Scaled Link Functions For Heterogeneous Ordinal Response Data *

Minge Xie
National Institute of Statistical Sciences

Douglas G. Simpson
University of Illinois at Urbana-Champaign

Raymond J. Carroll
Texas A&M University

ABSTRACT This paper describes a class ordinal regression models in which the link function has scale parameters that may be estimated along with the regression parameters. One motivation is to provide a plausible model for group level categorical responses. In this case a natural class of scaled link functions is obtained by treating the group level responses as threshold averages of possible correlated latent individual level variables. We find scaled link functions also arise naturally in other circumstances. Our methodology is illustrated through environmental risk assessment data where (correlated) individual level responses and group level responses are mixed.

Key words and phrases: Ordinal regression, Aggregated observations, Marginal modeling approach, Generalized likelihood inference.

1 Introduction

Data in which the individual responses are ordinal are frequently modeled by a generalized linear model of the form

$$\Pr(Y_i \geq s|x_i) = \begin{cases} H(\beta^T x_i - \alpha_s), & \text{if } s = 1, 2, \ldots, S; \\ 1, & \text{if } s = 0; \end{cases} \quad (1)$$

*The research was supported by NSF contract DMS 95-05290, and National Institute of Statistical Sciences Cooperative Agreement CR819638 and CR820897 with the U.S. EPA. R. J. Carroll's research was also supported by a grant from the National Cancer Institute (CA-57030). The authors thank Daniel Guth of the U.S. EPA for making available the data on tetrachloroethylene.

where Y_i is the ordinal response of the ith individual, $i = 1, \ldots, n$, x_i is a vector of explanatory variables, and H is a general cumulative distribution function. The inverse function H^{-1} of H is called a link function (McCullagh and Nelder, 1989). Common choices for the link function H^{-1} are the inverse function of the standard normal distribution Φ^{-1}, the logit function $H^{-1}(t) = \log\{t/(1-t)\}$, and the complimentary log–log function $H^{-1}(t) = \log\{-\log(t)\}$. If the link function has the logit form, model (1) is referred to as a proportional odds ratio model (McCullagh, 1980).

Our concern is with situations in which the assumed constancy of the link function across heterogeneous observations is untenable. For example, in a toxicology risk assessment example, due to various reasons, the responses may be reported at the group (e.g., cage or chamber) level rather than at the individual (animal) level. The number of individuals in each group (cage/chamber size) is known, but the information on individual incidences is missing. With groups of unequal size, the assumption of a common scale is suspect.

We propose a scaled link method in this paper. The method is easy to implement. The scaling of the link function is equivalent to a rescaling of the regression explanatory variables (including S columns of binary variables corresponding to the severity dependent intercepts). If the scale function contains unknown parameters, then the estimation of the regression parameters is easily nested within the maximum likelihood optimization for the scale parameters.

Depending on the scaling formula selected, the observations sometimes may supply only crude information about a scaling parameter. The estimation of the scaling parameter may not be precise. We provide sensitivity studies on the use of scaling parameters. The question we try to answer is how a simpler scaling with a wrong scale parameter would affect the estimation of the other parameters based on the data at hand. This analysis technique is similar to the analysis that Hjort (1994) used to study the t model versus the normal model.

An appealing feature of the scaled link function model for group level responses is that, although the observed elements are marginal responses, the parameters of the model are defined at the individual level and are therefore comparable across groups of different sizes. Moreover, we can combine group and individual level data by treating an individual as a group of size 1. Notice the observed individual level responses within the same group are considered correlated. In addition, some high level correlated structure, for example, correlated batches which are built upon the basic model, might also exist. We use a working likelihood as an estimating criterion, and adopt generalized estimating equation (GEE) techniques (Zeger and Liang, 1986) to deal with the correlation structure.

The rest of the paper is organized as follows. Section 2 provides some derivations of scaled link function models. Section 3 provides sensitivity studies for the scaled link analysis of group level responses. Section 4 briefly

reviews some results on generalized likelihood inference and generalizes our sensitivity analysis to correlated responses. In Section 5, an application of the methodology in toxicological risk assessment is discussed. Section 6 provides some further remarks.

