



6-2012

Rank-Based Estimation and Prediction for Mixed Effects Models in Nested Designs

Yusuf K. Bilgic

Western Michigan University, yekabe@hotmail.com

Follow this and additional works at: <https://scholarworks.wmich.edu/dissertations>



Part of the Mathematics Commons, Statistical Models Commons, and the Statistical Theory Commons

Recommended Citation

Bilgic, Yusuf K., "Rank-Based Estimation and Prediction for Mixed Effects Models in Nested Designs" (2012). *Dissertations*. 40.

<https://scholarworks.wmich.edu/dissertations/40>

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.



RANK-BASED ESTIMATION AND PREDICTION FOR MIXED EFFECTS
MODELS IN NESTED DESIGNS

by

Yusuf K. Bilgic

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Statistics
Advisor: Joseph W. McKean, Ph.D.

Western Michigan University
Kalamazoo, Michigan
June 2012

THE GRADUATE COLLEGE
WESTERN MICHIGAN UNIVERSITY
KALAMAZOO, MICHIGAN

Date 5/16/2012

WE HEREBY APPROVE THE DISSERTATION SUBMITTED BY

YUSUF K. BILGIC

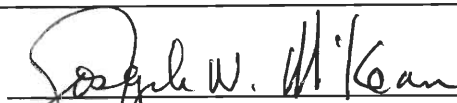
ENTITLED RANK-BASED ESTIMATION AND PREDICTION
FOR MIXED EFFECTS MODELS IN NESTED DESIGNS

AS PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF Doctor of Philosophy

Statistics

(Department)



JOSEPH W. MCKEAN
Dissertation Review Committee Chair

Statistics

(Program)


JEFFREY TERPSTRA
Dissertation Review Committee Member


JUNG C. WANG
Dissertation Review Committee Member


JON DAVIS
Dissertation Review Committee Member

APPROVED


Dean of The Graduate College

Date

June 2012

RANK-BASED ESTIMATION AND PREDICTION FOR MIXED EFFECTS MODELS IN NESTED DESIGNS

Yusuf K. Bilgic, Ph.D.

Western Michigan University, 2012

Hierarchical designs frequently occur in many research areas. The experimental design of interest is expressed in terms of fixed effects but, for these designs, nested factors are a natural part of the experiment. These nested effects are generally considered random and must be taken into account in the statistical analysis. Traditional analyses are quite sensitive to outliers and lose considerable power to detect the fixed effects of interest.

This work proposes three rank-based fitting methods for handling random, fixed and scale effects in k-level nested designs for estimation and inference. An algorithm, which iteratively obtains robust prediction for both scale and random effects, is used along with the proposed fitting methods including Joint Ranking (JR), Iteratively Reweighted Generalized Rank Estimate (GR), and Rank-based General Estimating Equation (GEER). For simplicity, a 3-level nested design that deals with students nested within sections in schools is handled. The asymptotic derivations for the proposed estimators are discussed. The results of a Monte Carlo evaluation of the methods, including comparisons with the traditional analysis are provided. The proposed methods compete with the traditional method under normal case, outperform it when random errors are contaminated, and inherit better efficiency properties of the estimates when outlier exists. The performance of the rank-based estimators of fixed parameters is more efficient than the REML. When random errors are contaminated, the intra-class correlation estimates in the proposed

algorithm are unbiased, while the REML estimates are biased. Also, real data examples of applications are presented.

© Yusuf K. Bilgic 2012

Acknowledgments

I owe great thanks to my advisor Dr. Joseph W. McKean for his knowledge, guidance and patience. He is such a great professor who has been very supportive and friendly. I also wish to thank the committee members, Dr. Terpstra, Dr. Wang and Dr. Davis for their valuable comments. I never forget the chair of the stat department and the acting chair of the math department, Dr. Magdalena Niewiadomska-Bugaj and Dr. Steven Ziebarth, for their forever support to me.

Finally, I wish to express my gratitude to my spouse, Selime, and beautiful daughters, Refia, Saime and Bedia, for their patience during my study.

Yusuf K. Bilgic

Table of Contents

ACKNOWLEDGMENTS	ii
LIST OF TABLES	vi
1 Introduction	1
1.1 Nested Designs	1
1.1.1 Correlation Structure Should Be Incorporated in Model	2
1.1.2 Common Research Questions	2
1.2 Existing Procedures: Techniques and Algorithms	3
1.2.1 Likelihood Procedures	5
1.2.2 Generalized Least Squares Estimates	7
1.2.3 Why the Need to Improve Existing Techniques	8
1.2.4 Research Goal	8
2 Methodology	10
2.1 Variance Component Estimators	10
2.1.1 Pseudo-Sample Approach	10
2.1.2 Rank Prediction Procedure (RPP)	11
2.1.3 Location Estimators: Hodges-Lehmann and Median	13
2.1.4 Scale Estimators	13
2.1.5 Intra- and Inter- Correlation Coefficients	14
2.2 Rank-Based Analysis	15

Table of Contents—Continued

3	Ranked-Based Estimate and Asymptotic Theory	17
3.1	Model and Assumptions	17
3.2	Joint Ranking Method (JR)	18
3.2.1	Studentized Residuals	23
3.2.2	General Linear Hypothesis	25
3.3	Generalized Rank Method (GR)	25
3.3.1	GR Algorithm	26
3.3.2	Pseudo-Estimates for the k-Step Estimator	26
3.3.3	One Step GR	28
3.3.4	Studentized Residuals	32
3.4	Generalized Estimating Equations Method (GEER)	32
3.5	Variance Component Estimators	37
3.5.1	For 2-level Nested Structure	38
3.5.2	For 3-level Nested Structure	44
3.5.3	Finite Correction for the Scale Estimates	46
4	Simulation Results and Real Data Analysis	47
4.1	Performance of GR Estimates	47
4.1.1	Simulation Results	50
4.1.2	Sensitivity	55
4.2	Empirical Results for JR, GR and GEER Estimators	57
4.2.1	Simulation Results	60
4.3	Application to Real World Data	65
4.3.1	PISA Data Set	65
4.3.2	PASSIVE Data Set	71
5	Discussion, Conclusion and Further Study	77

Table of Contents—Continued

REFERENCES	80
APPENDIX	84

List of Tables

1.1	School Data Analysis with Existing Techniques	6
4.1	Case-1 Simulation Results	51
4.2	Case-2 Simulation Results	52
4.3	Case-3 Simulation Results	53
4.4	Case-4 Simulation Results	54
4.5	Case-1 Simulation Results for Corrupted Data (Outlier)	56
4.6	$\theta = (1, 4, 9)$, Normal	60
4.7	$\theta = (9, 4, 1)$, Normal	60
4.8	$\theta = (1, 1, 9)$, Normal	61
4.9	$\theta = (1, 1, 1)$, Normal	61
4.10	$\theta = (0, 0, 1)$, Normal, nsims=1,000	61
4.11	$\theta = (0, 1, 1)$, Normal, nsims=1,000	62
4.12	$\theta = (1, 0, 1)$, Normal, nsims=1,000	62
4.13	$\theta = (100, 100, 1)$, Normal	62
4.14	$\theta = (100, 1, 100)$, Normal	63
4.15	$\theta = (1, 100, 100)$, Normal	63
4.16	$\theta = (100, 1, 1)$, Normal	63
4.17	$\theta = (1, 100, 1)$, Normal	64
4.18	$\theta = (1, 1, 3)$, CN at 20% with the Variance Ratio at 4	64
4.19	$\theta = (1, 1, 3)$, CN at 10% with the Variance Ratio at 4	64
4.20	$\theta = (1, 1, 1)$, CN at 20% with the Variance Ratio at 2	65

List of Tables—Continued

4.21	Estimates in PISA Data for Cognitive Scores with Two Covariates	67
4.22	The Q-Q Plots of PISA Data Analysis	68
4.23	The Q-Q Plots with Outlier	69
4.24	The Q-Q Plots with Outlier When Omitting the Outlier	70
4.25	Estimates for PASSIVE Data with Treatment	72
4.26	The Q-Q Plots of PASSIVE Data Analysis	73
4.27	The Q-Q Plots with Outlier	74
4.28	The Q-Q Plots with Outlier When Omitting the Outlier	75

Chapter 1

Introduction

Considering 'noisy' quantitative data sets in social science and educational research, there is much variability in the response variable(s) to be explained by statistical analyses. Additionally, the presence of non-normality or outliers makes a complete robust analysis an attractive alternative to traditional analyses.

1.1 Nested Designs

Behavioral and social data commonly have a nested structure and much of educational research deals with this nested structure regarding complex concepts such as learning, achievement, school and teacher effectiveness (Raudenbush & Bryk, 2002; O'Connell & McCoach, 2008; Sahai & Odeja, 2005). Observations taken from individuals within a cluster tend to be similar. Within these clusters, individuals are no longer independent. Nested random effects should be included in the models due to the similarity of each cluster so that correlated structure is taken into account. These models are often referred to as hierarchical models or the multilevel models.¹ Similar studies are being conducted in different fields as well, such as: survey sampling, meta-analysis, agriculture and health research. Survey sampling might happen within organizational units, communities, clusters, or hospitals. Also, meta-analysis is the pooling of results from separate studies.

¹Nested models are the subset of hierarchical models, which are the subsets of mixed models.

Using statistical techniques that employ both information about individuals and groups to which these individuals belong is one of the most challenging aspects of educational data analysis. Until the 1970s, nested structure in educational data was not incorporated in the model so that dependency and random effects were omitted in Ordinary and General Least Square Analyses (OLS and GLS). In contrast to fixed assumptions on clusters, random errors in nested design are dependent within clusters, so the OLS analysis is inappropriate. Taking cluster structure into account yields dependencies between observations within the same clusters regardless of fixed or random block assumptions. Also, to measure the impacts of educational interventions, schools are randomized in practice. This leads to models with group random effects or a correlated model rather than only fixed group effects models. In that regard, controversial inferences have been causing debates about school effectiveness data analysis (Leeuw & Meijer, 2008). Inevitably, reanalysis of data and disputes are common because existing methods lack the ability to handle violations of the assumptions (Raudenbush & Bryk, 2002).

1.1.1 Correlation Structure Should Be Incorporated in Model

The notion of individuals nested in the same groups leads to dependency between the individuals which is explained by *intra-class correlation coefficients* (ICC). Intra-class correlation coefficient is information on the degree of dependencies of the observations within the same cluster. It is a useful and contextual parameter associated with random effects of clusters that measures the proportion of the variability in the outcome to the total one. It is sometimes called cluster effect and applied only to random models (Raudenbush & Bryk, 2002). For example, independent observations within/between-cluster yield an ICC of zero.

1.1.2 Common Research Questions

Nested designs in educational/social-science research often address questions related to

- the examination of differences within and across groups or contexts such as classrooms, schools, neighborhoods on individual outcomes
- the investigation of the degree to which individuals within a group or cluster are similar as measured through the ICC
- the study of the factors that explain institution/school differences
- the effects of clusters and treatment on individual scores (e.g. student's academic achievement)
- the measurement of the impacts of interventions (e.g. educational) (Raudenbush & Bryk, 2002; O'Connell & McCoach, 2008)

1.2 Existing Procedures: Techniques and Algorithms

It is customary in educational research to use classical statistical methodology to conduct mixed (more specifically hierarchical) models analysis when the correlated structures and random effects are considered (Hill & Rowe, 1996; Leeuw & Meijer, 2008; Raudenbush & Bryk, 2002). Well-known statistical software such as SAS, SPSS, R and HLM use procedures such as Maximum Likelihood (ML), Restricted ML (REML) and MINQUE to evaluate variance components in mixed hierarchical models.

Consistent estimation of variance depends on large sample between-group sizes. Also, analyses require moderate within-group sample sizes (e.g., $n=30$). In application, this is often not the case (Leeuw & Meijer, 2008). Moreover, it is well known that the least squares procedures and Gaussian techniques lack statistical power in the presence of unbalanced data, heavy tailed or skewed error distributions. When unbalanced data occur in the design, these procedures will tend to be biased in variance-covariance (var-cov) components estimation. Further, they are quite sensitive to outliers and lose considerable power to detect effects of interest.

Perfectly balanced designs have closed-form expressions to estimate var-cov components. However, for likelihood estimators in practice, these are not available for most of the cases. Numeric optimization is done using iterative algorithms. These estimation and prediction algorithms integrate fixed effects, random effects and var-cov components in iterative schemes.

Example: To illustrate the issues involved in educational data analysis, consider a simple example of a normally distributed 3-level nested data set where students are nested within classrooms in different schools and measured on a variable of interest with a binary covariate (say, a treatment and a control group). Such a model is commonly used in the educational arena. It is known that the ICC is usually between .1 to .2 in the USA schools. In this condition, the data were generated.

The problem is summarized in the linear model as

$$y_{ijk} = x_{ijk}^T \beta + a_i + w_{j(i)} + \epsilon_{k(ij)},$$

$k = 1, \dots, n_{ij}; j = 1, \dots, J_i; i = 1, \dots, 7$, where a_i is the random effect for school i , $w_{j(i)}$ is the random effect for the j^{th} section of school i and k^{th} student, $\epsilon_{k(ij)}$ is the error effect. Table 1.1 shows that the 35% of the total variability on SAT score is due to section differences, the 23% is explained by school differences, and individual differences explain the 65%. The REML estimates obtained from different packages disagree on this whereas the ML, ANOVA and MINQUE agree.

The fixed effect results under normal case are consistent except on the estimates and standard errors in scale parameters. Thus, this produces contradictory significance test results and statistical inferences; see, for example, the standard errors for σ_{sch}^2 . In practice, statistical inferences are not trustworthy when data are unbalanced, small size, with outliers or not normally distributed. This is a lack of robust statistical analyses for hierarchical models. It is the purpose of this work to provide such analyses.

1.2.1 Likelihood Procedures

Traditional least squares procedures often require that error and random terms are normally distributed and use large-sample properties. Based on these assumptions, likelihood (L) procedures have been developed for the estimation of the fixed effects and the variance components in fitting linear mixed models. The basic idea in L is to choose estimates for which the likelihood of observing the actual data is maximum. ML has a shortcoming, that is, estimates of var-cov components are conditional upon point estimates of the fixed effects. REML estimates of var-cov components adjust for the uncertainty about the fixed effects that have potential to yield negative variance in ML.

There are many iterative algorithms that can be considered for computing the GLM, ML, REML, GEE and Bayesian estimates. For example, EM, Newton-Raphson, RIGLS (Goldstein, 1986), Fisher Scoring, and Fully Bayesian are most useful ones. These conceptually distinct approaches are described in Harville (1977), Goldstein (1986; 1995), Raudenbush & Bryk (2002), O'Connell & McCoach (2008) and Pinheiro et al (2011). Under balanced designs and normal errors, these methods coincide in variance component estimation. Depending on the degree to which the data are unbalanced and the particular type of the inference sought, traditional statistical inferences are not trustworthy (Raudenbush & Bryk, 2002), and no such exact tests exist.

Table 1.1: School Data Analysis with Existing Techniques

	Fixed Effect	Param	REML estimates with R package			ML estimates with R package			REML estimates with SAS PROC MIX			ML estimates with SAS PROC MIX			MINQUE estimates with SAS PROC MIX			2-Way Fixed ANOVA Estimates		
			est	se	p-value	est	se	p-value	est	se	p-value	est	se	p-value	est	se	p-value	est	se	p-value
Scale Parameter	Intercept	0	-1.37	1.24	0.27	-1.38	1.16	0.24	-1.37	1.24		-1.38	1.16		-1.46	1.11		0.63	1.05	0.55
	Covariate	0	0.48	0.72	0.5	0.5	0.71	0.48	0.48	0.72	0.5	0.5	0.71	0.48	0.62	0.75	0.4	1.42	0.81	0.08
	σ^2_{sch}	4	7.33	2.02		5.94	2.01		7.33	5.71		5.95	4.56		1.68	1.32				0.03
	σ^2_{sect}	4	4.08	2.71		4.05	2.44		4.08	2.78		4.05	2.74		6.01	3.62				0.59
Intra-Class Correlation	σ^2_{err}	16	21.2	4.62		21.2	4.6		21.3	2.34		21.2	2.32		23.5	2.6		21.3		
	Error p	67%	65%			68%			65%			68%			75%			100%		
	School p	17%	23%			19%			22%			19%			5%					
	Section p	33%	35%			32%			35%			32%			25%					
Model Fit	AIC		1124			1127			1120			1128			1122					
	BIC		1140			1144			1120			1128			1122					

1.2.2 Generalized Least Squares Estimates

One of the traditional methods is generalized least squares analysis to estimate fixed and scale effects in the linear model. Using Goldstein's (1995) Iterative Generalized Least-Square Estimation (IGLS) approach, the general mixed model can be written and estimated as follows:

$$Y = X\beta + e = X\beta + Zb + \epsilon, \quad (1.2.1)$$

where Y denotes an $n \times 1$ vector of responses, X is a $n \times p$ known fixed effects design matrix, β is a $p \times 1$ fixed effects parameter vector, Z is a $n \times k$ known random effects design matrix, b is a $k \times 1$ vector of random effects, and ϵ is an $n \times 1$ vector of random errors.

$E\{(Zb)(Zb)^T\} + \sigma^2\mathbf{I} = V$ and V is the covariance matrix of the response vector Y . In the estimation of the random variables, at each iteration $W = \text{vec}\{(Y - X\hat{\beta})(Y - X\hat{\beta})^T\}$ is regressed on design matrix. $E(W) = Z^*\theta$ where Z^* is the design matrix for the random parameters. When V is known then generalized least squares estimators for the fixed effects are

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \quad (1.2.2)$$

and $\text{cov}(\hat{\beta}) = (X^T V^{-1} X)^{-1}$. If β is known but V is unknown, then estimators θ of the parameters of V as $\hat{\theta} = (Z^{*T} M^{-1} Z^*)^{-1} Z^{*T} M^{-1} W$, $M = V \otimes V$ where \otimes is the Kronecker product. The covariance matrix of $\hat{\theta}$ is given by

$$(Z^{*T} M^{-1} Z^*)^{-1} Z^{*T} M^{-1} \text{cov}(W) M^{-1} Z^* (Z^{*T} M^{-1} Z^*)^{-1} \quad (1.2.3)$$

The iterative procedure iterates between (1.2.2) and (1.2.3) using the current estimates of the fixed and random parameters. The IGLS procedure produces biased estimates and a simple modification (called Restricted Unbiased IGLS or RIGLS) corrects the bias and leads to REML estimates. Random effects are obtained via the estimates of covariance

and residuals (Searle et al. 2006; Goldstein, 1995).

1.2.3 Why the Need to Improve Existing Techniques

Rank-based approaches are highly robust. They agree with traditional analyses when the random errors are normally distributed, and they are much less sensitive to outliers than the traditional analyses when the random errors are not normally distributed. For example, under location problems, these approaches achieve up to 95% efficiency relative to least squares methods when the data are normal and are much more efficient than the least squares methods for heavy tailed error distributions (Terpstra & McKean, 2005; Hettmansperger & McKean, 2011; Hollander & Wolfe, 1999).

In the literature, to our knowledge, robust concepts are employed for covariance estimation of fixed effects in ML setting which depends on normality assumptions (See p.42 in Leeuw & Meijer, 2008; p.407-425 in O’Connell & McCoach, 2008). Rank-based analysis in mixed model is theoretically introduced in Brunner & Denker (1994) and Akritas & Brunner (1997) using general score function. This is a comprehensive theory to test hypotheses, although not for fitting (estimation) and model checking. Kloke et al. (2009) developed rank-based estimation using the joint ranking method and testings for the fixed effects for models with dependent error structure. While their simple mixed model analysis covered a two-level nested design, it does not handle many-level nested structures.

Robust or nonparametric option for hierarchical linear models is not found in existing softwares such as HLM, MLwiN, SAS MIXED, and R packages (*i.e.*, *nlme/lme4*). Also, efficient techniques in checking for influential clusters or observations are not yet available for multilevel models (Goldstein, 1995).

1.2.4 Research Goal

In this dissertation, the goal is to investigate three proposed rank-based methods in handling random, fixed and scale effects in k-level random effect nested designs. Asymp-

tistical results will be derived. For the sake of simplicity, a 3-level nested design that deals with measurements nested within sections in schools will be used. This nested analog could be adopted for other organizational studies and settings. The validation and efficiency results of the proposed methods are presented via a Monte Carlo investigation including a comparison with the traditional analysis. Further, these are shown with real data examples.

Chapter 2

Methodology

2.1 Variance Component Estimators

Robust predictions of random effects have been discussed in several papers, including Groggel et al. (1988), Dubnicka (2004), and Kloeke et al. (2009). These techniques to predicting random effects based on clusters use robust estimators of the variance components. However, these consider models for cluster correlated data, namely here a 2-level random nested design. In this study, the proposed procedures for the rank-based predictions of the random effects use the similar algorithm as in the study of Groggel (1983), Groggel et al. (1988), and Dubnicka (2004) for k-level nested designs. Kloeke et al. (2009) developed the asymptotic theory for the rank-based estimates of fixed effects using the general rank theory of Thompson (1990), and Brunner and Denker (1994). This study extends the theory of Kloeke et al. (2009) for these fixed effects estimation. Also, the theory of consistency of var-cov estimators in the study of Hettmansperger and McKean (2011) and Groggel (1983) is extended.

2.1.1 Pseudo-Sample Approach

Groggel (1983) and Groggel et al. (1988) suggested the pseudo-samples of observations that are asymptotically equivalent to the samples of the random and error effects formed

using either sample means or sample medians in the simple mixed model. In particular, to understand it, consider the one-way random effects model

$$y_{ij} = \alpha + a_i + \epsilon_{ij} \quad (2.1.1)$$

for $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, n_i$ (say, k schools, n_i students in each). a_i and ϵ_{ij} are random cluster effects and error effects, respectively. We only observe the values of y_{ij} , the variables a_i and ϵ_{ij} are not observable. We begin formation of the pseudo-samples based on means as location estimate by defining the pseudo-effects A_i and E_{ij} . We define $y_{i.} = n_i^{-1} \sum_j y_{ij}$, $y_{..} = N^{-1} \sum_i \sum_j y_{ij}$, $\bar{\epsilon}_{i.} = n_i^{-1} \sum_j \epsilon_{ij}$, $\bar{\epsilon}_{..} = N^{-1} \sum_i \sum_j \epsilon_{ij}$ and $\bar{a}_{..} = k^{-1} \sum_i a_i$. As $n_i \rightarrow \infty$ and $k \rightarrow \infty$, the sample $E_{ij} = y_{ij} - y_{i.} = \epsilon_{ij} - \bar{\epsilon}_{i.}$ behaves like the ϵ_{ij} and the sample $A_i = y_{i.} - y_{..} = a_i + \bar{\epsilon}_{i.} - \bar{a}_{..} - \bar{\epsilon}_{..}$ behaves like a_i because $\bar{\epsilon}_{i.}$, $\bar{\epsilon}_{..}$ and $\bar{a}_{..}$ converge in probability to zero. The pseudo-samples E_{ij} and A_i behave like independent samples of the errors. In particular, the variances of them are equal to the variances of the errors in the limiting case (See Chapter 3 for the theory and the consistency.)

