	1.16	1.65	1.63	1.63	1.63	1.63
Step, level 2	0.58	1.06	2.01	3.60	6.77	0.66	1.38	2.82	5.22	10.03	0.66	1.36	2.77	5.12	9.83
	0.51	0.54	0.54	0.55	0.55	0.60	0.66	0.67	0.67	0.67	0.66	0.72	0.75	0.74	0.75
	0.65	0.65	0.65	0.65	0.65	1.08	1.08	1.08	1.08	1.08	1.53	1.51	1.51	1.51	1.51
Linear, then plateau	0.58	1.06	2.01	3.59	6.77	0.61	1.13	2.20	3.96	7.50	0.62	1.15	2.24	4.07	7.71
	0.51	0.54	0.54	0.54	0.54	0.65	0.72	0.73	0.72	0.72	0.73	0.84	0.85	0.86	0.85
	0.65	0.65	0.65	0.65	0.65	1.09	1.09	1.09	1.09	1.09	1.55	1.52	1.52	1.52	1.52
Step midway	0.45	0.45	0.45	0.45	0.45	0.54	0.83	1.41	2.39	4.35	0.58	1.02	1.92	3.41	6.39
	0.52	0.63	0.68	0.68	0.69	0.62	0.75	0.83	0.86	0.87	0.67	0.80	0.90	0.94	0.96
	0.69	0.69	0.69	0.69	0.69	1.11	1.11	1.11	1.11	1.11	1.57	1.54	1.54	1.54	1.54
$\beta = 0$	0.45	0.44	0.44	0.44	0.44	0.45	0.45	0.45	0.45	0.44	0.46	0.44	0.45	0.44	0.44
	0.41	0.41	0.41	0.41	0.41	0.51	0.51	0.51	0.51	0.51	0.60	0.57	0.58	0.58	0.57
	0.67	0.67	0.67	0.67	0.67	1.11	1.11	1.11	1.11	1.11	1.58	1.55	1.55	1.55	1.55

When $\beta = 1.5$ and the linear logit model holds, that model performs best for all n and r. For small n the order-restricted estimator performs well, but as n increases it shares the same poor behavior as the sample proportion, since the probability that it is identical to that estimator converges toward 1.0.

In cases (b) and (c), the true probabilities are strictly monotone but do not satisfy the linear logit model. In this case the linear logit estimates behave well when n is less than 100. Although they deteriorate as n increases and the inconsistency starts to take effect, they maintain their superiority over the order-restricted estimates until the sample size is several hundred, unless r is small. When r is large, they maintain superiority over the sample proportions unless n is very large (e.g., over a thousand for $r = 7$).

For the plateau cases (d)-(f), the order-restricted estimates are not asymptotically normal, but a mean squared error comparison still seems reasonable as a way of summarizing the quality of estimation. In these cases, the linear logit model estimates deteriorate considerably more seriously. For small samples they behave comparably to the order-restricted estimates, but by $n = 100$ they are much poorer, and for sample sizes in the hundreds they are even much poorer than the sample proportions. For these cases, the order-restricted methods behave well for all n and their superiority over the sample proportions increases as r increases.

Similar patterns occur for other values of β. For cases where the linear logit model does not hold, the deterioration of the logit-based estimates requires larger sample sizes as $|\beta|$ decreases. To illustrate, Table III shows results for the same six true relationships, when $\beta = .5$ and $r = 5$. Figure 2 illustrates the superiority of the order-restricted estimates for case (d) of a jump after the first level. That figure plots the average MSE values for the three estimators as a function of the sample size, when $r = 5$, both for $\beta = 1.5$ and .5.

5. COMPARISIONS FOR DEPARTURES FROM MONOTONICITY

The comparisons of the past two sections assumed that the order-restricted model truly holds, but the simpler linear logit model may not hold. This encompasses a broad class of situations encountered in practice. Another broad class of practical situations are ones having overall monotone trends but with occasional small departures from monotonicity or with less regular trends than linear, quadratic, or log. We next studied how adversely the order-restricted and linear logit analyses are affected by slight nonmonotonicity or irregularity in the trend.

TABLE III

Estimated Average Mean Squared Errors for Estimating Binomial Probabilities Using Linear Logit Model (first entry), Order-Restricted Estimates (second entry), and Sample Proportions (third entry), for Models with $r = 5$ and $\beta = 0.5$.