2 Derivation and Interpretation

To model group level ordinal responses and other forms of heterogeneous data we replace (1) by a model of the form

$$\Pr(Y_i \geq s|x_i) = \begin{cases} H\{w_i\ (\beta^T x_i - \alpha_s)\}, & \text{if } s = 1, 2, \ldots, S; \\ 1, & \text{if } s = 0; \end{cases} \tag{2}$$

where the design weights w_i are scaling factors and refer to the precision of latent observations. Effectively, we replace the inverse link function $H(\,\cdot\,)$ by the scaled function $H(\,w_i\,\cdot\,)$. Section 2.1 provides a derivation of model (2) under a heterosdastic setting. Section 2.2 considers critical mass model in detail for group level response data. The critical mass model is a special case of Section 2.1.

2.1 Heteroscedastic latent structure for ordinal data

¿From a standard quantal response analysis, Model (1) has an associated latent structure. For instance, one may assume that there exist a tolerance distribution (latent underlying continuous random variable) Z and a series of ordered thresholds $-\infty = \alpha_0 < \alpha_1 < \ldots < \alpha_S < \alpha_{S+1} = +\infty$, such that observation $y = s$ occurs if and only if the unobserved latent variable Z lies in the interval of $[\alpha_s, \alpha_{s+1})$, that is, $y = \sum_{s=1}^{S} \mathbf{1}_{\{Z \geq \alpha_s\}}$, where $\mathbf{1}_{\{C\}}$ is the indicator function which equals 1 if set C is true and equals 0 if set C is false. Suppose $Z - \beta^T x$ has cumulative distribution $1 - H(-t)$, then we can find that $\Pr(Y \geq s|x)$ has the same form as (1).

We generalize this standard latent structure to a heterosdastic setting. Suppose that a latent variable Z_i follows a heteroscedastic linear model,

$$(Z_i - \beta^T x_i)w_i \,|\, x_i \ \sim\ 1 - H(-t) \tag{3}$$

but we observe only $Y_i = \sum_{s=1}^{S} \mathbf{1}_{\{Z_i \geq \alpha_s\}}$ rather than Z_i. We then have the scaled link function model (2). The weights w_i are interpreted as inverse scale factors for the latent responses.

One special case is the model (5.4) of McCullagh and Nelder (1989, p154), where $(Z_i - \beta^T x_i)/\exp(\tau^T x_i)$ was assumed to follow the standard logistic distribution. In their model, the scale w_i has the form $w_i = \exp(-\tau^T x_i)$, and the linear regression $\tau^T x_i$ plays the role of the linear predictor for the overdispersion or variance. McCullagh and Nelder (1989) suggested this

model for testing the proportional odds ratio assumption against systematically increasing or systemically decreasing in score s.

Another interesting special case occurs under a hierarchical setup in which the latent variable follows a linear model with random coefficients, that is,

$$Z_i|x_i, b_i \sim N(-x_i^T b_i, 1), \quad b_i \sim N(\beta, \Gamma). \tag{4}$$

The marginal distribution of Z_i is then normal with mean $-\beta^T x_i$ and variance $1 + x_i^T \Gamma x_i$. For fully observed Z_i this is a standard growth curve model; see, for instance, Johansen (1984). Such a model might also be useful for observational data. If we observe only the $Y_i = \sum_{s=1}^{S} \mathbf{1}_{\{Z_i \geq -\alpha_s\}}$, then we obtain an ordinal regression model of the form (2) with $w_i = (1 + x_i^T \Gamma x_i)^{-1/2}$. In the binary case, this is a special case of the general latent model considered by Zeger et. al. (1988) and McCulloch (1994), who focused on the latent structure for modeling correlated binary outcomes. Here the observable responses are independent, but heteroscedastic.

In the preceding example the latent distribution for b_i is a way of expressing uncertainty about the linear model. The resulting link function model therefore controls the influence of extreme values of x_i. Note, in particular, that $\|w_i x_i\| \leq \lambda_{\min}^{-1/2}(\Gamma)$, where $\lambda_{\min}$ denotes the smallest eigenvalue. As a result, the maximum likelihood estimator has a bounded influence function in the sense of Hampel et. al. (1986).

2.2 Critical-mass model for ordinal group level responses

Let y_i denote the ordinal response of the ith group, $i = 1, \ldots, n$, and n_i be the group size. Suppose that the group ordinal response $y_i = s$ occurs if and only if the mean latent response of the individuals in the group falls into the threshold interval $[\alpha_s, \alpha_{s+1})$, that is $Y_i = \sum_{s=1}^{S} \mathbf{1}_{\{\bar{Z}_{i\cdot} \geq \alpha_s\}}$ where $\bar{Z}_{i\cdot} = n_i^{-1} \sum_{j=1}^{n_i} Z_{ij}$ and Z_{ij} is the latent response of the jth individual in the ith group.

