Journal of the Korean Statistical Society (2000), **29**: *2, pp 201–217*

Regression Analysis of Longitudinal Data Based on M-estimates

Sin-Ho Jung[1] and Terry M. Therneau[2]

ABSTRACT

The method of generalized estimating equations (GEE) has become very popular for the analysis of longitudinal data. We extend this work to the use of M-estimators; the resultant regression estimates are robust to heavy tailed errors and to outliers. The proposed method does not require correct specification of the dependence structure between observations, and allows for heterogeneity of the error. However, an estimate of the dependence structure may be incorporated, and if it is correct this guarantees a higher efficiency for the regression estimators. A goodness-of-fit test for checking the adequacy of the assumed M-estimation regression model is also provided. Simulation studies are conducted to show the finite-sample performance of the new methods. The proposed methods are applied to a real-life data set.

Key words: Generalized estimating equation; Goodness-of-fit test; Heteroscedasticity; Median regression model.

1. Introduction

M-estimators are so named because they are based on an estimating equation that is similar to that for a maximum likelihood estimator (MLE), and were proposed to estimate a 'center' of a distribution with heavy tails. Huber (1973) generalized the M-estimation method from a single population problem to regression problems with iid error terms. The method also has been so successful in the robust estimation of multivariate location parameters, see e.g. Rousseeuw and Leroy (1987). Applications of M-estimates to survival data have been proposed by Zhou (1992) and Lai and Ying (1994).

Repeated measurements data arise in applications where multiple response values are recorded for each subject, examples include growth curve data, pain relief over a long time window, and operating characteristics of mechanical devices.

[1]Division of Statistics, Indiana University School of Medicine, Indianapolis, IN 46202-2859, U.S.A.

[2]Section of Biostatistics, Mayo Clinic, Rochester, MN 55905, U.S.A.

A major issue with their analysis has been the correlation between observations for a subject, particularly if some measurements are missing. Liang and Zeger (1986) proposed a robust estimation method for regression parameters by specifying only the relationship between the marginal mean of the response variable and the covariates. Their method gives a consistent estimate of the regression coefficient vector and its asymptotic variance without specification of the correct correlation structure within a subject. Their methods are robust against the misspecification of dependence structure but not against outliers since the model is based on the mean of response variable.

Jung (1996) applied a median regression model to non-iid repeated measures data to gain robustness against outliers. In this paper, we apply M-estimation based estimating equations to regression analysis of longitudinal data; both the mean regression model of Liang and Zeger and the median regression model of Jung are special cases.

The M-estimation regression models are robust against heavy-tailed error distributions and outliers. Furthermore, the proposed method provides a consistent regression estimate of $\hat{\beta}$ and its asymptotic covariance matrix for any 'working' structure of dependence between the observations. If the working structures are correctly specified, the resulting regression estimator has the minimum variance within a family of consistent regression estimators.

When an M-regression model is misspecified, its estimator converges to different constants depending on the choice of working dependence and heterogeneity structures. This motivates a goodness-of-fit test of the assumed regression model. Extensive simulation results show that the new methods are very efficient in the presence of outliers. The proposed methods are illustrated with real-life data.

2. Estimating Equations for M-estimates

2.1 *Regression Models Based on M-estimation*

For subject i at measurement time j ($i = 1, ..., n; j = 1, ..., K_i$), let Y_{ij} be a continuous response variable and $Z_{ij} = (Z_{1ij}, ..., Z_{pij})^T$ be the corresponding $p \times 1$ covariate vector. Repeated observations over time on each subject are usually correlated with one another. The measurement times may vary subject by subject due to missed measurements, loss to follow-up or other causes. We assume that the measurement times are independent of previous observations in the sense of Rubin (1976). For simplicity of notation, we will omit the subscript i in K_i.

We consider the model

$$Y_{ij} = \beta_0^T Z_{ij} + \epsilon_{ij}, \tag{1}$$

where β_0 is a $p \times 1$ vector of unknown regression coefficients and $E\{\psi(\epsilon_{ij})\} = 0$ for a chosen M-score function ψ. The mean or L_2 regression model (e.g., Liang and Zeger, 1986) chooses $\psi(y) = y$ and the median or L_1 regression model (e.g., Jung, 1996) chooses $\psi(y) = 2I(y > 0) - 1$. In this paper, we will assume that ψ is monotone and absolutely continuous, in order to guarantee uniqueness and consistency of the M-estimator (see Freedman and Diaconis, 1982), and to have a continuous estimating equation. A continuous ψ-score is recommended to protect against the effect of round-off. The choice of a discontinuous score is briefly discussed in Section 6. For a limited influence of outliers, one also requires the ψ-score be finite. If the marginal error distributions are heteroscedastic and irregularly skewed, it may be difficult to find a proper score function and to interpret the regression coefficients. However if they are symmetric, an anti-symmetric score function will provide a high efficiency in the presence of outliers. When marginal error distributions are identical, the estimation of non-intercept parameters is robust to the choice of a score function.

