Statistica Sinica **7**(1997), 841-854

TRANSFORMATION MODELS FOR INTERVAL SCALE GROUPED DATA WITH APPLICATIONS

Henry Horng-Shing Lu and Fushing Hsieh

National Chiao Tung University and National Taiwan University

Abstract: Interval scale grouped data have peculiar structures of their own rights among various archetypes of polytomous data that deserve special statistical treatments. Maximum likelihood type approaches along with heteroscedastic and transformation models are adapted to take into account this kind of architecture with current state-of-art computation capabilities. Meanwhile, misclassification rates instead of sum of squared residuals are suggested for model fitting and selection in light of the data formation. Successful applications of these methods are demonstrated by a set of empirical data regarding the endotracheal tube size selection for small children in the emergency room of a hospital.

Key words and phrases: Constrained optimization, maximum likelihood estimators, misclassification rates, transformation models.

1. Introduction

There are three prototypes in polytomous (or grouped) data with categorical responses, namely nominal, ordinal and interval scales. They are investigated in depth by McCullagh and Nelder (1989). However, the categorical responses of interval scales are ordered by the clear-cut values of boundary points, that is, the lower and upper bounds. The information contained in the interval scales is more than ordinal scales because the boundaries between adjacent categories of ordinal scales are not ambiguous. Hence, if we assign any artificial numbers or labels to categories and perform analysis based on these artifacts, we sacrifice the genuine information contained in interval scales to that contained in ordinal scales. Furthermore, the beginning and ending categories may be censored (Turnbull (1976)). The regions of the first and the last categories may be unbounded. It is difficult to assign synthetic numbers or labels to left or right censored category. Therefore, we will use the lower and upper bounds of each category to implement analysis directly rather than utilize factitious numbers or labels.

The boundary points of all the categories define the territories. The lower and upper bounds of any category must be used simultaneously in order to specify the regions definitely. One way to evaluate the information contained in each category is to utilize the probability of each interval. The combined information of a full set of data can be described by the accumulate probabilities of all the observed intervals. By maximizing the accumulated probabilities, we can obtain a maximum likelihood estimator (MLE) for the parameters relating response and explanatory variables. Similar ideas can be at least traced back in the literature to Lindley (1950). A recent survey can be found in Heitjan (1989). Because of the constraints of computing power in the past, most researchers have restricted themselves to finding the MLE based on representative values on intervals, such as midpoints or means of intervals, and then doing the first order correction by the analogous Sheppard correction. In the literature of experimental design, the MLE with representative values without taking account of the censoring mechanism is called as the "quick and dirty method" (Hamada and Wu (1991)). It was demonstrated that this may lead to wrong decisions in highly fractionated experiments. However, if it is used with care, it may still reach a quality decision. This is demonstrated in Section 3. Thus, a neutral terminology for this kind of approach can be the *regula falsi* MLE, that is, the MLE with false position. The terminology *regula falsi* is borrowed from numerical methods where it is used to describe secant-like approximation of derivatives in the Newton-Raphson method. It must be used with caution.

As a matter of fact, with the growing power of computation, one can perform the MLE directly without any difficulty. This will be demonstrated in this article. Besides, the idea of MLE can be conveyed in advance with modern computation abilities. We will go further to have MLE cope with heteroscedastic and transformation models. Heteroscedastic and transformation models have been discussed admirably in Carroll and Ruppert (1988). Nevertheless, there are simpler ways to access interval scale grouped data. We will illustrate the methods in the following sections.

Because the pairs of lower and upper bounds are used instead of assigned numbers or labels, it is not suitable to use the sum of squared residuals to measure the goodness-of-fit of model. As the response variables are classified into categories, it is reasonable to use misclassification rates to detect model fitness. This criterion also provides guidelines for model selection and it will be demonstrated by an empirical study.

2. Empirical Data Set

This empirical study investigated the endotracheal tube (ETT) size selection for small children in emergency rooms of hospitals (Wong, Chen, Wu, Chang, Hsieh and Tan (1995)). There are 533 children that went through oral ETT general anesthesia for minor pediatric surgery in the emergency rooms of Chang Gung Memorial Hospital at Taiwan during October 1993 to August 1994. Their ages (to the nearest six months), body weights (in Kilograms), body lengths (i.e. heights, in centimeter), circumferences of the right fifth fingers (in centimeter), head girths (in centimeter) and genders were collected. These variables are recorded because they are reported to be very relevant in the literature (Wong, Chen, Wu, Chang, Hsieh and Tan (1995)). We will use these variables as predictor variables. For simplicity, we will use the number 0 to represent a girl and 1 for a boy.

Polyvinyl chloride ETT's made by the Mallinckrodt Laboratory in U.S.A. were used. The internal diameters (ID's) and external diameters (ED's) of ETT's are listed in Table 1.

Table 1. The ID's and ED's of Mallinckrodt ETT's (unit: mm).