For the 3-level nested structure, these pseudo random samples with means, school effects $A = \{A_i\}$, section effects $W = \{W_{ij}\}$ and error effects $E = \{E_{ijk}\}$, are defined as $A_i = y_{i..} - y_{...} = a_i + \bar{w}_{i.} + \bar{\epsilon}_{i..} - (\bar{a}_{..} + \bar{w}_{..} + \bar{\epsilon}_{...})$, $W_{ij} = y_{ij.} - y_{i..} = w_{ij} + \bar{\epsilon}_{ij.} - (\bar{w}_{.i} + \bar{\epsilon}_{i..})$, and $E_{ijk} = y_{ijk} - y_{ij.} = \epsilon_{ijk} - \bar{\epsilon}_{ij.}$. In case with covariates, y_{ijk} is replaced with $e_{ijk} = y_{ijk} - x_{ijk}^T \beta$. Here, the mean estimate can be replaced with another consistent location estimate with desired asymptotic properties (Groggel, 1983).

2.1.2 Rank Prediction Procedure (RPP)

For the k-level nested structure, we build pseudo-samples to predict random effects. For simplicity, much of our discussion is for a 3-level nesting model. But our results generalize to the k-level design. As described in the unpublished work by Terpstra & McKean (p.18), suppose that students' measurements are obtained from the sections. The sections are random and nested within schools, and we take the school effect as a

second random effect: I schools, J_i sections in school i , and n_{ij} students in section j of school i . Let y_{ijk} denote the response for the k^{th} student, in the j^{th} section of the i^{th} school and let x_{ijk} denote the vector of covariates. Then y_{ijk} follows the model

$$y_{ijk} = x_{ijk}^T \beta + a_i + w_{j(i)} + \epsilon_{ijk} \quad (2.1.2)$$

$k = 1, \dots, n_{ij}$; $j = 1, \dots, J_i$; $i = 1, \dots, I$, where a_i is the random effect for school i and $w_{j(i)}$ is the random effect for the j^{th} section of school i . That is, $\{a_i\}$ and $\{w_{j(i)}\}$ are the components of b in (3.1.1). Denote the variances of the random effects a_i and $w_{j(i)}$ by σ_a^2 and σ_w^2 , respectively. That is, $\theta = (\sigma_a^2, \sigma_w^2)^t$. Note that we can write the model as

$$y_{ijk} - x_{ijk}^T \beta = a_i + w_{j(i)} + \epsilon_{ijk} \quad (2.1.3)$$

Let $r_{ijk} = y_{ijk} - x_{ijk}^T \hat{\beta}$ denote raw residuals. Then, consider the model

$$r_{ijk} \doteq a_i + w_{j(i)} + \epsilon_{ijk} \quad (2.1.4)$$

With i and j fixed, this is a simple location model; hence, an estimate of that location is $\hat{u}_{ij} = \text{med}_k \{r_{ijk}\}$, which predicts $a_i + w_{j(i)}$. To separate a_i and $w_{j(i)}$, consider

$$\hat{u}_{i.} = \text{med}_{J_i} \{\hat{u}_{i1}, \hat{u}_{i2}, \dots, \hat{u}_{iJ_i}\}.$$

Then, for $j = 1, \dots, J_i$, the difference $\hat{u}_{ij} - \hat{u}_{i.}$ is free of a_i and, hence, is a predictor of $w_{j(i)}$. That is, the prediction of $w_{j(i)}$ is

$$\hat{w}_{j(i)} = \hat{u}_{ij} - \hat{u}_{i.}.$$

Finally, move this prediction to the left side of equation (2.1.4) to obtain the model

$$r_{ijk} - \hat{w}_{j(i)} \doteq a_i + \epsilon_{ijk}.$$

This simple location model yields as the prediction of a_i ,

$$\hat{a}_i = \text{med}_{jk}\{r_{ijk} - \hat{w}_{j(i)}\}.$$

Proceeding over all sections, we obtain the predictions of the random effects. Since we used the median as the location predictor, we adjust these with their common medians. $\hat{a}_i = \hat{a}_i - \text{med}_s\{\hat{a}_s\}$ and $\hat{w}_{j(i)} = \hat{w}_{j(i)} - \text{med}_{s,t}\{\hat{w}_{t(s)}\}$. Next, define the vectors of random effects as $\hat{a} = (\hat{a}_1, \dots, \hat{a}_I)$, and $\hat{w} = (\hat{w}_{1(1)}, \dots, \hat{w}_{J(I)})$. Hence, $\text{Disp}^2(\hat{a})$ and $\text{Disp}^2(\hat{w})$ serve as robust estimators of the variance components σ_a^2 and σ_w^2 , respectively. The RPP algorithm needs residuals from rank-based fittings for predictions of the random effects in random nested models. This recursive algorithm can handle a general number of nestings in a hierarchical mixed model as above. We discuss several location and scale estimators in the next section. Note that the pseudo-sample A 's and E 's are based on data, but for large sample sizes, A_i behaves as a_i and E_{ij} behaves as ϵ_{ij} . Groggel (1983) used these pseudo-samples in this sense. It is similar to the first order analysis of robust residuals in McKean et al. (1990). We connect these two approaches (See Groggel, 1983 and McKean et al., 1990).

2.1.3 Location Estimators: Hodges-Lehmann and Median

The $n(n+1)/2$ pairwise averages $\{(x_i + x_j)/2 : 1 \leq i \leq j \leq n\}$ are called Walsh averages Hodges & Lehmann (1963). The estimate of location parameter in one sample is the median of the Walsh averages, called the Hodges-Lehmann estimate (HL). Another estimate of location in rank prediction algorithm is the median estimate.

2.1.4 Scale Estimators

The fitted model in a Rank-based method yields residuals that inherit all random effects. We use them to estimate the variance components for each type of errors in

nested levels. The dispersion function is defined as

$$D(e) = \sum_i^n a(R(e_i))e_i. \quad (2.1.5)$$

The scores $a(k)$ are generated from the score function, $\varphi[k/(n+1)]$, where $\varphi(u)$ is a specified nondecreasing, square integrable function on interval $(0, 1)$. One of the robust candidates for scale parameter estimators is

$$Disp(e) = \frac{1}{n}D(e) = \frac{1}{n} \sum_{i=1}^n a(R(e_i)) \cdot e_i. \quad (2.1.6)$$

The other scale estimator is the median absolute deviation (MAD) defined as

$$MAD(e) = 1.483 \text{med}_i |e_i - \text{med}_j \{e_j\}|.$$

Also, Hettmansperger & McKean (2011) defines the functional corresponding to the dispersion for Wilcoxon scores for a random vector e defined as

$$Disp(e) = \frac{2\sqrt{\pi}}{n} \sum_{i=1}^n \left(\frac{R(e_i)}{n+1} - \frac{1}{2} \right) \cdot e_i. \quad (2.1.7)$$

This is a consistent estimator of the scale parameter σ_e when the errors have a normal distribution. Kloeke et al. (2009) suggest that the MAD is a consistent estimator for scale parameter in clustered correlated design.

2.1.5 Intra- and Inter- Correlation Coefficients

In a 3-level nested study (See Model 2.1.3), to measure each level's contribution to the variability, three intra-class correlations are defined as intra-error, $\frac{\sigma_e^2}{\sigma_e^2 + \sigma_{sch}^2 + \sigma_{sect(school)}^2}$, intra-school, $\frac{\sigma_{sch}^2}{\sigma_e^2 + \sigma_{sch}^2 + \sigma_{sect(school)}^2}$, and intra-section, $\frac{\sigma_{sect(school)}^2}{\sigma_e^2 + \sigma_{sch}^2 + \sigma_{sect(school)}^2}$. For a general number of nestings in a purely hierarchical model (k-level nested), these are defined in this sense.

2.2 Rank-Based Analysis

As discussed below, the rank-based analysis, the least squares (LS) Euclidean norm is replaced with another norm; hence, the geometry and intuition remain the same as in the traditional analysis. For example, in tests of linear hypotheses the reduction in sums of squares in the least squares based analysis is replaced by the reduction in dispersion of the rank-based analysis. The power of rank-based approach is insensitive to normality assumptions. The analysis doesn't require balanced design. It is robust to outliers in response space by choosing other score functions and rank estimators such as GR, WIL and Sign, and simple weighting schemes (HBR) yield resistance to outliers in factor space. Choosing score function depending on the information about distribution gives more powerful results. Rank-based approach is described in independent linear regression models, simple mixed models, models with correlated error structure in Hettmansperger & McKean (2011). The rank-based norm is

$$\|w\|_{\varphi} = \sum_i^n a((R(w_i))w_i, \quad (2.2.1)$$

$w \in R^n$, where the scores are generated as $a(i) = \varphi[i/(n+1)]$ for a nondecreasing function $\varphi(u)$, defined on the interval $(0, 1)$, and $R(w_i)$ is the rank of w_i among w_1, w_2, \dots, w_n . We assume without loss of generality that the scores sum to zero. Two of the most popular score functions are the Wilcoxon $\varphi(u) = \sqrt{12} \cdot (u - \frac{1}{2})$ and the sign $\varphi(u) = \text{sgn}[u-1/2]$. The joint rank-based estimate of β for the independent error model $Y = X\beta + e$ is given by

$$\hat{\beta}_{\varphi} = \text{Argmin} \|Y - X\beta\|_{\varphi} \quad (2.2.2)$$

Let $f(t)$ denote the pdf of error terms. In the iid linear model case, under regularity conditions

$$\hat{\beta} \sim N(\beta, \tau_{\varphi}^2 (X^T X)^{-1}), \quad (2.2.3)$$

$$\tau_{\varphi} = [\int \varphi(u) \varphi_f(u) du]^{-1}, \quad (2.2.4)$$

with $\varphi_f(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}$ (See Chapter 3, Hettmansperger & McKean, 2011). The parameter τ_φ is a scale parameter for the error terms e_i .

The estimator of the fixed effects is asymptotically normal shown in the study of Jaeckel (1972), Jureckova (1971), McKean & Hettmansperger (1976). Kloeke et al. (2009) obtained the asymptotic theory for the simple mixed model. In the next chapter, we obtain the asymptotic theory for the GR estimator of fixed effects in the hierarchical model and, further, we obtain the consistency of our robust variance component estimators.

We propose three new rank-based methods, Joint Ranking (JR), Generalized Ranking (GR) and Rank-based Generalized Estimating Equations (GEER), which iteratively obtain robust estimation for both the fixed and random effects, thus, intra-class correlation coefficients. These methods employ the proposed algorithm, called *Rank Prediction Procedure* (RPP), for predicting random errors and effect, and scale parameters. In particular, main interest is to estimate β as fixed effects using these three Rank-based methods, b as random effects and σ_l^2 's as scale parameters using the RPP with two alternative estimators. Hence, intra-class correlation coefficients for each level is estimated.

Chapter 3

Ranked-Based Estimate and Asymptotic Theory

In this chapter, we discuss the three Rank-based methods along with the random effect prediction algorithm for the k -level nested random effects model.

3.1 Model and Assumptions

Consider the general mixed model,

$$Y = X\beta + e = X\beta + Zb + \epsilon, \quad (3.1.1)$$

where Y denotes a $n \times 1$ vector of responses, X is a $n \times p$ known fixed effects design matrix, β is a $p \times 1$ fixed effects parameter vector, Z is a $n \times k$ known random effects design matrix, b is a $k \times 1$ vector of random effects, and ϵ is a $n \times 1$ vector of random errors.

We will make the following model assumptions: Random effects of each nested level are independent and identically distributed (iid) with mean 0 and variance σ_l^2 , $(0, \sigma_l^2)$, for levels $l = 1, \dots, k$. The components of ϵ are iid $(0, \sigma_\epsilon^2)$ and ϵ and b are independent. Hence, responses from individuals nested within a cluster are considered to be correlated while responses from individuals nested within highest level cluster are not.

The models with independent vector of errors e_i and three-level nested design can be

expressed as

$$Y_i = \alpha \mathbf{1}_{n_i} + X_i^T \beta + e_i \quad (3.1.2)$$

$$y_{ijk} = x_{ijk}^T \beta + a_i + w_{j(i)} + \epsilon_{ijk}, \quad (3.1.3)$$

where $X_i = [X_{i1}^T, X_{i2}^T, \dots, X_{iJ_i}^T]^T$ is $n_i \times p$ design and covariate matrix for school i , $Y_i = [Y_{i1}^T, Y_{i2}^T, \dots, Y_{iJ_i}^T]^T$ is $n_i \times 1$ response matrix for school i , the sum of observation in school i is $n_i = \sum_{j=1}^{J_i} n_{ij}$, and $N = \sum_{i=1}^I n_i$ is the sum of all observations. Schools, highest clusters, are independent. Also, the three-level nested error model is expressed as

$$e_{ijk} = a_i + w_{j(i)} + \epsilon_{ijk}, \quad k = 1, \dots, n_{ij}; j = 1, \dots, J_i; i = 1, \dots, I, \quad (3.1.4)$$

where ϵ_{ijk} is error, a_i is the random effect for school i and $w_{j(i)}$ is the random effect for the j^{th} section of school i . That is, $\{a_i\}$ and $\{w_{j(i)}\}$ are the components of b in (1.2.1). Denote the variances of the random effects a_i and $w_{j(i)}$ by σ_a^2 and σ_w^2 , respectively. Then the vector of variance components is $\theta = (\sigma_\epsilon^2, \sigma_a^2, \sigma_w^2)^t$.

3.2 Joint Ranking Method (JR)

Consider the Rank-based estimates for fitting linear models with dependent errors. The asymptotic theory for the clustered correlated model and simple mixed model in rank-based analysis is presented in detail in Chapter 5 of the book of Hettmansperger and McKean (2011). It is based on the work of Brunner & Denker (1994) and Kloeke et al. (2009). Here we only briefly summarize and discuss the modified form of the theorems in terms of the random nested model in (3.1.3). Also, we derive the standard errors of the fixed estimate.

The asymptotic linearity and quadratic dispersion for our case are derived in a manner similar to that of the study in the independent case. The theory for the Joint Rank (JR) estimate assumes that the marginal cdfs are the same; see Kloeke et al. (2009), which is true for our hierarchical model. We write the general result for the k -level design and

specialize it to the three-level nested model. Then, we obtain the asymptotic variance-covariance matrix of $\hat{\beta}_{JR}$, a consistent estimator of the asymptotic variance-covariance matrix of β_{JR} , and the Studentized residuals. Then, after fitting the model of interest, random effects and scale estimates are estimated via the RPP algorithm found in Section 2.1.

Assume Model (3.1.2) is true. Under the required assumptions, J1-J6, listed in the Appendix, consider Jaekel's (1972) dispersion function defined in terms of the pseudo-norm

$$D(\beta) = \|Y - X\beta\|_{\varphi} = \sum_{i=1}^I a \left[R(Y_i - X_i^T \beta) \right] \cdot (Y_i - X_i^T \beta). \quad (3.2.1)$$

This is the objective function. Since it is convex, the estimator of the fixed effects β is given by the value that minimizes $D(\beta)$ i.e.,

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

The negative of the gradient of $D(\beta)$ is

$$S(\beta) = X^T \cdot a[R(Y - X\beta)].$$

The scores are generated via a score function as $a(k) = \varphi[k/(N + 1)]$, where $\varphi(u)$ is a specified nondecreasing, square integrable function on interval $(0, 1)$. Without loss of generality, it is standardized so that $\int_0^1 \varphi(u) du = 0$ and $\int_0^1 \varphi^2(u) du = 1$.

Since the estimate is location and scale equivariant for the theory, without loss of generality, the true parameters are assumed to be zero. And our theory is for general score. The projection of the gradient is a random vector, with the covariance,

$$S_x(\beta) = X' \varphi[F(e)] \quad (3.2.2)$$

$$\text{cov}(S_x(\beta)) = X' \text{cov}(\varphi[F(e)]) X = \sum_{i=1}^I X_i' \Sigma_{\varphi,i} X_i,$$

where $\Sigma_{\varphi,i}$ is given by $\Sigma_{\varphi,i} = \text{cov}(\varphi(F(e_i)))$; see Klocke et al (2009). The linearity result is given by

$$S_x(\beta) = S_x(0) - \tau_\varphi^{-1} N^{-1} (X'X)^{-1} X' \beta + o_p(N^{1/2}) \quad (3.2.3)$$

for $N^{1/2} \|\beta\| \leq c$, $\forall c > 0$, where N is the total sample size. This leads to the asymptotic representation

$$N^{1/2} \hat{\beta} = \tau_\varphi N^{1/2} (X'X)^{-1} X' \varphi[F(e)] + o_p(1). \quad (3.2.4)$$

Thus the asymptotic variance of $\hat{\beta}_{JR}$ is

$$\text{var}(\sqrt{N} \hat{\beta}_{JR}) \doteq \tau_\varphi^2 (X'X)^{-1} (\lim_{I \rightarrow \infty} \sum_{i=1}^I X_i' \Sigma_{\varphi,i} X_i) (X'X)^{-1}, \quad (3.2.5)$$

where I is the number of independent clusters. In practice, it is expressed as

$$\hat{\beta}_{JR} \doteq \tau_\varphi (X'X)^{-1} X' \varphi[F(e)] \quad (3.2.6)$$

and

$$\text{var}(\hat{\beta}_{JR}) \doteq \tau_\varphi^2 (X'X)^{-1} (\Sigma_\varphi) (X'X)^{-1}, \quad (3.2.7)$$

where $\lim_{I \rightarrow \infty} N^{-1} \sum_{i=1}^I X_i' \Sigma_{\varphi,i} X_i \doteq \sum_{i=1}^I X_i' \Sigma_{\varphi,i} X_i = \Sigma_\varphi$ and τ_φ is defined in expression (2.2.4).

These results hold in general for the k -level nested designs with I independent clusters with k -level subclusters. Our model of interest is the 3-level nested model in (3.1.3). Note that the dispersion function can be written as

$$D_{JR}(\beta) = \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} a \left[R(y_{ijk} - x_{ijk}^T \beta) \right] \cdot (y_{ijk} - x_{ijk}^T \beta) \quad (3.2.8)$$

and that the gradient function is given by

$$S(\beta) = X^T \cdot a(R(Y - X\beta)) = \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} x_{ijk}^T a[R(y_{ijk} - x_{ijk}^T \beta)]. \quad (3.2.9)$$

Assume var-cov of $\varphi[F(e_i)]$, that is $\Sigma_{\varphi,i}$, and var-cov of e_i , that is Σ_i , exist. Covariance structure for the errors are selected as desired; i.e., structured (S), compound symmetry (CS), autoregressive-one AR(1), working independence (WI), or unstructured (UN). For example, in a three-level case, the three-parametered CS structures for $cov(\varphi[F(e)])$ and $cov(e)$ are defined and estimated as follows:

$$\begin{aligned} cov(\varphi[F(e)]) &= \sigma_\varphi^2 \cdot B(1, \rho_{\varphi,sec}, \rho_{\varphi,sch}) \\ &= \sigma_\varphi^2 \cdot diag\{B_1(1, \rho_{\varphi,sec}, \rho_{\varphi,sch}), B_2(1, \rho_{\varphi,sec}, \rho_{\varphi,sch}), \\ &\quad \dots, B_I(1, \rho_{\varphi,sec}, \rho_{\varphi,sch})\}_{N \times N}, \end{aligned} \quad (3.2.10)$$

where $B(1, \rho_{\varphi,sec}, \rho_{\varphi,sch})$ is the block-diagonal matrix in $N \times N$, and B_i 's are block matrices in $n_{i.} \times n_{i.}$ defined as $B_i(1, \rho_{\varphi,sec}, \rho_{\varphi,sch}) = cov(\varphi[F(e_i)])$, where $\rho_{\varphi,sch}$'s are off-block diagonal entries while $cov(\varphi[F(e_{ij})])$'s are block-diagonal matrices in $n_{ij} \times n_{ij}$. Since the scores are standardized and F is the cdf of e_{ijk} ,

$$var(\varphi[F(e_{ijk})]) = 1.$$

Also, random errors from different highest clusters are independent, so

$$cov(\varphi[F(e_{ijk})], \varphi[F(e_{i'j'k'})]) = 0.$$

For notation, define $\rho_{\varphi,sec}$ and $\rho_{\varphi,sch}$ as

$$cov(\varphi[F(e_{ijk})], \varphi[F(e_{ijk'})]) = E\{\varphi[F(e_{ijk})]\varphi[F(e_{ijk'})]\} = \rho_{\varphi,sec}$$

$$\text{and } cov(\varphi[F(e_{ijk})], \varphi[F(e_{i'j'k'})]) = E\{\varphi[F(e_{ijk})]\varphi[F(e_{i'j'k'})]\} = \rho_{\varphi,sch}.$$

Define intra-class correlation coefficients, $\rho_{sec(sch)}$ and ρ_{sch} , found in Section 2.1.5. For the error structure,

$$cov(e) = \Sigma = \sigma^2 \cdot B(1, \rho_{sec(sch)}, \rho_{sch}) \quad (3.2.11)$$

is a $N \times N$ block-diagonal matrix, $B = B(1, \rho_{sec(sch)}, \rho_{sch})$ is a block-diagonal matrix defined with $\rho_{sec(sch)}$ and ρ_{sch} instead of $\rho_{\varphi, sec}$ and $\rho_{\varphi, sch}$ as in (3.2.10), with the notations $var(e_{ijk}) = \sigma_e^2 + \sigma_a^2 + \sigma_w^2 = \sigma^2$, $cov(e_{ijk}, e_{ijk'}) = \sigma_a^2 + \sigma_w^2$, and $cov(e_{ijk}, e_{ij'k'}) = \sigma_a^2$. Let \hat{e}_{ijk} denote the residuals, i.e., $\hat{e}_{ijk} = y_{ijk} - x_{ijk}^t \hat{\beta}_{JR}$. Estimates of $\rho_{\varphi, sec}$ and $\rho_{\varphi, sch}$ are calculated using the simple moment estimator,

$$\hat{\rho}_{\varphi, sch} = \frac{1}{A} \sum_{i=1}^I \sum_{j>j'}^{J_i} \sum_{k=1}^{n_{ij}} (a[R(\hat{e}_{ijk})])(a[R(\hat{e}_{ij'k})]), \quad A = \sum_{i=1}^I \sum_{j=1}^{J_i} n_{ij} n_{ij} - p$$

$$\text{and } \hat{\rho}_{\varphi, sec} = \frac{1}{B} \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k>k'}^{n_{ij}} (a[R(\hat{e}_{ijk})])(a[R(\hat{e}_{ij'k'})]), \quad B = \sum_{i=1}^I \sum_{j=1}^{J_i} n_{ij}(n_{ij} - 1) - p.$$

