Rank-Based Estimation and Associated Inferences for Linear Models with Cluster Correlated Errors

John D. Kloke  Joseph W. McKean
Pomona College  Western Michigan University
M. Mushfiqur Rashid
FDA

Abstract

R estimators based on the joint ranks (JR) of all the residuals have been developed over the last twenty years for fitting linear models with independently distributed errors. In this paper, we extend these estimators to estimating the fixed effects in a linear model with cluster correlated continuous error distributions for general score functions. We discuss the asymptotic theory of the estimators and standard errors of the estimators. For the related mixed model with a single random effect, we discuss robust estimators of the variance components. These are used to obtain Studentized residuals for the JR fit. A real example is discussed which illustrates the efficiency of the JR analysis over the traditional analysis and the efficiency of a prudent choice of a score function. Simulation studies over situations similar to the example confirm the validity and efficiency of the analysis.
KEY WORDS: Compound symmetry; Joint rankings, Mixed models; Nonparametric; Rank regression scores; Robust; Wilcoxon procedures.

1 Introduction

In this article we develop R estimators of the fixed effects in an experiment done over clusters, blocks, groups, or subjects. Examples include repeated measure designs, split plot designs, multi-center clinical trials, randomized block designs, and two-stage cluster samples (Rao et al., 1993 and Rashid and Nandram, 1998). We will use the terms clusters and blocks interchangeably. The random errors are independent between blocks but are dependent within blocks. The design matrix for the fixed effects is general and can include classification factors, interactions, predictors and covariates. Such designs occur frequently in practice. These estimators are based on the joint ranking of all the residuals, so we refer to them as the JR estimators. The presentation is for an arbitrary score function; hence, the estimate can be optimized if information is available on the underlying distribution of the errors.

In Section 2, we discuss the geometry of the JR estimators and present their asymptotic theory under the assumption that the univariate marginal distributions within a block are the same. This is more general than the usual assumption of exchangeability with correlated clusters models, see Rao et al. (1993). Our assumption commonly occurs, for example, in repeated measure designs where the errors follow a stationary time series model. The asymptotic theory of the estimators is similar to that of R estimators for the usual linear model where all the random errors are independent of one another; hence, only a brief sketch is given. We discuss standard errors of the estimates and the associated Wald-type test of linear hypotheses. These standard errors include terms based on the underlying dependency. For brevity,
we only discuss highly efficient R estimators but the theory extends to the high breakdown estimators with breakdown of up to 50% discussed in Chang et al. (1999).

In Section 3, we discuss our estimators for the mixed model where block is taken as a random effect. In this case, we obtain simple robust estimates of the variance components of the model. As with all fitting of models, a residual analysis for model checking and determination of potential outliers is necessary. We present robust Studentized residuals to conduct this model checking. These residuals correct for both factor space and the underlying covariance structure of the model and offer a simple benchmark for declaration of potential outliers.

In Section 4, we examine a real data set that contains several outliers. The traditional least squares (LS) analysis (maximum likelihood as though multivariate normality is true) is sensitive to these outliers while our analysis is robust. In practice, the analyses would lead to different interpretations. This is a small data set, so we offer a simulation study which confirms the validity of our analysis (robustness of confidence level) and, for alternative cases with heavy-tailed error distributions, its robustness of efficiency over the LS procedure. Due to the quantity of outliers in this example, score functions less sensitive to outliers than the Wilcoxon scores might be more suitable. So, along with the Wilcoxon analysis, we present an analysis based on a Winsorized score function. For confirmation, we present the results of a simulation study comparing relative empirical efficiencies of the Wilcoxon and the Winsorized analyses over situations similar to the data of this example. In Section 5, we summarize the ARE and simulation comparisons that we made between JR and LS estimators and, further, between JR and a Friedman type of estimator.

There has been some work done on Wilcoxon procedures in location models under dependent error structure. Gastwirth and Rubin (1975) investigated the efficiency of the Hodges-Lehmann estimator in the one-sample location model when the obser-
observations are serially dependent. They showed that the efficiency increases over the
independent error case for a positive type correlation structure. For the same model,
Portnoy (1977) investigated M estimators for certain moving average schemes. Koul
(1977) consider R estimators for a simple linear model with stationary strongly mixing
time series errors. Hollander et al. (1974) investigated the Wilcoxon test statistic
in a two-sample problem which allowed for some of data to be paired between the
samples; i.e., a cluster of length two in our setting. They showed that the usual two-
sample Wilcoxon test statistic is conservative under a positive-type of correlation.
Our model, though, is a linear model with dependence within clusters and independ-
dence between clusters. In a work involving a clustered correlated model, Gerard
and Schucany (2007) considered the sign test. Their data were binary and, hence,
only testing was considered. In the case that the appropriate model had been linear
and the observations continuous, our proposed analysis with sign scores extends their
analysis to the fitting and inference of the linear model.