Assume that the latent variables themselves follow a latent variable model for a given x_i,

$$Z_{ij} = \beta^T x_i + a_i + e_{ij}, \quad a_i \sim N(0, \rho), \quad e_{ij} \sim N(0, 1 - \rho) \tag{5}$$

where we have standardized the marginal variance to be 1 due to identifiable consideration, and random effects a_i and error terms e_{ij} are independent. Then we have the model,

$$P(Y_i > s|x_i) = \Phi\{w_i \ (\beta^T x_i - \alpha_s)\}$$

where

$$w_i = \left\{ \frac{n_i}{1 + (n_i - 1)\rho} \right\}^{\frac{1}{2}} \tag{6}$$

The weight function w_i^2 may be interpreted as an effective group size. If $\rho = 0$ then $w_i^2 = n_i$, whereas a positive correlation shrinks the weight towards one. In the extreme, if $\rho = 1$, then the effective group size is 1.

Identifiability of the parameter ρ requires heterogeneity of the group sizes. If the group sizes were all the same, then the weighting would simply introduce a constant scale factor in the regression parameters. If the group sizes are heterogeneous, then the weights in (6) identify β, α and ρ at the individual level, $n_i = 1$. In this way the scaled model avoids attenuation effects that would otherwise occur in the combined analysis of group responses with heterogeneous group sizes.

The above derivation extends approximately to other link functions. For instance, the inverse logistic link function, $\Psi(t) = \{1 + \exp(-t)\}^{-1}$, is near to the standard normal distribution function in absolute terms, because

$$\sup_t |\Psi(1.702\, t) - \Phi(t)| < 0.01;$$

see, e.g., Baker (1992, p16). Thus, the preceding analysis might be used as a heuristic to suggest a weight function for scaled logistic regression.

3 Wide Model versus Narrow Model Asymptotics

In the group level response model of Section 2.2, we found an effective group size of $n_i^*(\rho) = n_i^{1/2}\{1 + \rho(n_i - 1)\}^{-1/2}$. If $\rho = 1$, then $n_i^* = 1$, and group level responses are indistinguishable from the responses of individuals. From large sample theory, estimation of ρ is not a problem as long as the group size n_i varies across groups and the sample size n is large enough. But in practice, sample sizes are often not large enough to estimate ρ precisely, since the threshold observations apply only crude information about scaling.

In the example discussed in Section 5, the ρ estimate is 1 but its confidence interval $(0.1, 1]$ covers most of the ρ parameter space; see Figure 1 and Section 5 for more detail. A question that naturally arises is whether we can use ordinary regression to replace regression with scaled link functions if ρ is close to 1. Although the parameter estimates would be biased, the variances of the estimates would be decreased and the analysis simplified. This dilemma is the familiar bias-variance tradeoff. In this section, we analyze the bias-variance tradeoff for the simplified model versus the more general model. Using local asymptotics, we determine how far ρ may deviate from 1 before it becomes necessary to include the extra parameter in the model.

Write $\eta = (\alpha_1, \ldots, \alpha_S, \beta)^T$ and let z_i be its associated ith covariates. Following Hjort's (1994) general prescription, let $\mu(\eta, \rho)$ denote any parameter of interest, and let $\hat{\mu}_{wide} = \mu(\hat{\eta}, \hat{\rho})$ denote the estimate with ρ unconstrained (within its natural parameter interval $[0, 1]$), $\hat{\mu}_{narr} = \mu(\hat{\eta}_{narr}, 1)$ denote the

estimate with ρ fixed at 1. We ask when the asymptotic mean square error of $\hat{\mu}_{narr}$ is smaller than or comparable to that of $\hat{\mu}_{wide}$. Of course if $\rho \neq 1$ is fixed regardless of the sample size then the answer is trivial, but such an analysis would shed little light on moderate deviations from ρ, which may not be detected with certainty in moderately sized samples. Therefore we consider sequences of the form $\rho = 1 - \delta n^{-1/2}$, where n is the sample size and δ is fixed and nonnegative. We derive the asymptotic mean square errors of the narrow and wide estimators of μ, and examine the ratio between them.