The intercept estimator, which is location and scale equivariance, is obtained from a consistent location estimator. One estimator is the median of the residuals; i.e., $\hat{\alpha} = \text{med}_{i,j,k}\{\hat{e}_{ijk}\}$. Its asymptotic representation is given by

$$\hat{\alpha} = \tau_s \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} \text{sgn}(e_{ijk}) + o_p(1/\sqrt{N}). \quad (3.2.12)$$

The asymptotic variance of the estimator is expressed as

$$\begin{aligned} var(\hat{\alpha}) &= var(\text{med}\{\hat{e}_{ijk}\}) = \tau_s^2 \frac{1}{N^2} \sum_{i=1}^I var\left(\sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} \text{sgn}(e_{ijk})\right) \\ &= \tau_s^2 \frac{1}{N^2} \sum_{i=1}^I \left\{ \sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} var(\text{sgn}(e_{ijk})) + \sum_{j=1}^{J_i} \sum_{k \neq k'}^{n_{ij}} cov(\text{sgn}(e_{ijk}), \text{sgn}(e_{ij'k'})) \right. \\ &\quad \left. + \sum_{j \neq j'}^{J_i} \sum_{k=1}^{n_{ij}} cov(\text{sgn}(e_{ijk}), \text{sgn}(e_{ij'k})) \right\}. \end{aligned}$$

In this 3-level case, letting $\tau_s = \frac{1}{2f(0)}$, this variance is

$$\begin{aligned}
\text{var}(\hat{\alpha}) &\doteq \tau_s^2 \frac{1}{N^2} \sum_{i=1}^I \left\{ n_i + \sum_{j=1}^{J_i} n_{ij}(n_{ij} - 1) \text{cov}(\text{sgn}(e_{ij1}), \text{sgn}(e_{ij2})) \right. \\
&\quad \left. + \sum_{j=j'}^{J_i} n_{ij} n_{ij'} \text{cov}(\text{sgn}(e_{ij1}), \text{sgn}(e_{ij'1})) \right\} \\
&\doteq \tau_s^2 \frac{1}{N^2} \left\{ N + \sum_{i=1}^I \sum_{j=1}^{J_i} n_{ij}(n_{ij} - 1) \text{cov}(\text{sgn}(e_{ij1}), \text{sgn}(e_{ij2})) \right. \\
&\quad \left. + \sum_{i=1}^I \sum_{j=j'}^{J_i} n_{ij} n_{ij'} \text{cov}(\text{sgn}(e_{ij1}), \text{sgn}(e_{ij'1})) \right\} \\
&\doteq \tau_s^2 \frac{1}{N} \left\{ 1 + \frac{k_1}{N} c_1 + \frac{k_2}{N} c_2 \right\}, \tag{3.2.13}
\end{aligned}$$

where $k_1 = \sum_{i=1}^I \sum_{j=1}^{J_i} n_{ij}(n_{ij} - 1)$, $k_2 = \sum_{i=1}^I \sum_{j=j'}^{J_i} n_{ij} n_{ij'}$, and $c_1 = \text{cov}(\text{sgn}(e_{ij1}), \text{sgn}(e_{ij2}))$, $c_2 = \text{cov}(\text{sgn}(e_{ij1}), \text{sgn}(e_{ij'1}))$. Consistent estimators for c_1 and c_2 are simple weighted moment estimators with school sample size. In the simulations, we use the weighted simple moment estimator with the sample size of school. As an alternative to this estimator, we can adjust the scores with its school averages by subtracting. The estimate of the scale parameter τ_φ is obtained from $\tau_\varphi = [\int \varphi(u) \varphi_f(u) du]^{-1}$ and $\varphi_f(u) = -f'(F^{-1}(u))/f(F^{-1}(u))$ (See Koul et al., 1987; Hettmansperger & McKean, 2011).

3.2.1 Studentized Residuals

As described in Hettmansperger and McKean (2012, Chapter 3), using the error model $\hat{e}_{JR} = Y - \hat{\alpha}_S 1_N - X \hat{\beta}_\varphi$, the asymptotic representation of $\hat{\alpha}_S$ in (3.2.12), and $\hat{\beta}_{JR}$ in (3.2.4)

for (3.1.4), we can write the residuals \hat{e}_{JR} to first order approximation s.t.

$$\hat{e}_{JR} = Y - \hat{\alpha}_S \mathbf{1}_N - \mathbf{X} \hat{\beta}_\varphi \quad (3.2.14)$$

$$\begin{aligned} &= Y - \frac{\tau_S}{N} \mathbf{1}' \varphi_S[\mathbf{F}(\mathbf{Y})] \mathbf{1} - \mathbf{X} \tau_\varphi (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X} \varphi[\mathbf{F}(\mathbf{e})] + \beta_0 \\ &= (Y - X \beta_0) - \frac{\tau_S}{N} \sum_{i=1}^{e_i} \text{sgn}(e_i) \mathbf{1} - \mathbf{X} \tau_\varphi (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X} \varphi[\mathbf{F}(\mathbf{e})] + \beta_0 \\ &= e - \tau_S \overline{\text{sgn}(e)} \mathbf{1} - \tau_\varphi \mathbf{H} \varphi[\mathbf{F}(\mathbf{e})] \\ &= e - \frac{\tau_S}{N} \mathbf{1}' \text{sgn}(\mathbf{e}) \mathbf{1} - \tau_\varphi \mathbf{H} \varphi[\mathbf{F}(\mathbf{e})]. \end{aligned} \quad (3.2.15)$$

Then,

$$\begin{aligned} \text{cov}(\hat{e}_{JR}) &= E((e - \frac{\tau_S}{N} \mathbf{1}' \text{sgn}(\mathbf{e}) \mathbf{1} - \tau_\varphi \mathbf{H} \varphi[\mathbf{F}(\mathbf{e})])(e - \frac{\tau_S}{N} \mathbf{1}' \text{sgn}(\mathbf{e}) \mathbf{1} - \tau_\varphi \mathbf{H} \varphi[\mathbf{F}(\mathbf{e})])') \\ &= E(ee') + \frac{\tau_S}{N} E(e \mathbf{1}' (\mathbf{1}' \text{sgn}(\mathbf{e}))') - \tau_\varphi \mathbf{E}(e \varphi[\mathbf{F}(\mathbf{e})]') \mathbf{H}' \\ &\quad - \frac{\tau_S}{N} E(\mathbf{1}' \text{sgn}(\mathbf{e}) \mathbf{1} e') + \frac{\tau_S^2}{N^2} \mathbf{E}(\mathbf{1}' \text{sgn}(\mathbf{e}) \mathbf{1} \mathbf{1}' (\mathbf{1}' \text{sgn}(\mathbf{e}))') \\ &\quad + \frac{\tau_S \tau_\varphi}{N} E(\mathbf{1}' \text{sgn}(\mathbf{e}) \mathbf{1} \varphi[\mathbf{F}(\mathbf{e})]') \mathbf{H}' - \tau_\varphi \mathbf{H} \mathbf{E}(\varphi[\mathbf{F}(\mathbf{e})] e') \\ &\quad + \frac{\tau_S \tau_\varphi}{N} H E(\varphi[\mathbf{F}(\mathbf{e})] \mathbf{1}' (\mathbf{1}' \text{sgn}(\mathbf{e}))') + \tau_\varphi^2 \mathbf{H} \mathbf{E}(\varphi[\mathbf{F}(\mathbf{e})] \varphi[\mathbf{F}(\mathbf{e})]') \mathbf{H}', \end{aligned} \quad (3.2.16)$$

where H is the projection matrix onto the the space spanned by the centered design matrix X . For $t = 1, 2, \dots, N$, the joint rank-based t -th Studentized marginal residual is

$$e_{JR,t}^* = \hat{e}_{JR,t} / \sqrt{\text{cov}(\hat{e}_{JR})_{t,t}}.$$

Using the similar approach in Hettmansperger and McKean (2012, Chapter 5) for the estimate of $\text{cov}(\hat{e}_{JR})$ with one level dependent error structure, we need some new parameters, which can be estimated as in (3.2.10), and notations defined as

$$E(ee') = \sigma^2 B(1, \rho_{sec}, \rho_{sch})$$

$$E(e\varphi[F(e)]') = B(E(e_{111}, \varphi[F(e_{111})]), E(e_{111}, \varphi[F(e_{112})]), E(e_{111}, \varphi[F(e_{121})]))H$$

$$E(\varphi[F(e)]\varphi[F(e)]') = HB(1, \rho_{\varphi, sec}, \rho_{\varphi, sch})H$$

$$E(e_{111}\varphi[F(e_{111})']) = \frac{1}{N} \sum_{i=1}^N e_{ijk}\varphi[F(e_{ijk})]$$

Likewise, $E(1'sgn(e)1\varphi[F(e)'])$ can be obtained from

$$\begin{aligned} E(1'sgn(e)1\varphi[F(e)']) &= \mathbf{11}'B(E(sgn(e_{111}), \varphi[F(e_{111})]), \dots \\ &\quad E(sgn(e_{111}), \varphi[F(e_{112})]), \dots \\ &\quad E(sgn(e_{111}), \varphi[F(e_{121})]))H \end{aligned}$$

3.2.2 General Linear Hypothesis

In regard to testing interest, the asymptotic distribution of $\hat{\beta}_{JR}$ in (3.2.6) suggests a Wald type test of the general null hypothesis in the form of $H_0 : H_{q \times p}\beta = 0$ based on the test statistics $T_\varphi = (H\hat{\beta})^T(H\hat{V}_{JR}H)(H\hat{\beta})$, which has an asymptotic distribution of $\chi^2(q)$. Under the local alternatives, noncentrality parameter is given by $\eta = (H\beta)^T(HV_{JR}H)(H\beta)$. (An alternative test is the reduction of dispersion test; see Hettmansperger and McKean, 2011).

3.3 Generalized Rank Method (GR)

In this section, the k-step estimate algorithm in the generalized rank method for the mixed models is introduced. This is an iteratively reweighted rank method based on the Newton-type approximation. McKean and Hettmansperger (1978) developed the asymptotic properties of linearized rank estimators for use in the linear model with the k-step Gauss-Newton approximation without reweighing, and it was proposed for independent case. Here, we extend this theory to the k-step GR method in the dependent case of mixed models; i.e., k-level nested models. It suffices to discuss the asymptotic linearity of the estimate for first step, because the argument is the same for each step. The theory

for one-step estimate suffices to show the behavior of the k-step iterations. Also, after the first step, the estimates are asymptotically equivalent to the independent case because residuals are no longer dependent due to reweighing with variance-covariance matrix.

3.3.1 GR Algorithm

Consider the general mixed model in (3.1.1). The proposed Iteratively Reweighted Generalized Rank-based algorithm (GR) is as follows.

- (0) Set $l = 0$
- (1) Obtain $\hat{\beta}^{(l)}$ by fitting $Y^* = \hat{\Sigma}_Y^{-\frac{1}{2}} \cdot Y$ to $X^* = \hat{\Sigma}_Y^{-\frac{1}{2}} \cdot X$ using a rank-based regression estimate. If $l = 0$ then use $\hat{\Sigma}_Y = I_n$; otherwise use $\hat{\Sigma}_Y = \Sigma_Y(\hat{\theta}^{(l-1)})$.
- (2) Use $\hat{\beta}^{(l)}$ to calculate the residuals, $\hat{e}^{(l)} = Y - X\hat{\beta}^{(l)}$.
- (3) Use $\hat{e}^{(l)}$ to obtain $\hat{b}^{(l)}$, the predictor of b .
- (4) Use $\hat{b}^{(l)}$ to estimate the variance components, $\hat{\theta}^{(l)}$.
- (5) If $\|\hat{\beta}^{(l)} - \hat{\beta}^{(l-1)}\| < TOL$ and $\|\hat{\theta}^{(l)} - \hat{\theta}^{(l-1)}\| < TOL$ then stop; let $\hat{\beta} = \hat{\beta}^{(l)}, \hat{\theta} = \hat{\theta}^{(l)}$ and $\hat{b} = \hat{b}^{(l)}$. Otherwise, let $l = l + 1$ and return to step 1.

The estimators of the asymptotic variance-covariance matrix of $\hat{\beta}_{GR}$ require consistent τ_S , τ_φ and Σ_Y which are obtained from the current estimate of weighted errors. When we specialize to the model of interest, the three level nested model, the estimate of Σ_Y is obtained via the RPP algorithm. The next section discusses the asymptotic theory and consistency of these estimates.

3.3.2 Pseudo-Estimates for the k-Step Estimator

Using the same assumptions as in JR, the model is

$$Y_i = \alpha 1_i + X_i \beta + e_i, \quad i = 1, 2, \dots, I.$$

Let $Y_i^* = \hat{\Sigma}_i^{-\frac{1}{2}} \cdot Y_i$, $X_i^* = \alpha \hat{\Sigma}_i^{-\frac{1}{2}} 1_i + \hat{\Sigma}_i^{-\frac{1}{2}} \cdot X_i$ and $e_i^* = \hat{\Sigma}_i^{-\frac{1}{2}} \cdot e_i$. $\hat{\Sigma}_i$ is the variance-covariance matrix of \hat{e}_i . Let $\beta^* = (\alpha, \beta)'$. New model is now the model passing through the origin

defined as

$$Y_i^* = X_i^* \beta^* + e_i^*. \quad (3.3.1)$$

It will be convenient to express the rank-based pseudo-estimate, which is free of the intercept estimate, after reweighting the first step in the next theorem.

Theorem 1. *The asymptotic representation of $\hat{\beta}^*$ is*

$$\tau_S(X^{*T} X^*)^{-1} X^{*T} H_1 \text{sgn}(e^*) + \tau_\varphi(X^{*T} X^*)^{-1} X^{*T} H_X \varphi[F(e^*)] + o_p(1/\sqrt{N}).$$

Then, the asymptotic pseudo-estimate and distribution are

$$\begin{aligned} \hat{\beta}_{GR} &\doteq \tau_S(X^{*T} X^*)^{-1} X^{*T} H_1 \text{sgn}(e^*) + \tau_\varphi(X^{*T} X^*)^{-1} X^{*T} H_X \varphi[F(e^*)] \\ &\quad + o_p(1/\sqrt{N}) \end{aligned} \quad (3.3.2)$$

$$\hat{\beta}_{GR} \doteq \tau_S(X^{*T} X^*)^{-1} X^{*T} H_1 \text{sgn}(e^*) + \tau_\varphi(X^{*T} X^*)^{-1} X^{*T} H_X \varphi[F(e^*)] \quad (3.3.3)$$

with variance

$$\text{var}(\hat{\beta}_{GR}^*) \doteq (X^{*T} X^*)^{-1} X^{*T} \begin{bmatrix} 1_N & X_c^* \end{bmatrix} \begin{bmatrix} \frac{\tau_S^2}{N} & 0 \\ 0 & \tau_\varphi^2(X_c^{*T} X_c^*)^{-1} \end{bmatrix} \begin{bmatrix} 1_N & X_c^{*T} \end{bmatrix}^T X^* (X^{*T} X^*)^{-1}, \quad (3.3.4)$$

where $X_c^* = X^* - H_1 X^*$, H_1 is the projection matrix on the intercept space, H_x is the projection matrix on the centered design matrix X^* space.

The argument follows that the projection of Y_i^* on the space spanned by 1_{n_i} and $X_{c,i}^*$ is reprojected on the desired space which is spanned by X^* keeping the independent properties on the β distribution (See Chapter 3 in Hettmansperger & McKean, 2011). For the asymptotic representation, equations (3.2.4) and (3.2.12) are modified along with the projections.

3.3.3 One Step GR

Consider the general mixed model,

$$Y = X\beta + e. \quad (3.3.5)$$

X is $N \times p$ including the intercept design, Y is $N \times 1$. Assume Σ is the var-cov of e , and without loss of generality the true β is 0. Consider the model with true weights Σ

$$\Sigma^{-1/2}Y = \Sigma^{-1/2}X\beta + \Sigma^{-1/2}e$$

$$Y^* = X^*\beta + e^*. \quad (3.3.6)$$

Let $\hat{\beta}^{(0)}$ be the JR estimate on model (3.3.5). Then $\sqrt{N}\hat{\beta}^{(0)}$ is $O(1)$, and it satisfies the asymptotic properties developed for the JR (See Sections 3.2.2, 3.2.4) and 3.2.6). After obtaining the estimated weights $\hat{\Sigma}$ based on $\hat{\beta}^{(0)}$, the model is

$$\hat{\Sigma}^{-1/2}Y = \hat{\Sigma}^{-1/2}X\beta + \hat{\Sigma}^{-1/2}e$$

$$Y^{**} = X^{**}\beta + e^{**}. \quad (3.3.7)$$

Let $\hat{\beta}^{**}$ denote the first Newton step rank-based estimate based on Model (3.3.7).

$$\hat{\beta}^{**} = \hat{\beta}^{(0)} + \tau(X^{**'}X^{**})^{-1}S_{X^{**}}(\hat{\beta}^{(0)}).$$

Using the linearity and quadratic results of the JR development (Jureckova, 1971; Chapter 3 in the book of Hettmansperger and McKean, 2011), denote \tilde{b}^* the pseudo-estimate of the approximation of the dispersion function on Model (3.3.6) defined in (3.3.3). This model is independent case. It suffices to obtain the results of Theorem 3.5.4 and 3.5.5 in Hettmansperger and McKean (2011); i.e.,

$$\sqrt{N}(\hat{\beta}^{**} - \tilde{b}^*) \rightarrow 0 \text{ in probability.}$$

Note that

$$\sqrt{N}\tilde{b}^* = \tau(X^{*'}X^*)^{-1}S_{X^*}(0).$$

Then we have

$$\sqrt{N}(\hat{\beta}^{**} - \tilde{b}^*) = \sqrt{N}\hat{\beta}^{(0)} + \tau\sqrt{N}(X^{**'}X^{**})^{-1}S_{X^{**}}(\hat{\beta}^{(0)}) - \tau\sqrt{N}(X^{*'}X^*)^{-1}S_{X^*}(0).$$

Now, we will show that the right hand side is $o_p(1)$. Assume the design matrix conditions for X , X^* and X^{**} are satisfied. This includes the Huber's condition. The assumptions on the leverage values. Given $\delta > 0$, assume that N_0 is large enough

$$N \geq N_0, \quad \|X^{**} - X^*\|_F < \delta$$

uniformly, where $\|\cdot\|_F$ denote the Frobenius norm, defined as the square root of the sum of the absolute squares of its elements for a given matrix or vector. In particular, assume $|x_{ij}^{**} - x_{ij}^*| < \delta \forall i, j$ and $i, j = 1, \dots, N$. The responses are transformed too. So similarly we assume that N_0 is sufficiently large so that

$$Y_i^{**} - Y_i^* = o_p(1) \text{ for } i = 1, \dots, N, \quad (3.3.8)$$

and

$$F^{**}(t) - F^*(t) = o_p(1) \text{ for } N \geq N_0. \quad (3.3.9)$$

It is convenient to consider the transformation of the design matrix and parameters. Let F be cdf of errors. The new notations are defined as

$$\Delta = (X^{*'}X^*)^{1/2}\beta$$

$$C^* = X^*(X^{*'}X^*)^{-1/2} = \frac{1}{\sqrt{N}}X\Sigma^{-1/2}$$

$$c_i^* = \frac{1}{\sqrt{N}}x_i^* \text{ s.t. the sum is 0 in probability since } x_i^* \text{ are centered.}$$

$$c_i^{**} = \frac{1}{\sqrt{N}} x_i^{**}.$$

Appendix A.3 of Hettmansperger and McKean (2011) develops the asymptotic linearity and quadracity of the R estimates for linear models. Based on this development it suffices to obtain our result for the simple linear case, because the convexity arguments used in the appendix hold here. What we need to show is the equivalence of $S^*(\Delta)$ in terms of c_i^{**} and F^{**} following the next lemmas. By Appendix A.3 in Hettmansperger and McKean (2011),

$$S^*(\Delta) = \sum_i^N c_i^* \varphi[F^*(Y_i^*)] - \frac{1}{\tau} \Delta + o_p(1). \quad (3.3.10)$$

Lemma 2. *Using the notations for models (3.3.6) and (3.3.7),*

$$S^*(\Delta) = \sum_i^N c_i^{**} \varphi[F^*(Y_i^*)] - \frac{1}{\tau} \Delta + o_p(1).$$

Proof. Knowing (3.3.10), we can add and subtract $\sum_i^N c_i^{**} \varphi[F^*(Y_i^*)]$ to it s.t.

$$S^*(\Delta) = \sum_i^N c_i^{**} \varphi[F^*(Y_i^*)] + \sum_i^N (c_i^* - c_i^{**}) \varphi[F^*(Y_i^*)] - \frac{1}{\tau} \Delta + o_p(1). \quad (3.3.11)$$

Choose N so large that $|x_i^{**} - x_i^*| < \delta \ \forall i, i = 1, \dots, N$. This implies $|c_i^{**} - c_i^*| < \frac{\delta}{\sqrt{N}}$. We next show that the second term in (3.3.11) is $o_p(1)$ with the following remark. Then the desired result follows. \square

Remark 3. $\sum_i^N (c_i^* - c_i^{**}) \varphi[F^*(Y_i^*)]$ is $o_p(1)$. To show it, recall that c_i^{**} of the order $\frac{x_i}{\sqrt{N}}$ as in Liang & Zeger (1986). Assume that the weights satisfy the expression

$$c_i^{**} = c_i^* + o_p\left(\frac{1}{\sqrt{N}}\right).$$

Hence, for arbitrary $\delta > 0, N \geq N_0$,

$$\sum_i^N (c_i^{**} - c_i^*) \varphi[F^*(Y_i)] = \frac{\delta}{\sqrt{N}} \sum_i^N \varphi[F^*(Y_i)] = \frac{o_p(1)}{\sqrt{N}}.$$

Since $\varphi[F^*(Y_i)]$ are independent and identically distributed (iid) with mean 0 by the Central Limit Theorem, the above is $o_p(1)$.

