Journal of Modern Applied Statistical Methods

Volume 3 | Issue 1 Article 5

5-1-2004

A Rank-based Estimation Procedure For Linear Models With Clustered Data

Suzanne R. Dubnicka
Kansas State University, dubnicka@stat.ksu.edu

Journal of Modern Applied Statistical Methods
May, 2004, Vol. 3, No. 1, 39-48

1538 – 9472/04/$95.00

A Rank-based Estimation Procedure For Linear Models With Clustered Data

Suzanne R. Dubnicka
Department of Statistics
Kansas State University

A rank method is presented for estimating regression parameters in the linear model when observations are correlated. This correlation is accounted for by including a random effect term in the linear model. A method is proposed that makes few assumptions about the random effect and error distribution. The main goal of this article is to determine the distributions for which this method performs well relative to existing methods.

Key words: R-estimate, random effect, pseudo-sample

Introduction

Consider a situation in which individuals selected for study are not independent of one another. In particular, we consider the situation in which clusters of individuals are observed. These clusters may be families, siblings, littermates, classmates in school, etc. Whatever the origin of the cluster, we consider individuals to be in the same cluster if these individuals are members of a group which, due to this group membership, are more likely to give similar responses than individuals in different groups. Therefore, responses from individuals within a cluster are considered to be correlated while responses from individuals in different clusters are not.

To account for this correlation within clusters, we add a random effect term to the usual linear regression model and consider the following model:

$$\boldsymbol{Y}_i = \alpha \mathbf{1}_{n_i} + \boldsymbol{X}_i \boldsymbol{\beta} + b_i \mathbf{1}_{n_i} + \boldsymbol{e}_i, \quad i = 1,...,m, \tag{1}$$

where $\boldsymbol{Y}_i$ is a $n_i \times 1$ vector of responses for cluster i, $\boldsymbol{X}_i$ is a $n_i \times p$ matrix with j^{th} row, $\boldsymbol{x}_{ij}^T$, corresponding to the p covariates for observation j in cluster i, α is the common unknown intercept, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, b_i is the random effect for cluster i, and $\mathbf{1}_{n_i}$ is a vector of ones of length n_i. We assume that b_1, ..., b_m are iid continuous random variables, that $e_{11},...,e_{mn_m}$ are iid continuous random variables, and that the b_i and the e_{ij} but these assumptions will depend on the method used for predicting the random effects. These assumptions are discussed in Section 2.1. Thus, there are m clusters with n_i observations within each cluster $(i = 1,...,m)$, and the total sample size is given by $N = \sum_i n_i$.

Our main interest is to estimate the unknown parameters α and $\boldsymbol{\beta}$. Linear models and generalized linear models with random effects have been studied extensively in a variety of parametric and semiparametric settings in which specific distributions are assumed for the random effects, b_i, and/or the random errors, e_{ij}. For example, Laird and Ware (1982), Ware (1985), Lindstrom and Bates (1988), Schall (1991), Zeger and Rezaul (1991), Waclawiw and Liang (1993), and Chen (2001) all provide methods for fitting such models. In addition, other approaches which also account for correclation within clusters, such as GEE, have

Suzanne R. Dubnicka is an Assistant Professor of Statistics at Kansas State University, 108B Dickens Hall, Manhattan, KS, 66506. Email her at Dubnicka@stat.ksu.edu.

also been developed. For example, see Zeger and Liang (1988) and Lin and Carroll (2001).

In this article, we propose a method for estimating the unknown regression parameters which does not assume a specific distributional form for either b_i or e_{ij}. The proposed method uses rank methods to estimate $\boldsymbol{\beta}$ and pseudo-samples to predict the random effects b_i. Chen (2001) presents a similar method in which the regression parameters are estimated via rank methods but the random effects are assumed to be normally distributed and are predicted using the best linear unbiased predictor under normality. In using pseudo-samples to estimate the random effects, we do not assume a specific distributional form for these random effects. In addition, unlike Chen, we do not need to estimate the variance of the b_is or the e_{ij}s with each iteration.

The main purpose of this paper is to evaluate the performance of the proposed method, relative to some existing methods, for a variety of distributions for the random effects and random errors. The more theoretical aspects relating to the proposed method, including asymptotics, are the subject of another paper currently in review (Dubnicka 2004).