ID	3.5	4.0	4.5	5.0	5.5	6.0	6.5
ED	4.9	5.6	6.2	6.9	7.5	8.2	8.7

Firstly, an ETT was selected by experience so that it can pass through the glottis without any difficulty. Then an anesthesiologist monitored the air leakage when the ventilation circuit was set to a 20-25 cm H_2O inflation pressure. If the air leakage was minimal, this ETT size was the right size and recorded. Otherwise, a bigger size was chosen and tested via the same procedures. The right ETT sizes will be response variables. They are discrete. Furthermore, if a 5.0 mm ID (6.9 mm ED) ETT is the right size for a children, it indicates that the unobserved true trachea size is between 6.9 and 7.5 mm ED ETT's. Therefore, the response variables actually are interval scales instead of ordinal scales.

Traditionally, the ID's of ETT's have been used as standard sizes for selection. However, the ED's are the real boundaries for the unobserved true trachea size. Therefore, we will investigate the selection rules based on the ED's of ETT's.

3. Maximum Likelihood Estimates

Suppose the unobserved response variables Y_i^* are independently and not identically distributed (i.n.i.d.) from a location and scale distribution density $f(x_i\beta, \sigma_i)$ for $i = 1, \ldots, n$, where the location is related to the predictor variables x_i, a 1 by p row vector, and the parameters β, a p by 1 column vector, are what we are interested in. The scale parameter σ_i has a positive value. Thus the unknown parameters are denoted by θ, which includes β and the parameters σ_i. If $\sigma_i = \sigma$, for all i, then this is a homoscedastic model and $\theta = (\beta,\ \sigma)$; otherwise, it is heteroscedastic. These setups are more general than those of generalized linear models in that the expectation values of Y_i^* are not required to exist. Thus, for instance, the family of Cauchy distributions can be included. Meanwhile, the location and scale density can be asymmetric or skewed.

The observed responses are the lower bounds and upper bounds of Y_i^*, Y_i^L and Y_i^U. That is, Y_i^* is only known to fall in the interval bounded by Y_i^L and Y_i^U. When the interval of Y_i^* is left censored, $Y_i^L = -\infty$ and $-\infty < Y_i^* \leq Y_i^U$. If the interval is right censored, $Y_i^U = \infty$ and $Y_i^L < Y_i^* < \infty$. For convenience, we will denote all kinds of intervals as $(Y_i^L,\ Y_i^U]$. Therefore, the information contained in the intervals can be described by cell probabilities, $P_\theta(Y_i^L < Y_i^* \leq Y_i^U)$, which is just the probability of observed intervals. The information for the complete random samples is the accumulation of cell probabilities, $\prod P_\theta(Y_i^L < Y_i^* \leq Y_i^U)$, which is the joint probability of all observed intervals. We will call this term the likelihood of all observed intervals. Hence, the MLE is defined as

$$\hat{\theta} = \text{argmax} \prod_{i=1}^{n} P_\theta(Y_i^L < Y_i^* \leq Y_i^U), \tag{1}$$

where the maximization is taken over the constrained set of arguments. For homoscedastic models, $\theta = (\beta,\ \sigma)$ and the MLE is

$$(\hat{\beta},\ \hat{\sigma}) = \underset{\sigma>0}{\text{argmax}} \prod_{i=1}^{n} \{F[(Y_i^U - x_i\beta)/\sigma] - F[(Y_i^L - x_i\beta)/\sigma]\}, \tag{2}$$

where $F(\cdot)$ is the distribution function of the density $f(\cdot)$. For example, if the underlying distribution is a normal distribution, then the MLE is

$$(\hat{\beta},\ \hat{\sigma}) = \underset{\sigma>0}{\text{argmax}} \prod_{i=1}^{n} \{\Phi[(Y_i^U - x_i\beta)/\sigma] - \Phi[(Y_i^L - x_i\beta)/\sigma]\}, \tag{3}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal density. This optimization can be solved by the EM algorithm (Dempster, Laird, and Rubin (1977)). Due to the constraints in the optimization problem, the approaches in the literature of constrained optimization are possible alternatives (Fletcher (1987)). They are accessible in numerical computation software, such as MATLAB, S-PLUS, ...

In heteroscedastic models, the variance function can be assumed to be $\sigma_i = \varsigma(x_i,\ \beta,\ \eta)$, for some additional nuisance parameters η, and η may contain more than one parameter. For instance, there are some functional forms suggested in literature as follows:

$$\sigma_i = (1 + \mid x_i\beta \mid)^\eta \quad \text{(Box and Hill (1974))}, \tag{4}$$

$$\sigma_i = \phi\exp(\lambda x_i\beta) \text{ and } \eta = (\phi,\ \lambda) \quad \text{(Bickel (1982))}, \tag{5}$$

$$\sigma_i = (1 + (x_i\beta)^2)^{1/2} \quad \text{(Jobson and Fuller (1980))}. \tag{6}$$

The MLE of $\theta = (\beta, \eta)$ for heteroscedastic models becomes:

$$(\hat{\beta},\ \hat{\eta}) = \text{argmax} \prod_{i=1}^{n} \{F[(Y_i^U - x_i\beta)/\sigma_i] - F[(Y_i^L - x_i\beta)/\sigma_i]\}, \tag{7}$$

where the maximization of arguments is taken over the constrained sets of parameters.

