Heteroscedastic Linear Model Estimation Based on Ranks: An Iterated Reweighted Least Squares Approach

Themba Louis Nyirenda
Western Michigan University

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

Recommended Citation
https://scholarworks.wmich.edu/dissertations/690
HETEROSCEDASTIC LINEAR MODEL ESTIMATION BASED ON RANKS: AN ITERATED REWEIGHTED LEAST SQUARES APPROACH

by

Themba Louis Nyirenda

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: Gerald Sievers, Ph.D.

Western Michigan University
Kalamazoo, Michigan
June 2009
For standard estimators, data that are heteroscedastic in nature contain outlying values which can lead to poor performance. In this study, we present a robust iterative method for estimating the location and scale parameters in the general linear model, using a rank based method. It is assumed that the errors are symmetric about 0 and the variance function model is nonlinear with respect to the scale coefficients and the design. The function is known up to a scale constant.

We propose taking the logarithm of the absolute values of the variance function to linearize it. The rank estimation of the scale coefficients amounts to regressing logs of absolute residuals from an initial rank based fit on to the design. The resulting scale coefficient estimates are used to form scale constants in a weighted signed-rank method. Thus, iterating between these two rank based methods leads to the desired estimates that are obtained from linear model fits for the both types of coefficients.

For the heteroscedastic linear model under consideration, this study has made the following contributions: (1) the asymptotic normality results that are established here show that the estimators are both consistent and highly efficient; (2) in each estimation problem, the Iterated Reweighted Least Squares (IRWLS) formulation for rank methods of Sievers and Abebe (2004) is employed with the other parameter substituted by their corresponding estimates from an appropriate iteration; (3) the high efficiency and good robustness qualities of the proposed method are confirmed by simulation trials that were conducted in two-sample problem, several groups and general linear models; (4) the inlier issue that is a consequence of employing the log transformation is also investigated and shown to be well curtailed by the proposed method and (5) finally, the method is
shown to outperform other methods when applied to real life data from a Psychiatric Clinical Trial containing two treatments, one covariate, and one confounding variable.

Thus, for samples larger than 20, the proposed method is highly robust and efficient under non-normal distributions.
UMI Number: 3364681

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction. In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
ACKNOWLEDGMENTS

I am very grateful to my advisor Dr. Gerald L. Sievers whose continued detailed guidance, patience, insights made this possible. I also wish to thank members of my advisory committee, Dr. Joseph W. McKean and Dr. Magdalena Niewiadomska-Bugaj, and Dr. Qiji Zhu for their comments and very supportive spirit that was always there for me.

I would like to thank all the staff members of the Department of Statistics and Department of Mathematics for assisting me with administrative matters, and financial support at various times during my program at WMU.

Finally, I wish to express my gratitude to my parents, siblings and friends for their very generous support, unfailing encouragement and patience.

This manuscript is dedicated to my parents, Dr. Stanley Nyirenda and Cynthia Nyirenda, my sister, Dr. Thandiwe Nyirenda, and brother, Stanley Nyirenda.

Themba Louis Nyirenda
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS ........................................................................... ii</td>
</tr>
<tr>
<td>LIST OF TABLES ................................................................................. ix</td>
</tr>
<tr>
<td>LIST OF FIGURES ................................................................................ xii</td>
</tr>
<tr>
<td>CHAPTER</td>
</tr>
<tr>
<td>I. HETEROSEDASTIC LINEAR MODELS .................................................. 1</td>
</tr>
<tr>
<td>1.1 Introduction ............................................................................... 1</td>
</tr>
<tr>
<td>1.2 Heteroscedastic Linear Regression .............................................. 2</td>
</tr>
<tr>
<td>1.3 Previous Work ............................................................................ 3</td>
</tr>
<tr>
<td>1.4 Research Problem ...................................................................... 4</td>
</tr>
<tr>
<td>1.4.1 Estimation Criterion ............................................................... 5</td>
</tr>
<tr>
<td>1.4.2 Weighted Least Squares Estimation .......................................... 8</td>
</tr>
<tr>
<td>1.4.3 Rank Estimates: Other Parameters are Specified ..................... 9</td>
</tr>
<tr>
<td>1.4.4 Rank Estimates: Other Parameters are Unknown ..................... 10</td>
</tr>
<tr>
<td>II. LOCATION AND SCALE MODELS .................................................... 12</td>
</tr>
<tr>
<td>2.1 Introduction .............................................................................. 12</td>
</tr>
<tr>
<td>2.2 Heteroscedastic Linear Model .................................................... 13</td>
</tr>
<tr>
<td>2.3 Regression Coefficient Estimation .............................................. 13</td>
</tr>
<tr>
<td>2.3.1 Residuals ................................................................................ 13</td>
</tr>
<tr>
<td>2.3.2 Estimation Criterion ............................................................... 14</td>
</tr>
<tr>
<td>2.3.3 Model Assumptions ................................................................. 16</td>
</tr>
<tr>
<td>2.3.4 Asymptotic Linearity of $S_{1N}(b, 0, 0)$ ................................ 18</td>
</tr>
<tr>
<td>2.4 Scale Parameter Estimation ....................................................... 22</td>
</tr>
<tr>
<td>2.4.1 Residuals .............................................................................. 23</td>
</tr>
<tr>
<td>2.4.2 Estimation Criterion ............................................................... 24</td>
</tr>
<tr>
<td>2.4.3 Additional Assumptions .......................................................... 24</td>
</tr>
</tbody>
</table>
Table of Contents—Continued

CHAPTER

2.4.4 Asymptotic Linearity of $S_{2N}(0, t)$ ........................................ 28

2.5 General Asymptotic Results .................................................. 30

III. RANK ESTIMATION OF REGRESSION PARAMETER, $\beta$ ............... 32

3.1 Introduction ............................................................................ 32

3.2 Model for Location Problem .................................................. 33

3.2.1 Scaling Transformation ..................................................... 34

3.2.2 Weighted Estimation Problem ............................................ 34

3.2.3 Model Assumptions .......................................................... 36

3.2.4 Linear Signed-Rank Statistic ............................................. 37

3.3 Asymptotic Properties of Estimator ......................................... 39

3.3.1 Probability Distribution of Errors ...................................... 39

3.3.2 Distributional Properties of Linear Signed-Rank Statistics ...... 40

3.3.3 Asymptotic Linearity of Shifted Process $S_{1N}(b, 0, 0)$ ............ 49

3.4 Asymptotic Linearity Result .................................................. 55

3.4.1 Asymptotic Properties of $S_{1N}(b, t, s)$ .............................. 60

3.4.2 Asymptotic Linearity of $S_{1N}(b, 0, s)$ ............................... 61

3.4.3 Asymptotic Linearity of $T_{1N}(b, t, s)$ ............................... 63

3.4.4 Asymptotic Linearity of $S_{1N}(b, t, s)$ ............................... 73

3.4.5 Asymptotic Uniform Linearity .......................................... 74

3.4.6 Application of Linearity Result ......................................... 76

3.5 Dispersion Function Criterion ............................................... 85

3.5.1 Asymptotic Uniform Quadraticity $D_{1N}(b, t)$ ..................... 87

3.6 Conclusion ............................................................................ 90

IV. RANK ESTIMATION OF SCALE PARAMETER, $\theta$ ..................... 91
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>91</td>
</tr>
<tr>
<td>4.2 Model for Scale Problem</td>
<td>92</td>
</tr>
<tr>
<td>4.2.1 Log Absolute Value Transformation</td>
<td>93</td>
</tr>
<tr>
<td>4.2.2 Estimation Problem</td>
<td>94</td>
</tr>
<tr>
<td>4.2.3 Model Assumptions</td>
<td>95</td>
</tr>
<tr>
<td>4.2.4 Linear Rank Statistic</td>
<td>97</td>
</tr>
<tr>
<td>4.3 Asymptotic Properties of Estimator</td>
<td>98</td>
</tr>
<tr>
<td>4.3.1 Distributive Properties of Errors</td>
<td>98</td>
</tr>
<tr>
<td>4.3.2 Distributional Properties of Linear Rank Statistics</td>
<td>100</td>
</tr>
<tr>
<td>4.3.3 Asymptotic Linearity of $S_{2N}(0, t)$</td>
<td>109</td>
</tr>
<tr>
<td>4.4 Asymptotic Linearity Result</td>
<td>113</td>
</tr>
<tr>
<td>4.4.1 Asymptotic Linearity of $S_{2N}(b, t)$</td>
<td>119</td>
</tr>
<tr>
<td>4.4.2 Asymptotic Linearity of $T_{2N}(b, t)$</td>
<td>120</td>
</tr>
<tr>
<td>4.4.3 Distributive Properties of Random Variables</td>
<td>124</td>
</tr>
<tr>
<td>4.4.4 Translation Properties</td>
<td>127</td>
</tr>
<tr>
<td>4.4.5 Asymptotic Linearity of $S_{2N}(b, t)$</td>
<td>130</td>
</tr>
<tr>
<td>4.4.6 Asymptotic Uniform Linearity of $S_{2N}(b, t)$</td>
<td>132</td>
</tr>
<tr>
<td>4.4.7 Application of the Linearity Result</td>
<td>133</td>
</tr>
<tr>
<td>4.5 Dispersion Function Criterion</td>
<td>141</td>
</tr>
<tr>
<td>4.5.1 Asymptotic Uniform Quadraticity of $D_{2N}(b, t)$</td>
<td>143</td>
</tr>
<tr>
<td>4.6 Conclusion</td>
<td>146</td>
</tr>
<tr>
<td>V. IMPLEMENTATION OF THE ITERATIVE METHOD</td>
<td>147</td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>147</td>
</tr>
<tr>
<td>5.2 Score Functions</td>
<td>147</td>
</tr>
</tbody>
</table>
Table of Contents—Continued

CHAPTER

5.3 Estimation of the Dispersion Parameters ........................................ 151
  5.3.1 Estimation of $\gamma_1$ ......................................................... 151
  5.3.2 Estimation of $\gamma_2$ ......................................................... 152

5.4 Iterative Estimation of $\beta$ and $\theta$ ........................................ 153
  5.4.1 Iterative Structure of the Procedure .................................... 154
  5.4.2 Iterative Weighted Estimation of $\beta$ ................................. 155
  5.4.3 Iterative Weighted Estimation of $\theta$ .................................. 157
  5.4.4 Iterative-IRWLS for Procedure for Estimation $\beta$ and $\theta$ ...... 159
  5.4.5 The Algorithm for $(\hat{\beta}^{(r+1)}, \hat{\theta}^{(r+1)})$ ............... 161

5.5 Conclusion ................................................................................. 164

VI. OTHER METHODS NOT BASED ON RANKS .................................... 165

6.1 Introduction ................................................................................. 165

6.2 The Two Sample Problem ............................................................ 165
  6.2.1 Estimation of Difference in Two Means, $\mu_2 - \mu_1$ ............... 165
  6.2.2 Estimation of Confidence Interval Estimate for $\mu_2 - \mu_1$ ....... 166
  6.2.3 Estimation of the Difference in Two Trimmed Means ............... 166
  6.2.4 Estimation of Confidence Interval for Trimmed $\vartheta_2/\vartheta_1$ .... 168

6.3 $M$-Estimation for $p$ Group Problem ........................................ 169
  6.3.1 Estimation of Group 1 to Groups Shifts in Location .................. 170
  6.3.2 Estimation of Ratio of Scale for Group to Group 1 .................... 171
  6.3.3 Algorithm for the Iterative $M$-estimation ............................... 172

6.4 Conclusion ................................................................................. 173

VII. SIMULATION EXPERIMENTS ..................................................... 174

7.1 Introduction ................................................................................. 174
| CHAPTER | 7.10.2 Estimates of Scale Coefficients | 215 |
| CHAPTER | 7.10.3 Estimated Relative Efficiency of the Iterative Methods | 217 |
| CHAPTER | 7.11 Concluding Remarks for the Simulation Trials | 219 |
| VIII. CASE STUDY: BEATING THE BLUES DATA | 223 |
| VIII. CASE STUDY: BEATING THE BLUES DATA | 8.1 A Psychiatric Clinical Trial on the Beating the Blues Intervention | 223 |
| VIII. CASE STUDY: BEATING THE BLUES DATA | 8.2 Background of the BtB study | 224 |
| VIII. CASE STUDY: BEATING THE BLUES DATA | 8.3 Model | 225 |
| VIII. CASE STUDY: BEATING THE BLUES DATA | 8.4 Boxplots of the BDI Scores by Length of Current Episode | 226 |
| VIII. CASE STUDY: BEATING THE BLUES DATA | 8.5 Scatterplots of BDI vs BDI-pre by Length of Current Episode | 227 |
| VIII. CASE STUDY: BEATING THE BLUES DATA | 8.6 Subjects Whose Current Episode was Less Than 6 Months | 229 |
| VIII. CASE STUDY: BEATING THE BLUES DATA | 8.7 Subjects Whose Current Episode was Greater Than 6 Months | 231 |
| VIII. CASE STUDY: BEATING THE BLUES DATA | 8.8 Residual Analysis by Length of Current Episode | 233 |
| VIII. CASE STUDY: BEATING THE BLUES DATA | 8.9 Score Function Validity by Length of Current Episode | 233 |
| VIII. CASE STUDY: BEATING THE BLUES DATA | 8.10 Concluding Remarks on the BtB Study | 235 |
| IX. CONCLUSION | 238 |
| IX. CONCLUSION | 9.1 Introduction | 238 |
| IX. CONCLUSION | 9.2 Concluding Remarks | 238 |
| IX. CONCLUSION | 9.3 Future Research | 240 |
| REFERENCES | 244 |
LIST OF TABLES

1. Summary of Methods Used in the Two Sample Location Simulation Trials ........................................... 178
2. Averages and Standard Deviations of Estimates of Difference in Location Based on 10,000 Simulations for $\mu_2 - \mu_1 = 2.0$, $\eta = \vartheta_2 / \vartheta_1 = 3.0$ under $CN(0.20, 9)$ Errors ................................................................. 179
3. Summary of Confidence Interval Estimation for Shift in Location for the Two Sample Case .................................................. 182
4. Empirical Levels for 90% and 95% Confidence Interval Estimates of Difference in Location Based on 10,000 Simulations for $\mu_2 - \mu_1 = 2.0$, $\eta = \vartheta_2 / \vartheta_1 = 3.0$ under $CN(0.20, 9)$ Errors ................................................................. 183
5. Summary of Methods Used in the Two Sample Scale Simulation Trials .......... 186
6. Averages and Standard Deviations of Ratio of Scale Estimates Based on 10,000 Simulations for $\mu_2 - \mu_1 = 2.0$, $\eta = \vartheta_2 / \vartheta_1 = 3.0$ under $CN(0.20, 9)$ Errors ................................................................. 187
7. Summary of the Confidence Interval Estimation for the Ratio of Spread for the Two Sample Case .................................................. 189
8. Empirical Levels for 90% and 95% Confidence Interval Estimates of Ratio of Scale Based on 10,000 Simulations for $\mu_2 - \mu_1 = 2.0$, $\eta = \vartheta_2 / \vartheta_1 = 3.0$ under $CN(0.20, 9)$ Errors ................................................................. 190
9. Summary of Methods Used in the $p$ Group Shift in Location Problem .......... 195
10. Summary of Methods Used in the Estimation of Ratio of Scale With Respect to Group 1 ............................................. 196
11. Average and Standard Deviation of Shifts of Each Group from Group 1 Based on 10,000 Simulations for case $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\vartheta_2 / \vartheta_1) = 1.5$, $\log(\vartheta_3 / \vartheta_1) = 3.0$, under the Normal and Contaminated Normal Distributions ............................................. 198
12. Average and Standard Deviation of Shifts of Each Group from Group 1 Based on 10,000 Simulations for case $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\vartheta_2 / \vartheta_1) = 1.5$, $\log(\vartheta_3 / \vartheta_1) = 3.0$, under the Slash and Laplace Distributions ............................................. 199
13. Estimated Relative Efficiency of Shift in Location Based on 10,000 Simulations where $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\vartheta_2 / \vartheta_1) = 1.5$, $\log(\vartheta_3 / \vartheta_1) = 3.0$, under the Normal and Contaminated Normal Distributions ............................................. 201
14. Estimated Relative Efficiency of Shift in Location Based on 10,000 Simulations where $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\vartheta_2/\vartheta_1) = 1.5$, $\log(\vartheta_3/\vartheta_1) = 3.0$, under the Slash and Laplace Distributions ........................................ 202

15. Averages and Standard Deviation of Logarithm of the Ratio of Scale for Each Group With Respect to Group 1, where $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\vartheta_2/\vartheta_1) = 1.5$, $\log(\vartheta_3/\vartheta_1) = 3.0$, under the Normal and Contaminated Normal Distributions ........................................................................ 202

16. Averages and Standard Deviation of Logarithm of the Ratio of Scale for Each Group With Respect to Group 1, where $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\vartheta_2/\vartheta_1) = 1.5$, $\log(\vartheta_3/\vartheta_1) = 3.0$, under the Slash and Laplace Distributions ........................................................................ 203

17. Estimated Relative Efficiency of Scale Coefficient With Respect to Group 1 Based on 10,000 Simulations for $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\vartheta_2/\vartheta_1) = 1.5$ and $\log(\vartheta_3/\vartheta_1) = 3.0$, under the Normal and Contaminated Normal Distributions ......................................................... 204

18. Estimated Relative Efficiency of Scale Coefficient With Respect to Group 1 Based on 10,000 Simulations for $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\vartheta_2/\vartheta_1) = 1.5$ and $\log(\vartheta_3/\vartheta_1) = 3.0$, under the Slash and Laplace Distributions ........................................................................ 205

19. Averages and Standard Deviations of Estimates of Logarithm of Ratio of Spread Based on 10,000 Simulations where $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\vartheta_2/\vartheta_1) = 1.5$, $\log(\vartheta_3/\vartheta_1) = 3.0$, under Inlier Contaminated Distribution ........................................ 207

20. Averages and Standard Deviations of Estimates of Logarithm of Ratio of Spread Based on 10,000 Simulations where $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\vartheta_2/\vartheta_1) = 1.5$, $\log(\vartheta_3/\vartheta_1) = 3.0$, under Inlier-Outlier Contaminated Distributions ........................................................................ 208

21. Estimated Relative Efficiency of Logarithm of Ratio of Spread Based on 10,000 Simulations where $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\vartheta_2/\vartheta_1) = 1.5$, $\log(\vartheta_3/\vartheta_1) = 3.0$, under Inlier Contaminated Distribution ........................................ 210

22. Estimated Relative Efficiency of Logarithm of Ratio of Spread Based on 10,000 Simulations where $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\vartheta_2/\vartheta_1) = 1.5$, $\log(\vartheta_3/\vartheta_1) = 3.0$, under Inlier-Outlier Contaminated Distributions ........................................................................ 211

23. Averages and Standard Deviations of Estimates of Regression Coefficients Based on 10,000 Simulations where $\beta_0 = 2.0$, $\beta_1 = 3.0$, $\beta_2 = 4.0$, $\theta_1 = 1.25$, $\theta_2 = 1.50$, under Standard Normal and Contaminated Error Distributions .. 214
24. Averages and Standard Deviations of Estimates of Regression Coefficients Based on 10,000 Simulations where $\beta_0 = 2.0, \beta_1 = 3.0, \beta_2 = 4.0, \theta_1 = 1.25, \theta_2 = 1.50$, under Slash and Laplace Error Distributions | 215

25. Averages and Standard Deviations of Estimates of Scale Coefficients Based on 10,000 Simulations where $\beta_0 = 2.0, \beta_1 = 3.0, \beta_2 = 4.0, \theta_1 = 1.25, \theta_2 = 1.50$, under Standard Normal, Contaminated Normal Error Distributions | 216

26. Averages and Standard Deviations of Estimates of Scale Coefficients Based on 10,000 Simulations where $\beta_0 = 2.0, \beta_1 = 3.0, \beta_2 = 4.0, \theta_1 = 1.25, \theta_2 = 1.50$, under Slash and Laplace Error Distributions | 217

27. Estimated Relative Efficiency of Regression Coefficients Based on 10,000 Simulations where $\beta_0 = 2.0, \beta_1 = 3.0, \beta_2 = 4.0, \theta_1 = 1.25, \theta_2 = 1.50$, under Standard Normal, Contaminated Normal $=CN(0.05, 100)$, Slash and Laplace Error Distributions | 218

28. Estimated Relative Efficiency of Scale Coefficients Based on 10,000 Simulations where $\beta_0 = 2.0, \beta_1 = 3.0, \beta_2 = 4.0, \theta_1 = 1.25, \theta_2 = 1.50$, under Standard normal-$N(0,1)$, contaminated normal-$CN(0.05, 100)$, Slash and Laplace Error Distributions | 219

29. Length of Current Episode Less Than 6 Months; Fitted Location Model Includes Constant BDI Value ($\beta_0$), and Effects: Treatment ($\beta_1$), BDI-pretreatment ($\beta_2$) and Interaction between Treatment and BDI-pretreatment ($\beta_3$) | 230

30. Length of Current Episode Less Than 6 Months; Fitted Scale Model Includes Effects: Treatment ($\theta_1$), BDI-pretreatment ($\theta_2$), and Interaction between Treatment and BDI-pretreatment ($\theta_3$) | 231

31. Length of Current Episode Greater Than 6 Months; Fitted Location Model Includes Constant BDI Value ($\beta_0$), and Effects: Treatment ($\beta_1$), BDI-pretreatment ($\beta_2$), and Interaction between Treatment and BDI-pretreatment ($\beta_3$) | 232

32. Length of Current Episode Greater Than 6 Months; Fitted Scale Model Includes Effects: Treatment ($\theta_1$), BDI-pretreatment ($\theta_2$), and Interaction between Treatment and BDI-pretreatment ($\theta_3$) | 233
LIST OF FIGURES

1. Plots of Score Function for the Location Problem under Symmetric Distributions: (a) Normal, (b) Logistic, and (c) Laplace ........................................... 149

2. Plots of Score Function for the Scale Problem under Symmetric Distributions: (a) Normal, (b) Logistic, and (c) Laplace ........................................... 150

3. Boxplots of BDI Scores at 2 Months by Treatment Group for (a) Length of Current Episode at Baseline Less Than 6 Months; (b) Length of Current Episode at Baseline Greater Than 6 Months ........................................... 227

4. Scatter Plot BDI Scores at 2 Months Versus BDI-pre Scores for (a) Subjects Whose Current Episode Less Than 6 Months and (b) Subjects Whose Current Episode Greater Than 6 Months at Baseline ........................................... 228

5. IRHET Fit Residual Plot of BDI for Subjects Whose Current Episode Less Than 6 Months at Baseline Against (a) BtB and (b) BDI-pre Less Than 6 Months at Baseline ........................................... 234

6. IRHET Fit Residual Plot of BDI for Subjects Whose Current Episode Greater Than 6 Months at Baseline Against (a) BtB and (b) BDI-pre Greater Than 6 Months at Baseline ........................................... 235

7. q-q Plot for Validation of Score Function Selection Fitting Model 1 Using IRHET for Subjects Whose (a) Current Episode Less Than 6 Months; (b) Current Episode Greater Than 6 Months at Baseline ........................................... 236
CHAPTER I
HETEROSCEDASTIC LINEAR MODELS

1.1 Introduction

In many scientific and practical applications, regression analysis provides an appropriate way to estimate or predict values of response variables from a set of predictors. In pursuit of fitting meaningful prediction model, it is customary to assume that the variances of the responses are constant from observation to observation. This underlying assumption is also known as the homoscedasticity of the error variances. Formally, in homoscedastic situations, the error variances of the distributions of the response variables for given values of the predictor variables are constant. In practice, heteroscedasticity may well be the typical case, particularly in pharmaceutical science and economics. For heteroscedastic situations, the errors are independent but do not have equal variances across the various levels of the responses. These levels are in the form of the design points.

In heteroscedastic case, we fit the desired models based on weighted observations and the weights also known as scaling constants are derived from the unequal variances. Thus, in heteroscedastic model fitting we seek to estimate parameters of both types: regression coefficients and scale constants. When the underlying distributions are known, as in the case of parametric models, there are several methods for estimating both parameters. One of these methods is the Weighted Least Squares estimation approach. It is well known that when the underlying distributions are non-normal, Least Squares estimators are very responsive to the presence of aberrant observations and high leverage points in the design space. The instability of the estimators becomes more pronounced when errors are heteroscedastic in nature. It is thus desirable to employ nonparametric methods that use ranks, as these methods are robust to outliers even under heteroscedastic error
settings. In this study, an estimation method for the parameters of interest that is resistant to outlying responses is introduced. It is worth noting that the problem of high leverage values when errors are heteroscedastic in nature is not addressed in this study. In what follows, the heteroscedastic model that is being considered in this study is presented.

### 1.2 Heteroscedastic Linear Regression

Consider the following model

\[ y_i = \beta_0 + x_i^T \beta_1 + \sigma_i e_i, \quad i = 1, \ldots, N, \tag{1.2.1} \]

where \(y_1, \ldots, y_N\) are responses, \(\beta_0\) and \(\beta_1\) are unknown regression parameters, \(x_i^T\) is the \(i\)th row of an \(N \times p\) matrix \(X\). \(X\) is centered with respect to its columns. The variables, \(e_1, \ldots, e_N\), are random errors that are assumed to be independently and identically distributed (iid) with a common cdf \(F\). Here, \(\sigma_1, \ldots, \sigma_N\), are scale constants that express heteroscedasticity through the relationship

\[ \sigma_i = \exp(x_i^T \theta), \quad i = 1, \ldots, N, \tag{1.2.2} \]

where \(\theta\) is a \(p \times 1\) vector of unknown scale parameters.

In the model specified by equations (1.2.1) and (1.2.2), estimates of both \(\beta = (\beta_0, \beta_1^T)^T\) and \(\theta\) are of interest. This is a member of the class of estimation problems of the location and scale type. In this class, it is not unusual to employ a method that iterates between a pseudo pure location estimation problem for \(\beta\) and a pseudo pure scale estimation problem for \(\theta\). Then for each problem, an estimation criterion is specified. In particular, a dispersion function in terms of \(\beta\) is defined for any given value of \(\theta\), and a dispersion function in terms of \(\theta\) is defined for any value of \(\beta\). It is worth noting in each of the problems, that if the given value of the other parameter is the true value, an
assumption rarely satisfied in practice, then we have a special case. For such special cases in each type problem, theoretical properties of the corresponding estimates from both classic and rank based methods have already been established. In this study, the theory of rank based estimation of the more general location and scale described above is developed. Before we embark on the discussion of this problem, a review of established results and recommended methods will be useful.

1.3 Previous Work

There has been substantial amount of research on obtaining estimates by the minimization of dispersion functions that yield a class of estimates broadly known as M-estimates which includes the well known Least Squares estimates (Huber, 1981)). With regard to heteroscedastic problems, Ruppert and Carroll (1982b), (1982a), (1988), proposed the M-estimation that provided the pioneering estimation methods for a member of the class of problem under consideration in this study. The authors considered the mean response model for the scale constants.

For the rank based estimation, the material in Jurečková (1969), (1971), Koul (1970), (1971), Kraft and van Eeden (1972), Adichie (1967), (1978) and Sen and Puri (1977) provided fundamental background for obtaining robust estimates. Jaeckel (1972) proposed the minimization of a suitably chosen dispersion function of residuals method and showed that the resultant estimates were asymptotically equivalent to those obtained using the method in Jurečková (1971). Hettmansperger and McKean (1977), (1978), (1998), extended on the linear rank statistic that can be derived from a suitably selected dispersion function. Sievers (1983) established the asymptotic linearity of a weighted rank method based on Gini's mean difference, which was the pioneer in rank methods that curtail high-leverage points in addition to outlying responses.

All of the aforementioned rank based methods focussed their attention on obtaining estimates of the regression parameters in homoscedastic linear models. Akritas (1996)
proposed a rank based method for estimating both regression and scale parameters in heteroscedastic models. In the author's paper, the scale constants were assumed to have a random coefficients model. Dixon and McKean (1996) proposed a rank analogue to the M-estimation method of Carroll and Ruppert (1982b) method. Dixon and McKean (1996) extended the linear rank test for heteroscedasticity proposed by Hájek and Šidák (1967) to the estimation problem and established asymptotic linearity based on the linear rank statistic. Carroll and Ruppert (1982b) and Dixon and McKean (1996) obtained a single scale parameter estimate.

There are several robust estimates of scale that have been proposed. For rank based scale parameter estimation, it is appealing to utilize the scale statistic proposed by Fligner and Killeen (1976), since its performance under non-normal distributions are well known. A good review of other methods can be found in Lax (1985). The results of the study of the class of scale estimators showed how well the trimmed standard deviation and Huber M-estimators performed in comparison to the highly efficient bi-weight functions. In this study, we only compare the proposed method to the first two methods.

1.4 Research Problem

In this study, as seen in (1.2.2), we investigate a heteroscedastic model in which the variance function is slightly different from the one used in Akritas (1996), in that, instead of a random coefficient model, our scale constants depend on the design of the model. Furthermore, due to the form of nonlinearity of the variance function, in terms of the coefficients, it is possible to linearize the relationship by applying a logarithmic transformation. The theoretical results for a unified rank based method for estimating both parameters in the model under study have yet to be established. Further, a simple technique utilizing suitably defined iterative method for the rank estimation of the parameters in the general linear problem would be a useful contribution to the analysis heteroscedastic models. Then the approach in this investigation is to obtain estimates of
both the location and the scale type parameters by minimizing a dispersion function of
residuals associated with some form of a linear model, as will be seen below.

1.4.1 Estimation Criterion

We now fix the idea behind the estimation criterion. In general linear models,
the dispersion function of residuals is a measure of the distance between the observed
responses and the fitted values. Hence, the goal of estimation is to obtain value of the
parameter of interest so that the distance between the observed responses and the fitted
values is minimized.

Following Jaeckel (1972), we establish the asymptotic linearity of the linear rank
statistic of residuals of linear model using the gradient of the dispersion function. Albeit,
this method was designed for estimating the regression coefficients in a linear model, the
linearized form of the variance function model being considered in this study satisfies
the conditions of the estimation problem in Jaeckel’s paper. Thus, the theory also holds
in estimating scale coefficients. However, to obtain an estimate of one parameter, an
estimate of the other parameter is required. Consequently, the method proposed in this
study will use an iterative scheme meeting some convergence criterion.

Let us rewrite model (1.2.1) as

\[ y_i = \beta_0 + x_i^T \beta_1 + e_i, \quad i = 1, \ldots, N, \]  

(1.4.1)

or

\[ y_i - \beta_0 - x_i^T \beta_1 = e_i, \quad i = 1, \ldots, N, \]  

(1.4.2)

Let us first consider the estimation of \( \beta \) using (1.4.1). If we assume that \( \theta \) is the
ture parameter value, so that \( \sigma_1, \ldots, \sigma_N \) are known,
then

\[ \frac{1}{\sigma_i} y_i = \frac{1}{\sigma_i} \beta_0 + \frac{1}{\sigma_i} x_i^T \beta_1 + e_i, \quad i = 1, \ldots, N, \tag{1.4.3} \]

equivalently,

\[ y_i^* = x_{i0}^* \beta_0 + x_i^{*T} \beta_1 + e_i, \quad i = 1, \ldots, N, \tag{1.4.4} \]

where \( y_i^* = \frac{1}{\sigma_i} y_i, \ x_{i0}^* = \frac{1}{\sigma_i}, \ x_i^{*T} = \frac{1}{\sigma_i} x_i^T, \) for \( i = 1, \ldots, N, \) is the transformed model that we fit to obtain an estimate of \( \beta. \) Let the residuals associated with this model be defined by

\[ r_i(b) = y_i^* - x_i^{*T} b, \quad i = 1, \ldots, N, \tag{1.4.5} \]

for \( b = (b_0, b_1), \ x_i^{*T} \) is the \( i \)th row of the matrix \( X_i^* = (x_{i0}^*, X^*), \ x_{i0}^* = (1/\sigma_1, \ldots, 1/\sigma_N). \) In matrix notation, we can write

\[ r(b) = y^* - x^*_0 b = (r_1(b), \ldots, r_N(b))^T. \tag{1.4.6} \]

Then, the following definition of the estimation criterion will be sufficient.

**Definition 1.** For any vector of residuals, \( r(b), \) a dispersion function is a function \( D(r(b)) \) given by

\[ D(r(b)) = \sum_{i=1}^{N} \varphi(r_i(b)), \tag{1.4.7} \]

where \( \varphi \) is continuous, nonnegative, and convex function of \( b. \)

In general, \( \varphi(r_i(b)) = \phi_N(i) \rho(r_i(b)), \) where \( \phi_N(1), \ldots, \phi_N(N) \) are nondecreasing scores. In rank estimation, we can select \( \phi_N(i) \) such that \( \sum_{i=1}^{N} \phi_N(i) = 0. \) Note that \( \rho(r_i(b)) \) is a convex function of \( b. \) Then it can be seen that rank estimation, the following relation-
ship holds

\[ D(ar(b) + c1) = |a|D(r(b)), \quad \text{(1.4.8)} \]

for any \( a, c \in \mathcal{R} \) and \( 1 \) is an \( N \times 1 \) column vector of ones.

Next, we consider the estimation of the \( \theta \), using model (1.4.2). Note that the model is nonlinear in \( \theta \), but taking the logarithm of the absolute values both sides of the equation linearizes it to

\[
\log |y_i - \beta_0 - x_i^T \beta_1| = x_i^T \theta + \log |e_i|, \quad i = 1, \ldots, N, \quad \text{(1.4.9)}
\]

which can also be written in the alternative matrix form,

\[
Z^* = X\theta + e^*, \quad \text{(1.4.10)}
\]

where \( Z^* = (\log |y_1 - \beta_0 - x_1^T \beta_1|, \ldots, \log |y_N - \beta_0 - x_N^T \beta_1|)^T \) and \( e^* = (\log |e_1|, \ldots, \log |e_N|)^T \).

It is seen that model (1.4.10) is similar to model (1.4.4). It is worth noting that the error terms \( \log |e_1|, \ldots, \log |e_N| \) follow a common distribution which is centered at some nonzero constant, \( \theta_0 \). Clearly, \( \theta \) excludes intercept-type parameters. For the moment, let us define the residuals for the current model as

\[
v_i(t) = z_i^* - x_i^T t, \quad i = 1, \ldots, N, \quad \text{(1.4.11)}
\]

for any \( t \in \mathcal{R}^p \), and \( x_i^T \) is the \( i \)th row of the design matrix \( X \). It may be to convenient to define the matrix form

\[
v(t) = z^* - Xt = (v_1(t), \ldots, v_N(t))^T. \quad \text{(1.4.12)}
\]

Since the residuals in (1.4.11) correspond to a linear model, given in (1.4.9), Defi-
nition 1 of a dispersion function applies with \( b \) replaced by \( t \) and, correspondingly, \( r(b) \) replaced by \( v(t) \).

Let us now turn to the basic idea behind the dispersion functions given above and how particular choices of the function affect the robustness of the desired estimates. We now describe two dispersion function based estimates.

### 1.4.2 Weighted Least Squares Estimation

The most commonly used estimation method is the Least Squares (LS) approach. When scale constants, \( \sigma_1, \ldots, \sigma_N \), that express heteroscedasticity are present in the model, the suitably transformed model in (1.4.4) can be fit by minimizing the Euclidian distance of the vector of residuals from the origin of \( \mathcal{R}^N \). Thus, the estimate of the regression parameter that is obtained using (LS) dispersion function is defined as

\[
\hat{\beta}_{WLS} = b = \text{Argmin} \| y^* - x_i^* b \|_2 = \text{Argmin} (\sum_{i=1}^{N} r_i(b)^2)^{1/2}. \quad (1.4.13)
\]

Under the asymptotic normality conditions, that will be given below, it can be shown that

\[
\hat{\beta}_{WLS} \text{ is asymptotically } N(\beta, \sigma^2(X^*_T X^*_1)^{-1}).
\]

Now consider the scale estimation based on LS dispersion function. For the suitably transformed model given in (1.4.9), the estimate of the 'shift like' parameter can be obtained by minimizing the vector of residuals. Thus, the location free estimate that is obtained using the LS dispersion function is given by

\[
\hat{\theta}_{LS} = t = \text{Argmin} \| z^* - x_t \|_2 = \text{Argmin} \left( \sum_{i=1}^{N} v_i(t)^2 \right)^{1/2}. \quad (1.4.14)
\]
Similarly, under some asymptotic normality conditions, it can be shown that

\[ \hat{\theta}_{LS} \text{ is asymptotically } N(\theta, \sigma^2(X^TX)^{-1}). \]

It is well known that the LS dispersion function is very responsive to outlying observations. Observe in definitions given in (1.4.13) and (1.4.14), respectively, that outlying values of \( y^*_i \) and \( z^*_i, \ i = 1, \ldots, N, \) can inflate the estimate in each case. This problem persists to even greater magnitude when heteroscedasticity is present in the model. To overcome this drawback of LS estimation methods, we use rank based methods.

1.4.3 Rank Estimates: Other Parameters are Specified

In this study, we employ rank dispersion functions to curb the effect of the outlying responses on the robustness of the desired estimates of \( \beta \) and \( \theta \). Rank dispersion functions can be viewed as sums of weighted residuals unlike their LS counterpart which are unweighted. The weights in rank dispersion function are suitably defined scores based on the ranks of the residuals such that large residuals have smaller score values. Thus, for the model in (1.4.4) can be fit by obtaining estimates of \( \beta \) such that

\[ \hat{\beta}_{\phi^*_i} = b = \text{Argmin} \|y^* - x^*_i b\|_{\phi_i^*} = \text{Argmin} \sum_{i=1}^{N} \phi_i^+(\frac{R(\|r_i\|)}{N + 1}) |r_i(b)|, \tag{1.4.15} \]

where \( \phi_i^+(u) \in (0, 1) \) is a nondecreasing score generating function and \( \phi_i^+(u) \) is positive valued.

Under the asymptotic normality conditions prescribed below, it can be shown that

\[ \hat{\beta}_{\phi_i^*} \text{ is asymptotically } N(\beta, \gamma_1^{-2}(X_1^TX_1^*)^{-1}). \]

Here, \( \gamma_1 \) is a dispersion parameter that depends on the score function through

\[ \gamma_1 = \int \phi_i^+(u) \phi_i^+(u, f) du \text{ and } \phi_i^+(u, f) = \frac{f'(F^{-1}(\frac{u+1}{2}))}{f(F^{-1}(\frac{u+1}{2}))^2}, \]

to be discussed later.
Let us consider the scale estimation problem using ranks. Observe that the suitably transformed model (1.4.4) can be fitted by obtaining location free estimates of \( \theta \) such that

\[
\hat{\theta}_{\phi^*} = t = \text{Argmin}_{z^*} \| z^* - xt \| = \text{Argmin} \sum_{i=1}^{N} \phi^*_2 \left( \frac{R(u_i(t))}{N + 1} \right) u_i(t),
\]

(1.4.16)

where \( \phi^*_2(u) \in (0, 1) \) is a nondecreasing score generating function and \( \sum_{i=1}^{N} \phi^*_2 \left( \frac{i}{N+1} \right) = 0 \).

Under some regularity conditions given below, and the true value of \( \beta \) is specified, it can be shown that

\[ \hat{\theta}_{\phi^*} \text{ is asymptotically } N(\theta, \gamma_2^{-2}(X^TX)^{-1}). \]

Here, \( \gamma_2 \) is a dispersion parameter that depends on the score function through

\[ \gamma_2 = \int \phi^*_2(u)\phi^*_2(u, h_{e^*}(u))du \text{ and } \phi^*_2(u, h_{e^*}(u)) = \frac{h_{e^*}(H_{e^*}(u))}{h_{e^*}(H_{e^*}(u))} \text{ and } h_{e^*} = dH_{e^*}(u) \text{ is density of the error terms in model (1.4.9)}. \]

Thus, the rationale for choosing to use rank based methods is that in these methods, the residuals associated with outlying responses are down weighted giving robust estimates. In LS based methods, on the other hand, the full effects of the outlying responses on their respective residual reverberate through the estimation process, yielding estimates with inflated standard errors. It is worth noting that, in this study, we restrict our attention to outlying responses and do not consider outlying design points.

1.4.4 Rank Estimates: Other Parameters are Unknown

While the estimation problems of regression and scale parameters have both been handled by a dispersion function for fitting a linear model, their difference is what drives the development of the theory in this study. In each of the dispersion functions, an estimate of the parameter of interest is obtained while assuming the other parameter is a fixed value. If the fixed value is the true value of the parameter, then we have the optimal
case where only one parameter is sought.

In practice, the true values are rarely known and have to be replaced by their corresponding estimates. The specific forms of the dispersion function associated with parameters \( \beta \) and \( \theta \), that permit estimated values of other underlying parameters, are discussed in chapters three and four, respectively. In these chapters, the asymptotic behavior of each estimator is characterized through asymptotic linearity of the linear rank statistics that correspond to the dispersion functions. Further, it is shown that the minimization of the dispersion function for each problem with the other parameter replaced by its estimates yields a robust and efficient estimator of the parameter of interest. It is worth noting that the statistics employed in this study, are reparametrized forms of the usual linear statistics of Kraft and van Eeden (1972) and Jurečková (1971).

In the section that follows, we furnish more details on the theoretical motivation of the proposed IRHET method.
2.1 Introduction

In this chapter, described in general terms are two types of estimation problems: (1) regression coefficient and (2) scale constant. For each type of estimation problem, existing fundamental asymptotic results and regularity conditions for the special case in which the value of the other parameter is specified, are presented. In addition, each of the problems is expounded to the cases in which the value of the other parameter is arbitrary but fixed to define the specific objective of the theoretical results being sought. Thus, in this chapter, previously established results that are relevant to each case are presented as a precursor to the detailed description of the estimation problems investigated in subsequent chapters.

The simple linear model with heteroscedasticity is specified in the next section. The regression estimation problem is presented in Section (2.3). The scale parameter estimation problem is presented in Section (2.4). Each of the sections includes the definition of residuals of a transformed form of the model, and estimation criterion. This is followed by linear rank statistic that satisfies the estimation criterion, regularity conditions of special case of the simple linear model and finally the well known asymptotic linearity results. In addition, the condition that is directly affected by the transformation employed in each estimation problem, is shown to be valid.
2.2 Heteroscedastic Linear Model

We furnish a simple linear model in which the heteroscedasticity is present in what follows. Suppose

\[ y_i = \beta_0 + \beta_1 x_i + e^{\theta x_i} e_i, \quad i = 1, \ldots, N, \quad (2.2.1) \]

where the \( e_i \) are i.i.d. with a common density \( f, f \) is symmetrical, \( \text{Var}(e_i) = \sigma^2_e \). Clearly, in this model, the variance of the responses depend on design through \( \sigma^2_i = e^{2\theta x_i} \) for \( i = 1, \ldots, N \), and this qualifies (2.2.1) a heteroscedastic linear model. Next, we describe theoretical results for estimates of \( \beta \) obtained from the model under certain conditions that are given below.

2.3 Regression Coefficient Estimation

This section begins with noting in (2.2.1) that if the equation is divided by \( e^{\theta x_i} \), we obtain

\[ y_i(e^{-\theta x_i}) = \beta_0(e^{-\theta x_i}) + \beta_1 x_i(e^{-\theta x_i}) + e_i, \quad i = 1, \ldots, N, \quad (2.3.1) \]

equivalently,

\[ y_i^* = \beta_0 x_i^* + \beta_1 x_i^* + e_i, \quad i = 1, \ldots, N, \quad (2.3.2) \]

where \( y_i^* = (e^{-\theta x_i})y_i, x_i^* = e^{-\theta x_i}, x_i^* = (e^{-\theta x_i})x_i \).

2.3.1 Residuals

To set up the estimation problem, we introduce arbitrary but fixed shifts from the parameters, \( \beta = (\beta_0, \beta_1)^T \) and \( \theta \). Suppose that \( \frac{1}{\sqrt{N}}b, b = (b_0, b_1) \) be an arbitrary but fixed
shift from the regression coefficient, $\beta$ and $\frac{1}{\sqrt{N}}, t \in \mathcal{R}$ is an arbitrary but fixed shift from the scale parameter, $\theta$. Rearranging terms in (2.3.2) and introducing the arbitrary shifts leads to

$$y^*_i - \beta_0 x^*_oi - \beta_1 x^*_1i - \frac{1}{\sqrt{N}} b_0 x^*_oi - \frac{1}{\sqrt{N}} b_1 x^*_1i = e_i - \frac{1}{\sqrt{N}} b_0 x^*_oi - \frac{1}{\sqrt{N}} b_1 x^*_1i, \quad i = 1, \ldots, N,$$

(2.3.3)

which defines residuals for the regression coefficient estimation problem. Observe that when $\beta = 0$, the residuals are given by

$$z_i(b, t) = y_i^* - \frac{1}{\sqrt{N}} b_0 x^*_oi - \frac{1}{\sqrt{N}} b_1 x^*_1i \quad i = 1, \ldots, N,$$

(2.3.4)

$$= y_i - b_0 x_i - \frac{1}{e^{(\theta + \frac{1}{\sqrt{N}}) x_i}}$$

(2.3.5)

$$= y_i - b_0 x_i - \frac{1}{\sigma_i(t)}.$$  

(2.3.6)

Given this definition of residuals for the regression coefficient problem, the estimation criterion can then be considered.

2.3.2 Estimation Criterion

Consider the residual defined in (2.3.4). The signed-rank estimate of $\beta, \hat{\beta}$, is the value $b$ that minimizes

$$D_{1N}(b, t) = \sum_{i=1}^{N} \phi^+_i \left[ \frac{R(|z_i(b, t)|)}{N + 1} \right] |z_i(b, t)|,$$

(2.3.7)

for any fixed $t$. 

Here, the function $\phi_i^+(u)$ is score generating function satisfying (S1). An example of this score is the signed-rank Wilcoxon scores, $\phi_i^+(u) = u$. $R(|z_i(b, t)|)$ is the rank of $|z_i(b, t)|$ among $|z_k(b, t)|$ for $k = 1, \ldots, N$.

For the objective function given in (2.3.7), the asymptotic behavior of the estimator, $b$, can be characterized by a suitably defined linear rank statistic. It is well known that the resulting estimates from both these processes are asymptotically equivalent, due to a result by Jaeckel (1972). The linear rank statistic that corresponds to (2.3.7) is given by

$$s_{1N}(b, t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i^+ \left[ \frac{R(|z_i(b, t)|)}{N + 1} \right] sgn(z_i(b, t)) \begin{bmatrix} x_{0i}^* \\ x_{1i}^* \end{bmatrix} e^{-\frac{1}{N} t x_i}, \quad (2.3.8)$$

where $S_{1N}(b, t) = (S_{1N1}(b, t), S_{1N2}(b, t))^T$, $x_{0i}^* = e^{-6x_i}$, $x_{1i}^* = (e^{-9x_i})x_i$. Observe that in the statistic $S_{1N}(b, t)$, the variables $z_i(b, t)$ are the source of randomness since for any fixed $t$. Further, at values $(0, 0)$, the statistic in (2.3.8) yields

$$s_{1N}(0, 0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i^+ \left[ \frac{R(|z_i(0, 0)|)}{N + 1} \right] sgn(z_i(0, 0)) \begin{bmatrix} x_{0i}^* \\ x_{1i}^* \end{bmatrix}, \quad (2.3.9)$$

Thus, the shifts from $(b, t)$ to $(0, 0)$ also affect the regression constants

$$\begin{bmatrix} x_{0i}^* \\ x_{1i}^* \end{bmatrix} e^{-\frac{1}{N} t x_i}, \quad i = 1, \ldots, N,$$

which are supposed to be fixed constants. To circumvent the problem of shifting regression constants, it may be convenient to introduce an arbitrary but fixed constant $s$. Then, constants can alternatively be expressed as

$$\begin{bmatrix} x_{0i}^{**} \\ x_{1i}^{**} \end{bmatrix} = \frac{1}{\sigma_i(s)} \begin{bmatrix} 1 \\ x_i \end{bmatrix}, \quad i = 1, \ldots, N, \quad (2.3.10)$$
where

$$\sigma_i(s) = e^{(e + \frac{1}{N})x_i}, \quad i = 1, \ldots, N.$$  

Consequently, the statistic in (2.3.8) has the alternative form

$$S_{1N}(b, t, s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i^+ \left[ \frac{R|z_i(b, t)|}{N+1} \right] \text{sgn}(z_i(b, t)) \begin{bmatrix} x_{i0}^* \\ x_{ii}^* \end{bmatrix}.$$  \hspace{1cm} (2.3.11)

Thus, it is seen that introducing the constant $s$ allows for arbitrary but fixed regression constants, while the stochastic properties of $S_{LN}(b, t, s)$ and underlying shifts of the process are sorely expressed by the variables $z_i(b, t)$, as expected in the model being considered.