Assuming independence, the log-likelihood for model (2) is given by

$$l(\eta, \rho | y, z) = \sum_{i=1}^{n} p_i\{w_i(\rho)\eta^T z_i\} \tag{7}$$

where $p_i = p_i\{w_i(\rho)\eta^T z_i\} = f(y_i|\eta, \rho; z_i, n_i)$, the log-likelihood term for the ith observation. We write

$$\bar{\mathcal{U}}_{\eta|\rho=1} = \{n^{-1}\frac{\partial}{\partial \eta} l(\eta, \rho | y, z)\}|_{\rho=1}, \quad \bar{\mathcal{U}}_{\rho|\rho=1} = \{n^{-1}\frac{\partial}{\partial \rho} l(\eta, \rho | y, z)\}|_{\rho=1}$$

and

$$\mathcal{J}_{wide} = \lim_{n\to\infty} var_0\{n^{\frac{1}{2}} \begin{pmatrix} \bar{\mathcal{U}}_{\eta|\rho=1} \\ \bar{\mathcal{U}}_{\rho|\rho=1} \end{pmatrix}\} = \begin{pmatrix} \mathcal{J}_{\eta\eta} & \mathcal{J}_{\eta\rho} \\ \mathcal{J}_{\rho\eta} & \mathcal{J}_{\rho\rho} \end{pmatrix}, \tag{8}$$

where the variance is taken under model (2) with $\rho = 1$, and some regularity conditions are needed to ensure the existence of the limit. Also, we denote $\hat{\eta}_{narr}$ to maximize $l(\eta, 1|y, z)$, the loglikelihood function of the narrow model, and $(\hat{\eta}^T, \hat{\rho})^T$ to maximize $l(\eta, \rho|y, z)$ with $\rho = 1 - \delta n^{-\frac{1}{2}}$, the likelihood function of the wide model. We have the following theorem:

Theorem 1 *Let $\mu(\eta, \rho)$ be smooth with continuous derivatives throughout the inner parameter space $(\eta, \rho) \in (-\infty, \infty) \times (0, 1)$. Assume its left derivative exists at $\rho = 1$. Under model (2) with $\rho = 1 - \delta n^{-\frac{1}{2}}$, then*

$$nE\{\mu(\hat{\eta}_{narr}, 1) - \mu(\eta, 1 - \delta n^{-\frac{1}{2}})\}^2 = b^2\delta^2 + \tau^2 + \mathbf{o}(1) \tag{9}$$

$$nE\{\mu(\hat{\eta}_{wide}, \hat{\rho}) - \mu(\eta, 1 - \delta n^{-\frac{1}{2}})\}^2 = b^2 E\{\mathbf{B}^2 \mathbf{1}_{(\mathbf{B}<\delta)} + \delta^2 \mathbf{1}_{(\mathbf{B}\geq\delta)}\} + \tau^2 + \mathbf{o}(1) \tag{10}$$

where $b = \frac{\partial}{\partial \rho}\mu - (\mathcal{J}_{\eta\eta}^{-1}\mathcal{J}_{\eta\rho})^T \frac{\partial}{\partial \eta}\mu$, $\tau^2 = (\frac{\partial}{\partial \eta}\mu)^T \mathcal{J}_{\eta\eta}^{-1} \frac{\partial}{\partial \eta}\mu$, and $\mathbf{B} \sim N(0, \mathcal{J}^{\rho\rho})$ with $\mathcal{J}^{\rho\rho} = (\mathcal{J}_{\rho\rho} - \mathcal{J}_{\eta\rho}^T \mathcal{J}_{\eta\eta}^{-1} \mathcal{J}_{\eta\rho})^{-1}$.

The basic idea of proof follows Xie (1996, Section 4.5-4.6 and Appendix 2).

Define λ as the ratio of the mean square errors between the narrow model and the wide model,

$$\lambda = \frac{n\mathrm{E}\{\mu(\hat{\eta}_{narr}, 1) - \mu(\eta, 1 - \delta n^{-\frac{1}{2}})\}^2}{n\mathrm{E}\{\mu(\hat{\eta}_{wide}, \hat{\rho}) - \mu(\eta, 1 - \delta n^{-\frac{1}{2}})\}^2}. \tag{11}$$

Based on the data set at hand, we estimate how λ varies as the true ρ or δ deviates away from the value which corresponds to the narrow model.