True	Sample Size				
Relation	25	100	250	500	1000
Linear logit	0.48	0.47	0.47	0.48	0.47
	0.55	0.59	0.64	0.71	0.79
	1.18	1.18	1.18	1.18	1.18
Quadratic	0.47	0.48	0.50	0.54	0.61
	0.54	0.59	0.64	0.70	0.78
	1.16	1.16	1.16	1.16	1.16
Log	0.48	0.48	0.50	0.52	0.57
	0.55	0.59	0.65	0.71	0.79
	1.19	1.19	1.19	1.19	1.19
Step, level 2	0.51	0.60	0.77	1.05	1.62
	0.57	0.62	0.67	0.72	0.73
	1.21	1.21	1.21	1.21	1.21
Linear, then plateau	0.50	0.56	0.69	0.91	1.34
	0.57	0.64	0.71	0.76	0.78
	1.19	1.19	1.19	1.19	1.19
Step, midway	0.49	0.53	0.61	0.73	0.99
	0.56	0.61	0.68	0.74	0.81
	1.20	1.20	1.20	1.20	1.20

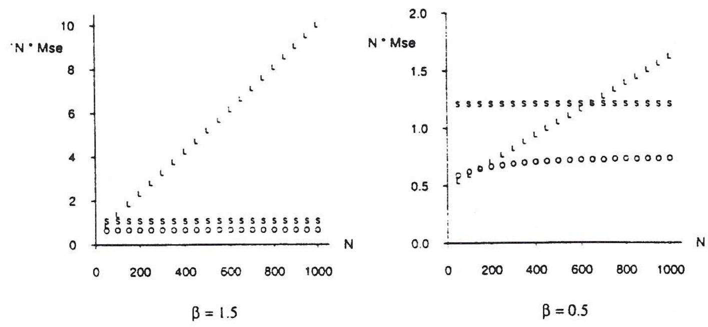

FIG. 2: N-Normalized MSE as a Function of N for Estimators Based on the Linear Logit Model (L), the Order-Restricted Model (O), and the Sample Proportions (S), When the Data are Generated from Model (d) with r=5 (Linear Logit and Order-Restricted MSE Values Based on 100,000 Simulations)

To study this situation, we used the following mixed-model extension of the linear logit model. Suppose that

$$\text{logit}(\pi_i) = \alpha + \beta(i-1)/(r-1) + \sigma z_i, \tag{3}$$

where $z_1, ..., z_r$ are independent N(0,1) variates. The expected value of the linear predictor is the linear logit model from the previous simulations, but adding the random effect to the model produces binomial probabilities that have an irregular increasing trend, with the potential of some departures from monotonicity. As σ increases, it is more likely that at least two of the binomial parameters are "out of order," and the degree of irregularity in the trend also increases.

We compared order-restricted, linear logit, and sample proportion inference assuming this model. To compare estimates, for each combination σ = 0, .125, .25, .50, .75, n = 25, 100, 250, 500, 1000, and r = 3, 5, 7, we generated data for 100,000 versions of the random effects model (i.e., for a random sample of 100,000 sets of standard normal variates $\{z_i\}$). We report results for the parameter values α = -.7 and β = 1.0 for the model. Note that the true probabilities differ for each of the 100,000 samples for each combination of σ, n, and r. For each sample, we compared the sample proportions, order-restricted estimates, and linear logit estimates to the true probabilities for the $\{z_i\}$ for that sample. We summarized the simulations for each combination of σ, n, and r by the sample-size-weighted mean squared error of each estimator around the true parameter values, averaged across the r response categories.

Table IV shows results. Specifically, if π_{ij} is the true binomial parameter for row i of the $r \times 2$ table in simulation j and if $\tilde{\pi}_{ij}$ is an estimate (sample proportion or order-restricted estimate or linear logit estimate), then Table IV reports

$$(n/100,000)\sum_j[\sum_i(\tilde{\pi}_{ij} - \pi_{ij})^2/r].$$

The estimated mean squared errors have standard error values ranging from .004 when $n \times \widehat{MSE} = 0.50$ to .005 when $n \times \widehat{MSE} = 1.00$ to .099 when $n \times \widehat{MSE} = 16.37$.