Our concern is the estimation of the fixed effects for models with dependent error
structure. If the residual analysis indicates that the model seems reasonable, then
test of hypotheses can be performed by Wald-type tests. In the literature there are
rank-based procedures for tests of hypotheses concerning the fixed effects in a mixed
effects model; see, for example, Akritas and Arnold (1994), Akritas and Brunner
(1997) and Brunner et al. (2002) for discussion. These procedures can also be used to
test hypotheses, although not for estimation of the parametric effects and standard
2 Robust Estimates of Fixed Effects

Consider an experiment done over \(m\) blocks (clusters), where block \(k\) has \(n_k\) observations. Within block \(k\), let \(Y_k, X_k,\) and \(e_k\) denote respectively the \(n_k \times 1\) vector of responses, the \(n_k \times p\) design matrix and the \(n_k \times 1\) vector of errors. Let \(1_{n_k}\) denote a vector of \(n_k\) ones. Then the model for \(Y_k\) is

\[
Y_k = \alpha 1_{n_k} + X_k \beta + e_k, \quad k = 1, \ldots, m,
\]

(2.1)

where \(\beta\) is the vector of regression coefficients and \(\alpha\) is the intercept parameter. Alternately, the model can be written as

\[
Y = 1_N \alpha + X \beta + e, \quad N = \sum_{k=1}^m n_k
\]

denotes the total sample size, \(Y = (Y'_1, \ldots, Y'_m)',\) \(X = (X'_1, \ldots, X'_m)',\) and \(e = (e'_1, \ldots, e'_m)'.\) Because an intercept parameter is in the model, we can assume that \(X\) is centered and that the true median of \(\varepsilon_{kj}\) is zero. Since we can always reparameterize, assume that \(X\) has full column rank.

For the theory discussed in Section 2.2, certain conditions are needed. Assume that the random vectors \(e_1, e_2, \ldots, e_m\) are independent. Assume that the univariate marginal distributions of \(e_k\) are continuous and are the same for all \(k\) and let \(F(x)\) and \(f(x)\) denote this common distribution function and density function. Assume that \(f(x)\) is absolutely continuous and that the usual regularity (likelihood) conditions hold; see, for example, Section 6.5 of Lehmann and Casella (1998).

For the design matrix \(X\), assume that Huber’s condition holds; i.e., the leverage values (diagonal entries of the projection matrix) get uniformly small as \(N \to \infty\). As with traditional theory, (see, eg, Liang and Zeger, 1986), assume the number of clusters go to \(\infty\) and that \(n_k \leq M\), for all \(k\), for some constant \(M\).
2.1 Geometry of Estimation

Because the design matrix is centered, the ordinary least squares (LS) estimator of $\beta$ is given by $\hat{\beta}_{LS} = \text{Argmin} \| Y - X\beta \|^2$, where $\| \cdot \|^2$ is the square of the usual Euclidean norm in $\mathbb{R}^N$. This, of course, is the same estimator as in the independent error case.

Replace the Euclidean norm by the pseudo-norm $\| \cdot \|_\varphi = \sum_{t=1}^N a[R(v_t)]v_t$, $v \in \mathbb{R}^N$, where $R(v_t)$ denotes the rank of $v_t$ among $v_1, \ldots, v_N$ and the scores are generated as $a[t] = \varphi[t/(N + 1)]$ for $\varphi(u)$ a nondecreasing bounded square-integrable function defined on $(0, 1)$ such that $\sum_t a[t] = 0$. Without loss of generality, standardize the score function so that $\int_0^1 \varphi(u) \, du = 0$ and $\int_0^1 \varphi^2(u) \, du = 1$. Popular scores include the Wilcoxon ($\varphi[u] = \sqrt{12} [u - (1/2)]$) and sign scores ($\varphi[u] = \text{sgn} [u - (1/2)]$).

The ordinary rank-based (JR) estimator of $\beta$ is

$$\hat{\beta}_\varphi = \text{Argmin} \| Y - X\beta \|_\varphi. \quad (2.2)$$

In the independent error case, these are the regression estimators proposed by Jurečková (1971) and Jaeckel (1972). Equivalently, $\hat{\beta}_\varphi$ is a solution to $S_X(\beta) = 0$, where $S_X(\beta) = -\nabla_\beta \| Y - X\beta \|_\varphi = X^T a[R(Y - X\beta)]$. Once $\beta$ is estimated, we estimate the intercept $\alpha$ by the median of the residuals, $\hat{\alpha}_s = \text{med}_{kj} \{ y_{kj} - x'_{kj}\hat{\beta}_\varphi \}$. Both estimators are regression and scale equivariant.

2.2 Fixed Effects Estimation Theory

Because of the invariances, without loss of generality, assume that the true regression parameters are zero in Model (2.1). Asymptotic theory for the fixed effects estimator involves establishing the distribution of gradient $S_X(0)$ and the asymptotic quadraticity of the dispersion function.
Consider Model (2.1) and assume the conditions given in that subsection. It then follows from Brunner and Denker (1994) that the projection of the gradient $S(Y - X\beta)$ is the random vector $X'\varphi[F(Y - X\beta)]$, where $\varphi[F(Y - X\beta)] = (\varphi[F(Y_{11} - x'_{11}\beta)], \ldots, \varphi[F(Y_{mn_m} - x'_{mn_m}\beta)])'$. We need to assume that the covariance structure of this projection is asymptotically stable; that is, the following limit exists and is positive definite:

$$\Sigma_\varphi = \lim_{m \to \infty} N^{-1} \sum_{k=1}^m X_k^T \Sigma_{\varphi,k} X,$$

(2.3)

where $\Sigma_{\varphi,k} = \text{Cov}\{\varphi[F(e_k)]\}$. (In likelihood methods, a similar assumption is made on the covariance structure of the errors).