4 Batch Correlated Observations

Batch correlated observations are common in many settings. For example in the toxicology data set considered in the next section, two levels of batch correlated structures are detected. First, the observed individual level responses within the same group are correlated. Second, these individual level responses as well as group level responses are nested within 24 studies. To handle this type of batch correlation, we use below a marginal modeling approach.

Assume that each observation satisfies marginally model (2), where $w_i = 1$ for individual level observations. We construct a working likelihood function $l(\eta, \rho)$ by multiplying their marginal likelihoods. (we actually include weights in the working log-likelihood function for individual level observations; see Simpson et al., 1996, for detail).

Write $\theta = (\eta, \rho)$. We denote the score covariance matrix and negative expected Hessian matrix by

$$\mathcal{J} = \mathrm{E}\{\mathcal{U}(\theta)\mathcal{U}(\theta)^T\} \text{ and } \mathcal{H} = -\mathrm{E}\{\frac{\partial}{\partial\theta}\mathcal{U}(\theta)\},$$

where $\mathcal{U}(\theta) = (\partial/\partial\theta)l(\theta)$ is the score function. In contrast to standard maximum likelihood theory, $\mathcal{H} \neq \mathcal{J}$ in the present context. Suppose the true parameters are in the interior of the parameter space. The parameter estimates obtained by maximizing $l(\theta)$ or setting $\mathcal{U}(\theta) = 0$ are consistent under minor regularity conditions; see for example, Li (1996). Based on the theories of Generalized Estimating Equations (GEE) (Zeger and Liang, 1986), the variance is approximated by a sandwich formula $\mathcal{H}^{-1}\mathcal{J}\mathcal{H}^{-1}$.

In order to make inference on ρ, we consider the profile likelihood analysis. The profile likelihood function $l(\rho) = l\{\hat{\eta}(\rho), \rho\}$ can be computed easily, since by setting $\tilde{z}_i = w_i(\rho)z_i$ we can $l(\rho)$ from maximizing a standard ordinal regression likelihood $l(\eta, 1|y, \tilde{x})$. Computing $-2l(\rho)$ over a grid of $[0, 1]$ provides a deviance profile as illustrated in Figure 1. If the observations are independent, a standard large sample theory ensures that $2[l(\hat{\eta}, \hat{\rho}) - l\{\hat{\eta}(\rho_0), \rho_0\}]$ has the standard χ^2 distribution where ρ_0 is the true parameter. But if the observations are batch correlated, a generalized profile likelihood theory should be used. We summarize the generalized profile likelihood theory in a following theorem. This theorem is a special case of the nonstandard asymptotic theorem of Simpson et al. (1994). For discussions on the nonstandard asymptotics, see Huber (1967), Kent (1982), Rotnitzky and Jewell (1990), and Li and McCullagh (1994).

Theorem 2 If $\rho = \rho_0$, then as the batch number increases, $2\{l(\hat{\eta}, \hat{\rho}) - l(\hat{\eta}(\rho_0), \rho_0)\}$ converges in distribution to μW, where W is distributed as χ^2_1 with 1 degree of freedom and $\mu = \mathcal{H}^{-1}_{\rho\rho}(\mathcal{H}^{-1}\mathcal{J}\mathcal{H}^{-1})_{\rho\rho}$.

The right corner index $\rho\rho$ indicates the diagonal matrix element associated with ρ.

We also ask whether the sensitivity analysis in last section also carries through the correlated case. Define $\mathcal{J}_{wide}$ the same as (8) and define

$$\mathcal{H}_{wide} = \lim_{n\to\infty} \mathrm{E}_0\{\frac{1}{n}\begin{pmatrix} \frac{\partial^2 \ell(\eta,\rho|y,z)}{\partial\eta\partial\eta^T}|_{\rho=1} & \frac{\partial^2 \ell(\eta,\rho|y,z)}{\partial\eta\partial\rho}|_{\rho=1} \\ \frac{\partial^2 \ell(\eta,\rho|y,z)}{\partial\rho\partial\eta^T}|_{\rho=1} & \frac{\partial^2 \ell(\eta,\rho|y,z)}{\partial^2\rho}|_{\rho=1} \end{pmatrix}\} = \begin{pmatrix} \mathcal{H}_{\eta\eta} & \mathcal{H}_{\eta\rho} \\ \mathcal{H}_{\rho\eta} & \mathcal{H}_{\rho\rho} \end{pmatrix}$$

where the expectation is taken under model (2) with $\rho = 1$, and some regularity conditions are needed to ensure the existence of the limit. Notice the n here denotes the number of independent batches instead of the total number of observations.