The existence of the MLE can be assured by the unimodality of the likelihood (Dharmadhikari and Joag-dev (1988)). If the finite valued Y_i^L and Y_i^U take values in $c^0 = \{c_1^0,\ c_2^0, \ldots, c_{m+1}^0\}$, where m is a positive integer, Pratt (1981) demonstrated the logconcavity in $(\beta,\ c_1^0,\ c_2^0, \ldots,\ c_{m+1}^0)$ for homoscedastic models with $\sigma = 1$. Burridge (1981, 1982) showed that the likelihood function is logconcave in $(\beta/\sigma,\ 1/\sigma)$ for homoscedastic models if and only if the density f is logconcave, which can be proved directly by the results of Prekopa (1973) as well. However, for heteroscedastic models with logconcave densities f, one can not obtain such nice results. It is necessary to check case by case. For non logconcave densities f, it is also necessary to check unimodality. For example, Copas (1975) has proved the unimodality of the location and scale Cauchy family. Silvapulle and Burridge (1986) have given necessary and sufficient conditions for the existence of a solution when there are grouped and ungrouped data at the same time. These conditions can be validated by linear programming methods. The uniqueness of the MLE can be established if the likelihood function is strictly logconcave. If the local maximizer of the likelihood function also satisfies the Kuhn-Tucker conditions, then it is the global maximizer in the constrained parameter space (e.g. see Fletcher (1987)). In practice, this can be programmed easily using the available constrained optimization subroutines in software packages. For instance, we will use MATLAB to perform the computation in the following analysis.

In order to see the major difference between the MLE and *regula falsi* MLE, we will perform a comparison study based on the empirical data. Suppose we use the lower bounds of ED's of ETT's as responses, i.e. $Y_i^* = Y_i^L$; then this MLE for β, namely $\hat{\beta}$, using lower bounds only is a *regula falsi* estimator of the actual MLE. Assuming that the empirical data come from a homoscedastic normal density, $f(x_i\beta,\ \sigma) = N(x_i\beta,\ \sigma)$, then, the *regula falsi* MLE using lower bounds only is equivalent to the ordinary least squared (OLS) estimate using lower bounds in this case. The OLS fits the lower bounds quite well with $R^2 = 82\%$ and the P-value for the F test is .0001, from the analysis of variance (ANOVA) results using the SAS package as shown in Table 2. The estimated coefficients, standard errors, and P-values are reported in Table 3. The t tests in Table 3 indicate that the gender and finger circumferences may be deleted in model selection. For the purpose of comparison, we keep all possible explanatory variables.

Table 2. The ANOVA results of OLS with the lower bounds of ED's.

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	261.83990	43.63998	411.911	0.0001
Error	526	55.72712	0.10595		
C Total	532	317.56702			

Root MSE	0.32549	R-square	0.8245
Dep Mean	6.76266	Adj R-sq	0.8225
C.V.	4.81307		

Table 3. The estimated coefficients, standard errors, and P-values of OLS with the lower bounds of ED's.

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> |T|
INTERCEP	1	2.625777	0.29916343	8.777	0.0001
AGE	1	0.008760	0.00192808	4.543	0.0001
WEIGHT	1	0.029918	0.00903041	3.313	0.0010
HEIGHT	1	0.018435	0.00302502	6.094	0.0001
FINGER	1	0.121069	0.05939629	2.038	0.0420
HEAD	1	0.029067	0.00778688	3.733	0.0002
GENDER	1	-0.035581	0.03866031	-0.920	0.3578