Under these assumptions, it follows from Theroem 3.1 of Brunner and Denker (1994) that $\frac{1}{\sqrt{N}} S_X(0) \xrightarrow{D} N_p(0, \Sigma_\varphi)$, where $\Sigma_\varphi$ is defined in expression (2.3). The linearity and quadraticity results respectively of Jurečková (1971) and Jaeckel (1972) for the fixed effects model can be extended to our model. The linearity result is $S_X(\beta) = S_X(0) - \tau_\varphi^{-1} N^{-1} X^T X \beta + o_p(\sqrt{N})$, uniformly for $\sqrt{N} \|\beta\|_2 \leq c$, for $c > 0$, where $\tau_\varphi$ is the scale parameter defined by

$$\tau_\varphi^{-1} = \int_0^1 \varphi(u) \left\{ -\frac{f'[F^{-1}(u)]}{f[F^{-1}(u)]} \right\} du. \quad (2.4)$$

Then $\sqrt{N} \hat{\beta}_{\varphi} = \tau_\varphi \sqrt{N}(X^T X)^{-1} X^T \varphi[F(e)] + o_p(1)$. From this, it follows that the asymptotic distribution of $\hat{\beta}_{\varphi}$ is normal with mean $\beta$ and covariance matrix

$$V_{\varphi} = \tau_\varphi^2(X^T X)^{-1} \left( \sum_{k=1}^m X_k^T \Sigma_{\varphi,k} X_k \right) (X^T X)^{-1}. \quad (2.5)$$

7
Letting $\tau_s = 1/2f(0)$, $\hat{\alpha}_S$ is asymptotically normal with mean $\alpha$ and variance

$$\sigma^2_1(0) = \tau_S^2 \frac{1}{N} \sum_{k=1}^{m} \left[ \sum_{j=1}^{n_k} \text{var}(\text{sgn}(e_{kj})) + \sum_{j \neq j'} \text{cov}(\text{sgn}(e_{kj}), \text{sgn}(e_{kj'})) \right]. \quad (2.6)$$

In this section, we have kept the model general; i.e., we have not specified the covariance structure. To conduct inference, we need an estimate of the covariance matrix of $\hat{\beta}_\varphi$. Define the residuals of the JR fit by

$$\hat{e}_{JR} = Y - \hat{\alpha}_s 1_N - X\hat{\beta}_\varphi. \quad (2.7)$$

Using these residuals, we can estimate the parameter $\tau_\varphi$ as proposed by Koul et al. (1987). Next, a nonparametric estimate of $\Sigma_{\varphi,k}$ is obtained by replacing the distribution function $F(t)$ in its definition by the empirical distribution function of the residuals. Based on these results, an asymptotic $(1 - \alpha)100\%$ confidence interval for $h'\beta$ for a vector $h \in \mathbb{R}^p$ is given by $h'\hat{\beta}_\varphi \pm z_{\alpha/2} \sqrt{h'\hat{V}_\varphi h}$.

On the other hand, in many applications the form of the covariance structure is assumed. For example, consider a simple mixed model with block as a random effect, which is discussed in detail in the Section 3. Here, besides an estimate of $\tau_\varphi$, only an additional covariance parameter is required to estimate $V_\varphi$. Another rich class of such models is the repeated measure designs, where block is synonymous with subject. The errors for a subject could have compound symmetrical structure, which would be the simple random effect model or the errors could follow a stationary time series model, for instance an autoregressive model. In this case, the univariate marginals would have the same distribution and, hence, the assumptions would hold for an initial JR fit. Based on this fit, R estimators of the autoregressive parameters of the error distribution can be obtained. These estimates could then be used in
the usual way to transform the observations and then a second (generalized) JR estimate could be obtained from the transformed observations. This is a robust analogue of the two-stage estimation procedure discussed for cluster samples in Rao et al. (1993). Such generalized JR estimators for a broad class of models are under current investigation.

Our discussion has been for general scores. At times, information about the distribution of the errors is known. For example, the error distributions appear to be very heavy-tailed, as in the example of Section 4, or the distribution may be skewed. In such cases appropriate score functions can be chosen. Further, the constant of proportionality of the covariance of the JR estimator is $\tau_\varphi$. Hence, if information on the form of the density $f(t)$ is known, a prudent choice of the score function can optimize the analysis, (Hettmansperger and McKean, 1998).

If the score function is bounded, the JR estimators have bounded influence in response space but not in factor space. However, for outliers in factor space, the high breakdown estimators of Chang et al. (1999) can be similarly extended.

