Statistica Sinica **19** (2009), 1247-1261

INFERENCE ON QUANTILE REGRESSION FOR HETEROSCEDASTIC MIXED MODELS

Huixia Judy Wang

North Carolina State University

Abstract: This paper develops two weighted quantile rank score tests for the significance of fixed effects in a class of mixed models with nonhomogeneous groups. One test is constructed by weighting the residuals to account for heteroscedasticity, while the other test is based on asymptotically optimal weights accounting for both heteroscedasticity and correlation. Without appropriate weights to account for heteroscedasticity, the quantile rank score tests often perform poorly. In finite samples, the test with optimal weights tends to provide marginal improvement over the one with simpler weights unless the intra-subject correlation is extremely high. The proposed methods are useful to accommodate nonparametric error distributions in studying the effect of covariates on any conditional quantile of the response distribution. We illustrate the value of the proposed methods by modeling several quantiles of the apnea duration of the elderly during normal swallowing. Our method suggests significant interaction effect between feeding type and viscosity in the upper quantiles of the apnea distribution, a result that tends to be overlooked by usual linear mixed model approaches.

Key words and phrases: Apnea duration, estimating equation, longitudinal data, rank score test, weighted quantile regression.

1. Introduction

Mixed models are widely used in the analyses of correlated and longitudinal data to incorporate both the inter-subject variation and the variation among measurements within a subject. In many applications, the within-subject errors exhibit heterogeneity, with variation depending on certain continuous or discrete covariates. Under the normality assumption, a number of methods were proposed to analyze longitudinal data with heterogeneity, for instance, Davidian and Giltinan (1993), Lin, Raz and Harlow (1997) and Kizilkaya and Tempelman (2005), among others.

Though the normality assumption provides mathematical convenience, it is not realistic in many applications, and its violation may lead to invalid statistical results; see Agresti, Caffo and Ohman (2004) and Hartford and Dividian (2000) for related discussions. One way to relax normality is to employ more general densities. Zhang and Davidian (2001) used a semi-nonparametric density, and Ghidey, Lesaffre and Eilers (2004) used a mixture of normal densities to approximate the random effects density. Cai and Dunson (2006) considered a Bayesian variable selection approach that accommodated non-normal response distributions. From a frequentist perspective, Zhou and He (2008) considered the skwed-t distributions for both random error and random effects; Jara, Quintana and Martín (2008) did the same from a Bayesian perspective. All these approaches are likelihood-based, and the inferences focus on the conditional mean change of the response variable Y given covariates X. The current paper considers an alternative quantile regression approach that focuses on the conditional quantiles of Y given X.

This paper is motivated by a study of swallowing. The purpose of the study was to determine how the apnea duration, a measurement of duration of breathing suspension, is affected by viscosity and different feeding types. In this study, the distribution of apnea duration was skewed to the right even after a log transformation. For such data, a quantile regression approach is able to capture the difference between groups at different locations of the response distribution, and thus can provide more comprehensive information than mean-based methods. As longer apnea durations are often due to pathological disorder, the upper quantiles are of more clinical importance. In addition, the apnea durations exhibit different variations in each viscosity condition, so appropriate quantile inference method is needed to account for this heteroscedasticity.

We develop two weighted quantile rank score tests for the significance of fixed effects in a class of mixed models with nonhomogeneous groups. One test is constructed by weighting the residuals by the original scale to account for heteroscedasticity, while the other test is constructed by incorporating both the scale and correlation structure of the residuals. The developed methods do not rely on any distributional assumptions, thus are useful to accommodate non-normal errors. With correctly specified weights to account for the heteroscedasticity among different groups, the proposed test statistics are asymptotically chi-square distributed. Confidence intervals for the fixed effects can be constructed by inverting the proposed rank score tests.

The next section of the paper introduces the motivating example, and establishes the weighted rank score tests. The performance of the proposed tests is illustrated through a simulation study in Section 3, and through the swallow study in Section 4. Section 5 provides some concluding remarks.

2. Data and Proposed Methods

2.1. Motivating swallow study

The swallow study was conducted by researchers at the University of Illinois at Urbana-Champaign; see Perlman, He, Barkmeier and Van Leer (2005) for clinical details. A total of 31 healthy females, aged 70−85, participated in the experiment. Each person was presented with 10ml of water or pudding under two conditions, self-fed and examiner-fed. The number of swallows varied from 4 to 8, and total 860 swallows were observed. The apnea duration (in seconds) was recorded during each swallowing. The study of apnea duration for elderly normal adults provides better understanding of the respiratory pattern and serves as a comparison to disordered swallows. Quantile regression has special value for analyzing this data set for two reasons: (1) preliminary studies show that the apnea duration distribution is highly non-normal even for log transformed data; (2) the lower and upper tails of the response distribution depends on the covariates.

Another important feature is that water and pudding exhibit different error variances. More specifically, the estimated median absolute deviation (MAD) of the error from drinking water was 0.3, while that from consuming pudding was 0.1. Ignoring this heteroscedasticity might lead to tests with either the wrong level or less power. Incorporating heteroscedasticity into the model gives

$$y_{ijkl} = \mu + V_j + F_k + VF_{jk} + a_i + \sigma_j e_{ijkl}, \tag{2.1}$$

where y_{ijkl} is the apnea duration, μ, V, F and VF stand for the intercept, viscosity effect, feeding type effect, and the interaction effect, respectively, a_i is the i.i.d. random subject effect, e_{ijkl} is the i.i.d. random error, and σ_j captures the block-wise heteroscedasticity in the error, $i = 1, \ldots, n$, $j = 1, \ldots, J$, $k = 1, \ldots, K$ and $l = 1, \ldots, m_{ijk}$. For the particular data studied, $J = K = 2$. We assume that a_i and e_{ijkl} are mutually independent. Model (2.1) represents a typical repeated-measures design with within-subject factors, where each subject is measured multiple times under J nonhomogeneous conditions (or blocks).