To characterize the asymptotic behavior of our estimate $b$, we can obtain a first order expansion of the process $S_{1N}(b, t, s)$ into the process $S_{1N}(0, 0, s)$, for which the asymptotic normality holds and a term that is linear with respect to $b$. To motivate the setting of the problem in this section, we first consider the conditions for asymptotic normality under the special case, $t = s = 0$.

2.3.3 Model Assumptions

In the sequel, the other regularity and design conditions will be specified. The next two conditions apply to the design in the regression estimation problem.

(D1*)

$$\lim_{N \to \infty} \max_{1 \leq i \leq N} x_{i1}^*(X_1^*X_1^*)^{-1}x_{i1}^* = 0.$$
\[(D2^*)\]

\[
\lim_{N \to \infty} \frac{1}{N} (X_1^T X_1^*) = \Sigma_1^*,
\]

where \(\Sigma_1^*\) is a \((p + 1) \times (p + 1)\) positive definite matrix.

The next two conditions apply to the distribution of errors,

(\textbf{F1}) \quad \text{(i) } f(y) \text{ is absolutely continuous.}

\text{(ii) } \int_{-\infty}^{\infty} (\frac{f''(y)}{f(y)})^2 f(y) dy < \infty.

Thus, if (i) and (ii) hold, then \(f(y)\) has a finite Fisher information \(I(f)\).

(\textbf{F2}) \quad f(y) = f(-y), \text{ that is, } f(\cdot) \text{ is symmetric about 0.}

Next, the restriction on the score function is described. Since, in the location problem, the errors, \(e_1, \ldots, e_N\), are symmetrically distributed around 0, it follows that the scores generating function satisfy the following:

(\textbf{S1}) \quad \text{Let } \phi_1^+(u) \text{ be a nondecreasing, positive-valued, and square integrable function defined on the interval } (0, 1). \text{ Further, } \sum_{i=1}^{N} \phi_1^+ \left( \frac{u}{N+1} \right) > 0. \text{ Furthermore, due to the square-integrability of } \phi_1^+(u), \text{ the scores can be standardized so that } \int_0^{1} (\phi_1^+(u))^2 \, du = 1.

Before discussing the asymptotic linearity result, the key condition that is directly affected by restricting the \(t = s = 0\) must be verified. That is, we need to show that for the regression constants in the statistic defined in \((2.3.2)\), under the restriction, the matrix \(\frac{1}{N} X_1^T X_1^*\) converges to a \(2 \times 2\) positive definite matrix, \(A^*\).

Remark 2.3.1. For \((D2^*)\) to be satisfied, the scaling constants \(\sigma_1, \ldots, \sigma_N\), must be bounded away from zero. Under this condition, \(\frac{1}{N} (X_1^T X_1^*)\) converges to positive definite matrix.

Proof. For finite scaling constants that are not close to zero the elements of \((X_1^T X_1^*)\) are fixed.
Consequently,

\[ \lim_{N \to \infty} \frac{1}{N}(X_1^TX_1) = \lim_{N \to \infty} \frac{1}{N} \begin{bmatrix} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \sum_{i=1}^{N} \frac{\varepsilon_i}{\sigma_i^2} \\ \sum_{i=1}^{N} \frac{\varepsilon_i}{\sigma_i^2} \sum_{i=1}^{N} \frac{\varepsilon_i^2}{\sigma_i^2} \end{bmatrix} = A^*, \]

which is a $2 \times 2$ positive definite matrix. For the general $(p + 1)$ linear model, the corresponding matrix, $\Sigma^*_1$, is of size $(p + 1) \times (p + 1)$. The proof is complete. \hfill \square

In what follows, the asymptotic linearity for the process $S_{1N}(b, 0, 0)$ is given in the spirit of Kraft and van Eeden (1972).

2.3.4 Asymptotic Linearity of $S_{1N}(b, 0, 0)$

For the general linear problem, observe in the function by which the scale constants, $\sigma_i$, are defined, that the scale parameter is a vector of size $p$, in correspondence with the number of independent variables. In what follows, the theorem is stated as a lemma. First, the conditions that were imposed in the aforementioned theorem are specified.

(KVF1) (i) $f(y) = dF(y)/dy$ exists and is absolutely continuous in the interval $(-\infty, \infty)$.  
(ii) $\int_0^1 [\phi^+_t(u, f)]^2 du = \int_0^1 [-f'(f^{(p-1)}(y)))^2 du < \infty$. That is, a finite Fisher information exists.

(KVF2) $f(y) = f(-y)$ for all $y$, that is, $f(\cdot)$ is symmetric about 0.

(KVD1) $\frac{\max_{1 \leq i \leq N} x_{ij}^2}{\sum x_{ij}^2} \to 0$ for each $j = 1, \ldots, p + 1$

(KVD2) $\frac{1}{N}X_1^TX_1 \to \Sigma^*_1$, where $\Sigma^*_1$ is a positive definite matrix.

(KVD3) For each pair $(j_1, j_2)(j_1 \neq j_2; j_1, j_2 = 1, \ldots, p + 1)$ there exist a number $\gamma_{j_1j_2} \neq 0$ such that, for $N > N_0$. 

\[ \frac{1}{N}X_1^TX_1 \to \Sigma^*_1 \]
(1.) \( x_{i1}^* (x_{i1}^* + \gamma_{i1} x_{i2}^* ) \geq 0 \) for all \( i \),

(2.) \( |x_{i1}^*| \) and \( |x_{i1}^* + \gamma_{i1} x_{i2}^*| \) are similarly ordered, where \( (x_0^*, x_1^*, \ldots, x_p^*) \) are column vectors of \( X_i^* \).

It is worth noting that the condition (KVD3) is a generalization of the concordance conditions imposed in Theorem 3.3 of van Eeden (1972) upon which the current linearity result is based. Furthermore, the concordance conditions have been shown by Tardif (1985) to be unnecessary for the asymptotical linearity result to hold.

Theorem 7.2 of Kraft and van Eeden (1972) is stated as a lemma in the sequel.

**Lemma 2.3.1.** Assume that (KVF1), (KVF2), (KVL1), (KVD1), (KVD2) and (KVD3) are satisfied. In addition, suppose that the errors \( e_1, \ldots, e_N \) are jointly distributed as \( p_N \). If \( p > 2 \), let

\[
S_{1N_j}(b, 0, 0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_{ij}^* \Phi_i^+(\frac{R(|z_i(b, 0)|)}{N + 1}) \text{sgn}(z_i(b, 0)).
\] (2.3.12)

Let \( \Sigma_i^{*}(j) \) denote \( j \) the column of the matrix \( \Sigma_i^* \).

Then, for each \( j = 1, \ldots, p \),

\[
\lim_{N \to \infty} P\left\{ \sup_{\|b\| \leq B_0} \left| S_{1N_j}(b, 0, 0) - S_{1N_j}(0, 0, 0) + \gamma_j b^T \Sigma_i^{*(j)} \right| > \epsilon \right\} = 0,
\] (2.3.13)

for each \( B_0 > 0 \) and \( \epsilon > 0 \).

This demonstrates that even for the special case \( p = 1 \), the result holds and direct substitution into the steps used by Kraft and van Eeden (1972), leads to the asymptotic normality of the estimate \( \hat{\beta} \) in a straightforward manner.

**Lemma 2.3.2.** Assume that \( (D1^*), (D2^*), (F1), (F2), \) and \( (S1) \) hold. In addition, suppose that
condition (B1) on the compactness of \(b\) to be given later is satisfied. Then,

\[
\hat{\beta} \xrightarrow{d} N_{p+1}(\beta, \gamma^{-2}(X_1^TX_1)^{-1}).
\]

(2.3.14)

In this study, it is of interest to obtain an analogous result under the general case, \(t, s \neq 0\). Observe from the regression constants in the alternative form of the linear rank statistic in (2.3.11), that it has yet to be demonstrated that the matrix \(\frac{1}{N}X_1^{**T}X_1^{**}\) converges to \(A^{**} = A^*\). Note that \(X_1^{**T}\) is the matrix whose elements are defined in (2.3.10).

**Remark 2.3.2.** Let us suppose that

\[
\frac{1}{N}(X_1^{**T}X_1^{**}) \rightarrow A^{**}, \text{ as } N \rightarrow \infty,
\]

where \(A^{**}\) is a \(2 \times 2\) positive definite matrix.

Then

\[
A^{**} = A^*.
\]

**Proof.** First, consider \(A_{11}^{**} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{N} x_{i1}^{**2}\). Note that

\[
\frac{1}{N} \sum_{i=1}^{N} x_{i1}^{**2} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} e^{-2\sqrt{N}z_i}
\]

(2.3.15)

\[
= \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} + \frac{1}{N} \sum_{i=1}^{N} (e^{-2\sqrt{N}z_i} - 1) \frac{1}{\sigma_i^2}
\]

(2.3.16)

In the second term, the following result will be useful. Let \(h(x) = e^x\), so that \(h(0) = 1\).

Then,

\[
h'(x) = e^x
\]

\[h(x) = h(0) + h'(\xi) \cdot x.
\]
Then, for $|\xi| \leq |x|$, 

$$e^x - 1 = e^\xi \cdot x$$

so that

$$\max_{1 \leq i \leq N} \frac{|x_i|}{\sqrt{N}} \to 0 \text{ as } N \to \infty.$$ 

Then

$$|e^{-2\sqrt{N}x_i} - 1| \leq \epsilon.$$ 

So the second term in (2.3.16) is

$$\leq \frac{1}{\sqrt{N}} \sum_{i=1}^{N} |e^{-2\sqrt{N}x_i} - 1| \cdot \frac{1}{\sigma_i^2}$$

$$\leq \epsilon \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^2}.$$ 

Since by Assumption (D2*) along with Remark 2.3.2, $\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^2}$ converges to a constant, the second term in (2.3.16) goes to zero, so that the $A_{ij}^{**}$ converges to $A_{ij}^*$ as $N$ goes to $\infty$.

Next, consider $A_{i2}^{**}$ and note that

$$A_{i2}^{**} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} x_i x_{i2}^{**}$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} e^{-2\sqrt{N}x_i} \cdot \frac{x_i}{\sigma_i^2}$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} + \frac{1}{N} \sum_{i=1}^{N} (e^{-2\sqrt{N}x_i} - 1) \cdot \frac{x_i}{\sigma_i^2}.$$ 

Thus, using an argument similar to that used in determining the limit in $A_{i1}^*$, observe that
the first term converges to a constant, by Assumption (D2\*) and Remark 2.3.2, whereas the second term goes to zero. Hence, $A_{12}^{**}$ converges to $A_{12}^{*}$ as $N$ goes to $\infty$. Similarly, as $N$ goes to $\infty$, $A_{21}^{**}$ converges to $A_{21}^{*}$. Further, by the symmetry of $X_1^{**}X_2^{**}$, $A_{21}^{*} = A_{12}^{*}$.

Next, consider $A_{22}^{**}$. Observe that

\[
A_{22}^{**} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} x_{i2}^{*2} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} e^{-2\sqrt{N}x_i} \cdot \frac{x_i^2}{\sigma_i^2} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} + \frac{1}{N} \sum_{i=1}^{N} (e^{-2\sqrt{N}x_i} - 1) \cdot \frac{x_i^2}{\sigma_i^2}.
\]

Thus, using an argument similar to that used in determining the limit in $A_{12}^{*}$, it can be seen that the first term converges to a constant by assumption (D2\*) and Remark 2.3.2, whereas the second term goes to zero. Hence, $A_{22}^{**}$ converges to $A_{22}^{*}$ as $N$ goes to $\infty$. Thus, this completes the proof for the remark.

The importance of Remark 2.3.2 is that this result demonstrates that the effect of replacing the scale parameter by its estimate $\theta + \frac{\theta^2}{\sqrt{N}}$ on the standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$ is asymptotically equivalent to that obtained when the scale parameter is $\theta$. In the general problem, with $p > 2$, it is the purpose of chapter three to show that the asymptotic linearity of $S_{1,2}(b, t, s)$ holds, and consequently, so does the asymptotic normality of the corresponding estimate, $\hat{\beta}$.

### 2.4 Scale Parameter Estimation

Let us consider the two sample model of (2.2.1) which can be written as the following error model

\[
y_i - \beta_0 - \beta_1 x_i = e^{\theta x_i} e_i, \quad i = 1, \ldots, N.
\]
Then an application of the logs of absolute value transformation gives

$$
\log|y_i - \beta_0 - \beta_1 x_i| = \theta x_i + \log|e_i|,
$$

(2.4.2)
equivalently,

$$
z^*_i = \theta x_i + e_i^*,
$$

(2.4.3)
e_i^*, \ldots, e_N^* are iid random variables with a common cdf $H_{e^*}(u)$ and density $h_{e^*}(u)$. Assume that $h_e$ has is centered around a nonzero constant, $\theta_0$.

2.4.1 Residuals

Similar to the regression problem, we consider arbitrary but fixed shifts from the parameters. Suppose that $\frac{1}{\sqrt{N}}b_0$ and $\frac{1}{\sqrt{N}}b_1$, respectively, are arbitrary fixed shifts from $\beta_0$ and $\beta_1$. Also let, $\frac{1}{\sqrt{N}}t$ be an arbitrary but fixed shift from $\theta$. Note $b_0, b_1, t \in \mathcal{R}$. Then, letting the vector of regression coefficient be $\beta + \frac{1}{\sqrt{N}}b$ so that

$$
z^*_i(b) = \log|y_i - \beta_0 - \beta_1 x_i - \left(\frac{b_0}{\sqrt{N}} + \frac{b_1}{\sqrt{N}} x_i\right)|, \quad i = 1, \ldots, N,
$$

(2.4.4)
the scale parameter be $\theta + \frac{1}{\sqrt{N}}t$, and rearranging terms in (2.4.3), we have residuals defined as

$$
v_i(b, t) = z^*_i(b) - x_i \frac{t}{\sqrt{N}}, \quad i = 1, \ldots, N.
$$

(2.4.5)
Having defined the residuals for the scale model problem, the estimation criterion for the scale parameter can now be described as we do in the sequel.
2.4.2 Estimation Criterion

The estimate of \( \theta \) that is being sought is the value, \( t \), that minimizes the objective function

\[
D_{2N}(b, t) = \sum_{i=1}^{N} \phi^*_2 \left( \frac{R[v_i(b, t)]}{N + 1} \right) v_i(b, t), \tag{2.4.6}
\]

where \( \phi^*_2(\cdot) \) satisfies assumption (S2), and an example of these scores are generated as, \( \phi^*_2(u) = u \cdot \log(\frac{u+1}{u-1}) \). Here, \( R[v_i(b, t)] \) is the rank of \( v_i(b, t) \) amongst \( v_k(b, t) \), \( k = 1, \ldots, N \).

For this objective function given in (2.3.7), which mirrors the dispersion function proposed by Jaeckel (1972), the asymptotic behavior of the estimator \( t \) can equivalently be characterized by a suitably defined linear rank statistic. This result is due to Theorem 3 of Jaeckel (1972), which demonstrated that estimates obtained by minimizing the dispersion function are asymptotically equivalent to those obtained by the solving for the zeros of the gradient of the dispersion function. For the objective function in (2.4.6), the suitably defined linear rank statistic is defined as

\[
S_{2N}(b, t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i \phi^*_2 \left( \frac{R[v_i(b, t)]}{N + 1} \right). \tag{2.4.7}
\]

The asymptotic behavior of the estimate, \( t \), can be obtained by deriving a first order expansion of the process \( S_{2N}(b, t) \) into a process \( S_{2N}(\cdot, \cdot) \) for which the asymptotic normality holds and a term that is linear with respect to \( t \). We set up the problem in this section by first specifying the conditions for asymptotic normality under the special case, \( b = 0 \).

2.4.3 Additional Assumptions

In the sequel, the other regularity and design conditions will be specified. The next two conditions apply to the design used when estimating the scale parameter.
(D1)

\[
\lim_{N \to \infty} \max_{1 \leq i \leq N} x_i^T (X^T X)^{-1} x_i = 0.
\]

(D2)

\[
\lim_{N \to \infty} \frac{1}{N} (X^T X) = \Sigma,
\]

where \( \Sigma \) is a \( p \times p \) positive definite matrix.

The next condition applies to the distribution of errors,

(F1) The same as previously defined in the Section 2.3.3. It should be noted that when the log of absolute value transformation is applied on the errors \( e_1, \ldots, e_N \), this condition carries over easily. Hence, the condition is not restated in terms of the cumulative distribution function (cdf) and probability distribution function (pdf) of \( \log |e_1|, \ldots, \log |e_N| \).

Next, the restriction on the score function is described. In this instance, this condition is applicable because the scale problem has been converted to a location type problem by the log of absolute residual transformation.

(S2) Let \( \phi_2(u) \) be a nondecreasing, square integrable function defined on the interval \((0, 1)\). Further, \( \int_0^1 \phi_2(u) du = 0 \) hence \( \sum_{i=1}^N \phi_2^{[\frac{1}{N+1}]} = 0 \). Furthermore, due to the square-integrability of \( \phi_2(u) \), the scores can be standardized so that \( \int_0^1 (\phi_2^*(u))^2 du = 1 \). In addition, it assumed that \( \phi_2^*(u) \) is bounded with bounded derivative.

Before we can state the asymptotic linearity and normality of the estimator of \( \theta \), we need to verify that the condition that is directly affected by the transformation employed in this section holds. In this regard, we seek to show that the error terms for the linearized model, the corresponding scores possess a property that is true for location score under
such errors. Then, it needs to be demonstrated that the working scores sum to zero so that condition (S2) is satisfied.

Recall that the errors $e_1, \ldots, e_N$ have the common cdf $F(u)$. Observe that the cdf of $\log|e_1|, \ldots, \log|e_N|$ is given by

$$J(t) = F(e^t) - F(-e^t) = 2F(e^t) - 1. \quad (2.4.8)$$

Thus, the density function is given by

$$j(t) = 2f(e^t)e^t. \quad (2.4.9)$$

Hence,

$$j'(t) = 2 \left[ f'(e^t)e^{2t} + f(e^t)e^t \right].$$

Then, the location score function corresponding to the density function, $j$, is given by

$$\frac{j'(t)}{j(t)} = \frac{f'(e^t)}{f(e^t)}e^t + 1,$$

where if we let $u = J(t) = 2F(e^t) - 1$ so that $J^{-1}(u) = t = \log[F^{-1}(\frac{1+u}{2})]$, we have

$$-\frac{j'(J^{-1}(u))}{j(J^{-1}(u))} = -\frac{f'(F^{-1}(\frac{1+u}{2}))}{f(F^{-1}(\frac{1+u}{2}))}F^{-1}\left( \frac{1 + u}{2} \right) - 1,$$

where

$$\phi_2(u) = -\frac{f''(F^{-1}(u))}{f(F^{-1}(u))}F^{-1}(u) - 1.$$
Observe that the expression on the right hand side of the first equation defines scale scores evaluated at \( \frac{1+u}{2} \). In particular, \( \phi_2(u) \) is nonincreasing on the interval \((-1, 0)\) and non-decreasing on the interval \((0, 1)\).

Let

\[
\phi_2^*(u) = \phi_2 \left( \frac{1 + u}{2} \right).
\]

Clearly, \( \phi_2^*(u) \) is defined on the interval \((0, 1)\). The remark that follows demonstrates that the transformed scale scores possess a property that is true for location scores used to fit a model in which the error terms are not symmetrical about zero. Recall that the error terms corresponds to the variables \( \log|e_1|, \ldots, \log|e_N| \).

**Remark 2.4.1.** The scores \( \phi_2^* \) satisfy \( \int_0^1 \phi_2^*(u)du = 0 \).

**Proof.** We need to first show that \( \int_0^1 \phi_2^*(u)du = \int_0^1 \phi_2 \left( \frac{1+u}{2} \right)du = 0 \). Then, recalling that

\[
\int_0^1 \phi_2 \left( \frac{1 + u}{2} \right)du = \int_0^1 \left[ - \frac{f'(F^{-1}(\frac{1+u}{2}))}{f(F^{-1}(\frac{1+u}{2}))} F^{-1} \left( \frac{1 + u}{2} \right) - 1 \right]du \\
= 2 \int_{\frac{1}{2}}^1 - \frac{f'(F^{-1}(v))}{f(F^{-1}(v))} F^{-1}(v)dv - 1,
\]

due to the change of variable \( v = \frac{1+u}{2} \). If we let \( x = F^{-1}(v) \) so that \( F(x) = v \) and \( dv = f(x)dx \) the right hand side of the last expression becomes

\[
-2 \int_0^\infty f'(x)xdx - 1 = -2 \int_0^\infty f(x)dx - 1 \\
= 2((1/2) - 1) = 0,
\]

where the first equality is due to an application of integration by parts with \( u = x, du = dx, \) and \( dv = f'(x)dx, v = f(x) \). This completes the proof. \( \square \)

Thus, the asymptotic results for the location problem subject conditions that are similar to those in this section apply. These results are furnished in the sequel.
2.4.4 Asymptotic Linearity of $S_{2N}(0, t)$

We now briefly discuss an asymptotic linearity result that will be useful in establishing the limiting behavior of the proposed estimator of $\theta$ in this study.

From (2.4.4) recall that the response variables $z_i^*(0), i = 1, \ldots, N,$ corresponds to the case when the regression coefficient is $\beta$. Then it follows that for the statistic $S_{2N}(0, t)$, a linearity result in terms of $t$ only, can be stated based on Theorem 3.1 of Jurečková. Note that presently, for the simple linear model we have only 1 independent variable, hence $p = 1,$ which is a special case of the Theorem 3.1 of Jurečková (1971).

We first begin with the conditions that were imposed for the model. To match out our notation take $z_i^*(0) = \log |e_i|,$ for $i = 1, \ldots, N.$

(J.1) $z_1^*(0), \ldots, z_N^*(0)$ are independent random variables having the distribution function

$$H_{e^*}(z^* - x_i^T \theta) \quad i = 1, \ldots, N$$

for $H_{e^*}$ with a finite Fisher’s information, and is centered around a nonzero value $\theta_0$.

(J.2) $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$ is a real vector parameter.

(J.3) $X = \{x_{ij}\}$ is an $N \times p$ matrix with rows $x_i^T$ for $i = 1, \ldots, N$ satisfying the concordance conditions (J.3 (a)), (J.3(b)), and (J.3(c)), which are not necessary as has been shown by Heiler and Willers (1988). In addition, the design conditions

(D1)

$$\lim_{N \to \infty} \max_{1 \leq i \leq N} x_i^T (X^T X)^{-1} x_i = 0.$$
\[
\lim_{N \to \infty} \frac{1}{N} (X^T X)^{-1} = \Sigma,
\]

where \(\Sigma\) is a \(p \times p\) positive definite matrix.

(J.4) \(R[v(0, t)] = (R[v_1(0, t)], \ldots, R[v_N(0, t)])^T\) is a vector of the ranks corresponding to the residuals \(v_i(0, t) = z_i^*(0) - x_i^T \frac{1}{\sqrt{N}} t\) for \(i = 1, \ldots, N\).

(J.5) Consider the linear statistics

\[
S_{2N_1}(0, t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_2^*(\frac{R[v_i(0, t)]}{N+1}) x_{ij}, \quad j = 1, \ldots, p.
\]

Also consider

\[
T_{2N_2}(0, t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_2^* \left( H_e^* (z_i^*(0) - x_i^T \frac{1}{\sqrt{N}} t) \right) x_{ij}, \quad j = 1, \ldots, p.
\]

(J.6) \(\phi_2^*(u)\) is non-constant, nondecreasing and square-integrable score generating function defined on \((0, 1)\)

Note that the condition (J.6) in Jurečková (1971) was based on \(\phi_1(u)\), which was generated by \(\phi_1\left(\frac{i}{N+1}\right)\) such that \(\sum_{i=1}^{N} \phi_1\left(\frac{i}{N+1}\right) = 0\). These conditions are fulfilled by our \(\phi_2^*(u)\). We next present Theorem 3.1 as a lemma

Lemma 2.4.1. Let \(P_N = \prod_{i=1}^{N} H_{z_i^*}(z_i^* - x_i^T \theta)\). Then, under assumptions (J.1) through (J.6),

\[
\lim_{N \to \infty} \left\{ \max_{||t|| \leq M} \left| S_{2N_1}(0, t) - S_{2N_1}(0, 0) + \gamma_2 t^T \Sigma^{(j)} \right| \geq \epsilon \right\} = 0,
\]

for any \(\epsilon > 0, M > 0, \Sigma^{(j)}\) is the \(j\)th column of the matrix \(\Sigma\) and \(j = 1, \ldots, p\).

Furthermore, the asymptotic normality of the estimator of \(\theta\) can be easily obtained by direct substitution into the steps of the proof for Theorem 3.1 of Jeackel (1972) to yield
Lemma 2.4.2. Assume that (D1), (D2), (F1) and (S2) hold. In addition, suppose that condition (A1) on the compactness of \( b \) to be given later is satisfied. Then

\[
\sqrt{N} \hat{\theta} \xrightarrow{d} N_p(0, \gamma^{-2}(X^T X)^{-1}).
\] (2.4.12)

2.5 General Asymptotic Results

In this section, we discuss asymptotic results that will be needed in developing the theory the behind the proposed estimator in chapter three and chapter four. We begin with a definition, due to Jaeckel (1972), of asymptotic equivalence of two estimators. Let \( \Delta_D \) and \( \hat{\Delta}_N \) be sequences of random vectors.

Definition 2. \( \Delta_D \) and \( \hat{\Delta}_N \) are said to be asymptotically equivalent if and only if the distance between corresponding terms converges to zero in probability. That is,

\[
\lim_{N \to \infty} P\left\{ \sup_{\Delta_D \in I(B_N)} \| \Delta_D - \hat{\Delta}_N \| \geq r \right\} = 0,
\]

for all \( r > 0 \).

Here, \( I(B_N) \) is an indicator function for the membership to the bounded set of possible solutions. The definition will be useful in establishing that the estimators proposed in this study are consistent for the respective parameters that they are targeting. That is, as \( N \) tends to be infinitely large, the values of the estimators approach the true values of the parameters in probability.

For the current estimation problem, in seeking to estimate one parameter when a certain optimal condition is imposed on the other parameter, strong consistency properties of estimator under such a condition, are well established. This also implies that the asymptotic distribution of the optimal solution is well known. However, since the true value of any parameter is rarely known, for the less than optimal solution being proposed here, it has yet to be established that the solution is asymptotically equivalent to
the optimal solution. That is, in chapter three and chapter four, we ascertain if

\[ \hat{\beta} - \hat{\beta}_{\text{sca}} = o_p(1), \]

and

\[ \hat{\theta} - \hat{\theta}_{\text{loc}} = o_p(1), \]

respectively. Here, \( \hat{\beta}_{\text{sca}} \) is the estimate obtained based on the signed-rank statistic (2.3.12) in Theorem 7.2 of Kraft and Eeden (1972), given here as Lemma 2.3.1. Recall from the remarks following the lemma it is assumed that the scale parameter is the true value, \( \theta \). \( \hat{\theta}_{\text{loc}} \) is the estimate that is obtained based on the rank statistic (2.4.11) in Theorem 3.1 of Jurečková (1971) given here as Lemma 2.4.1. Recall from the lemma that it is assumed that the regression parameter is the true value, \( \beta \).

Since the asymptotic linearity results for these optimal solution are well established in the aforementioned theorems, it is sufficient to show that the proposed estimators have the same asymptotic representations as their optimal analogues, to establish consistency. This is done in chapter three and chapter four, respectively, for the estimators of \( \beta \) and \( \theta \).
CHAPTER III

RANK ESTIMATION OF REGRESSION PARAMETER, $\beta$

3.1 Introduction

In this chapter, we restrict our attention to obtaining regression parameter estimates in a heteroscedastic linear model. In the present setting, the constants that express heteroscedasticity are multipliers of the error variables. It is customary to divide the equation by the constants so that the resultant error terms are homoscedastic. This transformation results in a model where the both the response variable and the independent variables are weighted by the dividing constants. In this study, the dividing constants also known as weights or scale constants depend on some scale parameters through a known variance function model. It is desirable to assume that the true values of the scale parameters are specified. As this is rarely the case in practice, in this chapter it is assumed that arbitrary but fixed parameter values are specified.

It is worth noting that, we seek robust residuals which are subsequently utilized to obtain the scale parameter estimates. Thus, it is important that the estimates of $\beta$ be robust in order to obtain robust subsequent scale parameter estimates. For robust estimates, the standard errors are not inflated in the presence of outlying responses. An increasingly popular approach for obtaining robust estimates is to use a rank based method. In the current study, the errors are assumed to be symmetrical around zero. In order to obtain robust estimates, we employ a signed-rank based estimation method. Since the model has been transformed by scale constants, weighted signed-rank statistic is considered.

The asymptotic theory of the estimator is based on the weighted rank statistic. Recall from chapter two that, in its present form, the standard signed-rank estimation method for homoscedastic linear models established by Kraft and van Eeden (1972) ex-
cludes the possibility of true scale parameters being unspecified. With some modification to the definition of the scale constants, a method that extends the result by Kraft and van Eeden (1972) to the heteroscedastic linear models is proposed.

It is worth noting that the assumptions that were imposed on the signed-rank theory included the concordance conditions of Jurečková (1971). These conditions have since been proven to be unnecessary for the asymptotic theory of the estimator to hold, due to the result in Tardif (1985).

In what follows, the true model that this investigation seeks to fit is presented.

3.2 Model for Location Problem

In this section, the heteroscedastic linear model being considered in this study is given. The model is multiplied by inverses of the scale constants to yield a homoscedastic linear model. Finally, the estimation problem is described.

Consider

\[ y_i = \beta_0 + x_i^T \beta_1 + \sigma_i e_i, \quad i = 1, \ldots, N, \]  

(3.2.1)

where \( y_1, \ldots, y_N \) are the responses, \( \beta_0 \in \mathcal{R} \), and \( \beta_1 \in \mathcal{R}^p \) are the regression parameters which we seek to estimate. Further, \( X \) is a design matrix with rows \( x_1^T, \ldots, x_N^T \), \( X \) is an \( N \times p \) matrix of known regression constants, \( X \) is centered, \( \sigma_1, \ldots, \sigma_N \) are scale constants which depend on the design matrix through the relationship

\[ \sigma_i = \exp\{x_i^T \theta\}, \quad i = 1, \ldots, N, \]  

(3.2.2)

where \( x_1^T, \ldots, x_N^T \) are rows of the \( N \times p \) design matrix, \( X, \theta \) is a \( p \times 1 \) vector of scale parameters. Here, the \( e_1, \ldots, e_N \), are iid random variables with common cdf, \( F(y) \), and
pdf, \( f(y) \). Note that \( f \) is symmetric about zero.

### 3.2.1 Scaling Transformation

From the equation in (3.2.1), the following alternative form is a consequence of dividing the model in (3.2.1) through by the scale constants, \( e^{x_i^T \theta}, \ldots, e^{x_i^T \theta} \).

\[
y_i e^{-x_i^T \theta} = \beta_0 (e^{-x_i^T \theta}) + e^{-x_i^T \theta} x_i^T \beta_1 + e_i, \quad i = 1, \ldots, N, \tag{3.2.3}
\]

which can be written as

\[
y_i^* = \beta_0 x_{0i}^* + x_i^T \beta_1 + e_i, \quad i = 1, \ldots, N, \tag{3.2.4}
\]

\( y_i^* = y_i e^{-x_i^T \theta}, \quad x_{0i}^* = e^{-x_i^T \theta}, \quad x_i^T = e^{-x_i^T \theta} x_i^T \), for convenience. Clearly, since \( e_1, \ldots, e_N \) are free of scale constants, the error terms in this model are homoscedastic insofar as the scale constants are accurately specified. Then standard rank methods can be employed to estimate the regression parameters. It should also be noted that the variance of estimators of \( \beta \) now depends on a transformed design \( W^{-1/2}(1, X) \), where \( W = \text{diag}\{\sigma_1^2, \ldots, \sigma_N^2\} \). In practice, the true values of the scale parameters are not specified. Thus, the scale parameters have to be replaced by their corresponding estimates. In anticipation of this possibility, the model given in (3.2.4) can be reparametrized. For the moment, we motivate the problem by assuming that the scale constants are arbitrary.

### 3.2.2 Weighted Estimation Problem

In this section, the problem of obtaining weighted estimates of \( \beta = (\beta_0, \beta_1^T)^T \) in the transformed model given in (3.2.4) is considered. Let \( X_1 = (1, X) \). For any \( b = (b_0, b_1^T)^T \),
where \( b_0 \in \mathcal{R} \) and \( b_1 \in \mathcal{R}^p \), define the residuals

\[
(y_1^* - x_{i1}^T b), \ldots, (y_N^* - x_{iN}^T b). \tag{3.2.5}
\]

For any given value of \( \theta \), we seek to obtain estimates of \( \beta \) such that the residuals are as small as possible. There are various methods for obtaining values of the desired parameter that minimize functions of the shifted errors. The decision on which method should be used depends on how the method behaves in the presence of outliers. Outlying responses can inflate residuals and ultimately cause the estimator to breakdown. In order to delimit the effects of such outliers on the estimate of \( \beta \), a linear rank statistic that is computed from the residuals defined in (3.2.5), can be employed. Recall that \( e_1, \ldots, e_N \) are assumed to be symmetrical about zero. Hence, it follows that a signed-rank process would be the suitable statistic to use when estimating \( \beta \) for a given value of \( \theta \).

Suppose \( R[|y_i^* - x_{i1}^T b|] \) denote the rank of \( |y_i^* - x_{i1}^T b| \) amongst \( |y_k^* - x_{i1}^T b| \), for \( k = 1, \ldots, N \). Consider

\[
S_1(b) = \sum_{i=1}^{N} \phi_1^+ \left( \frac{R[|y_i^* - x_{i1}^T b|]}{N + 1} \right) \text{sgn}(y_i^* - x_{i1}^T b)x_{i1}, \tag{3.2.6}
\]

Observe that \( S_1(b) = (S_{11}(b), \ldots, S_{1(p+1)}(b))^T \), and the set of positive-valued scores, \( \phi_1^+ \left( \frac{1}{N+1} \right) \), that is, \( \sum_{i=1}^{N} \phi_1^+ \left( \frac{1}{N+1} \right) > 0 \), where \( \phi_1^+(u) \) is a score generating function defined in an assumption given below. The suitability of the scores depends on the properties of the error distribution. The recommended approach for identifying which scores to use can be found in McKean and Sievers (1989). The asymptotic theory of the rank based estimates is developed by establishing the asymptotic linearity of the linear signed-rank statistic along the method of proof in Kraft and van Eeden (1972).

It is seen from the representation of the model given in (3.2.4) that it can be seen that \( e_i = y_i^* - x_{i1}^T \beta \) for \( i = 1, \ldots, N \). Since we are interested in estimating \( \beta \) when shifts in both location and scale are present, the residuals problem can be reexpressed in terms
of perturbations to $e_i$'s. In what follows, we consider the asymptotic behavior of the function of $e_i$ when the location shift, through $b$, and the scale shift, through $t$, have been induced. The asymptotic linearity is established under the condition that the variables $b$ and $t$, are arbitrary but fixed. Subsequently, we relax the "fixed values" restriction to accommodate estimates of location and scale parameters, which are random. Before defining such a statistic, the underlying assumptions required for obtaining an estimate of $\beta$ are specified.

3.2.3 Model Assumptions

In the sequel, the conditions that we impose on model in (3.2.4) are specified. The next two conditions apply to the design in this study.

(D1*)

$$\lim_{N \to \infty} \max_{1 \leq i \leq N} x_{ii}^* (X_1^T X_1^*)^{-1} x_{ii}^* = 0.$$ 

(D2*)

$$\lim_{N \to \infty} \frac{1}{N} (X_1^* T X_1^*) = \Sigma_1^*,$$

where $\Sigma_1^*$ is a $(p + 1) \times (p + 1)$ positive definite matrix. Recall that $X_1^* = W^{-1/2} X_1$, $W = diag\{\sigma_1^2, \ldots, \sigma_N^2\}$.

The next two conditions apply to the distribution of errors,

(F1) (i) $f(y)$ is absolutely continuous.

(ii) $\int_{-\infty}^{\infty} \left( \frac{f'(y)}{f(y)} \right)^2 f(y) dy < \infty$.

Thus, if (i) and (ii) hold, then $f(y)$ has a finite Fisher information $I(f)$. 

(F2) \( f(y) = f(-y) \), that is, \( f(\cdot) \) is symmetric about 0.

Next, the restriction on the score function is described. Since, in the location problem, the errors, \( e_1, \ldots, e_N \), are symmetrically distributed around 0, it follows that the scores generating function satisfy the following:

(S1) Let \( \phi_1^+ (u) \) be a nondecreasing, positive-valued, and square integrable function defined on the interval \((0,1)\). Further, \( \sum_{i=1}^{N} \phi_1^+ \left( \frac{i}{N+1} \right) > 0 \). Furthermore, due to the square-integrability of \( \phi_1^+ (u) \), the scores can be standardized so that \( \int_0^1 (\phi_1^+ (u))^2 \, du = 1 \). Note that in Hájek and Šidák (1967), it is assumed that \( \int_0^1 [\phi_1^+ (u) - \bar{\phi}_1]^2 \, du < \infty \), where \( \bar{\phi}_1 = \frac{1}{N} \sum_{i=1}^{N} \phi_1^+ \left( \frac{i}{N+1} \right) \).

Next, a restriction on the scaling constants \( \sigma_i \) is described, in view of Remark 2.3.1.

(W1) \( \sigma_1, \ldots, \sigma_N \) are bounded away from zero.

We now present the new notation that will be useful in developing the theory of this chapter.

3.2.4 Linear Signed-Rank Statistic

In this section, for the location component of the problem, we let \( b = (b_0, b_1^T) \), where \( b_0 \in \mathbb{R} \) and \( b_1 \in \mathbb{R}^p \) are any fixed parameter values. To obtain \( \sqrt{N} \)-consistent asymptotic results, we will be working with \( N^{-\frac{1}{2}} b \). Then, following Sievers (1978), p. 269, (1983), p. 1165, it should be noted that the vector \( N^{-\frac{1}{2}} b \) is a sequence of parameter values that converge to zero. Consider the scale component of the problem. Let \( t \in \mathbb{R}^p \) is an arbitrary but fixed vector. It is worth noting that similar to Sievers (1978), (1983), \( N^{-\frac{1}{2}} t \) is a sequence of parameter values that converge to zero. Thus, we define

\[
\sigma_i (t) = e^{\lambda^T \left( \theta + \frac{1}{\sqrt{N}} t \right)}, \quad i = 1, \ldots, N, \tag{3.2.7}
\]

\[
= \sigma_i e^{\frac{1}{\sqrt{N}} \lambda^T t}.
\]
Note that since \( t \) is fixed, it is selected so that \( \sigma_1(t), \ldots, \sigma_N(t) \) are also fixed constants. Define

\[
\begin{align*}
    z_i(b, t) &= \frac{y_i - x_i^T \beta - \frac{1}{\sqrt{N}} x_i^T b}{\sigma_i(t)}, \quad \text{for } i = 1, \ldots, N, \\
    &= \frac{e_i - \frac{1}{\sqrt{N}} x_i^T b}{\nu_i}, \\
    &= \frac{e_i - m_i^*}{\nu_i},
\end{align*}
\]

where \( m_i^* = \frac{1}{\sqrt{N}} x_i^T b, x_i^T \) is the \( i \)th row of the matrix \( X_i^* = W^{-1/2} X_i \), and \( g_i = \frac{1}{\sqrt{N}} x_i^T t \).

Then we establish asymptotic properties of the estimate of \( \beta \) based on the linear rank statistic

\[
S_{1N}(b, t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i^+ \left( \frac{R(\{|z_i(b, t)|\})}{N+1} \right) sgn(\nu_i) e^{-\frac{1}{2} x_i^T \nu_i^*},
\]

where \( S_{1N}(b, t) = (S_{1N1}^+(b, t), \ldots, S_{1N(N+1)}^+(b, t))^T \). Here, \( \phi_i^+(u) \) is a score generating function satisfying (S1), \( R(\{|z_i(b, t)|\}) \) is the rank of \( \{|z_i(b, t)|\} \) among \( \{|z_k(b, t)|\} \), for \( k = 1, \ldots, N \).

Recall from chapter two that we will be convenient to work with an alternative statistic

\[
S_{1N}(b, t, s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i^+ \left( \frac{R(\{|z_i(b, t)|\})}{N+1} \right) sgn(\nu_i) x_i^*,
\]

where \( S_{1N}(b, t, s) = (S_{1N1}(b, t, s), \ldots, S_{1N(N+1)}(b, t, s))^T \) and

\[
x_{i*}^* = e^{-\frac{1}{2} x_{i*}^T x_{i*}^*} x_{i*}^* = \frac{x_{i*}}{\sigma_i(s)} \quad \text{for } i = 1, \ldots, N.
\]

The difference between the statistics defined in (3.2.11) and (3.2.12) is the introduction of the third argument \( s \) which follows from the identity given in (3.2.13). The linear expansion of \( S(b, t) = S(b, t, s) \) that will be done later involve statistics of variables, \( z_i(0, 0) \). However, it is desirable to retain the regression constants as functions of \( t \) since the \( x \)s
are assumed to be fixed. The representation in (3.2.12) allows us to rename the fixed scale parameter value $t$ by $s$, hence, tying it with the fixed regression constants in $x_i^t$, while retaining the stochastic properties of the statistics through the variables $z_i(b, t)$, for $i = 1, \ldots, N$.

In what follows, we develop the theory on the asymptotic behavior of the statistic $S_{1N}(b, t, s)$.

3.3 Asymptotic Properties of Estimator

In this section, a result on the asymptotic normality of the estimating process $S_{1N}(b, t, s)$ is established based on an asymptotic linearity condition. From this result, the asymptotic distribution of the proposed estimator is derived. This parallels the approach used by Jurečková (1971), and Kraft and van Eeden (1972) in establishing the theory for the homoscedastic linear models. Here, transforming the model by dividing through by $\sigma_i(t)$ instead of $\sigma_i$ enables us to keep the theory simple for the moment. With this transformation, the well known asymptotic results established by the aforementioned authors are extended to accommodate the multi-parameter case that is of interest in this estimation problem. It is worth pointing out that for the single parameter case, an asymptotic linearity result for the signed-rank estimator was established by Kraft and van Eeden (1972).

Before the development of the theory behind the estimating process begins, it will be useful to describe of the distributive properties of the underlying random variables of the statistics considered in the sequel.

3.3.1 Probability Distribution of Errors

In this section, we derive the distribution functions of the underlying variables of the statistics utilized in obtaining the linearity result. Consider the random variables $z_i(b, t) = (e_i - m_i^*)e^{-\theta}$, and indeed observe that $z_i(0, 0) = \ldots$
$e_i$ for $i = 1, \ldots, N$. Thus,

$$z_i(0, 0) \mid_{e_i \sim F(v)} \text{ has cdf } F(v). \quad (3.3.1)$$

Further, $|z_1(0, 0), \ldots, z_N(0, 0)|$ have the distribution $F^+(v) = F(v) - F(-v)$. Recall that the random variables $e_1, \ldots, e_N$ are independent, identically distributed (iid) with common cdf, $F$. Let the joint densities of these random variables be defined as

$$p_N = \prod_{i=1}^{N} f(e_i), \quad (3.3.2)$$

where, $f$ is the corresponding common density function.

### 3.3.2 Distributional Properties of Linear Signed-Rank Statistics

As it will be seen in the linearity result that is given later, the asymptotic behavior of the empirical process $S_{1N}(b, t, s)$ depends on the limiting distribution of $S_{1N}(0, 0, s)$. Thus, in this section, the asymptotic normality of the process $S_{1N}(0, 0, s)$ is established assuming that the errors $e_1, \ldots, e_N$ are jointly distributed as $p_N$. To obtain this result, it is more convenient to work with $T_{1N}(0, 0, s)$, which is an approximation to the empirical process. Subsequently, the result is extended to the empirical process by demonstrating that the empirical process and its approximation are asymptotically equivalent. In addition, for another empirical process, $S_{1N}(0, 0, 0)$, that is defined below, it is shown that its approximation, $T_{1N}(0, 0, 0)$, is asymptotically equivalent to $T_{1N}(0, 0, s)$. Then the asymptotic equivalence of the corresponding empirical processes $S_{1N}(0, 0, s)$ and $S_{1N}(0, 0, 0)$ follows easily from this result.

Before these results can be established, it will be useful to provide two remarks on the relationship between the convergence of vectors of statistics and that of the respective components of those vectors.
Remark 3.3.1. Consider \((p + 1)\)-vectors \(S_{1N}\) and \(T_{1N}\). For any \(a \in \mathbb{R}^{p+1}\), define a linear combination \(X_i^* a = c^*\) such that
\[
c_i^* = a_1x_{i1}^* + a_2x_{i2}^* + \ldots + a_{p+1}x_{i(p+1)}^* \quad \text{for} \quad i = 1, \ldots, N.
\]

For a fixed \(h, 1 \leq h \leq (p + 1)\), denote \(A_h = \{a : a_j = 1 \text{ for } j = h; a_j = 0 \text{ for } j \neq h, j = 1, \ldots, p + 1\}\). Then, for this particular choice of \(a\), \(a^T S_{1N}(b, t, s)\) and \(a^T T_{1N}(b, t, s)\) selects the \(j\)'th components of statistics \(S_{1N}(b, t, s)\) and \(T_{1N}(b, t, s)\), respectively.

Remark 3.3.2. In considering convergence results of \((p + 1)\)-vectors \(S_{1N}(b, t, s)\) and \(T_{1N}(b, t, s)\), it is worth noting the well known result that componentwise convergence in probability implies convergence in probability of the entire vectors. Thus, convergence results that have been established for components of vectors can be extended to the entire vectors in view of this fact.

From the class of functions of random variables \(z_i(0, 0) = e_i\) for \(i = 1, \ldots, N\), consider an empirical process given by
\[
S_{1N}(0, 0, s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_1 \left( \frac{R[|e_i|]}{N+1} \right) sgn(e_i) x_{i*}^*, \quad (3.3.3)
\]
where \(S_{1N}(0, 0, s) = (S_{N1}(0, 0, s), \ldots, S_{N(p+1)}(0, 0, s))^T\). The distributive properties of the statistic \(S_{1N}(0, 0, s)\) can be obtained as an extension of the result by Hájek and Šidák [1967], p. 166]. It should be noted that assumption (S1) imposed on the scores for the location problem is a special case of the condition utilized by the aforementioned authors since \(\int_0^1 \phi_1^+(u) du = 0\) and \(\int_0^1 \phi_1^{+2}(u) du = 1\). For the empirical process in (3.3.3), the following approximation will be useful
\[
T_{1N}(0, 0, s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_1^+ \left( F^+([e_i]) \right) sgn(e_i) x_{i*}^*, \quad (3.3.4)
\]
where \(T_{1N}(0, 0, s) = (T_{N1}(0, 0, s), \ldots, T_{N(p+1)}(0, 0, s))^T\). The distributive properties of
this approximation follow from the proof of Theorem V.1.7 of Hájek and Šidák (1967).