2.2 *Working Iid Equation*

Under an iid working assumption on $(\epsilon_{ij}, i = 1, ..., n, j = 1, ..., K)$, the M-estimate $\hat{\beta} = \hat{\beta}_\psi$ of β_0 is obtained by solving the estimating equation

$$U_I(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{K} \psi(Y_{ij} - \beta^T Z_{ij}) Z_{ij} = 0. \tag{2}$$

If the errors ϵ_{ij} are iid with common density f, then the MLE corresponds to use of the function $\psi(x) = -f'(x)/f(x)$.

Let $Z_i = (Z_{i1}, \ldots, Z_{iK})^T$ and $\psi_i(\beta) = \{\psi(Y_{i1} - \beta^T Z_{i1}), \ldots, \psi(Y_{iK} - \beta^T Z_{iK})\}^T$. Then (2) can be written as

$$U_I(\beta) = \sum_{i=1}^{n} Z_i^T \psi_i(\beta) = 0.$$

The estimating equations can be easily solved by using the Newton-Raphson method,

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + A^{-1}(\hat{\beta}^{(k)}) U_I(\hat{\beta}^{(k)}),$$

where $\hat{\beta}^{(k)}$ is the estimate of β obtained at the k-th iteration, $A(\beta) = \sum_{i=1}^{n} Z_i^T D_i(\beta) Z_i$ and $D_i(\beta) = \text{diag}\{\psi'(Y_{i1} - \beta^T Z_{i1}), ..., \psi'(Y_{iK} - \beta^T Z_{iK})\}$. The recursion converges very quickly using an appropriate starting value. However, if the window of the score function ψ, i.e., $\{x : \psi'(x) \neq 0\}$, is very narrow, then we need a starting value close enough to the solution for $A(\beta)$ to be nonsingular at the starting value. When we use a regression model with a narrow score function, we may start from a regression model with a wide score function, and shrink the score function little by little using the estimate from a wider score-function, such as the indentity function for L_2 regression, as a starting value for a narrower one.

If $E\{\psi'(\epsilon_{ij})\} > 0$, then under mild conditions $\hat{\beta}$ is a consistent estimator of β_0 and $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically normal with mean 0 and covariance matrix

$$V = \lim_{n\to\infty} nA^{-1}(\beta_0)\{\sum_{i=1}^{n} Z_i^T \text{cov}(\psi_i) Z_i\}\{A^{-1}(\beta_0)\}^T,$$

where $\psi_i = \psi_i(\beta_0)$. The proof is straightforward and is omitted.

The covariance matrix V can be consistently estimated by

$$\widehat{V} = nA^{-1}(\hat{\beta})\{\sum_{i=1}^{n} Z_i^T \psi_i(\hat{\beta})\psi_i^T(\hat{\beta}) Z_i\}\{A^{-1}(\hat{\beta})\}^T.$$

Note that the estimation of $\hat{\beta}$ and $\widehat{V}$ does not require the distribution or the dependence structure of the error terms to be specified. Only the independence of ϵ_{ij} and $\epsilon_{i'j'}$ for $i \neq i'$ is needed, i.e., independence of the subjects. Once the regression model (1) is correctly specified, $\hat{\beta}$ and $\widehat{V}$ are consistent. However, if the observations within each subject are highly correlated we may lose some efficiency by estimating $\hat{\beta}$ via equation (2), which is motivated by an iid working assumption. In the next section, we discuss estimating equations leading to higher efficiency by incorporating working dependence and heterogeneity structure in constructing estimating equations.