Our aim is to test a subset of the covariate effects. We first partition the fixed effects as $(\alpha, \beta) \in (\mathbb{R}^p, \mathbb{R}^q)$. For instance, to test the interaction effect in Model (2.1), $\alpha \in \mathbb{R}^3$ corresponds to the main effects, and $\beta \in \mathbb{R}^1$. Without loss of generality, we assume a balanced design, that is, each subject has $m = \sum_{k=1}^{K} m_{ijk}$ repeated measurements under each of the nonhomogeneous conditions. To simplify the notation, we omit the subscript k in what follows. Let $M = Jm$ be the number of measurements of each subject, and $N = nJm$ be the total number of observations. We rewrite Model (2.1) in the more general form

$$y_{ijl} = x_{ijl}^T \alpha + z_{ijl}^T \beta + a_i + \sigma_j e_{ijl}, \quad i = 1, \ldots, n, \ j = 1, \ldots, J, \ l = 1, \ldots, m, \tag{2.2}$$

or

$$Y_i = X_i \alpha + Z_i \beta + U_i, \tag{2.3}$$

where $Y_i = (y_{i11}, \ldots, y_{iJm})^T$ denotes the response variable for the ith subject, $X_i = (x_{i11}, \ldots, x_{iJm})^T$ is an $M \times p$ design matrix, $Z_i = (z_{i11}, \ldots, z_{iJm})^T$ is an $M \times$

q design matrix, and $U_i = (u_{i11}, \ldots, u_{iJm})^T$ is the vector of composite error with $u_{ijl} = a_i + \sigma_j e_{ijl}$. We write $X = (x_{i11}, \ldots, x_{nJm})^T$ and $Z = (z_{i11}, \ldots, z_{nJm})^T$. For the purpose of identifiability, we assume that the τth quantile of u_{ijl} is 0. At any given quantile level $0 < \tau < 1$, we consider the conditional quantile of y given (x, z). We are interested in testing $H_0 : \beta = 0$.

We first review a simple quantile estimator that is obtained under the working assumption of independence and homoscedasticity. Under H_0, α can be estimated by

$$\hat{\alpha} = \underset{a \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{ijl} \rho_\tau(y_{ijl} - x_{ijl}^T a), \tag{2.4}$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is the quantile loss function and $I(\cdot)$ is the indicator function; see Koenker (2005) for a comprehensive review on quantile regression. He, Zhu and Fung (2002), Wei and He (2006) and Wang and Fygenson (2009), among others, investigated the behaviors of $\hat{\alpha}$ for longitudinal data under different contexts, and demonstrated that $\hat{\alpha}$ remains a consistent estimator of α for dependent data.

For linear mixed models with homogenous errors, Wang and He (2007) proposed a Quantile Rank Score test (QRS) based on $\hat{\alpha}$. However, for models with severe heteroscedasticity, we show in Section 3 that QRS produces inflated Type I error rates due to the incorrect weights used in the chi-square approximation. In the following subsections we propose two tests based on weighted estimators: one accounts for the heteroscedasticity, and the other accounts for both the dependence structure of the residuals and heteroscedasticity.

2.2. A simple weighted quantile approach

Let f_j denote the common density function of the u_{ijl} for the jth block. We consider weighting the residuals by the original block-wise scale. We find the weighted estimator $\tilde{\alpha}_1$ as a solution to

$$n^{-1} \sum_{i=1}^{n} X_i^T W_i \psi_\tau(Y_i - X_i \alpha) \approx 0, \tag{2.5}$$

where $W_i = diag\{f_1(0), \ldots, f_J(0)\} \otimes I_m$, $\psi_\tau(\varepsilon) = (\psi_\tau(\varepsilon_1), \ldots, \psi_\tau(\varepsilon_M))^T$ for any vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_M)^T$, $\psi_\tau(u) = \tau - I(u < 0)$, and I_m denotes the $m \times m$ identity matrix. In quantile regression, the weights are often taken to be proportional to the block-wise density evaluated at the quantile of interest, rather than as the reciprocals of the block-wise standard deviations. Since W_i is a diagonal matrix, solving (2.5) is equivalent to minimizing the weighted quantile objective function $\sum_{ijl} f_j(0) \rho_\tau(y_{ijl} - x_{ijl}^T a)$. Therefore $\tilde{\alpha}_1$ can be obtained directly by using the "rq" function in the R package *quantreg*. In practice, the unknown density function

f_j has to be estimated in order to compute the weighted estimators and perform the tests to be introduced. A specific choice is made in Remark 3.

From now on, we make the following assumptions.

A.1 For each j, the positive continuous density function f_j has a bounded first order derivative.

A.2 M is uniformly bounded as $n \to \infty$.

A.3 Let $\tilde{X} = (X, Z) = (\tilde{x}_{ijl})$, $\tilde{x}_{ijl}$ have uniformly bounded third moments for all i, j, l.