The method for estimating $\boldsymbol{\beta}$ proposed in this paper is an iterative procedure with two major components: the estimation of $\boldsymbol{\beta}$ given b_i and the prediction of b_i given $\boldsymbol{\beta}$. These two components are detailed in Methodology. In Simulations, we evaluate the proposed method and compare it to existing methods via computer simulations. We conclude with a summary of our findings.

Methodology

Consider the model given in (1). We estimate $\boldsymbol{\beta}$ and b_i using the following iterative steps until the convergence:

1. Estimate $\boldsymbol{\beta}$ as if the N subjects are independent by solving the usual rank estimating equations given below.
2. Predict the random effects, b_i, using a pseudo-sample approach.
3. Given the estimates of b_i, obtain a rank-based estimate $\boldsymbol{\beta}$ by solving (13).
4. Repeat steps 2 and 3 until convergence.

Steps 2 and 3 are detailed in the next two sections.

Prediction of the Random Effects

The random effects b_1, ..., b_m are predicted using pseudo-samples. Since we know only the Y_{ij}, the random effects b_i and the errors e_{ij} are not observable. However, we can use the information in the Y_{ij} to construct a sample of size m that, as $N \to \infty$, is asymptotically equivalent to the b_i. In particular, we follow the approach of Groggel, Wackerly and Rao (1988) who use pseudo-samples of random effects and random errors to conduct inference on the intraclass correlation in a one-way random effects model. They propose two methods for constructing pseudo-samples: one based on means and another based on medians. We modify their approach in order to predict the random effects in the linear model. The creation of such pseudo-samples requires only a few assumptions regarding the distributions of the random effects, b_i, and the random errors, e_{ij}. The particular assumptions depend on the method used to create the pseudo-samples and are discussed below.

The two methods for creating pseudo-samples proposed by Groggel, Wackerly, and Rao (1988) are the means method and the medians method. With a small adjustment, we can construct a pseudo-sample of the b_i using these methods. Let

$$U_{ij} = Y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{\beta}. \tag{2}$$

Then $U_{ij} = \_ + b_i + e_{ij}$ which is in the form of a one-way random effects model considered by Groggel, Wackerly, and Rao (1988).

Pseudo-samples based on means are given by

$$V_{ij} = U_{ij} - \overline{U}_{i\cdot} = e_{ij} - \bar{e}_{i\cdot} \tag{3}$$

$$W_i = \overline{U}_{i\cdot} - \overline{U}_{\cdot\cdot} = b_i + \bar{e}_{i\cdot} - \bar{b} - \bar{e}_{\cdot\cdot} \tag{4}$$

where $\overline{U}_{i\cdot} = n_i^{-1} \sum_j U_{ij}$, $\overline{U}_{\cdot\cdot} = N^{-1} \sum_i \sum_j U_{ij}$, $\bar{e}_{i\cdot} = n_i^{-1} \sum_j e_{ij}$, $\bar{e}_{\cdot\cdot} = N^{-1} \sum_i \sum_j e_{ij}$, and

$\bar{b} = m^{-1}\sum_i b_i$, If the random effects and the random errors distributions have mean 0, V_{ij} converges in distribution to e_{ij} and W_i converges in distributions to b_i (Dubnicka 2004; Groggel 1983).

Pseudo-samples based on medians are defined in an analogous manner. Let

$$V'_{ij} = U_{ij} - \hat{U}_i = e_{ij} - \hat{e}_i \tag{5}$$

$$W'_i = \hat{U}_i - \hat{U} = b_i - \hat{e}_i - \hat{b} \tag{6}$$

where $\hat{U}_i = med\{U_{i1},...,U_{in_i}\}$, $\hat{U} = med\{\hat{U}_1,...,\hat{U}_{m_i}\}$, $\hat{e}_i = med\{e_{i1},...,e_{in_i}\}$, and $\hat{b} = med\{b_1 + \hat{e}_1,...,b_m + \hat{e}_m\}$. If the random effects and the random errors distributions are bounded, V'_{ij} converges in distribution to e_{ij} and V'_{ij} converges in distribution to b_i (Dubnicka 2004; Groggel 1983). Note that for $n_i = 1$ or 2, $V_{ij} = V'_{ij}$.