Parallel results of Theorem 1 can be obtained, but the right hand sides of (9) and (10) are replaced by

$$\mathrm{E}\{(b\delta + c\mathbf{B})^2\} + \tau^2 + \mathbf{o}(1),$$

$$\mathrm{E}\{(b\mathbf{B} + c\mathbf{B})^2\mathbf{1}_{(\mathbf{B}<\delta)} + (b\delta + c\mathbf{B})^2\mathbf{1}_{(\mathbf{B}\geq\delta)}\} + \tau^2 + \mathbf{o}(1)$$

respectively, where $b = -(\mathcal{H}^{-1}_{\eta\eta}\mathcal{H}_{\eta\rho})^T(\partial/\partial\eta)\mu + \delta\ (\partial/\partial\rho)\mu$, $c = (\mathcal{H}^{-1}_{\eta\eta}\mathcal{H}_{\eta\rho} - D^{-1}_{\eta\eta}D_{\eta\rho})^T(\partial/\partial\eta)\mu$, $\tau^2 = \{(\partial/\partial\eta)\mu\}^T D^{-1}_{\eta\eta}(\partial/\partial\eta)\mu$, and $\mathbf{B} \sim N(0, D^{\rho\rho})$ with $D^{-1} = \mathcal{H}^{-1}_{wide}\mathcal{J}_{wide}\mathcal{H}^{-1}_{wide}$.

5 Example: Analysis of Toxicology Data

The data set analyzed here contains study results from multiple studies on the effects of acute inhalation exposure to tetrachloroethylene. The responses are scored as *No-effect*, *Adverse effect* and *Severe effect*. These outcomes were determined by a toxicologist during a thorough review of the literature. Some studies only provide partial response information and censored responses are reported. We will not explore the issue of censoring in this paper; see Simpson et al. (1996b) for detailed discussions.

The data set has total of 536 observations with 380 at individual levels and 156 at group levels. They are from a total of 24 studies which range in size from 1 group to 35 groups, with median of 4 per study. For group level responses, the group size are ranging from 2 to 30 and the median size is 10. Usually the reason for group responses rather than individual incidences is that the published results are simply descriptions of lesions and tissue damage typical of the group. This leads to data in which the

group size is known, and an overall group severity score can be assigned, but individual incidence information is missing.

In addition to the response, cluster and size information, we have information on the exposure concentration and duration, the species, and the gender. We fit a scaled link function model for logistic regression on $\log_{10}$(concentration) and $\log_{10}$(duration), and allow each species to have their own regression intercepts and concentration slope parameters. Of particular interest is the question of how much information group level responses provide as compared with individual outcomes. The latent correlation parameter, ρ, allows us to investigate this question.

Figure 1 shows the deviance profile function for ρ, along with the projected 95% generalized likelihood confidence interval. The cutoff value was computed following the result of Theorem 2. We see that the estimate of ρ is 1, but the confidence interval is rather wide. A naive confidence interval based on the usual χ^2 approximation is a bit narrower but suspect because of the likely correlation within studies. One conclusion we can draw from Figure 2 is that the within *group* correlation is significantly different from 0.

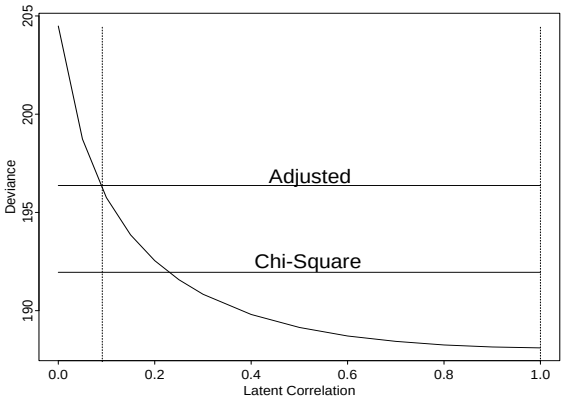

FIGURE 1. Deviance profile of latent correlation parameter for tetrachloroethylene data. Cutoff values are shown for the likelihood ratio (chi–square) and generalized likelihood (adjusted) tests.