We let $\tilde{Z}_{(1)} = (\tilde{z}_{ijl(1)}) = \{I_N - X(X^T W^2 X)^{-1} X^T W^2\} Z$, where $W = I_n \otimes W_1$, and $\tilde{Z}_{i(1)} = (\tilde{z}_{i11(1)}, \ldots, \tilde{z}_{iJm(1)})^T$. The weighted quantile rank score test $WQRS_1$ is based on

$$\tilde{S}_{(1)} = n^{-1/2} \sum_{i=1}^{n} \tilde{Z}_{i(1)}^T W_i \psi_\tau (Y_i - X_i \tilde{\alpha}_{(1)}).$$

Following a similar argument as in the proof of Lemma A.2 in Wang and He (2007), for some constant C,

$$\sup_{\|t\| \le C} \|r_n(t) - E\{r_n(t)\}\| = o_p(1), \tag{2.6}$$

where $r_n(t) = n^{-1/2} \sum_i \tilde{Z}_{i(1)}^T W_i \{\psi_\tau(U_i + n^{-1/2} X_i^T t) - \psi_\tau(U_i)\}$. Let F_j denote the CDF of residuals from the jth block. Expanding $E\{r_n(t)\}$ around 0, we have

$$\begin{aligned}\sup_{\|t\| \le C} \|E\{r_n(t)\}\| &= \sup_{\|t\| \le C} \left\| n^{-1/2} \sum_{ijl} \tilde{z}_{ijl(1)}^T f_j(0) \Big\{ F_j(0) - F_j(-n^{-1/2} x_{ijl}^T t) \Big\} \right\| \\ &= \sup_{\|t\| \le C} \left\| n^{-1} \tilde{Z}_{(1)}^T W^2 X t \right\| + o(1) = o(1), \end{aligned} \tag{2.7}$$

where the last step is due to the fact that $\tilde{Z}_{(1)}$ is orthogonal to $W^2 X$. With a trivial modification of the proof given in Theorem 1 of He et al. (2002), we obtain $\tilde{\alpha}_{(1)} - \alpha = O_p(n^{-1/2})$, which together with (2.6) and (2.7) yields

$$\tilde{S}_{(1)} - n^{-1/2} \sum_{i=1}^{n} \tilde{Z}_{i(1)}^T W_i \psi_\tau(u_{ijl}) = o_p(1). \tag{2.8}$$

The asymptotic normality of $\tilde{S}_{(1)}$ thus follows from the Lindberg-Feller Central Limit Theorem.

Under Model (2.2), $u_{ijl} = a_i + \sigma_j e_{ijl}$ has an exchangeable correlation structure within the same subject i and the same block j. With a tedious but otherwise routine variance-covariance calculation, we obtain the asymptotic variance-covariance matrix of $\tilde{S}_{(1)}$ as

$$\tilde{Q}_{(1)}(\Delta) = n^{-1} \sum_{i=1}^{n} \tilde{Z}_{i(1)}^T W_i V_i(\Delta) W_i \tilde{Z}_{i(1)}, \tag{2.9}$$

where

$$V_i(\Delta) = Cov(\psi_\tau(U_i)) = \begin{pmatrix} \Sigma_{11} & \delta_{12} 1_m & \cdots & \delta_{1J} 1_m \\ & \Sigma_{22} & \cdots & \delta_{2J} 1_m \\ & & \ddots & \\ & & & \Sigma_{JJ} \end{pmatrix} - \tau^2 1_M 1_M^T,$$

$\Sigma_{jj} = (\tau - \delta_{jj}) I_m + \delta_{jj} 1_m 1_m^T$, 1_m denotes the m-dimensional vector with all 1's, Δ is a collection of correlation parameters $\delta_{jj} = P(u_{ijl} < 0, u_{ijl'} < 0, l \neq l')$, and $\delta_{jj'} = P(u_{ijl} < 0, u_{ij'l'} < 0, j \neq j')$. The $V_i(\Delta)$ has a block structure consisting of total $J(J+1)/2$ unknown δ's. The δ_{jj} measures the dependence between repeated measurements from the same subject and the same block j, and $\delta_{jj'}$ measures the dependence between observations from the same subject but different blocks.

The weighted quantile rank score test statistic is

$$\tilde{T}_{(1)} = \tilde{S}_{(1)}^T \Big\{ \tilde{Q}_{(1)}(\Delta) \Big\}^{-1} \tilde{S}_{(1)}. \tag{2.10}$$

It follows from some routine technical details that, under H_0, $\tilde{T}_{(1)}$ is asymptotically χ_q^2 as $n \to \infty$.

In practice, Δ is unknown and has to be estimated. One natural and consistent estimator is $\tilde{\Delta} = (\tilde{\delta}_{jj}, \tilde{\delta}_{jj'})$ with

$$\begin{aligned} \tilde{\delta}_{jj} &= \underset{i,l \neq l'}{\text{Average}}\, I\Big\{ y_{ijl} - x_{ijl}^T \tilde{\alpha}_{(1)} < 0, y_{ijl'} - x_{ijl'}^T \tilde{\alpha}_{(1)} < 0 \Big\}, \\ \tilde{\delta}_{jj'} &= \underset{i,l,l'}{\text{Average}}\, I\Big\{ y_{ijl} - x_{ijl}^T \tilde{\alpha}_{(1)} < 0, y_{ij'l'} - x_{ij'l'}^T \tilde{\alpha}_{(1)} < 0 \Big\}, \quad j \neq j'. \end{aligned} \tag{2.11}$$

Note that, asymptotically, the $\tilde{\alpha}_1$ in (2.11) can be replaced by any other consistent estimator of α, such as $\hat{\alpha}$ and $\tilde{\alpha}_2$ to be introduced in Section 2.4. In our applications, these estimators of α yield very similar Δ estimates, even in finite samples.

Remark 1. In the special case with homoscedastic errors, $f_j = f$ is common across j, $W_i = f(0) I_M$, $\tilde{Z}_{(1)} = \{I_N - X(X^T X)^{-1} X^T\} Z \doteq Z^*$, and $\delta_{jj} = \delta_{jj'} = \delta$. The common δ measures the exchangeable intra-subject correlation. Therefore, we have $V_i(\Delta) = (\tau - \delta) I_M + (\delta - \tau^2) 1_M 1_M^T$. The $WQRS_1$ thus reduces to the QRS of Wang and He (2007).