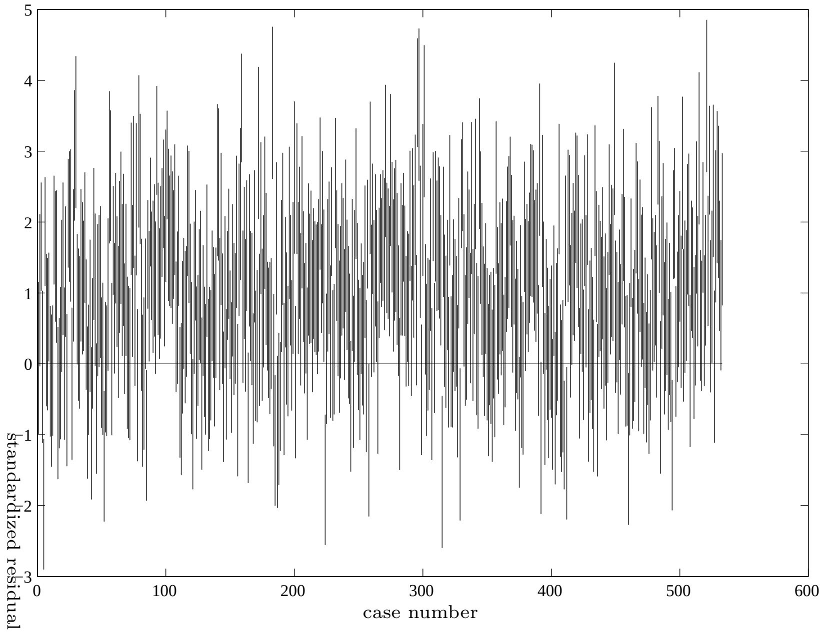

Figure 1. The standardized residual pair plot against case numbers of OLS fitting using lower bounds

The residual plot of $Y_i^L - x_i\hat{\beta}$ against case numbers looks fine. All the routine statistical tests and diagnoses show that the fitting status is quite good. However, this OLS fitting using lower bounds only is a serious under-fitting as the responses get close to the upper bounds Y_i^U. Therefore, the R^2 statistics does not reflect this situation and is not a good measure for goodness-of-fit for interval scale grouped data. The residual plot using the lower bounds only does not reveal the true fitting status for interval scale data.

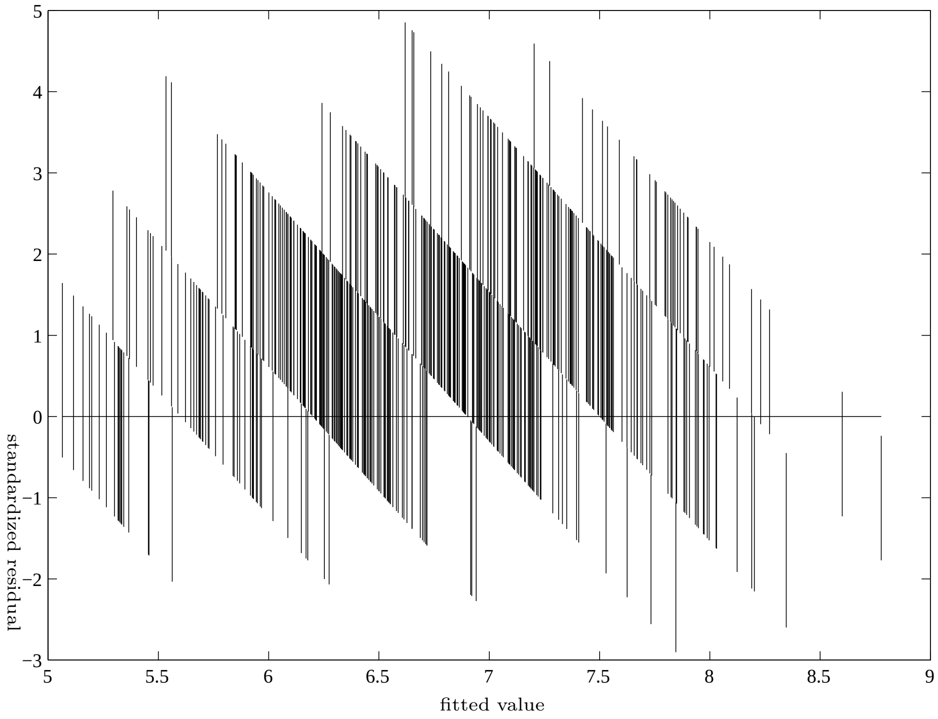

Figure 2. The standardized residual pair plot against fitted values of OLS fitting using lower bounds

In order to take into account of both residuals of lower and upper bounds, the pairs of corresponding residuals for lower and upper bounds are displayed and connected as in Figure 1. The residuals are divided by the estimated standard error, $\hat{\sigma}$, to obtain the standardized residuals in Figure 1. There are 533 line segments in total. The line segments have open endings at the bottom and close endings at the top. The lengths of line segments depend on the distance between contiguous categories. If the line segment intersects the zero line, it means that the fitted value $x_i\hat{\beta}$ is bigger than the lower bound, Y_i^L, and less than or equal to the upper bound, Y_i^U. The selection of ETT for this patient will be correct based on the OLS fitting. If the line segment is above the zero line, then $Y_i^L \geq x_i\hat{\beta}$ and the selection of ETT is underestimated. Similarly, if the line segment is below the zero line, then $Y_i^U < x_i\hat{\beta}$ and the selection of ETT is overestimated. Hence, there are 258 cases (48%) correctly classified, 263 cases (49%) under-classified, and 12 cases over-classified (2%). Among 263 under-classified cases, 250 cases (47%) are underestimated by 1 category and 13 cases (2%) are underestimated by 2 categories. All 12 over-classified cases are overestimated by 1 category. The total misclassification rate, including under- and over- classification cases, is $275/533 = 52\%$! This example exhibits that the misclassification rate is indeed a convincing measure of goodness-of-fit in interval scaled polytomous data. In order to find out the misclassified status in every group, we can plot the standardized residual pairs against fitted values in Figure 2. In each group, the slop of residual pairs is -1 because the residuals, $Y_i^L - x_i\hat{\beta}$ or $Y_i^U - x_i\hat{\beta}$, plus fitted values, $x_i\hat{\beta}$, are the same within each group. The residual pair plot in Figure 2 is more informative than Figure 1 in exposing the fitting status.