**Theorem 3.3.1.** Assume that (D1*), (D2*), (F1), (F2), (S1) and (W1) hold.

Then

\[ T_{1N}(0, 0, s) \xrightarrow{D} N_{p+1}(0, \Sigma_1^*), \]  

(3.3.5)

where \( \Sigma_1^* \) is a \((p + 1) \times (p + 1)\) positive definite matrix defined in assumption (D2*).

Note: "\( \xrightarrow{D} \)" means "converges in distribution to".

**Proof.** For any \( a \in \mathbb{R}^{p+1} \), employ an arbitrary linear combination so that

\[ a^T T_{1N}(0, 0, s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i^+(|e_i|) \text{sgn}(e_i)c_i^*, \]  

(3.3.6)

where \( c_i^* = X_i^*a \) is as defined in Remark 3.3.1. Note that, except for the constants \( c_i^* \),

\[ a^T T_{1N}(0, 0, s) \] is a sum of independent, identically distributed terms \( \phi_i^+(|e_i|) \) with mean, \( \int \phi_i^+(F^+(|e|)) \text{sgn}(e)dF(e) = 0 \) and variance, \( \int [\phi_i^+(F^+(|e|))]^2dF(e) = \int \{\phi_i^+(v)\}^2dv = 1 \), in light of assumption (S1).

Observe that

\[ E\left[ a^T T_{1N}(0, 0, s) \right] \xrightarrow{D} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i^* \int \phi_i^+(|e|) \text{sgn}(e)dF(e) \]

\[ = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i^* \int \phi_i^+(v)dv \]

\[ = 0, \]
and

\[
\text{Var} \left[ a^T T_{1N}(0, 0, s) \right]_{e \sim p_N} = \frac{1}{N} \sum_{i=1}^{N} c_i^{*2} \int \left[ \phi^+_1(F^+(|e|)) \right]^2 \text{sgn}(e) dF(e)
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} c_i^{*2} \int \{ \phi^+_1(v) \}^2 dv
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} c_i^{*2}.
\]

Furthermore, \( \frac{1}{N} \sum_{i=1}^{N} c_i^{*2} = a^T (\frac{1}{N} X_1^{*T} X_1^{**}) a \) tends to \( a^T \Sigma^*_1 a \) as \( N \to \infty \) due to assumption (D2*). Thus, by an application of Lindeberg Central Limit Theorem (Hájek and Šidák, p. 153 (1967)), we establish that \( a^T T_{1N}(0, 0, s)_{e \sim p_N} \xrightarrow{d} N(0, a^T \Sigma^*_1 a) \). Therefore, the desired result in (3.3.5) follows from this fact, which completes the proof. \( \square \)

In the sequel, the limiting distribution of the empirical process \( S_{1N}(0, 0, s) \) is considered. This result can be obtained if it can be shown that \( S_{1N}(0, 0, s) - T_{1N}(0, 0, s) \) goes to zero in probability. Note that an analogous result was established in Theorem 3.1 of Adichie [1967], p. 886], which considered a signed-rank process for the simple linear homoscedastic model type problem. To this end, we extend the result by Adichie to the weighted estimation problem under consideration in this study.

**Theorem 3.3.2.** Assume that (D1*), (D2*), (F1), (F2), (S1) and (W1) hold.

Then

\[
\left\| S_{1N}(0, 0, s) - T_{1N}(0, 0, s) \right\|_{e \sim p_N} = o_p(1). \tag{3.3.7}
\]

**Proof.** First, we show that \( E[|S_{1N}(0, 0, s)|]_{e \sim p_N} = 0 \). For any \( a \in \mathcal{R}^{p+1} \), we obtain an
arbitrary linear combination $a^T S_{1N}(0,0,s)$ such that

$$a^T S_{1N}(0,0,s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i^+ \left( \frac{R ||e_i||}{N+1} \right) sgn(e_i)c_i^*.$$  

Then

$$E \left[ a^T S_{1N}(0,0,s) \right]_{e \sim p_N} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i^* E \left[ \phi_i^+ \left( \frac{R ||e_i||}{N+1} \right) sgn(e_i) \right].$$

Observe that $|e_1|, \ldots, |e_N|$ and $sgn(e_1), \ldots, sgn(e_N)$ are mutually independent. Further, $E[sgn(e)] = 0$. Using these facts, it can be easily shown that $E[a^T S_{1N}(0,0,s)]_{e \sim p_N} = 0$. Then it suffices to show that

$$\left| a^T S_{1N}(0,0,s) - a^T T_{1N}(0,0,s) \right|_{e \sim p_N} = o_p(1).$$

This result can be obtained if we can demonstrate that

$$\text{Var} \left[ a^T S_{1N}(0,0,s) - a^T T_{1N}(0,0,s) \right]_{e \sim p_N} \to 0 \quad \text{as} \quad N \to \infty.$$ From the result above we have $E[a^T S_{1N}(0,0,s)]_{e \sim p_N} = 0$. Further, recall from Theorem 3.3.1 that $E[a^T T_{1N}(0,0,s)]_{e \sim p_N} = 0$.

Furthermore, following Hájek and Šidák, p. 153 (1967), we note that our $\phi_i^+ \left( \frac{i}{N+1} \right)$ defined as

$$\phi_i^+ \left( \frac{i}{N+1} \right) = E[\phi_i^+(U_i^1) | R_{1N}^+ = i], \text{ for } i = 1, \ldots, N,$$

is approximation of the scores $\phi_i^+ \left( \frac{R_i^+}{N+1} \right)$ where $U_i^+ = F^+(|Y_i^*|)$ are independent random variables that follow the uniform$(0,1)$ distribution, $R_i^+$ is the rank of the $|Y_i^*|$ amongst $|Y_k^*|$ for $k = 1, \ldots, N$. Note that $\phi_i^+ \left( \frac{i}{N+1} \right)$ corresponds to $a_N^\phi \left( \frac{1}{2} + \frac{R_i^+}{2(N+1)} \right)$ where $a_N^\phi(i)$ is
defined in Hájek and Šidákov (1967). Then,

$$\lim_{N \to \infty} E \left[ \phi_1^+ \left( \frac{R_{N1}^+}{N + 1} \right) - \phi_1^+(U_1^+) \right]^2 = 0,$$

(3.3.10)

by Theorem V.1.4(a) of Hájek and Šidákov (p. 157, 1967). Further,

$$\lim_{N \to \infty} \int_0^1 \left[ \phi_1^+ \left( \frac{(1 + [uN])/(N + 1)} - \phi_1^+(u) \right) \right] du^2 = 0,$$

(3.3.11)

by Theorem V.1.4(b) of Hájek and Šidákov (p. 158, 1967). Clearly,

$$E \left[ a^T S_{1N}(0, 0, s) - a^T T_{1N}(0, 0, s) \right] \xrightarrow{e \sim \rho_N} 0, \quad \text{as } N \to \infty,$$

due to the result in (3.3.10). It follows that

$$\text{Var} \left[ a^T S_{1N}(0, 0, s) - a^T T_{1N}(0, 0, s) \right] \xrightarrow{e \sim \rho_N}$$

$$= E \left[ a^T S_{1N}(0, 0, s) - a^T T_{1N}(0, 0, s) \right]^2 \xrightarrow{e \sim \rho_N}$$

$$= E \left[ \frac{1}{\sqrt{N}} \sum_{i=1}^N c_i \phi_1^+ \left( \frac{R_{i1}^+}{N + 1} \right) - \phi_1^+(U_i^+) \right]^2 \xrightarrow{e \sim \rho_N}$$

$$\leq \frac{N}{N - 1} \left\{ \frac{1}{N} \sum_{i=1}^N c_i^{*2} \right\} E \left[ \phi_1^+ \left( \frac{R_{i1}^+}{N + 1} \right) - \phi_1^+(U_i^+) \right]^2 \xrightarrow{e \sim \rho_N}$$

$$\to 1 \cdot a^T \Sigma_i^* a \cdot 0.$$

The second equality is due to the fact that \(|Y_1^*, \ldots, Y_N^*|\), \((R|Y_1^*, \ldots, R|Y_N^*)\) and \((\text{sgn}(Y_1^*), \ldots, \text{sgn}(Y_N^*))\), are mutually independent. Further, \(E[\text{sgn}(Y_i^*)] = 0\) and \(E[\text{sgn}(Y_i^*)]^2 = 1\). Note in the last inequality that \(\frac{1}{N} \sum_{i=1}^N c_i^{*2} \to a^T \Sigma_i^* a\) as \(N \to \infty\), in view of assumption (D2*). For fixed \(s\), the regression constants satisfy the conditions, (D1*) and (D2*), which were required in the theorem by Adichie. The bound in \(\frac{1}{N-1}\{\cdot\}\) is due to Lemma 2.3 of Hájek (1961). Now, the result \(E[\cdot]^2 \to 0\) is a direct consequence of an application of The-
It can be seen that the following theorem is an immediate consequence of the last theorem.

**Theorem 3.3.3.** Assume that \((D1^*), (D2^*), (F1), (F2), (S1) and (W1) hold. Then

\[
S_{1N}(0, 0, s) \xrightarrow{\text{e}^{|p|}} N_{(p+1)}(0, \Sigma_1^*). \tag{3.3.12}
\]

**Proof.** From Theorem 3.3.1, we have the result \(T_{1N}(0, 0, s) \xrightarrow{\text{e}^{|p|}} N_{(p+1)}(0, \Sigma_1^*)\). Since \(|S_{1N}(0, 0, s) - T_{1N}(0, 0, s)| \xrightarrow{\text{e}^{|p|}} o_p(1)\) in view of Theorem 3.3.2, the desired result follows from the fact that convergence in probability implies convergence in distribution. This completes the proof. \(\Box\)

Recall that the third argument in the statistic \(S_{1N}(0, 0, s)\) is fixed so that the design matrix is scaled by a function of the fixed-valued \(s\).

Consider the process

\[
S_{1N}(0, 0, 0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_1^+ \left( \frac{R|e_i|}{N + 1} \right) sgn(e_i) x_{1i}^*. \tag{3.3.13}
\]

Observe that, in this process, the errors are free of shift with respect to location or scale. It is of interest to ascertain if \(S_{1N}(0, 0, s) - S_{1N}(0, 0, 0) |_{e \sim p_N} \) goes to zero in probability. This is the goal of the next theorem.

**Theorem 3.3.4.** Assume that \((D1^*), (D2^*), (F1), (F2), (S1) and (W1) hold.
Then
\[ \left\| S_{1N}(0,0,s) - S_{1N}(0,0,0) \right\|_{\text{op}} = o_p(1). \] (3.3.14)

**Proof.** For any \( a \in \mathbb{R}^{p+1} \), define arbitrary linear combinations \( a^T S_{1N}(0,0,s) \) and \( a^T S_{1N}(0,0,0) \) so that
\[ a^T S_{1N}(0,0,s) - a^T S_{1N}(0,0,0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i^+(\frac{R[|e_i|]}{N+1}) \text{sgn}(e_i) d_i, \] (3.3.15)
where \( d_i = (e^{-\frac{1}{\sqrt{N}}x_i^T a} - 1) a^T x_i^* \). It is easy to demonstrate that for any \( \epsilon_0 > 0 \), \( |e^{-\frac{1}{\sqrt{N}}x_i^T a} - 1| < \epsilon_0 \), for sufficiently large \( N \).

Note that
\[ \frac{1}{N} \sum_{i=1}^{N} d_i^2 \leq \epsilon_0^2 \frac{1}{N} a^T \left( \frac{1}{N} X_1^* X_1^* \right) a. \]

Note in the last equality that \( \frac{1}{N} X_1^* X_1^* \) converges to the constant \( \Sigma_1^* \) on account of assumption (D2*). Further, since \( \epsilon_0 \) is fixed, it can always be chosen to arbitrarily small. Hence, it is clear that \( \frac{1}{N} \sum_{i=1}^{N} d_i^2 \) tends to zero as \( N \to \infty \). Except for the constants \( c_i^* \) being replaced by constants \( d_i \), the right hand side of (3.3.15), is the same as the definition in (3.3.8). Therefore, using the fact that \( \frac{1}{N} \sum_{i=1}^{N} d_i^2 \to 0 \), as \( N \to \infty \) along with Theorem 3.3.4 and Theorem 3.3.1, it is clear that \( a^T S_{1N}(0,0,s) - a^T S_{1N}(0,0,0) \mid_{e \sim p_N} \to N(0,0) \), which is the degenerate normal distribution. From the well known fact that convergence in distribution to degenerate normal implies convergence in probability, we conclude that
\[ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i^+(\frac{R[|e_i|]}{N+1}) \text{sgn}(e_i) d_i \mid_{e \sim p_N} = o_p(1). \] (3.3.16)

Since the result holds for every component of \( \|S_{1N}(0,0,s) - S_{1N}(0,0,0)\|_{e \sim p_N} \), the conver-
gence holds for the entire vector, in light of Remark 3.3.2, which terminates the proof. □

The next result is an immediate consequence of Theorem 3.3.4.

**Corollary 1.** Assume that \((D1^\ast), (D2^\ast), (F1), (F2), (S1)\) and \((W1)\) hold.

Then

\[
\| T_{1N}(0,0,s) - T_{1N}(0,0,0) \|_{e^\sim_P N} = o_P(1). \tag{3.3.17}
\]

**Proof.** The proof is similar to that of Theorem 3.3.4 in that its crux lie in factoring the coefficients \(d_i\) and an application of Theorem 3.3.1. □

It can be seen that Theorem 3.3.4 and Corollary 1 lead us to the result given in the next theorem.

**Theorem 3.3.5.** Assume that \((D1^\ast), (D2^\ast), (F1), (F2), (S1)\) and \((W1)\) hold.

Then

\[
\| S_{1N}(0,0,0) - T_{1N}(0,0,0) \|_{e^\sim_P N} = o_P(1). \tag{3.3.18}
\]

**Proof.** The proof uses the same arguments as those employed to establish Theorem 3.3.2, with \(X_1^\ast\) substituted for \(X_1^{\ast^\ast}\) so that for any \(a \in \mathbb{R}^{p+1}\), the linear combination \(X_1^\ast a = c^\ast\) instead of \(X_1^{\ast^\ast} a = c^{\ast^\ast}\). This completes the proof. □

This concludes the discussion on the distributive properties of the various processes of interest for the case in which no shift in location was assumed. In the sequel, the process for which the errors have been shifted or perturbed with respect to location is considered.
3.3.3 Asymptotic Linearity of Shifted Process $S_{1N}(b, 0, 0)$

In this section, the asymptotic behavior of the empirical process $S_{1N}(b, 0, 0)$ is examined. A first order expansion, yields a function that is linear in $b$, which in turn is used to derive the limiting distribution of the proposed estimator of $\beta$.

Consider the process

$$S_{1N}(b, 0, 0) = \left( S_{11}(b, 0, 0), \ldots, S_{1N+p+1}(b, 0, 0) \right)^T.$$

Then the approximation to this process is given by

$$T_{1N}(b, 0, 0) = \left( T_{11}(b, 0, 0), \ldots, T_{1N+p+1}(b, 0, 0) \right)^T.$$

where $T_{1N}(b, 0, 0) = (T_{1N1}(b, 0, 0), \ldots, T_{1N(p+1)}(b, 0, 0))^T$. For the shifted processes, it will be useful to define the joint distribution of the $e_1, \ldots, e_N$. First, note that, for $i = 1, \ldots, N$, we have

$$P(z_i(b, 0) \leq v) = P(e_i \leq m_i^* \leq v) = P(e_i \leq v + m_i^*) = F(v + m_i^*).$$

Here, $m_i^* = \frac{1}{\sqrt{N}} x_i^T b$, for $i = 1, \ldots, N$. If the random variables $e_1, \ldots, e_N$ follow the distribution $F(v + m_1^*), \ldots, F(v + m_N^*)$, then the joint distribution of these variables is given by

$$q_N(b, 0) = \prod_{i=1}^{N} f(e_i + m_i^*).$$

Suppose that we are interested in the distributive properties of $S_{1N}(b, 0, 0)$ under $p_N$. This result can be obtained from the distributive properties of $S_{1N}(0, 0, 0)$ under $q_N(b, 0)$, in view of the following translation property.
Lemma 3.3.1. For any fixed $b$,

$$S_{1N}(b, 0, 0) \overset{e \sim p_N}{=} S_{1N}(0, 0, 0) \overset{e \sim q_N(b, 0)}{=} .$$ (3.3.22)

Note that "$\overset{e}{\sim}$" means 'has the same distribution as'.

The proof of this property follows from the fact the left hand side of (3.3.22) is a function of $z_i(b, 0)$ and the right hand side of (3.3.22) is a function of $z_i(0, 0)$. It can be shown that assuming each $e_i$ follows $F(u)$, the random variables $z_i(b, 0)$ have the same distribution as $z_i(0, 0)$, assuming each $e_i$ follows $F(u + m_i^*)$. This fact can be obtained by using the same strategy as that employed in the more general result proven in lemma to be given later.

In the sequel, the distribution of the process $S_{1N}(0, 0, 0)$ under $q_N(b, 0)$ that is established, embraces a form of Theorem VI.2.5 of Hájek and Šidák (1967), p. 220) as a special case. In particular, Theorem VI.2.5 considered the case in which the regression constants are all ones. Since an analogous result for the general linear problem is needed, the regression constants are assumed to be arbitrary, following the method of proof in Theorem VI.2.4 of Hájek and Šidák (1967), p. 216). The latter theorem is postponed until chapter four, where it is directly applicable.

Theorem 3.3.6. Let be $c_i^*$ and $m_i^*$ be centered constants.

Consider the statistic

$$S_{1N}^{+} = \sum_{i=1}^{N} c_i^* \phi_i^+ \left( \frac{R_i^+}{N + 1} \right)$$

(3.3.23)

where $R_i^+$ is rank of observation $|y_i|$ amongst $|y_1|, \ldots, |y_N|$, $c_i$ satisfying condition

(C1): $\max_{1 \leq i \leq N} c_i^2 / \sum_{i=1}^{N} c_i^2 \to \infty$,

and $\phi_i^+(u)$ satisfy

(HS1): $\lim_{N \to \infty} \int_{0}^{1} [\phi_i^+ \left( \frac{1 + [uN]}{N} \right) - \phi_i^+(u)]^2 = 0$, $0 < u < 1$, $[uN]$ the largest integer not exceeding $u$. 


Let $q_N(b,0) = \prod_{i=1}^{N} f(e_i + m_0^*)$.

Then $S_{hs}^+$ is asymptotically normal with

\[
\text{mean} = \sum_{i=1}^{N} c_i^* m_i \int_0^1 \phi_1^+(u) \phi_1^+(u,f) du \quad \text{and variance} \quad \sum_{i=1}^{N} c_i^{*2} \int_0^1 \left[ \phi_1^+(u) - \overline{\phi_1^+} \right]^2 du,
\]

where

\[
\phi_1^+(u,f) = \frac{f'(F^{-1}(1+u/2))}{f(F^{-1}(1+u/2))},
\]

and $f$ is the pdf, $f(y) = dF(y)$.

It is worth noting that $c_i^*$ are centered, so that condition (C1) is readily satisfied under assumption (D1*), and (HS1) is a special case of assumption (S1). The proof for this result draws on fact that the statistic given in (3.3.23) converges to quadratic mean approximation which is derived by replacing the scores $\phi_1^+(R_1^+/N+1)$ by scores $\phi_1^+(U_i)$. Note that $U_1, \ldots, U_N$ are random variables from a uniform(0,1) obtained by utilizing the fact $U_i = F^+(Y_i)$. The asymptotic normality of the approximation is obtained by exploiting Le Cam's third lemma (Hájek and Šidák, 1967, p. 208). Finally, the result is extended to the statistic $S_{hs}^+$ through a contiguity argument and translation property. In the present problem, we seek to determine the distribution of $S_{1N}(0,0,0)$ under $q_N(b,0)$. Note that the translation property given in Lemma 3.3.1 can be used to obtain this result. Recall that $S_{1N}(b,0,0)$ can be approximated by $T_{1N}(b,0,0)$, under $p_N$. Then, similar to the former process, we have the following analogous translation property

**Lemma 3.3.2.** For any fixed $b$,

\[
T_{1N}(0,0,0) \bigg|_{b \sim q_N(b,0)} \stackrel{d}{=} T_{1N}(b,0,0) \bigg|_{b \sim p_N}, \tag{3.3.24}
\]

It follows from this lemma that, it suffices to establish distribution of $T_{1N}(0,0,0)$
under \( q_N(b, 0) \), in order to obtain the main result of this section. This result is furnished in the next theorem.

**Theorem 3.3.7.** Assume that \((D1^*), (D2^*), (F1), (F2), (S1)\) and \((W1)\) hold. Then

\[
T_{1N}(0, 0, 0) \bigg|_{e^{-q_N(b, 0)}} \overset{d}{\to} N_{p+1}(-\gamma_1 \Gamma^*_1 b, \Sigma^*_1),
\]

(3.3.25)

where

\[
\gamma_1 = \int \phi_1^+(u) \phi_1^+(u, f) du,
\]

(3.3.26)

such that

\[
\phi_1^+(u, f) = -f'(F^{-1}(\frac{1+u}{2}))/f(F^{-1}(\frac{1+u}{2})).
\]

(3.3.27)

**Proof.** For \( a \in \mathcal{R}(p+1) \), define an arbitrary linear combination \( a^T S_{1N}(0, 0, 0) \), where \( c^* = X_i^* a \) as described in Remark 3.3.1. Consider Theorem 3.3.6. For \( i = 1, \ldots, N \), let \( c_i^* \) be the regression constants and \( m_i^* = -\frac{1}{\sqrt{N}} x_i^* b \) be constants representing the shift in location. Following the proof of Theorem VI.2.4 of Hájek and Šidák (1967), we seek

\[
E[a^T S_{1N}(0, 0, 0)]|_{e^{-q_N(b, 0)}}.
\]

Recalling Lemma 3.3.2, it is seen that

\[
E\left[a^T T_{1N}(0, 0, 0) \right|_{e^{-q_N(b, 0)}} = E\left[a^T T_{1N}(b, 0, 0) \right|_{e^{-p_N}}.
\]

Thus, it suffices show that

\[
E\left[a^T T_{1N}(b, 0, 0) \right|_{e^{-p_N}} \to -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i^* m_i^* \int \phi_1^+(u) \phi_1^+(u, f) du \quad \text{as } N \to \infty.
\]

(3.3.28)

Also, recall from Theorem 3.3.1 that \( E[a^T T_{1N}(0, 0, 0)]|_{e^{-p_N}} = -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i^* \int_{0}^{1} \phi_1^+(|u|) dF^+(|u|) = 0.\)
Then observe that

\[ E \left[ a^T T_{1N}(b, 0, 0) \right]_{\mu \sim \mu_N} = E \left[ a^T T_{1N}(b, 0, 0) - a^T T_{1N}(0, 0, 0) \right]_{\mu \sim \mu_N} \]

\[ = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i^* \int_0^1 \left[ \phi_1^+ \left( F^+(|e_i - m^*_i|) \right) - \phi_1^+ \left( F^+(|e_i|) \right) \right] dF^+(|e_i|) \]

\[ = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i^* \int_0^1 \phi_1^{+/-} \left( F^+(w) \right) dF^+(w)(-m^*_i)dF^+(w) + o_p(1) \]

\[ = -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i^* m^*_i \int_0^1 \phi_1^{+/-} \left( F^+(w) \right) d(F^+(w))^2 + o_p(1), \]

where the second equality is a consequence of an application of the mean value theorem.

It is seen that

\[ -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i^* m^*_i = -\frac{1}{\sqrt{N}} a^T X^* X^*_i \frac{1}{\sqrt{N}} b = -a^T \left( \frac{1}{N} X^* X^*_i \right) b \rightarrow a^T \Sigma^*_i b \quad \text{as} \quad N \rightarrow \infty. \]

In addition, it can be shown by using integration by parts and substitution techniques that

\[ \int_0^1 \phi_1' \left( F^+(w) \right) d(F^+(w))^2 = \int_0^1 \phi_1^+ (u) \left( -\frac{f'(F^{-1}(1+u))}{f'(F^{-1}(1+u))} \right) du = \int_0^1 \phi_1^+ (u) \phi_1^+ (u, f) du, \]

which is denoted by \( \gamma_1 \), in view of (3.3.26).

Consider

\[ Var \left[ a^T T_{1N}(0, 0, 0) \right]_{\mu \sim \mu_N(b, 0)}. \]

It is sufficient to derive \( E[a^T T_{1N}(0, 0, 0)]^2_{\mu \sim \mu_N(b, 0)}. \) Similar to the proofs of Theorem VI.2.4 and Theorem VI.2.5 of Hájek and Šidák (1967), in light of assumption (S1), it can be
shown that

\[
E \left[ a^T T_{1N} (0, 0, 0) \right]_{e \sim q_N (b, 0)}^2 = \frac{1}{N} \sum_{i=1}^{N} c_i^2 \int_{0}^{1} [\phi_i^+ (u)]^2 du = a^T \left( \frac{1}{N} X_i^T X_i \right) a \int_{0}^{1} [\phi_i^+ (u)]^2 du \rightarrow a^T \Sigma_i^* a \cdot 1 \quad \text{as} \quad N \rightarrow \infty.
\]

Note that convergence was due to the fact that \( \left( \frac{1}{N} X_i^T X_i \right) \rightarrow \Sigma_i^* \) as \( N \rightarrow \infty \), on account of assumption (D2*). Furthermore, it is observed in the second equality that \( \int_{0}^{1} [\phi_i^+ (u)]^2 du = 1 \), in light of assumption (L1). Then, using LeCam’s third lemma as done in Theorem VI.2.4 and Theorem VI.2.5 of Hájek and Šidák (1967), we obtain the result

\[
a^T T_{1N} (0, 0, 0) \mid_{e \sim q_N (b, 0)} \xrightarrow{D} N_{p+1} (-\gamma_1 a^T \Sigma_1^* b, a^T \Sigma_1^* a).
\]

This result holds for every component of \( T_{1N} (0, 0, 0) \) under \( q_N (b, 0) \). Then, the limiting distribution for the entire vector \( T_{1N} (0, 0, 0) \) under \( q_N (b, 0) \) that we seek follows from this fact, along with Remark 3.3.2. The proof is complete.

It has remains to be shown that the limiting distribution in the last result also holds for the statistic \( S_{1N} (0, 0, 0) \) under \( q_N (b, 0) \). Recall from Theorem 3.3.5 that \( S_{1N} (0, 0, 0) - T_{1N} (0, 0, 0) \) goes to zero in probability under \( p_N \). Then, an appeal to the contiguity of \( q_N (b, 0) \) to \( p_N \) affords us the desired result. For convenience, a brief description of the principle of contiguity is provided by the next definition.

**Definition 3.** A sequence of densities \( q_N \) is contiguous to another sequence of densities \( p_N \) if for any sequence of events \( \{ A_N \} \),

\[
\int_{\{ A_N \}} p_N \rightarrow 0 \Rightarrow \int_{\{ A_N \}} q_N \rightarrow 0.
\]

It is shown in the monograph by Hájek and Šidák (1967, p.217) that the sequence
of densities $q_N(b,0)$ is contiguous to that of densities $p_N$. Since

$$
\left\| S_{1N}(0,0,0) - T_{1N}(0,0,0) \right\|_{e \sim p_N} = o_p(1),
$$

in view of Theorem 3.3.5, it follows that

$$
\left\| S_{1N}(0,0,0) - T_{1N}(0,0,0) \right\|_{e \sim q_N(b,0)} = o_p(1).
$$

Furthermore, since the convergence in probability implies convergence in distribution, it is clear that the next theorem follows from this discussion.

**Theorem 3.3.8.** Assume that $(D1^*)$, $(D2^*)$, $(F1)$, $(S1)$ and $(W1)$ hold. Then

$$
S_{1N}(0,0,0) \bigg|_{e \sim q_N(b,0)} \xrightarrow{d} N_{p+1}(-\gamma_1 \Sigma_1^* b, \Sigma_1^*),
$$

(3.3.28)

The expectation in the last result specifies the linearity term which is contained in the linear function that is used to approximate both of the processes $T_{1N}(b,0,0)$ and $S_{1N}(b,0,0)$ under $p_N$. In particular, an appeal to the appropriate translation property affords us this result. The next section focuses on the asymptotic linearity of the shifted process.

### 3.4 Asymptotic Linearity Result

In this section, asymptotic linearity results for both of the processes $T_{1N}(b,0,0)$ and $S_{1N}(b,0,0)$ are discussed. Note that both of these statistics represent the case where the errors have been perturbed with respect to location only. For these statistics, the standard asymptotic linear theory of Kraft and van Eeden (1972) is applicable. Following the strategy employed in Jurečková (1969), (1971), the asymptotic linearity of the approximating process is demonstrated first, then the result is extended to the empirical process.
In what follows, the motivation behind the method of proof for the asymptotic linearity result is presented. This is achieved by utilizing the result from the homoscedastic linear model, where $\theta$ is 0. The empirical process for this case is given by

$$S_{1N}(b) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi^+\left( R\left( \frac{|y_i - m_i|}{N+1} \right) \right) \text{sgn}(y_i - m_i)x_{1i}, \quad (3.4.1)$$

where $S_{1N}(b) = (S_{1N1}(b), \ldots, S_{1N(p+1)}(b))^T$, and $m_i = \frac{1}{N} x_{1i}^T b$.

Let

$$T_{1N}(b) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi^+\left( F^+\left( |y_i - m_i| \right) \right) \text{sgn}(y_i - m_i)x_{1i}, \quad (3.4.2)$$

be the approximation to the process $S_{1N}(b)$, for $T_N(b) = (T_{1N1}(b), \ldots, T_{1N(p+1)}(b))^T$.

Then, the standard linearity result for the approximation is given in the next theorem.

**Theorem 3.4.1.** Assume that (D1*), (D2*), (F1), (F2), (S1) and (W1) hold. In addition, suppose that $\beta = 0$ and $\theta = 0$.

$$\left\| T_{1N}(b) - T_{1N}(0) + \gamma_i \Sigma_1 b \right\|_{e-p_N} = o_p(1). \quad (3.4.3)$$

**Proof.** For any linear combination $X_1a = c$, given in Remark 3.3.1, it needs to be demonstrated that

$$\left\| a^TT_{1N}(b) - a^TT_{1N}(0) + \gamma_i a^T \Sigma_1 b \right\|_{e-p_N} = o_p(1). \quad (3.4.4)$$

Furthermore, for a suitably defined $a$ that selects any component, it suffices to obtain the linearity result for the chosen component. This is equivalent to demonstrating that

$$\text{Var} \left[ a^TT_{1N}(b) - a^TT_{1N}(0) \right] \to 0 \text{ as } N \to \infty, \quad (3.4.5)$$
due to the fact that \( E[a^T T_{1N}(0)]_{a \sim P_N} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i \int_0^1 \phi_1^+(v) dv = 0, \)
in light of assumption (S1). Observe that

\[
Var \left( a^T T_{1N}(b) - a^T T_{1N}(0) \right) \\
= \frac{1}{N} \sum_{i=1}^{N} c_i^2 \left( \phi_1^+ \left( F^+ \left( |y_i - m_i| \right) \right) sgn(y_i) - \phi_1^+ \left( F^+ \left( |y_i| \right) \right) sgn(y_i) \right). \tag{3.4.6}
\]

Since \( \frac{1}{N} \sum_{i=1}^{N} c_i^2 \) converges to a constant as \( N \) tends to be infinitely large, the proof rests on whether or not the second factor goes to zero as \( N \) becomes infinitely large. Thus,

\[
\frac{1}{N} \sum_{i=1}^{N} c_i^2 \left( \phi_1^+ \left( F^+ \left( |y_i - m_i| \right) \right) sgn(y_i - m_i) - \phi_1^+ \left( F^+ \left( |y_i| \right) \right) sgn(y_i) \right) \tag{3.4.7}
\]

\[
\leq \frac{1}{N} \sum_{i=1}^{N} c_i^2 \int \left( \phi_1^+ \left( F^+ \left( |y - m_i| \right) \right) sgn(y - m_i) - \phi_1^+ \left( F^+ \left( |y| \right) \right) sgn(y) \right)^2 dF(y) \to 0 \quad \text{as} \quad N \to \infty. \tag{3.4.8}
\]

The result was obtained by utilizing the continuity of \( F \), almost everywhere (a.e.), continuity of \( \phi_1(\cdot) \) and choosing an arbitrary \( \delta_1 > 0 \) so that for \( N \) sufficiently large

\[
\max_{1 \leq i \leq N} \{|m_i|\} < \delta_1. \tag{3.4.9}
\]

By this choice of \( \delta_1 \) along with any fixed \( \epsilon > 0 \), the expression in (3.4.8) was shown to be bounded. Hence, an application of the limit affords us the desired result for every component in \( (T_{1N1}(b), \ldots, T_{1N(p+1)}(b))^T \). Thus, in view of Remark 3.3.2, the convergence of the entire vector also holds, which terminates the proof for Theorem.

Next, under the heteroscedastic setting of this study, we consider a slightly more general linearity result for location problem where the errors are shifted with respect to location only.

**Theorem 3.4.2.** Assume that \((D1^*), (D2^*), (F1), (F2), (S1) and (W1) hold.**
Then
\[ \left\| T_{1N}(b,0,0) - T_{1N}(0,0,0) + \gamma_1 \Sigma_1^* b \right\|_{e^{-p_N}} = o_p(1). \] (3.4.10)

**Proof.** It follows from Remark 3.3.1 that it suffices to obtain linearity result for any selected component given by
\[ a^T T_{1N}(b,0,0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_1^+ \left( F^+ (|e_i - m_i^*|) \right) sgn(e_i - m_i^*) c_i^*. \] (3.4.11)

where \( X_i^* a = c^* \). On account of Corollary 1 and Theorem 3.3.1, it can easily be shown that
\[ E[a^T T_{1N}(0,0,0)]_{e^{-p_N}} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i^* \int_0^1 \phi_1^1 (v) dv = 0. \]
Then it suffices to show that
\[ Var \left[ a^T T_{1N}(b,0,0) - a^T T_{1N}(0,0,0) \right] \to 0 \] as \( N \to \infty. \)

Now
\[
Var \left[ a^T T_{1N}(b,0,0) - a^T T_{1N}(0,0,0) \right] \\
= \frac{1}{N} \sum_{i=1}^{N} c_i^{*2} Var \left[ \phi_1^+ \left( F^+ (|e_i - m_i^*|) \right) sgn(e_i - m_i^*) - \phi_1^+ \left( F^+ (|e_i|) \right) sgn(e_i) \right] \\
\leq \left( \frac{1}{N} \sum_{i=1}^{N} c_i^{*2} \right) E \left[ \phi_1^+ \left( F^+ (|e_i - m_i^*|) \right) sgn(e_i - m_i^*) - \phi_1^+ \left( F^+ (|e_i|) \right) sgn(e_i) \right]^2 \\
= \left( \frac{1}{N} \sum_{i=1}^{N} c_i^{*2} \right) \int \left[ \phi_1^+ \left( F^+ (|v - m_i^*|) \right) sgn(v - m_i^*) - \phi_1^+ \left( F^+ (|v|) \right) sgn(v) \right]^2 dF(v).
\]

Observe that \( \frac{1}{N} \sum_{i=1}^{N} c_i^{*2} \to a^T \Sigma_1^* a \), on account of assumption \( (D2^*) \). Thus, it only remains to prove that
\[
\limsup \left( \int \max_{1 \leq i \leq N} \left[ \phi_1^+ \left( F^+ (|v - m_i^*|) \right) sgn(v - m_i^*) - \phi_1^+ \left( F^+ (|v|) \right) sgn(v) \right]^2 dF(v) \right) = 0.
\]

Observe that the expression inside the square brackets is similar to that in square brackets.
of (3.4.8) in Theorem 3.4.2. Thus, with the same argument utilized in obtaining the result in Theorem 3.4.2, we can achieve the linearity condition for every component. Hence, in view of Remark 3.3.2, the convergence of the entire vector also holds. The proof is complete.

In the sequel, we extend linearity result of Theorem 3.4.2 to the empirical process $S_{1N}(b, 0, 0)$ by utilizing the result

$$
\left\| S_{1N}(0, 0, 0) - T_{1N}(0, 0, 0) \right\|_{e_{-pN}} = o_p(1),
$$

in light of Theorem 3.3.5.

Then an appeal to the contiguity of $q_{N}(b, 0)$ to $p_{N}$, along with an application of the translation properties given in Lemma 3.3.1 and Lemma 3.4.3 leads to the following consequence

$$
\left\| S_{1N}(b, 0, 0) - S_{1N}(0, 0, 0) + \gamma_{1} \Sigma_{i}^{*} b \right\|_{e_{-pN}} = o_p(1). \tag{3.4.12}
$$

It follows from (3.4.12) that the process $S_{1N}(b, 0, 0)$ can be approximated by linear function $S_{1N}(b, 0, 0) - \gamma_{1} \Sigma_{i}^{*} b$. It is standard to ascertain if the linearity result holds for all $b$ belonging to a bounded set. This result is given in the next theorem.

**Theorem 3.4.3.** Assume that $(D1^*)$, $(D2^*)$, $(F1)$, $(F2)$, $(S1)$ and $(W1)$ hold. Then

$$
\sup_{b \in B(\xi_{1})} \left\| S_{N}(b, 0, 0) - S_{1N}(0, 0, 0) + \gamma_{1} \Sigma_{i}^{*} b \right\|_{e_{-pN}} = o_p(1), \tag{3.4.13}
$$

where $B(\xi_{1}) = \{ b \in \mathcal{R}^{p+1} : \|b\| \leq \xi_{1} \}$ is the $(\xi_{1})$-ball centered at 0 for $\xi_{1} > 0$.

The proof of this theorem is obtained by invoking standard diagonal sequence
arguments which rely on the existence of an index set $\tilde{N}$ such that

$$S_{1N}(b, 0, 0) - \left(S_{1N}(0, 0, 0) - \gamma_1 \Sigma_1^* b\right) \to 0 \text{ almost surely},$$

for all $b \in \mathbb{R}^{p+1}$. Observe that on account of the same contiguity argument and the translation properties employed to link Theorem 3.4.2 and Theorem 3.4.3, it must also be true that

$$\sup_{b \in B(\xi_1)} \left\| T_{1N}(b, 0, 0) - T_{1N}(0, 0, 0) + \gamma_1 \Sigma_1^* b \right\|_{e^{-p_N}} = o_p(1). \quad (3.4.14)$$

This concludes the results on the process for the case where the errors are perturbed with respect to location only. A more typical situation is the case where the errors have been perturbed with respect to both location and scale. This main linearity result of this chapter is the focus of the next section.

3.4.1 Asymptotic Properties of $S_{1N}(b, t, s)$

Since in practice, the error variables are shifted with respect to both location and scale, it is useful to examine the local linearity of the empirical process $S_{1N}(b, t, s)$ defined in (3.2.6). Following the approach employed by Jurečková (1969), (1971), an approximation of $S_{1N}(b, t, s)$ is defined. It is demonstrated that the approximation satisfies an asymptotic linearity condition. Subsequently, utilizing a contiguity argument and a translation property to show that the empirical process and its approximation are asymptotically equivalent, affords us the linearity result of $S_{1N}(b, t, s)$. It turns out that the asymptotic behavior of $S_{1N}(b, t, s)$ can be obtained if it can be shown that the process is asymptotically equivalent to the statistic another process $S_{1N}(b, 0, s)$, defined in the sequel. Thus, we first consider the asymptotic linearity of the process $S_{1N}(b, 0, s)$. 
3.4.2 Asymptotic Linearity of $S_{1N}(b, 0, s)$

Define

$$S_{1N}(b, 0, s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i^+ \left( \frac{R[e_i - m_i^*]}{N+1} \right) \text{sgn}(e_i - m_i)x_{1i}^*, \quad (3.4.15)$$

where $S_{1N}(b, 0, s) = (S_{1N1}(b, 0, s), \ldots, S_{1N(p+1)}(b, 0, s))^T$. Then for this empirical process the following approximation is suggested.

$$T_{1N}(b, 0, s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i^+ \left( F^+([e_i - m_i^*]) \right) \text{sgn}([e_i - m_i^*])x_{1i}^*, \quad (3.4.16)$$

where $T_{1N}(b, 0, s) = (T_{1N1}(b, 0, s), \ldots, T_{1N(p+1)}(b, 0, s))^T$. The result in the next theorem will be useful in developing the asymptotic linearity we seek in this chapter.

**Theorem 3.4.4.** Assume that $(D1^*)$, $(D2^*)$, $(F1)$, $(F2)$, $(S1)$ and $(W1)$ hold.

Then

$$\left\| T_{1N}(b, 0, s) - T_{1N}(0, 0, s) + \gamma_1 \Sigma_1^* b \right\|_{\mathcal{L}^{\mathcal{P}N}} = o_p(1). \quad (3.4.17)$$

**Proof.** In view of the result in Theorem 3.4.2, it is sufficient to show that

$$\left\| T_{1N}(b, 0, s) - T_{1N}(b, 0, 0) \right\|_{\mathcal{L}^{\mathcal{P}N}} = o_p(1).$$

For any $a \in \mathcal{R}^{p+1}$, define linear combinations $a^T T_{1N}(b, 0, s)$ and $a^T T_{1N}(b, 0, 0)$, with $c_{**} = X_1^* a$ and $c_* = X_1^* a$, respectively.

It suffices to show that

$$\left\| a^T T_{1N}(b, 0, s) - a^T T_{1N}(b, 0, 0) \right\|_{\mathcal{L}^{\mathcal{P}N}} = o_p(1).$$
Observe that the left hand side of this representation is equivalent to

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_1^+ \left( F^+\left( |e_i - m_i^*| \right) \right) sgn(e_i - m_i^*) c_{i*}^{**} - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_1^+ \left( F^+\left( |e_i - m_i^*| \right) \right) sgn(e_i - m_i^*) c_{i*}^* \\
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_1^+ \left( F^+\left( |e_i - m_i^*| \right) \right) sgn(e_i - m_i^*) (c_{i*}^{**} - c_{i*}^*).
\]

In addition, note that \( c_i^{**} - c_i^* = a^T x_i^T \{ e^{\sqrt{\nu} x_i^T} - 1 \} \). Let \( \epsilon_0 > 0 \). It is seen that \( e^{\sqrt{\nu} x_i^T} < \epsilon_0 \) for sufficiently large \( N \); a fact established in the proof of Theorem 3.4.2. Then, using this fact, along with the contiguity of \( q_N(b, 0) \) to \( p_N \), it can be shown that \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_1^+ \left( F^+\left( |e_i - m_i^*| \right) \right) sgn(e_i - m_i^*) (c_{i*}^{**} - c_{i*}^*) \xrightarrow{p} n(0, 0) \), that is, the degenerate normal. Since convergence to the degenerate normal implies convergence in probability, the result holds componentwise. Further, in view of Remark 3.3.2, the result holds for the entire vector, which terminates the proof. \( \square \)

It follows from this discussion that the linearity above also holds for the empirical process, as the next theorem will show.

**Theorem 3.4.5.** Assume that \( (D1^*), (D2^*), (F1), (F2), (S1) \) and \( (W1) \) hold.

Then

\[
\left\| S_{1N}(b, 0, s) - S_{1N}(0, 0, s) + \gamma_1 \Sigma_1 b \right\|_{e \sim p_N} = o_p(1).
\]  

**Proof.** First, we show that

\[
\left\| S_{1N}(b, 0, s) - T_{1N}(b, 0, s) \right\|_{e \sim p_N} = o_p(1).
\]

Recall from the result in Theorem 3.3.2 that \( \left\| S_{1N}(0, 0, s) - T_{1N}(0, 0, s) \right\|_{e \sim p_N} = o_p(1) \). By
using the contiguity of \( q_N(b, 0) \) to \( p_N \), we obtain

\[
\left\| S_{1N}(0, 0, s) - T_{1N}(0, 0, s) \right\|_{e \sim p_N(b, 0)} = o_p(1). \tag{3.4.20}
\]

Then, an application of the translation properties,

\[
S_{1N}(b, 0, s) \left|_{e \sim p_N} \right. \equiv S_{1N}(0, 0, s) \left|_{e \sim q_N(b, 0)} \right.
\]

and

\[
T_{1N}(b, 0, s) \left|_{e \sim p_N} \right. \equiv T_{1N}(0, 0, s) \left|_{e \sim q_N(b, 0)} \right.
\]

to (3.4.17) leads us to the result in (3.4.19). An application of this result, together with the fact that \( \|S_{1N}(0, 0, s) - T_{1N}(0, 0, s)\|_{e \sim p_N} = o_p(1) \), due to Theorem 3.3.2 to

\[
\left\| T_{1N}(b, 0, s) - T_{1N}(0, 0, s) + \gamma_1 \Sigma^*_1 b \right\|_{e \sim p_N} = o_p(1),
\]

due to Theorem 3.4.4, affords us the desired result. Note that the last two translation properties hold on account of the results given in Lemma 3.3.1 and Lemma 3.4.3, since \( s \) is fixed and does not affect the distributive properties of the underlying variables \( z_i(b, 0) \) for \( i = 1, \ldots, N \). The proof is terminated.

Since this result has been completed, we can embark on the proof of the main linearity result of this chapter. This is provided in the next section.

3.4.3 Asymptotic Linearity of \( T_{1N}(b, t, s) \)

In this section, the asymptotic behavior of the approximation to the empirical process \( S_N(b, t, s) \) is examined. A first order expansion, yields functions linear in \( b \), which in turn are used to derive the limiting distribution of the proposed estimator of \( \beta \). Recall
Consider an approximation to the empirical process in (3.4.21). Define

\[ T_{1N}(b, t, s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_i^+ \left( \frac{R[|e_i - m^*_i|e^{-7^i}|]}{N + 1} \right) \text{sgn} \left( \frac{e_i - m^*_i}{e^{9_i}} \right) x_{li}^*. \]  

(3.4.22)

where \( T_N(b, t, s) = (T_{1N1}(b, t, s), \ldots, T_{1N(p+1)}(b, t, s))^T \). In the sequel, the asymptotic linearity result for the process \( T_{1N}(b, t, s) \) is obtained by using the same strategy as that employed to prove Theorem 3.4.2.

**Theorem 3.4.6.** Assume that \((D1^*), (D2^*), (F1), (F2), (S1) and (W1) hold. Then

\[ T_{1N}(b, t, s) - T_{1N}(0, 0, s) + \gamma_1 \Sigma^*_1 b \]  

(3.4.23)

\[ \rightarrow_p 0. \]

**Proof.** Consider the expression on the left hand side of (3.4.23). Adding and subtracting \( T_{1N}(b, 0, s) \) along with an application of the triangular inequality gives

\[
\left\| T_{1N}(b, t, s) - T_{1N}(0, 0, s) + \gamma_1 \Sigma^*_1 b \right\|_{e^P} = o_p(1).
\]

(3.4.23)

Observe that the second term on the right hand side of the last inequality, \( \| T_{1N}(b, 0, s) - T_{1N}(0, 0, s) + \gamma_1 \Sigma^*_1 b \|_{e^P} \) goes to zero in probability in light of Theorem 3.4.4. Thus, the result holds if it can be shown that the first term goes to zero in probability. In view of Remark 3.3.1, we apply a linear combination that selects any component of the vector in what follows. Thus, for \( a \in \mathbb{R}^{p+1} \) consider linear combinations \( a^T T_{1N}(b, 0, s) \) and \( a^T T_{1N}(b, t, s) \), where \( X_{li}^* a = c_i^* \). It was already demonstrated in proof of Theorem
3.4.4 that \(|a^T T_{1,N}(b, 0, s) - a^T T_{1,N}(b, 0, 0)|\) goes zero in probability. This fact together with the result \(E[a^T T_{1,N}(b, 0, 0)]\) goes zero as \(N \to \infty\), which was established in the proof of Theorem 3.3.8 implies that \(E[a^T T_{1,N}(b, 0, s)]\) goes zero as \(N \to \infty\).