Remark 2. The validity of the chi-square approximation, mentioned above, relies on the correct specification of the weight matrix W. With an incorrectly chosen W as in QRS, the first-order Taylor expansion of $E\{r_n(t)\}$ may not vanish, thus (2.8) no longer holds.

Remark 3. In order to compute $\tilde{\alpha}$ and to perform a valid test using the chi-square approximation, a consistent estimation of W or $f_j(0)$ is needed. In practice, for each j we estimate f_j using the Gaussian kernel density estimation method based on the estimated residuals $\hat{u}_{ijl} = y_{ijl} - x_{ijl}\hat{\alpha}$. Under Assumption A.1, it follows directly from Francq and Tran (2002) that our density estimate is still consistent for f_j even though the residuals are dependent.

2.3. The simplification of $\tilde{Q}_{(1)}(\Delta)$ in a special case

For the swallow data set, the asymptotic variance-covariance matrix $\tilde{Q}_{(1)}(\Delta)$ can be simplified as follows. Suppose we are interested in testing for the interaction effect between viscosity and feeding type, as in (2.1). Then we have $\tilde{Z}_{i(1)} = (\tilde{z}_{ijl(1)}) = c_1(v_1, -v_1, -v_2, v_2)^T \otimes 1_{m/2}$, $j = 1, 2$, $l = 1, \ldots, m$, and

$$\begin{aligned}\tilde{Q}_{(1)}(\Delta) &= n^{-1}\Big\{\sum_{ijl} \tilde{z}^2_{ijl(1)}\tau(1-\tau) + \sum_{ij,l\neq l'} \tilde{z}_{ijl(1)}\tilde{z}_{ijl'(1)}(\delta_{jj} - \tau^2) \\ &\quad + \sum_{i,j\neq j',l,l'} \tilde{z}_{ijl(1)}\tilde{z}_{ij'l'(1)}(\delta_{jj'} - \tau^2)\Big\} \\ &= c_2\Big\{\tau(1-\tau)(v_1^2 + v_2^2) - v_1^2(\delta_{11} - \tau^2) - v_2^2(\delta_{22} - \tau^2)\Big\},\end{aligned}$$

where $v_j = \{f_j(0)\}^{-2}$, and c_1 and c_2 are some constants. In this case, $\tilde{Q}_{(1)}(\Delta)$ mainly depends on the weights, and on the correlations between observations from the same subject and the same block.

2.4. Optimally weighted quantile approach

For median regression, Jung (1996) proposed to weight the residuals utilizing both scale and correlation structure information, and this weight matrix was shown to be optimal in terms of asymptotic efficiency. We extend Jung's idea to quantile regression and consider the optimally weighted estimator $\tilde{\alpha}_2$, as the solution to

$$n^{-1}\sum_{i=1}^{n} X_i^T W_i \Big\{V_i(\Delta)\Big\}^{-1} \psi_\tau(Y_i - X_i\alpha) \approx 0. \tag{2.12}$$

In practical applications, we estimate Δ using $\hat{\alpha}$ when solving (2.12).

Following the reweighted least squares algorithm, we can compute $\tilde{\alpha}_2$ by iterating

$$\tilde{\alpha}_2 \leftarrow \left[\sum_{i=1}^{n} X_i^T W_i \Big\{V_i(\Delta)\Big\}^{-1} A_i X_i\right]^{-1} \left[\sum_{i=1}^{n} X_i^T W_i \Big\{V_i(\Delta)\Big\}^{-1} A_i Y_i\right],$$

where $A_i = diag\{\psi_\tau(y_{ijl} - x_{ijl}^T\tilde{\alpha}_2)/(y_{ijl} - x_{ijl}^T\tilde{\alpha}_2)\}$, with the convention that $\psi_\tau(u)/u = 0$ when $u = 0$. For faster convergence, we suggest using $\hat{\alpha}$ as the starting value.

Based on $\tilde{\alpha}_2$, we construct the optimal weighted quantile rank score test, denoted by $WQRS_2$, as follows. The test statistic is

$$\tilde{T}_{(2)} = \tilde{S}_{(2)}^T \Big\{\tilde{Q}_{(2)}(\Delta)\Big\}^{-1} \tilde{S}_{(2)}, \tag{2.13}$$

where

$$\tilde{S}_{(2)} = n^{-1/2} \sum_{i=1}^{n} \tilde{Z}_{i(2)}^T W_i \Big\{V_i(\Delta)\Big\}^{-1} \psi_\tau(Y_i - X_i\tilde{\alpha}_{(2)}),$$

$$\tilde{Q}_{(2)}(\Delta) = n^{-1} \sum_{i=1}^{n} \tilde{Z}_{i(2)}^T W_i V_i(\Delta) W_i \tilde{Z}_{i(2)}, \quad \tilde{Z}_{i(2)} = (\tilde{z}_{i11(2)}, \ldots, \tilde{z}_{iJm(2)})^T,$$

$$\tilde{Z}_{(2)} = (\tilde{z}_{ijl(2)}) = \Big[I_N - X(X^T W^2 \{V(\Delta)\}^{-1} X)^{-1} X^T W^2 \{V(\Delta)\}^{-1}\Big] Z.$$

Similar to Section 2.3, we can show that, under H_0, $\tilde{T}_{(2)} \xrightarrow{D} \chi_q^2$ as $n \to \infty$.