2.3 *Efficient Estimating Equations*

Let Δ_i be a working dispersion matrix such that $\phi\Delta_i$ approximates

$$E\{D_i(\beta_0)\} = \text{diag}\{E\psi'(\epsilon_{i1}), \ldots, E\psi'(\epsilon_{iK})\}$$

for some constant ϕ. Also, for a positive constant σ^2, choose $\sigma^2 V_i(\beta)$ as a $K \times K$ working covariance matrix of ψ_i. Depending on the choice of ψ-score, heteroscedasticity of the error terms may be reflected in Δ_i as well as in V_i. For example, when $\psi(x) = \min(|x|, c)\text{sign}(x)$, then $E\{\psi'(\epsilon_{ij})\} = P(|\epsilon_{ij}| < c)$ decreases as the dispersion of ϵ_{ij} increases.

A generalized (weighted) estimating equation is defined as

$$U_w(\beta) = \sum_{i=1}^{n} Z_i^T W_i \psi_i(\beta) = 0, \tag{3}$$

where $W_i = \Delta_i V_i^{-1}$. The working iid equation of section 2.2 corresponds to $\Delta_i = V_i = I_K$, the identity matrix. Note also that, in the mean regression case, $\psi(x) = x$ and $\psi'(x) = 1$, for which estimating equation (3) reduces to the GEE of Liang and Zeger (1986). The equation can be recursively solved using Newton-Raphson method as in Section 2.2. The following result summarizes the asymptotic properties of the solution $\hat{\beta}_w$ of equation (3): if $E\{\psi'(\epsilon_{ij})\} > 0$, then as $n \to \infty$, $n^{1/2}(\hat{\beta}_w - \beta_0)$ is asymptotically normal with mean 0 and covariance matrix

$$V_w = \lim_{n\to\infty} n A_w^{-1}(\beta_0)\{\sum_{i=1}^{n} Z_i^T W_i \mathrm{cov}(\psi_i) W_i^T Z_i\}\{A_w^{-1}(\beta_0)\}^T$$

where $A_w(\beta) = \sum_{i=1}^n Z_i^T W_i D_i(\beta) Z_i$. The proof is similar to that of Theorem 2 of Liang and Zeger (1986) and is omitted.

A consistent estimate $\widehat{V}_w$ of V_w can be obtained by replacing $\mathrm{cov}(\psi_i)$ and β_0 by $\psi_i\psi_i^T$ and $\hat{\beta}_w$, respectively, in the expression of V_w. Since the estimating equations are invariant to the multiplication of a constant, σ^2 and ϕ are not involved in estimating $\hat{\beta}$ and $\mathrm{var}(\hat{\beta})$.

Δ_i and V_i may include other parameters α and γ which will be estimated from separate estimating equations. If their estimates are $n^{1/2}$-consistent, we can replace α and γ by their estimates without any modification in the estimating equation and the asymptotic results.

As in the iid case, the consistency of $\hat{\beta}_w$ and $\widehat{V}_w$ requires only the correct specification of the regression model (1), but not that of Δ_i and V_i. If Δ_i and V_i are correctly specified, then $n^{1/2}(\hat{\beta}_w - \beta_0)$ will have the asymptotic covariance matrix

$$V_o = \lim_{n\to\infty} n(\sum_{i=1}^{n} Z_i^T \Delta_i V_i^{-1} \Delta_i Z_i)^{-1}.$$

It is easy to show that, for a chosen ψ, this is an optimal estimator in the sense that $V_w - V_o$ is a non-negative definite matrix for any choice of working matrices. Obviously, a working matrix closer to the true one will improve the efficiency of the regression estimator.

Here are some choices of Δ_i and V_i:

Example 1. (iid case) The iid working assumption uses $\Delta_i = I$ and $V_i = I$.

Example 2. (modeling heteroscedasticity) When there exists an obvious trend in the dispersion of the errors, we may consider modeling the heteroscedasticity. Typically, the dispersion tends to depend on the mean, e.g., $E\psi'(\epsilon_{ij}) = \phi(\beta^T Z_{ij})^\gamma$. In this case, γ can be estimated by regressing $\psi'(\hat{\epsilon}_{ij})$ on $(\hat{\beta}^T Z_{ij})^\gamma$.

Example 3. (one-dependent case) Let $\gamma = (\gamma_1, ..., \gamma_{K-1})^T$, where $\gamma_j = \text{cov}(\psi_{ij}, \psi_{i,j+1})$ for $j = 1, ..., K-1$. An estimate of γ_j is

$$\hat{\gamma}_j = n^{-1} \sum_{i=1}^{n} \hat{\psi}_{ij} \hat{\psi}_{i,j+1},$$

where an initial estimate $\hat{\beta}$ may be obtained using the working iid assumption. The resulting matrix V will be tridiagonal with $V_{j,j+1} = V_{j+1,j} = \gamma_j$ and its j-th diagonal element estimated similarly by $\sum_{i=1}^n \hat{\psi}_{ij}^2/n$.