For the moment, suppose that \(E[a^T T_{1,N}(b, t, s)]\) goes zero as \(N \to \infty\). Then it suffices to show that the second moment of the difference between the two statistics goes to zero. That is,

\[
E \left[ a^T T_{1,N}(b, t, s) - a^T T_{1,N}(b, 0, s) \right]_{e \sim \mathcal{P}_N} \to 0, \text{ as } N \to \infty.
\]

Note here that the strategy that was employed to obtain a similar result in Theorem 3.4.2 cannot be applied directly, because we have additional terms, \(s, t\), that have to be accounted for. The left hand side of the last convergence result can be reexpressed as

\[
\var{a^T T_{1,N}(b, t, s) - a^T T_{1,N}(b, 0, s)}_{e \sim \mathcal{P}_N} \to 0 \text{ as } N \to \infty.
\]
this representation that the desired result can be obtained if it can be shown that

\[
\limsup \left( \int \left[ \phi_1^+ \left( F^+ |v - m_i^*|e^{-g_i} \right) \sgn((v - m_i^*)e^{-g_i}) \right. \\
- \phi_1^+ \left( F^+ |v - m_i^*| \right) \sgn(v - m_i^*) \right]^2 \, dF(v) \right) = 0.
\]

Let \( \varepsilon > 0 \) be given. Since \( \phi_1^+ (\cdot) \) is continuous almost everywhere, it can be assumed that it is continuous at \( F^+ |v - m_i^*| \). There exists a \( \delta_1 > 0 \) such that

\[
\left| \phi_1^+ \left( w_i \right) - \phi_1^+ \left( F^+ |v - m_i^*| \right) \right| < \varepsilon \text{ for } |w_i - F^+ |v - m_i^*|| < \delta_1.
\]

Further, by the uniform continuity of \( F_i^+ (|v - m_i^*|) \), choose \( \delta_2 > 0 \) such that

\[
\left| F^+ \left( |u - m_i^*| \right) - F^+ \left( |v - m_i^*| \right) \right| < \delta_1 \text{ for } |u - m_i^*| - |v - m_i^*| < \delta_2.
\]

Furthermore, by the uniform continuity of \( |v| \), choose \( \delta_3 > 0 \) such that

\[
\left| F^+ \left( |u - m_i^*| \right) - F^+ \left( |v - m_i^*| \right) \right| < \delta_1 \text{ for } |u - m_i^*| - |v - m_i^*| < \delta_2,
\]

where \( |u - v| < \delta_3 \).

Observe that letting \( |u - v| = |(e_i - m_i^*)(e^{-g_i} - 1)| = |e_i(e^{-g_i} - 1) - m_i^*(e^{-g_i} - 1)| \) for \( i = 1, \ldots, N \). Since both terms within the \( | \cdot | \) contain \((e^{-g_i} - 1)\), it suffices to consider the second term. Let \( |m_i^*| < \delta_4 \) for \( \delta_4 > 0 \). Recall from the proof of Theorem 3.3.4 that for any \( \delta_5 > 0 \), \( |e^{-g_i} - 1| < \delta_5 \) for sufficiently large \( N \). Note that since \( \delta_4 \) and \( \delta_5 \) are fixed, they can always be selected to be such that \( \delta_4 \cdot \delta_5 = \delta_3 \). There exists an \( N_0 \) such that \( N > N_0 \) implies that

\[
\max_{1 \leq i \leq N} \{ |m_i^*| \} < \delta_4 \text{ and } \max_{1 \leq i \leq N} \{ |e^{-g_i} - 1| \} < \delta_5
\]
Then, if we select $N > N_0$, it follows that

$$\left| (e_i - m_i^*)(e^{-g_i} - 1) \right| < \delta_3 \text{ for } i = 1, \ldots, N.$$ 

Further, for $N > N_0$, it follows that

$$\left| e_i - m_i^* e^{-g_i} - |e_i - m_i^*| \right| < \delta_2, \text{ for } i = 1, \ldots, N.$$ 

Furthermore, for $N > N_0$, it follows that

$$\left| F^+(|e_i - m_i^* e^{-g_i}|) - F^+(|e_i - m_i^*|) \right| < \delta_1 \text{ for } i = 1, \ldots, N.$$ 

Then, for $N > N_0$, it follows that

$$\left| \phi_1^+ \left( F^+(|e_i - m_i^* e^{-g_i}|) \right) - \phi_1^+ \left( F^+(|e_i - m_i^*|) \right) \right| < \varepsilon \text{ for } i = 1, \ldots, N.$$ 

Hence, for $N > N_0$, it is clear that

$$\max_{1 \leq i \leq N} \left[ \phi_1^+ \left( F^+(|e_i - m_i^* e^{-g_i}|) \right) - \phi_1^+ \left( F^+(|e_i - m_i^*|) \right) \right]^2 < \varepsilon^2.$$ 

Therefore,

$$\limsup \left( \int \max_{1 \leq i \leq N} \left[ \phi_1^+ \left( F^+(|v - m_i^* e^{-g_i}|) \right) - \phi_1^+ \left( F^+(|v - m_i^*|) \right) \right]^2 \, dF(v) \right) \leq \varepsilon^2.$$ 

Since $\varepsilon$ is fixed, it can be chosen to be arbitrarily small in order that this bound holds for any chosen component. Thus, it also holds for every component of the vector of the difference $T_{1N}(b, t, s) - T_{1N}(b, 0, s)$. Therefore, the convergence is achieved for the entire vector in view of Remark 3.3.2, which completes the proof.  

In the sequel, it is shown that the linearity result above holds for the empirical pro-
cess $S_{1N}(b, t, s)$. Since the underlying variables of this statistic are $z_1(b, t), \ldots, z_N(b, t)$, the distributive properties of these variables are considered in what follows.

**Lemma 3.4.1.** Assume that the errors $e_1, \ldots, e_N$ are jointly distributed as $p_N$. Then, for $i = 1, \ldots, N$, the random variables $z_i(b, t)$ are distributed as

\[ G_{b,t}(v) = P(z_i(b, t) \leq v) = P(e_i \leq ve^{b_i} + m_i^*) = F(ve^{b_i} + m_i^*) \]

Thus, from this lemma, we can ascertain the joint distribution of the random variables $e_1, \ldots, e_N$, assuming that each $e_i$ has the cdf $F(e^{b_i}e + m_i^*)$. Let this joint distribution be defined as

\[ q_N(b, t) = \prod_{i=1}^{N} f(e^{b_i}e_i + m_i^*)e^{b_i}. \quad (3.4.24) \]

The proof of the following result is deferred to chapter four. We state the contiguity condition.

**Remark 3.4.1.** The sequence of densities $q_N(b, t)$ defined in (3.4.24) is contiguous to the sequence $p_N$.

This contiguity condition is used to establish the asymptotic linearity of the process $S_{1N}(b, t, s)$. In addition, the final step in obtaining the desired result, involves drawing on the relationships between $T_{1N}(b, t, s)$ and $T_{1N}(0, 0, 0)$ as well as $S_{1N}(b, t, s)$ and $S_{1N}(0, 0, 0)$. A distributive relationship between the underlying variables is required before translation properties are furnished.
Lemma 3.4.2. For any fixed $b$ and $t$

\[
    z_i(b, t) \bigg|_{e_i \sim F(v)} \overset{d}{=} z_i(0, 0) \bigg|_{e_i \sim F(e^{a^i}v + m^*_i)}, \text{ for } i = 1, \ldots, N. \quad (3.4.25)
\]

Proof. Since under $p_N$, $z_i(b, t)$ has the cdf $G_{b,t}(v)$ in view of Lemma 3.4.1, it suffices to show that

\[
    z_i(0, 0) \bigg|_{e_i \sim F(e^{a^i}v + m^*_i)} \text{ has the cdf } G_{b,t}(v) \text{ for } i = 1, \ldots, N.
\]

Observe that, provided each $e_i$ is drawn from the distribution $F(e^{a^i}v + m^*_i)$,

\[
    P(z_i(0, 0) \leq v) = P(e_i \leq v) = F(e^{a^i}v + m^*_i) = G_{b,t}(v).
\]

This terminates the proof.

Then the next translation property follows directly from this result.

Lemma 3.4.3. For any fixed $b$ and $t$,

\[
    T_{1N}(b, t, s) \bigg|_{e \sim p_N} \overset{d}{=} T_{1N}(0, 0, s) \bigg|_{e \sim q_N(b, t)}, \quad (3.4.26)
\]

Proof. The left hand side of (3.4.26) is a function of the random variables $z_i(b, t)$, whereas the right hand side is a function of random variables $z_i(0, 0)$, for $i = 1, \ldots, N$. Since $z_i(b, t) \big|_{e_i \sim F(v)} \overset{d}{=} z_i(0, 0) \big|_{e_i \sim F(e^{a^i}v + m^*_i)}$, in view of Lemma 3.4.2, along with the fact that the distribution properties of the random variables extend to their corresponding statistics, the desired result is easily afforded. The proof is complete.

In the next lemma, it is shown that an analogous translation property also holds
for the empirical processes $S_{1N}(b, t, s)$ and $S_{1N}(0, 0, s)$.

**Lemma 3.4.4.** For any fixed $b$ and $t$,

$$S_{1N}(b, t, s)_{e~p_N} \not \equiv S_{1N}(0, 0, s)_{e~q_N(b, t)}, \quad (3.4.27)$$

**Proof.** Similar to Lemma 3.4.3, the result holds in view of the fact $S_{1N}(b, t, s)_{e~p_N}$ and $S_{1N}(0, 0, s)_{e~q_N(b, t)}$ are functions of random variables $z_i(b, t)$ and $z_i(0, 0)$, respectively. Then the result follows from the relationship between these random variables in light of Lemma 3.4.2. The proof is terminated.

Recall that it had been assumed that $E[S_{1N}(b, t, s)_{e~p_N}] \rightarrow -\gamma_1 \Sigma_1 b$ as $N \rightarrow \infty$. However, this has yet to be verified. Consider the proof of the analogous result $E[T_{1N}(b, t, s)_{e~p_N}] \rightarrow -\gamma_1 \Sigma_1 b$ as $N \rightarrow \infty$. This is done in the next theorem.

**Theorem 3.4.7.** Assume that $(D1^*)$, $(D2^*)$, $(F1)$, $(F2)$, $(S1)$ and $(W1)$ hold. Then

$$E[T_{1N}(b, t, s)_{e~p_N}] \rightarrow -\gamma_1 \Sigma_1 b \quad \text{as } N \rightarrow \infty. \quad (3.4.28)$$

**Proof.** For $a \in \mathcal{R}^{p+1}$, define an arbitrary linear combination $a^T T_{1N}(b, t, s)$, where $c^* = X_i^* a$ as described in Remark 3.3.1. It suffices to show

$$E[a^T T_{1N}(b, t, s)_{e~p_N}] \rightarrow -\gamma_1 a^T \Sigma_1^* b \quad \text{as } N \rightarrow \infty. \quad (3.4.29)$$
Observe that

\[
E\left[ a^T T_{1N}(b, t, s) \right]_{e^\sim_{PN}} = E\left[ a^T T_{1N}(b, t, s) - a^T T_{1N}(b, 0, s) + a^T T_{1N}(b, 0, s) - a^T T_{1N}(0, 0, 0) \right]_{e^\sim_{PN}} \\
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon^* E\left[ \phi_1^+\left( F^+\left(|e_i - m| e^{-g}\right) \right) - \phi_1^+\left( F^+\left(|e_i - m| e^{-g}\right) \right) \right] - \gamma_1 a^T \Sigma_1^* b + o_p(1),
\]

where the result \( E[a^T T_{1N}(b, 0, s)]_{e^\sim_{PN}} \rightarrow -\gamma_1 a^T \Sigma_1^* b + o_p(1) \) as \( N \rightarrow \infty \) follows from the fact that \( \|T_{1N}(b, 0, s) - T_{1N}(b, 0, 0)\|_{e^\sim_{PN}} = o_p(1) \), in light of Theorem 3.4.4, and \( E[a^T T_{1N}(b, 0, 0)]_{e^\sim_{PN}} \rightarrow -\gamma_1 a^T \Sigma_1^* b, \) on account of Theorem 3.3.8. Thus, it suffices to show that \( E[a^T T_{1N}(b, t, s) - a^T T_{1N}(b, 0, s)]_{e^\sim_{PN}} \) goes to zero as \( N \) becomes infinitely large. It is seen that

\[
E\left[ a^T T_{1N}(b, t, s) - a^T T_{1N}(b, 0, s) \right]_{e^\sim_{PN}} \\
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_{i}^* \int_0^1 \left[ \phi_1^+\left( F^+\left(|e_i - m_i^\ast| e^{-g_i}\right) \right) - \phi_1^+\left( F^+\left(|e_i - m_i^\ast| e^{-g_i}\right) \right) \right] dF^+(|v|) \\
= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_{i}^* \int_0^1 \left[ \phi_1^+\left( F^+(|v|) \right) f'(v) \left( (v - m_i^\ast)(e^{-g_i} - 1) \right) \right] dF^+(|v|),
\]

where the second equality is a consequence of an application of the mean value theorem. Observe that \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_{i}^* \) converges to a constant, \( \phi_1^+/\cdot) \) and \( f'(\cdot) \) are both bounded. Finally, from the proof of Theorem 3.4.4, recall that for any \( \epsilon_0 > 0, \|e^{-g_i} - 1\| \leq \epsilon_0 \) for \( N \) sufficiently large. In addition, it was shown above that \( \|v - m_i^\ast\| (e^{-g_i} - 1) \| < \delta_3 \), for \( \delta_3 > 0 \). Then it is clear that \( E[a^T T_{1N}(b, t, s) - a^T T_{1N}(b, 0, s)]_{e^\sim_{PN}} \) goes to zero as \( N \) becomes infinitely large. Hence, that desired result in (3.4.29) holds. Therefore, in view of Remark 3.3.2, the expectation of the vector, \( T_{1N}(b, t, s) \), follows from this fact, and this terminates the proof.

In the sequel, an analogous result for the empirical process \( S_{1N}(b, t, s) \) is consid-
Theorem 3.4.8. Assume that \((\text{D}1^*), (\text{D}2^*), (\text{F}1), (\text{F}2), (\text{S}1)\) and \((\text{W}1)\) hold. Then

\[
E \left[ S_{1N}(b, t, s) \right] \bigg|_{e \sim p_N} \to -\gamma_1 \Sigma_1^* b \quad \text{as } N \to \infty. \quad (3.4.30)
\]

Proof. Recall that from Theorem 3.3.2 that

\[
\left\| S_{1N}(0, 0, s) - T_{1N}(0, 0, s) \right\|_{e \sim p_N} = o_p(1), \quad (3.4.31)
\]

By the contiguity of \(q_N(b, t)\) to \(p_N\), it follows that

\[
\left\| S_{1N}(0, 0, s) - T_{1N}(0, 0, s) \right\|_{e \sim q_N(b, t)} = o_p(1). \quad (3.4.32)
\]

In the left hand side of the last equality, consider \(T_{1N}(0, 0, s) \mid_{e \sim q_N(b, t)} \). From Lemma 3.4.3, we have \(T_{1N}(0, 0, s) \mid_{e \sim q_N(b, t)} \overset{\text{def}}{=} T_{1N}(b, t, s) \mid_{e \sim p_N}\). Thus, it follows that

\[
E \left[ T_{1N}(0, 0, s) \right] \mid_{e \sim q_N(b, t)} = E \left[ T_{1N}(b, t, s) \right] \mid_{e \sim p_N}
\]

\[
\quad \to -\gamma_1 \Sigma_1^* b, \quad \text{as } N \to \infty,
\]

where the second equality is due to the result in Theorem 3.4.7. In view of this result, together with the fact that convergence in probability implies convergence in distribution, it follows from (3.4.32) that

\[
E \left[ S_{1N}(0, 0, s) \right] \mid_{e \sim q_N(b, t)} \to -\gamma_1 \Sigma_1^* b \quad \text{as } N \to \infty. \quad (3.4.33)
\]

Then the desired result follows on account of the translation property

\(S_{1N}(0, 0, s) \mid_{e \sim q_N(b, t)} \overset{\text{def}}{=} S_{1N}(b, t, s) \mid_{e \sim p_N}\) given in Lemma 3.4.4 to terminate the proof. \(\square\)
Thus, all the pieces required to derive the asymptotic linearity that is extended to the empirical process \( S_{1N}(b, t, s) \) have been established. Obtaining the linearity result is done in next section.

3.4.4 Asymptotic Linearity of \( S_{1N}(b, t, s) \)

In this section, using contiguity of \( q_N(b, t) \) to \( p_N \) along with the translation properties given above, we establish that the empirical process \( S_{1N}(b, t, s) \) satisfies a linearity result condition. The main result is given in the theorem that follows.

**Theorem 3.4.9.** Assume that \((D1^*), (D2^*), (F1), (F2), (S1) and (W1) hold. Then

\[
S_{1N}(b, t, s) - S_{1N}(0, 0, s) + \gamma_1 \Sigma_1^* b \overset{d}{\rightarrow} 0. \tag{3.4.34}
\]

**Proof.** Recall from Theorem 3.4.6 that

\[
\left\| T_{1N}(b, t, s) - T_{1N}(0, 0, s) + \gamma_1 \Sigma_1^* b \right\|_{e \sim p_N} = o_p(1). \tag{3.4.35}
\]

Then, in view of Theorem 3.3.2, observe that

\[
\left\| S_{1N}(0, 0, s) - T_{1N}(0, 0, s) \right\|_{e \sim p_N} = o_p(1).
\]

Thus, by the contiguity of the densities \( q_N(b, t) \) to densities \( p_N \), it follows that

\[
\left\| S_{1N}(0, 0, s) - T_{1N}(0, 0, s) \right\|_{e \sim q_N(b, t)} = o_p(1).
\]

Consider the left hand side of the last representation. An application of the translation property \( S_{1N}(b, t, s) \mid_{e \sim p_N} \nless S_{1N}(0, 0, s) \mid_{e \sim q_N(b, t)} \) given in Lemma 3.4.4 and
\( T_{1N}(b, t, s) \mid e^{-p_N} \leq T_{1N}(0, 0, s) \mid e^{-q_N(b,t)} \) given in Lemma 3.4.3 yields
\[
\left\| S_{1N}(b, t, s) - T_{1N}(b, t, s) \right\|_{e^{-p_N}} = o_p(1).
\]

Then, taking the left hand side of (3.4.35), to which this result along with the fact
\( \| S_{1N}(0, 0, s) - T_{1N}(0, 0, s) \|_{e^{-p_N}} \) goes to zero in probability due to Theorem 3.3.2, are applied, leads to the desired result in (3.4.34). The proof is complete. □

It is clear from the previous theorem that the linear function \( S_{1N}(0, 0, s) - \gamma_1 \Sigma_i^* b \) is asymptotically equivalent to the desired empirical process \( S_{1N}(b, t, s) \). Hence, the linear function can be used to approximate the process.

### 3.4.5 Asymptotic Uniform Linearity

In this section, we extend the asymptotic uniform linearity of the empirical process \( S_{1N}(b, 0, 0) \) established in Theorem 3.4.3 to the process \( S_{1N}(b, t, s) \). This result is given in the next theorem.

**Theorem 3.4.10.** Assume that \((D1^*), (D2^*), (F1), (F2), (S1) and (W1) hold.

Then
\[
\sup_{b \in \mathcal{B}(\xi_1)} \sup_{t \in \mathcal{K}(\xi_2)} \sup_{s \in \mathcal{K}(\xi_3)} \left\| S_{1N}(b, t, s) - S_{1N}(0, 0, s) + \gamma_1 \Sigma_i^* b \right\|_{e^{-p_N}} = o_p(1). \tag{3.4.36}
\]

**Proof.** This is equivalent to proving that
\[
\sup_{b \in \mathcal{B}(\xi_1)} \sup_{t \in \mathcal{K}(\xi_2)} \sup_{s \in \mathcal{K}(\xi_3)} \left\| S_{1N}(b, t, s) - S_{1N}(b, 0, s) \right\|_{e^{-p_N}} = o_p(1). \tag{3.4.37}
\]

Recall from Theorem 3.4.3 that
\[
\sup_{b \in \mathcal{B}(\xi_1)} \left\| S_{1N}(b, 0, 0) - S_{1N}(0, 0, 0) + \gamma_1 \Sigma_i^* b \right\|_{e^{-p_N}} = o_p(1).
\]
If we select \( b \) to be any fixed vector in the bounded set \( B(\xi_1) \), it suffices to demonstrate that

\[
\sup_{t \in \mathcal{K}(\xi_2)} \sup_{s \in \mathcal{K}(\xi_3)} \left\| S_{1N}(b, t, s) - S_{1N}(b, 0, s) \right\|_{o_p(N)} = o_p(1). \tag{3.4.38}
\]

In view of Remark 3.3.1, define a linear combination such that an arbitrary component of the difference on the left hand side of this equality is selected. That is, it is enough to show that

\[
\sup_{t \in \mathcal{K}(\xi_2)} \left\| a^T S_{1N}(b, t, s) - a^T S_{1N}(b, 0, s) \right\|_{o_p(N)} = o_p(1). \tag{3.4.39}
\]

Recall that \( B(\xi_1) = \{ b : \| b \| \leq \xi_1 \} \). Following Tardif (1985), Heiler and Willers (1988), utilizing standard diagonal sequence arguments, it is possible to obtain an index set, \( \mathcal{N} \), such that

\[
a^T S_{1N}(b, t, s) - a^T S_{1N}(0, 0, s) \to 0 \text{ almost surely,}
\]

for all \( t \in \mathcal{R}^p \) and \( s \in \mathcal{R}^p \). Note that for convenience, the special case \( t = s \) can always be selected. It follows that this convergence is uniform on every compact subset of \( \mathcal{R}^{p+1} \). Further, the uniform convergence also holds for any bounded subsets on \( \mathcal{K}(\xi_2) \) and \( \mathcal{K}(\xi_3) \). Similar to Heiler and Willers, (1988), p. 179), this compact uniform convergence is equivalent to the result in (3.4.36). Since the result holds for every component of \( S_{1N}(b, t, s) - S_{1N}(b, 0, s) \), the convergence of the entire vector is immediate on account of Remark 3.3.2. The proof is terminated. \( \square \)

This concludes the series of theoretical results that are required before establishing the asymptotic behavior of the proposed estimator in this chapter. In the sequel, we apply the previous results in order to obtain the asymptotic distribution of the estimator.
3.4.6 Application of Linearity Result

In this section, the asymptotic linearity results established above are utilized to determine the asymptotic distribution of the proposed estimator of $\beta$.

In addition to the conditions set by assumptions (D1*), (D2*), (F1), (F2), (S1) and (W1), we assume that

(B1) we have an estimate of $\theta$ such that $\sqrt{N}(\hat{\theta} - \theta) = O_p(1)$.

Suppose we have a $\sqrt{N}$-consistent estimate of the scale parameter, $\theta$, satisfying (B1). Then, a weighted estimate of the regression parameter, $\hat{\beta}$ can be obtained by minimizing the signed-rank dispersion function that has $\hat{\theta}$ substituted for $\theta$. In what follows, it is demonstrated that such an estimate of $\beta$ is asymptotically equivalent to the estimate obtained by solving $S_N(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\theta} - \theta)) = 0$. In addition, the estimate is shown to be $\sqrt{N}$-consistent for $\beta$.

Firstly, it is demonstrated that when estimated scale constants are employed, the asymptotic properties established above are still valid. In order to show this result, the general strategy that is pursued, in the sequel, is similar to that used to establish the result in Lemma A.3.12 of Hettmansperger and McKean (1998). The authors considered a special case of the unweighted linear model problem previously investigated by Jurečková (1971). In the results from both of these citations, consistent, unweighted residuals were used to estimate the intercept parameter. In contrast, the current application explores the validity of the asymptotic linearity results for the location problem when an estimate satisfying condition (B1) has been substituted for the scale parameter. It is seen that the standard signed-rank process for the location problem, cannot be a suitable tool for ascertaining this result. The next theorem furnishes an analogous result for the current weighted signed-rank process for the location problem.

**Theorem 3.4.11.** Assume that (D1*), (D2*), (F1), (F2), (S1) and (W1) hold. In addition, suppose
that assumption (B1) is satisfied.

Then, for any $b \in \mathcal{R}^{p+1}$,

$$
\|S_{1N}(b, \sqrt{N}(\theta - \theta), \sqrt{N}(\theta - \theta)) - S_{1N}(b, 0, 0)\|_{e \sim p_N} = o_p(1).
$$

Proof. Let $\varepsilon > 0$ be given. From Theorem 3.4.9 we have the result $\|S_{1N}(b, t, s) - S_{1N}(0, 0, s) + \gamma_1 \Sigma^*_1 b\|_{e \sim p_N} = o_p(1)$. When $t = s = 0$, the result holds in view of Theorem 3.4.2. For any $\xi_2 > 0, \xi_3 > 0$ and $\varepsilon > 0$ observe that

$$
\lim_{N \to \infty} P\left( \max_{t \in T(\xi_2)} \max_{s \in S(\xi_3)} \|S_{1N}(b, t, s) - S_{1N}(b, 0, 0)\|_{e \sim p_N} \geq \varepsilon \right) = 0,
$$

for all $\varepsilon > 0$. Consider the expression in $P(\cdot)$. Observe that adding and subtracting $S_{1N}(0, 0, s) - \gamma_1 \Sigma^*_1 b$ and $S_{1N}(0, 0, 0)$ to the expression within the $\| \cdot \|$, together with an application of the triangular inequality yields

$$
\|S_{1N}(b, t, s) - S_{1N}(b, 0, 0)\| \leq \|S_{1N}(b, t, s) - S_{1N}(0, 0, s) + \gamma_1 \Sigma^*_1 b\| + \|S_{1N}(0, 0, s) - S_{1N}(0, 0, 0)\|.
$$

Then

$$
P\left( \max_{t \in T(\xi_2)} \max_{s \in S(\xi_3)} \|S_{1N}(b, t, s) - S_{1N}(b, 0, 0)\|_{e \sim p_N} \geq \varepsilon \right) \leq P\left( \max_{t \in T(\xi_2)} \max_{s \in S(\xi_3)} \|S_{1N}(b, t, s) - S_{1N}(0, 0, s) + \gamma_1 \Sigma^*_1 b\|_{e \sim p_N} \geq \varepsilon / 3 \right)
$$

$$
+ P(\|S_{1N}(b, 0, 0) - S_{1N}(0, 0, 0) + \gamma_1 \Sigma^*_1 b\|_{e \sim p_N} \geq \varepsilon / 3) + P(\|S_{1N}(0, 0, s) - S_{1N}(0, 0, 0)\|_{e \sim p_N} \geq \varepsilon / 3).
$$
Let $||t|| \leq \xi_2$ and $||s|| \leq \xi_3$ for any $\xi_2 > 0$ and $\xi_3 > 0$. Since $\varepsilon > 0$ is given, it can be selected to be arbitrarily small. Consider the right hand side of the last inequality. The first term holds on account of Theorem 3.4.10. The second term holds in view of the fact $||S_{1N}(b, t, s) - S_{1N}(0, 0, s) + \gamma_1 \Sigma^*_1 b||_{o_p} = o_p(1)$ from Theorem 3.4.2. The third term holds since $||S_{1N}(0, 0, s) - S_{1N}(0, 0, 0)||_{o_p} = o_p(1)$ in light of Theorem 3.3.4. Thus, for $N$ sufficiently large, the three terms on the right hand side of the last inequality are arbitrarily small. Since $\sqrt{N}(\hat{\theta} - \theta)$ is bounded in probability, on account of (B1), the desired result follows from this to complete the proof. \[\Box\]

Next we apply this result to show that the uniform asymptotic linearity in $b$ still holds when the scale parameter is replaced by its estimate, $\hat{\theta}$.

**Theorem 3.4.12.** Assume that $(D1^*)$, $(D2^*)$, $(F1)$, $(F2)$, $(S1)$ and $(W1)$ hold. In addition, suppose that assumption $(B1)$ is satisfied.

Then, for any $b \in \mathcal{R}^{p+1}$,

$$\sup_{b \in B(\xi_1)} \left\| S_{1N}(b, \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\theta} - \theta)) - S_{1N}(0, 0, \sqrt{N}(\hat{\theta} - \theta)) + \gamma_1 \Sigma^*_1 b \right\|_{o_p} = o_p(1). \quad (3.4.40)$$

**Proof.** Observe that

$$\begin{align*}
\left\| S_{1N}(b, \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\theta} - \theta)) - S_{1N}(0, 0, \sqrt{N}(\hat{\theta} - \theta)) + \gamma_1 \Sigma^*_1 b \right\| \\
\leq \left\| \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\theta} - \theta)) - S_{1N}(0, 0, 0) \right\| \\
+ \left\| S_{1N}(b, 0, 0) - S_{1N}(0, 0, 0) + \gamma_1 \Sigma^*_1 b \right\| + \left\| S_{1N}(0, 0, 0) - S_{1N}(0, 0, \sqrt{N}(\hat{\theta} - \theta)) \right\|.
\end{align*}$$


by an application of the triangular inequality. Thus, for any $\varepsilon > 0$, write

\[
P\left( \max_{b \in B(\xi_1)} \left\| S_{1N}(b, \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\theta} - \theta)) - S_{1N}(0, 0, \sqrt{N}(\hat{\theta} - \theta)) + \gamma_1 \Sigma^*_b \right\|_{e^{-p_N}} \geq \varepsilon \right) 
\leq P\left( \max_{b \in B(\xi_1)} \left\| S_{1N}(b, \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\theta} - \theta)) - S_{1N}(b, 0, 0) \right\|_{e^{-p_N}} \geq \varepsilon/3 \right) 
+ P\left( \max_{b \in B(\xi_1)} \left\| S_{1N}(b, \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\theta} - \theta)) - S_{1N}(b, 0, 0) + \gamma_1 \Sigma^*_b \right\|_{e^{-p_N}} \geq \varepsilon/3 \right) 
+ P\left( \left\| S_{1N}(b, 0, 0) - S_{1N}(0, 0, \sqrt{N}(\hat{\theta} - \theta)) \right\|_{e^{-p_N}} \geq \varepsilon/3 \right).
\]

Consider the right hand side of this last inequality. Since $\varepsilon > 0$ is fixed, it can be selected to be arbitrarily small. The first and third terms hold in view of Theorem 3.4.11 and the second term hold on account of Theorem 3.4.3. Thus, for $N$ sufficiently large, the three terms on the right hand side of the last inequality are arbitrarily small. Letting $t^* = s^* = \sqrt{N}(\hat{\theta} - \theta)$ for $\hat{\theta}$ satisfying condition (B1), and substituting them for $t$ and $s$, yields us the desired result. \hfill \Box

Now under a certain condition on the stochastic nature of the bound on $b^*$ shown below, we can let $b^* = \gamma_1^{-1} \Sigma^{-1} S_{1N}(0, 0, \sqrt{N}(\hat{\theta} - \theta))$ and substitute it for $b$ in the result Theorem 3.4.12 so that

\[
\left\| S_{1N}(b^*, \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\theta} - \theta)) \right\|_{e^{-p_N}} = o_p(1)
\]

is immediate. This result would suggest that the estimate of $\beta$, $\hat{\beta}$, which we seek can be obtained by solving

\[
S_{1N}(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\theta} - \theta)) = 0.
\]

(3.4.41)
Observe that the empirical process

\[ S_{1N}(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\theta} - \theta)) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_{y}^{+} \left( \frac{N}{N+1} F_{N}^{+} \left( |\tilde{\sigma}_{i}^{-1}(y_{i} - x_{i}^{T}\hat{\beta})| \right) \right)(\tilde{\sigma}_{i}^{-1}x_{i}), \]

where \( F_{N}(|\tilde{\sigma}_{i}^{-1}(y_{i} - x_{i}^{T}\hat{\beta})|) \) is an empirical distribution of the estimated absolute residuals

\[ z_{i}(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N}(\hat{\theta} - \theta)) = |\tilde{\sigma}_{i}^{-1}(y_{i} - x_{i}^{T}\hat{\beta})|, \text{ where } \tilde{\sigma}_{i}^{-1} = e^{-x_{i}^{T}\theta}, \quad i = 1, \ldots, N. \]

If the result in (3.4.41) holds then, by the asymptotic equivalence of this solution to that of the minimization of the dispersion function \( D_{1}(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N}(\hat{\theta} - \theta)) \), to be given later, implies that we have an alternative method for estimating \( \beta \) with \( \theta \) replaced by \( \hat{\theta} \) that satisfies condition (B1). Several results have yet to be proved. This is done in what follows.

It is verified that \( \|S_{1N}(b, t, s)\|_{e-p_{N}} \) goes to zero in probability. Recall that \( S_{1N}(b, t, s) \) can be approximated by the linear function \( S_{1N}(0, 0, s) - \gamma_{l}\Sigma_{i}'b \). Thus, we first prove that \( S_{1N}(0, 0, s) \) is bounded in probability and this is done in the next lemma.

**Lemma 3.4.5.** Assume that \((D1^*), (D2^*), (F1), (F2), (S1)\) and \((W1)\) hold. Then

\[ \|S_{1N}(0, 0, s)\|_{e-p_{N}} = O_{p}(1). \quad (3.4.42) \]

**Proof.** It was proved in Theorem 3.3.3 that \( S_{1N}(0, 0, s)_{e-p_{N}} \overset{\Delta}{\rightarrow} N_{p+1}(0, \Sigma_{i}) \). This result, along with the fact that convergence in distribution implies that the variable under consideration is bounded in probability, establishes the result being sought in (3.4.42). The proof is complete. \( \square \)

**Corollary 2.** Assume that \((D1^*), (D2^*), (F1), (F2), (S1)\) and \((W1)\) hold. In addition, suppose
(B1) is satisfied.

Then

\[ \left\| S_{1N}(\sqrt{N}(\beta - \beta), \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\theta} - \theta)) \right\|_{e\sim p_N} = o_p(1). \]  

(3.4.43)

Proof. Let \( b^* \in \mathcal{R}^{p+1} \) be such that

\[ b^* = \gamma_1^{-1} \Sigma_1^{* -1} S_{1N}(0, 0, s). \]  

(3.4.44)

Observe that \( \gamma_1 \) and \( \Sigma_1^* \) are constants. Then, since \( S_{1N}(0, 0, s)|_{e\sim p_N} \) is bounded in probability, due to Lemma 3.4.5, it follows that \( b^* \) is bounded in probability. Now, we know from Theorem 3.4.10 that

\[ \sup_{b \in B_1} \left\| S_{1N}(b, t, s) - S_{1N}(0, 0, s) + \gamma_1 \Sigma_1^{*} b \right\|_{e\sim p_N} = o_p(1). \]  

(3.4.45)

Then, substituting \( b^* \) for \( b \) in (3.4.45), it is seen that for any \( b^*, t \) and \( s \),

\[ \left\| S_{1N}(b^*, t, s) \right\|_{e\sim p_N} = o_p(1), \]

since \( S_{1N}(0, 0, s) - \gamma_1 \Sigma_1^{*} b^* = 0 \) in light of the definition in (3.4.44). Therefore, from the definition in (3.4.41), we obtain

\[ \left\| S_{1N}(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\theta} - \theta)) \right\|_{e\sim p_N} = o_p(1). \]

The proof is terminated. \( \square \)

Then the question that might arise is whether the proposed estimate of \( \beta \) is asymptotically equivalent to the estimate of \( \beta \) for the case when the true value of the scale parameter, \( \theta \), is specified. This regression parameter estimate shall be denoted by \( \hat{\beta}_{\text{sc}} \).
For the standard homoscedastic linear model fitting problem, Jurečková established that \( \sqrt{N}(\hat{\beta} - \beta) = O_p(1) \) in her Lemma 5.2. In what follows, the signed-rank analogue for the case of an unweighted estimate of \( \beta \) is presented. Subsequently, the result is extended to the case where a weighted estimate of \( \beta \) is being sought. Recall that for the former case, the estimating process was given in (3.4.1). Next, the lemma is restated without proof.

**Lemma 3.4.6. Signed Rank Modified Lemma 5.2 : Jurečková (1977)** Assume that \( (D1^*, D2^*), (F1), (F2), (S1) \) and \( (W1) \) hold.

Then, to any \( \epsilon > 0 \), correspond \( \xi_\delta > 0, \eta > 0 \), and a positive integer \( N_0 \) such that

\[
P \left\{ \min_{\|b\| \geq 64} \left\| S_{1N}(b) \right\| \leq \eta \right\} < \epsilon
\]

(3.4.46)

holds for \( N > N_0 \).

Then, with the aid of an extended form of this lemma, the asymptotic equivalence of the weighted estimates, \( \hat{\beta} \) and \( \hat{\beta}_{\text{soa}} \), can be established. This result achieved in the next theorem. The method of proof for the result that follows is similar to that used in Akritas (1996).

**Theorem 3.4.13.** Assume that \( (D1^*), (D2^*), (F1), (F2), (S1) \) and \( (W1) \) hold. In addition, suppose that condition \( (B1) \) is satisfied.

Then

\[
\sqrt{N}(\hat{\beta} - \hat{\beta}_{\text{soa}}) \xrightarrow{P} 0.
\]

(3.4.47)

**Proof.** Let us assume that

\[
\sqrt{N}(\hat{\beta} - \beta) = O_p(1)
\]

(3.4.48)
holds for the moment. Then from Theorem 3.4.10 and Corollary 2, it is clear that

$$\left\| \sqrt{N}(\hat{\beta} - \beta) - \gamma_1^{-1}\Sigma_1^{*-1}S_{1N}(0, 0, \sqrt{N}(\hat{\theta} - \theta)) \right\|_{e \sim p_N} = o_p(1). \quad (3.4.49)$$

Equivalently,

$$\left\| \sqrt{N}(\hat{\beta} - \beta) - \gamma_1^{-1}\Sigma_1^{*-1}S_{1N}(0, 0, 0) \right\|_{e \sim p_N} = o_p(1). \quad (3.4.50)$$

This equivalence follows from the fact that $\|S_{1N}(0, 0, s) - S_{1N}(0, 0, 0)\|_{e \sim p_N} = o_p(1)$, in light of Theorem 3.3.4. Furthermore, it can be shown that

$$\left\| \sqrt{N}(\hat{\beta}_{sca} - \beta) - \gamma_1^{-1}\Sigma_1^{*-1}S_{1N}(0, 0, 0) \right\|_{e \sim p_N} = o_p(1), \quad (3.4.51)$$

by the standard asymptotic linearity argument of Kraft and van Eeden (1972), so that the desired result holds. It still remains to be proved that the convergence in (3.4.49) is valid. However, in view Corollary 2, this result holds if for each $\eta > 0$ and $\epsilon > 0$ and $\xi_2, \xi_3$, there exists $\xi_4$ satisfying

$$P\left\{ \inf_{b \in B^*(\xi_4), \xi_2 < \xi, s \in K(\xi)} \left\| S_{1N}(b, t, s) \right\|_{e \sim p_N} > \eta \right\} > 1 - \epsilon,$$

where $B^*(\xi_4) = \{b \in \mathbb{R}^{p+1} : \|t\| \geq \xi_4\}$ is exterior to the $(\xi_4)$-ball centered at 0 for $\xi_4 > 0$. This result can be achieved by a method of proof that is similar to that employed in Lemma 3.4.6. This completes the proof. \qed

Recall that the linear function $S_{1N}(0, 0, s) - \gamma_1\Sigma_1^*b$ can be used to approximate the process $S_{1N}(b, t, s)$. We know from Theorem 3.3.3 that $S_{1N}(0, 0, 0)_{e \sim p_N} \overset{L}{\rightarrow} N_{p+1}(0, \Sigma_1^*)$. Then the convergence in probability in $\|S_{1N}(0, 0, s) - S_{1N}(0, 0, 0)\|_{e \sim p_N} = o_p(1)$ of Theo-
rem 3.3.4 implies

\[ S_{1N}(0, 0, 0) \xrightarrow{e \sim p_N} N_{p+1}(0, \Sigma_1^*). \]  

(3.4.52)

Furthermore, that it suffices to use \( S_{1N}(0, 0, 0) - \gamma_1 \Sigma_1^* \) b to approximate \( S_{1N}(b, t, s) \). Since \( \|S_{1N}(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N}(\hat{\theta} - \theta), \sqrt{N}(\hat{\gamma} - \gamma))\|_{e \sim p_N} = o_p(1) \) on account of Corollary 2, it follows that

\[ \|\gamma_1 \Sigma_1^* \sqrt{N}(\hat{\beta} - \beta) - S_{1N}(0, 0, 0)\|_{e \sim p_N} = o_p(1). \]  

(3.4.53)

Thus,

\[ \|\sqrt{N}(\hat{\beta} - \beta) - \gamma_1^{-1} \Sigma_1^{*-1} S_{1N}(0, 0, 0)\|_{e \sim p_N} = o_p(1). \]  

(3.4.54)

In view of the result in (3.4.52), it is seen that

\[ \gamma_1^{-1} \Sigma_1^{*-1} S_{1N}(0, 0, 0) \xrightarrow{e \sim p_N} N_{p+1}(0, \gamma_1^{-2} \Sigma_1^{*-1}). \]

When this result, along with the fact that convergence in probability implies convergence in distribution are applied to (3.4.54), the following theorem is immediate.

**Theorem 3.4.14.** Assume that (D1*), (D2*), (F1), (F2), (S1) and (W1) hold. In addition, suppose that condition (B1) is satisfied.

Then

\[ \sqrt{N}(\hat{\beta} - \beta) \xrightarrow{e} N_{p+1}(0, \gamma_1^{-2} \Sigma_1^{*-1}). \]  

(3.4.55)

Furthermore, the next result follows directly from this theorem.

**Theorem 3.4.15.** Assume that (D1*), (D2*), (F1), (F2), (S1) and (W1) hold. In addition, suppose
that condition (B1) is satisfied.

Then

\[ \hat{\beta} \overset{\mathcal{L}}{\sim} N_{p+1}(\beta, \gamma_1^{-2}(X_1^*T X_1^*)^{-1}) \]  \hspace{1cm} (3.4.56)

There­fore, this theorem provides us with a general result for the asymptotic dis­tribution of weighted estimate, \( \hat{\beta} \). That is, \( \hat{\beta} \) follows an asymptotic distribution that is a 

\((p + 1)\)-variate normal with mean \( \beta \) and variance \( \gamma_1^{-2}(X_1^*T X_1^*)^{-1} \). An immediate conse­quence of this theorem is that \( \hat{\beta} \) is consistent for \( \beta \).

Before concluding this chapter, a description of the signed- rank dispersion func­tion is discussed. This is a function of the absolute residuals whose minimization prob­lem is equivalent to the one considered in the linear signed-rank based method present above. In particular, it is demonstrated that asymptotic uniform linearity and asymptotic uniform quadraticity are equivalent conditions. The latter condition is described as such because the function is quadratic with respect to the parameter of interest, \( \beta \). It turns out that it is much simpler to obtain the estimate of \( \beta \) as a minimizer of the dispersion function by reformulating the function as a Weighted Least Squares problem. In light of these facts, the theory behind estimating \( \beta \) on the basis of a suitably defined dispersion function is an appealing and important discussion that this study seeks to contribute to. This is presented in the section that follows.

3.5 Dispersion Function Criterion

This section begins with a brief description of the dispersion function. A dispersion function is a suitably defined function of residuals that is minimized in pursuit of an estimate of the parameter of interest. The signed-rank dispersion function is a sum of weighted absolute residuals, where the weights are some suitably defined signed-rank scores that are based on the ranks of the absolute residuals. In this current problem,
the residuals are scaled by the constants that express the heteroscedasticity. Let $b = (b_0, b_1^T)^T \in \mathbb{R}^{p+1}$. Then, for model given in (3.2.4), the estimate of $\beta$, that we seek is such that for a suitably defined and given $t$, there is a $b$ minimizes the function

$$D(b, t) := \sum_{i=1}^{N} \phi_1^+ \left( \frac{R[[e_i - x_{ii}^T b]e^{-x_{ii}^T t}]}{N + 1} \right) \left( \left\| \frac{e_i - x_{ii}^T b}{e_{ii}^T t} \right\| \right),$$

(3.5.1)

where $R[[e_i - x_{ii}^T b]e^{-x_{ii}^T t}]$ is the rank of the $|(e_i - x_{ii}^T b)e^{-x_{ii}^T t}|$ amongst $|(e_k - x_{ik}^T b)e^{-x_{ik}^T t}|$, for $k = 1, \ldots, N$, and the score generating function $\phi_1^+(u)$ that satisfies assumption (S1). Recall that $\sigma_i = e^{x_i^T \theta}$ for $i = 1, \ldots, N$. Then, note in (3.5.1) that $e_i = (1/\sigma_i)(y_i - x_{ii}^T \beta)$ and $x_{ii}^T = (1/\sigma_i)x_{ii}^T$ is the $i$th row the matrix, $X_i$.

The objective function given in (3.5.1), is nonnegative, piecewise linear and convex in $b$. In general, the solution to the minimization problem is not unique. However, in view of a result by Jaeckel (1972), the diameter of the set of possible solutions shrinks to zero as $N$ tends infinitely large. Furthermore, at the point at which the function is minimized for a given $t$, the partial derivatives of $D(b, t)$ with respect to $b$ should be approximately zero. Except at finite points, the partial derivatives of $D(b, t)$ exist almost everywhere and are given by

$$\frac{\partial D(b, t)}{\partial b_j} = -\sum_{i=1}^{N} \phi_1^+ \left( \frac{R[[e_i - x_{ii}^T b]e^{-x_{ii}^T t}]}{N + 1} \right) \text{sgn} \left( \frac{e_i - x_{ii}^T b}{e_{ii}^T t} \right) x_{ij}^* \quad \text{for } j = 1, \ldots, p + 1.$$

Note that $S_{1N}(b, t) = -\partial D(b, t)/\partial b_1, \ldots, -\partial D(b, t)/\partial b_{p+1}$ define the components of the gradient of the objective function given in (3.5.1).

Observe that $-\nabla D(b, t)$ is a special case of the linear rank statistic $S_{1N}(b, t, s)$ defined in (3.4.21). Recall that it has been demonstrated previously that the linear function $S_{1N}(0, 0, s) - \gamma_1 \Sigma_i^* b$ could serve as a suitable approximation to $S_{1N}(b, t, s)$ in the asymptotic sense. Similar to Jaeckel (1972), a quadratic function can serve as an approximation to $D_{1N}(b, t)$, a statistic that is given below. A description of the suitably defined quadratic function and
its related asymptotic properties are presented in the next section.

3.5.1  Asymptotic Uniform Quadraticity $D_{1N}(b, t)$

In this section, we define the objective function, $D_{1N}(b, t)$, that is used to obtain an estimate of $\beta$ in this study. In addition, another quadratic function, $Q_{1N}(b, t)$, that is obtained by performing a first order expansion of $D_{1N}(b, t)$ is also described. An asymptotic uniform quadratic result is established from these two functions. It is demonstrated that the asymptotic uniform quadraticity condition is equivalent to asymptotic uniform linearity condition obtained above. Furthermore, it is seen that the minimization of the quadratic function yields an estimate that is asymptotically equivalent to the minimizer of the objective function under consideration.