2.5. Construction of confidence intervals

The confidence interval for the quantile coefficient $\beta \in \mathbb{R}^1$ at a given quantile level τ can be constructed by inverting the weighted rank score test statistics. Let $\tilde{T}(\beta_0)$ be the test statistic of $WQRS_1$ or $WQRS_2$ for testing $H_0 : \beta = \beta_0$. Note that the corresponding quantile rank score $\tilde{S}(\beta_0)$ is monotone in β_0. Therefore, for a fixed Δ, the set $\{\beta_0 : \tilde{T}(\beta_0) \le \chi^2_{1-\alpha,1}\}$ is a $(1-\alpha)$ confidence interval for β, where $\chi^2_{1-\alpha,1}$ is the $(1-\alpha)$th quantile of χ^2_1. For models with independent errors, Koenker (1994) discussed confidence intervals via inversion of a rank score test, that is a special case of $WQRS_1$ with $V_i(\Delta) = (\tau - \tau^2)I_M$. For dependent data, the variance of $\tilde{S}(\beta_0)$ involves the unknown correlation parameters Δ. The empirical estimator $\tilde{\Delta}$ based on the estimated residuals obtained under H_0 depends on β_0, and it is asymptotically consistent under the local alternative. When Δ is estimated by $\tilde{\Delta}$, we employ a two-step procedure to obtain an approximate confidence interval for β, whose coverage probability converges to $1-\alpha$ as $n \to \infty$. First we obtain a rough $(1-\alpha)$ confidence interval (a_0, b_0) for β, assuming independence. Then we perform the weighted quantile rank score test around a_0 by moving a small distance to the left or right to find the lower confidence limit a at which the test statistic value is within 0.05 of the critical value. Following a similar procedure we can find the upper confidence limit b. In empirical studies, we find that a fairly small number of searches, around 20, is needed.

3. Simulation Study

We conducted a simulation study to assess the performance of proposed methods in both homoscedastic and heteroscedastic mixed models. The simulation study mimics the apnea duration experiment analyzed in this paper to study the effects of viscosity and feeding type. For comparative purpose, besides QRS, we also considered the rank score test (denoted by $Indep$) implemented in the R package *quantreg*, assuming independent data.

The simulated data sets were generated from the model

$$y_{ijkl} = 1 + x_{ijkl}\alpha + w_{ijkl}\beta + z_{ijkl}\gamma + a_i + \sigma_j e_{ijkl} - 0.4\sqrt{1+\sigma_j^2}\Phi^{-1}(\tau), \quad (3.1)$$

where $i = 1, \ldots, 30$, $j = 1, 2$, $k = 1, 2$, $l = 1, \ldots, 6$, $x_{i1kl} = w_{ij1l} = 0$, $x_{i2kl} = w_{ij2l} = 1$, $z_{ijkl} = x_{ijkl} \cdot w_{ijkl}$, and Φ is the cdf of $N(0, 1)$. The α, β, and γ correspond to the viscosity, feeding type, and viscosity-feeding interaction effects in the swallow study. As an illustration, we generated both random subject effects a_i and random errors e_{ijkl} from $N(0, 0.4^2)$. Note that under this setting, the marginal distribution of $u_{ij} = a_i + \sigma_j e_{ijkl}$ was $N(0, 0.4^2(1 + \sigma_j^2))$. Therefore, $u_{ij} - 0.4\sqrt{1+\sigma_j^2}\Phi^{-1}(\tau)$ has zero τth quantile for each i and j, satisfying the assumption made in Model (2.3) for the identifiability purpose. To assess the Type I errors of different methods, we set $(\alpha, \beta, \gamma) = (0, 1, 0)$ for testing $H_0 : \alpha = 0$, and $(\alpha, \beta, \gamma) = (2, 0, 0)$ for testing $H_0 : \beta = 0$. For testing $H_0 : \gamma = 0$, we let $\alpha = 2$, $\beta = 1$ and varied γ from 0 to 0.5 to assess the Type I error and power.

In this simulation, four different cases were considered. Case 1 was a homoscedastic model with $\sigma_1 = \sigma_2 = 1$, so that the intra-subject correlation, denoted by ϱ, was 0.5, and $\delta_{11} = \delta_{22} = 0.333$ at median. Cases 2−4 were heteroscedastic models. In Case 2, we set $\sigma_1 = 1$ and $\sigma_2 = 0.5$, resulting in $\delta_{11} = 0.333$ and $\delta_{22} = 0.398$ (corresponding to $\varrho = 0.8$) at median. Case 3 had a larger degree of heteroscedasticity with $\sigma_1 = 2$ and $\sigma_2 = 0.5$, resulting in $\delta_{11} = 0.282$ (corresponding to $\varrho = 0.2$) and $\delta_{22} = 0.398$ at median. Case 4 was similar to Case 3 but with $\sigma_2 = 0.25$ associated with an extremely large intra-subject correlation $\varrho = 0.941$ in the second viscosity group. We focused on the quantile levels $\tau = 0.25$ and 0.50, and repeated the simulation 500 times for each case.

Table 1 summarizes the Type I errors of different methods for testing α, β, and γ under the respective null hypotheses. The nominal significance level

Table 1. The Type I errors of *Indep*, QRS, $WQRS_1$ and $WQRS_2$ in Cases 1−4 for testing three different null hypotheses. The significance level is 0.05.