Figure 2 performs a sensitivity study on how a narrow model with $\rho = 1$ behaves if the true ρ grids over the range of 0 to 1. The three parameters investigated are the 10% *effective concentration level* (EC_{10}) at unit duration (1 hour) for human, rat and mice species, where the 10% effective concentration level, EC_{10}, is defined as the dose level which would leads to 10% rate of adverse or severe effect at a given duration and for a given species. Figure 2 plots the mean square error (MSE) ratio λ, defined in (11), versus the latent group correlation coefficient ρ for these three parameters. All three MSE ratios increase from below 1 to larger than 1 as ρ decreases

from 1 to 0. The MSE ratio for human EC_{10} almost overlaps the ratio=1 line. It implies that no matter where the true ρ is, the use of a narrow model ($\rho = 1$) would have little effect on the performance of the estimated human one hour exposure EC_{10} in the sense of comparing MSEs. The MSE ratio for rat unit exposure EC_{10} can strike as high as 1.3, and the MSE ratio for mice unit exposure EC_{10} are mostly below 1. If we are willing to accept any MSE ratio less than 1.1, then from Figure 2 we see that ρ larger than 0.43 is sufficient to reduce to the narrow model. Observe that, according to this analysis, group level responses contain roughly the same information as a single individual level response. It implies that individual data should be obtained whenever possible.

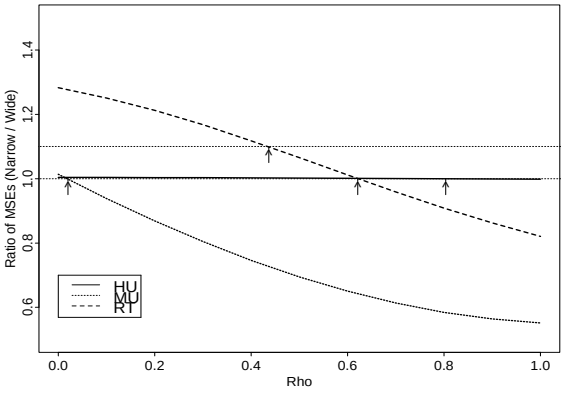

FIGURE 2. Sensitivity analysis on the one hour EC_{10}'s for humans (solid curve), mice (dotted curve), and rats (dashed curve). The x-axis is the assumed true value of ρ grid over [0,1]; the y-axis is the ratio of Mean Square Errors (MSE) as defined in (11). The three arrows on the ratio=1 line indicate the crossing points of the MSE ratios for estimated human, rat and mice EC_{10} parameters from right to left. The fourth arrow indicates the point where the MSE ratio for rat EC_{10} crosses 1.1.

By adopting the assumption that $\rho = 1$, we fit the data set to model (1), and computed the EC_{10} conditional on the duration. The solid lines in Figure 3 are effective concentration levels estimated from the fitting. These lines are called the EC_{10} lines. The dotted lines in Figures 3 (a) to (c) are their confidence bands. The slopes of concentration parameters for humans and rats are parallel, and they are both different from the mouse slope parameter. Notice that the confidence intervals for mice ED_{10} are quite large. Simpson *et al.* (1996a,b) reported the same or similar results. Their robust analysis (1996a) indicted that the problem was the lack of design points of mice at lower exposure level.