Consider general linear hypotheses of the form $H_0 : \mathbf{H} \mathbf{\beta} = \mathbf{0}$ versus $H_A : \mathbf{H} \mathbf{\beta} \neq \mathbf{0}$, where $\mathbf{H}$ is a $q \times p$ matrix of rank $q$. We offer two test statistics. First, the asymptotic distribution of $\hat{\mathbf{\beta}}_\varphi$ suggests a Wald type test of $H_0$ based on the test statistic $T_{W,\varphi} = (\mathbf{H} \hat{\mathbf{\beta}}_\varphi)^T [\mathbf{H} \hat{\mathbf{V}}_\varphi \mathbf{H}^T]^{-1} (\mathbf{H} \hat{\mathbf{\beta}}_\varphi)$. Under $H_0$, $T_{W,\varphi}$ has an asymptotic $\chi^2_q$ distribution with $q$ degrees of freedom. Hence, a nominal level $\alpha$ test is to reject $H_0$ if $T_{W,\varphi} > \chi^2_\alpha(q)$. As in the independent error case (see Hettmansperger and McKean, 1998), this test is consistent for all alternatives of the form $\mathbf{H} \mathbf{\beta} \neq \mathbf{0}$. For efficiency results consider a sequence of local alternatives of the form $H_{An} : \mathbf{H} \mathbf{\beta}_n = \frac{\beta}{\sqrt{n}}$, where $\mathbf{\beta} \neq \mathbf{0}$. Under this sequence of alternatives $T_{W,\varphi}$ has an asymptotic noncentral $\chi^2_q$-distribution with noncentrality parameter $\eta = (\mathbf{H} \mathbf{\beta})^T [\mathbf{H} \hat{\mathbf{V}}_\varphi \mathbf{H}^T]^{-1} (\mathbf{H} \mathbf{\beta})$. A second test utilizes the reduction in dispersion, $RD_\varphi = D(\text{Red}) - D(\text{Full})$, where $D(\text{Full})$
and $D$(Red) are respectively the minimum values of the dispersion function under the full and reduced (full model constrained by $H_0$) models. The asymptotically correct standardization depends on the dependence structure of the errors.

### 3 Simple Mixed Model

In this section, we discuss a mixed model with block as a random effect. So consider Model (2.1), but for each block $k$, model the error vector $e_k$ as $e_k = 1_{n_k}b_k + \epsilon_k$, where the components of $\epsilon_k$ are independent and identically distributed and $b_k$ is a continuous random variable which is independent of $\epsilon_k$. Assume that the random effects $b_1, \ldots, b_m$ are independent and identically distributed random variables. It follows that the distribution of $e_k$ is exchangeable and all marginal distributions of $e_k$ are the same; so, the theory of Section 2 holds. For this model, though, the asymptotic variance-covariance matrix of $\hat{\beta}_\varphi$, (2.5) simplifies to

$$
\tau_\varphi^2(X'X)^{-1} \sum_{k=1}^{m} X_k' \Sigma_{\varphi,k} X_k (X'X)^{-1},
\Sigma_{\varphi,k} = (1 - \rho_\varphi)I_{n_k} + \rho_\varphi J_{n_k},
$$

(3.1)

and $\rho_\varphi = \text{cov}\{\varphi[F(e_{11})], \varphi[F(e_{12})]\} = E\{\varphi[F(e_{11})]\varphi[F(e_{12})]\}$. Also, the asymptotic variance of the intercept (2.6) simplifies to $N^{-1}\tau_\varphi^2(1+n^*\rho_\varphi^*S)$, for $\rho_\varphi^* = \text{cov}[\text{sgn}(e_{11}), \text{sgn}(e_{12})]$ and $n^* = N^{-1}\sum_{k=1}^{m} n_k(n_k - 1)$. As with LS, for positive definiteness, we need to assume that each of $\rho_\varphi$ and $\rho_\varphi^*S$ exceeds $\max_k\{-1/(n_k - 1)\}$. Let $M = \sum_{k=1}^{m}\binom{n_k}{2} - p$. A simple moment estimator of $\rho_\varphi$ is

$$
\hat{\rho}_\varphi = M^{-1} \sum_{k=1}^{m} \sum_{i>j} a[R(\hat{e}_{ki})]a[R(\hat{e}_{kj})].
$$

(3.2)
Plugging this into (3.1) and using the estimate of \( \tau_\phi \) discussed in Section 2, we have an estimate of the asymptotic covariance matrix of the JR estimators.

### 3.1 Variance Component Estimators

Assume that the variances of the errors exist. Let \( \Sigma_{e_k} \) denote the variance-covariance matrix of \( e_k \). Under the model of this section, the variance-covariance matrix of \( e_k \) is compound symmetric having the form

\[
\Sigma_{e_k} = \sigma^2 A_k(\rho) = \sigma^2 [(1 - \rho) I_{n_k} + \rho J_{n_k}],
\]

where \( \sigma^2 = \text{Var}(\epsilon_{ki}) \), \( I_{n_k} \) is the identity matrix of order \( n_k \), and \( J_{n_k} \) is a \( n_k \times n_k \) matrix of ones. Letting \( \sigma^2_b \) and \( \sigma^2_\varepsilon \) denote respectively the variances of the random effect \( b \) and the error \( \varepsilon \), we have that \( \sigma^2 = \sigma^2_\varepsilon + \sigma^2_b \) and, hence, that \( \rho = \sigma^2_b / (\sigma^2_\varepsilon + \sigma^2_b) \). These parameters, \( (\sigma^2_\varepsilon, \sigma^2_b, \sigma^2) \), are referred to as the variance components.