	$\tau = 0.25$				$\tau = 0.5$			
	Indep	QRS	$WQRS_1$	$WQRS_2$	*Indep*	QRS	$WQRS_1$	$WQRS_2$
$H_0 : \alpha = 0$								
Case 1	0.018	0.044	0.044	0.048	0.016	0.054	0.048	0.052
Case 2	0.010	0.068	0.050	0.046	0.026	0.070	0.044	0.032
Case 3	0.054	**0.098**	0.036	0.038	0.050	**0.102**	0.058	0.046
Case 4	0.082	**0.166**	0.074	0.036	0.070	**0.142**	0.064	0.035
$H_0 : \beta = 0$								
Case 1	0.018	0.042	0.040	0.042	0.010	0.034	0.044	0.060
Case 2	0.004	0.030	0.028	0.034	0.002	0.048	0.048	0.048
Case 3	0.008	0.034	0.042	0.036	0.008	0.034	0.044	0.048
Case 4	0.044	0.064	0.040	0.054	0.018	0.068	0.034	0.034
$H_0 : \gamma = 0$								
Case 1	0.026	0.050	0.038	0.034	0.006	0.044	0.044	0.038
Case 2	0.020	0.048	0.050	0.052	0.004	0.032	0.024	0.026
Case 3	0.042	**0.080**	0.048	0.044	0.016	**0.072**	0.034	0.038
Case 4	0.044	**0.104**	0.040	0.054	0.018	**0.088**	0.024	0.024

was 0.05. Note that under the current simulation design, for inference on β, $Z^* = 1_{30} \otimes (1, 1, -1, -1)^T \otimes 1_6$ was orthogonal to W and thus the chi-square approximation used in QRS was valid. However, this would not be true for general designs. For inferences on α and γ, QRS showed reasonable robustness to modest heteroscedasticity (Case 2), but it produced inflated levels at both quantile levels in Cases 3−4 where the homoscedasticity assumption was substantially violated. On the other hand the *Indep* test, was too conservative; see Figure 1 for the power curves of all four tests at $\tau = 0.5$ for testing γ. Both $WQRS_1$ and $WQRS_2$ gave decent Type I errors and competitive powers, and $WQRS_2$ was slightly more powerful than $WQRS_1$ only in Case 4, where the intra-subject correlation was extremely high.

Besides the testing performance, we also compared the efficiency of the working independence estimator and two weighted quantile estimators for estimating the viscosity effect α and the feeding type effect β when the data was generated with $(\alpha, \beta, \gamma) = (2, 1, 0)$. The results at median are summarized in Table 2. The three estimators performed similarly for estimating α, while the two weighted estimators had higher efficiency (and less bias) for estimating β in Cases 2−4. One explanation is that in these heteroscedastic models, residuals from the same j (viscosity) are identically distributed, but those from the same k (feeding type) are non-identically distributed. Therefore, weighting the residuals by the original scale reduces the heterogeneity within the same k, and thus improves the efficiency of the feeding type effect estimation in a more obvious way. $\tilde{\beta}_2$ is clearly

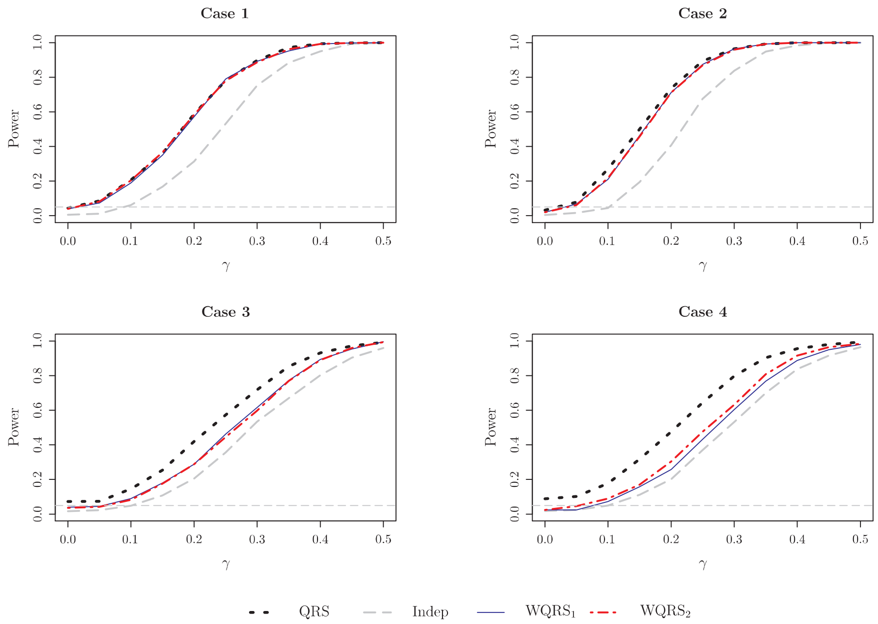

Figure 1. Power curves of $Indep$, QRS, $WQRS_1$ and $WQRS_2$ for testing γ in Cases 1-4 at $\tau = 0.5$. The dashed horizontal line is the nominal 0.05 level.

Table 2. The comparison of finite sample mean squared errors (MSE) and average bias (Bias) of three estimators: working independence estimator ($\hat{\alpha}$, $\hat{\beta}$), the simple weighted estimator ($\tilde{\alpha}_1$, $\tilde{\beta}_1$), and the optimally weighted estimator ($\tilde{\alpha}_2$, $\tilde{\beta}_2$). The α stands for the viscosity effect and β is for the feeding type effect.