Example 4. (continuous AR(1)) Assuming homogeneity, the continuous time AR(1) model specifies $\text{cov}(\epsilon_{it}, \epsilon_{it'}) = \sigma^2 \gamma^{|t-t'|}$. Using an initial estimate of $\gamma = 0$ (equivalent to the working iid equations) and the resulting value of $\hat{\beta}$, an estimate of σ^2 is given by

$$\hat{\sigma}^2 = \sum_i \sum_t \psi(\hat{\epsilon}_{it})^2/nK\,.$$

Based on the relation $E\{\psi(\hat{\epsilon}_{it})\psi(\hat{\epsilon}_{it'})\} \approx \sigma^2 \gamma^{|t-t'|}$, an estimate of γ can be obtained by solving the equation

$$\sum_i \sum_{t<t'} (\hat{\psi}_{it}\hat{\psi}_{it'} - \hat{\sigma}^2 \gamma^{|t-t'|})|t-t'|\gamma^{|t-t'|} = 0$$

with respect to γ. Continuous AR(1) working covariance matrix will be useful in modeling a dependence structure when measurement times vary among subjects.

Example 5. (unspecified dependence and dispersion) When the dependence structure is totally unspecified, V can be estimated by $n^{-1}\sum_{i=1}^n \hat{\psi}_i \hat{\psi}_i^T$. Also, $E\psi'(\epsilon_{ij})$ can be estimated by $n^{-1}\sum_{i=1}^n \psi'(\hat{\epsilon}_{ij})$. If measurement times are fixed for all patients and the error distribution depends only on measurement time, then these working matrices will give an optimal estimator and its covariance matrix will be asymptotically equivalent to V_o.

3. A Goodness-of-fit Test for M-estimation Model

When the assumed model (1) is adequate, estimators are consistent whatever weighting systems we choose, and should therefore be "close" each other in value in some sense. However, the estimators will differ when the model is inadequate. In particular, let $\hat{\beta}_1$ and $\hat{\beta}_2$ be estimates of β_0 obtained from estimating functions

$U_k(\beta) = \sum_{i=1}^n Z_i^T W_{ki}\psi_i(\beta)$ for $k = 1, 2$ respectively. Then, under mild conditions, $\hat{\beta}_k$ converges in probability to a constant β_k^*, the unique solution of the equations $u_k(\beta) = 0$, which may differ from β_0, where $u_k(\beta) = \lim_{n\to\infty} n^{-1}U_k(\beta)$. This motivates a goodness-of-fit test based on the difference between two estimators.

If model (1) holds, we have

$$\hat{\beta}_k - \beta_0 = A_k^{-1}(\beta_0)U_k(\beta_0) + o_p(n^{-1/2}),$$

so that

$$\begin{aligned}\hat{\beta}_1 - \hat{\beta}_2 &= A_1^{-1}(\beta_0)U_1(\beta_0) - A_2^{-1}(\beta_0)U_2(\beta_0) + o_p(n^{-1/2}) \\ &= \sum_{i=1}^n B_i(\beta_0)\psi_i + o_p(n^{-1/2}),\end{aligned} \tag{4}$$

where $A_k(\beta) = \sum_{i=1}^n Z_i^T W_{ki} D_i(\beta) Z_i$ and $B_i(\beta) = A_1^{-1}(\beta)Z_i^T W_{1i} - A_2^{-1}(\beta)Z_i^T W_{2i}$. Since ψ_i are independent random vectors with mean 0, by the multivariate central limit theorem applied to (4), $n^{1/2}(\hat{\beta}_1 - \hat{\beta}_2)$ is asymptotically normal with mean 0. And its covariance matrix can be consistently estimated by

$$\hat{\Sigma} = n\sum_{i=1}^n B_i(\hat{\beta}_1)\psi_i(\hat{\beta}_1)\psi_i(\hat{\beta}_1)^T B_i^T(\hat{\beta}_1).$$

The goodness-of-fit for the model (1) can be tested based on the quadratic form

$$Q = n(\hat{\beta}_1 - \hat{\beta}_2)^T\hat{\Sigma}^{-1}(\hat{\beta}_1 - \hat{\beta}_2),$$

which has an asymptotic χ^2 distribution with p degrees of freedom when model (1) holds. This test statistic is $n^{1/2}$-consistent against any model misspecification under which $\beta_1^* \neq \beta_2^*$ or equivalently $u_k(\beta_{k'}^*) \neq 0$ for $k \neq k'$.