Recall that the statistic $S_{1N}(b, t, s)$ is a function of the variables $\{z_i(b, t) = (e_i - m_i^*)e_i : i = 1, \ldots, N\}$. Then, to match the notation of the statistic so as to obtain a $\sqrt{N}$-consistent estimate, define the working objective function as

$$D_{1N}(b, t) = \sum_{i=1}^{N} \phi_i^T \left( \frac{R||e_i - m_i^* e_i||}{N + 1} \right) \left( \left| \frac{e_i - m_i^*}{e_i} \right| \right). \tag{3.5.2}$$

In the sequel, note that $S_{1N}(0, t) = S_{1N}(0, 0, s)$ for $t = s$. Let

$$Q_{1N}(b, t) := -\gamma_1 b^T \Sigma_i^* b(1/2) + b^T S_{1N}(0, 0, s) + D_{1N}(0, 0), \tag{3.5.3}$$

be a convex function that is quadratic in $b$. This function will serve as an approximation to $D_N(b, t)$. Then, following Heiler and Willers (1988), utilizing diagonal sequence arguments we obtain the asymptotic uniform quadraticity, which is given in the next theorem.

**Theorem 3.5.1.** Assume that $(D1^*)$, $(D2^*)$, $(F1)$, $(F2)$, $(S1)$ and $(W1)$ hold.
Then, for any \( \varepsilon > 0 \),

\[
\lim_{N \to \infty} P \left( \sup_{b \in B(\xi_1), t \in \mathcal{X}(\xi_2)} \left| Q_{1N}(b, t) - D_{1N}(b, t) \right| \geq \varepsilon \right) = 0. \tag{3.5.4}
\]

**Proof.** The crux of the proof lies in the fact that, for a given \( t \), the functions \( D_{1N}(b, t) \) and \( Q_{1N}(b, t) \) are both proper convex with respect to all \( b \in \mathcal{R}^{p+1} \). Further, the gradients for these functions are given by

\[
\nabla D_{1N}(b, t) = -S_{1N}(b, t) = -S_{1N}(b, t, s) \quad \text{for} \quad t = s
\]

and

\[
\nabla Q_{1N}(b, t) = -\gamma_1 \Sigma_i^s b + S_{1N}(0, 0, s),
\]

respectively. Then putting these results together leads to

\[
\nabla \left( D_{1N}(b, t) - Q_{1N}(b, t) \right) = - \left[ S_{1N}(b, t, s) - S_{1N}(0, 0, s) + \gamma_1 \Sigma_i^s b \right]. \tag{3.5.5}
\]

There exists a diagonal infinite index set of nested sequences, denoted by \( \mathcal{N} \). With the aid of standard diagonal arguments, in the spirit of Heiler and Willers (1988, p. 179), it can be shown that

\[
\nabla \left( D_{1N}(b, t) - Q_{1N}(b, t) \right) \to 0, \quad \text{almost surely,}
\]

\[
S_{1N}(b, t, s) - (S_{1N}(0, 0, s) - \gamma_1 \Sigma_i^s b) \to 0, \quad \text{almost surely,}
\]

where both convergence results are for \( N \in \mathcal{N} \) and uniformly on \( C_o = \{ b \in B(\xi_1), t \in \mathcal{N} \}. \)
Since $N$ is arbitrary, it follows that

$$S_{1N}(b, t, s) - (S_{1N}(0, 0, s) - \gamma_1 \Sigma_1^* b) \overset{p}{\to} 0,$$

uniformly on $C_\omega$. Furthermore, from the last uniform convergence results, it is seen that

$$\lim_{N \to \infty} P \left( \sup_{b \in B(1), \; s \in C(\xi_2)} \left\| S_{1N}(b, t, s) - S_{1N}(0, 0, s) + \gamma_1 \Sigma_1^* b \right\| \geq \varepsilon \right)_{b \sim p_N} = 0. \quad (3.5.6)$$

$$\lim_{N \to \infty} P \left( \sup_{b \in B(1), \; t \in K(\xi_2)} \left| Q_{1N}(b, t) - D_{1N}(b, t) \right| \geq \varepsilon \right)_{b \sim p_N} = 0. \quad (3.5.7)$$

Thus, the proof is complete.

Although $D_{1N}(b, t)$ and $Q_{1N}(b, t)$ are asymptotically equivalent, the approximation, $Q_{1N}(b, t)$ cannot be used for estimating $\beta$ because its minimum depends on unspecified quantities $\gamma_1$ and $\beta$. However, the minimum provides us with the asymptotic normality result upon which the limiting distribution the estimator, the minimizer of $D_{1N}(b, t)$, depends. It should be noted that besides showing that $D_{1N}(b, t)$ and $Q_{1N}(b, t)$ approach each other as $N$ tends to be large, the result also demonstrates that asymptotic uniform linearity and asymptotic uniform quadraticity are equivalent.

Therefore, it follows from the argument above that an estimate of $\beta$ that is being sought can also be obtained as a solution to the minimization problem of the objective function given in (3.5.2). Through a suitable (IRWLS) formulation of the rank dispersion function, the minimizer is obtained in a simple manner following Sievers and Abebe (2004).
3.6 Conclusion

In this chapter, a signed-rank weighted estimator of the regression parameter, $\beta$, was proposed. It attains asymptotic normality and is consistent for $\beta$. Furthermore, it is asymptotically efficient in that its asymptotic variance is equivalent to the variance corresponding to the estimator of $\beta$ where the true scale parameter, $\theta$ is completely specified.
CHAPTER IV

RANK ESTIMATION OF SCALE PARAMETER, $\theta$

4.1 Introduction

In this chapter, we focus our attention on obtaining estimates of scale parameters which express the heteroscedasticity that is inherent to the model under consideration in this study. Simple algebraic manipulation of the heteroscedastic linear model yields a variance function model that is exponential with respect to the scale parameter and design. Clearly, this an intrinsically linear type of nonlinear model. For estimation problems involving this type of nonlinearity, it is intuitively appealing to linearize the model by applying a log transformation (Draper and Smith, p.222, (1981)). In this study, a log of absolute value transformation is applied to the nonlinear model.

In pursuit of the scale parameter estimates, a rank based estimation method is utilized to fit the linearized model. Thus, the proposed estimate of the scale parameter is obtained by regressing the transformed residuals from a weighted robust fit onto the design matrix. To obtain an initial estimate of the scale parameter, it suffices to utilize transformed residuals from an initial unweighted robust fit. For the standard linear models, linear rank statistic based methods of Jurečková (1971), Jaeckel (1972) and Hettmansperger and McKean (1998) can be used to obtain estimates of the sought scale coefficients as regression coefficients in the usual sense.

Since the responses are residuals based on an initial fit, this chapter extends these well known asymptotic results to rank scale estimation that accounts for the estimates of the underlying location parameters. Since the errors for the linearized model are asymmetric, the linear rank statistic uses rank scores with properties of location models under this class of distributions. A good review of the analysis of suitable choices of scores for
linear models with asymmetric errors can be found in McKean and Sievers (1989).

Recall that the heteroscedasticity being considered in this study is expressed through a variance function model that is known up to a constant. This model is described in the next section.

4.2 Model for Scale Problem

In this section, the heteroscedastic linear model is reexpressed so that the suitably aligned responses are functions of the scale constants which in turn depend on the scale parameter values and the design.

Consider heteroscedastic linear model

\[ y_i = \beta_0 + x_i^T \beta_1 + \sigma_i e_i, \quad i = 1, \ldots, N, \tag{4.2.1} \]

where \( y_1, \ldots, y_N \) are the responses, \( \beta_0 \in \mathcal{R} \), and \( \beta_1 \in \mathcal{R}^p \) are the regression parameters. Further, \( X \) is a centered design matrix with rows \( x_1^T, \ldots, x_N^T \), \( X \) is an \( N \times p \) matrix of known regression constants, \( \sigma_1, \ldots, \sigma_N \) are scale constants which can be modeled by the variance function

\[ \sigma_i = \exp\{x_i^T \theta\}, \quad i = 1, \ldots, N, \tag{4.2.2} \]

where \( \theta \) is a \( p \times 1 \) vector of scale parameters that we seek. The random variables \( e_1, \ldots, e_N \), given in (4.2.1) are independent identically distributed (i.i.d.) with common a distribution function \( F(y) \) and density function \( f(y) \).

From the equation in (4.2.1), the following alternative form is immediate

\[ y_i - \beta_0 - x_i^T \beta_1 = e_i e_i, \quad i = 1, \ldots, N. \tag{4.2.3} \]
It is clear that the right hand side of the last equation is nonlinear with respect to $\theta$. For models with this property, two strategies for obtaining estimates are commonly used. Nonlinear methods of estimation can be employed directly. Alternatively, due to the form of nonlinearity being considered here, an appropriate linearization can be applied in order to utilize existing standard methods for fitting linear models. The latter strategy for formulating the scale problem as a linear model was exploited by Hettmansperger and McKean (1998, p. 118) in a two sample problem where the given location is assumed to be the true value. In the sequel, we employ the strategy in more general settings where the given location is arbitrary.

4.2.1 Log Absolute Value Transformation

Consider the form of the model given in (4.2.3). Observe that for given $\beta_0$ and $\beta_1$, (4.2.3) depicts that the observations $y_1 - \beta_0 - x_1^T \beta_1, \ldots, y_N - \beta_0 - x_N^T \beta_1$, are related to the errors $e_1, \ldots, e_N$, which are weighted by exponential functions of design points through the coefficients $\theta_1, \ldots, \theta_p$. This nonlinear relationship can be linearized by an application of logs to both sides of (4.2.3). It is worth noting that for the case where the given locations of the samples are the true values, it is sufficient to assume that location is zero in order to establish the theory. Under this assumption, Hettmansperger and McKean (1998) illustrated that for the two sample scale problem, well known theoretical results of the linear rank statistic of Hájek and Šidák (1967) carry through. For the more general problem, it is not unusual to align the samples by replacing the location parameters by their corresponding estimates. In anticipation of this possibility, the left hand side of (4.2.3) can be reparametrized. For the moment, we motivate the problem by assuming that the given location is arbitrary. Taking natural logs of absolute values of (4.2.3) leads to the linear model

$$\log |y_i - \beta_0 - x_i^T \beta_1| = x_i^T \theta + \log |e_i|, \quad i = 1, \ldots, N,$$

(4.2.4)
where the \( \log |y_i - \beta_0 - x_i^T \beta_1| \) for \( i = 1, \ldots, N \), are the transformed responses related to the centered design matrix \( X \) and the error terms \( \log |e_i| \). It can be seen in (4.2.4) that the parameter of interest \( \theta \) takes the role of a regular vector of regression coefficients.

It should be pointed out that random variables \( \log |e_1|, \ldots, \log |e_N| \), are i.i.d. with a common cdf \( H^*_e(u) \), a distribution to be defined later. It should also be noted, however, that \( H^*_e(u) \) has a median \( \theta_0 \), hence the model in (4.2.4) contains an intercept type parameter which is subsumed in the errors. Following Jaeckel (1972), the dispersion function method cannot be used to estimate an intercept type parameter. In particular, utilizing a centered design that is full column rank in order to achieve a bounded set of solutions (Jaeckel, 1972, p. 1451), excludes the possibility of an intercept being estimated by the method. In this chapter, the matrix form of (4.2.4) that is often used is given by

\[
\log |Y - \beta_0 1 - X \beta_1| = X \theta + e^*,
\]

where \( Y = (y_1, \ldots, y_N)^T \), \( 1 \) is an \( N \times 1 \) column of ones, and \( e^* = (e_1^*, \ldots, e_N^*)^T \) with the correspondence \( e_i^* = \log |e_i| \) for \( i = 1, \ldots, N \).

### 4.2.2 Estimation Problem

Having converted the relationship from the nonlinear type in (4.2.3) to the linear type in (4.2.4), we proceed with the problem of estimating \( \theta \) utilizing standard methods for fitting linear models. Let \( \beta = (\beta_0, \beta_1^T)^T \). Define the residuals as

\[
\log |y_i - x_i^T \beta| - x_i^T t, \quad \text{for} \quad i = 1, \ldots, N.
\]

Any pursuit of estimates of \( \theta \), for any given \( \beta \), involves obtaining values that make the residuals as small as possible. There are various methods for obtaining values of the desired parameter that minimize functions of the residuals in (4.2.6). The decision
on which method should be used can be influenced by how the method behaves in the presence of outliers. Outlying responses can lead to large residuals which in turn causes the estimation process to breakdown. In this regard, it is desirable to use methods that are resistant to outliers. In order to curtail the effects of outliers on the estimate of $\theta$, a linear rank statistic that is a function of the variables defined in (4.2.6) can be employed.

Let $X_1 = (1, X)$. To obtain an estimate of $\theta$ for a given value of $\beta$, let $R[\log |y_i - x_i^T \beta| - x_i^T t]$ denote the rank of $\log |y_i - x_i^T \beta| - x_i^T t$ amongst $\log |y_k - x_k^T \beta| - x_k^T t$, for $k = 1, \ldots, N$. Note that $x_i^T$ is the $i$th row of $X_1$. We consider the function

$$S_2(t) = \sum_{i=1}^{N} \phi_2^*(\frac{R[\log |y_i - x_i^T \beta| - x_i^T t]}{N+1})x_i,$$  

(4.2.7)

where $S_2(t) = (S_{21}(t), \ldots, S_{2p}(t))^T$, the set of scores $\phi_2^*(\frac{i}{N+1})$ satisfying $\sum_{i=1}^{N} \phi_2^*(\frac{i}{N+1}) = 0$, in view of Remark 2.4.1, and $\phi_2^*(u)$ is a score generating function defined in an assumption given below.

For the residuals being utilized in the statistic in (4.2.7), it can be seen directly from the definition given in (4.2.3) that $\log |y_i - x_i^T \beta| - x_i^T \theta = \log |e_i|$ for $i = 1, \ldots, N$. Since we are interested in estimating $\theta$ when shifts in both location and scale are present, the residuals problem can be reexpressed in terms of perturbations to $\log |e_i|$'s. In what follows, we consider the asymptotic behavior of the function of $\log |e_i|$ when the location shift, through $b$, and the scale shift, through $t$, have been induced. A statistic that includes estimates of location and scale as special cases is defined. Before defining such a statistic, the underlying assumptions required for obtaining an estimate of $\theta$ are specified.

### 4.2.3 Model Assumptions

The next two conditions apply to the design matrix utilized in this chapter.
\(\lim_{N \to \infty} \max_{1 \leq i \leq N} x_i^T (X^T X)^{-1} x_i = 0.\)

\(\lim_{N \to \infty} \frac{1}{N} (X^T X) = \Sigma,\)

where \(\Sigma\) is a \(p \times p\) positive definite matrix. The next condition applies to the distribution of errors, \(e_1, \ldots, e_N\), whose probability density \(f\) satisfies

**F1**  
(i) \(f(y)\) is absolutely continuous.  
(ii) \(\int_{-\infty}^{\infty} \left( \frac{f'(y)}{f(y)} \right)^2 f(y) \, dy < \infty.\)

Thus, if (i) and (ii) hold, then \(f(y)\) has a finite Fisher information \(I(f)\).

Since the log of absolute value transformation is applied to the model, the resulting error variables \(\log |e_1|, \ldots, \log |e_N|\), follow an asymmetric distribution. Furthermore, with the scale problem having been converted to a location type problem, a condition on location-type scores is all that is needed for the rank estimation to be valid.

**S2** Let \(\phi_2^*(u)\) be a nondecreasing, square integrable function defined on the interval \((0, 1)\). Further, due to the symmetry of the underlying distribution of the errors, \(\phi_2^*\) satisfies \(\int_0^1 \phi_2^*(u) \, du = 0\), on account of Remark 2.4.1. It can be assumed that \(\int_0^1 (\phi_2^*(u))^2 \, du = 1\). The scores are approximated by \(\phi_2^*\left(\frac{i}{N+1}\right)\) so that \(\sum_{i=1}^N \phi_2^*\left(\frac{i}{N+1}\right) = 0\) holds.

Next, a restriction on the scaling constants \(\sigma_i\) is described, in view of Remark 2.3.1.

**W1** \(\sigma_1, \ldots, \sigma_N\) are bounded away from zero.
To proceed with the theory for the estimating process associated with the function S, additional notation is introduced and the linear rank statistic is defined in the next section.

4.2.4 Linear Rank Statistic

In order to obtain the $\sqrt{N}$-consistent asymptotic results, we will be working with $N^{-\frac{1}{2}}b$ for the location component of the problem. Observe that the vector $N^{-\frac{1}{2}}b$ is a sequence of parameter values that converge to zero [Sievers ((1978), p.629), (1983), p. 1165)]. Define

$$z_i^*(b) = \log |e_i - N^{-\frac{1}{2}}x_i^Tb|, \quad i = 1, \ldots, N,$$

$$= \log |e_i - m_i|.$$

Next, consider the scale component of the problem which, due to the log transformation, has been converted to the shift in location problem. To obtain the $\sqrt{N}$-consistent asymptotic results, we will be working with $N^{-\frac{1}{2}}t$. It is worth noting that, similar to Sievers ((1978), p. 629), (1983), p. 1165), $N^{-\frac{1}{2}}t$ is a sequence of parameter values that converge to zero. Then, for the estimation problem being considered in this chapter, define the random variables

$$v_i(b, t) = z_i^*(b) - \frac{1}{\sqrt{N}}x_i^Tt, \quad i = 1, \ldots, N, \quad (4.2.8)$$

$$= z_i^*(b) - g_i. \quad (4.2.9)$$

Then, for the estimation problem under consideration, let

$$S_{2N}(b, t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_2^*(R[z_i^*(b) - g_i] \frac{1}{N + 1})x_i \quad (4.2.10)$$
be the linear rank statistic that this study utilizes to develop the theory behind the proposed estimator of \( \theta \). Observe that \( S_{2N}(b, t) = (S_{2N1}(b, t), \ldots, S_{2Np}(b, t))^T \). In the next section, the asymptotic behavior of the proposed estimator is established based on the asymptotic linearity of \( S_{2N}(b, t) \).

4.3 Asymptotic Properties of Estimator

In the sections that follow, we establish a strong result on the asymptotic normality of the estimating process \( S_{2N}(b, t) \). Utilizing an asymptotic linearity result of the process, we derive the limiting distribution of the proposed estimator of \( \theta \), following the approach by Jurečková (1971). Before embarking on developing the theory behind the estimating process, it will be useful to describe the distributive properties of the underlying random variables of the statistics considered in this chapter.

4.3.1 Distributive Properties of Errors

In this section, we derive the distribution functions of the underlying variables of the statistics that are free of any type of shifts from the parent distribution of the errors.

Recall the definition of \( v_i(b, t) \) given in (4.2.9). In the rest of this chapter, it will be convenient to utilize the alternative form

\[
v_i(b, t) = \log|e_i - m_i| - g_i, \quad i = 1, \ldots, N.
\]
Observe that

\[ v_i(0, 0) = \log |e_i| = e_i^* \quad i = 1, \ldots, N, \]

\[ \Rightarrow |e_i| = e^{e_i^*} \]

\[ \Rightarrow e_i = \begin{cases} 
  e_i^* & \text{if } e_i > 0 \\
  -e_i^* & \text{if } e_i < 0 
\end{cases} \quad (4.3.1) \]

Thus, it is clear that \( e_i^* = \log |e_i| \) is not a one-to-one transformation, hence the \( e_i \) cannot directly be expressed in terms of \( e_i^* \). Also recall from assumption (F1) that the random variables, \( \{e_i : i = 1, \ldots, N\} \), have a common cdf \( F(u) \).

Next, we consider the random variables \( v_i(0, 0) = \log |e_i| \) for \( i = 1, \ldots, N \).

**Lemma 4.3.1.** Assume that the i.i.d. random variables \( e_1, \ldots, e_N \) have a common cdf \( F(u) \). Then, the random variables \( v_i(0, 0) = z_i^*(0) = \log |e_i| = e_i^* \) have the cdf \( H_{e_i^*}(u) = F(e_i^u) - F(-e_i^u) \).

**Proof.**

\[ H_{e_i^*}(u) = P(e_i^* \leq u) = P(\log |e_i| \leq u) \quad \text{for any } u \]

\[ = P(|e_i| \leq e^u) \]

\[ = P(e_i \leq e^u) - P(e_i \leq -e^u) \]

\[ = F(e^u) - F(-e^u). \]

Thus,

\[ v_i(0, 0) \bigg|_{e_i \sim F(u)} \text{ has the cdf } H_{e_i^*}(u) \text{ for } i = 1, \ldots, N. \quad (4.3.2) \]

\[ \square \]

Note that observations are composed of the deterministic component, which is the focus of the estimation process, and an error component upon which the stochastic nature
of the observations depend. Since the statistics involve joint occurrence of observations, it is useful to define a joint distribution function of the errors. Let \( p_N \) be the joint distribution of the random variables \( e_1, \ldots, e_N \), assuming that each \( e_i \) has the cdf \( F(u) \). That is,

\[
p_N = \prod_{i=1}^{N} f(e_i),
\]

where \( f \) is the common density of the random variables \( e_1, \ldots, e_N \).

It is worth noting that when the statistics being considered are functions of the distributions given above, then their distributive properties readily follow from the previous discussion. However, if the statistics are functions of the empirical distributions for the variables considered above, it is more convenient to establish the asymptotic behavior of those statistics through the approximations of the empirical processes. For the empirical processes considered in this study, it is possible to obtain suitable approximations that are functions of the variables. This is presented in the next section.

### 4.3.2 Distributional Properties of Linear Rank Statistics

The asymptotic distribution properties of the empirical process \( S_{2N}(b, t) \) depend on limiting behavior of \( S_{2N}(0, 0) \). In this section, an asymptotic normality result for the empirical process \( S_{2N}(0, 0) \) is established assuming that the errors \( e_1, \ldots, e_N \), are jointly distributed as \( p_N \). We derive the results for the asymptotic behavior of statistic \( T_{2N}(0, 0) \), which is an approximation to \( S_{2N}(0, 0) \) and then subsequently show that the results also hold for the empirical process itself. We begin this section by providing a general result describing the relationship between the convergence of vectors of statistics and the convergence of the respective components of those vectors. Except for being expressed in terms of \( S_{2N}(b, t) \) and \( T_{2N}(b, t) \), the remarks are essentially the same as those presented at the beginning of chapter three.

**Remark 4.3.1.** Consider \( p \)-vectors \( S \) and \( T \). For any \( a \in \mathbb{R}^p \), define a linear combination \( Xa = c \)
such that

\[ c_i = a_1 x_{i1} + a_2 x_{i2} + \ldots + a_p x_{ip}, \quad \text{for } i = 1, \ldots, N. \]

For a fixed \( h, 1 \leq h \leq p \), denote \( A_h = \{ a : a_j = 1 \text{ for } j = h; a_j = 0 \text{ for } j \neq h, j = 1, \ldots, p \}. \) Then, for \( a \in A_h, a^T S_{2N}(b, t) \) and \( a^T T_{2N}(b, t) \) selects the \( j \)th components of statistics \( S_{2N}(b, t) \) and \( T_{2N}(b, t) \), respectively.

**Remark 4.3.2.** In considering convergence results of \( p \)-vectors \( S_{2N}(b, t) \) and \( T_{2N}(b, t) \), it is worth noting the well known result that componentwise convergence in probability implies convergence in probability of the entire vectors. While many results have been established for components of vectors, the analog results for the entire vectors can be obtained easily by utilizing this well known result.

Next, consider the process \( S_{2N}(0, 0) \), a function of random variables \( v_i(0, 0) = z_i^*(0) = \log |e_i| \text{ for } i = 1, \ldots, N. \) Define

\[ S_{2N}(0, 0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_2^* \left( \frac{R_i \log |e_i|}{N + 1} \right) x_{i}, \tag{4.3.4} \]

where \( S_{2N}(0, 0) = (S_{2N1}(0, 0), \ldots, S_{2Np}(0, 0))^T. \) The distributive properties of the \( S_{2N}(0, 0) \) can be described following results by Hájek and Šidák (1967, p. 61). It is worth noting that assumption (S2) used in this study is a special case of the score function imposed by the authors since we have further assumed \( \int_0^1 \phi_2^*(u) du = 0 \) and \( \int_0^1 \phi_2^{*2}(u) du = 1; \) Hájek and Šidák (1967) considered \( \int_0^1 [\phi_2^*(u) - \overline{\phi_2^2}]^2 du < \infty. \)

**Theorem 4.3.1.** *(Theorem II.3.1c: Hájek and Šidák (1967))* Assume that \( (c_1, \ldots, c_N) \) and \( (\phi_2^*(\frac{1}{N+1}), \ldots, \phi_2^*(\frac{N}{N+1})) \) be arbitrary vectors. Let \( R_i \) be the rank of \( z_i^* \) among \( z_1^*, \ldots, z_N^* \).

Consider the statistic

\[ S_{hs} = \sum_{i=1}^{N} c_i \phi_2 \left( \frac{R_i}{N + 1} \right). \tag{4.3.5} \]
Let \( \sigma^2_{\phi_2} = \frac{1}{N-1} \sum_{i=1}^{N} \left[ \phi_2^*(\frac{R_i}{N+1}) - \phi^2_{2N} \right]^2 \), \( \phi^2_{2N} = \frac{1}{N} \sum_{i=1}^{N} \phi_2^*(\frac{R_i}{N+1}) \).

Then, under \( p_N \),

\[
E[S_{hs}] = \phi^2_{2N} \sum_{i=1}^{N} c_i,
\]

and

\[
Var[S_{hs}] = \sigma^2_{\phi_2} \sum_{i=1}^{N} (c_i - \bar{c})^2.
\]

In what follows, the distributive properties of \( S_{2N}(0, 0) \) are presented.

**Remark 4.3.3.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold.

Then

\[
E \left[ S_{2N}(0, 0) \right] \bigg|_{e^{-p_N}} = 0 \quad \text{and} \quad Var \left[ S_{2N}(0, 0) \right] \bigg|_{e^{-p_N}} = \frac{\sigma^2_{\phi_2}}{N} X^T X.
\]

Further,

\[
Var \left[ S_{2N}(0, 0) \right] \bigg|_{e^{-p_N}} \rightarrow \Sigma, \quad \text{as} \quad N \rightarrow \infty.
\]

**Proof.** For any \( a \in \mathbb{R}^p \), employ an arbitrary linear combination \( c = Xa \) so that

\[
a^T S_{2N}(0, 0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_2^* \left( \frac{R_i \log |e_i|}{N+1} \right) c_i.
\]

Then, substituting \( \frac{1}{\sqrt{N}} c_i \) for \( c_i \), taking \( R_i = R_i \log |e_i| \), and applying Theorem 4.3.1, it is seen that

\[
E \left[ a^T S_{2N}(0, 0) \right] \bigg|_{e^{-p_N}} = \phi^2_{2N} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i \bigg|_{e^{-p_N}} = 0,
\]
where the last equality is due to fact that \( \sum_{i=1}^{N} \phi_2^*(\frac{i}{N+1}) = 0 \) in light of assumption (S2), and

\[
\text{Var} \left[ a^T S_{2N}(0,0) \right]_{\psi \rightarrow \psi_N} = \sigma_2^2 \frac{1}{N} \sum_{i=1}^{N} c_i^2.
\]

Further,

\[
\sigma_2^2 \frac{1}{N} \sum_{i=1}^{N} c_i^2 \rightarrow 1 \cdot a^T \Sigma a, \quad \text{as } N \rightarrow \infty,
\]

where \( \sigma_2^2 \rightarrow 1 \) as \( N \rightarrow \infty \) since \( \sigma_2^2 = \int_0^1 [\phi_2^*(u)]^2 du \), which is equal to 1 under the assumption (S2). Note that \( \frac{1}{N} \sum_{i=1}^{N} c_i^2 = a^T (\frac{1}{N} X^T X) a \rightarrow a^T \Sigma a \) as \( N \rightarrow \infty \), on account of assumption (D2). Thus, these results hold for each component of \( S_{2N}(0,0) \). Further, the desired result for the entire \( S_{2N}(0,0) \) follows in view of Remark 4.3.2. The proof is complete.

Consider the following approximation of the empirical process \( S_{2N}(0,0) \).

\[
T_{2N}(0,0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_2^* \left( H_{\psi^*}(\log |e_i|) \right) x_i.
\]

(4.3.6)

Here, note that \( T_{2N}(0,0) = (T_{2N1}(0,0), \ldots, T_{2Np}(0,0))^T \). For this approximation, the distributive properties follow from Corollary V.1.6 of Hájek and Šidák (1967). This corollary is stated without proof in the sequel.

**Corollary 3. (Corollary V.1.6 of Hájek and Šidák (1967))**

Let \( c_i \) be any centered constants satisfying condition

\((HC1): \sum_{i=1}^{N} c_i^2 / \max_{1 \leq i \leq N} c_i^2 \rightarrow \infty \text{ as } N \rightarrow \infty.\)

Assume that the independent random variables \( y_1, \ldots, y_N \) follow any distribution function, \( F \), with the density \( f \) such that

\((HF1): f \text{ is absolutely continuous and } \int \left( \frac{f(y)}{F(y)} \right)^2 f(y) dy < \infty.\)
Assume that the score function, \( \phi_2^*(u) \), satisfies

(HS1): \( \lim_{N \to \infty} \int_0^1 [\phi_2^*(\frac{1+[u/N]}{N}) - \phi_2^*(u)]^2 du = 0 \), [uN] the largest integer not exceeding uN and \( \int_0^1 [\phi_2^*(u) - \overline{\phi_2^*}]^2 du < \infty \).

Let \( R_i \) be the rank of \( y_i \) amongst \( y_1, \ldots, y_N \).

Define

\[
S_{hsf} = \sum_{i=1}^N c_i \phi_2^* \left( \frac{R_i}{N+1}, f \right),
\]

where \( \phi_2^* \left( \frac{R_i}{N+1}, f \right) \) are approximate scores corresponding to the density, \( f \).

Then, under \( p_N \), \( S_{hsf} \) is asymptotically normal with

\[
\text{mean} = \bar{c} \sum_{i=1}^N \phi_2^* \left( \frac{i}{N+1}, f \right),
\]

and

\[
\text{variance} = \int_0^1 [\phi_2^*(u)]^2 du \sum_{i=1}^N (c_i - \bar{c})^2.
\]

Observe that in view of the condition (HS1), \( \phi_2^* \left( \frac{i}{N+1}, f \right) \) goes to 0 as \( N \to \infty \), where \( U_i = F(Y_i) \). Then, it follows that the limiting distribution result holds when the scores \( \phi_2^* \left( \frac{i}{N+1}, f \right) \) are replaced by \( \phi_2^*(U_i) \). Further, it suffices to obtain the extension based on the latter scores, and this is given in the next theorem.

**Theorem 4.3.2.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold.

Then

\[
T_{2N}(0, 0) \bigg|_{\theta = p_N} \overset{d}{\to} N_p(0, \Sigma).
\]
Proof. For any $a \in \mathbb{R}^p$, employ an arbitrary linear combination so that

$$a^T T_{2N}(0, 0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_2^i \left( H_{e^*} \left( \log |e_i| \right) \right) c_i,$$

where $c = Xa$ as described in Remark 4.3.1. Observe that, except for the constants, $c_i$, $a^T T_{2N}(0, 0)$ is a sum of independent, identically distributed terms $\phi_2^i \left( H_{e^*} \left( \log |e_i| \right) \right)$ with mean

$$E \left[ a^T T_{2N}(0, 0) \right] \bigg|_{e \sim p_N} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i \int \phi_2^i \left( H_{e^*} (u) \right) dH_{e^*} (u),$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i \int \phi_2^i (v) dv$$

$$= 0,$$

and variance

$$Var \left[ a^T T_{2N}(0, 0) \right] \bigg|_{e \sim p_N} = \frac{1}{N} \sum_{i=1}^{N} c_i^2 \int \left[ \phi_2^i \left( H_{e^*} (u) \right) \right]^2 dH_{e^*} (u),$$

$$= \frac{1}{N} \sum_{i=1}^{N} c_i^2 \int \phi_2^{i^2} (v) dv$$

$$= \frac{1}{N} \sum_{i=1}^{N} c_i^2.$$

Observe that $\int \phi_2^i (v) dv = 0$ and $\int \phi_2^{i^2} (v) dv = 1$, in light of assumption (S2). Furthermore, $\frac{1}{N} \sum_{i=1}^{N} c_i^2 = a^T \left( \frac{1}{N} X^T X \right) a \to a^T \Sigma a$ as $N$ becomes infinitely large, on account of assumption (D2). Since $a^T T_{2N}(0, 0)$ is sum of i.i.d. random variables, except for the $c_i$'s, it follows from an application of the Lindeberg Central Limit Theorem that $a^T T_{2N}(0, 0) \nrightarrow N(0, a^T \Sigma a)$ for every component of $T_{2N}(0, 0)$, in view of Remark 4.3.1. Thus, the desired convergence for the entire vector follow from this result in light Remark 4.3.2. Thus, the proof is complete. \qed
To determine the distribution of $S_{2N}(0, 0)$, we consider the convergence result for the simple linear rank established in Theorem V.1.5a of Hájek and Šidák (1967), p. 160. This theorem is restated next.

**Theorem 4.3.3. (Theorem V.1.5a: Hájek and Šidák (1967))** Assume that conditions (HC1), (HF1), and (HS1) given in Corollary 3 hold. Let $R_i$ be the rank of $y_i$ amongst $y_1, \ldots, y_N$.

Define

$$S_{hs} = \sum_{i=1}^{N} c_i \phi_2^* \left( \frac{R_i}{N+1} \right), \quad (4.3.8)$$

Then, under $p_N$, $S_{hs}$ is asymptotically normal with

$$\text{mean} = \overline{\phi_2^*} \sum_{i=1}^{N} c_i,$$

and

$$\text{variance} = \sum_{i=1}^{N} c_i^2 \int_0^1 [\phi_2^*(u) - \overline{\phi_2^*}]^2 du.$$

A direct application of the last result leads to the asymptotic behavior of $S_{2N}(0, 0)$, under $p_N$, and this is furnished in the next theorem.

**Theorem 4.3.4.** Assume that conditions (D1), (D2), (F1), (F2), (S2) and (W1) hold.

Then

$$S_{2N}(0, 0) \bigg|_{e \sim \rho_N} \overset{\mathcal{L}}{\rightarrow} N_p(0, \Sigma). \quad (4.3.9)$$

**Proof.** For $a \in \mathcal{R}^p$, define an arbitrary linear combination $a^T S_{2N}(0, 0)$, where $c = Xa$ as described in Remark 4.3.1. Then it suffices to show that $a^T S_{2N}(0, 0) |_{e \sim p_N} \overset{\mathcal{L}}{\rightarrow} N(0, a^T \Sigma a)$. Note that $a^T S_{2N}(0, 0)$ corresponds to $S_{hs} = \sum_{i=1}^{N} d_i \phi_2^* \left( \frac{R_i}{N+1} \right)$ with $d_i = \frac{1}{\sqrt{N}} c_i$, $R_i = R(c_i^*)$, for $i = 1, \ldots, N$, and $f$ replaced by $h_{c^*}$. It is seen from the result in Theorem 4.3.3 that,
under $p_N$, $a^T S_{2N}(0,0)$ is asymptotical normal with

$$\text{mean} = \phi_2^* \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i = 0,$$

(4.3.10)

where the last equality is due to the fact the $\sum_{i=1}^{N} \phi_2^*(\frac{1}{N+1}) = 0$, in light of assumption (S2), and

$$\text{variance} = \frac{1}{N} \sum_{i=1}^{N} c_i^2 \int_{0}^{1} [\phi_2^*(u) - \overline{\phi_2^*}]^2 du$$

$$\to a^T \Sigma a \cdot 1 \text{ as } N \to \infty,$$

since $\frac{1}{N} \sum_{i=1}^{N} c_i^2 \to a^T \Sigma a$ as $N \to \infty$, in view of assumption (D2), and $\int_{0}^{1} [\phi_2^*(u) - \overline{\phi_2^*}]^2 du = \int_{0}^{1} [\phi_2^*(u)]^2 du = 1$, due to assumption (S2). Then it can be seen that this result holds for every component of $S_{2N}(0,0)|_{e \sim p_N}$, in view of Remark 4.3.1. Further, for the entire vector, the convergence sought in (4.3.9) holds in light of Remark 4.3.2. This terminates the proof. \hfill \square

In the proof for Theorem V.1.5a of Hájek and Šidák (1967), it is demonstrated that, under $p_N$, the statistic $S_{hs}$ converges to its approximation in probability. This result is stated in what follows as a lemma.

**Lemma 4.3.2. (Corollary to Theorem V.1.5a: Hájek and Šidák (1967))** Assume that (HC1), (HF1), and (HS1) of Theorem 3 hold. Consider $S_{hs}$ defined in (4.3.8).

Define

$$T_{hs} = \sum_{i=1}^{N} c_i \phi_2^*(U_i),$$

(4.3.11)

where $U_i = F(Y_i)$ for $i = 1 \ldots, N$. Note that $R_1, \ldots, R_N$ are independent of $U^{(i)}$. Let $\sigma_c^2 = \sum_{i=1}^{N} c_i^2 \int_{0}^{1} [\phi_2^*(u) - \overline{\phi_2^*}]^2 du$. Suppose that $\sum_{i=1}^{N} c_i^2 / \max_{1 \leq i \leq N} c_i^2 \leq N$. 


Then, under $p_N$,

$$\lim_{c} P \left( \left| \frac{S_{hs} - T_{hs}}{\sigma_c} \right| > \epsilon \right) = 0,$$

(4.3.12)

for any $\epsilon > 0$.

Then this result can be employed to establish that, under $p_N$, $S_{2N}(0, 0)$ converges to $T_{2N}(0, 0)$. This is done in what follows.

**Theorem 4.3.5.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold.

Then

$$\left\| S_{2N}(0, 0) - T_{2N}(0, 0) \right\|_{\sigma \sim p_N} = o_p(1).$$

(4.3.13)

**Proof.** For $a \in \mathbb{R}^p$, define arbitrary linear combinations $a^T S_{2N}(0, 0)$ and $a^T T_{2N}(0, 0)$, where $c = Xa$ is as described in Remark 4.3.1. It is enough to show that

$$\left\| a^T S_{2N}(0, 0) - a^T T_{2N}(0, 0) \right\|_{\sigma \sim p_N} = o_p(1).$$

(4.3.14)

Observe that

$$a^T S_{2N}(0, 0) - a^T T_{2N}(0, 0) \bigg|_{\sigma \sim p_N} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i \left[ \phi_2^* \left( \frac{R(e_i^*)}{N+1} \right) - \phi_2^* \left( H_{e_i}(e_i^*) \right) \right].$$

Then, from Lemma 4.3.2 with constants $\frac{1}{\sqrt{N}} c_i$ substituted for $c_i$, $R_i = R(e_i^*)$ and $U_i = H_{e_i}(e_i^*)$, for $i = 1, \ldots, N$, it is seen that

$$\lim_{c} P \left( \left| \frac{a^T S_{2N}(0, 0) - a^T T_{2N}(0, 0)}{\sigma_c} \right| > \epsilon \right) = 0,$$

(4.3.15)

where $\sigma_c^2 = \frac{1}{N} \sum_{i=1}^{N} c_i^2 \int_0^1 [\phi_2^*(u) - \bar{\phi}_2^2]^2 du$. Observe that condition (HC1) is readily satisfied under assumption (D1). In view of assumption (D2), $\frac{1}{N} \sum_{i=1}^{N} c_i^2 \rightarrow a^T \Sigma a$ as $N \rightarrow \infty$, and
on account of assumption (S2), \( \int_0^1 [\phi_2^*(u) - \phi_2^0]^2 du = \int_0^1 [\phi_2^*(u)]^2 du = 1 \), so that \( \sigma_e^2 \to a^T \Sigma a \) as \( N \to \infty \). Then it can be seen that this result holds for every component of \( S_{2N}(0,0) - T_{2N}(0,0) \) in view of Remark 4.3.1. Further, for the entire vector the convergence sought in (4.3.13) holds in light of Remark 4.3.2 to terminate the proof.

In the sequel, we consider the case where the errors \( \log |e_i| \) have been perturbed with respect to scale, that is, through \( \tau \).

4.3.3 Asymptotic Linearity of \( S_{2N}(0, t) \)

In this section, linear functions of \( t \) needed to establish the limiting distribution of the proposed estimator of \( \theta \) are derived. Consider the shifted empirical process given by

\[
S_{2N}(0, t) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_2^\tau \left( \frac{R [\log |e_i| - g_i]}{N+1} \right) x_i,
\]

where \( S_{2N}(0, t) = (S_{2N1}(0, t), \ldots, S_{2Np}(0, t))^T \). The statistic in (4.3.16) can be approximated by

\[
T_{2N}(0, t) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \phi_2^\tau \left( H_{e\tau} (\log |e_i| - g_i) \right) x_i
\]

where \( T_{2N}(0, t) = (T_{2N1}(0, t), \ldots, T_{2Np}(0, t))^T \). In pursuit of the limiting distributions of the processes assuming that the errors jointly follow \( p_N \), it will useful to define the joint distribution of the errors that correspond to the shifted processes. Let the random variables \( e_1, \ldots, e_N \) follow the distributions \( F(e^{g_1} a), \ldots, F(e^{g_N} a) \). Then the joint distribution of these variables is given by

\[
q_N(0, t) = \prod_{i=1}^N f_i(e^{g_i} a) e^{g_i}.
\]

\[4.3.18\]
Consider the distribution of $S_{2N}(0, t)$ under $p_N$. An application of the translation property

**Lemma 4.3.3.** For any fixed $t$,

$$S_{2N}(0, t) \overset{d}{=} S_{2N}(0, 0),$$  \hspace{1cm} (4.3.19)

affords us the distribution of the shifted process. Except for letting $b = 0$, the proof for this property is the same as that used in the proof for the analogue of the property for the case $b \neq 0$ which is given later. Now, the asymptotic distribution of the right hand side of the relationship in (4.3.19) can be obtained by utilizing Theorem VI.2.4 of Hájek and Šidák (1967). For convenience, the theorem is stated in what follows.

**Theorem 4.3.6.** (VI.2.4: Hájek and Šidák (1967)) Assume that (HC1), (HF1) and (HS1) given in Corollary 3 hold. Let be $c_i$ and $g_i$ be centered constants. Consider the statistic $S_{hs}$ that is defined in 4.3.8. Then, under $q_N(0, t) = \prod_{i=1}^{N} e^{g_i} f(e_i e^{g_i})$, $S_{hs}$ is asymptotically normal with

$$\text{mean} = \sum_{i=1}^{N} c_i d_i \int_{0}^{1} \phi_{2}^*(u) \phi_{2}^*(u, h_{e^{*}}) du \quad \text{and variance} = \sum_{i=1}^{N} c_i^2 \int_{0}^{1} \left[ \phi_{2}^*(u) - \phi_{2}^* \right]^2 du,$$

where $\phi_{2}^*(u, h_{e^{*}}) = \phi_{2}^*(u, h_{e^{*}}) = -1 - F^{-1}\left(\frac{1+u}{2}\right) \frac{f'(F^{-1}\left(\frac{1+u}{2}\right))}{f(F^{-1}\left(\frac{1+u}{2}\right))}$, $f$ is the pdf, $f(t) = dF(t)$.

It is worth noting that $c_i$ and $d_i$ are centered. Thus, condition (HC1) is readily satisfied under assumption (D1), (S2) and (W1) is a special case of assumption (HS1). The proof for this result draws on the fact that $S_{hs}$ converges in probability to the approximation $T_{hs}$ in view of Lemma 4.3.2. Thus, the mean and variance of the shifted process are obtained based on its approximation. Further, the asymptotic normality of the approximation is obtained by exploiting LeCam's third lemma (Hájek and Šidák, 1967, p. 208). Finally, the result is extended to the statistic $S_{hs}$ using a contiguity argument and translation property.
In this section, it will be useful to first ascertain the distribution of $S_{2N}(0, 0)$ under $q_N(0, t)$. This result is given in the next theorem.

**Theorem 4.3.7.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold. Then

$$S_{2N}(0, 0) \underset{e \sim q_N(0, t)}{\overset{d}{\rightarrow}} N_p(-\gamma_2 \Sigma t, \Sigma), \quad (4.3.20)$$

where

$$\gamma_2 = \int \phi_2^*(u) \phi_2^*(u, h_{e^*}) du, \quad (4.3.21)$$

such that

$$\phi_2^*(u, h_{e^*}) = - \frac{h_{e^*}'(H_{e^*}^{-1}(\frac{1+u}{2}))}{h_{e^*}'(H_{e^*}^{-1}(\frac{1+u}{2}))}, \quad h_{e^*}'(t) = dH_{e^*}(t). \quad (4.3.22)$$

**Proof.** For $a \in \mathcal{R}^p$, define an arbitrary linear combination $a^T S_{2N}(0, 0)$, where $c = Xa$ as described in Remark 4.3.1. Then it suffices to show that

$$a^T S_{2N}(0, 0) \underset{e \sim q_N(0, t)}{\overset{d}{\rightarrow}} N_p(-\gamma_2 a^T \Sigma t, a^T \Sigma a). \quad (4.3.23)$$

Note that $a^T S_{2N}(0, 0)$ corresponds to $S_{hs} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i \phi_2^*(\frac{R_i}{N+1})$ with $R_i = R(e_i^*)$, for $i = 1, \ldots, N$, and $\phi_2^*(u, h_{e^*}) = \phi_2^*(u, f)$. Let $d_i = - \frac{1}{\sqrt{N}} x_i^T t$, for $i = 1, \ldots, N$. Then it is seen from the result in Theorem 4.3.6 that under $q_N(0, t)$, $a^T S_{2N}(0, 0)$ is asymptotical normal
with

\[
\text{mean} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i \int_0^1 \phi_2^*(u) \phi_2^*(u, f) du \\
= \frac{1}{\sqrt{N}} a^T X^T \left( - \frac{1}{\sqrt{N}} X t \right) \gamma_2 = -a^T \left( \frac{1}{N} X^T X \right) t \gamma_2 \\
\rightarrow -\gamma_2 a^T \Sigma t \quad \text{as } N \rightarrow \infty,
\]

and

\[
\text{variance} = \frac{1}{N} \sum_{i=1}^{N} c_i^2 \int_0^1 [\phi_2^*(u) - \overline{\phi_2}]^2 du \\
= a^T \left( \frac{1}{N} X^T X \right) a \int_0^1 [\phi_2^*(u)]^2 du \\
\rightarrow a^T \Sigma a \cdot 1 \quad \text{as } N \rightarrow \infty,
\]

since \( \frac{1}{N} X^T X \rightarrow \Sigma \) in view of assumption (D2), \( \gamma_2 = \int_0^1 \phi_2^*(u) \phi_2^*(u, f) du, \int_0^1 [\phi_2^*(u) - \overline{\phi_2}]^2 du = \int_0^1 [\phi_2^*(u)]^2 du = 1 \) in view of assumption (S2). Since the result holds for any component of \( S_{2N}(0, 0) \) when \( a \) is suitably defined as in Remark 4.3.1, it follows that the converge holds for the entire vector due to Remark 4.3.2. This leads to the desired result to complete proof. \( \square \)

Next, it is demonstrated that the limiting distribution in the last result also holds for the statistic \( T_{2N}(0, 0) \) under \( q_N(0, t) \). This is achieved by exploiting the contiguity of \( q_N(0, t) \) to \( p_N \). For convenience, a brief description of the principle of contiguity is provided by the next definition.