Therefore, under general conditions, the W_i and $W'_i\,(i = 1,...,m)$ asymptotically equivalent to the true random effects b_i, $i = 1,...,m$. Thus, the W_i and W'_i represent pseudo-samples which predict the random effects b_i, $i = 1,\ldots,m$. Throughout the remainder of this paper $\hat{b}_i$ represents the predicted value of b_i based on one of these two methods; that is, $\hat{b}_i = W_i$ or W'_i. Note that in creating pseudo-samples to predict b_i, we can also create pseudo-samples which predict the e_{ij}. These predicted errors are provided by the V_{ij} and the V'_{ij}. However, we do not need these pseudo-samples in our iterative estimation procedure.

Estimation of Regression Parameters

In the proposed iterative procedure, the regression parameters are estimated using rank methods. Rank-based regression requires only very general assumptions on the underlying error distribution. There are several rank-based regression methods from which we can choose. We present several variations below but focus on the most basic approach.

Consider the linear model

$$Y_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$$

where the e_i are iid random variables. Then the most common rank-based estimate of $\boldsymbol{\beta}$, introduced by Jaeckel (1972), is found by minimizing the dispersion function

$$D^*(\boldsymbol{\beta}) = \sum_{i=1}^{n} R\left(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)\left(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right) \tag{7}$$

where $R\left(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)$ is the rank of $Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}$ among $Y_1 - \boldsymbol{x}_1^T\boldsymbol{\beta},...,Y_n - \boldsymbol{x}_n^T\boldsymbol{\beta}$. Estimates of $\boldsymbol{\beta}$ found by minimizing (7) are called R-estimates.

One generalization of (7) is given by

$$D_a(\boldsymbol{\beta}) = \sum_{i=1}^{n} a\left[R\left(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right]\left(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right) \tag{8}$$

where $a(1) \le a(2) \le \cdots a(n)$ is a set of scores generated by $a(i) = \phi\left[i/(n+1)\right]$ for some nondecreasing score function $\phi(u)$ which is defined on (0,1) and satisfies $\int \phi(u)du = 0$ and $\int \phi^2(u)du = 1$. Two commonly used score functions are Wilcoxon scores and sign scores given by $\phi_W(u) = \sqrt{12}(u - 1/2)$ and $\phi_S(u) = sgn(u - 1/2)$, respectively.

Using Wilcoxon scores produces a dispersion function which is equivalent to (7) and which will produce the usual R-estimate for $\boldsymbol{\beta}$. Sign scores will produce the L_1 estimate of $\boldsymbol{\beta}$. Other score functions which are optimal for specific error distributions have also been proposed. In addition, there are score functions which may be more appropriate for asymmetric errors (Hettmansperger and McKean 1998).

Note that minimizing $D^*(\boldsymbol{\beta})$ is equivalent to minimizing $D(\boldsymbol{\beta}) = \sum_{i<j}\sum \left|\left(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right) - \left(Y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}\right)\right|$. (9)

That is, minimizing $D(\boldsymbol{\beta})$ will also provide the R-estimate for $\boldsymbol{\beta}$. A related approach, introduced by Sievers (1983) and further developed by Naranjo and Hettmansperger (1994), estimates $\boldsymbol{\beta}$ by minimizing

$$D_b(\boldsymbol{\beta}) = \sum\sum_{i<j} \left|\left(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right) - \left(Y_j - \boldsymbol{x}_j^T\boldsymbol{\beta}\right)\right| \quad (10)$$

where $b_{ij} = b(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Properly chosen weights b_{ij} will produce estimates of $\boldsymbol{\beta}$ with a bounded influence function and high breakdown. The bounded-influence estimate, however, tends to be less efficient than the usual R-estimate. Estimates produced by minimizing $D_b(\boldsymbol{\beta})$ are called generalized rank estimates, or GR-estimates.

The proposed iterative procedure can be performed using R-estimation, general score R-estimation, or GR-estimation. In practice, one would carefully evaluate the particular application to determine which is most appropriate. For the remainder of this paper, however, we will use the more common R-estimates of $\boldsymbol{\beta}$.

Return now to our model (1) which includes the random effect. Let $Y_{ij}^* = Y_{ij} - b_i$. Then, given the random effects, we can estimate the regression parameters using the usual rank estimating equations where the Y_{ij} s are replaced by Y_{ij}^* s.