For a general model testing, we may choose the iid working covariance and any working structure such as unspecified covariance structure. This method may be useful when the model is so complicated that it is not easy to display the data by a plot. However, if we want to test the adequacy of a fitted model against a specific alternative model, possibly found from a data plot, we may be able to choose two weighting systems based on the marginal models under consideration. For example, suppose we want to test the adequacy of the linear model $Y_{it} = \beta_0 + \beta_1 t + \epsilon_{it}$ against the piecewise linear model $Y_{it} = \gamma_0 + \gamma_1 t + (\gamma_3 + \gamma_4 t)I(t \geq t_0) + \epsilon_{it}$ for a specific t_0 value. Then we may partition the data into two sets, $\{Y_{it}, t < t_0, i = 1, ..., n\}$ and $\{Y_{it}, t \geq t_0, i = 1, ..., n\}$, and compare the estimates of linear models fitting two subsets of data using a simple working covariance structure such as iid. In this case, two weighting systems are represented as

$$W_{1i} = \begin{bmatrix} I_{1i} & 0 \\ 0 & 0 \end{bmatrix} \text{ and } W_{2i} = \begin{bmatrix} 0 & 0 \\ 0 & I_{2i} \end{bmatrix},$$

where I_{ki} for $k = 1, 2$ are the identity matrices corresponding to the data points of $\{Y_{it}, t < t_0\}$ and $\{Y_{it}, t \geq t_0\}$, respectively. Two subsets of data do not have to be disjoint.

A goodness-of-fit test using two different weighting systems has been used for a two-sample test of proportional hazards by Gill and Schumacher (1987), and for the Cox model by Lin (1991).

4. Application to Real Data

Our methods are illustrated with a study of labor pain; the data are reported by Davis (1991) and are used by Jung (1996). In this study, $n = 83$ women in labor were randomly assigned to a pain medication group (43 women) or placebo group (40 women). At 30 minutes intervals, the self-reported amount of pain was marked on a 100mm line, where 0 = no pain and 100 = very much pain. The maximum number of measurements for each woman was 6, but there are numerous missing values at later measurement times. We assume random missing here. A boxplot taken from Jung (1996) appears in Figure 1 to show the change of pain level over time for two groups.

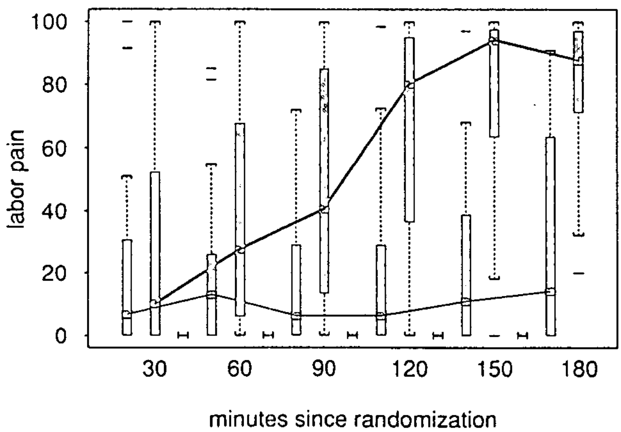

Figure 1. Boxplot of Labor Pain Data (═══ placebo; ─── control)

Let Y_{ij} be the amount of pain for patient i at time j, R_i be the treatment indicator taking 1 for placebo and 0 for treatment and $T_{ij} = j$ be the measurement time divided by 30 minutes. We consider the regression model

$$Y_{ij} = \beta_0 + \beta_1 R_i + \beta_2 T_{ij} + \beta_3 R_i T_{ij} + \epsilon_{ij} \tag{5}$$

containing a separate slope and intercept for each treatment. Note that the predicted value for the treatment group is $\beta_0 + \beta_2 T_{ij}$ and that for the placebo group is $(\beta_0 + \beta_1) + (\beta_2 + \beta_3)T_{ij}$.