**Definition 4.** A sequence of densities \( q_N \) is contiguous to another sequence of densities \( p_N \) if for any sequence of events \( \{A_N\} \),

\[
\int_{\{A_N\}} p_N \rightarrow 0 \Rightarrow \int_{\{A_N\}} q_N \rightarrow 0.
\]
It is shown in the monograph by Hájek and Šidák (1967, p.217) that the sequence of densities $q_N(0, t)$ is contiguous to that of densities $p_N$. Since

$$\begin{align*} \left\| S_{2N}(0, 0) - T_{2N}(0, 0) \right\|_{e \sim p_N} = o_p(1), \end{align*}$$

in light of Theorem 4.3.5, it follows that the convergence also holds under $q_N(0, t)$ on account of the contiguity of densities $q_N(0, t)$ to $p_N$. Thus, we have

$$\begin{align*} \left\| S_{2N}(0, 0) - T_{2N}(0, 0) \right\|_{e \sim q_N(0, t)} = o_p(1). \end{align*}$$

Furthermore, since the convergence in probability implies convergence in distribution, it follows that $T_{2N}(0, 0)|_{e \sim q_N(0, t)} \xrightarrow{d} S_{2N}(0, 0)|_{e \sim q_N(0, t)}$. Thus, it is clear that the next theorem follows from this result, along with the fact that $S_{2N}(0, 0)|_{e \sim q_N(0, t)} \xrightarrow{d} N_p(-\gamma_2 \Sigma t, \Sigma)$, on account of Theorem 4.3.7.

**Theorem 4.3.8.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold. Then

$$\begin{align*} T_{2N}(0, 0)|_{e \sim q_N(0, t)} \xrightarrow{d} N_p(-\gamma_2 \Sigma t, \Sigma). \end{align*}$$

(4.3.24)

It is worth noting that the expectation in Theorem 4.3.8, hence Theorem 4.3.7, specifies the linearity term which is contained in the linear function that is used to approximate the process $T_{2N}(0, t)$. Thus, with this result established, we can now focus on the asymptotic linearity of the shifted process.

**4.4 Asymptotic Linearity Result**

In this section, the asymptotic linearity results for both of the processes $T_{2N}(0, t)$ and $S_{2N}(0, t)$ are presented. Following the strategy employed in Jurečková (1969), (1971), the asymptotic linearity of the former process is demonstrated first before establishing the
result for the latter process. Subsequently, the linearity result for $S_{2N}(0, t)$ is obtained by invoking contiguity argument and translation properties on its analogue for $T_{2N}(0, t)$.

Recall that $g_i = \frac{1}{\sqrt{N}}x_i^Tt$ for $i = 1, \ldots, N$. Observe that $\bar{g} = \frac{1}{N} \sum_i^N g_i = 0$ follows from the fact that $X$ is a centered design matrix. Further, the constants representing shifts also satisfy the conditions that follow.

$(g_1)$: For a nonzero vector, $t$

$$\lim_{N \to \infty} \sum_{i=1}^{N} g_i^2 = t^T \Sigma t.$$  

This follows from noting that $\sum_{i=1}^{N} g_i^2 = \sum_{i=1}^{N} (\frac{1}{\sqrt{N}}x_i^Tt)^2 = \frac{1}{N} t^T (X^TX)t = t^T (\frac{1}{N}X^TX)t$  

$\to t^T \Sigma t$, as $N \to \infty$, where the last convergence is due to the fact that $\frac{1}{N}X^TX \to \Sigma$, as $N \to \infty$, in view of assumption (D2).

$(g_2)$: and the condition

$$\lim_{N \to \infty} \max_{1 \leq i \leq N} \frac{g_i^2}{\sum_{i=1}^{N} g_i^2} = 0,$$

which follows from the fact that $\lim_{N \to \infty} \max_{1 \leq i \leq N} x_i^T (X^TX)^{-1} x_i = 0$, on account of assumption (D1). To see this, fix $j$ in $X = \{(x_{ij})\}$ and $t = \{(t_j)\}$ for $j = 1, \ldots, p$, and observe that

$$\frac{\max_{1 \leq i \leq N} g_i^2}{\sum_{i=1}^{N} g_i^2} = \frac{\max_{1 \leq i \leq N} (\frac{1}{\sqrt{N}}x_{ij}t_j)^2}{\sum_{i=1}^{N} (\frac{1}{\sqrt{N}}x_{ij}t_j)^2} = \frac{t_j^2 \cdot \max_{1 \leq i \leq N} (\frac{1}{N}x_{ij}^2)}{t_j^2 \frac{1}{N} \sum_{i=1}^{N} x_{ij}^2} = \frac{\max_{1 \leq i \leq N} (x_{ij}^2)}{\sum_{i=1}^{N} x_{ij}^2},$$

for $i = 1, \ldots, N$, $j = 1, \ldots, p$. Now,

$$\frac{x_{ij}^2}{\sum_{i=1}^{N} x_{ij}^2} \leq h_{Ni} \quad \text{for} \quad i = 1, \ldots, N, \quad j = 1, \ldots, p,$$
where $h_{Nii} = x_i^T(X^TX)^{-1}x_i$, that is, the $i$th diagonal of $H = X(X^TX)^{-1}X^T$. Further,

$$\max_{1 \leq j \leq p} \max_{1 \leq i \leq N} \frac{x_{ij}^2}{\sum_{i=1}^N x_{ij}^2} \leq \max_{1 \leq i \leq N} h_{Nii}.$$ 

Then, due to the fact that Huber's condition given in assumption (D1) implies

$$\lim_{N \to \infty} \max_{1 \leq i \leq N} h_{Nii} = 0,$$

Noether's condition. In addition, since $\lim_{N \to \infty} \frac{1}{N} x_i x_j^T = \sigma_{ij}$ and

$$\Sigma = [\sigma_{ij}]_{i,j=1}^p = [\sigma^{(1)}, \ldots, \sigma^{(p)}], \; \sigma^{(j)} \text{ is the } j \text{th column of the } \Sigma, \text{ by assumption (D2), then for fixed } j, \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N x_{ki}^2 = \sigma_j^2 \text{ for } j = 1, \ldots, p. \text{ Then the convergence of the vector } g = (g_1, \ldots, g_N) \text{ follows from these facts.}$$

In what follows, we present the result for the scale estimation problem with linear formulation where the errors have been have been perturbed with respect to scale only. In seeking the asymptotic linearity result of the process $T_{2N}(0, \tau)$, we first note a general linearity result for the regression problem that is established in Hettmansperger and McKean (1998).

**Theorem 4.4.1. (Theorem A.2.5: Hettmansperger and McKean (1998)).** Let $c_1, \ldots, c_N$ be any centered regression constants satisfying the conditions

(HMC1): $\max_{1 \leq i \leq N} c_i^2 / \sum_{i=1}^N c_i^2 \to 0$ as $N \to \infty,$

(HMC2): $N^{-1} \sum_{i=1}^N c_i^2 \to \sigma_c^2 > 0.$

Assume that the independent random variables $y_1, \ldots, y_N$ follow a distribution, $F$ with the density $f$ such that

(HMF1): $f$ is absolutely continuous and $\int (f'(y))^2 f(y)dy < \infty.$

The score function satisfies the condition

(HMS1): $\phi_1(u)$ is nondecreasing, square-integrable, and bounded function defined on the interval $(0, 1)$ such that $\int_0^1 \phi_1(u) du = 0$ and $\int_0^1 [\phi_1(u)]^2 du = 1.$

In addition let the constants representing the shifts $d_1, \ldots, d_N$ be such that
(hmd1): \( \sum_{i=1}^{N} d_i = 0 \),

(hmd2): \( \sum_{i=1}^{N} d_i^2 \to \sigma_d^2 > 0 \), as \( N \to \infty \),

(hmd3): \( \max_{1 \leq i \leq N} d_i^2 \to 0 \), as \( N \to \infty \).

(hmd4): \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i d_i \to \sigma_{cd} \), as \( N \to \infty \).

Define the approximation

\[
T_{hm} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i \phi_1(F(y_i)),
\]

and the shifted process' approximation

\[
T_{hmd} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i \phi_1(F(y_i - d_i)).
\]

Then

\[
\left| T_{hmd} - T_{hm} - E \left[ T_{hmd} \right] \right|_{\omega_{-\rho_N}} = o_p(1),
\]

where \( E[T_{hmd}]_{\omega_{-\rho_N}} \to -\gamma_{1d} \sigma_{cd} \) as \( N \to \infty \), and

\[
\gamma_1 = \int_0^1 \phi_1(u) \phi_1(u, f) du; \phi_1(u, f) = -f'(F^{-1}(u))/f(F^{-1}(u)).
\]

This result can be utilized in seeking to obtain the linearity result of \( T_{2N}(0, t) \) as is seen in what follows.

**Theorem 4.4.2.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold.

Then

\[
\left| T_{2N}(0, t) - T_{2N}(0, 0) + \gamma_2 \Sigma t \right|_{\omega_{-\rho_N}} = o_p(1).
\]

**Proof.** Observe that in view of Remark 4.3.1, it is sufficient to establish this result for any
selected component given by

\[ a^T T_{2N}(0, t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i \phi_2^\prime \left( H_{e^*}(\log |e_i| - g_i) \right). \]  

(4.4.5)

Then it suffices to show that

\[ \left\| a^T T_{2N}(0, t) - a^T T_{2N}(0, 0) + \gamma_2 a^T \Sigma t \right\|_{\text{e}^{-p_N}} = o_p(1). \]  

(4.4.6)

Now, recall Theorem 4.4.1, and observe that conditions (HMC1) and (HMC2) are special cases of assumptions (D1) and (D2), (HMF1) and (HMS1) are equivalent to assumptions (F1) and (S2), respectively. Then apply Theorem 4.4.1 in view of the correspondence that follows. Let \( c_i \) be components of \( Xa \), \( d_i = g_i = \frac{1}{\sqrt{N}} x_i^T t \) for \( i = 1, \ldots, N \). Further, substitute \( F(y_i) \) by \( H_{e^*}(\log |e_i|) \), so that \( \gamma \) is substituted by \( \gamma_2 \). Note that \((hmd_2)\) and \((hmd_3)\) are readily satisfied when \((g_1)\) and \((g_2)\) hold and the constants \( g_i \) satisfy the conditions \((g_1)\) and \((g_2)\) due to assumptions (D2) and (D1), respectively. Observe that \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i g_i = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} ax_i (\frac{1}{\sqrt{N}} x_i^T t) = a^T (\frac{1}{N} X^T X) t \rightarrow a^T \Sigma t \) as \( N \rightarrow \infty \). Clearly, for this case \( \sigma_{cd} = a^T \Sigma t \) so that condition \((hmd_4)\) is satisfied. Further, \( E[a^T T_{2N}(0, t)]_{\text{e}^{-p_N}} \rightarrow -\gamma_2 \sigma_{cd} \). Then it can be seen that the result also holds for every component of vector on the left hand side of (4.4.4), in view of Remark 4.3.1. Then, for this entire vector, the result also holds in light of Remark 4.3.2. The proof is complete.

The last result can be extended to the empirical process \( S_{2N}(0, t) \) by utilizing the result

\[ \left\| S_{2N}(0, 0) - T_{2N}(0, 0) \right\|_{\text{e}^{-p_N}} = o_p(1), \]  

(4.4.7)
in light of Theorem 4.3.5 and using the contiguity of \( q_N(0, t) \) to \( p_N \) to yield

\[
\left\| S_{2N}(0, 0) - T_{2N}(0, 0) \right\|_{\eta^\sim q_N(0, t)} = o_p(1), \quad (4.4.8)
\]

To obtain this result, we apply the translation property given in Lemma 4.3.3 and its analogue, the latter of which is presented in what follows.

**Lemma 4.4.1.** For any fixed \( t \),

\[
T_{2N}(0, t) \left|_{\eta^\sim p_N} \right. \equiv T_{2N}(0, 0) \left|_{\eta^\sim q_N(0, t)} \right.. \quad (4.4.9)
\]

The result in (4.4.9) can be easily shown to hold as a special case of a translation property to be given later. Thus, suitably applying these properties to (4.4.8) gives

\[
\left\| S_{2N}(0, t) - T_{2N}(0, t) \right\|_{\eta^\sim p_N} = o_p(1),
\]

so that, this result fact, along with the convergence given in (4.4.7), due to Theorem 4.3.5 leads to

\[
\left\| S_{2N}(0, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{\eta^\sim p_N} = o_p(1). \quad (4.4.10)
\]

It is clear from (4.4.10) that the statistic \( S_{2N}(0, t) \) can be approximated by linear function \( S_{2N}(0, 0) - \gamma_2 \Sigma t \). Now, suppose that we seek this convergence result for the case where \( t \) is a member of a bounded set. This is equivalent to establishing that the asymptotic uniform linearity of the process \( S_{2N}(0, t) \) holds for every value of \( t \) belonging to the bounded set.

**Theorem 4.4.3.** Assume that \( (D1), (D2), (F1), (F2), (S2) \) and \( (W1) \) hold.

Then

\[
\sup_{t \in \mathcal{K}(\xi_2)} \left\| S_{2N}(0, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{\eta^\sim p_N} = o_p(1), \quad (4.4.11)
\]
where $\mathcal{K}(\xi_2) = \{t \in \mathcal{R}^p : \|t\| \leq \xi_2\}$ is the $(\xi_2)$-ball centered at 0 for $\xi_2 > 0$.

The proof of this theorem is obtained by invoking standard diagonal sequence arguments based on the existence of an index set $\tilde{N}$ such that

$$S_{2N}(0, t) - \left( S_{2N}(0, 0) - \gamma_2 \Sigma t \right) \rightarrow 0 \text{ almost surely},$$

for all $t \in \mathcal{R}^p$. This result is equivalent to the convergence given in (4.4.11), following Heiler and Willers (1988). It should be noted that, on account of Theorem 4.4.2 and Theorem 4.3.5, it must also be true that

$$\sup_{t \in \mathcal{K}(\xi_2)} \left\| T_{2N}(0, t) - T_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{\mathcal{L}^p} = o_p(1). \quad (4.4.12)$$

This concludes the results on the process for the case in which the errors are perturbed with respect to scale only. A more typical situation is the case where the errors have been perturbed with respect to both location and scale. This is the focus of the next section.

4.4.1 Asymptotic Linearity of $S_{2N}(b, t)$

In this section, the linearity of the empirical process $S_{2N}(b, t)$ defined in (4.2.10) is examined. In what follows, it is worth recalling that the location type parameter $b$ does not affect the distribution of the scale parameter estimates. In considering a shift in location as well as scale, it is seen that the linearity result is shown to be bounded by a sum of two expressions, one of which is that considered in Theorem 4.4.2. Then the problem reduces to showing that the remaining expression, which captures the location aspect of the problem, goes to zero in probability. In particular, the asymptotic behavior of $S_{2N}(b, t)$ is obtained by showing that the process is asymptotically equivalent to the statistic $S_{2N}(0, t)$ discussed above.
4.4.2 Asymptotic Linearity of $T_{2N}(b, t)$

Consider the approximation to the empirical process $S_{2N}(b, t)$.

Define

$$T_{2N}(b, t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi_2^*(H_{e^*}(\log |e_i - m_i| - g_i))x_i,$$  \hspace{1cm} (4.4.13)

where $T_{2N}(b, t) = (T_{2N1}(b, t), \ldots, T_{2Np}(b, t))^T$. In the sequel, the asymptotic linearity result for the process $T_{2N}(b, t)$ is established. We extend result in Theorem 4.4.2 to a two parameter case of $b$ and $t$. It is worth recalling that $m_i = \frac{1}{\sqrt{N}} x_{1i}^T b$ for $i = 1, \ldots, N$. Further,

$(m_1)$: the constants $m_i$ have the property

$$\lim_{N \to \infty} \max_{1 \leq i \leq N} m_i^2 = 0,$$

which follows from the assumption $\lim_{N \to \infty} \max_{1 \leq i \leq N} x_{1i}^T (X_1^T X_1)^{-1} x_{1i} = 0$, which is a special case of the assumption (D1*) specified in chapter three.

**Theorem 4.4.4.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold.

Then

$$\left\| T_{2N}(b, t) - T_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{e \sim p_N} = o_p(1).$$  \hspace{1cm} (4.4.14)

**Proof.** Consider the expression on the left hand side of (4.4.14). Adding and subtracting
$T_{2N}(0, t)$ along with an application of the triangular inequality gives

$$\left\| T_{2N}(b, t) - T_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{e \sim p_N} \leq \left\| T_{2N}(b, t) - T_{2N}(0, t) \right\|_{e \sim p_N} + \left\| T_{2N}(0, t) - T_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{e \sim p_N}. $$

It is clear that the second term on the right hand side of the last inequality goes to zero in probability, on account of Theorem 4.4.2. Then it only remains to show that the vector $\left\| T_{2N}(b, t) - T_{2N}(0, t) \right\|_{e \sim p_N}$ goes to zero in probability. Consider the case where $t = 0$. It can be shown that it suffices to demonstrate that

$$\left\| T_{2N}(b, 0) - T_{2N}(0, 0) \right\|_{e \sim p_N} = o_p(1). \quad (4.4.15)$$

Consider any linear combination that selects any component of $T_{2N}(b, 0) - T_{2N}(0, 0)$, in view of Remark 4.3.1. Then the result in (4.4.15) holds if it can be shown that

$$\left\| a^T T_{2N}(b, 0) - a^T T_{2N}(0, 0) \right\|_{e \sim p_N} = o_p(1). \quad (4.4.16)$$

Further, this result holds if it can be demonstrated that

$$\text{Var} \left[ a^T T_{2N}(b, 0) - a^T T_{2N}(0, 0) \right]_{e \sim p_N} \rightarrow 0 \text{ as } N \rightarrow \infty. \quad (4.4.17)$$
Observe that

\[
\text{Var} \left[ a^T T_{2N}(b, 0) - a^T T_{2N}(0, 0) \right]_{e \sim p_N}
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} c_i^2 \text{Var} \left[ \phi^*_2 \left( H_{e^*} \left( \log |e_i - m_i| \right) \right) - \phi^*_2 \left( H_{e^*} \left( \log |e_i| \right) \right) \right]^2
\]

\[
\leq \frac{1}{N} \sum_{i=1}^{N} c_i^2 E \left[ \phi^*_2 \left( H_{e^*} \left( \log |e_i - m_i| \right) \right) - \phi^*_2 \left( H_{e^*} \left( \log |e_i| \right) \right) \right]^2
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} c_i^2 E \left[ \phi^*_2 \left( F(|e_i - m_i|) - F(-|e_i - m_i|) \right) - \phi^*_2 \left( F(|e_i|) - F(-|e_i|) \right) \right]^2
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} c_i^2 \int \left[ \phi^*_2 \left( F(|v - m_i|) - F(-|v - m_i|) \right) - \phi^*_2 \left( F(|v|) - F(-|v|) \right) \right]^2 dF(|v|)
\]

\[
\leq \left( \frac{1}{N} \sum_{i=1}^{N} c_i^2 \right) \left( \int \max_{1 \leq i \leq N} \left[ \phi^*_2 \left( F(|v - m_i|) - F(-|v - m_i|) \right) - \phi^*_2 \left( F(|v|) - F(-|v|) \right) \right]^2 dF(|v|) \right)
\]

Since \( \frac{1}{N} \sum_{i=1}^{N} c_i^2 \) converges to \( a^T \Sigma a \), as \( N \to \infty \), in view of assumption (D2), it is clear that the result we seek can be obtained if it can be shown that

\[
\limsup \left( \int \max_{1 \leq i \leq N} \left[ \phi^*_2 \left( F(|v - m_i|) - F(-|v - m_i|) \right) - \phi^*_2 \left( F(|v|) - F(-|v|) \right) \right]^2 dF(|v|) \right) = 0 \quad (4.4.18)
\]

Let \( \epsilon > 0 \) be given. Since \( \phi^*_2(v) \) is continuous a.e., it can be assumed to be continuous at \( F(|v|) - F(-|v|) \). Then there exists a \( \delta_1 > 0 \) such that \( |\phi^*_2(z) - \phi^*_2(F(|v|) - F(-|v|))| < \epsilon \) for \( |z - (F(|v|) - F(-|v|))| < \delta_1 \). By the uniform continuity of \( F \) choose \( \delta_2 > 0 \) such that \( |F(|w|) - F(-|w|) - (F(|v|) - F(-|v|))| < \delta_1 \) for \( 2 \cdot |w - v| < \delta_2 \). Observe that

\[
|F(|w|) - F(-|w|) - (F(|v|) - F(-|v|))| < |F(|w|) - F(|v|)| + |F(-|w|) - F(-|v|)| < \delta_1
\]

for \( |w - v| + |v - w| < \delta_2 \). Since it has been assumed that the constants \( m_i \) satisfy the
condition \((m_1)\), we can select \(N_0\) so that for \(N > N_0\) implies

\[
2 \cdot \max_{1 \leq i \leq N} \{|m_i|\} < \delta_2. \tag{4.4.19}
\]

This follows from the fact that \(|m_i| \leq m_i^2\) for \(i = 1, \ldots, N\). Thus for \(N > N_0\),

\[
\left| F(|v - m_i|) - F(-|v - m_i|) - (F(|v|) - F(|v|)) \right| < \delta_1, \quad \text{for } i = 1, \ldots, N.
\]

and, hence,

\[
\left| \phi_2^* \left( F(|v - m_i|) - F(-|v - m_i|) \right) - \phi_2^* \left( F(|v|) - F(-|v|) \right) \right| < \epsilon, \quad \text{for } i = 1, \ldots, N.
\]

Thus, for \(N > N_0\),

\[
\max_{1 \leq i \leq N} \left[ \phi_2^* \left( F(|v - m_i|) - F(-|v - m_i|) \right) - \phi_2^* \left( F(|v|) - F(-|v|) \right) \right]^2 < \epsilon^2.
\]

Therefore,

\[
\lim \sup \left( \int \max_{1 \leq i \leq N} \left[ \phi_2^* \left( F(|v - m_i|) - F(-|v - m_i|) \right) - \phi_2^* \left( F(|v|) - F(-|v|) \right) \right]^2 \, dF(|v|) \right) \leq \epsilon^2.
\]

Since \(\epsilon\) can always be chosen to be arbitrarily small, it follows that result in 4.4.18 holds for every component of \(T_{2N}(b, 0) - T_{2N}(0, 0)\) in view of Remark 4.3.1. Furthermore, it must also be true that \(\|T_{2N}(b, 0) - T_{2N}(0, 0)\|_{e \sim p_N}\) goes to zero in probability on account of Remark 4.3.2.

Next, consider the case \(t \neq 0\). In light of the contiguity of the sequence of densities
We obtain the result

\[ \left\| T_{2N}(b, 0) - T_{2N}(0, 0) \right\|_{e \sim q_N(0,t)} = o_p(1) \quad (4.4.20) \]

Recall the translation property in Lemma 4.3.3

\[ T_{2N}(0, t) \left|_{e \sim p_N} \right. \overset{d}{=} T_{2N}(0, 0) \left|_{e \sim q_N(0,t)} \right. \quad (4.4.21) \]

Since the shift under consideration is only through \( t \), it also follows that

\[ T_{2N}(b, t) \left|_{e \sim p_N} \right. \overset{d}{=} T_{2N}(b, 0) \left|_{e \sim q_N(0,t)} \right. \quad (4.4.22) \]

Then observe in (4.4.20) that an application of the translation property given in (4.4.22) to \( T_{2N}(b, 0) \) and the property in (4.4.21) to \( T_{2N}(0, 0) \) leads to

\[ \left\| T_{2N}(b, t) - T_{2N}(0, t) \right\|_{e \sim p_N} = o_p(1). \quad (4.4.23) \]

This completes the proof.

In the sequel, it is shown that the result above holds for the statistic \( S_{2N}(b, t) \). To obtain this result, it will be useful to ascertain how the underlying variables of this statistic are distributed.

### 4.4.3 Distributive Properties of Random Variables

In order to establish the distributive properties of \( S_{2N}(b, t) \), we consider the distribution of the random variables \( v_1(b, t), \ldots, v_N(b, t) \). The distribution function of these transformed variables is given in the next lemma.

**Lemma 4.4.2.** Assume that the i.i.d. random variables \( e_1, \ldots, e_N \), have a common cdf \( F(u) \).

Then, for \( i = 1, \ldots, N \), the random variables:
(i) \( e_i - m_i \), the cdf is \( G_i(u) \), where

\[
G_i(u) = P(e_i - m_i \leq u) \quad \text{for any } u
\]
\[
= P(e_i \leq u + m_i)
\]
\[
= F(u + m_i);
\]

(ii) \(|e_i - m_i|\), the cdf is \( H_i(u) \), where

\[
H_i(u) = P(|e_i - m_i| \leq u) \quad \text{for } u > 0
\]
\[
= P(-u \leq e_i - m_i \leq u)
\]
\[
= G_i(u) - G_i(-u).
\]
\[
= F(u + m_i) - F(-u + m_i);
\]

(iii) \( \log |e_i - m_i| \), the cdf is \( H_{3i}(u) \), where

\[
H_{3i}(u) = P(\log |e_i - m_i| \leq u) \quad \text{for any } u
\]
\[
= P(|e_i - m_i| \leq e^u)
\]
\[
= G_i(e^u) - G_i(-e^u).
\]
\[
= F(e^u + m_i) - F(-e^u + m_i);
\]

(iv) \( v_i(b, t) = \log |e_i - m_i| - g_i \), the cdf is \( H_{2i}(u) \).
where

\[ H_{2i}(u) = P(\log |e_i - m_i| - g_i \leq u) \quad \text{for any } u \]
\[ = P(\log |e_i - m_i| \leq u + g_i) \]
\[ = H_{1i}(u + g_i) \]
\[ = G_i(e^{u+g_i}) - G_i(-e^{u+g_i}) \]
\[ = F(e^{u+g_i} + m_i) - F(-e^{u+g_i} + m_i). \]

Thus, we have the result

\[ v_i(b, t) \mid e_i \sim F \]

\[ \text{have the cdf } H_{2i}(u) = F(e^{u+g_i} + m_i) - F(-e^{u+g_i} + m_i). \]

Next, consider the distribution of the errors under the shifted process.

**Lemma 4.4.3.** Assume that the random variables \( e_1, \ldots, e_N \), follow a cdf \( P(e_i \leq a) \)
\[ = F(ae^{g_i} + m_i) = F_{1i}(a). \] Then the random variables \( v_i(0, 0) = e_i^* = \log|e_i| \) have the cdf \( H_{2i}(u) \).

**Proof.**

\[ P(e_i^* \leq u) = P(\log |e_i| \leq u) \quad \text{for any } u \]
\[ = P(e_i \leq e^u) - P(e_i \leq -e^u) \]
\[ = F_{1i}(e^u) - F_{1i}(-e^u) \]
\[ = F(e^{u+g_i} + m_i) - F(-e^{u+g_i} + m_i) \]
\[ = H_{2i}(u). \]

Thus,

\[ v_i(0, 0) \mid e_i \sim F_{1i}(a) = F(ae^{g_i} + m_i) \]

\[ \text{has the cdf } H_{2i}(u) \text{ for } i = 1, \ldots, N. \quad (4.4.24) \]
From the last lemma, we specify the joint distribution of the random variables $e_1, \ldots, e_N$, assuming that each $e_i$ follows the cdf $F(e^{\theta_i}e_i + m_i)$. Let this joint distribution be given by

$$q_N(b, t) = \prod_{i=1}^{N} f(e^{\theta_i}e_i + m_i)e^{\theta_i}. \quad (4.4.25)$$

Then, the sequence of densities $q_N(b, t)$ defined in (4.4.25) is contiguous to the sequence $p_N$. Then, the sequence of densities $q_N(b, t)$ defined in (4.4.25) is contiguous to the sequence $p_N$, due to an argument similar to that given previously. It is worth pointing out that for the scale problem, the contiguity of $q_N(b, t)$ to $p_N$ holds if the contiguity of $q_N(0, t)$ to $p_N$ holds. Presently, the latter contiguity condition has already been established above. Then, this contiguity condition can be utilized to extend the asymptotic linearity of the process $T_{2N}(b, t)$ to that $S_{2N}(b, t)$. To achieve this, we also need to ascertain the relationships between $T_{2N}(b, t)$ and $T_{2N}(0, 0)$ as well as $S_{2N}(b, t)$ and $S_{2N}(0, 0)$; these are provided in form of translation properties.

### 4.4.4 Translation Properties

Before the translation properties are furnished, we consider the distributive relationships between the underlying variables.

**Lemma 4.4.4.** For any fixed $b$ and $t$

$$v_i(b, t) \bigg|_{e_i \sim F(a)} \overset{d^*}{=} v_i(0, 0) \bigg|_{e_i \sim F(e^{\theta_i}a + m_i)}, \text{ for } i = 1, \ldots, N. \quad (4.4.26)$$

**Proof.** Since under $p_N$, $v_i(b, t)$ has the cdf $H_{2i}(u)$, due to Lemma 4.4.2 (iv), it suffices to
show that

\[ v_i(0,0) \mid_{c_i \sim F(e^{\pi^2 t_a + t_1})} \text{ has the cdf } H_{2i}(u). \]

However, this result holds on account of Lemma 4.4.3 to complete the proof. \( \square \)

Then the next translation property follows directly from this result.

**Lemma 4.4.5.** For any fixed \( b \) and \( t \),

\[ T_{2N}(b, t) \mid_{e \sim p_N} \overset{\mathcal{L}}{=} T_{2N}(0, 0) \mid_{e \sim q_N(b, t)}, \tag{4.4.27} \]

**Proof.** The left hand side of (4.4.27) is a function of random variables \( v_i(b, t) \) whereas the right hand side is a function of random variables \( v_i(0, 0) \), for \( i = 1, \ldots, N \). Since the distribution properties of the random variables extend to their corresponding statistics, an application of Lemma 4.4.4 leads to the desired result, which terminates the proof. \( \square \)

The next step is to prove that a similar translation property also holds for the empirical processes \( S_{2N}(b, t) \) and \( S_{2N}(0, 0) \).

**Lemma 4.4.6.** For any fixed \( b \) and \( t \),

\[ S_{2N}(b, t) \mid_{e \sim p_N} \overset{\mathcal{L}}{=} S_{2N}(0, 0) \mid_{e \sim q_N(b, t)}, \tag{4.4.28} \]

**Proof.** Similar to Lemma 4.4.5, the result holds in view of the fact \( S_{2N}(b, t) \mid_{e \sim p_N} \) and \( S_{2N}(0, 0) \mid_{e \sim q_N(b, t)} \) are functions of random variables \( v_i(b, t) \) and \( v_i(0, 0) \), respectively, which satisfy the relationship in Lemma 4.4.4. Thus, the proof is terminated. \( \square \)

Thus, we have obtained almost all the pieces required to extend the asymptotic linearity of the statistic \( T_{2N}(b, t) \) to the empirical process \( S_{2N}(b, t) \). It still remains to show that \( E[a^T T_{2N}(b, t)] \mid_{e \sim p_N} \to -\gamma a^T \Sigma t \) as \( N \to \infty \). This is done in the next theorem.
**Theorem 4.4.5.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold. Then

\[ E \left[ T_{2N}(b, t) \right]_{e \sim p_N} \rightarrow -\gamma_2 \Sigma t \quad \text{as } N \rightarrow \infty. \]  
(4.4.29)

**Proof.** From Theorem 4.4.4, we obtain the result

\[ \left\| T_{2N}(b, t) - T_{2N}(0, t) \right\|_{e \sim p_N} = o_p(1). \]

This, along with the fact that \( E[T_{2N}(0, t)]_{e \sim p_N} \rightarrow -\gamma_2 \Sigma t \) as \( N \rightarrow \infty \), which follows from Theorem 4.3.8, and an application of the result \( T_{2N}(0, t)_{e \sim p_N} \nless T_{2N}(0, 0)_{e \sim q_N(0, t)} \) on account of Lemma 4.4.5, leads us to the desired result. This terminates the proof. \( \square \)

In the sequel, an analogous result for the empirical process \( S_{2N}(b, t) \) is considered.

**Theorem 4.4.6.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold. Then

\[ E \left[ S_{2N}(b, t) \right]_{e \sim p_N} \rightarrow -\gamma_2 \Sigma t \quad \text{as } N \rightarrow \infty. \]  
(4.4.30)

**Proof.** Recall from Theorem 4.3.5 that

\[ \left\| S_{2N}(0, 0) - T_{2N}(0, 0) \right\|_{e \sim p_N} = o_p(1), \]  
(4.4.31)

By the contiguity of \( q_N(b, t) \) to \( p_N \), it follows that

\[ \left\| S_{2N}(0, 0) - T_{2N}(0, 0) \right\|_{e \sim q_N(b, t)} = o_p(1), \]  
(4.4.32)

From Lemma 4.4.5, we know that \( T_{2N}(0, 0)_{e \sim q_N(b, t)} \nless T_{2N}(b, t)_{e \sim p_N} \). Then it follows
that

\[
E \left[ T_{2N}(0, 0) \right] \bigg|_{e \sim q_N(b, t)} = E \left[ T_{2N}(b, t) \right] \bigg|_{e \sim p_N} \\
\rightarrow -\gamma_2 \Sigma t, \text{ as } N \rightarrow \infty,
\]

where the last convergence is due to the result in Theorem 4.4.5. In view of this result together with the fact that convergence in probability implies convergence in distribution, it is seen in (4.4.32) that

\[
E \left[ S_{2N}(0, 0) \right] \bigg|_{e \sim q_N(b, t)} \rightarrow -\gamma_2 \Sigma t \text{ as } N \rightarrow \infty. \quad (4.4.33)
\]

Then the desired result follows from the translation property

\[
S_{2N}(0, 0) \big|_{e \sim q_N(b, t)} \not\equiv S_{2N}(b, t) \big|_{e \sim p_N} \text{ given in Lemma 4.4.6, which completes the proof.} \quad \square
\]

Therefore, we have completed establishing all of the results needed to obtain the linearity result of empirical process \( S_{2N}(b, t) \). Establishing the linearity of this process is main objective of the next section.

4.4.5 Asymptotic Linearity of \( S_{2N}(b, t) \)

In this section, using contiguity of \( q_N(b, t) \) to \( p_N \) along with the translation properties given above, we establish the linearity result of \( S_{2N}(b, t) \). This result is furnished by the next theorem.

**Theorem 4.4.7.** Assume that \((D1), (D2), (F1), (F2), (S2) and (W1) hold. Then

\[
\left\| S_{2N}(b, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{e \sim p_N} = o_p(1). \quad (4.4.34)
\]
Proof. Recall that it has already been established that

$$\left\| T_{2N}(b, t) - T_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{e \sim p_N} = o_p(1), \quad (4.4.35)$$

on account of Theorem 4.4.4. Observe that, in view of Theorem 4.3.5, it is true that

$$\left\| S_{2N}(0, 0) - T_{2N}(0, 0) \right\|_{e \sim p_N} = o_p(1).$$

Then by the contiguity of the densities $q_N(b, t)$ to densities $p_N$, it follows that

$$\left\| S_{2N}(0, 0) - T_{2N}(0, 0) \right\|_{e \sim q_N(b, t)} = o_p(1).$$

Consider the left hand side of the last representation. An application of the translation properties $S_{2N}(b, t)|_{e \sim p_N} \overset{c}{=} S_{2N}(0, 0)|_{e \sim q_N(b, t)}$ given in Lemma 4.4.6, and $T_{2N}(b, t)|_{e \sim p_N} \overset{c}{=} T_{2N}(0, 0)|_{e \sim q_N(b, t)}$ given in Lemma 4.4.5, affords us the result

$$\left\| S_{2N}(b, t) - T_{2N}(b, t) \right\|_{e \sim p_N} = o_p(1).$$

An application of this result, along with the fact $\| S_{2N}(0, 0) - T_{2N}(0, 0) \|_{e \sim p_N} = o_p(1)$, on account of Theorem 4.3.5, to the left hand side of (4.4.35) leads us to the desired result. The proof is complete. \qed

The result in the previous theorem shows that the linear function $S_{2N}(0, 0) - \gamma_2 \Sigma t$ is asymptotically equivalent to the empirical process $S_{2N}(b, t)$. Hence, the linear function can be used to approximate the process. In addition, the theorem illustrates a well known fact: the asymptotic distribution of statistic for estimating the scale parameter is independent of the location parameter.
4.4.6 Asymptotic Uniform Linearity of $S_{2N}(b, t)$

In this section, we extend the asymptotic uniform linearity of the empirical process $S_{2N}(0, t)$ established in Theorem 4.4.3 to the process $S_{2N}(b, t)$. This result is given in the next theorem.

**Theorem 4.4.8.** Assume that $(D1)$, $(D2)$, $(F1)$, $(F2)$, $(S2)$ and $(W1)$ hold. Then

\[
\sup_{b \in B(\xi_1)} \left\| S_{2N}(b, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{e \sim p_N} = o_p(1). \tag{4.4.36}
\]

**Proof.** Similar to the proof of Theorem 4.4.4, it suffices to show that

\[
\sup_{b \in B(\xi_1)} \left\| S_{2N}(b, t) - S_{2N}(0, t) \right\|_{e \sim p_N} = o_p(1). \tag{4.4.37}
\]

In view of Remark 4.3.1, we can define a linear combination such that any component of the difference on the left hand side of this equality is selected. Note that we already have the result

\[
\sup_{t \in K(\xi_2)} \left\| S_{2N}(0, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{e \sim p_N} = o_p(1), \tag{4.4.38}
\]

due to Theorem 4.4.2. Then, if we select $t$ to be any fixed vector in the bounded set $K(\xi_2)$, it is enough to demonstrate that

\[
\sup_{b \in B(\xi_1)} \left\| a^T S_{2N}(b, t) - a^T S_{2N}(0, 0) \right\|_{e \sim p_N} = o_p(1). \tag{4.4.39}
\]

Recall that $B(\xi_1) = \{b : \|b\| \leq \xi_1\}$. Following Heiler and Willers (1988), by employing
standard diagonal sequence arguments, it is possible to obtain an index set $\mathcal{N}$ such that

$$a^T S_{2N}(b, t) - a^T S_{2N}(0, t) \to 0 \text{ almost surely,}$$

for all $b \in \mathcal{R}^{p+1}$. It follows that this convergence is uniform on every compact subset of $\mathcal{R}^{p+1}$. Further, the uniform convergence also holds for any bounded subsets on $B(\xi_i)$. This compact uniform convergence is equivalent to the result in (4.4.39) on account of Heiler and Willers, (1988), p. 179. Since the result holds for every component of $S_{2N}(b, t) = S_{2N}(0, t)$, the convergence of the entire vector is immediate on account of Remark 4.3.2. The proof is terminated.

Having completed establishing the previous results, we can now embark upon analyzing the asymptotic distribution of the proposed estimator in this chapter. This is the focus of the next section.

4.4.7 Application of the Linearity Result

In this section, an application of the asymptotic linearity result above is utilized to establish the asymptotic distribution of the proposed estimate of $\theta$. In addition to the regularity condition under which the result was established, it also is assumed that:

(A1) we have an estimate of $\beta$ such that $\sqrt{N}(\hat{\beta} - \beta) = O_p(1)$.

(A2) the true scale parameter is 0 without loss of generality. So assume $\theta = 0$.

Suppose we have a $\sqrt{N}$-consistent estimate of the regression parameter, $\beta$, satisfying (A1). Then, an estimate of the scale parameter $\theta$ based on log transformed and aligned responses can be obtained by minimizing the rank dispersion function that has $\hat{\beta}$ substituted for $\beta$. It is demonstrated that when transformed residuals, $\log |y_i - x_{it}^T \hat{\beta}|$, are utilized as responses instead of the transformed true errors, $\log |y_i - x_{it}^T \beta|$, which are never really
known, the asymptotic properties established above are still valid. The proof for this result follows a strategy that is similar to that used to obtain the result in Lemma A.3.12 of Hettmansperger and McKean (1998). The authors considered a special case of the unweighted linear model problem previously examined by Jurečková (1971), which was based on signed-rank process for the location problem. In the results from both of these citations, consistent, unweighted residuals were used to estimate the intercept parameter. In the present problem, the logs of the absolute, unweighted residuals are needed to obtain an estimate of the scale parameter based on a rank process which has properties that are the same as those from the location problem. Clearly, the signed-rank process mentioned above would not be suitable for the scale parameter estimation under consideration. In what follows, an analogous result for the scale problem is presented.

**Theorem 4.4.9.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold. In addition, suppose that assumptions (A1) and (A2) are satisfied.

Then, for any $t \in \mathcal{R}^p$,

$$
\mathbb{E}^{0,t} \left\{ \left\| S_{2N}(\sqrt{N}(\hat{\beta} - \beta), t) - S_{2N}(0, t) \right\|_{c_{pN}} \right\} = o_p(1).
$$

**Proof.** Let $\varepsilon > 0$ be given. Keeping $t$ as an arbitrary but fixed vector in the set $\mathcal{K}(\xi_2)$, and $\|b\| \leq \xi_1$. Observe that $\xi_1 > 0$ and $\xi_2 > 0$. From the proof of Theorem 4.4.8, note that the representation in (4.4.37) can also be written as

$$
\lim_{N \to \infty} P \left( \max_{b \in \mathcal{B}(\xi_1)} \left\| S_{2N}(b, t) - S_{2N}(0, t) \right\|_{c_{pN}} \geq \varepsilon \right) = 0.
$$

Consider the expression in $P(\cdot)$. Observe that adding and subtracting $S_{2N}(0, 0) - \gamma_2 \Sigma t$ to the expression within the $\| \cdot \|$, along with an application of the trian-
gular inequality yields

\[ \|S_{2N}(b, t) - S_{2N}(0, t)\| \leq \|S_{2N}(b, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t\| \]
\[ + \|S_{2N}(0, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t\|. \]

Then

\[
P \left( \max_{b \in \mathcal{B}(\xi_1)} \left\| S_{2N}(b, t) - S_{2N}(0, t) \right\|_{e^{-p_N}} \geq \varepsilon \right)
\leq P \left( \max_{b \in \mathcal{B}(\xi_1)} \left\| S_{2N}(b, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{e^{-p_N}} \geq \varepsilon/2 \right)
\]
\[ + P \left( \left\| S_{2N}(0, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{e^{-p_N}} \geq \varepsilon/2 \right). \]

Since \( \varepsilon > 0 \) is fixed, it can be selected to be arbitrarily small. Consider the right hand side of the last inequality. Observe that the first term, \( \|S_{2N}(b, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t\|_{e^{-p_N}} \), goes to zero in probability in view Theorem 4.4.8. Furthermore, the second term, \( S_{2N}(0, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t\|_{e^{-p_N}} \), goes to zero in probability on account of Theorem 4.4.3. Thus, for \( N \) sufficiently large, the two terms on the right hand side of the last inequality are arbitrarily small. Letting \( b^* = \sqrt{N}(\bar{\beta} - \beta) \) for \( \bar{\beta} \) satisfying condition (A1) and substituting it for \( b \) in the last expression so that applying the limit leads us to the desired result. \( \square \)

We can now use the result in the last theorem to show that the uniform asymptotic linearity of the process \( S_{2N}(\cdot, \cdot) \) still holds when \( \beta \) is replaced by \( \bar{\beta} \). This we do in the sequel.

**Theorem 4.4.10.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold. In addition, suppose that assumptions (A1) and (A2) are satisfied.

Then, for any \( t \in \mathcal{R}^p \),

\[
\sup_{t \in \mathcal{K}(\xi_2)} \left\| S_{2N}(\sqrt{N}(\bar{\beta} - \beta), t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{e^{-p_N}} = o_p(1). \tag{4.4.40}
\]
Proof. Observe that adding and subtracting terms and an application of the triangular inequality yields

\[
\left\| S_{2N}(\sqrt{N}(\beta - \beta), t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\|
\leq \left\| S_{2N}(\sqrt{N}(\beta - \beta), t) - S_{2N}(0, 0) \right\| + \left\| S_{2N}(0, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\|
\]

Then, for any \( \epsilon > 0 \),

\[
P\left( \max_{t \in K(\xi_2)} \left\| S_{2N}(\sqrt{N}(\beta - \beta), t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\| \geq \epsilon \right) 
\leq P\left( \max_{t \in K(\xi_2)} \left\| S_{2N}(\sqrt{N}(\beta - \beta), t) - S_{2N}(0, t) \right\| \geq \epsilon/2 \right) 
+ P\left( \max_{t \in K(\xi_2)} \left\| S_{2N}(0, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\| \geq \epsilon/2 \right)
\]

Since \( \epsilon > 0 \) is fixed, it can be selected to be arbitrarily small. Consider the right hand side of the last inequality. On account of Theorem 4.4.9 along with the fact that \( \|T_{2N}(b, 0) - T_{2N}(0, 0)\|_{e-p_N} \) goes to zero in probability obtained from the proof of Theorem 4.4.4, it seen that both the first term of the right hand side of the last inequality hold true. The second term \( \max_{t \in K(\xi_2)} \|S_{2N}(0, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t\|_{e-p_N} \) goes to zero in probability in view of Theorem 4.4.3. Thus, for \( N \) sufficiently large, the two terms on the right hand side of the last inequality are arbitrarily small. Applying the limit, as \( N \to \infty \), to both sides of the inequality, right hand sides goes to zero which leads us to the desired result. This completes the proof.

Now under a certain condition on the stochastic nature of the bound on \( t^* \) shown below, we can let \( t^* = \gamma_2^{-1}\Sigma^{-1}S_{2N}(0, 0) \) and substitute it for \( t \) in the result of Theorem 4.4.8 so that

\[
\left\| S_{2N}(\sqrt{N}(\beta - \beta), t^*) \right\|_{e-p_N} = o_p(1)
\]
is immediate. These results suggest that the estimate of $\theta$ that we seek solves

$$S_{2N}(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N} \hat{\theta}) = 0. \quad (4.4.41)$$

Observe that the empirical process $S_{2N}(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N} \hat{\theta})$ is a function of the empirical distribution of the estimated residuals $v_i(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N} \hat{\theta}) = \log |y_i - x_i^T \hat{\theta}| - x_i^T \hat{\theta}$ for $i = 1, \ldots, N$. If the result in (4.4.41) holds then, by the asymptotic equivalence of this solution to that obtained as minimization of the corresponding dispersion function $D_1(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N}(\hat{\theta} - \theta))$ to be given later, we have obtained a fundamental result of the application. That is, this gives an alternative way to obtaining an estimate of $\theta$ when $\beta$ has been replaced with $\hat{\beta}$ satisfying condition (A1).

We wish to demonstrate that when the estimating process is evaluated at these estimates, it goes to zero in probability. This is demonstrated in the next corollary.

**Corollary 4.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold. In addition, suppose that conditions (A1) and (A2) are satisfied. Then

$$\left\| S_{2N}(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N} \hat{\theta}) \right\|_{e \sim p_N} = o_p(1). \quad (4.4.42)$$

**Proof.** Let $t^* = \gamma_2^{-1} \Sigma^{-1} S_{2N}(0, 0)$ for $t^* \in \mathbb{R}^p$. Since $S_{2N}(0, 0)_{e \sim p_N} \xrightarrow{p} N_p(0, \Sigma)$, in view of Theorem 4.3.4, it implies that $S_{2N}(0, 0)_{e \sim p_N}$ is bounded in probability. Thus, it follows that $t^*$ is bounded in probability, since $\gamma_2$ and $\Sigma$ are constants. Recall from Theorem 4.4.8 that

$$\sup_{b \in B(\zeta_1)} \left\| S_{2N}(b, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{e \sim p_N} = o_p(1). \quad (4.4.43)$$
Then, substituting \( t^* \) for \( t \) in (4.4.43), it is seen that

\[
\left\| S_{2N}(b, t^*) \right\|_{e \sim \mathcal{P}_N} = o_p(1).
\]

Thus, the desired result follows from this fact and the definition of the estimate of \( \theta \) given in (4.4.41). The proof is terminated.