Note that the case $\sigma = 0$ is the ordinary linear logit model. Table IV shows that when $\sigma > 0$, unless σ is large, the linear logit or order-restricted estimates are better than the sample proportions when the sample size is small. Naturally, the model-based estimates deteriorate as n or σ increase. The deterioration is worse for the linear logit model than the order-restricted model. For a nonsevere potential degree of nonmonotonicity or irregularity (e.g., $\sigma \leq .25$), the order-restricted estimates

TABLE IV

Estimated Average Mean Squared Errors for Estimating Binomial Probabilities Using Linear Logit Model (first entry), Order-Restricted Estimates (second entry), and Sample Proportions (third entry), for Model (3) with Random-Effect and $\beta = 1.0$.

	r = 3					r = 5					r = 7				
	Sample Size														
σ	25	100	250	500	1000	25	100	250	500	1000	25	100	250	500	1000
0.00	0.47	0.47	0.47	0.47	0.47	0.48	0.48	0.48	0.48	0.48	0.50	0.48	0.48	0.48	0.48
	0.50	0.60	0.67	0.70	0.71	0.59	0.69	0.80	0.91	1.03	0.68	0.75	0.87	0.99	1.14
	0.72	0.72	0.71	0.72	0.71	1.20	1.20	1.20	1.20	1.20	1.73	1.69	1.69	1.69	1.69
.125	0.48	0.50	0.55	0.62	0.77	0.49	0.53	0.61	0.75	1.02	0.51	0.54	0.64	0.80	1.13
	0.50	0.59	0.65	0.68	0.70	0.60	0.71	0.81	0.90	1.00	0.68	0.78	0.91	1.04	1.19
	0.71	0.71	0.71	0.71	0.71	1.20	1.20	1.20	1.20	1.20	1.72	1.69	1.67	1.68	1.68
0.25	0.50	0.58	0.76	1.06	1.65	0.53	0.68	1.00	1.54	2.61	0.55	0.73	1.11	1.75	3.01
	0.50	0.58	0.64	0.67	0.72	0.61	0.75	0.89	1.05	1.30	0.71	0.86	1.08	1.35	1.79
	0.71	0.70	0.70	0.70	0.70	1.18	1.18	1.19	1.19	1.18	1.70	1.67	1.67	1.67	1.67
0.50	0.56	0.89	1.54	2.65	4.82	0.65	1.25	2.44	4.42	8.39	0.71	1.40	2.82	5.18	9.92
	0.51	0.66	0.87	1.21	1.86	0.68	1.00	1.55	2.38	4.11	0.80	1.22	1.99	3.20	5.56
	0.68	0.68	0.68	0.68	0.68	1.14	1.14	1.14	1.14	1.14	1.64	1.61	1.60	1.60	1.60
0.75	0.64	1.30	2.61	4.78	9.19	0.83	2.03	4.43	8.45	16.37	0.93	2.34	5.19	9.96	19.48
	0.56	0.90	1.56	2.63	4.80	0.78	1.47	2.80	5.01	9.37	0.93	1.84	3.58	6.48	12.23
	0.64	0.64	0.64	0.64	0.64	1.08	1.08	1.08	1.08	1.08	1.55	1.52	1.52	1.51	1.52

have the advantage of performance like the linear logit estimates for small n and like the sample proportions for large n.

For comparing performance of the tests, Table V shows results for the cases $r = 3, 5, 7$, $n = 100, 250$, $\beta = 0, .5, 1.0, 1.5$, for the same σ values for the random effect. Again, for small σ and n the model-based approaches have the advantage, as long as somewhat of a positive trend remains (i.e., $\beta > 0$). As σ increases and the frequency and extent of nonmonotonicity increase, the model-based methods lose their advantage. Again, the linear-logit inference deteriorates somewhat more quickly.

6. CONCLUSION

The comparisons of the previous two sections may be summarized as follows, for the moderate-sized values of r commonly encountered in practice (say, around 4 or 5): If the linear logit model appears to fit well, use inference based on it for any n. If the true probabilities are likely to be strictly monotone but deviate somewhat from the linear logit model, it is still best to use the logit-based inference unless the sample size is very large. For this large sample-size case, the order-restricted estimates perform slightly better than the sample proportions (unless the sample size is extremely large, in which case they are equivalent for estimation), though using the latter provides a protection against nonmonotonicity.

Suppose the true probabilities are likely to assume only a couple of levels, rather than be strictly increasing. Then inference based on the linear logit-model is fine for small sample sizes (e.g., 25 or 50), but with moderate to large samples it is better to use order-restricted methods. Finally, if the true probabilities tend to increase but need not be monotone, then the logit-based and order-restricted methods are better than inference based on sample proportions if the sample size is small, but as the sample size or the departure from monotonicity increases, the linear-logit based estimates deteriorate more quickly then the order-restricted estimates. In fact, for purposes of estimation, if the degree of potential nonmonotonicity is very small, then the order-restricted estimates have the advantage of performance like the linear logit estimates when n is small and like the sample proportions when n is large.