Lemma 4.

$$S^*(\Delta) = \sum_i^N c_i^{**} \varphi[F^{**}(Y_i^{**})] - \frac{1}{\tau} \Delta + o_p(1).$$

Proof. Knowing Lemma 3.3.10, $S^*(\Delta) = \sum_i^N c_i^{**} \varphi[F^*(Y_i^*)] - \frac{1}{\tau} \Delta + o_p(1)$, we can rewrite the first term as

$$\sum_i^N c_i^{**} \varphi[F^*(Y_i^*)] = \sum_i^N c_i^{**} \varphi[F^{**}(Y_i^{**})] + \sum_i^N c_i^{**} \{\varphi[F^*(Y_i^*)] - \varphi[F^{**}(Y_i^{**})]\}.$$

The claim is that the second term is $o_p(1)$. To do this, we again use the assumptions of the weights. Assume $N \geq N_0$ so that

$$\frac{1}{\sqrt{N}} Y_i^{**} = \frac{1}{\sqrt{N}} Y_i^* + o_p\left(\frac{1}{\sqrt{N}}\right), \quad i = 1, \dots, N.$$

and

$$\frac{1}{\sqrt{N}} F^{**} = \frac{1}{\sqrt{N}} F^* + o_p\left(\frac{1}{\sqrt{N}}\right), \quad i = 1, \dots, N.$$

Because the composite function $\varphi \circ F$ is uniformly continuous, it follows from the expression (3.3.9), for any arbitrary $\delta > 0$, $\exists N_0$, $N \geq N_0$ and

$$\varphi[F^*(Y_i^*)] - \varphi[F^{**}(Y_i^{**})] < \delta.$$

Then,

$$\begin{aligned} \sum_i^N c_i^{**} \{\varphi[F^*(Y_i^*)] - \varphi[F^{**}(Y_i^{**})]\} &\leq \sum_i^N |c_i^{**}| \delta \\ &= \sum_i^N \frac{1}{\sqrt{N}} |x_i^{**}| \delta = \delta O_p(1) = o_p(1) \end{aligned}$$

where we have used the assumption that $\sum \frac{1}{\sqrt{N}} x_i^{**}$ is bounded. Hence, the proof is completed. \square

3.3.4 Studentized Residuals

For the GR, theoretically the transformed model is an independent model. Assume that the transformed errors are uncorrelated. Hence, the Studentized residual for independent case in Rank-based fitting is valid (See Hettmansperger and McKean, 2011, p. 222).

3.4 Generalized Estimating Equations Method (GEER)

Considering an alternative generalization of GLMs for correlated data in estimates, Abebe et al. (2010) implemented the general estimating equations (GEE) method of Liang and Zeger (1986) for simple mixed models, and derived the asymptotic normality of the rank estimators. Note that this estimator does not require the same marginal cdfs. Using the notations and approach of Abebe et al. (2010) for the model of interest in (3.1.3), this GEE method assumes that the marginal density of y_{ijk} , univariate variables, is of the exponential class of distributions and is given by

$$f(y_{ijk}) = \exp\{y_{ijk}\theta_{ijk} + a(\theta_{ijk}) + b(y_{ijk})\}\phi,$$

where $\theta_{ijk} = h(\eta_{ijk})$, $\eta_{ijk} = x_{ijk}\beta$, $\phi > 0$, thus, the first two moments of y_{ijk} are given by

$$E(y_{ijk}) = a'(\theta_{ijk}) \text{ and } \text{var}(y_{ijk}) = a''(\theta_{ijk}).$$

Any link function, $h^{-1}o(a')^{-1}$, can be used to make a link between the marginal mean $E(Y_i)$ and the design matrix X_i . For example, when the outcome is categorical, we can use the logit link function coupled with logistic model. To increase efficiency, the GEE takes the correlation into account employing the 'working' covariance of Y_i given by

$$V_i = A_i^{1/2} R_i A_i^{1/2}, \quad (3.4.1)$$

which is a $n_i \times n_i$ symmetric matrix (See Liang & Zeger, 1986; Abebe et al., 2010). Here, $A_i = \text{diag}\{a''(\theta_{i1}), a''(\theta_{i2}), \dots, a''(\theta_{in_i})\}$ and R_i is 'working' correlation matrix for school i . In particular, V_i need not be the variance of Y_i .

For Liang & Zeger's (1986) estimator, the general estimating equations are defined by

$$\sum_{i=1}^I D_i^T \hat{V}_i^{-1} [Y_i - a'(\theta_i)] = 0, \quad (3.4.2)$$

where $D_i = \frac{\partial a'(\theta_i)}{\partial \beta}$ is $n_i \times p$ Hessian matrix. When the identity link function $a(x)$ is assumed, $(\sum_{i=1}^I D_i^T V_i^{-1} D_i)^{-1} (\sum_{i=1}^I D_i^T V_i^{-1} Y_i)$ is the LS solution of the equation (3.4.2) for β . The iterative procedure works as follows: $V_i^{(0)}$ requires an initial estimate of fixed effect β and estimation of scale effects. After $V_i^{(0)}$ is estimated, the fixed effect $\beta^{(1)}$ is estimated as in (3.4.3). These iterations continue until a desired tolerance is achieved. Updated estimate can be expressed using the Gauss-Newton method as

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} - \left(\sum_{i=1}^I D_i^{T(k)} V_i^{-1(k)} D_i^{(k)} \right)^{-1} \left(\sum_{i=1}^I D_i^{T(k)} V_i^{-1(k)} (Y_i - a'(\theta_i)^{(k)}) \right). \quad (3.4.3)$$

The variance of $\hat{\beta}_{GEE}$ is

$$\left(\sum_{i=1}^I D_i^T V_i^{-1} D_i \right)^{-1} \left\{ \sum_{i=1}^I D_i^T V_i^{-1} \text{cov}(Y_i) V_i^{-1} D_i \right\} \left(\sum_{i=1}^I D_i^T V_i^{-1} D_i \right)^{-1}. \quad (3.4.4)$$

This method is called GEE, or Iteratively Reweighted Generalized LS estimates (IRLS).

Alternatively, we can rewrite expression (3.4.2) in terms of the Euclidean norm

$$D_{GEE}(\beta) = \sum_{i=1}^I (Y_i^* - d_i(\beta))^2, \quad (3.4.5)$$

where $n_i \times 1$ vectors $Y_i^* = \hat{V}_i^{-1/2} \cdot Y_i$ and $d_i(\beta) = \hat{V}_i^{-1/2} \cdot a'(\theta_i)$. Liang & Zeger (1986) showed that the fixed estimates $\hat{\beta}_{GEE}$, which is a nonlinear least squares (LS) estimator, is consistent and the consistent variance estimates \hat{V}_i are possible under the weak assumption that the estimated variance converges. Combining the results of two theories

from Liang & Zeger (1986) and Brunner & Denker (1994), the consistency of the estimate of Σ_i (the same notation with V_i in the GR) requires the cluster number $I \rightarrow \infty$, and the cluster size $n_{i.} \rightarrow \infty$. Consistency of scale estimates are discussed in Liang & Zeger (1986), saying that these do not depend on the correct choice of working variance-covariance matrix, so the consistency of Σ_i and Σ is not a matter of issue.

Abebe et al. (2010) developed a class of nonlinear robust estimators minimizing rank-based pseudo-norm of the residuals defined in (3.2.1) utilizing a 'working' covariance structure in the rank-based fitting as an analogue of the GEE.

Next, we will make the rank-based objective function analogous to the GEE in the model of interest, the 3-level nested model, for estimating fixed effects with the dispersion function

$$D_R(\beta) = \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} \varphi \left[\frac{R(y_{ijk}^* - d_{ijk}(\beta))}{N+1} \right] \cdot [y_{ijk}^* - d_{ijk}(\beta)], \quad (3.4.6)$$

where $d_{ijk}(\beta)$ is (jk)-th of the vector $d_i(\beta)$. We can obtain the R estimator and write (3.4.6) as to be in the form of (3.4.2), thus, these two representations will be used to employ to estimate and to derive the R asymptotic theory. The equivalence of these dual representations is shown by

$$\begin{aligned} D_R(\beta) &= \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} \varphi \left[\frac{R(e_{ijk}^*)}{N+1} \right] \cdot [e_{ijk}^*] \\ &= \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} \varphi \left[\frac{R(e_{ijk}^*)}{N+1} \right] \cdot [e_{ijk}^* - m(\beta)] \\ &= \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} \frac{\varphi \left[\frac{R(e_{ijk}^*)}{N+1} \right]}{e_{ijk}^* - m(\beta)} \cdot [e_{ijk}^* - m(\beta)]^2 \\ &= \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} w_{ijk}(\beta) [e_{ijk}^* - m(\beta)]^2 \\ &= \sum_{i=1}^I \sum_{j=1}^{J_i} \sum_{k=1}^{n_{ij}} [w_{ijk}^{1/2}(\beta) e_{ijk}^* - w_{ijk}^{1/2}(\beta) m(\beta)]^2 \end{aligned} \quad (3.4.7)$$

$$= \sum_{i=1}^I D_i^T V_i^{-1/2} W_i V_i^{-1/2} [Y_i - a'(\theta_i) - m^*(\beta)]^2, \quad (3.4.8)$$

where $e_{ijk}^* = y_{ijk}^* - d_{ijk}(\beta)$, $w_{ijk} = \frac{\varphi\left[\frac{R(e_{ijk}^*)}{N+1}\right]}{e_{ijk}^* - m(\beta)}$, $W_i = \text{diag}\{w_{ijk}\}$, $m(\beta) = \text{med}_{i,j,k}\{e_{ijk}^*\}$, $m^*(\beta) = m(\beta) \cdot V_i^{1/2} \cdot 1_{n_i}$. Assuming that the score function is odd about $\frac{1}{2}$, the use of the median in the weights ensures that the weights are non-negative. As discussed in Abebe et al. (2010) other score functions, not odd about $\frac{1}{2}$, can be used by changing to another quantile. The median is used for centering so the weights are skewed about 0. Hence, the weights are non-negative. If the weight is zero, we replace it with the maximum value of the weight matrix W . When the score $a\left[\frac{R(e_{ijk}^*)}{N+1}\right]$ is zero, again, the weight is adjusted with the maximum value. Thus, the estimates are obtained by the usual iterated re-weighted least squares (IRLS) algorithm applied to the Rank-based fitting in (3.4.8). We call these estimates the GEER estimates following with the next theorem.

Theorem. Define $\hat{\beta}_{GEER}$ as the solution to the IRLS equation in (3.4.8). Letting $M_i = \hat{V}_i^{-1/2} W_i \hat{V}_i^{-1/2}$, and $n_i \times n_i$ vector $\varphi_i = \varphi_i(R(e_i^*))$ then the asymptotic representation and variance of $\hat{\beta}_{GEER}$ are

$$\hat{\beta}_{GEER} = \left(\sum_{i=1}^I D_i^T M_i D_i\right)^{-1} \left(\sum_{i=1}^I D_i^T M_i (Y_i - a'(\theta_i) - m^*(\beta))\right) \quad (3.4.9)$$

and

$$\text{var}(\hat{\beta}_{GEER}) = \left(\sum_{i=1}^I D_i^T M_i D_i\right)^{-1} \left\{ \sum_{i=1}^I D_i^T \hat{V}_i^{-1/2} \text{var}(\varphi_i) \hat{V}_i^{-1/2} D_i \right\} \left(\sum_{i=1}^I D_i^T M_i D_i\right)^{-1}, \quad (3.4.10)$$

respectively.

An elegant proof is sketched in the Appendix of the study in Abebe et al. (2010) involving a Taylor series expansion similar to Liang & Zeger (1986)'s approach. The GEER algorithm iterates between (3.4.9) and (3.4.1). For estimate of $\text{var}(\varphi_i)$, we propose three covariance structures via the RPP: Nonparametric (NP) estimator, Approximation of the NP estimator (AP) and Improved AP with Compound Symmetric Structure (CS). For the nested design, the AP and CS structure in the variance estimates of fixed effects

in the model is our choice used in the simulation study. Abebe et al. (2010) implemented NP and AP for the simple mixed model. To get the NP estimator, $\widehat{var}(\varphi_i)$ is obtained from simple moment estimator using the scores $a(i)$. AP assumes the true weighted errors are independent. Asymptotically, $var(\varphi_i)$ converges to I_{n_i} and W_i converges to $(1/\hat{\tau}_\varphi)I_{n_i}$ for the nested model, however, there is a need for the investigation of convergence under different models. Thus, the variance estimates of the fixed effect estimate will be defined by

$$var(\hat{\beta}_{GEER-AP}) \doteq \hat{\tau}_\varphi^2 \left(\sum_{i=1}^I D_i^T \hat{V}_i^{-1} D_i \right)^{-1} \left\{ \sum_{i=1}^I D_i^T \hat{V}_i^{-1} D_i \right\} \left(\sum_{i=1}^I D_i^T \hat{V}_i^{-1} D_i \right)^{-1}. \quad (3.4.11)$$

The other one is

$$\begin{aligned} var(\hat{\beta}_{GEER-CS}) &\doteq \hat{\tau}_\varphi^2 \left(\sum_{i=1}^I D_i^T \hat{V}_i^{-1} D_i \right)^{-1} \left\{ \sum_{i=1}^I D_i^T \hat{V}_i^{-1/2} B_i(1, \rho_{\varphi, sec}, \rho_{\varphi, sch}) \hat{V}_i^{-1/2} D_i \right\} .. \\ &.. \left(\sum_{i=1}^I D_i^T \hat{V}_i^{-1} D_i \right)^{-1}. \end{aligned}$$

The GEER methods allow any link function. For the convergence of the weights used here, see Sievers and Abebe (2004).

So far, all three Rank-based methods (JR, GR and GEER) allow for arbitrary scores in the rank dispersion function that is appropriate for the error distribution. Hence, knowledge of the underlying distribution of the errors should increase the efficiency of the estimators.

3.5 Variance Component Estimators

The Rank-based theories developed in this work require the consistency of the estimate of scale parameters. Asymptotic theory for the dispersion function and its estimate are discussed here in two steps. First, working with true errors for the 2-level nested model, we discuss it for the RPP algorithm. Here, we need the results of mean and median estimates in Groggel's (1983) thesis. Second, after fitting the model with Rank-based estimates, the asymptotic properties and convergence of residual dispersion function are handled using the quadratic approximation of the dispersion function and the first order approximation of the rank-based fit and its residuals. Later, these results are generalized for the 3-level nested model. Similar arguments will lead to the consistency of the scale estimators for each type of random errors in the k-level nested models.

Definition 5. The functional parameter corresponding to the dispersion function in (2.1.5) evaluated at the errors e_i for $i = 1, \dots, n$ is

$$\bar{D}_e = \int \varphi(F(x)) x dF(x).$$

Lemma 6. \bar{D}_e is a scale parameter (or, $\bar{D}_{ae+b} = a\bar{D}_e$, $a > 0$, b is real.)

Proof. Assuming the independent case, letting F_n and F denote the empirical distribution function and distribution function of the errors e_1, \dots, e_n , respectively. Now, let $u = ae + b$, and $F_u(ae + b) = P(e \leq \frac{u-a}{b}) = F(\frac{u-a}{b})$. Thus, $\bar{D}_u = \int \varphi(F_u(x)) x dF_u(x) = \int \varphi(v) F_u^{-1}(v) dv$ with $F_u(x) = v$. We can show $F_u^{-1}(x) = aF^{-1}(x) + b$. Thus, $\bar{D}_u = \int \varphi(v) F_u^{-1}(v) dv = \int \varphi(v) (aF^{-1}(v) + b) dv = a[\int \varphi(v) (F^{-1}(v) + b) dv]$. Because the score function is standardized, this simplifies to $a\bar{D}_e$. \square

Theorem 7. (The consistency of the dispersion function with general scores) Assuming that the errors e_1, \dots, e_n are iid,

$$\frac{1}{n} \sum_{i=1}^n a(R(e_i)) \cdot e_i \rightarrow \int \varphi(F(x)) x dF(x) = \bar{D}_e \text{ in probability.}$$

Proof. This part is shown in Hettmansperger & McKean (2011, pg.201). \square

Corollary 8. *(The consistency of the dispersion function with Wilcoxon scores) Assuming the score function is the Wilcoxon. Then,*

$$\frac{2\sqrt{\pi}}{n} \sum_{i=1}^n \left(\frac{R(e_i)}{n+1} - \frac{1}{2} \right) \cdot e_i \rightarrow \int \varphi(F(x)) x dF(x) = \bar{D}_e.$$

Also, $\bar{D}_e = \sqrt{\frac{3}{\pi}} \sigma_e$ under the normal errors e_i with the standard deviation σ_e .

The argument in the proof follows the same argument in the previous proof. Only we need to show that \bar{D}_e is equivalent to $\frac{3}{\pi} \tau_\varphi$ using the Wilcoxon scores. Replacing $a(u) = \sqrt{12}(u - \frac{1}{2})$ in the dispersion function with general scores yields the desired results.

3.5.1 For 2-level Nested Structure

Consider the balanced dependent case, for a 2-level nested structure or simple mixed error model, the errors can be expressed as

$$e_{ij} = a_i + \epsilon_{ij} \tag{3.5.1}$$

for $i = 1, 2, \dots, k$, $j = 1, 2, \dots, n$ and $N = n + k$ (say, k schools, n students observations in each school). Errors a_i and ϵ_{ij} are random cluster effects and error effects, respectively. We only observe the values of e_{ij} , the individual a_i and ϵ_{ij} are not observable. We begin formation of the pseudo-samples by defining the pseudo-errors E_{ij} and A_i using the mean as location estimate. We define $\bar{e}_{i.} = n^{-1} \sum_j e_{ij}$, $\bar{e}_{..} = N^{-1} \sum_i \sum_j y_{ij}$, $\bar{\epsilon}_{i.} = n^{-1} \sum_j \epsilon_{ij}$, $\bar{\epsilon}_{..} = N^{-1} \sum_i \sum_j \epsilon_{ij}$ and $\bar{a}_{.} = k^{-1} \sum_i a_i$. Throughout this chapter, the bar symbol might substitute with any consistent location estimate averaging/collapsing over the dot subscripts; in particular, it might be the mean, the median or the HL location estimate. The pseudo-error E_{ij} is defined as

$$E_{ij} = e_{ij} - \bar{e}_{i.} = \epsilon_{ij} - \bar{\epsilon}_{i.}. \tag{3.5.2}$$

As $n \rightarrow \infty$, $\bar{\epsilon}_{i.} \rightarrow 0$ in probability, so for large n , E_{ij} behaves like the ϵ_{ij} . The pseudo-random error A_i is defined as

$$A_i = \bar{e}_{i.} - \bar{e}_{..} = a_i + \bar{\epsilon}_{i.} - \bar{a}_{.} - \bar{\epsilon}_{..} \quad (3.5.3)$$

As $n \rightarrow \infty$ and $k \rightarrow \infty$, A_i behaves like a_i , since $\bar{\epsilon}_{i.}$, $\bar{\epsilon}_{..}$ and $\bar{a}_{.}$ converge in probability to zero. Let

$$A = \{A_1, A_2, \dots, A_k\} \text{ and } E_1 = \{E_{11}, E_{12}, \dots, E_{1n}\} \quad (3.5.4)$$

be the pseudo-samples. In our nested structure, the highest level clusters are independent and down to levels and sample sizes of these levels, they have the same distribution. So without loss of generality, we will generally work with the first cluster. The analogs of these random and error vectors are as follows:

$$a = \{a_1, a_2, \dots, a_k\} \text{ and } \epsilon_1 = \{\epsilon_{11}, \epsilon_{12}, \dots, \epsilon_{1n}\}.$$

Under the conditions of bounded underlying density functions of these errors, and for k and n large, the pseudo-samples converge for the estimates of random and error effects such that

$$A \rightarrow a \text{ and } E_1 \rightarrow \epsilon_1,$$

if consistent location estimates with desired asymptotic properties are used (Groggel, 1983). Also, these pseudo-samples behave like independent samples of the errors so that the variance of them is equal to the variances of the errors in the limiting case as $n \rightarrow \infty$ and $k \rightarrow \infty$. What we substitute for the estimates of the true errors (a_i and ϵ_{1j}) in model (3.5.1) is the pseudo-sample estimates of the pseudo-samples.

Now, to establish the consistency of the residual dispersion function, we need the following lemma and corollary. Before that, we write the results of Groggel's study (1983) that we need. We call Facts.

Fact 9. Assume errors a_i and ϵ_{1j} in Section 3.5.1 have the empirical cdf $F_k(x)$ and $G_n(x)$,

and cdf $F(x)$ and $G(x)$, respectively. Assume each cdf is symmetric around zero. The results are obtained as follows as $n \rightarrow \infty$ and $k \rightarrow \infty$:

(a) $P(\epsilon_{1j} - \bar{\epsilon}_1. < x) \rightarrow P(\epsilon_{1j} < x) = G(x)$, $P(a_i + \bar{\epsilon}_i. < x) \rightarrow P(a_i < x) = F(x)$, and $P(a_i + \bar{\epsilon}_i. - \bar{a}_. - \bar{\epsilon}_.. < x) \rightarrow P(a_i < x) = F(x)$ in probability by Slutsky's Theorem.

(b) $G_n(x) \rightarrow G(x)$ and $F_k(x) \rightarrow F(x)$ in distribution.

(c) $|\varphi[F_k(a_i + \bar{\epsilon}_i.)] - \varphi[F_k(a_i)]| \rightarrow 0$ and $|\varphi[F_k(a_i + \bar{\epsilon}_i. - \bar{a}_. - \bar{\epsilon}_..)] - \varphi[F_k(a_i)]| \rightarrow 0$, since $\varphi \circ F$ is a uniformly bounded and continuous function, and part (a) and (b).

(d) $|\varphi[G_n(\epsilon_{1j} + \bar{\epsilon}_1.)] - \varphi[G_n(\epsilon_{1j})]| \rightarrow 0$, since $\varphi \circ F$ is a uniformly bounded and continuous function, and part (a) and (b).

(e) $\bar{\epsilon}_i. = o_p(1)$, $(n+k)^{1/2}\bar{a}_. = O_p(1)$, $(n+k)^{1/2}\bar{\epsilon}_i. = O_p(1)$ due to the Markov inequality and $E[(n+k)\bar{\epsilon}_i.^2]$ is uniformly bounded due to the finite Fisher assumption.

(f) Two lemmas from Groggel (1983) are summarized here as.

$$\frac{1}{n} \sum_j^n [G_n(\epsilon_{1j} - \bar{\epsilon}_1.) - G_n(\epsilon_{1j})] = o_p\left(\frac{1}{\sqrt{n+k}}\right)$$

and

$$\frac{1}{k} \sum_i^k [F_k(a_i + \bar{\epsilon}_i. - \bar{a}_. - \bar{\epsilon}_..) - F_k(a_i)] = o_p\left(\frac{1}{\sqrt{n+k}}\right).$$

(g) $\varphi[F(\epsilon_{1j})]\bar{\epsilon}_1. \rightarrow 0$ and $\varphi[F(a_i)]\bar{\epsilon}_i. \rightarrow 0$ in probability since $\bar{\epsilon}_1. \rightarrow 0$ in probability, $\varphi \circ F$ is a uniformly bounded, and Slutsky's theorem.

Groggel (1983) has shown part (f) using the mean and median estimate. For the median, $\bar{a}_. + \bar{\epsilon}_..$ is replaced by $\text{med}_{i,j}\{a_i + \text{med}_j(\epsilon_{ij})\}$. For the HL estimate, it has the desired tightness to get the similar convergence results.