Since the linear function \( S_{2N}(0, 0) - \gamma_2 \Sigma t \) can be used to approximate the process \( S_{2N}(b, t) \), along with the fact that \( \| S_{2N}(\sqrt{N}(\hat{\beta} - \beta), \sqrt{N}\hat{\theta}) \| = o_p(1) \), on account of Corollary 4, it follows that \( \| \gamma_2 \Sigma \sqrt{N}\hat{\theta} - S_{2N}(0, 0) \|_{e \sim \mathcal{P}_N} = o_p(1) \). Thus,

\[
\left\| \sqrt{N}\hat{\theta} - \gamma_2^{-1}\Sigma^{-1}S_{2N}(0, 0) \right\|_{e \sim \mathcal{P}_N} = o_p(1).
\] (4.4.44)

Since, \( S_{2N}(0, 0)|_{e \sim \mathcal{P}_N} \xrightarrow{\mathcal{L}} N_p(0, \Sigma) \), in light of Theorem 4.3.4, it follows that

\[
\gamma_2^{-1}\Sigma^{-1}S_{2N}(0, 0)|_{e \sim \mathcal{P}_N} \xrightarrow{\mathcal{D}} N_p(0, \gamma_2^{-2}\Sigma^{-1}).
\]

When this result along with the fact that convergence in probability implies convergence in distribution are applied to (4.4.44), the next theorem is immediate.

**Theorem 4.4.11.** Assume that (D1), (D2), (F1), (S2) and (W1) hold. In addition, suppose that conditions (A1) and (A2) are satisfied.

Then

\[
\sqrt{N}\hat{\theta} \xrightarrow{\mathcal{D}} N_p(0, \gamma_2^{-2}\Sigma^{-1}).
\] (4.4.45)

It is of interest to obtain a more general result for the case where \( \theta \neq 0 \). Since \( S_{2N}(b, t)|_{e \sim \mathcal{P}_N} \xrightarrow{\mathcal{L}} S_{2N}(0, 0)|_{e \sim \mathcal{Q}_N(b, t)} \), in view of Lemma 4.4.6, it follows from this property of \( S_{2N}(b, t) \) that the estimate \( \hat{\theta} \) also possesses a translation invariance property. Then the result in the next lemma follows from the property.
Lemma 4.4.7. Assume that (D1), (D2), (F1), (S2) and (W1) hold. In addition, suppose assumption (A1) is satisfied. Then

\[ \sqrt{N}(\hat{\theta} - \theta) \overset{\mathcal{L}}{\to} N_p(0, \gamma_2^{-2}(X^T X)^{-1}). \]

Thus, it is clear that the next result easily follows from this discussion in view of the asymptotic normality result \( \sqrt{N}\hat{\theta} \overset{\mathcal{D}}{\to} N_p(0, \gamma_2^{-2}\Sigma^{-1}) \) given in Theorem 4.4.11.

Theorem 4.4.12. Assume that (D1), (D2), (F1), (S2) and (W1) hold. In addition, suppose assumption (A1) is satisfied. Then

\[ \hat{\theta} \overset{\mathcal{D}}{\to} N_p(\theta, \gamma_2^{-2}(X^T X)^{-1}). \] (4.4.46)

It is seen that a general result for the limiting distribution of \( \hat{\theta} \) is afforded by the last theorem. That is, \( \hat{\theta} \) follows an asymptotic distribution that is a \( p \)-variate normal with mean \( \theta \) and variance \( \gamma_2^{-2}(X^T X)^{-1} \). As a consequence of this theorem, it is true that \( \hat{\theta} \) is consistent for \( \theta \).

In addition, it is of interest to show that the proposed estimate of \( \theta \) is asymptotically equivalent to the estimate of \( \theta \) for the case when the true value of the regression coefficient is specified. The latter shall be denoted by \( \hat{\theta}_{loc} \). For the standard homoscedastic linear model fitting problem, Jurečková (1977) established that \( \sqrt{N}(\hat{\beta} - \beta) = O_p(1) \) in her Lemma 5.2. In what follows, the lemma’s analog in the present problem of estimating \( \theta \) is given. We now restate the lemma without proof.

Lemma 4.4.8. (Lemma 5.2: Jurečková (1977)) Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold. In addition, suppose that condition (A2) are satisfied.
Then, to any $\epsilon > 0$, correspond $\xi_3 > 0$, $\eta > 0$, and a positive integer $N_0$ such that

$$P\left\{ \min_{\|t\| \geq \xi_3} \left\| S_{2N}(0, t) \right\|_{e \sim p_N} < \eta \right\} < \epsilon$$

(4.4.47)

holds for $N > N_0$.

Then in what follows, utilizing this result, we furnish the application that establishes the asymptotic equivalence of $\hat{\theta}$ and $\hat{\theta}_{loc}$.

**Theorem 4.4.13.** Assume that (D1), (D2), (F1), (F2), (S2) and (W1) hold. In addition, suppose that conditions (A1) and (A2) are satisfied.

Then,

$$\sqrt{N}(\hat{\theta} - \hat{\theta}_{loc}) \xrightarrow{P} 0.$$  

(4.4.48)

**Proof.** Suppose that

$$\sqrt{N}\hat{\theta} = O_p(1)$$

(4.4.49)

is true for the moment. Then, from Theorem 4.4.8 and Corollary 4, it can be seen that

$$\left\| \sqrt{N}\hat{\theta} - \gamma_2^{-1}\Sigma^{-1}S_{2N}(0, 0) \right\|_{e \sim p_N} = O_p(1).$$

(4.4.50)

Note that it can easily be shown that

$$\left\| \sqrt{N}\hat{\theta}_{loc} - \gamma_2^{-1}\Sigma^{-1}S_{2N}(0, 0) \right\|_{e \sim p_N} = O_p(1),$$

(4.4.51)

by employing the standard asymptotic linearity argument of Jurečková (1971), so that the desired result holds. It has yet to be shown that the convergence given in (4.4.50) is valid. However, in view of Corollary 4, the result holds if for each $\eta > 0$ and $\epsilon > 0$ and $\xi_1$, there
exists $\xi_3$ satisfying

$$P\left\{ \inf_{b \in B(\xi_3)} \left\| S_{2N}(b, t) \right\|_{e^{-p_N}} > \eta \right\} > 1 - \epsilon,$$

where $K^*(\xi_3) = \{ t \in \mathcal{R}^p : \|t\| \geq \xi_3 \}$ is exterior to the $(\xi_3)$-ball centered at 0 for $\xi_3 > 0$. This result can be achieved by using arguments that are similar to those employed to establish the result in Lemma 4.4.8 (Lemma 5.2 of Jurečková, 1977). This completes the proof. \( \square \)

Before concluding this chapter, a description of the dispersion function is presented. This is a function of the residuals whose minimization problem is equivalent to the one considered when the linear rank statistic $S_{2N}(b, t)$ is utilized to obtain an estimate of the scale parameter, $\theta$. This is the focus of the next section.

### 4.5 Dispersion Function Criterion

In this section, we present an extension to the analogously defined objective function $D_{2N}(b, t)$ of shifted errors and this function is given below. In particular, it is demonstrated that asymptotic uniform linearity and asymptotic uniform quadraticity are equivalent conditions. As a consequence of this equivalence, it is shown that minimizers of $D_{2N}(b, t)$ and its quadratic approximation are asymptotically equivalent. It is seen that given the regression parameter satisfying (A1), the estimate of $\theta$ is such that $D_{2N}(b, t)$ reduces to the dispersion function of residuals. It turns out that it is much simpler to obtain the estimate of $\theta$ as a minimizer of the dispersion function than solving for the roots of $S_{2N}(b, t)$, as will be seen below.

First, we recall from chapter one the idea behind the dispersion function. A dispersion function is a suitably defined function of residuals that is minimized in pursuit of an estimate of the parameter of interest. The rank dispersion function is a sum of weighted residuals, where the weights are some suitably defined scores that are based on the ranks of the residuals. Let $b = (b_0, b_1^T) \in \mathcal{R}^{p+1}$. Then, for model given in (4.2.4), the estimate of
that we seek is such that for a suitably defined and given \( b \), there is a \( t \) that minimizes the function

\[
D(b, t) := \sum_{i=1}^{N} \phi_{\sigma}^2 \left( \frac{R[\log |e_i - x_{1i}^T b| - x_i^T t]}{N + 1} \right) (\log |e_i - x_{1i}^T b| - x_i^T t),
\]

(4.5.1)

where \( R[\log |e_i - x_{1i}^T b| - x_i^T t] \) is the rank of the \( \log |e_i - x_{1i}^T b| - x_i^T t \) amongst \( \log |e_k - x_{1k}^T b| - x_k^T t, \) for \( k = 1, \ldots, N \), and the score generating function \( \phi_{\sigma}^2(u) \) that satisfies assumption (S2).

The objective function given in (4.5.1), is nonnegative, piecewise linear and convex in \( t \). In general, the solution to the minimization problem is not unique. However, as remarked by Sievers (1983), "under mild conditions, the diameter of the set of solutions goes to zero asymptotically". At the point at which the function is minimized for a given \( b \), the partial derivatives of \( D(b, t) \) with respect to \( t \) should be approximately zero. Except at finite points, the partial derivatives of \( D(b, t) \) exist almost everywhere and are given by

\[
\frac{\partial D(b, t)}{\partial t_j} = -\sum_{i=1}^{N} \phi_{\sigma}^2 \left( \frac{R[\log |e_i - x_{1i}^T b| - x_i^T t]}{N + 1} \right) x_{ij} \quad \text{for } j = 1, \ldots, p.
\]

Note that \( S(b, t) = -\partial D(b, t)/\partial t_1, \ldots, -\partial D(b, t)/\partial t_p \) define the components of the gradient of the function that is being considered here.

Observe that \(-\nabla D(b, t)\) is a special case of the linear rank statistic \( S_{2N}(b, t) \) defined in (4.2.10). Recall that the linear function \( S_{2N}(0, 0) - \gamma_2 \Sigma t \) can serve as a suitable approximation of \( S_{2N}(b, t) \) in the asymptotic sense. Following Jaeckel (1972), a quadratic function can be used to approximate the statistic \( D_{2N}(b, t) \) is defined below. A description of the suitably defined quadratic function and its related asymptotic properties are presented in the next section.
4.5.1 Asymptotic Uniform Quadraticity of $D_{2N}(b, t)$

In this section, we define the objective function, $D_{2N}(b, t)$, that is used to obtain an estimate of $\theta$ in this study. In addition, another quadratic function, $Q_{2N}(b, t)$, that is obtained by performing a first order expansion of $D_{2N}(b, t)$ is also given. An asymptotic uniform quadratic result is established from these two functions. It is demonstrated that the asymptotic uniform quadraticity condition is equivalent to asymptotic uniform linearity condition obtained above. Furthermore, it is shown that the estimate that is obtained as a minimizer of the quadratic function, is asymptotically equivalent to the minimizer of the objective function.

Recall that the statistic $S_{2N}(b, t)$ is function of the perturbed errors \( \{v_i(b, t) = \log |e_i - m_i| - g_i : i = 1, \ldots, N\} \) were utilized. Then, to match with the empirical process, $S_{2N}(b, t)$, from which a $\sqrt{N}$-consistent estimate is obtained, define the working objective function as

\[
\phi^*(b, t) = \sum_{i=1}^{N} \phi_2^* \left( \frac{R|\log |e_i - m_i| - g_i|}{N+1} \right) \left( \log |e_i - m_i| - g_i \right).
\]

Let

\[
Q_{2N}(b, t) := -\gamma^2 t^T \Sigma t(1/2) + t^T S_{2N}(0, 0) + D_{2N}(b, 0),
\]

be a convex function that is quadratic in $t$. This function will serve as an approximation of $D_{2N}(b, t)$. Then, following Heiler and Willers (1988), utilizing standard diagonal sequence arguments, we obtain the asymptotic uniform quadraticity which is given in the next theorem.

**Theorem 4.5.1.** Assume that (D1), (D2), (F1), (S2) and (W1) hold.
Then, for any $\varepsilon > 0$,

$$
\lim_{N \to \infty} P \left( \sup_{b \in B(\xi_1), \, \, t \in C(\xi_2)} \left| Q_{2N}(b, t) - D_{2N}(b, t) \right|_{e_{\sim p_N}} \geq \varepsilon \right) = 0. \quad (4.5.4)
$$

**Proof.** The proof rests on the fact that, for a given $b$, the functions $D_{2N}(b, t)$ and $Q_{2N}(b, t)$ are both proper convex with respect to all $t \in \mathcal{R}^p$. Further, the gradients for these functions are given by

$$
\nabla D_{2N}(b, t) = -S_{2N}(b, t)
$$

and

$$
\nabla Q_{2N}(b, t) = -\gamma_2 \Sigma t + S_{2N}(0, 0),
$$

respectively. Then it is seen that putting these results together, we obtain

$$
\nabla \left( D_{2N}(b, t) - Q_{2N}(b, t) \right) = - \left[ S_{2N}(b, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right]. \quad (4.5.5)
$$

There exists a diagonal infinite index set of nested sequences, denoted by $\tilde{N}$. Following the method in Heiler and Willers (1988), p. 179), by the utilization of standard diagonal arguments, it can be shown that

$$
\nabla \left( D_{2N}(b, t) - Q_{2N}(b, t) \right) \to 0, \quad \text{almost surely},
$$

$$
S_{2N}(b, t) - (S_{2N}(0, 0) + \gamma_2 \Sigma t) \to 0, \quad \text{almost surely},
$$

where both convergence results are valid for $N \in \tilde{N}$ and uniformly on $C = \{ b \in B(\xi_1), \, t \in \mathcal{R}^p \}$. 
Since \( \mathcal{N} \) is arbitrary, it is seen that

\[
S_{2N}(b, t) - (S_{2N}(0, 0) + \gamma_2 \Sigma t) \xrightarrow{p} 0, \text{ uniformly on } \mathcal{C}.
\]  

(4.5.6)

Furthermore, from the uniform convergence results, we obtain

\[
\lim_{N \to \infty} P \left( \sup_{t \in \mathcal{N}(\xi_2)} \left\| S_{2N}(b, t) - S_{2N}(0, 0) + \gamma_2 \Sigma t \right\|_{\# \mathbb{P}_N} \geq \varepsilon \right) = 0,
\]  

(4.5.7)

and

\[
\lim_{N \to \infty} P \left( \sup_{t \in \mathcal{N}(\xi_2)} \left\| Q_N(b, t) - D_{2N}(b, t) \right\|_{\# \mathbb{P}_N} \geq \varepsilon \right) = 0.
\]  

(4.5.8)

Thus, the proof is complete.

Although \( D_{2N}(b, t) \) and \( Q_{2N}(b, t) \) are asymptotically equivalent, the approximation, \( Q_{2N}(b, t) \) cannot be used for estimating \( \theta \) because its minimum depends on unspecified quantities \( \gamma_2 \) and \( \theta \). However, the minimum provides us with the asymptotic normality result upon which the limiting distribution the estimator, the minimizer of \( D_{2N}(b, t) \), depends. It should be noted that besides showing that \( D_{2N}(b, t) \) and \( Q_{2N}(b, t) \) approach each other as \( N \) tends to be large, the result also demonstrates that asymptotic uniform linearity (4.5.7) and asymptotic uniform quadraticity (4.5.8) are equivalent.

Therefore, it follows from the argument above that an estimate of \( \theta \) that is being sought can also be obtained as a solution to the minimization problem of the objective function given in (4.5.2). Through a suitable (IRWLS) formulation of the rank dispersion function, the minimizer is obtained in a simple manner following Sievers and Abebe (2004).
4.6 Conclusion

In this chapter, an estimator of the scale parameter, \( \theta \), based on a log transformation, was proposed. Due to the transformation, it was anticipated that the estimator would enjoy the properties of estimators of the parameter in the standard linear model where location scores are employed. Consequently, it has been demonstrated that the proposed estimator attains the asymptotic normality and is consistent for \( \theta \). In addition, the proposed estimator for \( \theta \) is asymptotically equivalent to the optimal solution, that is, its asymptotic variance is the same as that which would be obtained if the regression parameters \( \beta \) were known. Since the estimator is based on rank type estimation, it provides us with estimates of \( \theta \) that are resistant to outlying responses, hence making the estimates robust. Therefore, employing the log transformation when heteroscedasticity is assumed to be of the form \( \sigma_i = exp(x_i^T \theta) \) affords us an alternative route for developing a complete inference theory for \( \theta \). Presently, only the estimation component has been established in this chapter.

Having completed establishing useful results the asymptotic theory for \( \beta \) in Chapter 3 and \( \theta \) in the current chapter, it is appealing to analyze the iterative nature of the estimation process for these parameters. Describing the iterative scheme of this study is the focus of the next chapter. \( E[sgn(Y_i^*)] = 0 \)
CHAPTER V

IMPLEMENTATION OF THE ITERATIVE METHOD

5.1 Introduction

In this chapter, the main results of chapter three and chapter four are implemented. Consequently, the estimation process for obtaining the desired estimates in a cohesive framework is presented in form of a single algorithm.

This chapter has a four-fold purpose: (1) review of score functions; (2) describe estimates of the dispersion parameter for each process; (3) present IRWLS formulations of the objection function utilized to compute the estimates from each process; and (4) prescribe an algorithm for iteratively obtaining the estimates.

5.2 Score Functions

In pursuit of robust rank estimators, several score functions for the location problem are considered. The suitability of a particular set of scores for analyzing a data set depends on the nature of the underlying distribution. In the absence of knowledge of this true nature, suitable scores computed from the data following the recommendations in McKean and Sievers (1989). Although the formal strategy for determining the most suitable scores that the authors prescribed was designed for the homoscedastic linear models, it is applicable to the scale problem under consideration since the underlying variance function has been linearized, and the resulting error terms are homoscedastic. Let us first consider the estimation of the location parameter $\beta$ under the symmetry condition of the errors. Recall that under this condition, signed-rank scores are the most appropriate.

If the errors come from a normal distribution, then the optimal scores for estimat-
ing $\beta$, are the standardized signed-rank normal scores are given by

$$\phi^+_1(u) = \Phi^{-1}\left(\frac{1+u}{2}\right), \quad 0 < u < 1,$$

(5.2.1)

If the errors come from the logistic distribution, signed-rank Wilcoxon scores

$$\phi^+_1(u) = \sqrt{3}(u), \quad 0 < u < 1,$$

(5.2.2)

are optimal.

If the errors follow the Laplace (double exponential) distribution, then the optimal scores are the signed-rank sign scores defined by

$$\phi^+_1(u) = 1, \quad 0 < u < 1,$$

(5.2.3)

The score generating functions of the three aforementioned distributions are depicted in Figure 1.

It is worth noting that the IRWLS routine that is used in this study requires that scores be distinct values. Clearly, this condition is not satisfied by the signed-rank sign scores. Without any modification to these scores, different scores that have at least two values have to be employed, albeit, they would be less than optimal.

Next, we consider the scale estimation problem. Recall from chapter two that the upshot of the linearization of this study is that the suitable scores are generated by function that is non-decreasing, square-integrable and defined the on interval $(0,1)$, as it should be expected in location type problem. Examples of these score generating functions are given next.

If the errors follow the normal distribution, then the optimal scores for estimating
\( \theta \), are the rank normal scores for the scale problem given by
\[
\phi_2^*(u) = \left( \Phi^{-1} \left( \frac{1 + u}{2} \right) \right)^2 - 1, \quad 0 < u < 1, \tag{5.2.4}
\]

If the errors come from the logistic distribution, the optimal scale are given by
\[
\phi_2^*(u) = u \cdot \log \left( \frac{u + 1}{1 - u} \right) - 1, \quad 0 < u < 1, \tag{5.2.5}
\]

If the errors follow the Laplace distribution then the optimal scores are the rank sign scores for the scale defined as
\[
\phi_2^*(u) = -\log(1 - |u|) - 1. \tag{5.2.6}
\]

The graphs for the optimal score functions for the scale condition to the transformation of this study are presented in Figure 2. Observe that all three of them are non-
decreasing as would be anticipated for location problems.

For the two sample problem, Fligner and Killeen (1976) developed the test statistic that was based on the scores \( \phi_2(u) = (\Phi^{-1}(\frac{1+u}{2}))^2 \). It was demonstrated in Conover and Johnson (1976) that, under less restrictive conditions of non-normal distributions, the Fligner-Killeen test statistic outperformed the other more widely used tests of homogeneity of variance. In the sequel, we exploit the centered form of these scores to obtain an estimate of \( \theta \) following the findings in Conover and Johnson (1976) as our motivation. Hetmansperger and McKean used these scores to estimate \( \theta \) in a two sample problem assuming \( \beta \) is known. In this study, the true value of \( \beta \) was assumed to be unknown treating the two sample case as a special case.

Next, the estimates of \( \gamma_1 \) and \( \gamma_2 \) are presented. As seen in each of the covariance matrices listed in any of the results above, \( \gamma_1 \) and \( \gamma_2 \) are crucial components in the computation of standard errors of \( \beta \) and \( \theta \). The estimation of these dispersion parameters is presented first before proceeding to the main estimating procedure.

Figure 2. Plots of Score Function for the Scale Problem under Symmetric Distributions: (a) Normal, (b) Logistic, and (c) Laplace.
5.3 Estimation of the Dispersion Parameters

In this study, we use the robust estimator proposed by Koul, Sievers and McKean (1987), which shall be referred to as the KSM estimator is suitable for bounded score functions and it is a density-type estimator which is based on residuals. The KSM estimator of $\gamma$ is consistent for symmetric and asymmetric error distributions. Furthermore, it is uniformly consistent for $\gamma$. It is also worth noting that, in view of the transformation procedures that are invoked for the estimation of both regression and scale parameters, the resulting error terms have constant variance over all observed responses. To see this, recall that the model being considered is

$$\log |y_i - x_{i1}^T \beta| = x^T \theta + e_i^*, \quad i = 1, \ldots, N.$$ 

It is seen from the error terms of the suitably transformed model of both problems being considered that they were free of heteroscedasticity. Hence, consistent residuals from each model should suffice to obtain estimates for $\gamma_1$ and $\gamma_2$.

5.3.1 Estimation of $\gamma_1$

Recall that for the regression parameter estimates the model being considered is

$$y_i^* = x_{i1}^T \hat{\beta} + e_i, \quad i = 1, \ldots, N. \quad (5.3.1)$$

For the location problem, the residuals are given by

$$\hat{e}_i = (y_i - x_{i1}^T \hat{\beta})/\hat{\sigma}_i \quad i = 1, \ldots, N.$$
Following definitions in Koul, et.al. (1987), we let $w_{N,\alpha}$ be the $\alpha^{th}$ quantile of $J_{1N}$, where

$$J_{1N}(y) = \frac{1}{N} \sum_{i=1}^{N} \sum_{l=1}^{N} \left\{ \phi_{\alpha}^+(\frac{l}{N}) - \phi_{\alpha}^+(\frac{l-1}{N}) \right\} I\left( |\hat{e}_{(i)}| - |\hat{e}_{(i)}| \leq y \right),$$

where $|\hat{e}_{(1)}| \leq \ldots \leq |\hat{e}_{(N)}|$ are ordered absolute residuals, and $I(A)$ is the indicator for the event $A$. Then

$$\hat{\gamma}_{1,N,\alpha} = \frac{\left( \phi_{\alpha}^+(1) - \phi_{\alpha}^+(0) \right) J_{1N}(w_{N,\alpha}/\sqrt{N})}{(2w_{N,\alpha}/\sqrt{N})}, \quad 0 < \alpha < 1,$$

is the KSM estimator of $\gamma_1$. Furthermore,

$$\hat{\tau}_{\phi_{\alpha}^+} = \sqrt{\frac{N}{N - p}} \hat{\gamma}_{1,N,\alpha}^{-1},$$

is the consistent estimator of $\tau_{\phi_{\alpha}^+}$, the analogue to least squares dispersion estimator, $\hat{\sigma}_{LS}$ for the model in (5.3.1).

### 5.3.2 Estimation of $\gamma_2$

Consider the scale parameter estimates, and recall that the model being considered is

$$\log |y_i - x_{1i}^T \beta| = x_i^T \theta + e_i^*, \quad i = 1, \ldots, N. \quad (5.3.2)$$

For the scale problem, the residuals are defined as

$$\hat{e}_i^* = \log |y_i - x_{1i}^T \hat{\beta}| - x_i^T \hat{\theta},$$
Similar to the location problem, we let \( w_{N,\alpha} \) be the \( \alpha \)th quantile of \( J_{2N} \), where

\[
J_{2N}(y) = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{N} \left\{ \phi_2^* \left( \frac{1}{N} \right) - \phi_2^* \left( \frac{1-1}{N} \right) \right\} I \left( \left| \hat{e}_{(i)}^* - \hat{e}_{(t)}^* \right| \leq y \),
\]

where \( \hat{e}_{(1)}^* \leq \ldots \leq \hat{e}_{(N)}^* \) are ordered residuals, and \( I(A) \) is the indicator for the event \( A \). Then

\[
\hat{\gamma}_{2,N,\alpha} = \frac{\phi_2^*(1) - \phi_2^*(0) J_{2N}(w_{N,\alpha}/\sqrt{N})}{2 w_{N,\alpha}/\sqrt{N}}, \quad 0 < \alpha < 1,
\]

is the KSM estimator of \( \gamma_2 \). Furthermore,

\[
\hat{\tau}_{\phi_2} = \sqrt{\frac{N}{N-p-1} \hat{\gamma}_{2,N,\alpha}^{-1}}
\]

is the consistent estimator of \( \tau_{\phi_2} \), the analogue to least squares dispersion estimator, \( \hat{\sigma}_{LS} \) for the model in (5.3.2).

It should be noted that Koul, Sievers, and McKean (1987), Sievers and Abebe (2004) recommended values of \( \alpha \) to be such that

\[
\alpha = \begin{cases} 
0.80 & \text{if } N > 5p, \\
0.90 & \text{otherwise.}
\end{cases} \quad (5.3.3)
\]

Here, \( p \) is the dimension for the full model. This concludes the estimation of the dispersion parameters. In what follows, the iterative procedure for obtaining an estimate of \( \beta \) and \( \theta \) is presented.

### 5.4 Iterative Estimation of \( \beta \) and \( \theta \)

In this section, we present the iterative method that is used to obtain estimates of \( \beta \) and \( \theta \). Recall that it is the objective of this study that at each iteration, obtaining these es-
timates is achieved by employing an Iterated Reweighted Least Squares (IRWLS) method in the spirit of Sievers and Abebe (2004). It is worth noting that the authors considered the problem of minimizing the regular rank dispersion function in pursuit of an estimate of the regression coefficient in a homoscedastic linear model. For this problem, they proposed a weighted least squares formulation resulting in an iterative procedure.

In this study, we employ this method when minimizing the objective functions of both regression coefficients as well as scale parameters. This notwithstanding, the method is still applicable since each function is minimized with respect to one parameter assuming an estimate of the other parameter is available as was shown in chapter three and four. Moreover the residuals in each estimation problem consistent for the respective true errors. Thus, in this regard, the dispersion functions are well defined.

Since the rank based estimation problem is formulated as an (IRWLS) problem, the method is simple to use and can be implemented easily by users from any computer. This is due to the fact that it utilizes routines that are available in standard statistical packages including R, which is also free.

The section proceeds with a general overview of the procedure used in this investigation. The IRWLS techniques for computing the rank estimate of each parameter are described in Section 5.4.2 and Section 5.4.3. Finally these components are combined to form the iterative scheme for both parameters in Section 5.4.5, which also includes the algorithm which has been prescribe for the proposed method.

5.4.1 Iterative Structure of the Procedure

The iterative method of this study cycles back and forth between computing estimates of $\beta$ and $\theta$. We let $r$ denote the number of iterations the method computes the desired pair $(\hat{\beta}, \hat{\theta})$. Note that in each iteration, the estimation of $\theta$ is preceded by that $\beta$. Further, to compute an estimate of each parameter in each iteration, a step-estimator is employed. Thus, within the $r$th iteration, an estimate of $\beta$ is computed using the IRWLS
technique in $k_r$ steps to produce $\hat{\beta}^{(r)}$ upon convergence of the step estimates. Similarly, within the $r$th iteration, an estimate of $\theta$ is computed using IRWLS technique in $l_r$ steps to yield $\hat{\theta}^{(r)}$ upon convergence of the step estimates. It should be noted that in the $r$th iteration, for all values of $k$, the $k$-step estimates of $\beta$ are computed using the estimate of $\theta$ from the previous iteration, $\hat{\theta}^{(r-1)}$. Analogously, for all values of $l$, the $l$-step estimate of $\theta$ in the $r$th iteration, is computed using the estimate of $\beta$ from the current iteration, $\hat{\beta}^{(r)}$.

5.4.2 Iterative Weighted Estimation of $\beta$

Consider the regression coefficient estimation problem of chapter three. Recall that the estimating process was based on residuals defined as

$$z_i(b, t) = (\sigma_i(t))^{-1}\left(y_i - x_i^T \beta - \frac{1}{\sqrt{N}} x_i^T b\right), \quad i = 1, \ldots, N,$$

(5.4.1)

where $\sigma_i(t) = \exp\{x_i^T(\theta + \frac{1}{\sqrt{N}} t)\}$, $i = 1, \ldots, N$. Here, $b \in \mathbb{R}^{p+1}$ and $t \in \mathbb{R}^p$ are fixed vectors, and $x_i^T$ is the $i$th row of the matrix $X_1 = (1, X)$. Then, based on these residuals, the estimate of $\beta$ was obtained by minimizing the function

$$D_{1N}(b, t) = \sum_{i=1}^{N} \phi^+_1\left(\frac{R(|z_i(b, t)|)}{N + 1}\right) |z_i(b, t)|,$$

(5.4.2)

with respect to $b$. Observe that $\phi^+_1(u)$ is a positive valued function satisfying Assumption (S1), $R(|z_i(b, t)|)$ is the rank of the absolute residual $|z_i(b, t)|$ amongst $|z_1(b, t)|, \ldots, |z_N(b, t)|$. Next, the IRWLS formulation of this objective function is furnished.
Weighted Least Squares Formulation for Objective Function for Estimating $\beta$

If we let

$$K_{1i}(b, t) = \phi^*_1 \left( \frac{R(|z_i(b, t)|)}{N + 1} \right),$$

then the objective function given in (5.4.2) can be written as

$$D_{1N}(b, t) = \sum_{i=1}^{N} K_{1i}(b, t) |z_i(b, t)|$$

$$= \sum_{i=1}^{N} w_i(b, t) [z_i(b, t)]^2,$$

(5.4.3)

where

$$w_i(b, t) = \begin{cases} \frac{K_{1i}(b, t)}{|z_i(b, t)|} & \text{if } z_i(b, t) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

(5.4.4)

It is clear that the estimate we seek can be obtained by invoking an iterative process as suggested by the expression given in (5.4.3). Let $k_r$ denote the $k$th step for obtaining an estimate of $\beta$ in the $r$th iteration. Then, given the $k_r$-th step estimate, $b_{(k_r)}$, the $(k_r + 1)$th step estimate of $\beta$ is the minimum of the $k_r$-step dispersion

$$D_{1N}(b | b_{(k_r)}, t) := \sum_{i=1}^{N} w_i(b_{(k_r)}, t) [z_i(b, t)]^2, \quad k_r = 0, 1, \ldots,$$

(5.4.5)

where the weights $\{w_i(b_{k_r}, t) : i = 1, \ldots, N\}$, are defined in (5.4.4) with $b = b_{k_r}$. Since $K_{1i}(b, t)$ is positive valued and $|z_i(b, t)|$ is nonnegative, it is obvious that the weights are nonnegative.
5.4.3 Iterative Weighted Estimation of $\theta$

Consider the scale parameter estimation problem of chapter four. Recall that the estimating process was based on residuals were given by

$$v_i(b, t) = z_i^*(b) - g_i, \quad i = 1, \ldots, N,$$

(5.4.6)

where $z_i^*(b) = \log |e_i - \frac{1}{N}x_i^T b|$, and $g_i = \frac{1}{N}x_i^T t$ for $i = 1, \ldots, N$. Observe that $b \in \mathcal{R}^{p+1}$ and $t \in \mathcal{R}^p$ are fixed vectors as defined previously. Then, using the residuals defined in (5.4.6), the estimate of $\theta$ was obtained by minimizing the function

$$D_{2N}(b, t) = \sum_{i=1}^{N} \phi_2^*(\frac{R[v_i(b, t)]}{N+1}) v_i(b, t),$$

(5.4.7)

with respect to $t$. Observe that $\phi_2^*(u)$ is a nondecreasing function such that $\sum_{i=1}^{N} \phi_2^*(\frac{1}{N+1}) = 0$, satisfying Assumption (S2), $R[v_i(b, t)]$ is the rank of the residual $v_i(b, t)$ amongst $v_1(b, t), \ldots, v_N(b, t)$. In the sequel, the IRWLS formulation of the objective function, $D_{2N}(b, t)$ is provided.

Weighted Least Squares Formulation for Objective Function for Estimating $\theta$

In pursuit of a simple estimate, $\hat{\theta}$, we formulate the dispersion function given in (5.4.7) as a WLS problem. In recalling that the intercept parameter is not estimated by the dispersion function in (5.4.7), a centering constant for the residuals is required to develop the IRWLS scheme that follows. Let $m_v(b, t)$ be the $v$th quantile of $\{v_i(b, t) : i = 1, \ldots, N\}$, where $v$ is such that $\phi_2^*(u) < 0$ ($\phi_2^*(u) > 0$) if $u < v$ ($u > v$). This definition of the centering constant is the analogue to that recommended by Sievers and Abebe (2004) for the homoscedastic linear model. It is worth noting that when $b = 0$, then $m_v(b, t)$ is similar to that considered by the aforementioned authors. For a detailed discussion of this choice of the centering constant, their 2004 paper is a good source.
Let
\[ K_{2i}(b, t) = \phi^*_2 \left( \frac{R(v_i(b, t))}{N + 1} \right), \quad i = 1, \ldots, N. \]

Since \( \sum_{i=1}^{N} \phi^*_2 \left( \frac{i}{N+1} \right) = 0 \), the dispersion function given in (5.4.7) can be written as
\[ D_{2N}(b, t) = \sum_{i=1}^{N} K_{2i}(b, t) [v_i(b, t) - m_v(b, t)] \]
\[ = \sum_{i=1}^{N} w_i(b, t) [v_i(b, t) - m_v(b, t)]^2, \tag{5.4.8} \]
where
\[ w_i(b, t) = \begin{cases} \frac{K_{2i}(b, t)}{w_i(b, t) - m_v(b, t)} & \text{if } v_i(b, t) \neq m_v(b, t) \\ 0 & \text{otherwise} \end{cases} \tag{5.4.9} \]

Since, as a centering constant of the residuals, \( m_v(b, t) \) is defined based on \( \phi^*_2 \left( \frac{i}{N+1} \right) \) summing to zero, it is clear that the weights, \( w_i(b, t) \) are nonnegative. It is seen in (5.4.8) that the desired estimate of \( \theta \) can be obtained by the use of an iterative process. Let \( l_r \) denote the \( l \)th step for obtaining an estimate of \( \theta \) in the \( r \)th iteration. Then, given the \( l_r \)th step estimate, \( t_{l_r} \), the \( (l_r + 1) \)th step estimate of \( \theta \) is the value that minimizes the \( l_r \)-step dispersion
\[ D_{2N}(b, t | t_{l_r}) := \sum_{i=1}^{N} w_i(b, t_{l_r}) [v_i(b, t) - m_v(b, t_{l_r})]^2, \quad l_r = 0, 1, \ldots, \tag{5.4.10} \]
where the weights \( \{w_i : i = 1, \ldots, N\} \), are as defined in (5.4.9). Since the weights are nonnegative, this function is convex in \( t \).
5.4.4 Iterative-IRWLS for Procedure for Estimation $\beta$ and $\theta$

In this section, the step estimators of $\beta$ and $\theta$, respectively, given in objection functions (5.4.5) and (5.4.10) are applied in an iterative scheme constructed for joint estimation of both parameters. The formulations of the objective functions utilizes suitably defined estimates of the secondary parameter in each case. Thus, to obtain the $(k_r + 1)$-step estimate of $\beta$, the $(k_r)$-step dispersion $\{D_{1N}(b|b_{(k_r)}), t) : b_{(k_r)} = \sqrt{N}(\hat{\beta}_{(k_r)} - \beta), t = \sqrt{N}(\hat{\theta}_{(r-1)} - \theta)\}$ is minimized with respect to $b$. This is repeated until the estimates converge to a value which is taken to be $\hat{\beta}_{(r)}$. Similarly, to obtain the $(l_r + 1)$-step estimate of $\theta$, the $(l_r + 1)$-step dispersion $\{D_{2N}(b, t|t_{(l_r)}) : b = \sqrt{N}(\hat{\beta}_{(r)} - \beta), t_{(l_r)} = \sqrt{N}(\hat{\theta}_{(l_r)} - \theta)\}$ is minimized with respect to $t$. This is repeated until the estimates converge to a value which is taken to be $\hat{\theta}_{(r)}$. These formulations are presented more detailed expressions in what follows.

**Computation of $\hat{\beta}_{(r)}$**

First note that $\hat{\beta}_{(r)}$ is value to which the $(k_r + 1)$-step estimates converge. The estimate of $\beta$ in the $(k_r + 1)$th step is the value, such that, for $k_r = 0, 1, \ldots$,

$$D_{1N}\left(\sqrt{N}(\hat{\beta}_{(k_r+1)} - \beta) \bigg| \sqrt{N}(\hat{\beta}_{(k_r)} - \beta), \sqrt{N}(\hat{\theta}_{(r-1)} - \theta)\right) = \min.$$  \hspace{1cm} (5.4.11)

Equivalently, $\hat{\beta}_{(k_r+1)}$ is a value such that

$$\sum_{i=1}^{N} w_i(\sqrt{N}(\hat{\beta}_{(k_r)} - \beta), \sqrt{N}(\hat{\theta}_{(r-1)} - \theta)) z_i(\sqrt{N}(\hat{\beta}_{(k_r+1)} - \beta), \sqrt{N}(\hat{\theta}_{(r-1)} - \theta))^2 = \min$$  \hspace{1cm} (5.4.12)
where

\[
 w_i(\sqrt{N}(\hat{\beta}^{(k_r)} - \beta), \sqrt{N}(\hat{\theta}^{(r-1)} - \theta)) = \begin{cases} 
 \phi_i \left( \frac{(y_i - x_{ii}^T \hat{\beta}^{(k_r)})/\hat{\sigma}_i^{(r-1)}}{|(y_i - x_{ii}^T \hat{\beta}^{(k_r)})/\hat{\sigma}_i^{(r-1)}|} \right) & \text{if } |(y_i - x_{ii}^T \hat{\beta}^{(k_r)})/\hat{\sigma}_i^{(r-1)}| \neq 0, \\
 0 & \text{otherwise.} 
\end{cases}
\]

Then

\[
 z_i(\sqrt{N}(\hat{\beta}^{(k_r+1)} - \beta), \sqrt{N}(\hat{\theta}^{(r-1)} - \theta)) = (y_i - x_{ii}^T \hat{\beta}^{(k_r+1)})/\hat{\sigma}_i^{(r-1)},
\]

and

\[
 \hat{\sigma}_i^{(r-1)} = \exp\{x_i^T \hat{\theta}^{(r-1)}\}.
\]

**Computation of \( \hat{\theta}^{(r)} \)**

Consider the estimate, \( \hat{\theta}^{(r)} \). This is value to which the \((l_r + 1)\)-step estimates converge. The estimate of \( \theta \) in the \((l_r + 1)\)th step is the value, such that, for \( l_r = 0, 1, \ldots \),

\[
 D_{2N}\left(\sqrt{N}(\hat{\beta}^{(r)} - \beta), \sqrt{N}(\hat{\theta}^{(l_r+1)} - \theta)\Big|\sqrt{N}(\hat{\theta}^{(l_r)} - \theta)\right) = \min. \quad (5.4.13)
\]

That is, \( \hat{\theta}^{(l_r+1)} \) is a value such that

\[
 \sum_{i=1}^{N} w_i(\sqrt{N}(\hat{\beta}^{(r)} - \beta), \sqrt{N}(\hat{\theta}^{(l_r)} - \theta)) v_i(\sqrt{N}(\hat{\beta}^{(r)} - \beta), \sqrt{N}(\hat{\theta}^{(l_r+1)} - \theta)) \\
 - m_u(\sqrt{N}(\hat{\beta}^{(r)} - \beta), \sqrt{N}(\hat{\theta}^{(l_r)} - \theta))^2 = \min \quad (5.4.14)
\]
where

\[
w_i(\sqrt{N}(\hat{\beta}(r) - \beta), \sqrt{N}(\hat{\theta}(l) - \theta)) = \begin{cases} 
\phi_2^*(\log|y_i - x^T_{ii}\hat{\beta}(r)| - x^T_{ii}\hat{\theta}(l)) & \text{if } \log|y_i - x^T_{ii}\hat{\beta}(r)| - x^T_{ii}\hat{\theta}(l) \neq m_0^*(\hat{\beta}(r), \hat{\theta}(l)) \\
0 & \text{otherwise.}
\end{cases}
\]

\[
v_i(\sqrt{N}(\hat{\beta}(r) - \beta), \sqrt{N}(\hat{\theta}(l^{r+1}) - \theta)) = \log|y_i - x^T_{ii}\hat{\beta}(r)| - x^T_{ii}\hat{\theta}(l^{r+1}),
\]

and \(m_0^*(\hat{\beta}(r), \hat{\theta}(l)) = m_u(\sqrt{N}(\hat{\beta}(r) - \beta), \sqrt{N}(\hat{\theta}(l) - \theta))\) is the \(u\)th quantile of \(\log|y_i - x^T_{ii}\hat{\beta}(r)| - x^T_{ii}\hat{\theta}(l), \ldots, N\), where \(u\) is such that \(\phi_2(u) < 0(\phi_2(u) > 0)\) if \(u < v(u > v)\).

It is seen that the algorithm that follows combines both of these formulations to yield the joint estimates, \((\hat{\beta}(r), \hat{\theta}(r))\), being sought under the Iterative Rank Heteroscedastic (IRHET) method of this study.

### 5.4.5 The Algorithm for \((\hat{\beta}(r+1), \hat{\theta}(r+1))\)

In this section, the algorithm that is employed in this study in obtaining the iterative estimates of this study is described. The estimate we seek are obtained by using the following procedure:

1. Obtain initial estimates the regression coefficient and scale parameter.

   (i) Obtain a preliminary estimate of \(\beta, \hat{\beta}(0)\), so that the scale parameter is a zero vector. Observe that when \(t = -\sqrt{N}\theta, \sigma_i = \exp\{x^T_i(\theta + \frac{1}{\sqrt{N}}t)\}\bigg|_{t=-\sqrt{N}\theta} = \exp\{x^T_i(\theta - \theta)\} = 1\) for \(i = 1, \ldots, N\). Thus, the \((k_0+1)\)-step preliminary estimate
of $\beta$ is the value, $\hat{\beta}^{(k_0+1)}$ such that

$$\sum_{i=1}^{N} w_i(\sqrt{N}(\hat{\beta}^{(k_r)} - \beta), \sqrt{N}(\hat{\theta}^{(r-1)} - \theta))[z_i(\sqrt{N}(\hat{\beta}^{(k_r+1)} - \beta), -\sqrt{N}\theta)]^2 = \min$$

(5.4.15)

where

$$w_i(\sqrt{N}(\hat{\beta}^{(k_0)} - \beta), -\sqrt{N}\theta) = \left\{ \begin{array}{ll} \phi_i^*(R[|y_i - x_i^T\hat{\beta}^{(k_0)}|/(N+1)]) & \text{if } |y_i - x_i^T\hat{\beta}^{(k_0)}| \neq 0, \\ 0 & \text{otherwise}. \end{array} \right.$$ 

$$z_i(\sqrt{N}(\hat{\beta}^{(k_0+1)} - \beta), -\sqrt{N}\theta) = (y_i - x_i^T\hat{\beta}^{(k_0+1)}).$$

Repeat (i) until convergence is achieved to yield $\hat{\beta}^0$.

(ii) Obtain a preliminary estimate of $\theta, \hat{\theta}^{(0)}$. In the case where $l_0 = 0$, $t_{(0)} = -\sqrt{N}\theta$, so that the scale parameter is a zero vector. Then, the $(l_0 + 1)$-step preliminary estimate of $\theta$ is the value, $\hat{\theta}^{(l_0+1)}$, such that

$$\sum_{i=1}^{N} w_i(\sqrt{N}(\hat{\beta}^{(0)} - \beta), -\sqrt{N}\theta)[v_i(\sqrt{N}(\hat{\beta}^{(0)} - \beta), \sqrt{N}(\hat{\theta}^{(l_0+1)} - \theta))$$

$$- m_v(\sqrt{N}(\hat{\beta}^{(0)} - \beta), -\sqrt{N}\theta)]^2 = \min$$

(5.4.16)
where

\[ w_i\left(\sqrt{N}(\hat{\beta}^{(0)} - \beta), -\sqrt{N}\theta\right) = \begin{cases} \phi_2^+(\frac{R\log|y_i - x_i^T\hat{\beta}^{(0)}|}{(N+1)}) & \text{if } \log|y_i - x_i^T\hat{\beta}^{(0)}| \neq m^*_v(\hat{\beta}^{(0)}, 0), \\ 0 & \text{otherwise.} \end{cases} \]

\[ v_i(\sqrt{N}(\hat{\beta}^{(0)} - \beta), -\sqrt{N}\theta) = \log|y_i - x_i^T\hat{\beta}^{(0)}| - x_i^T\hat{\theta}^{(l_0+1)}, i = 1, \ldots, N, \]

and \( m^*_v(\hat{\beta}^{(0)}, 0) = m_v(\sqrt{N}(\hat{\beta}^{(0)} - \beta), -\sqrt{N}\theta) \) is the \( v \)th quantile of \( \log|y_i - x_i^T\hat{\beta}^{(0)}|, i = 1, \ldots, N, \)

where \( \phi_2^+(u) < 0(\phi_2^+(u) > 0) \) if \( u < u(u > u) \).

Repeat (ii) for \( l_0 = 1, 2, \ldots \) so that \( t(0) = \sqrt{N}(\hat{\theta}^{(l_0)} - \theta) \) until convergence is achieved to produce \( \hat{\theta}^0 \).

(iii) Form initial estimated scaling constants computed using \( \hat{\delta}_i^{(0)} = \exp\left(x_i^T\hat{\theta}^{(0)}\right), \)

for \( i = 1, \ldots, N. \)

(2) Obtain \( \hat{\beta} \) in the \( r \)th iteration.

(i) Compute the \((k_r + 1)\)-step estimate, \( \hat{\beta}^{(k_r+1)} \) such that

\[ D_{1N}\left(\sqrt{N}(\hat{\beta}^{(k_r+1)} - \beta)\right) = \min \]

where \( D_{1N}(\cdot) \) is as defined in (5.4.12).

(ii) Execute (i) for \( k_r = 0, 1, 2, \ldots \) until convergence is attained to yield \( \hat{\beta}^{(r)} \).

(3) Obtain \( \hat{\theta} \) in the \( r \)th iteration.
(i) Compute the \((l_r + 1)\)-step estimate, \(\hat{\theta}^{(l_r+1)}\) such that

\[
D_{2N}\left(\sqrt{N}(\hat{\beta}^{(r)} - \beta), \sqrt{N}(\hat{\theta}^{(l_r+1)} - \theta) \right) \min,
\]

where \(D_{2N}(\cdot|\cdot)\) is as defined in (5.4.14).

(ii) Execute (i) for \(l_r = 0, 1, 2, \ldots\) until convergence is attained to yield \(\hat{\theta}^{(r)}\).

(iii) Form estimated scaling constants, \(\hat{\sigma}_i^{(r)} = \exp\{x_i^T \hat{\theta}^{(r)}\}\), for \(i = 1, \ldots, N\).

(4) Repeat steps 2 and 3 until convergence is achieved. The convergence criterion is to terminate iteration when \(\|\hat{\theta}^{(r+1)} - \hat{\theta}^{(r)}\| < 0.0001\).