					Bias (in percentage)					
	$\frac{MSE(\tilde{\alpha}_1)}{MSE(\hat{\alpha})}$	$\frac{MSE(\tilde{\alpha}_2)}{MSE(\hat{\alpha})}$	$\frac{MSE(\tilde{\beta}_1)}{MSE(\hat{\beta})}$	$\frac{MSE(\tilde{\beta}_2)}{MSE(\hat{\beta})}$	$\hat{\alpha}$	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	$\hat{\beta}$	$\tilde{\beta}_1$	$\tilde{\beta}_2$
Case 1										
$\tau = 0.25$	0.97	1.00	1.00	1.01	0.08	0.04	0.08	-0.64	-0.59	-0.66
$\tau = 0.5$	0.98	0.99	1.01	1.00	0.16	0.15	0.17	-0.54	-0.54	-0.52
Case 2										
$\tau = 0.25$	0.98	0.97	0.90	0.88	0.21	0.20	0.16	-0.21	-0.17	-0.20
$\tau = 0.5$	0.99	0.96	0.98	0.95	-0.05	-0.05	-0.06	-0.39	-0.36	-0.34
Case 3										
$\tau = 0.25$	0.99	0.98	0.75	0.69	0.36	0.36	0.29	-0.45	-0.26	-0.32
$\tau = 0.5$	0.99	0.98	0.79	0.75	-0.24	-0.18	-0.19	-0.60	-0.56	-0.44
Case 4										
$\tau = 0.25$	1.00	0.99	0.57	0.38	0.56	0.54	0.40	-0.33	-0.17	-0.26
$\tau = 0.5$	1.00	1.00	0.61	0.48	-0.25	-0.23	-0.22	-0.43	-0.35	-0.23

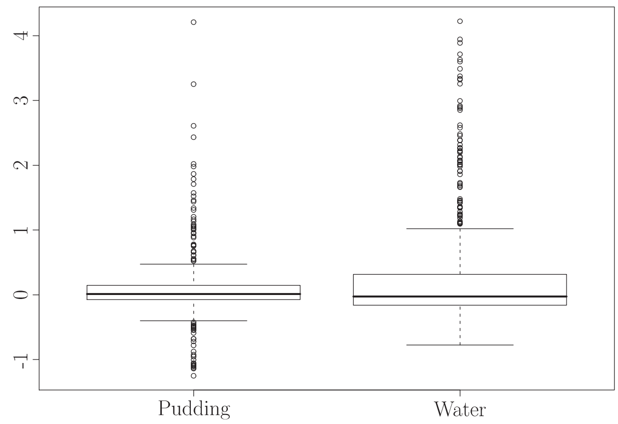

Figure 2. The boxplots of the estimated errors in the swallow study.

more efficient than $\tilde{\beta}_1$ when the intra-subject correlation was extremely high as, in Case 4.

4. Application to the Swallow Study

We now apply the proposed method to the swallow data set introduced in Section 2.1 to study the effects of viscosity and feeding type. For easy interpretation of the results, we work on the raw scale of the apnea duration data. We focus on three different quantile levels $\tau = 0.1$, 0.5 and 0.9. In swallow studies, the upper quantiles are of more clinical importance since longer apnea durations in the elderly are often due to pathological disorder, or to healthy age-related function changes.

Figure 2 shows the boxplots of the estimated errors for two viscosity conditions, obtained by fitting model (2.1) at median. The plot suggests that the apnea duration distribution is skewed to the right (this is true even for the log transformed data), and the water and pudding swallows exhibit very different variabilities. These should all be taken into account in the statistical inference.

For this data set, as the estimated intra-subject correlation is 0.2 for pudding and 0.4 for water, the $WQRS_1$ and $WQRS_2$ give similar results and we focus on $WQRS_1$ in the following analysis. Table 3 summarizes the point estimates and 95% confidence intervals of viscosity (pudding=1 and water=0), feeding type (self-fed=1 and examiner-fed=0), and viscosity-feeding interaction effects at three quantile levels. For comparative purposes, we also report the mean regression results from PROC MIXED assuming unequal variances for water and pudding swallows. The QRS and $WQRS_1$ yield similar significance results. However, $WQRS_1$ provides shorter confidence intervals than QRS for the feeding and interaction effects. For the viscosity effect, QRS yields shorter confidence intervals,

Table 3. The point estimates and 95% confidence intervals (CI) of viscosity, feeding type and viscosity-feeding interaction effects on apnea duration at three quantiles. The last row gives the mean regression results from PROC MIXED.

	Viscosity		Feeding		Interaction	
	Estimate	95%CI	Estimate	95%CI	Estimate	95%CI
$\tau = 0.1$						
$WQRS_1$	-0.09	(-0.18, -0.02)	-0.04	(-0.08, 0.02)	0.02	(-0.03, 0.07)
QRS	-0.09	(-0.14, -0.05)	-0.04	(-0.08, 0.04)	0.02	(-0.03, 0.08)
$\tau = 0.5$						
$WQRS_1$	-0.24	(-0.35, -0.14)	-0.04	(-0.09, 0.02)	0.04	(-0.04, 0.10)
QRS	-0.24	(-0.31, -0.17)	-0.04	(-0.11, 0.02)	0.04	(-0.05, 0.11)
$\tau = 0.9$						
$WQRS_1$	-2.53	(-3.27, -1.28)	-1.53	(-2.07, -1.11)	1.71	(1.16, 2.32)
QRS	-2.53	(-3.13, -1.94)	-1.53	(-2.09, -1.08)	1.71	(1.04, 2.31)
Mean	-0.61	(-0.76, -0.47)	-0.27	(-0.45, -0.10)	0.31	(0.11, 0.51)

Table 4. Estimated quantiles of apnea duration from different conditions.

	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
Water, examiner-fed	0.762	1.106	4.038
Water, self-fed	0.726	1.070	2.504
Pudding, examiner-fed	0.670	0.870	1.506
Pudding, self-fed	0.658	0.874	1.684

possibly due to the obvious over-rejection as observed in the simulation study. For this data set, the lower and upper tails of the apnea distribution behave differently. At the lower quantile ($\tau = 0.1$) and at the median, only viscosity is significant, while at the upper quantile ($\tau = 0.9$) both main and interaction effects are significant. More specifically, $WQRS_1$ yields a p-value of 0.0009 for the interaction effect at $\tau = 0.9$, suggesting significance even after multiple testing adjustment at three quantile levels. The exclusive use of the analysis on the mean differences from PROC MIXED would overlook the different covariate effects on the upper and lower quantiles of the apnea duration distribution.