A simple way to correct the above approach is to use the middle value, $(Y_i^L + Y_i^U)/2$, as the response after we recognize the censoring nature of the data formation. Fortunately, the upper bound of this observed data is bounded and this approach works well for this data set. The ANOVA results are similar to those in Table 2. The R^2 increases from 82.45% in Table 2 to 82.60% only. The routine statistical tests and diagnoses are similar to those of OLS fitting with lower bounds only with slightly higher significance. However, the misclassified cases are reduced significantly to 168 cases (32%)! There are 85 cases (16%) underclassified by 1 category and 83 cases (16%) overclassified by 1 category. The phenomenon of underclassification by 2 categories disappears. The residual pair plot is more centralized around the central line than that in Figure 2. The middle values are much better than the lower (or upper) bounds in representing the intervals. The resulting regression fits and predicts the middle points of the intervals, not the lower bounds of the intervals. Hence, the *regula falsi* MLE with middle points can lead to suitable fitting and prediction. A more refined approach for the *regula falsi* MLE can use the imputation values as in Hamada and Wu (1991). Hence, the *regula falsi* MLE must be used carefully.

Alternatively, the MLE can be applied to configure the *bona fide* information contained in the pairs of lower and upper bounds. Fitting MLE with homoscedastic models in the empirical data set, one can obtain the standardized residual pair plot in Figure 3. The residual pairs are central around the zero line. More precisely, there are 84 cases (16%) under-classified by 1 category and 86 cases (16%) over-classified by 1 category. Therefore, the total misclassification rate has been reduced from 52% to 32%! Furthermore, the serious underestimation phenomenon in OLS with lower bounds only has been conformed into an even distribution. The cases that have been underestimated by 2 categories are corrected at the same time. To see the details of the situation of misclassification in each group, one can check Figure 3. In practice, this can provide finer adjustments in each group for different fitted values. Furthermore, one can actually pull out those misclassified cases and carry out the diagnosis. Therefore, the *regula falsi* MLE must be used with caution because it may lead to wrong decisions if one does not use proper representation values for censored data. The MLE will always account for the censoring mechanism.

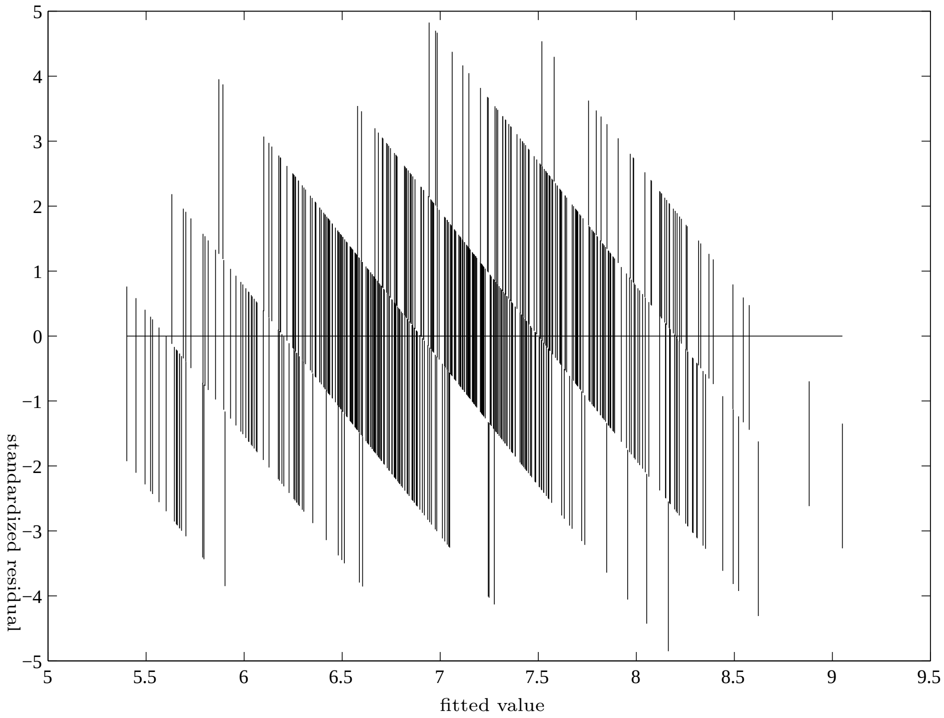

Figure 3. The standardized residual pair plot against fitted values of MLE fitting

4. Transformation Models

Suppose the original responses Y_i^* are not i.n.i.d. from a location and scale distribution density. One can search for a transformation so that the transformed responses come from a location and scale density. That is, we suppose that there exists an unknown increasing function, $H(\cdot)$, such that $Y_i = H(Y_i^*)$ are i.n.i.d. from $g(x_i\beta, \sigma_i)$ for $i = 1, \ldots, n$. Thus, the possible class of distribution functions for the responses can be larger than the family of location and scale distributions.