To simplify notation, let $\{Y_1^*, Y_2^*, ..., Y_N^*\}$ represent $\{Y_{11}^*, ..., Y_{1n_1}^*, Y_{21}^*, ..., Y_{2n_2}^*, ..., Y_{m1}^*, ..., Y_{mn_m}^*\}$. The vectors of covariates corresponding to these rsponses can be written in an analogous manner. Then $\hat{\boldsymbol{\beta}}_R$ is the estimators of β which minimizes

$$D(\boldsymbol{\beta}|\boldsymbol{b}) = \sum\sum_{l<k} \left|\left(Y_l^* - \boldsymbol{x}_l^T\boldsymbol{\beta}\right) - \left(Y_k^* - \boldsymbol{x}_k^T\boldsymbol{\beta}\right)\right|. \quad (11)$$

The gradient of $D(\boldsymbol{\beta}|\boldsymbol{b})$ is given by

$$S(\boldsymbol{\beta}|\boldsymbol{b}) = -\nabla D(\boldsymbol{\beta}|\boldsymbol{b}) = \sum\sum_{l<k} (\boldsymbol{x}_l - \boldsymbol{x}_k)\, sgn \left[\left(Y_l^* - \boldsymbol{x}_l^T\boldsymbol{\beta}\right) - \left(Y_k^* - \boldsymbol{x}_k^T\boldsymbol{\beta}\right)\right] \quad (12)$$

As $D(\boldsymbol{\beta}|\boldsymbol{b})$ is a piecewise linear, continuous, convex function, minimizing $D(\boldsymbol{\beta}|\boldsymbol{b})$ is equivalent to solving

$$S(\boldsymbol{\beta}|\boldsymbol{b}) \doteq \mathbf{0}. \quad (13)$$

Note that it is unlikely $S(\boldsymbol{\beta}|\boldsymbol{b})$ will equal **0** for any value of $\boldsymbol{\beta}$. In the case of one covariate, $S(\beta|\boldsymbol{b})$ is a nondecreasing step function of β which steps down at each sample slope. There may be an interval of solutions $S(\beta|\boldsymbol{b}) = 0$ or $S(\beta|\boldsymbol{b})$ may "step across" the horizontal axis. We let $\hat{\boldsymbol{\beta}}_R$ denote this rank estimate of $\boldsymbol{\beta}$ in either case.

Once an estimate for $\boldsymbol{\beta}$ has been obtained, α can be estimated by solving

$$S_1(\alpha|\boldsymbol{\beta}, \boldsymbol{b}) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} sgn\left(Y_{ij} - \boldsymbol{x}_{ij}^T\hat{\boldsymbol{\beta}}_R - \hat{b}_i\right) \doteq 0 \quad (14)$$

where $\hat{\boldsymbol{\beta}}_R$ is the estimate of $\boldsymbol{\beta}$ obtained from solving (13) and $\hat{b}_i$ is the predicted value of b_i using one of the pseudo-sample methods of the previous section. The solution to equation (14) is simply the median of the residuals $Y_{ij} - \boldsymbol{x}_{ij}^T\hat{\boldsymbol{\beta}}_R - \hat{b}_i$. That is,

$$\hat{\alpha}_R = median\left\{Y_{ij} - \boldsymbol{x}_{ij}^T\hat{\beta}_R - \hat{b}_i\right\} \quad (15)$$

Simulations

Simulations were conducted to evaluate the performance of our proposed method. These simulations were performed with the intent of answering two questions:

1. How large must m and the n_i be to produce "good" estimates of α and $\boldsymbol{\beta}$?

2. How does this method perform as the random effects distribution and the random error distribution vary?

Recall that the pseudo-samples are only asymptotically equivalent to the true random effects. For small samples, there is some concern that this method will not produce estimates of β and α which are reasonably on target. In this first simulation study, we focus on the cluster sizes and number of clusters rather than the distributions of the random effects and random errors. Therefore, with m clusters of n subjects per cluster, a single covariate $x \sim$ lognormal(2, 0.5^2), and $(\alpha,\beta) = (2, 2)$, 1000 samples were generated in which the random effects and the random errors were both normally distributed: $b_i \sim N(0, 0.5^2)$ and $e_{ij} \sim N(0, 0.4^2)$. Note that, for simplicity, we have chosen all of the clusters sizes to be the same $(n_1 = ... = n_m)$.

For comparison, estimates of α and β were also obtained using maximum likelihood (ML) and restricted maximum likelihood (REML) since these methods are included in most existing statistical software.

Results of this first simulation study appear in Table 1. For each method, the means of the 1000 estimates are given with the standard deviations of the 1000 estimates below the estimates in parentheses.