Table 4 shows the estimated quantiles of apena duration from four different viscosity-feeding conditions. Generally speaking, normal senior people tend to have longer apnea duration when drinking water than when consuming pudding. At the upper quantile, the viscosity exhibits a larger effect on the apnea duration when subjects were fed by examiners than when they were consuming by themselves.

5. Discussion and Conclusions

In this paper, we proposed two weighted quantile rank score tests for mixed models with block-wise heteroscedasticity. The tests do not impose any parametric assumptions on the response distributions and thus are robust in performance. We showed that without appropriate weights to account for the heteroscedasticity, the rank score test QRS based on an asymptotic chi-square distribution is no longer valid. We also demonstrated that the δ adjustment employed in the proposed tests is crucial for dependent data. Simply ignoring the intra-subject correlation led to conservative tests in our simulation study, while under some other circumstances for testing the between-subject factors, it led to tests with inflated Type I errors. Solving the estimating equation as in $WQRS_2$ is computationally demanding, because it no longer corresponds to minimization of a convex objective function. This requires sufficient data for the asymptotic optimality of $WQRS_2$ to take effect. In finite samples, the asymptotically optimal weights that incorporate both correlation and heteroscedasticity provide marginal improvement over the simpler weights, except in situations where the intra-subject correlation is extremely high. Based on the finite-sample performance and the computational complexity, for clustered data with nonhomogeneous groups we would recommend $WQRS_1$ when the intra-subject correlation is moderate, and $WQRS_2$ for cases with extremely high intra-subject correlation ($\varrho \geq 0.90$).

Our proposed weights are proportional to the error densities evaluated at the target quantiles. In models such as (2.1), where we can assume common densities within blocks, the desired weights can be estimated consistently. The apnea duration study used in the paper is typical, as this type of data is encountered frequently in applications. Under more general heteroscedasticity, the error densities would be difficult to estimate, and we might consider using the wild bootstrap to obtain a reference distribution for the rank score test statistic, by treating each subject as a whole unit. More details on the wild bootstrap can be found in Liu (1988) and Wu (1986). The validity and performance of such bootstrap-based procedures need further investigation.

Acknowledgements

This work is supported by the National Science Foundation Award DMS-0706963. The author would like to thank Dr. Perlman for providing the swallow data, and two referees for insightful comments and suggestions on the first draft.

References

Agresti, A., Caffo, B. and Ohman, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Comput. Statist. Data Anal.* **47**, 639-653.

Cai, B. and Dunson, D.B. (2006). Bayesian covariance selection in generalized linear mixed models. *Biometrics* **62**, 446-457.

Davidian, M. and Giltinan, D. M. (1993). Some simple methods for estimating intra-individual variability in nonlinear mixed effects models. *Biometrics* **49**, 59-73.

Francq, C. and Tran, L. T. (2002). Nonparamatric estimation of density, regression and dependence coeffcients. *Nonparametric Statistics* **14**, 729-747.

Ghidey, W., Lesaffre, E. and Eilers, P. (2004). Smooth random effects distribution in a linear mixed model. *Biometrics* **60**, 945-953.

Hartford, A. and Dividian, M. (2000). Consequences of misspecifying assumptions in nonlinear mixed effects models. *Comput. Statist. Data Anal.* **34**, 139-164.

He, X., Zhu, Z. Y. and Fung, W. K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* **89**, 579-590.

Jara, A., Quintana, F. and Martín, E. S.(2008). Linear mixed models with skew-elliptical distributions: A Bayesian approach. *Comput. Statist. Data Anal.* **52**, 5033-5045.

Jung, S. (1996). Quasi-likelihood for median regression models. *J. Amer. Statist. Assoc.* **91**, 251-257.

Kizilkaya, K. and Tempelman, R. J. (2005). A general approach to mixed effects modeling of residual variances in generalized linear mixed models. *Genetics, Selection, Evolution* **37**, 31-56.

Koenker, R. (1994). Confidence intervals for regression quantiles. In *Proceedings of the 5th Prague Symposium on Asymptotic Statistics*, 349-359, Springer-Verlag, New York.

Koenker, R. (2005). *Quantile Regression*, Cambridge University Press, Cambridge.

Lin, X., Raz, J. and Harlow, S. D. (1997). Linear mixed models with heterogeneous within-cluster variances. *Biometrics* **53**, 910-923.

Liu, R. Y. (1988). Bootstrap procedures under some non-iid models. *Ann. Statist.* **16**, 1696-1708.

Perlman, A. L., He, X., Barkmeier, J. and Van Leer, E. (2005). Bolus location associated with videofluoroscopic and respirodeglutometric events. *Journal of Speech, Language, and Hearing Research* **48**, 21-33.

Wang, H. and Fygenson, M. (2009). Inference for censored quantile regression models in longitudinal studies. *Ann. Statist.* **37**, 756-781.

Wang, H. and He, X. (2007). Detecting differential expressions in GeneChip microarray studies: a quantile approach. *J. Amer. Statist. Assoc.* **102**, 104-112.

Wei, Y. and He, X. (2006). Conditional growth charts. *Ann. Statist.* **34**, 2069-2097.

Wu, C. F. J. (1986). Jackknife bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14**, 1261-1295.

Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* **57**, 795-802.

Zhou, T. and He, X. (2008). Three-step estimation in linear mixed models with skew-t distributions. *J. Statist. Plann. Inference* **138**, 1542-1555.

Department of Statistics, North Carolina State University, Raleigh NC 27695, U.S.A.

E-mail: wang@stat.ncsu.edu

(Received September 2007; accepted March 2008)