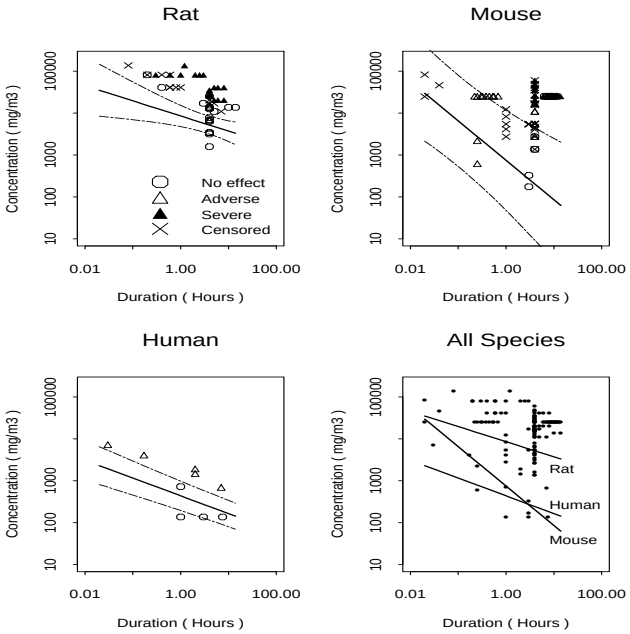

FIGURE 3. Tetrachloroethylene data showing the ordinal response versus concentration and duration for (a) rats only, (b)mice only, (c) humans only, and (d) pooled data. Lines in (a)-(c) are EC10's (solid) and their 95% confidence bounds (dotted). Lines in (d) are EC10's for humans, mice and rats.

6 Discussion

If both group and individual level data are available, the latent model for the group responses provides the basis for combining these sources of information. The key is that the parameters are defined at the individual level rather than at the group level. In particular (6) implies that the individual ordinal responses follow model (2) with $w_i = 1$. Therefore the regression parameters of the group and individual response models have the same interpretation.

In applications, however, some caution is required in combining group level and individual level responses when the individuals are also subject to group correlations. The group level responses might well appear to have a high within group (latent) correlation in part because the group level reporting itself inflates the apparent agreement among elements of the group. Thus, although the latent responses appear to be highly correlated, this need not imply that fully observed individual responses would exhibit the same high correlation. For this reason it would be a good idea to model the within group correlations separately for individual level and group level data.

The scaled models generalize the unscaled models just as weighted least squares regression generalizes ordinary least squares. The simple technique of weighting the regression variables may be useful in a variety of problems. In the group ordinal model, group size and within group correlation are natural effects on which to model the scale of the link function. In other settings it might be more natural to key on covariate information to model heterogeneity of the scale.

7 References

[1] Baker, F.B. (1992). *Item Response Theory.* Marcel Dekkar, New York.

[2] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

[3] Hjort, N.L.(1994). The exact amount of t-ness that the normal model can tolerate , *Journal of the American Statistical Association*, 89, 665-675.

[4] Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the 5th Berkeley Symposium*, 1, 221-233.

[5] Johansen, S. (1984). *Functional Relations, Random Coefficients, and Nonlinear Regression with Application to Kinetic Data.* Lecture Notes in Statistics, 22, Springer-Verlag, New York.

[6] Kent, J.T. (1982). Robust properties of likelihood ratio tests. *Biometrika*, 69, 19-27.

[7] Li, B. and McCullagh, P. (1994). Potential functions and conservative estimating functions. *Annals of Statistics*, 22, 340-356.

[8] Li, B. (1996). A minimax approach to consistency and efficiency for estimating equations. To appear in *Annals of Statistics.*

[9] McCullagh, P. (1980) Regression models for ordinal data , *Journal of the Royal Statistical Society, Series B*, 42, 109-142.

[10] McCullagh, P. and Nelder,J.A. (1989). *Generalized Linear Models* (2nd edition), London: Chapman and Hall.

[11] McCulloch, C.E. (1994). Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association*, 89, 330-335.

[12] Rotnitzky, A. and Jewell, N.P.(1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data, *Biometrika*, 77, 485-497.

[13] Simpson, D.G., Carroll, R.J. and Xie, M. (1994). Scaled link functions and generalized likelihood inferences for heterogeneous binary response data. Technical Report #24, National Institute of Statistical Sciences.

[14] Simpson, D.G., Carroll, R.J., Xie, M. and Guth, D.L. (1996a). Weighted logistic regression and robust analysis of diverse toxicology data. *Communications in Statistics – theorem and Method*, 25(11), 2615-2632..

[15] Simpson, D. G., Carroll, R.J., Zhou, H. and Guth, D.L. (1996b). Interval censoring and marginal analysis in ordinal regression. *Journal of Agricultural, Biological and Environmental Statistics*, 1, 354-376.

[16] Xie, M. (1996). *Regression Modeling: Latent Structure, Theories and Algorithms*. Ph.D. Dissertation, University of Illinois at Urbana-Champaign.

[17] Zeger, S.L., Liang, K.-Y. (1986). Longitudinal Data analysis for discrete and continuous outcomes, Biometrics, 42, 121-130.

[18] Zeger, S.L., Liang, K.-Y. and Albert, P.S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 44, 1049-1060.