For analyses, we considered Winsorized mean regression (Huber,1964) with a score function $\psi(x) = \min(|x|, c)\text{sign}(x)$ for $c = 5$, 30, 50 and ∞. The regression coefficients were estimated based on two different working correlation structures, iid and unspecified. Analysis results are displayed in Table 1. Overall, β_3, the difference in slope between two groups, is highly significant, whereas β_1 and β_2 are insignificant. It is natural that β_1, the difference in baseline pain score level

Table 1. Analysis results of Labor pain data using Model (4) with score function $\psi(x) = \min(|x|, c)\text{sign}(x)$: regression estimates and their standard errors in parentheses, and values of goodness-of-fit test statistic

		Working covariance		Goodness-of-fit
c	β	Iid	Unspecified	Test statistic, χ^2_4
5	β_0	5.387 (3.708)	4.614 (3.302)	2.556*
	β_1	−9.414 (15.445)	−8.677 (13.724)	2.322**
	β_2	0.988 (0.952)	1.485 (1.155)	
	β_3	15.793 (2.491)	15.092 (2.344)	
30	β_0	11.894 (3.346)	12.483 (3.317)	3.801
	β_1	−7.013 (7.497)	−4.912 (5.982)	2.160
	β_2	1.276 (1.058)	1.559 (0.940)	
	β_3	13.107 (1.783)	12.237 (1.484)	
50	β_0	12.428 (3.638)	12.856 (3.542)	2.846
	β_1	0.676 (7.837)	1.172 (6.740)	2.360
	β_2	1.749 (1.180)	2.002 (1.072)	
	β_3	10.334 (2.029)	10.149 (1.750)	
∞	β_0	13.429 (3.915)	14.381 (4.077)	2.595
	β_1	2.229 (7.693)	1.575 (7.262)	2.641
	β_2	1.751 (1.237)	1.930 (1.150)	
	β_3	9.576 (2.035)	9.500 (1.790)	

*: global, **: piecewise

between two groups, should be near zero since this study was a randomized one. Insignificance of β_2 implies that the women in the treatment group have a constant pain level in time.

Regression estimates based on the two working correlations are very close to each other, but use of an unspecified correlation structure tends to give a smaller standard error. When we change the window size c of the M-score function the standard error of regression estimators seems to be minimized at $c^* = 30$. This might be an optimal choice for the marginal error distribution of the data set, see Section 5 for details. From Figure 1 we observe that bands connecting medians ± 30 cover the major part of data points. Although not shown in this paper, residual plot based on Liang and Zeger's GEE leads to the same finding, using the interval $(-30, 30)$ in this case.

A goodness-of-fit test for model (5) was also performed by using iid and unspecified working covariance structures, called 'global method'. From Figure 1, we observe that the median pain level for the control group increases steeply until 120 minutes ($j = 4$) and is flat after then. This motivates a model checking by partitioning the time span into two pieces by $1 \le j \le 4$ and $3 \le j \le 6$ and fitting two linear models to subsets of data with the iid working weights. We call it 'piecewise method'. Note that two subsets are not disjoint in this case. The results are reported in the last column of Table 1. For each choice of c, the test statistics have a χ^2-distribution with 4 degrees of freedom when the model is valid, and none of the tests approach significance.

5. Simulation Studies

The methods proposed in Sections 2 and 3 are based on asymptotic theory. Hence, we conducted simulation studies to investigate finite-sample properties of the procedures.

In the first set of simulations, we want to check the performance of M-estimator in the presence of outliers. For $n = 80$ and $K = 6$, we consider the model

$$Y_{ij} = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2ij} + \beta_3 Z_{1i} Z_{2ij} + \epsilon_{ij},$$

where $Z_{1i} = 0$ if $1 \le i \le n/2$ and $= 0$ otherwise, $Z_{2ij} = j$, $\beta_0 = \beta_3 = 0$, and $\beta_1 = \beta_2 = 1$. Heavy-tailed error terms $\epsilon_{i1}, ..., \epsilon_{iK}$ were generated by adding outliers 5 or -5 with a probability 0.1 each to 0-mean unit-variance random noises with ar(1) process with autocorrelation coefficient ρ or with exchangeable structure with correlation ρ. The regression parameters were estimated using the Winsorized mean score $\psi_c(x) = \min(|x|, c)\text{sign}(x)$ and the iid working correlation or the unspecified covariance weights. Under the given marginal error distribution, an optimal value $c^* = 0.862$ is obtained by minimizing the variance, $\text{var}\{\psi_c(\epsilon_{ij})/\text{E}\psi_c'(\epsilon_{ij})\}$, of the location estimator obtained using ψ_c. In the regression analyses, we use $c = c^*/2, c^*, 2c^*$, or ∞. For each simulated data set, $\hat{\beta}_3$ and its standard error was calculated and $H_0 : \beta_3 = 0$ was tested with the nominal significance level 0.05. The average bias (BIAS) and standard error (SE) of $\hat{\beta}_3$ and the empirical power (PWR) of the testing based on $N = 1000$ simulation data sets are given in Table 2. Note that bias is ignorable most cases. Although SE is minimized when $c = c^*$, it does not change much in a wide range of window sizes around this optimal value. For a fixed M-score function, the unspecified working covariance structure is optimal under the simulation model, so that SE by this working structure is always slightly smaller than that of iid working structure. However, the tests based on the former working structure are anticonservative in many cases, while those based on the latter have PWR close to the nominal significance level 0.05. We found that the standard error formula by the unspecified working covariance structure tends to slightly underestimate the true standard error. Also note that the decrease in SE by using an optimal correlation structure is much smaller than that by using an optimal M-score function. Hence, in the presence of outliers, we may ignore the dependence structure of the error distributions by using iid working covariance structure in estimation. Instead, we had better put more effort on investigating the marginal distribution to find an optimal M-score function, whose window covering major part of the data points.