Lemma 10. *For the 2-level nested model, under the notations of Fact 9, the pseudo-sample error vectors A and E_1 are defined in (3.5.4). Then, the pseudo-samples have dispersion functions s.t.*

$$\text{Disp}(A) \rightarrow \bar{D}_a \text{ and } \text{Disp}(E_1) \rightarrow \bar{D}_\epsilon$$

Proof. First part, as $k \rightarrow \infty$,

$$\begin{aligned}
Disp(A) &= \frac{1}{k} \sum_{i=1}^k a[R(A_j)] A_j \\
&= \frac{1}{k} \sum_{i=1}^k a[R(\bar{e}_i - \bar{e}_{..})] (\bar{e}_i - \bar{e}_{..}) \\
&= \frac{1}{k} \sum_{i=1}^k a[R(a_i + \bar{e}_i - \bar{a}_{..} - \bar{e}_{..})] (a_i + \bar{e}_i - \bar{a}_{..} - \bar{e}_{..}) \\
&\doteq \frac{1}{k} \sum_{i=1}^k \varphi \left[\frac{k}{k+1} F_k(a_i + \bar{e}_i - \bar{a}_{..} - \bar{e}_{..}) \right] (a_i + \bar{e}_i - \bar{a}_{..} - \bar{e}_{..}), \text{ by (a)} \\
&\doteq \frac{1}{k} \left\{ \sum_{i=1}^k \varphi \left[\frac{k}{k+1} F_k(a_i + \bar{e}_i - \bar{a}_{..} - \bar{e}_{..}) \right] (a_i) + \right. \\
&\quad \left. + \sum_{i=1}^k \varphi \left[\frac{k}{k+1} F_k(a_i + \bar{e}_i - \bar{a}_{..} - \bar{e}_{..}) \right] (\bar{e}_i - \bar{a}_{..} - \bar{e}_{..}) \right\}, \text{ by (a), (c) and (f)} \\
&\doteq \frac{1}{k} \sum_{i=1}^k \varphi \left[\frac{k}{k+1} F_k(a_i) \right] (a_i), \text{ by (c) and Slutsky Theorem} \\
&\doteq \frac{1}{k} \sum_{i=1}^k \varphi[F(a_i)](a_i), \text{ by (b)} \\
&\doteq \int \varphi[F(x)](x) dF(x) = \bar{D}_a. \tag{3.5.5}
\end{aligned}$$

For the second part, as $k \rightarrow \infty$ and $n \rightarrow \infty$,

$$\begin{aligned}
Disp(E_1) &= \frac{1}{n} \sum_{j=1}^n a[R(E_{1j})] E_{1j} \\
&= \frac{1}{n} \sum_{j=1}^n a[R(e_{1j} - \bar{e}_{1.})] (e_{1j} - \bar{e}_{1.}) \\
&= \frac{1}{n} \sum_{j=1}^n a[R(\epsilon_{1j} - \bar{\epsilon}_{1.})] (\epsilon_{1j} - \bar{\epsilon}_{1.}) \\
&\doteq \frac{1}{n} \sum_{j=1}^n \varphi \left[\frac{n}{n+1} G_n(\epsilon_{1j} - \bar{\epsilon}_{1.}) \right] (\epsilon_{1j} - \bar{\epsilon}_{1.}) \text{ by (a)} \\
&\doteq \frac{1}{n} \sum_{j=1}^n \varphi \left[\frac{n}{n+1} G_n(\epsilon_{1j}) \right] (\epsilon_{1j}) \text{ by (d), (f) and (g)} \\
&\doteq \frac{1}{n} \sum_{j=1}^n \varphi[G(\epsilon_{1j})] (\epsilon_{1j}) \text{ by (b)} \\
&\doteq \int \varphi[G(x)](x) dG(x) = \bar{D}_\epsilon. \tag{3.5.6}
\end{aligned}$$

□

We consider to summarize the consistency of the dispersion function with pseudo-errors and true errors in the following corollary. Next, the consistency of the rank-based

residual dispersion function is stated in the theorem.

Corollary 11. As a result of the previous lemma and independent error types, it can be summarized that $Disp(A) \rightarrow \bar{D}_a$, $Disp(E_1) \rightarrow \bar{D}_\epsilon$, $Disp(a) \rightarrow \bar{D}_a$, $Disp(\epsilon_1) \rightarrow \bar{D}_\epsilon$ as well as $Disp(e_i) \rightarrow \bar{D}_e$, where e_i is error vector in any highest level cluster i .

Theorem 12. Consider Model (2.1.1) and the assumptions in Lemma 10. Suppose the residuals \hat{e} are obtained from a rank-based method fit s.t. $\hat{e}_{ij} = y_{ij} - x_{ij}^T \hat{\beta}_\varphi$, and one cluster residuals vector is \hat{e}_1 , $\hat{e}_1 = (\hat{e}_{11}, \dots, \hat{e}_{1n})$. If the pseudo residual vectors \hat{A} and \hat{E}_1 are obtained from the median location estimate, where $\hat{A}_i = \bar{\hat{e}}_{i.} - \bar{\hat{e}}_{..}$, $\hat{E}_{ij} = \hat{e}_{ij} - \bar{\hat{e}}_{i.}$ in the RPP, then the corresponding residual dispersion functions are consistent, i.e.

$$|Disp(\hat{A}) - Disp(A)| \rightarrow 0,$$

$$|Disp(\hat{E}_1) - Disp(E_1)| \rightarrow 0 \text{ and}$$

$$|Disp(\hat{e}_1) - Disp(e_1)| \rightarrow 0$$

in probability.

Note: Here, the bar symbol substitutes with the median or HL estimate averaging/collapsing over the dot subscripts.

Proof. We need the sufficiency of one cluster residuals (\hat{e}_1) obtained from the Rank-based fit to establish the consistency of the residual dispersion due to the RPP structure in estimating random effects (See Section 2.1.2). For example, the difference $\hat{E}_{ij} = \hat{e}_{ij} - \bar{\hat{e}}_{i.}$ is the predictor of ϵ_{ij} , defined with $\bar{\hat{e}}_{i.} = \text{med}_j\{\hat{e}_{i1}, \hat{e}_{i2}, \dots, \hat{e}_{in}\}$. Also, the difference $\hat{A}_i = \bar{\hat{e}}_{i.} - \bar{\hat{e}}_{..}$ is the predictor of a_i , which is $\text{med}_j\{\hat{e}_{i1}, \hat{e}_{i2}, \dots, \hat{e}_{in}\} - \text{med}_i\{\text{med}_j\{\hat{e}_{ij}\}\}$. In particular, to get the random effects, we get the intercept estimate in a Rank-based model using one cluster residuals. Once we show the scale estimate obtained from this cluster is consistent, we can generalize it for these predictors. The next argument discusses this.

The following case suffices to generalize for many clusters in the model $Y = X\beta + e$. Without loss of generality, we consider the first cluster. Suppose \mathbf{Y}_1 , \mathbf{X}_1 and \mathbf{e}_1 are the

response vector, design matrix, and the error vector, respectively for cluster 1. Let \mathbf{Y}_2 , \mathbf{X}_2 and \mathbf{e}_2 denote the respective parts for the rest of the model. Then, without loss of generality, we write

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}. \quad (3.5.7)$$

So that $(\mathbf{Y}_1, \mathbf{X}_1)$ and $(\mathbf{Y}_2, \mathbf{X}_2)$ define clusters 1 and 2, respectively. Let (\mathbf{Y}, \mathbf{X}) be the long model; $(\mathbf{Y}_1, \mathbf{X}_1)$ be the short model. Let n_1 and n_2 denote the cluster sample sizes and let $n = n_1 + n_2$. Let $\lambda_1 = \lim_{n \rightarrow \infty} n_1/n$ and assume that $0 < \lambda_1 < 1$. Now we have the same model so statements such as the following are true.

$$\lim \mathbf{X}_1' \mathbf{X}_1 / n_1 = \boldsymbol{\Sigma}$$

$$\lim \mathbf{X}_2' \mathbf{X}_2 / n_2 = \boldsymbol{\Sigma}$$

$$\lim \mathbf{X}' \mathbf{X} / n = \boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ is positive definite. Assume without loss of generality that the true $\boldsymbol{\beta}$ is $\mathbf{0}$. Suppose we get any of our fits (the JR, GR or GEER) fit of the entire model, i.e., the long model Y and X . Then we know that $\sqrt{n}\hat{\beta}_L = O_p(1)$. But then we also have that $\sqrt{n_1}\hat{\beta}_L = O_p(1)$. That is, the estimate of the long model is uniformly bounded for the short model too. This is all we need to establish the theory for the intercept in the short model. For the short model, the residuals based on the long model fit are $Y_1 - X_1\hat{\beta}_L$. The results in Hettmansperger and McKean (Lemma 3.5.1 and A.3.2, 2011) hold because $(X_1'X_1)^{1/2}\hat{\beta}_L$ is $O_p(1)$. Thus Theorem 3.5.6 of Hettmansperger & McKean (p.186, 2011) is true and the results on theory for the intercept as the median of the residuals $Y_1 - X_1\hat{\beta}_L$ holds.

In the same way the theory holds for the HL estimate of the residuals. The linearity result on Theorem A.2.11 (Hettmansperger and McKean, p.460 and p.190, 2011) and the boundedness of $\sqrt{n_1}\hat{\beta}_L = O_p(1)$ establishes the linearity. Using the above notations and

results, since $\sqrt{n_1}\hat{\beta}_L = O_p(1)$, we have:

$$D(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta}_L) = Q(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta}_L) + o_p(1), \quad (3.5.8)$$

where $Q(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta}_L) = (1/(2\tau))\hat{\beta}_L'\mathbf{X}_1'\mathbf{X}_1\hat{\beta} - \hat{\beta}_L S_1(\mathbf{e}_1) + D_1(\mathbf{e}_1)$

and $S_1(\mathbf{e}_1) = X_1\varphi[F(\mathbf{e}_1)]$. In Equation (3.5.8), the first term on the right is $O_p(1)$. For the second term, using the pseudo estimate based on $Q()$, quadratic approximation to the dispersion function, in the long model we have:

$$\hat{\beta}_L = \tau(1/n)\Sigma^{-1}S(\mathbf{e}) + o_p(1/\sqrt{n}). \quad (3.5.9)$$

After some algebra, the second term is

$$\begin{aligned} \hat{\beta}_L' S_1(\mathbf{e}_1) &= \tau\varphi[F(\mathbf{e})]'\mathbf{X}(1/n)\Sigma^{-1}[\mathbf{X}_1\varphi[F(\mathbf{e}_1)] + \mathbf{0}'\varphi[F(\mathbf{e}_2)]] + o_p(1) \\ &= \tau\varphi[F(\mathbf{e})]'\mathbf{X}(1/n)\Sigma^{-1}[\mathbf{X}_1'\mathbf{0}']\varphi[F(\mathbf{e})] + o_p(1). \end{aligned}$$

The second term in (3.5.8) is uniformly bounded. It follows from Equation (3.5.8) that $D(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta}_L)$ is asymptotically equivalent to $D_1(\mathbf{e}_1)$. It follows that $(1/\sqrt{n})[D(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta}_L) - D(\mathbf{e}_1)] \rightarrow 0$ in probability. Also the weaker result holds, i.e., $(1/n)[D(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta}_L) - D(\mathbf{e}_1)] \rightarrow 0$. Since $\frac{1}{n}D(\mathbf{e}_1)$ is a consistent estimate of \bar{D}_e , so is $\frac{1}{n}D(\hat{e})$. $(1/n)D(\mathbf{Y}_1 - \mathbf{X}_1\hat{\beta}_L)$ because $D_1(\mathbf{e}_1)$ is consistent. This result can be generalized to obtain the consistency of the residual dispersion of other types since the algorithm in the RPP uses either the median or HL estimate. Also similar arguments obtain the consistency of the asymptotic distributions of the location estimate based on the median and the HL estimator. \square

3.5.2 For 3-level Nested Structure

In this section, using the similar arguments in the previous section, we can obtain, for the 3-level nested structure, that the pseudo random samples, which are pseudo school effects vector $A = \{A_i\}$, pseudo section effects vector $W_i = \{W_{ij}\}$, and pseudo error effects

vector $E_{ij} = \{E_{ijk}\}$, are represented as $A_i = e_{i..} - e_{...} = a_i + \bar{w}_{i.} + \bar{\epsilon}_{i..} - (\bar{a}_{.} + \bar{w}_{..} + \bar{\epsilon}_{...})$, $W_{ij} = e_{ij.} - e_{i..} = w_{ij} + \bar{\epsilon}_{ij.} - (\bar{w}_{i.} + \bar{\epsilon}_{i..})$, and $E_{ijk} = e_{ijk} - e_{ij.} = \epsilon_{ijk} - \bar{\epsilon}_{ij.}$ for $i = 1, \dots, I$, $j = 1, \dots, J$ and $k = 1, \dots, n$. Besides the cdf $F(x)$ for random term a_i and $G(x)$ for w_{ij} , $H(x)$ is defined for ϵ_{ijk} . Under balanced case, using the similar argument discussed for the 2-level case for the choice of either of the estimates (mean, median or any consistent location estimator), $A_i \rightarrow a_i$, $W_{ij} \rightarrow w_{ij}$, and $E_{ijk} \rightarrow \epsilon_{ijk}$ in probability as I and nk go to ∞ . The desired results are summarized in Claims.

Claim 13. (Consistency of Residual Dispersion Function in the 3-level model) Under Model 3.1.3 and the pseudo-sample notations in Section 3.5, the following results are obtained.

1. Using true errors vector a , w_1 and ϵ_{11} , $Disp(a) \rightarrow \bar{D}_a$, $Disp(w_1) \rightarrow \bar{D}_w$, $Disp(\epsilon_{11}) \rightarrow \bar{D}_\epsilon$, and $Disp(e_1) \rightarrow \bar{D}_e$.
2. Using pseudo errors vector A , W_1 and E_{11} , $Disp(A) \rightarrow \bar{D}_a$, $Disp(W_1) \rightarrow \bar{D}_w$ and $Disp(E_{11}) \rightarrow \bar{D}_\epsilon$, $Disp(e_1) \rightarrow \bar{D}_e$.
3. Using the estimates of pseudo errors vector \hat{A} , \hat{W}_1 and \hat{E}_{11} obtained from the RPP with the Rank-based fit, $Disp(\hat{A}) \rightarrow \bar{D}_a$, $Disp(\hat{W}_1) \rightarrow \bar{D}_w$, $Disp(\hat{E}_{11}) \rightarrow \bar{D}_\epsilon$, and $Disp(\hat{e}_1) \rightarrow \bar{D}_e$.

As in the 2-level case argument, we have consistency of location based on residuals. To these analogous quantities of true errors and consistency of the scale estimator (the dispersion function). Using similar asymptotical and convergence equivalences regarding with additional pseudo-sample terms described below and in Facts 14, we can generalize the consistency of residual dispersion functions to the the 3-level case and k-level case. In case of using all errors in estimates by forming pseudo random samples, the average for the same level clusters is a consistent estimate.

Claim 14. By Slutsky Theorem and convergence of the location estimate,

- (a) $G(w_{1j} + \bar{\epsilon}_{1j.} - \bar{w}_{1.}) \rightarrow G(w_{1j})$, $\varphi[F(a_i)]\bar{\epsilon}_{i.} \rightarrow 0$ in probability since $\bar{\epsilon}_{i.} \rightarrow 0$, $\varphi[F(a_i)]$ is a uniformly bounded.

(b) $\varphi[F(w_{ij} + \bar{\epsilon}_{ij.})](\bar{\epsilon}_{ij.}) \rightarrow 0$ in probability.

Using residuals from the fit with the Rank-based methods, there is a need to investigate the finite sample corrections for score functions and various location-scale estimates in RPP algorithm.

3.5.3 Finite Correction for the Scale Estimates

Groggel et al. (1988) developed finite sample corrections for median-based pseudo-samples in the 2-level nested model. We offer the correction for the variance components σ_ϵ^2 , σ_{sch}^2 and σ_{sect}^2 as in the traditional case. The variance terms in the pseudo random sample structure using the mean estimate can be expressed as $var(A_i) = \frac{I-1}{I}\sigma_{sch}^2 + \frac{I-1}{I \cdot \bar{J}_i}\sigma_{sect}^2 + \frac{I-1}{N}\sigma_\epsilon^2$, $var(W_{j(i)}) = \frac{\bar{J}_i-1}{\bar{J}_i}\sigma_{sect}^2 + \frac{\bar{J}_i-1}{\bar{n}_i}\sigma_\epsilon^2$ and $var(E_{ijk}) = \frac{\bar{n}_{ij}-(p+1)}{\bar{n}_{ij}}\sigma_\epsilon^2$ for the estimates of σ_ϵ^2 , σ_{sch}^2 and σ_{sect}^2 where $p+1$ is the rank of the design matrix X . In an analogous manner to this correction, in the HL, median, MAD and $Disp$ estimators, it might worth trying similar ways with which \bar{n}_{ij} and \bar{J}_i can be calculated from the harmonic mean of the number of observations in section and the number of sections within school.

Chapter 4

Simulation Results and Real Data Analysis

To illustrate the proposed methods, we first present the results of a Monte Carlo investigation comparing the performance of the GR method and the traditional likelihood method, REML. For the GR method, we also compare the empirical behaviors of two alternative procedures of location-scale estimator pairs (med-MAD and HL-Disp) for the RPP algorithm. For the second performance of our study, we compare the fittings including the JR, GR, GEER and REML fitting methods over various situations of the strengths of intra and inter correlation. The second part contains a validation and efficiency study. Further, two real data sets are analyzed with these methods. Throughout the simulations, the Wilcoxon scores are used. Initial fit is obtained from the JR estimate with the Wilcoxon scores. For the GEER analysis, one step iteration is run due to the fast convergence. Also, the Rank-based regression estimates are obtained from R packages *wwest* (Terpstra and McKean, 2005).

4.1 Performance of GR Estimates

In this section, the performance of fixed effect estimates in the GR fitting are investigated and scale estimates using the RPP are discussed. In this experiment, the package *nlme/lmer* in R software developed by Pinheiro et al. (2011) is used to obtain the REML and ML estimates.

Simulations are conducted to evaluate the performance of our iterative GR fitting with the RPP algorithm found in (2.1.2) for random, fixed, scale effects and intra-class correlations in the 3-level nested model,

$$y_{ijk} = x_{ijk}^T \beta + a_i + w_{j(i)} + \epsilon_{k(ij)}. \quad (4.1.1)$$

We assume that the section effects $w_{j(i)}$ are random and nested within schools with random effects a_i . With unbalanced design, the total sample size is 338 observations in the total of seven schools with varying sections (four to five) within schools. The total measurements in each section vary from 8 to 20. We have two error types in the model, overall error ($\epsilon_{k(ij)}$) and random errors (a_i and $w_{j(i)}$). Fixed parameters relating to the fixed effects, namely intercept and covariate, are set at zero, $\beta = (\beta_0, \beta_1) = (0, 0)$, because the methods are regression and scale equivariant.

For non normal cases, contaminated (CN) errors are set at 20% contamination percentage, and the ratio of the contaminated variance to the uncontaminated is set at 16. Scale parameters, $\theta = (\sigma_{school}^2, \sigma_{section(school)}^2, \sigma_{error}^2)$, are set at (1, 4, 9) when no contamination is employed. The scale parameter θ becomes (4, 16, 36) when both random and error terms are contaminated.

Using the two location estimates (med and HL) and scale estimates (Disp and MAD), we investigate empirically the two proposed RPP procedures (HL-Disp and med-MAD) in the GR method under the four cases: Case 1: Both errors are normal. Case 2: Only error terms are contaminated. Case 3: Only random errors are contaminated. Case 4: Both errors are contaminated.

Also, for each case, one observation or one section are made up as outliers (call it *corrupted data*) to observe the sensitivity and efficiency of the methods. In the experiment, the standard errors (*se*) and empirical asymptotic relative efficiencies (ARE) without bias correction were obtained. The empirical AREs were obtained from the ratios of mean squared errors (MSE). The *se*'s of fixed effects, scale and intra-class estimates are either competitive under normal case, and are mostly better than the REML when CN or cor-

rupted data are considered. Surprisingly, the empirical *se* for covariate estimates except in Case 1 are observed smaller than that in the REML throughout the simulations (See Tables 4.1 to 4.4). Moreover, the estimates in the REML method are biased under some situations. When both errors are contaminated as specified, the intra-class estimates in the RPP are unbiased, while the REML estimates are biased. On the other side, we have not made any degrees of freedom corrections for the RPP estimators of scale. In general, they appear to be biased. Small sample corrections are under investigation. The HL-Disp procedure outperforms better than the med-MAD procedure in the RPP. Our small simulation study in nested design suggests that Disp as a scale estimator is more efficient for nested designs than MAD.

We compare on performance of the intra-class parameter estimates for only Case 1 and Case 4. Under these situations, the scale estimates $\hat{\sigma}$, Disp and Mad are all consistent to the same parameter. In Case 2 and 3, it is fact that they are not consistent due to the mixed contaminated and normal errors. Moreover, the scale estimates using REML were highly variable in all cases (See *se*'s for the scale and intra-class parameter estimates in Tables 4.1 to 4.4.). We did not use a bias correction for these estimators. Intraclass correlation parameters vary depending on the amount of contamination as in four cases above: This is for Cases 1 and 4, $\rho = (64.3\%, 7.1\%, 35.7\%)$, for Case 2, $\rho = (87.8\%, 2.4\%, 12.2\%)$, for Case 3, $\rho = (31\%, 13.8\%, 69\%)$. The intra-class correlation estimators of the RPP are more efficient when contaminated or corrupted data are employed. However, the RPP and REML estimators are biased for contaminated data. Surprisingly, in all cases the REML does poorly, except when the errors are normal. Even in Case 1, the intra-school is biased using the REML estimator.

The empirical AREs indicate that our proposed procedures, HL-Disp and med-MAD in the RPP algorithm with the GR fitting method, are generally competitive with or outperforms the REML estimators. Also, the results indicate that random effects for school and section in both methods (GR and REML) capture the parameter values.

4.1.1 Simulation Results

In the tables below, HL-Disp means that the location and scale estimators are respectively HL and Disp. Med-MAD is likewise. These are the RPP procedures in the GR fitting we work. ARE values which are greater than 1 indicates the results in the GR with the RPP is favored over the REML in efficiency. Also, (*) means any cluster effect among the others. When the data is corrupted, (*) is the effect of the cluster related with the corrupted data.