To estimate these variance components, we proceed similar to Rashid and Nandram (1998); see, also, Gerand and Schucany (2007) in prediction of the random effect in their model. In block \( k \), rewrite model (2.1) as

\[
y_{kj} - [\alpha + x'_{kj}\beta] = b_k + \varepsilon_{kj}, \quad j = 1, \ldots, n_k.
\]

Since the residuals \( \hat{\varepsilon}_{kj} \), (2.7), estimate the left-side, a predictor of \( b_k \) is given by \( \hat{b}_k = \text{med}_{1 \leq j \leq n_k} \{ \hat{\varepsilon}_{kj} \} \). Let MAD denote the median of the absolute deviations from the median, i.e.,

\[
\text{MAD}(v) = 1.483 \text{med}_{1 \leq i \leq l} |v_i - \text{med}_{1 \leq j \leq l} \{ v_j \}|, \quad v \in \mathbb{R}^l.
\]

We have tuned MAD so that it is a consistent robust estimator of the standard deviation at the normal distribution. Hence, a robust estimator of \( \sigma^2_b \) is \( \hat{\sigma}^2_b = (\text{MAD}_{1 \leq k \leq m} \{ \hat{b}_k \})^2 \). The residuals for the errors \( \varepsilon_{kj} \) are \( \hat{\varepsilon}_{kj} = \hat{\varepsilon}_{kj} - \hat{b}_k, \quad j = 1, \ldots, n_k, k = 1, \ldots, m \) and a robust estimate of \( \sigma^2_\varepsilon \) is \( \hat{\sigma}^2_\varepsilon = (\text{MAD}_{1 \leq j \leq n_k, 1 \leq k \leq m} \{ \hat{\varepsilon}_{kj} \})^2 \). Thus,

\[
\hat{\sigma}^2 = \hat{\sigma}^2_\varepsilon + \hat{\sigma}^2_b \quad \text{and} \quad \hat{\rho} = \hat{\sigma}^2_b / \hat{\sigma}^2.
\]
3.2 Studentized Residuals

Studentized residuals are fundamental to diagnostic analyses of linear models. They correct for both the model (factor space) and the underlying covariance structure and allow for a simple benchmark rule for designating potential outliers. McKean, Sheather and Hettmansperger (1990) proposed Studentized residuals for robust analyses of linear models for the independent error case based on first order effects of the asymptotic representations of the robust estimators. In this section, we Studentize the JR residuals, \( \hat{e}_{JR} \) given in expression (2.7).

Because the block sample sizes \( n_k \) are not necessarily the same, some additional notation simplifies the presentation. Let \( \nu_1 \) and \( \nu_2 \) be two parameters and define the block-diagonal matrix \( B(\nu_1, \nu_2) = \text{diag}\{B_1(\nu_1, \nu_2), \ldots, B_m(\nu_1, \nu_2)\} \), where \( B_k(\nu_1, \nu_2) = (\nu_1 - \nu_2)I_{n_k} + \nu_2J_{n_k}, \ k = 1, \ldots, m \). Using this notation, \( \text{Var}(e) = \sigma^2B(1, \rho) \).

Using the asymptotic representation for \( \hat{\beta}_\varphi \) given in Section 2, the approximate covariance matrix of \( \hat{e}_{JR} \) is given by

\[
C_{JR} = \sigma^2B(1, \rho) + \frac{\tau^2}{N^2}J_NB(1, \rho^*_S)J_N + \tau^2H_cB(1, \rho_\varphi)H_c - \frac{\tau_s}{N}J_NB(\delta^*_{11}, \delta^*_{12}) - \tau H_cB(\delta_{11}, \delta_{12}) - \frac{\tau_s}{N}H_cB(\gamma_{11}, \gamma_{12})J_N,
\]

where \( H_c \) is the projection matrix onto the column space of the centered design matrix \( X_c, J_N \) is the \( N \times N \) matrix of all ones, \( \delta^*_{11} = E[e_{11}\text{sgn}(e_{11})], \delta^*_{12} = E[e_{11}\text{sgn}(e_{12})], \delta_{11} = E[e_{11}\varphi(F(e_{11}))], \delta_{12} = E[e_{11}\varphi(F(e_{12}))], \gamma_{11} = E[\text{sgn}(e_{11})\varphi(F(e_{11}))], \gamma_{12} = E[\text{sgn}(e_{11})\varphi(F(e_{12}))] \) and \( \rho_\varphi \) and \( \rho^*_S \) are defined in (2.3) and (2.5), respectively.