The second set of simulations were conducted to investigate the performance of the goodness-of-fit test. The true regression model was

$$Y_{ij} = \beta(t_{ij} - 5)^2 + \epsilon_{ij},$$

where $t_{ij} = j$ for $i = 1, ..., 80$, $j = 1, ..., 9$. The error terms were generated as in the first set of simulations. The goodness-of-fit for a linear model was tested by comparing two sets of linear regression estimates obtained by iid working weights and unspecified covariance structures (global method) or by fitting two subsets of data $\{Y_{ij}, 1 \le i \le n, 1 \le j \le 5\}$ and $\{Y_{ij}, 1 \le i \le n, 5 \le j \le 9\}$ (piecewise method) with iid working weights. We set β at 0 and 0.05 for checking type I and II error probabilities respectively. Rejection probability was calculated based on the goodness-of-fit test for linearity with $\alpha = 0.05$ using 1,000 simulation samples. Regression estimates were obtained using Winsorized mean score function with $c = c^*$ (Huber) or $c = \infty$ (L_2). From Table 3 (a), we observe that global method is slightly conservative especially under exchangeable correlation structure, while piecewise method has type I error probability close to the nominal significance level overall. When the true model is quadratic (Table 3 (b)), goodness-of-fit tests using Huber's regression model and piecewise method have high power. Whether we use piecewise method or global method, tests using Huber's regression model have higher power than those using L_2 regression model. Since both iid and unspecified working covariance structures give relatively symmetric weight around $t = 5$ in our simulation setting, two sets of linear regression estimates should be very close, and consequently global method has a low power.

6. Concluding Remarks

We have proposed a regression method of longitudinal data which is robust not only to outliers and heavy-tailed distributions but also to misspecification of the covariance and dispersion structure of the repeated measurements. It has been known that using a working covariance structure close to the true one can improve the sufficiency of regression estimator, see for example Liang and Zeger (1986). However, in the presence of outliers, the improvement attained by using an optimal working covariance structure is rather marginal. Furthermore, from our experience, we found that, compared to the choice of working iid matrix, the choice of optimal working structure (or the true dispersion and correlation matrices) does not improve the efficiency of regression estimators much. Choice of some sophisticated working structure can ruin the efficiency of the estimators. Crowder (1995) and Sutradhar and Das (1999) provide excellent discussions on this issue in L_2 regression case. Instead, we recommend to choose an optimal M-score function to decrease the influence of outliers based on marginal error distribution, which can be visualized by residual plots based on L_2 regression methods such as GEE. By specifying the true error distributions, we may be able to find an optimal score function analytically. In real data analysis, however, this can be very impractical, especially if the error distributions are heterogeneous. Through extensive data analyses and simulation studies, we found that efficiency can be almost equally improved by wide choices of M-score function covering major part of data points.