Table 4.1: Case-1 Simulation Results

CASE-1: Normal case (No contamination)			Loc-Scale Param		No corrupted data												Empirical AREs to REML	
					HL-DISP						Med-MAD							
					median			mean			median			mean				
			est	se	est	se	est	se	est	se	est	se	est	se	est	se		
Fixed Effect	β_0	0	0.0056	0.0274	0.0023	0.0222	0.007	0.0306	0.0213	0.0252	-0.012	0.0178						
	β_1	0	-0.0126	0.0131	-0.0064	0.0108	-0.0001	0.0153	0.0003	0.0112	-0.0011	0.0106						
Scale Param.	σ^2_{sch}	1	1.1447	0.0291	1.3244	0.0253	1.658	0.0728	2.3704	0.0696	1.0875	0.0369						
	σ^2_{sect}	4	3.5621	0.0406	3.651	0.0325	2.8072	0.0613	3.1001	0.0518	3.9618	0.0414						
	σ^2_{err}	9	8.0349	0.0282	8.0512	0.0207	6.8944	0.0433	6.9569	0.0317	9.0436	0.0228						
Intra-class Corr	Error p	0.6429	0.6252	0.0028	0.6253	0.0023	0.5803	0.0059	0.5837	0.004	0.649	0.0024						
	School p	0.0713	0.0893	0.0021	0.0996	0.0017	0.1429	0.0061	0.1733	0.0039	0.0731	0.0023						
	Section p	0.3571	0.3773	0.0028	0.3779	0.0021	0.4179	0.0048	0.4183	0.004	0.351	0.0024						
Random Effect	School*	0	-0.0065	0.03	-0.0305	0.0409	0	0	-0.0211	0.0491	-0.0066	0.0243						
	Section*	0	-0.0056	0.0331	0.0152	0.0623	0	0	0.1298	0.0674	0.0032	0.0548						
			RPP Methods						Med-MAD									
			HL-DISP															
			0.6429						0.9633						0.4989			
			1.6227						1.2132						0.5173			
			1.0888						1.8304						0.3478			
			1.3061						0.353						0.2449			
			0.7737						0.6611									

Table 4.2: Case-2 Simulation Results

CASE-2: Only error term contaminated			Loc-Scale Param		No corrupted data												Empirical AREs to REML	
					HL-DISP						Med-MAD							
					median		mean		median		mean		median		mean			
			est	se	est	se	est	se	est	se	est	se	est	se	est	se		
Fixed Effect	β_0	0	0.0038	0.0295	-0.0259	0.0223	0.0352	-0.0007	0.0352	0.0302	-0.0263	0.0302	-0.0036	0.0223	1	0.5453		
	β_1	0	0.0128	0.0248	-0.007	0.0154	0.0108	0.017	0.017	0.0149	0.01	0.0149	-0.0022	0.0218	2.0039	2.1406		
Scale Param	σ^2_{sch}	1	1.31	0.0368	1.4635	0.0266	0.0741	1.8192	0.0741	2.5672	0.079	0.079	1.2191	0.0435	2.6743	0.3032		
	σ^2_{sect}	4	4.194	0.0524	4.3337	0.0396	3.1368	0.0666	3.4292	0.0586	0.0625	3.8598	0.0625	2.491	1.1375			
	σ^2_{err}	36	24.4048	0.1598	24.68	0.117	10.6927	0.0775	10.7814	0.0509	35.8277	0.1884	2.5929	13.7				
Intra-class Corr	Error p	0.878	0.8131	0.0017	0.8106	0.0016	0.6661	0.0052	0.6622	0.0038	0.8754	0.0017	1.1289	0.2001				
	School p	0.0244	0.0446	0.0012	0.0483	0.0009	0.116	0.0043	0.1415	0.0033	0.0296	0.001	1.2346	0.0918				
	Section p	0.122	0.1863	0.0019	0.191	0.0015	0.3387	0.0046	0.3424	0.0036	0.1246	0.0017	1.2844	0.223				
Random Effect	School*	0	-0.0088	0.0395	-0.0287	0.0455	0	0	0.0484	0.0499	-0.0203	0.0269	0.3495	0.2906				
	Section*	0	0.0012	0.0311	0.0984	0.0687	0	0	0.001	0.0709	0.0492	0.0454	0.4367	0.41				

Table 4.4: Case-4 Simulation Results

CASE-4: Both error and random terms contaminated			Loc-Scale Param		No corrupted data												Empirical AREs to REML	
					HL-DISP						Med-MAD							
					median		mean		se		median		mean		se			
			est	se	est	se	est	se	est	se	est	se	est	se	est	se		
Fixed Effect	β_0	0	0.0495	0.0521	0.0153	0.0399	0.114	0.0677	0.1151	0.0476	-0.0223	0.036	0.8141	0.572				
	β_1	0	0.0196	0.0204	0.0044	0.0158	0.0222	0.0226	0.0134	0.015	0.013	0.0213	1.8174	2.0164				
Scale Param	σ^2_{sch}	4	3.0875	0.1012	4.0079	0.1062	3.3202	0.1312	4.9641	0.1568	4.3722	0.1914	3.2481	1.49				
	σ^2_{sect}	16	10.2757	0.1537	11.2067	0.1589	4.752	0.1302	5.4098	0.1006	15.6117	0.2689	2.8637	7.1447				
	σ^2_{err}	36	24.5321	0.1427	24.6752	0.1133	10.7023	0.0573	10.8339	0.0539	35.8894	0.1833	2.6174	11.565				
Intra-class Corr	Error p	0.6429	0.6361	0.005	0.6306	0.0031	0.5427	0.0057	0.5437	0.0046	0.6587	0.0037	1.4246	0.647				
	School p	0.0713	0.0779	0.0021	0.0965	0.0022	0.1787	0.0063	0.2052	0.0046	0.0724	0.0028	1.6198	0.3705				
	Section p	0.3571	0.3681	0.0036	0.3701	0.0031	0.4577	0.0068	0.4562	0.0044	0.3413	0.0037	1.4246	0.7071				
Random Effect	School*	0	-0.0121	0.0539	-0.0613	0.0733	0	0	-0.0919	0.0779	-0.024	0.0501	0.4672	0.4136				
	Section*	0	0	0.039	-0.07	0.1283	0	0	0.3854	0.1364	-0.1152	0.1175	0.8387	0.7421				

4.1.2 Sensitivity

Rank-based inference inherits better efficiency properties of the estimates when outlier exists. In the simulation study, when corrupted data on one observation or one subcluster (section) were employed, it was observed that the RPP procedure was more efficient and less sensitive in the cases (See Table 4.5). The estimates of fixed effects using the REML were highly variable in case of having any outlier. Also, little biases were observed in small sample sizes in both methods.

Table 4.5: Case-1 Simulation Results for Corrupted Data (Outlier)

CASE-1: Normal case (No contamination)		Loc-Scale		1 corrupted observation						1 corrupted section						Empirical AREs to REML (1 obs)				Empirical AREs to REML (1 sect)								
		HL-DISP			Med-MAD			REML			HL-DISP			Med-MAD			REML		RPP Methods		RPP Methods		HL-DISP		Med-Disp			
		est	mean	se	est	mean	se	est	se	est	mean	se	est	mean	se	est	se	est	se	est	se	est	se	est	se	est	se	
Fixed Effect																												
β_0	0	-0.4376	0.0226	0.0273	-0.4357	0.0225	0.0225	0.2395	0.0209		0.7005	0.034	0.034	-1.4226	0.0442	0.0182	3.0348	0.0182	0.86	0.81	0.29	0.27	0.29	0.27	0.29	0.27	0.29	0.27
β_1	0	0.0019	0.0113	0.0248	0.009	0.0114	0.0221	0.0221	0.0225		-0.0018	0.0115	0.0115	-0.0117	0.012	0.0128	-0.009	0.0128	3.96	3.90	1.24	1.31	1.24	1.31	1.24	1.31	1.24	1.31
σ^2_{sch}	1	1.3213	0.0248	0.0279	1.5429	0.0279	0.0279	1.1105	0.0413		1.4392	0.0258	0.0258	2.4443	0.0686	0.1699	3.4977	0.1699	2.77	2.09	43.37	32.29	43.37	32.29	43.37	32.29	43.37	32.29
Scale Param.																												
σ^2_{sect}	4	3.6145	0.0319	0.0351	3.8933	0.0351	0.0351	3.3707	0.0582		46.6105	0.1318	0.1318	3.2736	0.056	0.5859	305.181	0.5859	3.33	2.67	19.76	16.98	19.76	16.98	19.76	16.98	19.76	16.98
σ^2_{err}	9	11.0975	0.0273	0.0273	11.3487	0.0273	0.0273	39.1636	0.1938		9.7666	0.0304	0.0304	7.3982	0.0334	0.0547	12.3253	0.0547	50.39	49.30	3.24	3.11	3.24	3.11	3.24	3.11	3.24	3.11
Error p	0.6429	0.7162	0.0021	0.0021	0.6993	0.0022	0.0022	0.8963	0.0015		0.1676	0.0006	0.0006	0.6622	0.0047	0.0002	0.0385	0.0002	0.51	0.46	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
Intra-class Corr	School p	0.0835	0.0014	0.0014	0.0926	0.0015	0.0015	0.0253	0.0009		0.0246	0.0004	0.0004	0.2056	0.0051	0.0005	0.0109	0.0005	0.41	0.36	1.56	1.00	1.56	1.00	1.56	1.00	1.56	1.00
	Section p	0.3138	0.0019	0.0019	0.3286	0.0021	0.0021	0.1037	0.0015		0.8207	0.0007	0.0007	0.4757	0.0049	0.0002	0.9615	0.0002	0.62	0.51	0.08	0.06	0.08	0.06	0.08	0.06	0.08	0.06
	School*	0	0.0736	0.0422	-0.0078	0.045	0.045	0.66	0.0274		0.9571	0.047	0.047	0.4979	0.0535	0.0429	0.9039	0.0429	0.42	0.32	0.83	0.68	0.83	0.68	0.83	0.68	0.83	0.68
Random Effect																												
	Section*	0	0.4373	0.0648	0.3525	0.0693	0.0693	3.7027	0.0638		98.9339	0.1074	0.1074	99.5627	0.118	0.1007	95.7805	0.1007	0.97	0.90	0.88	0.71	0.88	0.71	0.88	0.71	0.88	0.71

4.2 Empirical Results for JR, GR and GEER Estimators

In this section, the validity using the empirical level for alpha at 5% and efficiency using comparative confidence interval (CI) lengths (or ARE's) are investigated for the proposed JR, GR and GEER estimators. Thirteen situations were formed with respect to the strength of intra and inter correlation in clusters (school) and subclusters (section within school). For each situation 10,000 simulations were used. Model is the 3-level nested model with unbalanced size, and the same model (4.1.1) is used. The total sample size is 185 observations in the total of seven schools with two to three sections within schools. The total measurements in each section vary from 6 to 20. Two error types in the model were generated from normal distribution. For non-normal situations, contaminated errors were generated at 10% to 20% contamination percentage and at 4 to 16 as the variances ratio (See Section 4.1). The variables of interests are response (a continuous data obtained from the errors), treatment (treatment or control group) and covariate (a normal continuous data).

The Monte Carlo results indicate that the performance of the Rank-based estimators of the fixed parameters $\beta = (\beta_0, \beta_1, \beta_2)$ is more efficient than the REML. This traditional estimator of the fixed parameters in the nested random model mostly meets the true alpha level at 5% when the error distribution is normal, however not optimal for some situations. For example, when the scale parameters $\theta = (\sigma_{school}^2, \sigma_{section(school)}^2, \sigma_{error}^2)$ were generated as (1,1,9) and (0,1,1), it yielded somewhat liberal results for the empirical level in the covariate (See Tables 4.8 and 4.11). Also, the empirical levels for intercept in both methods under each situation was liberal (>5% to 10%) The exception to this fact is that the GR generated slight conservative results (See Tables 4.6, 4.8, 4.9, 4.11, 4.6, 4.15 and 4.17). Overall, the GR and GEERs among the Rank-based methods are the best and closest to the true alpha level. One reason to explain why the GR is conservative (less than 5%) in some situations is that the standard error calculations and the estimates using the projected space in Theorem 1 were obtained without the correction for degrees

of freedom. Hence, this would make the error wider. However, Rank-based estimators have consistency for this type of estimators (Hettmansperger and McKean, 2011). This part can be further studied to get close results to the true level.

The ARE results in the proposed methods are competitive with the REML under normal errors, and better than it when the data follow contaminated errors. Some small degree of loss is observed in rank-based estimators. Under normal errors, the efficiency loss is about 5% in fixed effects, and 30-50% in intercept estimates. These empirical AREs are expected results for the estimator of intercept using L_1 (median) norm and the estimator of fixed effects using the Wilcoxon scores.

One weakness in the JR estimators was observed in the empirical level results of the treatment variable when observations within schools were highly correlated ($>30\%$) in Tables 4.7, 4.9, 4.12, 4.13, 4.14, 4.16 and 4.17. In the future, we are planning to investigate the JR method by fitting the random effects as fixed effects in addition to their estimation of the random effects. This appeared to be a positive correction in the studies of Kloke et al. (2009) The empirical alpha level in this simulation was close to the true level, 5%, in the GR and GEER methods for the treatment effect estimators.

The covariate variable used in this simulation was continuous data, and overall our methods behaved very well in the empirical level and efficiency, better than the traditional method in the AREs. In this variable, the true alpha level for alpha was gained in general.

Under the highly correlated school and section situations (high intra and inter correlation), the empirical levels of the results in the GR, GEER-AP and GEER-CS were quite close to the true levels, whereas the REML was either slightly liberal or conservative. For example, Tables 4.7 and 4.16 show the outperformance of these methods. One of the most powerful results about the proposed methods can be observed here. The GR with the procedure med-MAD and HL-Disp is the winner among the rank methods.

The REML and the Rank-based methods had somewhat liberal empirical alpha level ($>5\%$) in covariate estimates in the situation of zero intra-school and intra-section correlation (See Table 4.10). In this case, the GEERs were observed better among the other

rank methods.

Tables 4.12, 4.14 and 4.16 summarize the results of the situation of little or zero correlation for nested subclusters, indicating that the GR is better, whereas the REML has liberal level in treatment and intercept. Even in the intercept estimate, the GR is precise.

Table 4.11 illustrates the results for independent situation for school clusters (zero intra-school correlation). The GR and GEER were observed better in empirical alpha level, however, the REML has lost some precision.

When contaminated errors were employed, the relative efficiencies of the estimators in the GR and GEER relative to the REML estimators were in favor of the Rank-based estimators. Tables 4.18, 4.19 and 4.20 illustrate that the GR and GEER methods outperformed the REML in the empirical AREs. These 3-level nested model results are in an agreement with the study of Hettmansperger & McKean (2011) and Kloke et al. (2009) for the independent and simple mixed model simulations.

Table 4.6: $\theta = (1, 4, 9)$, Normal

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0643	1.0000	0.0536	1.0000	0.0549	1.0000
JR	0.1324	1.0276	0.0397	0.6883	0.0519	0.8250
GR (med-mad)	0.0178	0.5020	0.0603	0.9870	0.0526	0.9659
GR (hl-disp)	0.0176	0.6158	0.0599	0.9975	0.0536	0.9726
GEER-AP (med-mad)	0.0686	0.7264	0.0522	0.9550	0.0500	0.9550
GEER-AP (hl-disp)	0.0839	0.8841	0.0505	0.9606	0.0509	0.9606
GEER-CS (med-mad)	0.1903	1.0095	0.0520	0.9537	0.0487	0.9538
GEER-CS (hl-disp)	0.1645	1.0529	0.0506	0.9594	0.0509	0.9594

Table 4.7: $\theta = (9, 4, 1)$, Normal

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0996	1.0000	0.0532	1.0000	0.0497	1.0000
JR	0.1386	0.8707	0.0004	0.1449	0.0489	0.3274
GR (med-mad)	0.0468	0.5594	0.0462	0.9257	0.0438	0.9243
GR (hl-disp)	0.0485	0.7047	0.0484	0.9485	0.0457	0.9470
GEER-AP (med-mad)	0.1397	0.8272	0.0497	0.9193	0.0482	0.9193
GEER-AP (hl-disp)	0.1627	1.0454	0.0554	0.9424	0.0505	0.9423
GEER-CS (med-mad)	0.1908	0.9665	0.0555	0.9472	0.0551	0.9471
GEER-CS (hl-disp)	0.1690	1.0347	0.0594	0.9561	0.0543	0.9561

4.2.1 Simulation Results

10,000 simulation yields the margin of error of 0.0043 at 5% alpha level, and the acceptable interval on level is $[0.0457, 0.0543]$. For 1,000 simulations, the margin of error becomes 0.0135 and the acceptable confidence interval for the alpha level is $[0.0365, 0.0635]$. Throughout the tables, the variance components are stated as $\theta = (\sigma_{school}^2, \sigma_{section(school)}^2, \sigma_{\epsilon}^2)$. By changing the parameters in θ , new situations are formed in regard to the strength of intra and inter correlation coefficients ρ_{school} and $\rho_{section(school)}$. Each table illustrates the situations in this simulation study.

Table 4.8: $\theta = (1, 1, 9)$, Normal

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0740	1.0000	0.0564	1.0000	0.0576	1.0000
JR	0.1409	1.0571	0.0594	0.8584	0.0552	0.9094
GR (med-mad)	0.0207	0.5180	0.0695	1.0137	0.0525	0.9799
GR (hl-disp)	0.0253	0.6366	0.0696	1.0227	0.0581	0.9818
GEER-AP (med-mad)	0.0763	0.7379	0.0564	0.9650	0.0484	0.9650
GEER-AP (hl-disp)	0.0928	0.8929	0.0535	0.9649	0.0533	0.9649
GEER-CS (med-mad)	0.1840	0.9981	0.0541	0.9565	0.0464	0.9566
GEER-CS (hl-disp)	0.1520	1.0235	0.0516	0.9589	0.0524	0.9589

Table 4.9: $\theta = (1, 1, 1)$, Normal

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0874	1.0000	0.0518	1.0000	0.0475	1.0000
JR	0.1205	0.9369	0.0061	0.3537	0.0518	0.6239
GR (med-mad)	0.0311	0.5337	0.0485	0.9546	0.0446	0.9469
GR (hl-disp)	0.0367	0.6664	0.0515	0.9668	0.0485	0.9583
GEER-AP (med-mad)	0.1076	0.7865	0.0473	0.9403	0.0441	0.9403
GEER-AP (hl-disp)	0.1334	0.9799	0.0510	0.9516	0.0478	0.9515
GEER-CS (med-mad)	0.1917	0.9839	0.0497	0.9505	0.0456	0.9504
GEER-CS (hl-disp)	0.1697	1.0375	0.0523	0.9578	0.0490	0.9578

Table 4.10: $\theta = (0, 0, 1)$, Normal, nsims=1,000

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0552	1.0000	0.0491	1.0000	0.0662	1.0000
JR	0.2070	1.2987	0.1976	1.4287	0.0562	0.9632
GR (med-mad)	0.0150	0.4880	0.0772	1.0467	0.0712	0.9832
GR (hl-disp)	0.0160	0.5914	0.0720	1.0750	0.0720	0.9898
GEER-AP (med-mad)	0.0381	0.6637	0.0522	0.9588	0.0622	0.9596
GEER-AP (hl-disp)	0.0480	0.7749	0.0360	0.9608	0.0630	0.9607
GEER-CS (med-mad)	0.1665	1.0118	0.0502	0.9450	0.0542	0.9455
GEER-CS (hl-disp)	0.1450	1.0363	0.0350	0.9495	0.0610	0.9494

Table 4.11: $\theta = (0, 1, 1)$, Normal, nsims=1,000

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0540	1.0000	0.0520	1.0000	0.0600	1.0000
JR	0.1373	1.0366	0.0340	0.5914	0.0540	0.7348
GR (med-mad)	0.0170	0.4930	0.0680	0.9691	0.0550	0.9540
GR (hl-disp)	0.0180	0.5956	0.0450	0.9841	0.0520	0.9671
GEER-AP (med-mad)	0.0700	0.7202	0.0620	0.9448	0.0540	0.9447
GEER-AP (hl-disp)	0.0950	0.8685	0.0460	0.9573	0.0500	0.9573
GEER-CS (med-mad)	0.2143	1.0397	0.0670	0.9517	0.0580	0.9513
GEER-CS (hl-disp)	0.2022	1.0953	0.0460	0.9591	0.0500	0.9591

Table 4.12: $\theta = (1, 0, 1)$, Normal, nsims=1,000

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.1020	1.0000	0.0630	1.0000	0.0550	1.0000
JR	0.1190	0.9467	0.0050	0.4015	0.0570	0.7367
GR (med-mad)	0.0490	0.5832	0.0590	0.9806	0.0550	0.9675
GR (hl-disp)	0.0700	0.7467	0.0700	0.9797	0.0420	0.9651
GEER-AP (med-mad)	0.1430	0.8564	0.0590	0.9591	0.0560	0.9590
GEER-AP (hl-disp)	0.1750	1.0927	0.0630	0.9561	0.0440	0.9562
GEER-CS (med-mad)	0.2000	1.0246	0.0520	0.9545	0.0560	0.9545
GEER-CS (hl-disp)	0.1630	1.0329	0.0650	0.9543	0.0440	0.9544

Table 4.13: $\theta = (100, 100, 1)$, Normal

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0980	1.0000	0.0410	1.0000	0.0590	1.0000
JR	0.1570	0.8427	0.0010	0.0486	0.0410	0.1064
GR (med-mad)	0.0320	0.5269	0.0460	0.9125	0.0540	0.9123
GR (hl-disp)	0.0370	0.6764	0.0330	0.9442	0.0540	0.9441
GEER-AP (med-mad)	0.1110	0.7837	0.0760	0.9073	0.0790	0.9072
GEER-AP (hl-disp)	0.1290	0.9975	0.0590	0.9377	0.0770	0.9377
GEER-CS (med-mad)	0.1955	0.9542	0.0870	0.9489	0.0830	0.9490
GEER-CS (hl-disp)	0.1580	1.0327	0.0610	0.9524	0.0810	0.9524

Table 4.14: $\theta = (100, 1, 100)$, Normal

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0930	1.0000	0.0532	1.0000	0.0546	1.0000
JR	0.1160	0.9551	0.0037	0.4087	0.0482	0.7395
GR (med-mad)	0.0536	0.5943	0.0550	0.9800	0.0542	0.9663
GR (hl-disp)	0.0553	0.7462	0.0569	0.9801	0.0514	0.9655
GEER-AP (med-mad)	0.1389	0.8711	0.0524	0.9580	0.0527	0.9579
GEER-AP (hl-disp)	0.1647	1.0899	0.0520	0.9569	0.0502	0.9569
GEER-CS (med-mad)	0.1817	1.0209	0.0511	0.9541	0.0519	0.9542
GEER-CS (hl-disp)	0.1542	1.0379	0.0514	0.9551	0.0498	0.9551