To compute the Studentized residuals, estimates of the parameters in \( C_{JR}, (3.4) \),
are required. First, consider the matrix $\sigma^2 B(1, \rho)$. In Section 3.1, we obtained robust estimators $\hat{\sigma}^2$ and $\hat{\rho}$ given in expression (3.3). Substituting these estimators for $\sigma^2$ and $\rho$ into $\sigma^2 B(1, \rho)$, we have a robust estimator of $\sigma^2 B(1, \rho)$ given by $\hat{\sigma}^2 B(1, \hat{\rho})$.

Expression (3.2) gives a simple moment estimator of $\rho_\varphi$. The parameters $\rho_\varphi^*, \delta_{11}, \delta_{12}, \delta_{11}^*, \delta_{12}^*, \gamma_{11},$ and $\gamma_{12}$ can be estimated in the same way. Substituting these estimators into the matrix $C_{JR}$, let $\hat{C}_{JR}$ denote the resulting estimator.

For $t = 1, \ldots, N$, let $\hat{c}_{tt}$ denote the $tth$ diagonal entry of the matrix $\hat{C}_{JR}$. Then the $tth$ Studentized residual based on the JR fit is $\hat{e}_{JR,t}^* = \hat{e}_{JR,t}/\sqrt{\hat{c}_{tt}}$. Traditional benchmarks used with Studentized residuals are the limits $\pm 2$; that is, Studentized residuals which exceed 2 in absolute value are called potential outliers.

4 Example and Simulation Studies

Cobb (1998) presented an example of a complete block design concerning the weight of crab grass. The fixed factors in the experiment are the density of the crabgrass (four levels) and the levels (two) of the three nutrients nitrogen, phosphorus, and potassium. Two complete blocks of the experiment were carried out, so altogether there are $N = 64$ observations. Here block is a random factor and we assume the simple mixed model of Section 3. Under each set of experimental conditions, crab grass was grown in a cup. The response is the dry weight of a unit (cup) of crab grass, in milligrams. The data are presented in Cobb (1998). We are interested in the JR estimates of the effects and their standard errors. We also present the results of simulation studies which verify the validity of the analysis.

We discuss other scores for this example below, but for now we consider the rank-based analysis based on Wilcoxon scores. For the main effects model, Table 1 displays the estimated effects (contrasts) and standard errors for the Wilcoxon and
LS analyses. For the nutrients, these effects are the differences between the high and low levels, while for the density the three contrasts reference the highest density level. There are major differences between the Wilcoxon and the LS estimates. For the Wilcoxon estimates, the nutrients nitrogen and phosphorus are significant and the contrast between the low and high levels of density is highly significant. Nitrogen is the only significant effect for the high level analysis. The Wilcoxon test statistic for the testing density effect $T_{W,\varphi} = 20.55$ with $p = 0.002$, while the LS test statistic is $F_{\text{lmE}} = 0.82$ with $p = 0.490$. The robust estimates of the variance components are: $\hat{\sigma}^2 = 206.33$, $\hat{\sigma}_b^2 = 20.28$, and $\hat{\rho} = 0.098$

An outlier accounts for much of this dramatic difference between the robust and LS analyses. Originally, one of the responses was mistyped; instead of the correct value 97.25, the response was typed as 972.5; see Cobb (1998) for discussion. Upon replacing the outlier with its correct value, the Wilcoxon and LS analyses are similar; although, the Wilcoxon analysis is still more precise; see the discussion below on the other outliers in this data set. This is true too of the test for the factor density: $T_{W,\varphi} = 23.23$ ($p = 0.001$) and $F_{\text{lmE}} = 6.33$ with $p = 0.001$. The robust estimates of the variance components are: $\hat{\sigma}^2 = 209.20$, $\hat{\sigma}_b^2 = 20.45$, and $\hat{\rho} = 0.098$ These are essentially unchanged from their values on the original data. If on the original data the experimenter had run the robust fit and compared it with the LS fit, then the

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Wilcoxon Est.</th>
<th>SE</th>
<th>LS Est.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nit</td>
<td>39.90</td>
<td>4.08</td>
<td>69.76</td>
<td>28.7</td>
</tr>
<tr>
<td>Pho</td>
<td>10.95</td>
<td>4.08</td>
<td>-11.52</td>
<td>28.7</td>
</tr>
<tr>
<td>Pot</td>
<td>-1.60</td>
<td>4.08</td>
<td>28.04</td>
<td>28.7</td>
</tr>
<tr>
<td>D_{34}</td>
<td>3.26</td>
<td>5.76</td>
<td>57.74</td>
<td>40.6</td>
</tr>
<tr>
<td>D_{24}</td>
<td>7.95</td>
<td>5.76</td>
<td>8.36</td>
<td>40.6</td>
</tr>
<tr>
<td>D_{14}</td>
<td>24.05</td>
<td>5.76</td>
<td>31.90</td>
<td>40.6</td>
</tr>
</tbody>
</table>

Table 1: Wilcoxon and LS Estimates and SEs of Effects for the Crabgrass.
outlier would have been discovered immediately.