For instance, this unknown transformation can be a shifted power transformation (Box and Cox (1964)), then the whole problem setup is parametric. If this unknown transformation is only assumed to be increasing, then the setup is semiparametric. This kind of idea has been applied in practice, such as the signal-detection theory in psychology (Dorfman and Alf (1968, 1969)).

MLE can be applied to transformation models in the same fashion. Suppose that the finite valued Y_i^L and Y_i^U belong to the set $c^0 = \{c_1^0,\ c_2^0, \ldots, c_{m+1}^0\}$; then the transformed values are $c_i = H(c_i^0)$. Without loss of generality, we assume $c_1^0 < c_2^0 < \cdots < c_{m+1}^0$ and $c_1 < c_2 < \cdots < c_{m+1}$. We also set $H(-\infty) = -\infty$ and $H(\infty) = \infty$. For a power transformation with shift, one needs to estimate the power and shift parameters at the same time. For an arbitrary increasing transformation, one also needs to estimate the nuisance parameters, $\{c_1,\ c_2, \ldots,\ c_{m+1}\}$, at the same time. Since the number of nuisance parameters, $m+1$, is fixed and does not increase with sample size, n, it will not cause any trouble in estimation. For instance, the MLE with a semiparametric transformation model (MLE/STM) is

$$\hat{\theta} = \text{argmax} \prod_{i=1}^{n} P_\theta(C_i^L < Y_i \leq C_i^U), \tag{8}$$

where $C_i^L = H(Y_i^L)$ and $C_i^U = H(Y_i^U)$, which take values in the set of $\{c_1, c_2, \ldots, c_{m+1}\}$. Take a homoscedastic MLE/STM, $\sigma_i = \sigma$, for all i, as an example. Then $\theta = (\beta, \sigma, c_1, c_2, \ldots, c_{m+1})$ and we have

$$\begin{aligned}
&(\hat{\beta},\ \hat{\sigma},\ \hat{c}_1,\ \hat{c}_2,\ \ldots,\ \hat{c}_{m+1}) \\
&= \underset{\sigma>0,\ c_1<c_2<\cdots<c_{m+1}}{\text{argmax}} \prod_{i=1}^{n} P_\theta(C_i^L < Y_i \leq C_i^U) \\
&= \underset{\sigma>0,\ c_1<c_2<\cdots<c_{m+1}}{\text{argmax}} \prod_{i=1}^{n} \{G[(C_i^U - x_i\beta)/\sigma] - G[(C_i^L - x_i\beta)/\sigma]\},
\end{aligned} \tag{9}$$

where $G(\cdot)$ is the distribution function of the density $g(\cdot)$. The constrained optimization over $c_1 < c_2 < \cdots < c_{m+1}$ can be performed via reparametrization by $\theta = (\beta,\ \sigma,\ c_1,\ d_1,\ d_2, \ldots,\ d_m)$, where $d_i = c_{i+1} - c_i$, for $i = 1, \ldots,\ m$, are the increments of boundary points. Then the estimator can be obtained by solving the constrained optimization problem,

$$\begin{aligned}
&(\hat{\beta},\ \hat{\sigma},\ \hat{c}_1,\ \hat{d}_1,\ \hat{d}_2, \ldots, \hat{d}_m) \\
&= \underset{\sigma>0,\ d_1>0,\ d_2>0,\ d_m>0}{\text{argmax}} \prod_{i=1}^{n} \{G[(C_i^U - x_i\beta)/\sigma] - G[(C_i^L - x_i\beta)/\sigma]\}.
\end{aligned} \tag{10}$$

Similarly, one can consider heteroscedastic MLE/STM. MLE with power transformation models (MLE/PTM) can be utilized as well. Other types of transformation models can be considered equally. We will call this kind of MLE with transformation models MLE/TM. The existence and uniqueness of MLE/TM follows analogous arguments to those of MLE.

Applying MLE/STM to the empirical data, we will assume that the transformed responses, Y_i, follow a logistic density with location parameters $x_i\beta$ and scale parameters $\sigma_i = 1$, for all i. The choice of logistic distribution is just to demonstrate that this popular distribution can be combined into MLE/STM with ease. Other choice of distributions can be associated with MLE/STM without any difficulty. The standardized residual pair plot is shown in Figure 4. In particular, the residual pairs are jacked up mostly in the first and last categories, which are results of the transformation of scales. Consequently, the misclassification rates is reduced to 159 cases (30%)! There are 82 cases (15%) underestimated by 1 category and 77 cases (14%) overestimated by 1 category.