Table 4.15: $\theta = (1, 100, 100)$, Normal

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0553	1.0000	0.0508	1.0000	0.0523	1.0000
JR	0.1349	1.0369	0.0329	0.5848	0.0501	0.7324
GR (med-mad)	0.0157	0.4934	0.0540	0.9686	0.0516	0.9544
GR (hl-disp)	0.0161	0.5962	0.0516	0.9823	0.0537	0.9659
GEER-AP (med-mad)	0.0695	0.7182	0.0496	0.9457	0.0505	0.9457
GEER-AP (hl-disp)	0.0760	0.8665	0.0459	0.9565	0.0520	0.9565
GEER-CS (med-mad)	0.1987	1.0255	0.0513	0.9518	0.0508	0.9517
GEER-CS (hl-disp)	0.1817	1.0971	0.0465	0.9585	0.0525	0.9585

Table 4.16: $\theta = (100, 1, 1)$, Normal

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0957	1.0000	0.0496	1.0000	0.0469	1.0000
JR	0.1975	0.7838	0.0002	0.0684	0.0350	0.1840
GR (med-mad)	0.0713	0.6008	0.0437	0.9341	0.0469	0.9339
GR (hl-disp)	0.0636	0.7578	0.0443	0.9495	0.0435	0.9493
GEER-AP (med-mad)	0.1660	0.8887	0.0559	0.9288	0.0580	0.9288
GEER-AP (hl-disp)	0.1811	1.1295	0.0602	0.9445	0.0527	0.9445
GEER-CS (med-mad)	0.1811	0.9937	0.0620	0.9493	0.0620	0.9492
GEER-CS (hl-disp)	0.1518	1.0357	0.0642	0.9563	0.0556	0.9563

Table 4.17: $\theta = (1, 100, 1)$, Normal

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0556	1.0000	0.0521	1.0000	0.0525	1.0000
JR	0.1687	0.8593	0.0104	0.0812	0.0455	0.1384
GR (med-mad)	0.0151	0.4803	0.0417	0.9166	0.0440	0.9163
GR (hl-disp)	0.0178	0.5983	0.0479	0.9514	0.0505	0.9512
GEER-AP (med-mad)	0.0721	0.7110	0.0648	0.9110	0.0646	0.9109
GEER-AP (hl-disp)	0.0904	0.8860	0.0710	0.9457	0.0649	0.9457
GEER-CS (med-mad)	0.2050	0.9944	0.0743	0.9491	0.0737	0.9491
GEER-CS (hl-disp)	0.2025	1.1185	0.0748	0.9558	0.0670	0.9558

Table 4.18: $\theta = (1, 1, 3)$, CN at 20% with the Variance Ratio at 4

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0642	1.0000	0.0530	1.0000	0.0541	1.0000
JR	0.1236	1.2669	0.0430	1.0035	0.0511	1.2035
GR (med-mad)	0.0331	0.6837	0.0613	1.4218	0.0490	1.3825
GR (hl-disp)	0.0541	0.9465	0.0643	1.4466	0.0521	1.3857
GEER-AP (med-mad)	0.0941	0.9127	0.0524	1.3639	0.0479	1.3636
GEER-AP (hl-disp)	0.1297	1.2370	0.0499	1.3603	0.0487	1.3604
GEER-CS (med-mad)	0.1618	1.0603	0.0522	1.3625	0.0468	1.3623
GEER-CS (hl-disp)	0.1298	1.1786	0.0512	1.3687	0.0494	1.3687

Table 4.19: $\theta = (1, 1, 3)$, CN at 10% with the Variance Ratio at 4

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0645	1.0000	0.0520	1.0000	0.0543	1.0000
JR	0.1281	1.2347	0.0480	1.0008	0.0484	1.1600
GR (med-mad)	0.0327	0.6527	0.0596	1.3301	0.0524	1.2904
GR (hl-disp)	0.0447	0.8576	0.0615	1.3578	0.0546	1.3002
GEER-AP (med-mad)	0.0968	0.9004	0.0499	1.2719	0.0484	1.2720
GEER-AP (hl-disp)	0.1246	1.1523	0.0473	1.2763	0.0493	1.2763
GEER-CS (med-mad)	0.1664	1.0558	0.0492	1.2700	0.0477	1.2701
GEER-CS (hl-disp)	0.1416	1.1519	0.0479	1.2798	0.0500	1.2799

Table 4.20: $\theta = (1, 1, 1)$, CN at 20% with the Variance Ratio at 2

Methods	Intercept		Treatment		Covariate	
	level	ARE	level	ARE	level	ARE
REML	0.0867	1.0000	0.0525	1.0000	0.0518	1.0000
JR	0.1209	0.9761	0.0052	0.3667	0.0507	0.6475
GR (med-mad)	0.0339	0.5575	0.0485	1.0185	0.0446	1.0106
GR (hl-disp)	0.0405	0.6993	0.0496	1.0098	0.0497	1.0009
GEER-AP (med-mad)	0.1097	0.8017	0.0474	1.0033	0.0456	1.0034
GEER-AP (hl-disp)	0.1354	1.0144	0.0491	0.9937	0.0488	0.9938
GEER-CS (med-mad)	0.1872	0.9780	0.0496	1.0170	0.0480	1.0169
GEER-CS (hl-disp)	0.1658	1.0464	0.0517	1.0029	0.0510	1.0029

4.3 Application to Real World Data

4.3.1 PISA Data Set

The data set was obtained from the OECD Programme for International Student Assessment (PISA) in 2009 (OECD, 2010). Every three years, math, science and reading comprehension skills of students of age 15 in 65 countries are assessed. This is a very large data set. We consider a subset of the data concerning the private schools in USA (334 observations in 11 private schools across the country) on the educational outcome of metacognitive scores. This is an index measure of the metacognitive aspect of learning, which has a skewed distribution. The research interests are how metacognitive scores depend on gender and age, and how the variability of the scores are explained by region differences (here, the regions are Northeast, Midwest, South and West) and school differences. Student scores are nested in the private schools within the regions. Assuming the schools within the same regions are correlated; students within the same schools are correlated; hence, regions and schools could be considered to have random effects on observations in the random effects nested model. (These are clusters so that correlation structures are inherited.) This design is hierarchical with three levels (students in schools within regions). In a 3-level nested study, to measure each level's contribution to the total variability, intra-class correlations are defined as intra-region,

$$\rho_{Region} = \frac{\sigma_{Region}^2}{\sigma^2 + \sigma_{Region}^2 + \sigma_{School(Region)}^2}, \text{ and intra-school, } \rho_{School(Region)} = \frac{\sigma_{Region}^2 + \sigma_{School(Region)}^2}{\sigma^2 + \sigma_{Region}^2 + \sigma_{School(Region)}^2}.$$

In this data set, the equity of educational systems and outcomes are assessed via the calculation of percentage of the variance in student performance explained by cluster levels or hierarchical levels (i.e. for regional or locational differences, intra-region correlation coefficient helps to answer it). The coherence is the total variance between schools expressed as a percentage of the total variance within the region, which is defined as the intra-school correlation coefficient. We show the precision of the rank-based GR analysis with this data set by obtaining contextual parameter estimates. In this analysis, the sampling weights were not considered.

The fixed and scale estimates are reported in Table 4.21. Both the REML and GR agree on the fixed effect inference with little difference. The standard errors confirm the same or slightly better precision of the R estimates compared to the REML analysis. Moreover, in these estimates, there is gain, about 4% to 14%, in favor of the Rank-based analysis. The GR result indicates that 15.1% of the total variability is due to the region differences, 16.6% is due to the school differences within region. These are almost the same in the REML results. For the checking of the quality of both fits, the standardized residuals and the q-q plots in Table 4.22 indicate that the standardized residuals in the GR fit are scattered as in the REML. Also, both q-q plots suggest the skewness of the residuals.

For the sensitivity analysis, an outlier at 100 standard deviations away from the center is introduced in the response space. When this new data set is analyzed, its conclusions and inferences on the contextual interests might be corrupt and misleading. Doing the same analysis with the new data set, PISA results using the REML analysis in Table 4.21 dramatically changed. The fixed effect estimates remained the same for the GR estimates, whereas the REML results changed. In the REML, the age effect was calculated negative, indicating that as the age increases the metacognitive score decreases, -which is unrealistic-. The sign of gender effect in the new data set was reversed, indicating the conflict with the original data set results. The lack of precision of these estimates worsened in the REML, indicating the GR estimates were 900% to 4,500% more efficient.

Table 4.21: Estimates in PISA Data for Cognitive Scores with Two Covariates

Original Data	REML		GR	
<i>Fixed effects</i>	<i>est</i>	<i>se</i>	<i>est</i>	<i>se</i>
Intercept	0.082	0.278	0.350	0.267
Gender	-0.264	0.109	-0.295	0.107
Age	0.246	0.172	0.226	0.161
<i>Variance</i>				
Region	0.275		0.147	
Private School (Region)	0.001		0.015	
Student	0.855		0.812	
<i>Intra-Inter Corr</i>	<i>est</i>		<i>est</i>	
ρ_{Region}	24.3%		15.1%	
$\rho_{School(Region)}$	24.4%		16.6%	
With One Outlier	REML		GR	
<i>Fixed effects</i>	<i>est</i>	<i>se</i>	<i>est</i>	<i>se</i>
Intercept	0.267	0.516	-0.093	0.176
Gender	0.133	0.640	-0.324	0.106
Age	-0.376	1.033	0.185	0.154
<i>Variance</i>				
Region	0.000		0.137	
Private School (Region)	0.000		0.018	
Student	31.085		2.049	
<i>Intra-Inter Corr</i>	<i>est</i>		<i>est</i>	
ρ_{Region}	0%		6.2%	
$\rho_{School(Region)}$	0%		7.0%	

Table 4.22: The Q-Q Plots of PISA Data Analysis

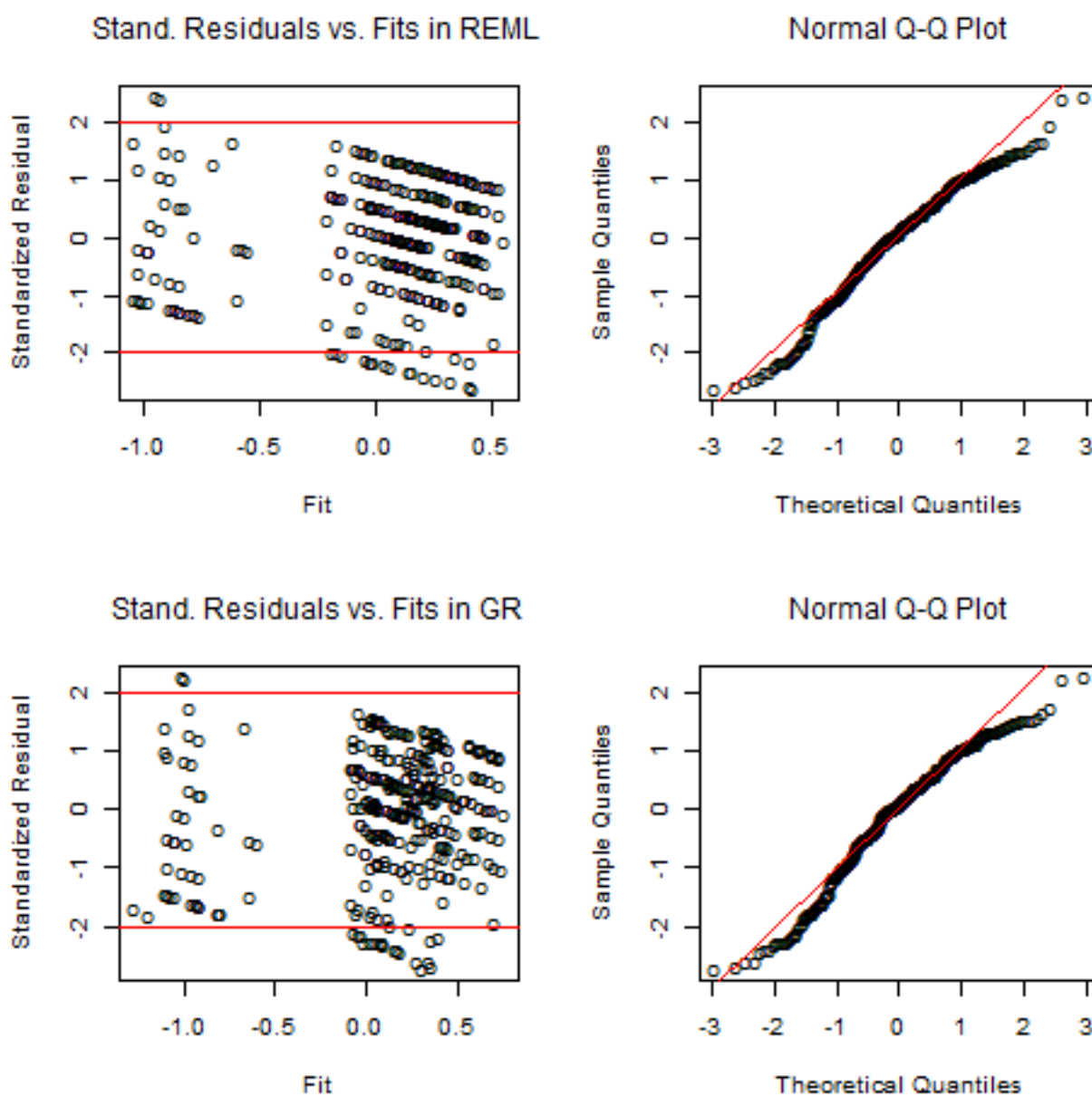


Table 4.23: The Q-Q Plots with Outlier

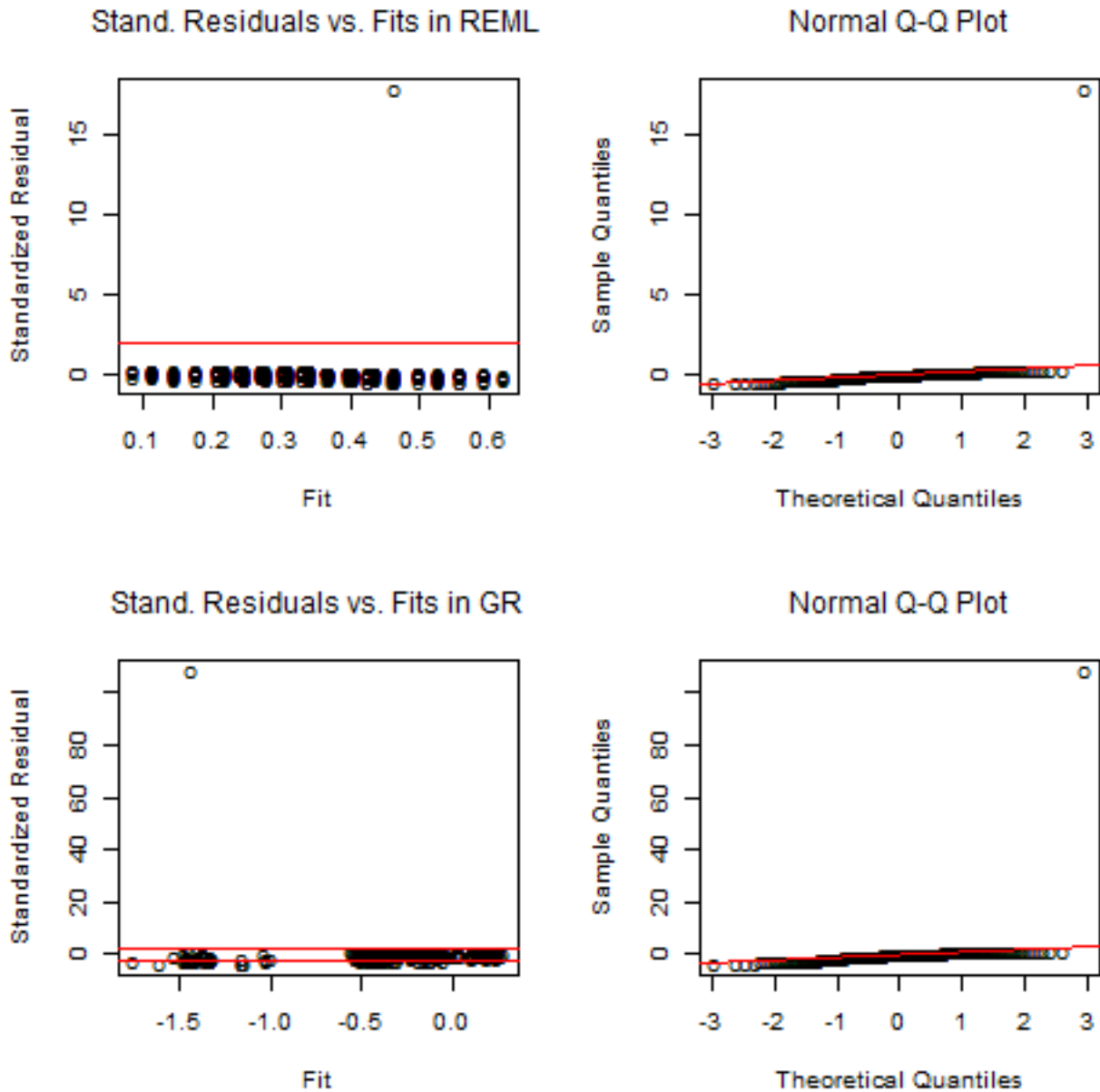
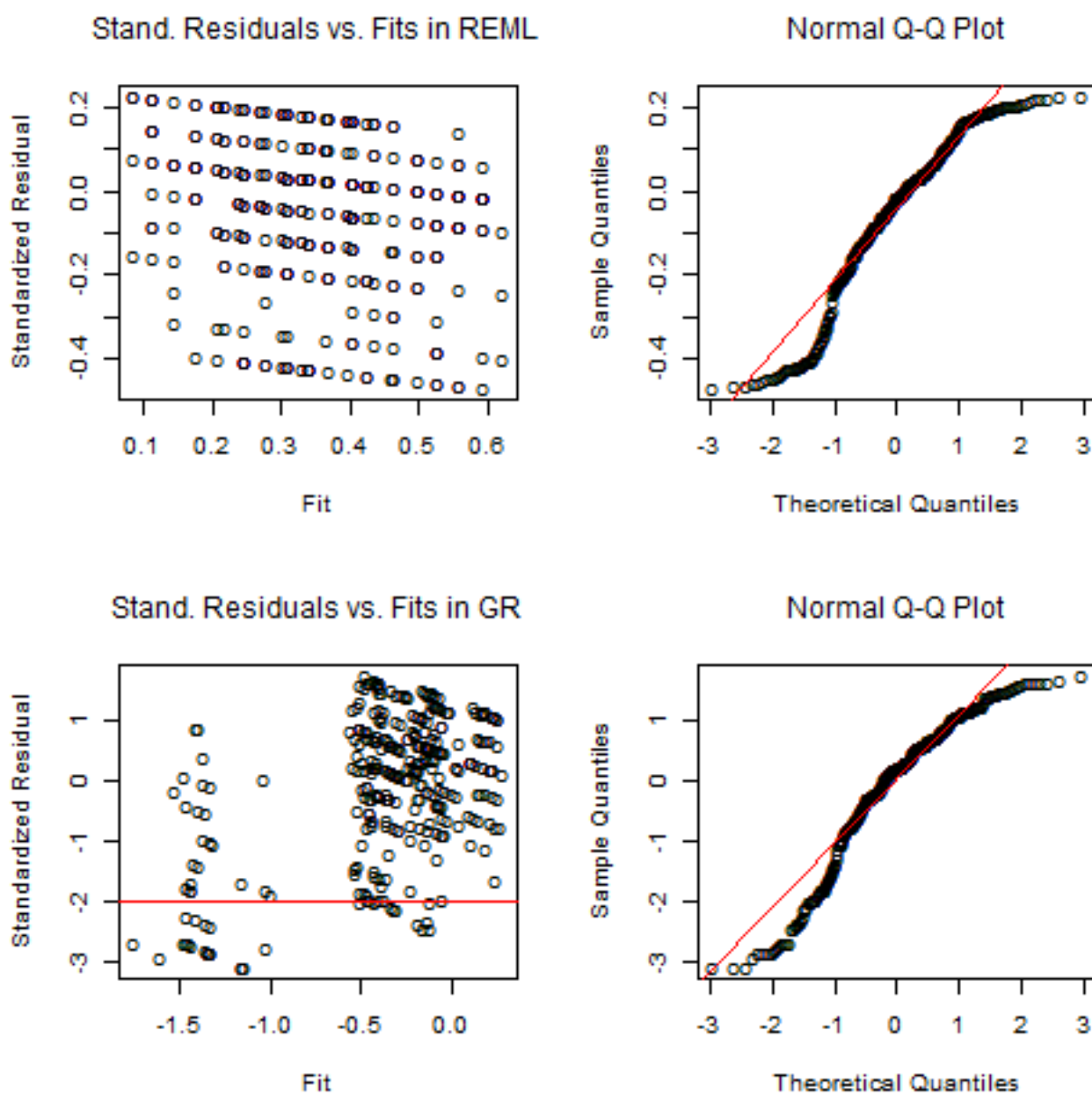


Table 4.24: The Q-Q Plots with Outlier When Omitting the Outlier



As shown in the REML results with the data with one outlier, Table 4.21 indicates that the regions and schools had 0% contribution to the total variability, which is debatable. The GR analysis is not nearly as sensitive to the outlier effect as far the REML analysis (See Tables 4.22, 4.23, and 4.24). Standardized residuals in the REML with outlier analysis and fitted values were distorted and forced to be within the range of $[-.5, .2]$ and $[.1, .6]$, respectively, except the residual's value pertaining to the outlier. These ranges mislead the judgment of the goodness of individual fits and the prediction, indicating the REML fit is clearly poor when the data has an outlier. However, the GR were not affected that much and kept almost the same values of the standardized residuals and fitted values as in the original fit (See Tables 4.22 and 4.24).

4.3.2 PASSIVE Data Set

Next, we consider an application of the nested models with random and treatment effect in industry. A data set consisting of highly correlated data with 72 observations is discussed in Lit tell et al. (2006, p.81). The data are from a study in the semiconductor industry where the objective is to estimate the variance components to determine assignable causes for the observed variability. The outcomes are thicknesses of the oxide layer on silicon wafers determined at three randomly selected sites on each wafer. The variable 'source' has two treatments, one is control, hence, the treatment variable is binary. The wafers stem from eight different lots, and each lot consists of 3 wafers. The process consisted of randomly selecting eight lots, then 3 wafers were selected from each lot for use in the oxide deposition process. After the layer of oxide was deposited, the thickness of the layer was determined at three randomly selected sites on each wafer. The structure of the study involves three sizes of experimental units in the design structure with a uniform application of a single treatment in the treatment structure. Table 4.25 indicates that the estimates of the treatment and scale effects in both methods are consistent. The GR is more precise.