Table 1: Parameter Estimates for Various m and n, $b_i \sim N(0,0.5^2)$, $e_{ij} \sim N(0,0.4^2)$

	Mean Method	Median Method	ML	REML
(m,n)	$\hat{\alpha},\hat{\beta}$	$\hat{\alpha},\hat{\beta}$	$\hat{\alpha},\hat{\beta}$	$\hat{\alpha},\hat{\beta}$
	(StDev)	(StDev)	(StDev)	(StDev)
(5,2)	2.00902,1.99751	2.015,1.99747	2.00592,1.99819	2.003934,1.99783
	(0.53457,0.0625)	(0.54577,0.06239)	(0.48297,0.05278)	(0.47995,0.05272)
(5,5)	1.99484,2.0008	2.00185,2.00077	1.99704,2.00077	1.99697,2.0078
	(0.29678.0.02253)	(0.31546,0.02366)	(0.28749,0.02159)	(0.28703,0.02155)
(5,8)	2.00099,2.00031	2.00178,2.00029	2.00078,2.00019	2.00079,2.00019
	(0.27815,0.01747)	(0.28995,0.01755)	(0.27389,0.01702)	(0.27386,0.01699)
(15,2)	1.98592,2.00158	1.98717,2.00158	1.98907,2.00135	1.98894,2.00136
	(0.26826,0.0278)	(0.28243,0.0278)	(0.24222,0.0239)	(0.24184,0.02386)
(15,5)	1.99979,2.00004	2.00222,2.00002	1.99953,2.00008	1.99954,2.00008
	(0.17499,0.01245)	(0.18033,0.01283)	(0.16884,0.01215)	(0.16884,0.01215)
(15,8)	2.00682,1.99933	2.0056,1.99942	2.00641,1.99942	2.00643,1.99942
	(0.159,0.00919)	(0.16353,0.00925)	(0.15728,0.00901)	(0.15724,0.009)
(30,2)	1.99458,2.00071	1.99312,2.00071	1.99318,2.00088	1.9932,2.00088
	(0.1843,0.01814)	(0.19196,0.01814)	(0.16987,0.01601)	(0.16981,0.016)
(30,5)	1.99352,2.00001	1.99409,2.00012	1.99282,2.0007	1.99284,2.00007
	(0.11875,0.00846)	(0.12621,0.00875)	(0.1152,0.00813)	(0.11519,0.00813)
(30,8)	2.00049,2.00001	2.0003,2	2.00023,2.00005	2.00023,2.00005
	(0.11341,0.00648)	(0.11765,0.00661)	(0.11055,0.00628)	(0.1055,0.00628)

Note that the estimates of and obtained from the proposed iterative method using either mean or median pseudo-samples seem to be reasonably unbiased even for small *m* and *n*; see Table 1. As one would expect, when both the random effects and the random errors are normally distributed the standard deviations of the estimates obtained through maximum likelihood and REML are smaller. However, the standard deviations of the estimates obtained through the proposed iterative method are not much larger.

Although the proposed method, using mean or median pseudo-samples, provides estimates which are reasonably on target for small *m* and *n*, the procedure failed to converge for some samples regardless of the pseudo-sample method used. Table 2 shows the percentage of times that the mean method and the median method converge for each of the combinations of *m* and *n* in the first simulation study. Upon closer investigation, we found that for some of the samples the estimates of continued to increase (or decrease) as more iterations were completed. For some of the samples, however, the estimates of seemed to "bounce" between two values. This happened more frequently when both *m* and *n* were small

Table 2: Convergence Percentage for Various *m* and *n*, $b_i \sim N(0,0.5^2)$, $e_{ij} \sim N(0,0.4^2)$

(*m*,*n*)	Mean Method	Median Method
(5,2)	99.1%	99.1%
(5,5)	96.7%	99.8%
(5,8)	99.0%	93.0%
(15,2)	99.7%	99.7%
(15,5)	98.2%	100%
(15,8)	99.1%	96.5%
(30,2)	99.8%	99.8%
(30,5)	98.8%	100%
(30,8)	99.7%	98.4%

The remaining simulations were designed to help answer the second question. That is, we wanted to determine the distributions under which the proposed method is superior to the existing methods considered. In these simulations, a variety of distributions for both the random effects and the random error were used. Table 3 gives the abbreviations for the particular distributions chosen for these simulations.