In summary, we found one situation in which order-restricted methods showed substantial improvement over linear logit and sample proportion estimates and thus would seem to be worth serious consideration. This is when the true relationship is close to a plateau following some initial jump, such as when all doses of a drug are

TABLE V

Estimated Powers for Likelihood-Ratio Tests Based on Logit Model with Linear (first entry), Order-Restricted (second entry), and Saturated (third entry) Alternatives for Model (3) with Random-Effect.

	r = 3				r = 5				r = 7			
	Size of Effect (Beta)											
σ	0	.5	1.0	1.5	0	.5	1.0	1.5	0	.5	1.0	1.5
	n=100											
0.00	.051	.241	.621	.905	.052	.216	.543	.836	.051	.202	.498	.801
	.052	.244	.616	.893	.054	.205	.507	.801	.054	.195	.466	.757
	.053	.134	.421	.771	.057	.101	.258	.533	.068	.093	.207	.426
.125	.061	.253	.614	.891	.058	.225	.539	.826	.057	.206	.499	.793
	.065	.261	.613	.884	.063	.220	.518	.796	.062	.204	.477	.754
	.073	.157	.435	.764	.074	.120	.280	.543	.079	.111	.228	.444
0.25	.087	.281	.596	.857	.074	.245	.531	.800	.067	.220	.496	.771
	.100	.307	.614	.861	.088	.262	.536	.792	.081	.239	.499	.753
	.132	.222	.473	.751	.125	.179	.340	.581	.127	.162	.288	.493
0.50	.159	.338	.564	.769	.125	.286	.513	.731	.105	.255	.479	.710
	.206	.400	.620	.808	.185	.365	.584	.779	.159	.331	.554	.756
	.344	.418	.573	.745	.341	.401	.528	.682	.333	.378	.491	.635
0.75	.224	.370	.541	.702	.177	.320	.493	.668	.152	.284	.470	.651
	.301	.460	.626	.769	.288	.450	.620	.765	.261	.421	.598	.756
	.548	.587	.670	.766	.590	.631	.700	.785	.593	.632	.700	.776
1.00	.267	.391	.525	.653	.220	.343	.483	.622	.190	.307	.457	.605
	.370	.498	.626	.740	.375	.510	.645	.761	.351	.492	.632	.759
	.680	.706	.751	.802	.758	.784	.820	.861	.782	.809	.840	.876
	n=250											
0.00	.053	.477	.933	.999	.050	.393	.865	.993	.051	.357	.829	.988
	.051	.448	.922	.999	.052	.356	.832	.989	.053	.325	.784	.978
	.052	.274	.820	.993	.054	.168	.583	.929	.055	.133	.459	.846
.125	.076	.479	.903	.997	.065	.398	.842	.989	.060	.364	.812	.982
	.081	.471	.901	.996	.075	.386	.824	.984	.069	.352	.784	.974
	.099	.326	.802	.986	.092	.218	.611	.922	.088	.181	.499	.851
0.25	.131	.482	.839	.980	.101	.409	.793	.968	.088	.378	.773	.963
	.160	.513	.854	.983	.138	.452	.812	.970	.121	.421	.783	.962
	.250	.437	.778	.961	.234	.369	.677	.912	.213	.325	.604	.865
0.50	.238	.483	.728	.894	.193	.430	.699	.887	.164	.403	.688	.889
	.317	.566	.786	.923	.316	.565	.796	.930	.287	.548	.790	.933
	.581	.658	.794	.909	.635	.702	.818	.919	.638	.707	.819	.918
0.75	.307	.482	.657	.806	.258	.441	.635	.799	.226	.416	.633	.806
	.412	.592	.748	.865	.442	.628	.785	.894	.430	.626	.796	.908
	.759	.792	.842	.902	.849	.876	.912	.947	.878	.903	.934	.962

uniformly equal but better than a zero dose, and when the sample size is moderate to large. Figures 1 and 2 illustrated this case.