Figure 1 contains the Wilcoxon Studentized residual plot and $q-q$ plot for the original data (containing the large outlier). We have removed the large outlier, so that we can focus on the remaining data. The "vacant middle" in the residual plot is an indication that interaction may be present. For the hypothesis of interaction between the nutrients, the value of the Wald type test statistic is $T_{W,\phi} = 30.61$, with $p = 0.000$. Hence, the R analysis strongly confirms that interaction is present. On the other hand, the LS likelihood ratio test statistic for this interaction is $2.92$, with $p = 0.404$. In the presence of interaction, many statisticians would consider interaction contrasts instead of a main effects analysis. Hence, for such statisticians, the robust and LS analyses would have different practical interpretations.

![Studentized Residual Plot, Outlier Deleted](image1)

![Normal q-q Plot, Outlier Deleted](image2)

Figure 1: Studentized Residual and $q-q$ Plots, Minus Large Outlier.
4.1 Simulation Studies of Validity

In this data set, the number of blocks is two. Hence, to answer questions concerning the validity of the Wilcoxon analysis, we conducted a small simulation study. Table 2 summarizes the empirical confidences and AREs of this study for two situations, normal errors and contaminated normal errors (20% contamination and the ratio of the contaminated variance to the uncontaminated variance at 25). For each situation, we used the same design with the correlation structure as estimated by the Wilcoxon analysis. We used the same design, so, in particular, we used only $m = 2$ blocks. Our discussion of this example focused on the estimates and their standard errors. Thus from our study, we report the empirical confidence of the asymptotic 95% confidence intervals of the form $\text{Estimate} \pm 1.96 \times \text{SE}$, where $\text{SE}$ denotes the standard errors of the estimates. The number of simulations was 10,000 for each situation, therefore, the error in the table based on the usual 95% confidence interval for a proportion is 0.004. The empirical confidences for the Wilcoxon are quite good with the target of 0.95 usually within range of error. They were perhaps a little conservative at the the contaminated normal situation. Hence, the Wilcoxon analysis appears to be valid for this design. The intervals based on LS are slightly liberal. Turning our attention to the empirical AREs, LS is more efficient at the normal but the efficiencies are close to the value 0.95 for the independent error case. The Wilcoxon analysis is much more efficient over the contaminated normal situation.

One referee suggested a comparison between the JR analysis of Section 2 and the R analysis for the independent errors (IR) as discussed in McKeon and Vidmar (1994); see Rao et al. (1993) for a comparison of traditional analyses. We are planning a larger study but for this example we ran 10,000 simulations using the model of the example of this section. Wilcoxon scores were used for both analyses. We considered
normal error distributions, setting the variance components at the values of the robust estimates. The JR and IR fits are the same, so we consider the differences in their inferences of the six effects listed in Table 1. For 95% nominal confidence, the average empirical confidences over these six contrasts are 95.32% and 96.12%, respectively for the JR and IR procedures. Hence, both procedures appear valid. For a measure of efficiency, we averaged, across the contrasts, the averages of squared lengths of the confidence intervals. The ratio of the JR to the IR averages is 0.914; hence for the simulation, the JR inference is about 9% more efficient than the IR inference. Similar results for traditional analyses are reported in Rao et al. (1993).

### 4.2 Simulation Study of Other Score Functions

Besides the large outlier, there are six other potential outliers. This quantity of outliers suggests the use of score functions which are more optimal than the Wilcoxon score function for very heavy-tailed error structure. To investigate this, we considered a symmetric Winsorized Wilcoxon score function. While the Wilcoxon score function is linear over the entire interval \((0,1)\), the Winsorized Wilcoxon score function is a continuous nondecreasing function which is linear in the middle and constant on the ends. Thus, relatively, less weight is given to outlying residuals in the fitting process. These scores are optimal for error distributions with exponential left and

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Norm. Errors</th>
<th>Cont. Norm. Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wilc.</td>
<td>LS</td>
</tr>
<tr>
<td>Nit</td>
<td>0.948</td>
<td>0.932</td>
</tr>
<tr>
<td>Pho</td>
<td>0.953</td>
<td>0.934</td>
</tr>
<tr>
<td>Pot</td>
<td>0.948</td>
<td>0.927</td>
</tr>
<tr>
<td>D_{34}</td>
<td>0.950</td>
<td>0.929</td>
</tr>
<tr>
<td>D_{24}</td>
<td>0.951</td>
<td>0.929</td>
</tr>
<tr>
<td>D_{14}</td>
<td>0.952</td>
<td>0.930</td>
</tr>
</tbody>
</table>
right tails and a logistically distributed middle; see p. 100 of Hettmansperger and McKe (1998) and McKean et al. (1999). Hence, for the Cobb data they seem more appropriate than the Wilcoxon scores. In the article by McKean et al. (1989), Winsorized Wilcoxon scores for skewed error distributions are investigated and the Monte Carlo study showed that they were more efficient than the Wilcoxon scores, in the presence of skewed error distributions.