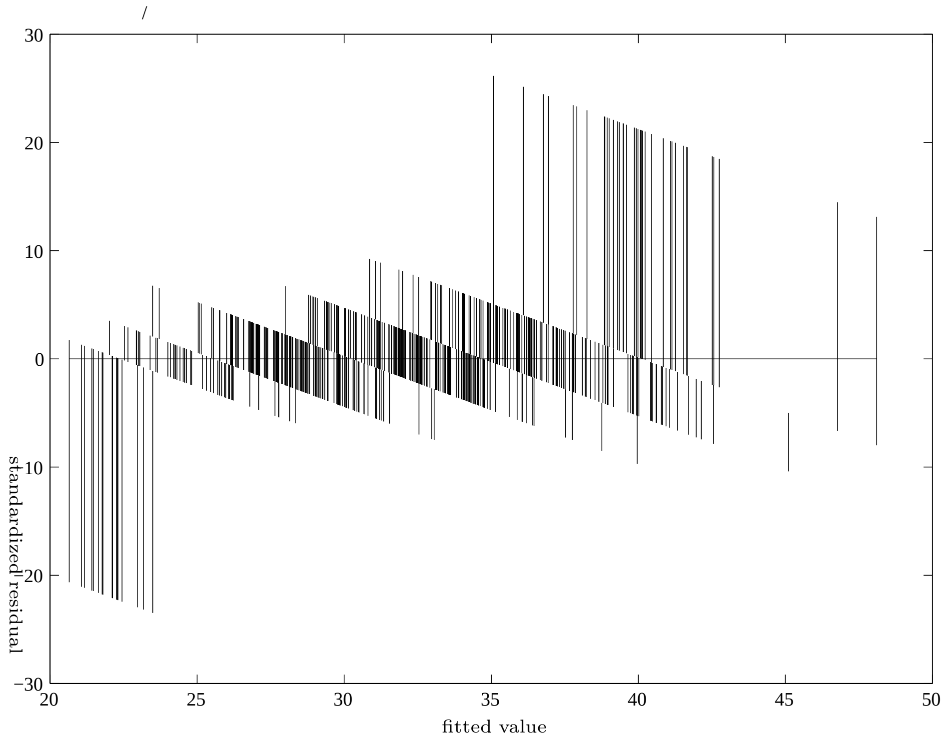

Figure 4. The standardized residual pair plot against fitted values of MLE/STM fitting

5. Theoretical Properties

We will demonstrate the asymptotic results for MLE/STM in homoscedastic models, equation (10), for illustration. Similar results can be obtained for other cases in the same way. Since the boundary values of response categories are nuisance parameters, we can assume that the boundary values of adjacent categories are connected without loss of generality. Suppose that there are a fixed number of response categories and the finite valued $C_i^L = H(Y_i^L)$ and $C_i^U = H(Y_i^U)$ belong to the set $c = \{c_1,\ c_2, \ldots, c_{m+1}\}$. Now $\theta = (\beta,\ \sigma,\ c_1,\ d_1,\ d_2, \ldots, d_m)$, where $d_j = c_{j+1} - c_j$, for $j = 1, \ldots,\ m$. One can obtain the asymptotic consistency and normality of MLE/STM, $\hat{\theta}$.

Theorem 1. *Suppose that the true parameter, θ_0, is in the constrained parameter space, Θ.*

(1) *Under the conditions* C1–C5 *in Hoadley* (1971), $\hat{\theta} \ \rightarrow_P \ \theta_0$.

(2) *Consider the conditions* N1–N9 *in Hoadley* (1971). *In particular, the condition* N7 *assumes that the average Fisher information matrix converges to a positive definite information matrix, $\overline{I}(\theta)$, for all $\theta \in \Theta$. Then $n^{1/2}(\hat{\theta} - \theta_0) \ \rightarrow_D N(0,\ \overline{I}^{-1}(\theta_0))$.*

Proof. Define a new random variable Z_i from $Y_i = H(Y_i^*)$ as follows,

$$Z_i = \begin{cases} j, & \text{if } c_j < Y_i \le c_{j+1},\ j = 1, \ldots, m, \\ 0, & \text{otherwise,} \end{cases}$$

for $i = 1, \ldots, n$. Then the probability density function of Z_i is a multinomial distribution with cell probabilities, $\{G[(c_{j+1} - x_i\beta)/\sigma] - G[(c_j - x_i\beta)/\sigma]\}$, for all $j = 1, \ldots, m$. The random variables, Z_i, for all $i = 1, \ldots, n$, are i.n.i.d. Since the likelihoods of Y_i are the same as those of Z_i, one can prove this theorem based on Theorems 1 and 2 in Hoadley (1971).