7. SUGGESTIONS FOR FUTURE RESEARCH

At the beginning of the paper, we noted that although the order-restricted literature is substantial, the methods do not seem to be used much in practice. Besides the reasons cited there, the lack of application may reflect the fact that the development of order-restricted methods has, for the most part, been limited to fairly elementary problems. For multinomial data, for instance, the major emphasis has been on comparing two groups or inference about probabilities for a single multinomial distribution (e.g., Chap. 5 in Robertson et al. (1988)). Moreover, most of this work has focused on hypothesis testing, for which there may indeed be little benefit to using order-restricted methods. Table I suggests, for instance, that a test based on a simple model may perform adequately and even have greater power than the order-restricted test unless reality departs quite markedly from the model.

There seems to be considerable scope, on the other hand, for the further development of order-restricted methods for estimating parameters in generalized linear models. For instance, does the good performance of order-restricted estimates when the trend is monotone but irregular or with potential slight departures from monotonicity (shown in Table IV) extend to other generalized linear models? How can one reduce the bias in the ML order-restricted estimates? How can one construct confidence intervals using the order-restricted estimates? (Robertson et al. (1988) noted the difficulty of this problem and devoted only two pages.) How can one incorporate order restrictions in a Bayesian framework? What are the effects of including additional variables as stratification factors? For examples of some recent uses of order-restricted methods in estimation and modeling, see Disch (1981), Schmoyer (1984), Agresti, Chuang, and Kezouh (1987), Morris (1988), McDonald and Diamond (1990), Wax and Gilula (1990), Bacchetti (1991), Gelfand and Kuo (1991), Geyer (1991), Ritov and Gilula (1991), Silvapulle (1994), and Eddy et al. (1995).

Furthermore, order-restricted methods could be extended to additional types of applied problems. For instance, there seems to have been little work on order-restricted methods for repeated measurement data, as opposed to independent samples. A variety of applications may benefit from order-restricted analyses in a generalized linear mixed model format.

ACKNOWLEDGEMENT

We are appreciative of a referee's comments. This research was partially supported by a grant from the National Institutes of Health.

BIBLIOGRAPHY

Agresti, A., and Coull, B. A. (1996). "Order-restricted tests for stratified comparisons of binomial proportions," *Biometrics*, 52, 1103-1111.

Bacchetti, P. (1989). "Additive isotonic models," *Journal of the American Statistical Association*, 84, 289-293.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions*, New York: Wiley.

Bartholomew, D. J. (1959). "A test of homogeneity for ordered alternatives," *Biometrika*, 46, 36-48.

Das Gupta, S., and Perlman, M. D. (1974). "Power of the noncentral F-test: Effect of additional variates on Hotelling's T^2 test," *Journal of the American Statistical Association*, 69, 174-180.

Disch, D. (1981). "Bayesian nonparametric inference for effective doses in a quantal-response experiment," *Biometrics*, 37, 713-722.

Eddy, W. F., Quan, S., and Sampson, A. (1995). "Isotonic probability modeling with multiple covariates," *Computer Science and Statistics: Proceedings of the 27th Symposium on the Interface*, 27, 500-505.

Gelfand, A. E., and Kuo, L. (1991). "Nonparametric Bayesian bioassay including ordered polytomous response," *Biometrika*, 78, 657-666.

Geyer, C. J. (1991). "Constrained maximum likelihood exemplified by isotonic convex logistic regression," *Journal of the American Statistical Association*, 86, 717-724.

Morris, M. D. (1988). "Small-sample confidence limits for parameters under inequality constraints with application to quantal bioassay," *Biometrics*, 44, 1083-1092.

Robertson, T., and Wegman, E. J. (1978). "Likelihood ratio tests for order restrictions in exponential families," *Annals of Statistics*, 6, 485-505.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*, New York: Wiley.

Schmoyer, R. L. (1984). "Sigmoidally constrained maximum likelihood estimation in quantal bioassay," *Journal of the American Statistical Association*, 79, 448-453.

Shi, N.-Z. (1991). "A test of homogeneity of odds ratios against order restrictions," *Journal of the American Statistical Association*, 86, 154-158.

Silvapulle, M. J. (1992). "Robust tests of inequality constraints and one-sided hypotheses in the linear model," *Biometrika*, 79, 621-630.

Silvapulle, M. J. (1994). "On tests against one-sided hypotheses in some generalized linear models," *Biometrics*, 50, 853-858.

Wax, Y., and Gilula, Z. (1990). "A score test for monotonic trend in hazard rates for grouped survival data in stratified populations," *Communications in Statistics, Part A – Theory and Methods*, 19, 3375-3386.

Received December, 1996; Revised July, 1997.