The simulations conducted are divided into three cases: (1) the random effects distribution is normal and the error distribution varies, (2) the error distribution is normal and the random effects distribution varies, and (3) both distributions are nonnormal but from the same family of distributions. As with the first simulation, 1000 random samples were generated with a single covariate $x \sim$ lognormal(2, 0.5^2) and $(\alpha,\beta) = (2,2)$. Furthermore, each sample consists of $m = 50$ clusters of $n = 3$ subjects per cluster. For comparison, estimates of α and β were also obtained using Chen's method and restricted maximum likelihood (REML). Recall that Chen's method differs from the proposed method in that the random effect is assumed to be normally distributed. Maximum likelihood estimates were also computed but there were almost identical to the REML estimates.

Table 4 shows the simulation results for normally distributed random effects. Since Chen's method assumes normality for the b_i but does not assume a specific distribution for the e_{ij}, one would expect Chen's method to perform better than the proposed methods and REML. To some extent, the simulations support this theory. When the errors follow a contaminated normal or double exponential distribution, the standard deviations of the β estimates using Chen's method are smaller than those of the other methods. When the errors follow a Cauchy distribution, the standard deviation of the β estimates based on the proposed method with median pseudo-samples is smaller than that of the other approaches. Notice, however, that the standard deviation of the α estimates is smaller for the median method than the other methods. In particular, the estimates of α using the mean method and REML were highly variable in the case of Cauchy errors.

Table 3: Distributions used in Simulations

Abbreviation	Name of Distribution	Description
CN1	Contaminated Normal	0.9 $N(0,0.4^2)$ + 0.1 $N(0,1.2^2)$
CN2	Contaminated Normal	0.9 $N(0,0.3^2)$ + 0.1 $N(0,0.9^2)$
DE1	Double Exponential	$\frac{1}{2}\lambda \exp(-\lambda\|x\|), \lambda = 2.5$
DE2	Double Exponential	$\frac{1}{2}\lambda \exp(-\lambda\|x\|), \lambda = 3$
C1	Cauchy	0.16 *Cauchy*(0.1)
C2	Cauchy	0.12 *Cauchy*(0,1)
U1	Uniform	*Uniform*(-1.2,1.2)
U2	Uniform	*Uniform*(-0.9,0.9)

Table 4: Parameter Estimates for $m = 50$, $n = 3$, $b_i \sim N(0, 0.3^2)$

	Mean Method	Median Method	Chen	REML
Error Distribution	$\hat{\alpha}, \hat{\beta}$ (StDev)	$\hat{\alpha}, \hat{\beta}$ (StDev)	$\hat{\alpha}, \hat{\beta}$ (StDev)	$\hat{\alpha}, \hat{\beta}$ (StDev)
*CN*1	1.99406,2.00087 (0.11152,0.01085	1.9978,2.00054 (0.10595,0.01008)	1.99548,2.00074 (0.12488,0.00974)	1.99809,2.00047 (0.10788,0.01052)
*DE*1	1.99851,2.00024 (0.11973,0.01232)	1.99718,2.00015 (0.11261,0.01101)	1.99368,2.00024 (0.12384,0.01081)	1.99763,2.00033 (0.11517,0.01183)
*C*1	2.12036,2.00034 (4.5088,0.01205	1.99724,2.00004 (0.09558,0.00919)	1.99743,1.99999 (0.10268.0.00988)	2.1836,1.99199 (7.35446,0.68516)
*U*1	2.00417,1.99981 (0.16375,0.01667)	2.00281,1.99973 0.15796,0.0147)	2.0028,1.99971 (0.18557,0.01459)	2.0028,1.99974 (0.13293,0.01367)

Table 5: Convergence Percentage $m = 50$, $n = 3$, $b_i \sim N(0, 0.3^2)$

Error Distribution	Mean Method	Median Method	Chen
*CN*1	98.5%	100%	96.2%
*DE*1	98.9%	100%	98.8%
*C*1	99.2%	100%	66.3%
*U*1	97.8%	100%	92.1%

The major disadvantage of Chen's method is this situation is that it does not always converge. This is also true, to a lesser extent, for the proposed method with mean pseudo-samples. Table 5 gives the convergence percentage of the three rank-based methods for the simulations in Table 4. Notice that the median method always converged. The mean method and Chen's method converged most of the time when the error distribution was a contaminated normal, double-exponential, or uniform. The mean method also converged most of the time when the error distribution is Cauchy but Chen's method had difficulty converging in this case. Chen (2001) also notes this problem. The main source of the problem is that the Chen's method requires the estimation of the error variance (and the random effect variance) at each iteration, and convergence of the algorithm depends on the convergence of the error variance. In distributions for which the variance is undefined, convergence problems will exist for Chen's method.