For the Cobb data, we chose the score function which is linear over the interval $(0.2, 0.8)$; hence, there is 20\% Winsorizing on both sides and we shall denote it by $WW_2$. For the parameters as in Table 1, the $WW_2$ estimates and standard errors (in parentheses) are: $39.16 (3.78), 10.13 (3.78), -2.26 (3.78), 2.55 (5.35), 7.68 (5.35)$, and $23.28 (5.35)$. The estimate of the scale parameter $\tau$ is 14.97 compared to the Wilcoxon estimate which is 15.56. This indicates that an analysis based on the $WW_2$ fit has more precision than one based on the Wilcoxon fit.

To investigate this gain in precision, we ran a small simulation study. We used the same model and the same correlation structure as estimated by the Wilcoxon fit. We considered normal and contaminated normal errors, with the percent of contamination at 20\% and the relative variance of the contaminated part at 25. For each situation 10,000 simulations were run. The AREs were very similar for all six parameters, so we will only report their averages. For the normal situation the average ARE between the $WW_2$ and Wilcoxon estimates was 0.90; hence, the $WW_2$ estimate was 10\% less efficient for the normal situation. For the contaminated normal situation, though, this average was 1.21; hence, the $WW_2$ estimate was 20\% more efficient than the Wilcoxon estimate for the contaminated normal situation.

Computation of the $WW_2$ estimates poses no additional problems, because we need only change the score function in the algorithm. In the same way, scores could be chosen for skewed error distributions. For example, if the error distribution had
a heavy right tail, the WW estimate which is linear over (0, 0.80) and constant over (0.80, 1) would be less sensitive to large outliers on the right than the Wilcoxon estimate.

There are other families of score functions besides the Winsorized Wilcoxon scores. Gastwirth (1966) presents several families of score functions appropriate for classes of distributions with tails heavier than the exponential distribution. For certain cases, he selects a score based on a maxi-min strategy. We are investigating adaptive procedures to choose the scores optimally from such families.

5 Comparisons

For the simple mixed model of Section 3, we have explored the ARE’s between the JR analysis and the traditional analysis based on LS (maximum likelihood analysis under normality) and the JR analysis and a rank procedure where the ranking of residuals are within blocks (Friedman-type). We offer a quick summary of our investigations.

If we assume that the design is centered within blocks then the ARE between the JR and LS analyses can be obtained in closed form. For Wilcoxon scores, this ARE is

\[ ARE(F_{W,\varphi}, F_{LS}) = \frac{(1 - \rho)(1 - \rho)}{(1 - \rho^2)}\frac{12\sigma^2}{\left(\int f^2(t) dt\right)^2}, \]

where \( \rho_{\varphi} \) is defined under expression (3.2) and \( \rho \) is the correlation within a block. If the random vectors in a block follow the multivariate normal distribution, then this ARE lies in \([0.8660, 0.9549]\) when \(0 < \rho < 1\). The lower bound is attained when \( \rho \rightarrow 1 \). The upper bound is attained when \( \rho = 0 \) (the independent case). This, of course, is the usual high efficiency of the Wilcoxon to LS analysis at the normal distribution. When \(-1 < \rho < 0\), this ARE lies in \([0.9549, 0.9662]\) and the upper bound is attained when \( \rho = -0.52 \) and the lower bound is attained when \( \rho \rightarrow -1 \). Generally, the high efficiency properties of the Wilcoxon analysis to LS analysis in the independent errors case extend to the
Wilcoxon analysis for this mixed model design.

ARE results were obtained for several common error distributions comparing the JR estimator with one other rank estimator, the MR, which is based on a Friedman-type of ranking and was proposed by Rashid (1995). Except in extreme cases the JR estimator does quite well. For example the JR estimator is more efficient than the MR estimator at the normal distribution except when the block size is large or when the correlation is high. We confirmed these ARE results by a Monte Carlo simulation study. This simulation study also confirms that for heavier tailed distributions (contaminated normals) the JR estimator is more efficient than LS.

6 Conclusion

In this paper we extended Jaeckel’s (1972) ordinary rank (JR) estimator to models with dependent block error structure. Standard errors and tests of linear hypotheses were also discussed. The results are for general models and scores. For the mixed model with one random effect, in addition to the estimation of the fixed effects, we proposed practical, robust estimates of the variance components. We further developed Studentized residuals which can be used to determine potential outliers, as well as to check the quality of fit of the model. We considered highly efficient estimates, but the high breakdown estimates presented in Chang et al. (1999) for the independent error case can be extended in the same way to these models.

We illustrated the robustness of the procedures on a real data set which contained several outliers. Our robust procedures were much less sensitive to the effect of the outliers than the traditional analysis based on LS. A simulation study over situations similar to that of the data set confirmed the validity of our analysis. The study also showed the robustness of efficiency of our approach over that of the tra-
ditional analysis. The quantity of outliers in this data set suggest that rank scores more suited for very heavy-tailed error structure would be more powerful than the Wilcoxon analysis. We presented the results of the analysis based on one such score, a Winsorized Wilcoxon. It was more precise than the Wilcoxon analysis. The results of a simulation study confirmed the validity of the Winsorized Wilcoxon analysis and its efficiency edge over the Wilcoxon analysis for very heavy-tailed error structure. We intend to do more in depth studies comparing the JR analyses with traditional and other rank-based analyses in the future.

References