For the sensitivity analysis, an outlier at 15 standard deviations away from the center is

Table 4.25: Estimates for PASSIVE Data with Treatment

	REML		GR	
<i>Fixed effects</i>	<i>est</i>	<i>se</i>	<i>est</i>	<i>se</i>
Intercept	-5.04	5.77	-3.88	8.83
Source	10.08	8.16	8.62	7.40
<i>Variance</i>	<i>est</i>	<i>sd</i>	<i>est</i>	<i>sd</i>
lot	119.89		79.36	
wafer(lot)	35.86		21.47	
error	12.56		8.15	
<i>Intra-Inter Corr</i>	<i>est</i>		<i>est</i>	
ρ_{lot}	71.2%		72.8%	
$\rho_{wafer(lot)}$	92.5%		92.5%	
With One Outlier				
	REML		GR	
<i>Fixed effects</i>	<i>est</i>	<i>se</i>	<i>est</i>	<i>se</i>
Intercept	-5.04	8.28	-4.88	5.04
Source	14.64	11.71	-0.42	4.31
<i>Variance</i>	<i>est</i>	<i>sd</i>	<i>est</i>	<i>sd</i>
lot	194		124	
wafer(lot)	104		62	
error	409		59	
<i>Intra-Inter Corr</i>	<i>est</i>		<i>est</i>	
ρ_{lot}	27%		51%	
$\rho_{wafer(lot)}$	42%		76%	

Table 4.26: The Q-Q Plots of PASSIVE Data Analysis

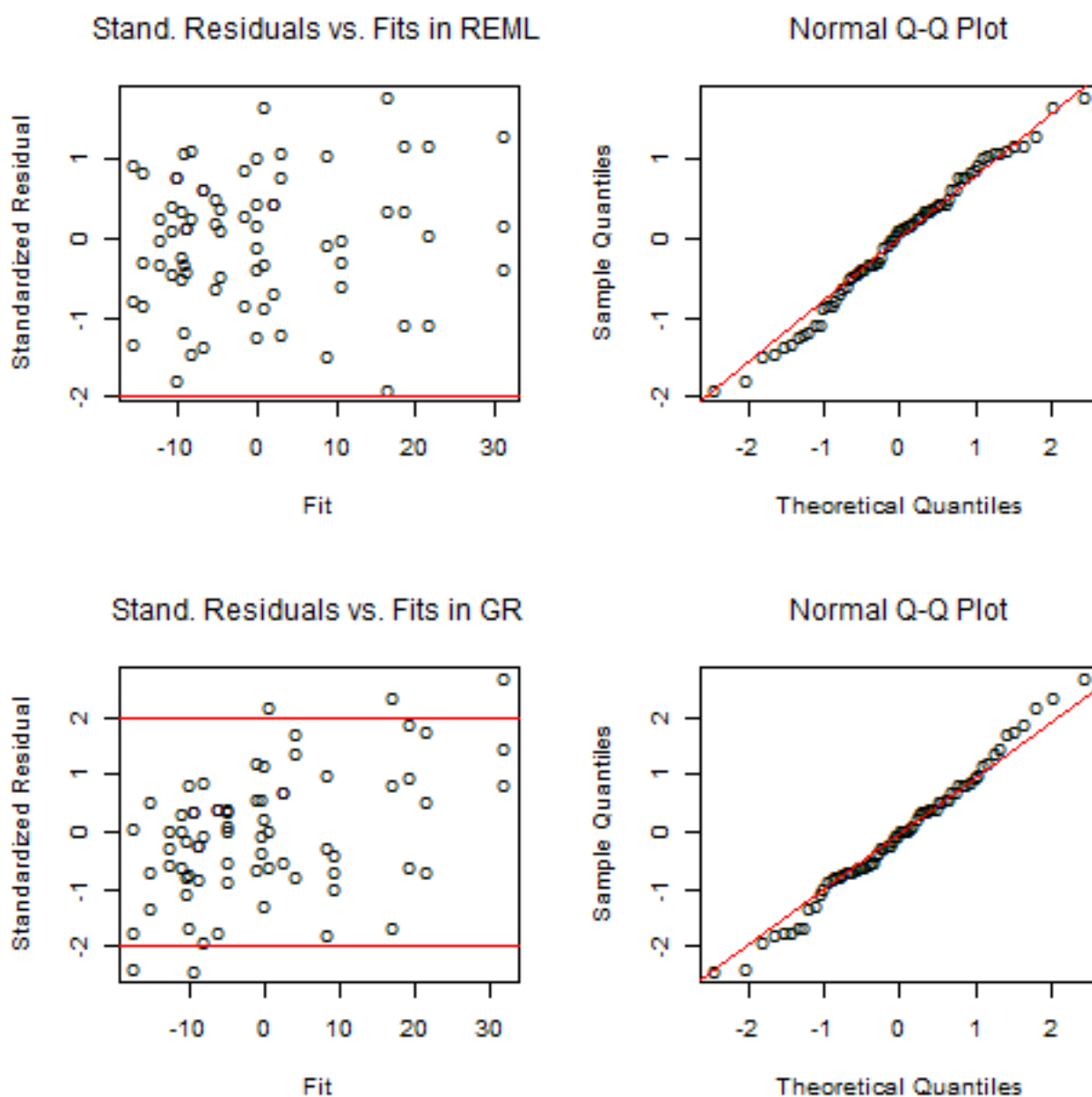


Table 4.27: The Q-Q Plots with Outlier

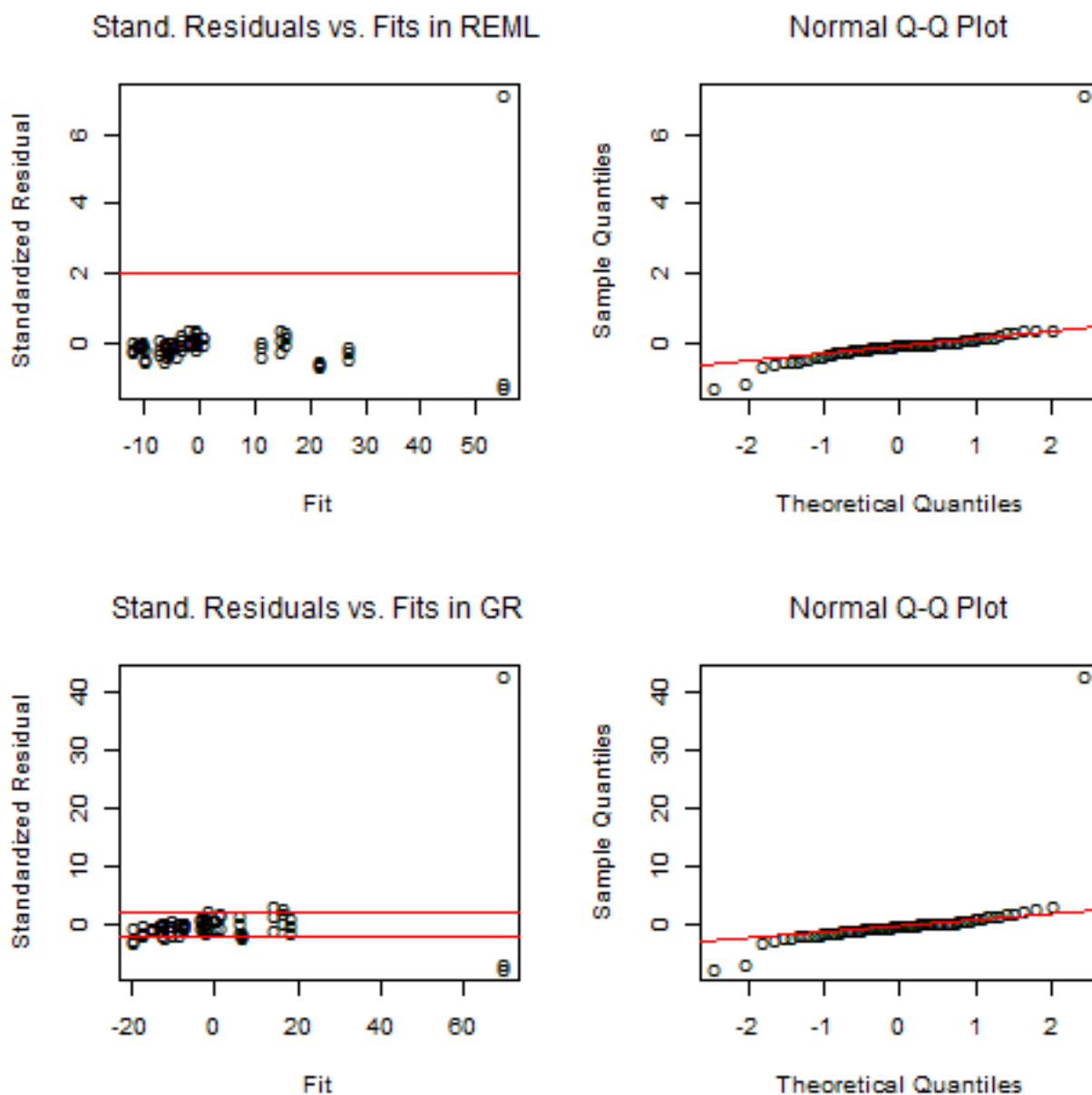
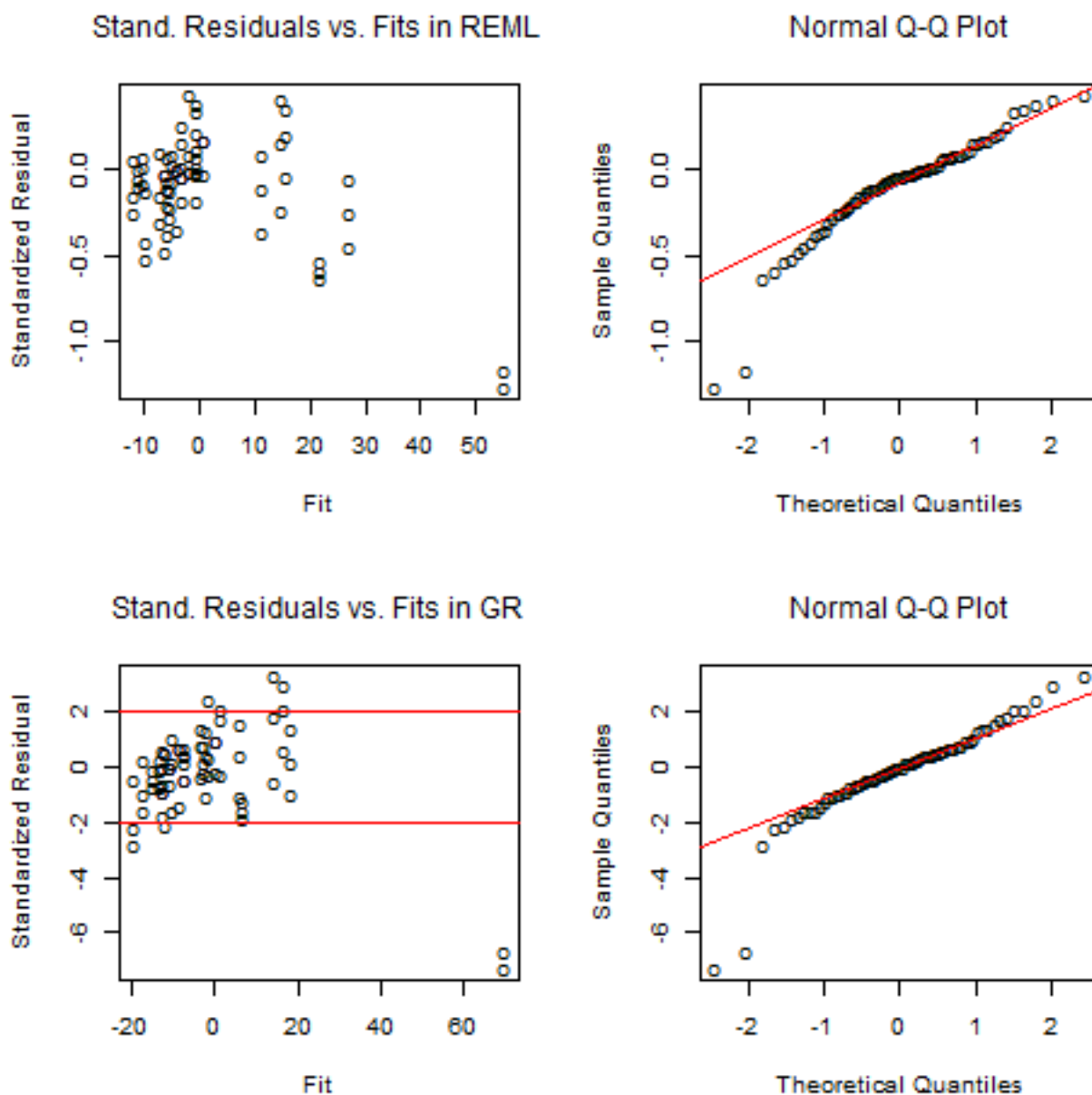


Table 4.28: The Q-Q Plots with Outlier When Omitting the Outlier



introduced in the response space. Passive data results dramatically changed in inferences of the REML analysis. The fixed effect estimates changed in both methods as well. The variance estimates in the GR method were not as sensitive as in the REML's method. The estimate of the lot-to-lot variance of 194 in the REML is 1.5 times larger than the GR. The scale estimate of residual variance of 409 in the REML is 6.9 times larger than the GR. The precision of the scale and fixed effects in the GR was observed better than that of the REML (See Tables 4.25, 4.26, 4.27 and 4.28). In particular, standardized residuals in the REML fit with outlier were affected as to be correlated in Table 4.28 and thus the goodness of the fit is said poor. Also, the fits for the observations in the same cluster with the outlier were distorted. Table 4.28 illustrates that the GR fit still kept the randomness of the distribution of the standardized residuals the same as in the original fit. The outlier did not affect much the other fits and residuals in the GR except the intercept estimate, however, the REML was affected badly in these parts.

Chapter 5

Discussion, Conclusion and Further Study

This work proposes new rank-based fitting methods in handling random, fixed and scale effects in k -level nested designs for estimation and inference. Also, the proposed Rank-based fitting methods including Joint Ranking (JR), Generalized Rank Estimate (GR) and Rank-based General Estimating Equation (GEER), are investigated along with an algorithm for Rank Prediction Procedure (RPP) with two alternatives. In the RPP, the Hodges-Lehman (HL) and median (med) location estimators, and the Dispersion Function (Disp) and Median Absolute Deviation (MAD) scale estimators are employed. These methods with the RPP obtain robust estimation and prediction. Estimation includes fixed effects and their standard errors. Prediction includes random effects. The methods further obtain estimates of the variance components. The RPP is an iterative procedure, hence the GR estimation procedure is a k -step estimate. In this work, asymptotic distribution theory of the GR estimation is developed. Furthermore, the consistency of scale estimate using dispersion function is discussed. For simplicity, a 3-level nested design that deals with students nested within sections in schools is used to employ our methods. These methods can be applied to any mixed models, i.e. k -level nested random models with covariate. For hierarchical models, the RPP algorithm should be adjusted according to the design of random effects.

In this work, the validation and efficiency results of a Monte Carlo investigation including a comparison with the traditional analysis are presented. The proposed methods

are competitive with the traditional method when the errors have normal distributions. Under non-normal situations, these methods are less sensitive than the REML method. When random errors are contaminated, the intra-class correlation estimates in the RPP are unbiased, while the REML estimates are biased.

For the Wilcoxon score, the empirical precision of the estimates based on a Monte Carlo investigation exhibited empirical AREs that indicate the loss is close to 5% relative to that of the REML method when the error distributions are normal. In fact, this is what is expected from the Wilcoxon scores. When correlation structure is increased within nested groups, the GR and GEER outperform the traditional likelihood method (REML). In little correlated or highly correlated data, the validity of the proposed methods is confirmed, using the empirical significance level at 5%. The proposed methods have the empirical alpha level of nearly 5% in treatment and covariate estimates. However, for the intercept estimates, it is observed around 10%, which is liberal ($>5\%$). This occurs in the REML as well.

Based on our investigations, the REML analysis is quite sensitive to outliers, whereas the proposed Rank-based methods are robust. Two real data analyses further validate this. When an outlier is introduced to response space in these examples, it is observed that the Rank-based methods are less sensitive to the outlier, hence, more reliable inferences are obtained relative to the traditional likelihood method.

Further research is recommended on the power, finite correction, geometry for the intercept in Rank-based analysis, the convergence of residual weights in the GEER, and fitting without intercept in the GR. Also, the growth rate of variance of empirical distribution function should be investigated. The extension of k-level nested designs with random effect and covariates should be applied to random coefficient models in hierarchical designs. Also, poor empirical alpha level and efficiency results of the JR analysis for correlated data could be fixed by augmenting the random effects as fixed to the model employed in some conventional techniques (see Pinheiro et al., 2011). This will be the next investigation.

Computation of the rank estimators are easily accessible for R Software users. We aim to develop a complete R Software package, called *R-HLM* for two- and three-level nested cases, which are common in practice. This will make use of the R Software package *Rfit* developed by Kloke and McKean (2012).

References

- [1] Akritas, M.G. and Brunner, E. (1997). A Unified Approach To Rank Tests For Mixed Models. *J. Statist. Plan. Infer.* 61, 249-277
- [2] Abebe, A., McKean, J. W., and Kloke, J. D. (2010). Iterated Reweighted Rank-Based Estimates for GEE Models. Submitted.
- [3] De Leeuw, J. and Meijer E. (2008). *Handbook of Multilevel Analysis*, New York: Springer.
- [4] Brunner E. and Denker M. (1994). Rank statistics under dependent observations and applications to factorial designs. *J. Statist. Plan. Infer.* 42, 353-378.
- [5] Dubnicka, S. R. (2004). A rank-based estimation procedure for linear models with clustered data. *Journal of Modern Applied Statistical Methods*, 3 (1), 39–48.
- [6] Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73 (1), 43–56.
- [7] Goldstein, H. (1995). *Multilevel Statistical Models*. Edward Arnold Publishers Ltd.
- [8] Groggel, D. J. (1983). “Asymptotic nonparametric confidence intervals for the ratio of scale parameters in balanced one-way random effects models”, Department of Statistics, University of Florida.
- [9] 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.

- [10] Harville, David A. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, Vol. 72 (358): 320–338.
- [11] Herbert T., David. and Nagaraja, H. (2003). *Order Statistics*. Wiley Series in Probability and Statistics.
- [12] Hettmansperger, T. P. and McKean, J. W. (2011). *Robust nonparametric statistical methods*. 2nd Ed., New York:Chapman-Hall.
- [13] Hill, P. W. and Rowe, K. J. (1996). Multilevel Modelling in School Effectiveness Research. *School Effectiveness and School Improvement: An International Journal of Research*, 7 (1), 1-34.
- [14] Hollander, M. and Wolfe, D. A. (1999). *Nonparametric statistical methods*. Second ed., Wiley Series in Probability and Statistics: Texts and References Section. New York: John Wiley & Sons Inc. A Wiley-Interscience Publication.
- [15] Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann Math Stat*, 1972, 43:1449-1458.
- [16] Jureckova, A. (1971). Nonparametric estimate of regression coefficients. *Annals of Mathematical Statistics*, 42, 1328-1338.
- [17] Kloeke, J. D. and McKean, J. M. (2012). “Rfit: Rank Estimation for Linear Models”. *R Package Version 1*.
- [18] Kloeke, J. D., McKean, J.W., and Rashid, M. M. (2009). Rank-based estimation and associated inferences for linear models with cluster correlated errors. *Journal of the American Statistical Association*, 104 (485), 384–390.
- [19] Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.

- [20] Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). *SAS for Mixed Models*, Second Edition, Cary, NC: SAS Institute Inc.
- [21] McKean, J. W. and Hettmansperger, T. P. (1978). A robust analysis of the general linear model based on one-step R-estimates. *Biometrika*, 65, 571-9.
- [22] McKean, J. W., Sheather, S. J. and Hettmansperger, T. P. (1990). Regression diagnostics for rank-based methods, *Journal of the American Statistical Association*, 85, 1018-1028
- [23] O'Connell, A.A. and McCoach, D.B. (Eds.) (2008). *Multilevel Modeling of Educational Data*. Charlotte, NC: Information Age Publishing.
- [24] OECD (2010) PISA 2009 Results. <http://www.oecd.org/>
- [25] Pinheiro, J., Bates, D., DebRoy S., Sarkar D., and R Development Core Team. (2011). "nlme: Linear and Nonlinear Mixed Effects Models", *R package version* 3.1-98.
- [26] Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. 2nd edition. Newbury Park, CA: Sage.
- [27] Robert J. Serfling. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics.
- [28] Sahai, H. and Ojeda, M. M. (2005). *Analysis of Variance for Random Models Volume 2: Unbalanced Data, Theory, Methods, Applications and Data Analysis*. Boston: Birkhauser.
- [29] Searle, S. R., Casella, G., and McCulloch, C. E. (2006). *Variance Components* (Wiley Series in Probability and Statistics). New York: Wiley-Interscience.
- [30] Sievers, G. L. & Abebe, A. (2004). Rank Estimation of Regression Coefficients Using Iterated Reweighted Least Squares. *Journal of Statistical Computation and Simulation*, 74, 821–831.

- [31] Thompson, G. L. (1990). Asymptotic Distributions of Rank Statistics Dependencies with Multivariate Applications. *Journal of Multivariate Analysis*, 33, 183-211.
- [32] Terpstra, J. and McKean, J. (2005). Rank-based analyses of linear models using R. *Journal of Statistical Software*, 14 (7), 1–26.
- [33] Terpstra, J. and McKean, J. Powerful Rank-Based Procedures for Educational Research Designs. *Unpublished Manuscript*. Western Michigan University.

Appendix

Assumptions

Assumptions required for the asymptotic theory in the rank-based analysis:

The regularity conditions are:

J1. $N^{-1}X^TX$ converges to a positive definite matrix where X is centered design matrix. $n_i^{-1} \cdot X_i^T \cdot X_i \rightarrow \Sigma_i > 0$ as $n_i \rightarrow \infty$.

J2. Leverage values converge to zero. (leverage value assumption of the projection matrix onto range X_{ij})

J3. $f(t)$ and $F(t)$ denote continuous pdf and cdf of errors ϵ and $f(t)$ has finite Fisher information.

J4. Exchangeable: Assuming e_i 's distribution is exchangeable: i.e.,

$$L(e_{i1}, e_{i2}, \dots, e_{iJ_i}) = L(e_{i\alpha_1}, e_{i\alpha_2}, \dots, e_{i\alpha_i})$$

L denotes distribution and α 's are permutation of integers $1, \dots, J_i$.

J5. (This is more general than J4.) All univariate marginal distributions of the random errors are the same.

J6. Errors are continuous variables; X_i full column design matrix. Observations between schools are independent. Random errors are uncorrelated.