Table 6 gives the results for cases in which the errors are normally distributed but the distribution of the random effects is non-normal. In addition, Table 7 gives the convergence percentages of the rank-based methods for these simulations. For the four situations considered, the standard deviations of the REML estimates of β are the smallest. This seems to imply that REML is a relatively efficient method for estimating β even when the random effect distribution is non-normal. Notice that the standard deviations of the median method β estimates are the largest of the four methods but they are not much larger than the REML standard deviations. Also, REML estimates of α also tend to be more precise (smallest standard deviation of the α estimates) except when the random effects distribution is Cauchy.

When the random effects follow a Cauchy distribution the median method provides the most precise estimate of α.

As in the previous simulations, Chen's method did not converge for all samples. In fact, when the error distribution was normal and the random effects distribution was Cauchy, Chen's method only converged half of the time. Notice again that the proposed method with median pseudo-samples always converged, and the proposed method with mean pseudo-samples converged most of the time.

Finally, we consider situations in which neither the error distribution nor the random effects distribution are non-normal. For each situation, the error and random effects distributions are from the same family of distributions. The results appear in Tables 8 and 9. In these situations, there is no clear "winner" with respect to the estimation of β. Under the contaminated normal distributions and double exponential distributions, Chen's β estimates have the smallest standard deviations. Under Cauchy distributions and uniform distributions, REML estimates of β are less variables. However, the proposed method with median pseudo-samples provided the most precise estimates of α under the distributions considered. Again note that the median method converged for all samples while the mean method converged most of the time and Chen's method converged most of the time except under the Cauchy distributions. Under Cauchy distributions, Chen's method only converged half of the time.

Conclusion

The paper introduced a new rank-based method for parameter estimation in linear model with a random effect term. Such a model is useful in accounting for the correlation between subjects that are correlated, as is the case when clusters of subjects are observed. The proposed method uses rank-based regression to estimate the parameters of the linear model and pseudo-samples to predict the random effects. As a result the proposed method requires few assumptions regarding the underlying distributions of the errors and the random effects.

Table 6: Parameter Estimates for $m = 50$, $n = 3$, $e_{ij} \sim N(0, 0.4^2)$

	Mean Method	Median Method	Chen	REML
Error Distribution	$\hat{\alpha}, \hat{\beta}$ (StDev)	$\hat{\alpha}, \hat{\beta}$ (StDev)	$\hat{\alpha}, \hat{\beta}$ (StDev)	$\hat{\alpha}, \hat{\beta}$ (StDev)
*CN*2	2.00057,2.00014	1.99799,2.00033	2.00262,2.00023	1.99829,2.00035
	(0.10673,0.00946)	(0.11103,0.00984)	(0.22225,0.00911)	(0.10245,0.00883)
*DE*2	1.99463,2.00048	1.99765,2.00025	2.00305,2.00039	1.99458,2.00037
	(0.10983,0.00934)	(0.11096,0.01001)	(0.28775,0.00898)	(0.10303,0.00859)
*C*2	2.0442,2.00018	1.99758,1.99998	1.99871,2.00012	2.04528,1.99997
	(4.77186,0.00953)	(0.10331,0.00993)	(0.83112,0.00947)	(4.77105,0.00919)
*U*2	2.00176,1.99979	1.99978,1.99998	1.98411,1.99978	2.00285,1.99982
	(0.10811,0.00911)	(0.13954,0.01103)	(0.32091,0.00884)	(0.10519,0.00871)

Table 7: Convergence Percentage $m = 50$, $n = 3$, $b_i \sim N(0, 0.4^2)$

Error Distribution	Mean Method	Median Method	Chen
*CN*2	98.3%	100%	89.8%
*DE*2	98.6%	100%	87.4%
*C*2	99.3%	100%	47.9%
*U*2	98.7%	100%	83.8%