So far, we have assumed that ψ is absolutely continuous. More generally, a score function has the form $\psi(x) = \psi_1(x) + \psi_2(x)$, where ψ_1 is absolutely continuous and ψ_2 is a step function with finitely many jumps of finite jump sizes. For example, the score $\psi(x) = xI(|x| \leq c)$ for the Huber's (1964) trimmed mean is expressed as a sum of $\psi_1(x) = \min(|x|, c)\text{sign}(x)$ and $\psi_2(x) = \text{sign}(-x)cI(|x| \leq c)$. In this case, the estimating function for β is not continuous and hence, in estimation, efficient algorithms based on its differentiability cannot be used. Furthermore, $\psi'(\epsilon_{ij})$ in the variance formula should be replaced by $\psi_1'(\epsilon_{ij}) + \sum_l d_l f_{ij}(a_l)$, where the summation is over jump points $a_1, ..., a_m$ of ψ_2 with jump sizes $d_1, ..., d_m$ and f_{ij} is the marginal density function of ϵ_{ij}. Hence, a consistent estimate of asymptotic covariance matrix of $\hat{\beta}$ requires density estimation using the residuals $\hat{\epsilon}_{ij}$. When the error distributions are complicatedly heterogeneous, it may be very difficult to estimate the density functions.

Table 2. BIAS, SE and PWR of the inferences on β_3

Covariance			Window size			
True	Working		$c^*/2$	c^*	$2c^*$	∞
Exchangeable						
$\rho = .4$	Iid	BIAS	−.004	−.004	−.004	−.007
		SE	.072	.069	.076	.125
		PWR	.047	.047	.050	.057
	Unspecified	BIAS	−.005	−.004	−.003	−.006
		SE	.070	.067	.074	.121
		PWR	.066	.066	.066	.073
$\rho = .8$	Iid	BIAS	−.004	−.003	−.004	−.005
		SE	.062	.059	.068	.120
		PWR	.039	.038	.040	.057
	Unspecified	BIAS	−.003	−.003	−.003	−.004
		SE	.059	.057	.065	.116
		PWR	.044	.044	.048	.074
AR(1)						
$\rho = .4$	Iid	BIAS	−.001	.000	.000	−.003
		SE	.086	.083	.088	.132
		PWR	.047	.045	.045	.050
	Unspecified	BIAS	.001	.002	.001	−.002
		SE	.082	.080	.086	.128
		PWR	.065	.062	.043	.058
$\rho = .8$	Iid	BIAS	−.002	−.003	−.003	−.004
		SE	.083	.078	.083	.129
		PWR	.033	.041	.043	.052
	Unspecified	BIAS	−.002	−.002	−.001	−.003
		SE	.078	.076	.080	.125
		PWR	.053	.048	.047	.058

Table 3. Goodness-of-fit test for linearity

(a) Type I error probabilities with the true model $Y_{ij} = \epsilon_{ij}$

True		Global		Piecewise	
covariance	ρ	Huber	L_2	Huber	L_2
Exchangeable	.4	.033	.036	.055	.057
	.8	.028	.035	.048	.044
AR(1)	.4	.040	.038	.052	.052
	.8	.042	.035	.049	.054

(b) Powers with the true model $Y_{ij} = .05(j - 4.5)^2 + \epsilon_{ij}$

True		Global		Piecewise	
covariance	ρ	Huber	L_2	Huber	L_2
Exchangeable	.4	.433	.247	.988	.597
	.8	.462	.257	.995	.637
AR(1)	.4	.471	.247	.956	.555
	.8	.515	.265	.980	.588

Acknowledgement

This research was supported by NON DIRECTED RESEARCH FUND, Korea Research Foundation.

References

Crowder, M. (1995). On the use of working correlation matrix in using generalized linear models for repeated measures. *Biometrika* **82**, 407-410.

Davis, C.S. (1991). Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in Medicine* **10**, 1959-1980.

Freedman D.A. and Diaconis P. (1982). On inconsistent M-estimators. *Annals of Statistics* **10**, 454-461.

Gill R. and Schumacher M. (1987). A simple test of the proportional hazards assumption. *Biometrika* **74**, 289-300.

Huber P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics* **35**, 73-101.

Huber P.J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Annals of Statistics* **1**, 799-821.

Jung, S.H. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association* **91**, 251-257.

Lai T.L. and Ying Z. (1994). A missing information principle and M-estimators in regression analysis with censored and truncated data. *Annals of Statistics* **22**, 1222-1255.

Liang K.Y. and Zeger S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

Lin D.Y. (1991). Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association* **86**, 725-728.

Rousseeuw P.J. and and Leroy A.M.(1987). *Robust regression and outlier detection.* New York: Wiley.

Rubin D.B. (1976). Inference and missing data. *Biometrika* **63**, 81-92.

Sutradhar, B.C. and Das, K. (1999). On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika* **86**, 459-465.

Zhou, M. (1992). M-estimation in censored linear models. *Biometrika* **79**, 837-841.