Mixed types of interval scale grouped and ungrouped data can be handled by the product of cell probabilities and probability densities (or masses); the idea of maximum likelihood can go through without any difficulty (see Burridge (1981) for instance). Actually, the asymptotic results in Theorem 1 also hold for the maximum likelihood type estimator in this case using the results in Hoadley (1971) after combining the Borel measure and counting measure into a new dominated measure.

6. Discussion

Throughout this article, we distinguish the critical differences between interval scaled and other scaled polytomous data. The idea of MLE can be enhanced in interval scaled polytomous data with current computation ability. Equipped with heteroscedastic and transformation models, MLE can be more powerful. The improvements of MLE over *regula falsi* MLE (or OLS) can be detected by misclassification rates, including under- and over- classification rates. Together with the standardized residual pair plots, one can explore the real status of model fitting in each category. These are demonstrated in the empirical studies in this article. Further analyses of this empirical data are studied in Lu, Hsieh, Chang and Wong (1996).

As for the computation aspects, the other technique available is the EM algorithm in Dempster, Laird, and Rubin (1977). Because of the constraint nature of MLE, especially the increasing orderings of the transformation in MLE/STM, we considered constrained optimization methods. The methods considered here can be generalized to multivariate responses as in Burridge (1982). Combination of interval scale grouped and ungrouped data can be managed in the same fashion. Utilizing the likelihoods of observed intervals, we can go ahead with the MLE type approach with better model fitting and selection for interval scale grouped data.

Acknowledgements

The authors would like to thank Doctors T. K.-m. Wong and T.-C. Chang for providing the empirical data set, Professors A. Chao and J. S. Huang for their helpful discussions. We also thank the editor, associate editor and referees for their valuable comments and suggestions, especially in pointing out the related literature of Hamada and Wu (1991). This research was supported by the National Science Council of Taiwan.

References

Bickel, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10**, 647-671.

Box, G. E. P. and Hill, W. J. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics* **16**, 385-389.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *J. Roy. Statist. Soc. Ser. B* **26**, 211-252.

Burridge, J. (1981). A note on maximum likelihood estimation for regression models using grouped data. *J. Roy. Statist. Soc. Ser. B* **43**, 41-45.

Burridge, J. (1982). Some unimodality properties of likelihoods derived from grouped data. *Biometrika* **69**, 145-151.

Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression.* Chapman and Hall, London.

Copas, J. B. (1975). On the unimodality of the likelihood for the Cauchy distribution. *Biometrika* **62**, 701-704.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.

Dharmadhikari, S. W. and Joag-dev, K. (1988). *Unimodality, Convexity, and Applications.* Academic Press, Boston.

Dorfman, D. D. and Alf, E., Jr. (1968). Maximum-likelihood estimation of parameters of signal-detection theory–a direct solution. *Psychometrika* **33**, 117-124.

Dorfman, D. D. and Alf, E., Jr. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals-rating-method data. *J. Math. Psych.* **6**, 487-496.

Fletcher, R. (1987). *Practical Methods of Optimization.* Second edition. Wiley, New York.

Hamada, M. and Wu, C. F. J. (1991). Analysis of censored data from highly fractionated experiments. *Technometrics* **33**, 25-38.

Heitjan, D. F. (1989). Inference from grouped continuous data: a review. *Statist. Science* **4**, 164-183.

Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distribution case. *Ann. Math. Statist.* **42**, 1977-1991.

Jobson, J. D. and Fuller, W. A. (1980). Least squared estimation when covariance matrix and parametric vectors are functionally related. *J. Amer. Statist. Assoc.* **75**, 176-181.

Lindley, D. V. (1950). Grouping corrections and maximum likelihood equations. *Proc. Cambridge Philos. Soc.* **46**, 106-110.

Lu, H. H.-S., Hsieh, F., Chang, T.-C. and Wong, T. K.-m. (1996). Model selection in interval scale grouped data: a case study. Technical report. National Chiao Tung University, Taiwan.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* Second edition. Chapman and Hall, London.

Prekopa, A. (1973). On logarithmic concave measures and functions. *Acta. Sci. Math.* **34**, 335-343.

Pratt, J. W. (1981). Concavity of the log likelihood. *J. Amer. Statist. Assoc.* **76**, 103-106.

Silvapulle, M. J. and Burridge, J. (1986). Existence of maximum likelihood estimates in regression models for grouped and ungrouped data. *J. Roy. Statist. Soc. Ser. B* **48**, 100-106.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser. B* **38**, 290-295.

Wong, T. K.-m., Chen, C., Wu, R. S. C., Chang, T.-C., Hsieh, F. and Tan, P. P. C. (1995). Guidelines for endotracheal tube size selection in small Chinese children: a prospective study of 533 cases. Technical report. Chang Gung Memorial Hospital, Taiwan.

Institute of Statistics, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu 30050, Taiwan.
E-mail: hslu@stat.nctu.edu.tw

Department of Mathematics, National Taiwan University, Taipei, Taiwan.
E-mail: fushing@math.ntu.edu.tw

(Received February 1996; accepted February 1997)