Table 8: Parameter Estimates for $m = 50$, $n = 3$

	Mean Method	Median Method	Chen	REML
Distributions	$\hat{\alpha}, \hat{\beta}$ (StDev)	$\hat{\alpha}, \hat{\beta}$ (StDev)	$\hat{\alpha}, \hat{\beta}$ (StDev)	$\hat{\alpha}, \hat{\beta}$ (StDev)
$b \sim CN2$, $e \sim CN1$	2.00216.2.00001	2.0031, 1.99981	2.00374, 1.99999	2.00268,1.99983
	(0.11928, 0.01108)	(0.11469,0.01081)	(0.17199,0.01014)	(0.1196,0.01118)
$b \sim DE2$, $e \sim DE1$	1.99352,2.00026	1.9962,2.0001	1.99604,2.00036	1.99336,2.0002
	(0.13154,0.01244)	(0.1285,0.01238)	(0.2035,0.0116)	(0.13173,0.01256)
$b \sim C2$, $e \sim C1$	1.07711,1.99969	2.00552,1.9996	1.99952,1.99971	1.07733,1.99966
	(27.68038,0.00957)	(0.09885,0.00989)	(0.82259,0.00941)	(27.68138,0.00913)
$b \sim U2$, $e \sim U1$	2.01073,1.99879	2.00585,1.99932	2.02594,1.99903	2.00635,1.99907
	(0.17553,0.01733)	(0.17511,0.01654)	(0.33254,0.01616)	(0.15331,0.01504)

Table 9: Convergence Percentage $m = 50$, $n = 3$, $b_i \sim N(0, 0.4^2)$

Distributions	Mean Method	Median Method	Chen
$b \sim CN2,\ e \sim CN1$	98.6%	100%	98.0%
$b \sim DE2,\ e \sim DE1$	98.3%	100%	98.3%
$b \sim C2,\ e \sim C1$	99.0%	100%	48.5%
$b \sim U2,\ e \sim U1$	98.4%	100%	96.6%

Results from the simulation studies showed that REML often provided estimates $\boldsymbol{\beta}$ which were less variable than those of other methods. If the goal of a study is to see how the response changes as the predictors change, then REML might provide the best means for assessing this. However, if the goal is to predict a response for certain values of the predictors, REML may provide inaccurate predictions under some distributions since the REML estimate of α can be highly variable. The three rank-based methods considered (mean pseudo-samples, median pseudo-samples, and Chen) all produce estimates of $\boldsymbol{\beta}$ with comparable precision to REML. Only the median method seems to provide consistently precise estimates of α under all distributions considered. In general, the proposed method with median pseudo-samples is robust to the underlying distribution of the random effects and errors as it is relatively efficient for all distributions considered. Therefore, if prediction of the goal of study, the proposed method with median pseudo-samples is recommended.

As a final note, it may be possible for the proposed method to perform better that REML with properly chosen scores in (8), but this has not yet been explored.

References

Chen, C. C. (2001). Rank estimating equations for random effects models. *Statistics and Probability Letters*, *54*, 5-12.

Dubnicka, S. R. (2004). Rank-based Regression with Clustered Data, in review.

Groggel, D. J. (1983). Asymptotic Nonparametric Confidence Intervals for the Ratio of Scale Parameters in Balanced One-Way Random Effects Models. Ph.D. thesis, University of Florida.

Groggel, D. J., Wackerly, D., & Rao, P. (1988). Nonparametric estimation in one-way random effects models. *Communications in Statistics - Simulation and Computation 17*, 887-903.

Hettmansperger, T. P., & McKean , J. W. (1998). *Robust Nonparametric Statistical Methods*. Arnold.

Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Annals of Mathematical Statistics*, *43*, 1449-1458.

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics*, *38*, 962-974.

Lin, X., & Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, 96, 1045-1056.

Lindstrom, M. J., & Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated measures. *Journal of the American Statistical Association*, *83*, 1014-1022.

Naranjo, J. D., & Hettmansperger, T. P. (1994). Bounded influence rank regression. *Journal of the Royal Statistical Society, Series B*, 56, 209-220.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, *78*, 719-727.

Sievers, G. L. (1983). A weighted dispersion function for estimation in linear models. *Communications in Statistics - Theory and Methods*, *12*, 1161-1179.

Waclawiw, M. A., & Liang, K. Y. (1993). Prediction of random effects in the generalized linear model. *Journal of the American Statistical Association*, *88*, 171-178.

Ware, J. H. (1985). Linear models for the analysis of longitudinal data. *The American Statistician*, *39*, 95-101.

Zeger, S. L., & Liang, K. Y. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, *44*, 1049-1060.

Zeger, S. L., & Rezaul, K. M. (1991). Generalized linear models with random effects: a Gibbs sampling approach. *Journal of the American Statistical Association*, *86*, 79-86.