5.5 Conclusion

The minimization problems of the objective functions given in chapters three and four, are formulated into Iterated Reweighted Least Squares type problems. In light of the formulations, the \('k'\)-step estimators for both parameters, \(\beta\) and \(\theta\). Finally, an iterative procedure for obtaining these estimates is prescribed. In the chapters that follow, an analysis of the method presented here is performed using simulated data in order to assess how well it works under different conditions. Finally, a real life data example from clinical trials in psychiatry is utilized in the application of the estimation method that has been proposed by this investigation.
CHAPTER VI

OTHER METHODS NOT BASED ON RANKS

6.1 Introduction

In this chapter, methods used in the simulations examples that presented in the next chapter are described in detail. The methods are divided into two sections: (1) Two Sample Problem and (2) General Linear Model Problem using M-estimation.

6.2 The Two Sample Problem

Suppose \( y_{11}, \ldots, y_{1n_1} \) is a random sample from population with location, \( \mu_1 \) and variance, \( \sigma_1^2 \). Let \( y_{21}, \ldots, y_{2n_2} \) be another sample from the population with location, \( \mu_2 \) and variance, \( \sigma_2^2 \). Define the scale coefficient \( \theta \) to be such that \( e^\theta = \sigma_2 / \sigma_1 \). Assume that both populations follow the same family of distributions \( F \), with a symmetric density function \( f \). In this problem, we seek an estimate of the difference in location, \( \mu_2 - \mu_1 \) and its corresponding standard error.

6.2.1 Estimation of Difference in Two Means, \( \mu_2 - \mu_1 \)

A more likely situation than the one assumed in pooled variance method is that the population variance are not equal. Suppose, we have a random sample \( y_{11}, \ldots, y_{1N_1} \), from population 1 and \( y_{21}, \ldots, y_{2N_2} \), from population 2. Let \( \mu_1 \) and \( \sigma_1^2 \) be the mean and variance of the first population and \( \mu_2 \) and \( \sigma_2^2 \) be the mean and variance of the second population. Since \( \sigma_1 \neq \sigma_2 \) is not satisfied, the pooled variance is not valid. Thus, it common to use the
standard error of the estimate of difference $\mu_2 - \mu_1$ which is given by

$$S_{WEL} = \sqrt{\left(\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}\right)}.$$  

6.2.2 Estimation of Confidence Interval Estimate for $\mu_2 - \mu_1$

In addition to obtaining the point estimate of the mean difference of the two means, it may be of interest to compute a confidence interval for the estimate. For this problem, it is worth recalling that the statistic $((\bar{y}_2 - \bar{y}_1) - (\mu_2 - \mu_1))/S_{WEL}$ follows the $t$ distribution, in which, due to the result by Welch (1937), the degrees of freedom are adjusted to be

$$df_{WEL} = \frac{(d_1 + d_2)^2}{d_1^2/(N_1 - 1) + d_2^2/(N_2 - 1)},$$

where $d_j = S_j^2/N_j$, for $j = 1, 2$. Then the $100 \times (1 - \alpha)\%$ confidence interval for the shift $\mu_2 - \mu_1$ is given by

$$\bar{y}_2 - \bar{y}_1 \pm t_{\alpha/2, df_{WEL}} \times S_{WEL}.$$  

6.2.3 Estimation of the Difference in Two Trimmed Means

It is known that the classic estimate of the difference when the two parent population variance are unequal is very sensitive to departures from normality of the random variables. Furthermore, this estimate is still sensitive to outliers non robust irrespective of what distribution the data came from. The trimmed mean is much more resistant to the outlying values. It has been studied and applied in many applications. Furthermore, to obtain an interval estimate Yuen (1974) proposed an interval estimate based on extension of the standard estimator for the difference in two means under the heteroscedastic cases to the trimmed mean. This led to the development of a more robust trimmed means procedure by Yuen (1974). Under this procedure, the location is estimated by the trimmed
mean and the standard deviation is based on the Winsorized variance. This is briefly described in what follows.

Consider the ordered observations \( Y_{(1)j} \leq Y_{(2)j} \leq \ldots \leq Y_{(N_j)j} \) for the \( j \)th sample for \( j = 1, 2 \). Let \( g_j = \lfloor \lambda N_j \rfloor \) define the case \( \lambda N_j \) is rounded down to the nearest integer, where \( \lambda \) is the proportion of observations to be trimmed in each tail of the corresponding distribution. Then the effective sample for the \( j \)th population becomes \( h_j = N_j - 2g_j \). Let the \( j \)th trimmed mean be given by

\[
\bar{y}_{tj} = \frac{1}{h_j} \sum_{i=g_j+1}^{n_j-g_j} Y_{ij} \quad \text{for} \ j = 1, 2.
\]

To obtain a Winsorized variance, a Winsorized mean is required and is given by

\[
\bar{y}_{wj} = \frac{1}{N_j} \sum_{i=1}^{N_j} X_{ij},
\]

where

\[
X_{ij} = \begin{cases} 
Y_{(g_j+1)j} & \text{if } Y_{ij} \leq Y_{(g_j+1)j} \\
Y_{ij} & \text{if } Y_{(g_j+1)j} < Y_{ij} \leq Y_{(N_j-g_j)j} \\
Y_{(N_j-g_j)j} & \text{if } Y_{ij} \geq Y_{(N_j-g_j)j}
\end{cases}
\]

Then the \( j \)th Winsorized variance is given by

\[
S^2_{Wj} = \frac{1}{N_j - 1} \sum_{i=1}^{N_j} (X_{ij} - \bar{y}_{wj})^2, \quad j = 1, 2.
\]

Further, the sample standard error of the trimmed mean is obtained by

\[
\sqrt{(N_j - 1)S^2_{Wj}/[h_j(h_j - 1)]}.
\]
Let \( d_{wj} = \frac{(N_j-1)s_{wj}^2}{h_j(h_j-1)} \), so that \( S_{YT} = \sqrt{d_{wj} + d_{w2}} \). Then, in view of Yuen (1974), the statistic \( (\bar{y}_2 - \bar{y}_1)/S_{YT} \) follows the \( t \) distribution, with the degrees of freedom

\[
d_{f_{1Y}} = \frac{(d_{w1} + d_{w2})^2}{d_{w1}/(h_1 - 1) + d_{w2}/(h_2 - 1)},
\]

(6.2.2)

Then the \( 100 \times (1 - \alpha)\% \) confidence interval for the shift \( \mu_{t2} - \mu_{t1} \) is given by

\[
\bar{y}_{t2} - \bar{y}_{t1} \pm t_{\alpha/2, df_{1Y}} \times S_{YT},
\]

6.2.4 Estimation of Confidence Interval for Trimmed \( \theta_2/\theta_1 \)

It is well known that if the two independent samples described in Section 6.2.1 come from normal distributions \( N(\mu_1, \sigma_1^2) \) and \( N(\mu_2, \sigma_2^2) \), then

\[
\frac{s_{2}^2/\theta_2^2}{s_{1}^2/\theta_1^2}
\]

follows an \( F \) distribution with \( n_2 - 1, n_1 - 1 \) degrees of freedom.

Following this reason, it seems intuitive that in the case of trimmed samples,

\[
\frac{s_{2}^2/\theta_2^2}{s_{1}^2/\theta_1^2}
\]

follows an approximate \( F \) distribution with \( h_2 - 1, h_1 - 1 \) degrees of freedom.

However, it is well known that trimming alter the distribution of the underlying responses. This would lead to modified distribution which is not the same as the \( F \) random variable above. Moreover, trimming has its limitations in that if scale estimation is being sought, “too much” trimming will lead to loss of information that is in the tail of the distributions hence render the estimator less meaningful. In practice, the problem involving scale parameter, such as a ratio of variance, it is common to utilize bootstrap sample distribution to obtain confidence interval estimates instead of parametric models which are very sensitive to departures from normality (Wilcox, (2003)). Our “approximate F” statistic has been included in this experiment to explore how well the resulting interval
estimates perform given its limitation mentioned. In particular, it is of interest to observe whether or not the inherent curtailing of outlying responses improves the efficiency of the resulting interval estimates. Moreover, when \( \lambda = 0.0 \), the ratio of trimmed sample variances reduces to the standard ratio of sample variance for which the strengths and weaknesses with regard to efficiency are well documented.

It can be deduced from the discussion above that the 100 \( \times (1 - \alpha) \)% confidence interval for the ratio \( \vartheta_{M2}/\vartheta_{M1} \) is given by

\[
\left( \sqrt{\frac{S_{M2}^2}{S_{M1}^2} F_{(\frac{\alpha}{2}),df_1,df_2}}, \sqrt{\frac{S_{M2}^2}{S_{M1}^2} F_{(1-\frac{\alpha}{2}),df_1,df_2}} \right),
\]

where \( S_{Mj}^2 \) is the \( \lambda \)-trimmed variance for the \( j \)th sample, \( j = 1, 2, \) and \( F_{(\frac{\alpha}{2}),df_1,df_2} \) is the 100(1 – \( \frac{\alpha}{2} \))th percentile of the \( F \) distribution with \( df_1 = \) denominator degrees of freedom and \( df_2 = \) numerator degrees of freedom. The trimming rate is fixed at \( \lambda \). Observe that when \( \lambda = 0.0 \), the 100 \( \times (1 - \alpha) \)% confidence interval for the ratio \( \vartheta_2/\vartheta_1 \) is given by

\[
\left( \sqrt{\frac{S_j^2}{S_1^2} F_{(\frac{\alpha}{2}),df_1,df_2}}, \sqrt{\frac{S_j^2}{S_1^2} F_{(1-\frac{\alpha}{2}),df_1,df_2}} \right),
\]

where \( S_j^2 \) is the classic variance for the \( j \)th sample, \( j = 1, 2, \) and \( F_{(\frac{\alpha}{2}),df_1,df_2} \) is the 100 \( \times (1 - \frac{\alpha}{2}) \)th percentile of the \( F \) distribution with denominator and numerator degrees of freedom, \( df_1 \) and \( df_2 \), respectively, where \( df_j = n_j - 1 \) for \( j = 1, 2 \).

6.3 \( M \)-Estimation for \( p \) Group Problem

In this section, an iterative methods for estimating shift in location and ratio of spread for the general \( p \)-group problem is presented. As a natural extension of the two sample group discussed above, we suppose that the shifts \( \mu_2 - \mu_1, \mu_3 - \mu_1, \ldots, \mu_p - \mu_1 \) are of interest.
6.3.1 Estimation of Group 1 to Groups Shifts in Location

The model given in 7.3.1 can be extended to the form

\[ \frac{y_i}{\sigma_i} = \left( \frac{x_{1i}^T}{\sigma_i} \right) \beta + e_i, \quad \sigma_i = e^{x_i^T \theta}, \text{ for } i = 1, \ldots, N, \]  

(6.3.1)

where \( \beta = (\beta_0, \beta_1, \ldots, \beta_{p-1})^T = (\mu_1, \mu_2 - \mu_1, \ldots, \mu_p - \mu_1)^T \), and \( \theta = (\log(\sigma_2/\sigma_1), \log(\sigma_3/\sigma_1), \ldots, \log(\sigma_p/\sigma_1))^T \). \( \sigma_j^2 \) is the variance of the \( j \)th group, \( j = 1, \ldots, p \). Note that by this definition, the ratio of the standard deviation of each group to that of group 1 is given by \( e^{\theta_1}, e^{\theta_2}, \ldots, e^{\theta_{p-1}} \).

In standard least squares method for tackling the location estimation problem, the value \( b \) that one seeks is such that, for a given \( \sigma_i^2 = e^{x_i^T \theta^o} \),

\[ \sum_{i=1}^{N} \left( \frac{(y_i - x_{1i}^T b)}{\sigma_i^o} \right)^2, \]

(6.3.2)

is minimized. We can take \( \theta^o = \hat{\theta}^{(r-1)} \), where \( \hat{\theta}^{(r-1)} \) is an estimate \( \theta \) from the \( (r - 1) \)th iteration. It is clear from this why LS estimator is nonresistant to outliers. \( M \)-estimation addresses this problem by applying a function to the residuals, \( (y_1 - x_{11}^T b)/\sigma_1^o, \ldots, (y_N - x_{1N}^T b)/\sigma_N^o \) that delimits the extreme values. In particular, Huber’s \( M \)-estimate employs a function \( \rho(x) \) such that only the squared residuals within a certain range give a full contribution of their effect to the sum in the objective function. That is, the estimate, \( b \), is the value such that

\[ \sum_{i=1}^{N} \rho((y_i - x_{1i}^T b)/\sigma_i^o) \]
is minimized. Note that $\rho(z)$ is nonnegative, convex, piecewise function

$$\rho(z) = \begin{cases} \frac{1}{2}z^2 & \text{for } |z| \leq k \\ k|z| - \frac{1}{2}k^2 & \text{for } |z| > +k \end{cases}$$  \hspace{1cm} (6.3.3)$$

where $z = ((\sigma_i^*)^{-1}(y_i - x_{i1}^*b))/\hat{\tau}_1$ is standardized value based on some constant $k$. Note that $\rho(z)$ is a continuous function that has a continuous and nondecreasing first derivative

$$\psi(z) = \begin{cases} -m & \text{for } z < -m \\ z & \text{for } -m < z < +m \\ +m & \text{for } +m < z \end{cases}$$  \hspace{1cm} (6.3.4)$$

where $\hat{\tau}_1 = 1.4826 \times MAD$, $MAD$ is Median Absolute Deviation of the residuals, $(y_i - x_{i1}^*\beta)/\sigma^*$. $\sigma^*$ is taken to be $\hat{\sigma}_1^{(r-1)} = \exp\{x_i^T\hat{\beta}^{(r-1)}\}$, $\hat{\beta}^{(r-1)}$ is the $(r - 1)$th iteration $M$-estimate of $\beta$ obtained from fitting model (6.3.5), below.

### 6.3.2 Estimation of Ratio of Scale for Group to Group 1

Let us now consider the estimation of the ratios of spread for the general $p$ group problem. An estimate the scale coefficient is analogously obtained in the following manner. The linearized model used for the two sample problem can be extended to

$$\log |y_i - x_{i1}^T\beta^*| = x_i^T\theta + e_i^*, \quad i = 1, \ldots, N, \hspace{1cm} (6.3.5)$$

with a view to obtain $(\hat{\theta}_2/\hat{\theta}_1, \hat{\theta}_3/\hat{\theta}_1, \ldots, \hat{\theta}_p/\hat{\theta}_1) = (e^{\hat{\theta}_1}, e^{\hat{\theta}_2}, \ldots, e^{\hat{\theta}_{p-1}})$. Note that the variables $e_i^* = \log |e_i|, i = 1, \ldots, N$ have a common distribution that is centered at some constant, $\theta_0$. In this model, for a specified $\beta^* = \hat{\beta}^{(r-1)}$, let $\hat{\beta}^{(r-1)}$, the $(r - 1)$th iteration $M$-estimate of $\beta$ obtained from fitting model (6.3.1). Then the estimate of $\theta$ we seek is the
value \( t \) that minimizes the function

\[
\sum_{i=1}^{N} \rho \left( \log |y_i - x_i^T \beta^0| - x_i^T t \right).
\]

Further, \( \rho(v) \) is a continuous, function that has a continuous and nondecreasing first derivative

\[
\psi(v) = \begin{cases} 
-l & \text{for } v < -l \\
v & \text{for } -l < v < +l \\
+l & \text{for } +l < v
\end{cases},
\]

where \( v = \frac{(\log |y_i - x_i^T \beta^0| - x_i^T \hat{t})}{\hat{\tau}_2}, \hat{\tau}_2 = 1.4826 \times \text{MAD} \), \( \text{MAD} \) is the Median Absolute Deviation of the residuals, \( \log |y_i - x_i^T \beta^0| - x_i^T \hat{t}, i = 1, \ldots, N \).

It is worth noting that when

\[
\rho\left(\frac{(y_i - x_i^T \hat{\beta})}{\sigma^0}\right) = \left(\frac{(y_i - x_i^T \hat{\beta})}{\sigma^0}\right)^2,
\]

and

\[
\rho\left(\log |y_i - x_i^T \beta^0| - x_i^T \hat{t}\right) = \left(\log |y_i - x_i^T \beta^0| - x_i^T \hat{t}\right)^2,
\]

the corresponding \( M \)-estimates are exactly the iterative LS estimates under heteroscedasticity, referred to below as ILSHET estimates. Furthermore, if in \( \sigma^0 = 1 \) in the first equation, then the \( M \)-estimation reduces to LS estimate under homoscedasticity, referred to later as LSHOM.

6.3.3 Algorithm for the Iterative \( M \)-estimation

To obtain the IRWLS \( M \)-estimates of \( \beta \) and \( \theta \), the following method is utilized:
(1) Obtain a preliminary unweighted M-estimate of $\hat{\beta}^{(0)}$ by minimizing
$$\sum_{i=1}^{N} \rho(y_i - x_{i1}^T b).$$
Using $\beta^o = \hat{\beta}^{(0)}$, form estimated transformed errors
$$(\log |y_1 - x_{11}^T \beta^o|, \ldots, \log |y_N - x_{1N}^T \beta^o|)^T.$$

(2) Obtain a preliminary scale parameter estimate, $\hat{\theta}^{(0)}$, by minimizing
$$\sum_{i=1}^{N} \rho(\log |y_N - x_{iN}^T \beta^o| - x_i^T t), \beta^o = \beta^{(0)}.$$ 
Using $\theta^o = \hat{\theta}^{(0)}$, form scaling constants $\sigma_i^o = \exp\{x_i^T \theta^o\}$ for $i = 1, \ldots, N$.

(3) Obtain weighted M-estimate of $\beta, \hat{\beta}^{(1)}$, by minimizing
$$\sum_{i=1}^{N} \rho(|y_i - x_{i1}^T b|/\sigma_i^o), \text{ given } \sigma_i^o = \exp\{x_i^T \theta^o\}, \theta^o = \theta^{(0)}.$$ 
Using $\beta^o = \hat{\beta}^{(1)}$, form updated transformed errors $(\log |y_1 - x_{11}^T \beta^o|, \ldots, \log |y_N - x_{1N}^T \beta^o|)^T.$

(4) Steps 2 and 3, respectively, may be repeated using $\log |y_1 - x_{1N}^T \beta^o|, \ldots, \log |y_N - x_{1N}^T \beta^o|$, $\beta^o = \hat{\beta}^{(r)}$ and $\sigma_i^o = \hat{\sigma}_i^{(r-1)} = \exp\{x_i^T \theta^{(r-1)}\}$, for $r = 0, 1, 2, \ldots$, until the estimates have converged. The convergence criterion iteration are terminated once $\|\hat{\theta}^{(r+1)} - \hat{\theta}^{(r)}\| < 0.0001.$

6.4 Conclusion

The methods that are not based on ranks that were performed in this investigation are in general described by the summary presented in this chapter. Below there are some variations such as those based on trimmed absolute residuals, are straight forward modifications of the $M$-estimates and Rank estimates. These were be briefly described as needed in the sequel.
CHAPTER VII

SIMULATION EXPERIMENTS

7.1 Introduction

In this chapter, the analysis of four simulations experiments that demonstrate performance of IRHET in comparison with other standard methods are presented. The experiments consist of the (1) two sample problem, (2) three group problem model, (3) Inlier-Outlier Contaminated Normal in three group problem and (4) multiple regression with 2 independent variables.

7.2 Example 1: Two Sample Problem

Consider a two sample problem under the designs: unbalanced design (20,40) and balanced design (40,40). This allocation allows us to study the performance of the IRHET and other methods taking into account the sizes of samples with in view of gains from robustness and efficiency. Let \( \mu_1 \) and \( \sigma_1^2 \) denote the location and variance of the population 1. Let \( \mu_2 \) and \( \sigma_2^2 \) denote the location and variance of the population 2. The difference in location that we seek is \( \mu_2 - \mu_1 \) and the ratio of the spread is \( \theta = \frac{\sigma_2}{\sigma_1} \). Then, the responses from the combined sample can be represented by following linear model

\[
y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, N. \tag{7.2.1}
\]

Observe that \( x_i \) is the \( i \)th element of vector of zeros and ones. The theory of estimator, of \( \theta \) required the \( x_i \)s to be centered. However, since a ratio of scale is independent of location so the process adding and subtracting \( \overline{x} \) does not alter the asymptotic properties of the estimate of \( \theta \). Further, it is remarked in Kraft and van Eeden (1972) that the estimate of
the shift in parameter that is of interest when obtained under centered and uncentered design are exactly equivalent for any $N$. These two facts support that fact that writing the model as done in 7.2.1 does not lead to the loss of validity of the asymptotic properties of the resulting estimates of parameters of interest in this study.

Using signed-rank Wilcoxon scores for the location problem, and normal scores for the scale problem, the proposed estimation method was employed to obtain estimates of $\beta = (\beta_0, \beta_1)^T$ and the ratio of standard deviations given by $e^\theta$. Iteratively Reweighted Least Square (IRWLS), Least Squares for homoscedastic and heteroscedastic cases, Trimmed Mean for the heteroscedastic cases rank estimation methods have been included so that meaningful comparisons between the proposed estimates of $\beta$ and $\theta$ and their IRWLS analogues can be made. Similarly, comparisons between the homoscedastic LS and rank estimates are considered. The experiment is repeated 10,000 times subject to the following conditions:

(i) the true regression parameter values were $\beta_0 = 0.0$ and $\beta_1 = 2.0$. The location for population 1 is $\mu_1 = 0.0$ where as for population 2, it is assumed to be $\mu_2 = 0.0 + 2.0 = 2.0$, hence of interest is the shift in location of 2.0.

(ii) the ratio of the standard deviation of sample 2 to that of sample 1, $\eta = e^\theta$ is 3, which is equivalent to considering the case $\theta = 1.098612$. This turns out to be the case $\theta_1 = 1$, and $\theta_2 = exp\{1.098612\} = 3.0$.

(iii) the errors are drawn from a $N(0, 1)$ for the uncontaminated normal distribution case where as for the contaminated normal distribution case, the errors are drawn from $CN(\epsilon, 9)$ where $\epsilon = (0.01, 0.10, 0.20)$. The foregoing presentation only considers the results from 20% level contamination.

(iv) For each method considered in this study, a fixed seed was used so that the results can be replicated.
Our experience shows that the proposed method does not perform well when sample sizes are smaller than 20, since there are not enough observations to fit the variance function. In this discussion, we present the results from the cases with sample designs (20,40) and (40,40). The former was used to demonstrate performance of the methods for the unbalanced design in which, appropriately, more observations were drawn from the population with larger variance. It should be pointed out that the methods performed well even in the ill-allocated unbalanced design (40,20). The balanced case was included to demonstrate how well the proposed method performs in the design which the homoscedastic methods are known do well.

7.3 Analysis of the Estimates of $\mu_2 - \mu_1$

In this section, the estimates of the shift in location, $\mu_2 - \mu_1$ for the methods under consideration are analyzed. Averages and standard deviations of the estimates from 10,000 trials were computed using the several methods under considerations. Before discussing the results, a brief summary of the methods is provided. RHOM and LSHOM were included as controls since these methods assume that the sample are homoscedastic in nature. LSHET, LSHETtr1 and LSHETtr2 were included since they are suited for heteroscedastic cases, however, with these methods, one can realize the estimate of the shift in location without specifying the underlying variance function model.

7.3.1 Rank Estimation Under Heteroscedasticity

It is worth recalling that the average of the estimates of the shift in location, $\mu_2 - \mu_1$ was computed based on estimates of the coefficients of the signed-rank fit of the re-scaled model

$$\frac{y_i}{\sigma_i} = (x_i^T/\sigma_i)\beta + e_i, \quad \sigma_i^o = e^{\theta x_i}, i = 1, \ldots, N,$$ (7.3.1)
where $\beta = (\beta_0, \beta_1)^T$ and $\beta_1 = \mu_2 - \mu_1$. Let $\theta^o = \hat{\beta}^{(r-1)}$, where $\hat{\beta}^{(r-1)}$ is the estimate of $\beta$ obtained from the $(r-1)$th iteration. Suppose $\mathbf{b} \in \mathcal{R}^2$, so that the $(y_i - \mathbf{x}_{i1}^T \mathbf{b})/\sigma_i^o$ for $i = 1, \ldots, N$, are the residuals. Then, rank estimate of $\beta$ in the $r$th iteration, $\hat{\beta}^{(r)}$, minimizes

$$D_{1N}(\mathbf{b}, \theta^o) := \sum_{i=1}^{N} \phi_i^+(R||y_i - \mathbf{x}_{i1}^T \mathbf{b})/\sigma_i^o||/\sqrt{(N + 1)}((y_i - \mathbf{x}_{i1}^T \mathbf{b})/\sigma_i^o).$$  

(7.3.2)

Further, observe that this objection function can be formulated as a IRWLS problem

$$D_{1N}(\mathbf{b}, \theta^o) := \sum_{i=1}^{N} w_i^+(\mathbf{b}, \theta^o)[(y_i - \mathbf{x}_{i1}^T \mathbf{b})/\sigma_i^o]^2,$$  

(7.3.3)

where

$$w_i^+(\mathbf{b}, \theta^o) = \frac{\phi_i^+(R||y_i - \mathbf{x}_{i1}^T \mathbf{b})/\sigma_i^o||/(N + 1))}{||y_i - \mathbf{x}_{i1}^T \mathbf{b})/\sigma_i^o||} \text{ for } |(y_i - \mathbf{x}_{i1}^T \mathbf{b})/\sigma_i^o| \neq 0$$

$$= 0 \text{ elsewhere.}$$

Then given the step in the $r$th iteration, $k_r$, the $(k_r + 1)$-step estimate of $\beta$, $\hat{\beta}_N^{(k_r)}$, minimizes the $k_r$th step dispersion given by

$$D_{1N}^*(\mathbf{b}|\mathbf{b}_{k_r}, \theta^o) := \sum_{i=N}^{N} w_i^+(\mathbf{b}_{k_r}, \theta^o)[(y_i - \mathbf{x}_{i1}^T \mathbf{b})/\sigma_i^o]^2, \quad k_r = 0, 1, \ldots$$  

(7.3.4)

Table 1 gives a summary of all the methods that were considered in the simulation trial. It can be seen that the first three methods account for outlier by down weighting the extreme values through the rank based scores. The next pair of methods handle aberrant observations by trimming them. The bottom pair does not curtail the effect of outliers. The methods RHOM and LSHOM are homoscedastic and are included as controls for the estimation method that admits heteroscedasticity. Finally, the methods LSHETtr1, LSHETtr2 and LSHET accommodate heteroscedasticity with out the additional knowledge of the
<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET</td>
<td>Iterating between estimation of $\beta = \mu_2 - \mu_1$ and $\theta$ leads to final estimate, $\hat{\theta}^{(k)}$. Then, the final estimate of $\beta$ is the value that minimizes the $k$-step dispersion function in (7.3.4), with $\theta^0 = \hat{\theta}^{(k)}$.</td>
</tr>
<tr>
<td>RHETsca</td>
<td>Let $\theta^0$ denote the true value of the scale coefficient. Then, the desired estimate of $\beta = \mu_2 - \mu_1$ is the value that minimizes the $k$-step dispersion function in (7.3.4), with $\theta^0 = \theta$.</td>
</tr>
<tr>
<td>RHOM</td>
<td>For constant variance, $\sigma_i = 1$ for all $i$ which is equivalent to $\theta = 0$. Then, the estimate of $\beta = \mu_2 - \mu_1$ is the value that minimizes the $k$-step dispersion function in (7.3.4), with $\theta^0 = 1$. Pooled estimate of variance is used.</td>
</tr>
<tr>
<td>LSHETtr1</td>
<td>10% of the smallest and largest values are trimmed in each sample. Then, the estimate of $\mu_2 - \mu_1$ is the difference in means of the remaining values. Unequal variance is estimated by winsorized sample variances.</td>
</tr>
<tr>
<td>LSHETtr2</td>
<td>20% of the smallest and largest values are trimmed in each sample. Then, the estimate of $\mu_2 - \mu_1$ is the difference in means of the remaining values. Unequal variance is estimated by winsorized sample variances.</td>
</tr>
<tr>
<td>LSHET</td>
<td>There is no trimming and all the observations from each sample are used. Then, the estimate of $\mu_2 - \mu_1$ is the difference in the two sample means. Unequal variance is estimated by common sample variances.</td>
</tr>
<tr>
<td>LSHOM</td>
<td>There is no trimming and all the observations from each sample are used. Then, the estimate of $\mu_2 - \mu_1$ is the difference in the two sample means. Pooled variance is estimated by pooled sample variances.</td>
</tr>
</tbody>
</table>

variance function which being considered in this study.

7.3.2 Estimation Results under $CN(0.20, 9)$ Distribution

In this section, we consider the results of estimates of difference in location from all the methods for the cases in which the errors come from the 20% contaminated normal, $CN(0.20, 9)$. It is known that the Least Square type methods are superior when responses come from the standard normal $N(0, 1)$. The proposed method performed almost as well as the LS squares. Since our interest lies in the robustness of the methods when outliers
are introduced, we chose the highest possible contaminated normal, \( CN(0.20, 9) \). This seemed reasonable since the signed-rank Wilcoxon scores that were used to estimate \( \mu_2 - \mu_1 \), are known to have a breakdown of 25\% in the homoscedastic cases.

**Location Shift Estimates under \( N(0.20, 9) \)**

Consider results given in Table 2. All of the methods reported averages close to the true value, 2.0. The differences were observed in the variability. The proposed method, IRHET was resistant to outliers and reported standard deviations that was close to those from RHETsca in the balanced case. Observe that the LSHETtr2 yielded the standard deviations that were almost equivalent to those reported by RHETsca. In both the designs, due to the high level of contamination, the LSHOM and LSHET methods reported standard deviations that were so much larger than those produced by the other five robust methods. Observe that amongst the robust, RHOM and LSHETtr1 reported the largest

<table>
<thead>
<tr>
<th>Method</th>
<th>Sample Sizes : ((n_1, n_2))</th>
<th>(20, 40)</th>
<th>(40, 40)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>IRHET</td>
<td></td>
<td>2.006</td>
<td>0.685</td>
</tr>
<tr>
<td>RHETsca</td>
<td></td>
<td>1.993</td>
<td>0.675</td>
</tr>
<tr>
<td>RHOM</td>
<td></td>
<td>1.986</td>
<td>0.686</td>
</tr>
<tr>
<td>LSHETtr1</td>
<td></td>
<td>1.992</td>
<td>0.684</td>
</tr>
<tr>
<td>LSHETtr2</td>
<td></td>
<td>1.990</td>
<td>0.673</td>
</tr>
<tr>
<td>LSHET</td>
<td></td>
<td>2.014</td>
<td>0.843</td>
</tr>
<tr>
<td>LSHOM</td>
<td></td>
<td>2.007</td>
<td>0.851</td>
</tr>
</tbody>
</table>

\(I=\)Iterated, \(R=\)Rank, \(LS=\)Least Squares, 
\(HET=\)Heteroscedastic, \(HOM=\)Homoscedastic, 
\(sca=\)with scale coefficient specified.

LSHETtr1 is computed using 10\% trimming rate, 
LSHETtr2 is computed using 20\% trimming rate.
standard deviations. This can be attributed to the fact the former method does not account for heteroscedasticity, and 10% trimming rate in the latter was not sufficient to curtail the outlier effects in the contaminated distribution.

7.4 Coverage of 90% and 95% Confidence Interval for Shift in Location

In this section, for $0 < \alpha < 1$ the results of empirical levels of the $100 \times (1 - \alpha)\%$ confidence intervals are analyzed. We considered two values of $\alpha$, 0.10 and 0.05, corresponding to 90% and 95% confidence levels. Based on the 10,000 trials, the proportion of intervals that contained the parameter value of was enumerated for each method and used as the corresponding empirical coverage probability.

The interval estimates for $\mu_2 - \mu_1$ that are classified into three categories: (A) Rank based heteroscedastic and homoscedastic, (B) trimmed samples heteroscedastic, (C) Classic Homoscedastic.

(A) The 90% and 95% confidence interval estimates based on ranks were calculated using the formula

$$CI_{1-\alpha} (\mu_2 - \mu_1) := \mu_2 - \mu_1 \pm t_{(1-\alpha),df} \hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$  

$$\hat{\mu_2} - \hat{\mu_1} = \hat{\beta}_1, \quad t_{(1-\alpha),df} \text{ is the 95th and 97.5th percentile of } t \text{ with } df = n_1 + n_2 - 2,$$

$$\hat{\sigma} = (\hat{\gamma}_1)^{-1}, \quad \hat{\gamma}_1 \text{ is the estimate of } \gamma_L \text{ which is based on the residuals } (y_i - x_i^T \hat{\beta})/\sigma_i,$$

for $i = 1, \ldots, N$.

(B) Under the general Least Squares method, the 90% and 95% confidence interval estimates were calculated using the formula
\[ CI_{1-\frac{\alpha}{2}}(\mu_{\lambda 2} - \mu_{\lambda 1}) := \mu_{\lambda 2} - \mu_{\lambda 1} \pm t_{(1-\frac{\alpha}{2})} \sqrt{d_{y1} + d_{y2}}, \]  

(7.4.2)

\[ \mu_{\lambda 2} - \mu_{\lambda 1} = \bar{Y}_{\lambda 2} - \bar{Y}_{\lambda 1}, \quad \bar{Y}_{\lambda j} \text{ is the } \lambda \text{-trimmed mean of the } j \text{ th sample}, \]

\[ t_{(1-\frac{\alpha}{2})} \] is the 95th and 97.5th percentile of \( t \) with \( df_y = \frac{(d_{y1} + d_{y2})^2}{\frac{d_{y1}^2}{h_1 - 1} + \frac{d_{y2}^2}{h_2 - 1}} \)

\[ d_{yj} = \frac{(n_j - 1)S^2_{\lambda wj}}{h_j(h_j - 1)}, \quad \text{where } h_j = n_j - 2g_j, g_j = \lfloor \lambda n_j \rfloor, \]

where \( \lfloor p \rfloor \) is the greatest integer part of \( p \),

\[ S^2_{\lambda wj} \text{ is the } \lambda \text{-winsorized variance for the } j \text{ th sample, for } j = 1, 2, \]

and \( \lambda \) is the trimming rate.

(C) This is the classic LS analogue to the interval estimation in (A) with \( \mu_2 - \mu_1 \) estimated by \( \bar{Y}_2 - \bar{Y}_1 \), where \( \bar{Y}_j \) is mean of the \( j \)th sample, \( j = 1, 2 \). Further, \( \hat{\tau}_1 \) in (A) is replaced by \( \hat{\tau}_{LS} = \hat{\tau}_p \), where \( \hat{\tau}_p \) is the pooled variance defined by \( S^2_p = S^2_{S_P} = \frac{(n_1 - 1)S^2_j + (n_2 - 1)S^2_j}{n_1 + n_2 - 2} \), \( S^2_j \) is variance of the \( j \)th sample, for \( j = 1, 2 \) where \( S_j \) is the variance of the \( j \)th sample, \( j = 1, 2 \).

In seeking to construct the interval estimates of the shift in location, Table 3 presented below briefly describes how this is accomplished under each method.

7.4.1 Coverage Probability of Interval Estimates for \( \mu_2 - \mu_1 \)

In this section, we consider the empirical level obtained using all the methods for the shift in location problem at both the 90% and 95% nominal levels, when responses come from the \( CN(0.20, 9) \). Observe in Table 4 that the methods considered in the experiment yielded estimates that were reasonably close to the nominal levels, 90% and 95%, in all the designs with the exception of SRHOM in the balanced design and LSHOM in the
Table 3
Summary of Confidence Interval Estimation for Shift in Location for the Two Sample Case

<table>
<thead>
<tr>
<th>Method (Category)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IRHET (A)</strong></td>
<td>$\sigma_i^0 = \hat{\theta}_0^{(k)}$ for $i = 1, \ldots, N$, where $\hat{\theta}_0^{(k)} = {\theta^{(k)}}^{\hat{\theta}}_i$, $\hat{\theta}_i^{(k)}$ is final estimate of $\theta$ in the iterative scheme. Re-scaling the residuals by $e^{\hat{\theta}_i^{(k)}}$, the interval was computed using (7.4.1), based on a pooled estimate of dispersion $\hat{\tau}_1$.</td>
</tr>
<tr>
<td><strong>RHETsca (A)</strong></td>
<td>$\sigma_i^0 = \sigma_i$, for $i = 1, \ldots, N$, where $\sigma_i = e^{\theta x_i}$, the specified value of $\theta$ is the true value. Re-scaling the residuals by $e^{-\theta x_i}$, the interval was computed using (7.4.1), based on a pooled estimate of dispersion $\hat{\tau}_1$.</td>
</tr>
<tr>
<td><strong>RHOM (A)</strong></td>
<td>$\sigma_i^0 = 1$ for $i = 1, \ldots, N$, a special case of IRHETsca. Taking the residuals, $(y_i - x_i^0 \hat{\theta})$, the interval was computed using (7.4.1), based on a pooled estimate of dispersion $\hat{\tau}_w$.</td>
</tr>
<tr>
<td><strong>LSHETtr1 (B)</strong></td>
<td>10%-Trimmed Mean Difference under the Heteroscedastic case. The interval estimates were computed using formula (7.4.2) with $\lambda = 0.10$. Note that the standard error is a function of unweighted sum of 10% winsorized sample variances.</td>
</tr>
<tr>
<td><strong>LSHETtr2 (B)</strong></td>
<td>20%-Trimmed Mean Difference under the Heteroscedastic case. The interval estimates were computed using formula (7.4.2) with $\lambda = 0.20$. Note that the standard error is a function of unweighted sum of 20% winsorized sample variances.</td>
</tr>
<tr>
<td><strong>LSHET (B)</strong></td>
<td>Standard Mean Difference under the Heteroscedastic case. The interval estimates were computed using formula (7.4.2) with $\lambda = 0.00$. Note that the standard error is a function of unweighted sum of usual sample variances.</td>
</tr>
<tr>
<td><strong>LSHOM (C)</strong></td>
<td>There is no trimming and all the observations from each sample are utilized. The standard error is estimated based on a weighted sum of sample variances (pooled estimate).</td>
</tr>
</tbody>
</table>

unbalanced design.

7.5 Analysis of the Estimates of $\eta$

In this section, the estimates of the ratio of the scale for the two samples, $\vartheta_2/\vartheta_1$, for the methods under consideration are analyzed. Before discussing the results, a brief
Table 4

Empirical Levels for 90% and 95% Confidence Interval Estimates of Difference in Location Based on 10,000 Simulations for \( \mu_2 - \mu_1 = 2.0, \eta = \frac{\theta_2}{\theta_1} = 3.0 \) under \( CN(0.20, 9) \) Errors

<table>
<thead>
<tr>
<th>Method</th>
<th>Sample Sizes : ((n_1, n_2))</th>
<th>90%</th>
<th>95%</th>
<th>90%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(20,40)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRHET</td>
<td>0.895</td>
<td>0.888</td>
<td>0.945</td>
<td>0.945</td>
<td></td>
</tr>
<tr>
<td>RHETsca</td>
<td>0.917</td>
<td>0.923</td>
<td>0.945</td>
<td>0.947</td>
<td></td>
</tr>
<tr>
<td>RHOM</td>
<td>0.934</td>
<td>0.804</td>
<td>0.971</td>
<td>0.874</td>
<td></td>
</tr>
<tr>
<td>LSHETtr1</td>
<td>0.904</td>
<td>0.902</td>
<td>0.954</td>
<td>0.951</td>
<td></td>
</tr>
<tr>
<td>LSHETtr2</td>
<td>0.899</td>
<td>0.901</td>
<td>0.949</td>
<td>0.949</td>
<td></td>
</tr>
<tr>
<td>LSHET</td>
<td>0.901</td>
<td>0.903</td>
<td>0.948</td>
<td>0.952</td>
<td></td>
</tr>
<tr>
<td>LSHOM</td>
<td>0.969</td>
<td>0.898</td>
<td>0.989</td>
<td>0.950</td>
<td></td>
</tr>
</tbody>
</table>

I=Iterated, R=Rank, LS=Least Squares, HET=Heteroscedastic, HOM=Homoscedastic, sca=with scale coefficient specified.
LSHETtr1 is computed using 10% trimming rate,
LSHETtr2 is computed using 20% trimming rate.

summary of the methods is provided.

7.5.1 Rank Estimation under Heteroscedasticity

The average of the estimates of the ratio, \( \eta = \frac{\theta_2}{\theta_1} \), is obtained from the rank estimation of the slope coefficient in the linearized model

\[
\log |y_i - x_i^T \hat{\beta}| = \theta x_i + \log |e_i|, \quad i = 1, \ldots, N, \tag{7.5.1}
\]

where it is assumed \( \log |e_i| \) follow a distribution function that is centered at some non-zero constant \( \theta_0 \), and finally computing

\[
\hat{\eta} = e^{\hat{\theta}}.
\]
Let $\beta^{(r)} = \hat{\beta}^{(r)}$ be the estimate of $\beta$ in the $r$th iteration. Then, for any $t \in \mathcal{R}$, the residuals $\log |y_i - x_{1i}^T \beta^{(r)}| - x_i t$, the rank estimate of $\theta$ in the $r$th iteration, $\hat{\theta}(r)$, minimizes

$$D_{2N}(\beta^{(r)}, t) := \sum_{i=1}^{N} \phi_2^*(\frac{R[\log |y_i - x_{1i}^T \beta^{(r)}| - x_i t]}{N + 1})(\log |y_i - x_{1i}^T \beta^{(r)}| - x_i t),$$

(7.5.2)

This objective function can be reexpressed as the IRWLS formulation

$$D_{2N}(\beta^{(r)}, t) := \sum_{i=1}^{N} w_i(\beta^{(r)}, t)[v_i(\beta^{(r)}, t) - m_v(\beta^{(r)}, t)]^2,$$

(7.5.3)

where the weights,

$$w_i(\beta^{(r)}, t) = \frac{\phi_2^*(R[v(\beta^{(r)}, t) - m_v(\beta^{(r)}, t)]/(N + 1))}{[v(\beta^{(r)}, t) - m_v(\beta^{(r)}, t)]} \quad \text{for } v(\beta^{(r)}, t) \neq m_v(\beta^{(r)}, t)$$

$$= 0 \quad \text{elsewhere.}$$

and $v_i(\beta^{(r)}, t) = \log |y_i - x_{1i}^T \beta^{(r)}| - x_i t$ and $m_v(\beta^{(r)}, t)$, is the centering constant defined as the $\nu$th quantile of the residuals $v(\beta^{(r)}, t)$. Note that $v$ is such that $\phi(u) < 0(\phi(u) > 0)$ if $u < v(u > v)$.

Then, given the $l$th step in the $r$th iteration, $l_r$, $(l_r + 1)$-step estimate of $\theta$, $\hat{\theta}^{(l_r)}$, minimizes the $l$-step dispersion given by

$$D_{2N}(\beta^{(0)}, t|t_{l_r}) := \sum_{i=1}^{N} w_i(\beta^{(0)}, t_{l_r})[v_i(\beta^{(r)}, t) - m_v(\beta^{(r)}, t_{l_r})]^2.$$

(7.5.4)

We now turn to the estimate of the ratio of standard deviations of $\lambda$-trimmed samples. The average of the estimates of the ratio, $\eta = \vartheta_{2}/\vartheta_{1}$, is obtained by estimating the ratio of the $\lambda$-trimmed standard deviations, $\sigma_{\lambda 2}/\sigma_{\lambda 1}$ by the ratio of $\lambda$-trimmed sample standard deviations. That is,

$$\hat{\eta}_{\lambda w} = S_{\lambda 2}/S_{\lambda 1},$$

(7.5.5)
where

\[
S_{\lambda 1}^2 = \frac{1}{n_1 - 2g_1} \sum_{i=1}^{n_1 - 2g_1} (X_{i1} - \bar{Y}_{\lambda 1})^2, \quad \text{and} \quad S_{\lambda 2}^2 = \frac{1}{n_2 - 2g_2} \sum_{j=1}^{n_2 - 2g_2} (X_{j2} - \bar{Y}_{\lambda 2})^2,
\]

where \( \bar{Y}_{\lambda j} \) is the mean of the \( j \)th \( \lambda \)-trimmed sample and \( g_j = \lfloor \lambda n_j \rfloor \). Note that this estimator ignores the known variance model assumed in this study. It is worth noting that LSVHETr1 and LSVHETr2 have been included as benchmarks as to well the down weighting utilized in the proposed method performs. This study did not consider a detailed analysis of their asymptotic properties. Once the outer elements have been trimmed, the remaining \( n_j - 2g_j \) are being assumed to be independent in order to utilize the Snedecor's \( F \) distribution to obtain the confidence interval of the ratio of scale. However, nonparametric interval estimates of the ratio of scales from the trimmed samples can be obtained using bootstrap and other methods.

The following table gives a summary of all the methods that were considered in the simulation trial.

7.5.2 Estimation Results under \( CN(0.20,9) \) Errors

In this section, we consider the results of estimates of ratio of scale from all the methods for the cases in which the 20% contaminated normal was employed to generate the responses.

Ratio of Scale Estimates under \( N(0.20,9) \)

Observe in Table 6 that generally the methods reported values that were reasonably close to 3.0 in the (40,20) and (40,40) designs. It is seen in the (20,40) design that all the methods with the exception of LSVHETr1 and LSVHETr2, reported average estimate that were somewhat further from the true value of 3.0. In the balanced design, the IRHET and RHETsca yielded the values that were closest to the true value of 3.0.
**Table 5**

Summary of Methods Used in the Two Sample Scale Simulation Trials

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET</td>
<td>general estimate of scale parameter under heteroscedasticity. The method iterates between estimating $\beta$ and $\theta$. Then, for a $(k - 1)$-step, estimate $\hat{\beta}^{(k-1)}$, an estimate of the latter is computed as the value that minimizes the function (7.5.4) with $\beta^* = \hat{\beta}^{(k-1)}$.</td>
</tr>
<tr>
<td>RHETloc</td>
<td>special case estimate of scale parameter under heteroscedasticity. The method assumes that a specified value of $\beta^<em>$ is supplied to obtain an estimate of $\theta$. Then, for the specified value, $\beta$, which in this case is the true value, the estimate of $\theta$, is the value $\theta$ which minimizes the function (7.5.4) with $\beta^</em> = \beta$.</td>
</tr>
<tr>
<td>LSVHETtr1</td>
<td>10% of the smallest and largest values are trimmed in each sample. Then, the estimate of the ratio of spread, $\eta$, is estimated by ratio of standard deviations of the remaining values from the two samples, as with in (7.5.5) with $\lambda = 0.10$.</td>
</tr>
<tr>
<td>LSVHETtr2</td>
<td>20% of the smallest and largest values are trimmed in each sample. Then, the estimate of the ratio of spread, $\eta$, is estimated by ratio of standard deviations of the remaining values from the two samples, as in (7.5.5) with $\lambda = 0.20$.</td>
</tr>
<tr>
<td>LSVHET</td>
<td>There is no trimming and all the observations from each sample are used. The estimate of the ratio of spread, $\eta$, is estimated by ratio of standard deviations from the two samples, as in (7.5.5) with $\lambda = 0.0$.</td>
</tr>
<tr>
<td>TrSDHET</td>
<td>For each sample, a 10% trimmed mean is computed and used to derive the deviations, $d_i = x_i - \bar{x}_t$ which are used to obtain sample standard deviation in the usual way. Based on these estimates of the trimmed standard deviation, the desired ratio is computed.</td>
</tr>
</tbody>
</table>

In every design, RHETloc yielded the smallest standard deviations while LSVHET, reported the largest standard deviations. The IRHET method yielded standard deviations that were slightly larger than those obtained from RHETloc. Observe that the standard deviations of the estimate obtained from the proposed method and the ratio of trimmed standard deviations were nearly equal under both designs.
Table 6

Averages and Standard Deviations of Ratio of Scale Estimates Based on 10,000 Simulations for $\mu_2 - \mu_1 = 2.0$, $\eta = \theta_2/\theta_1 = 3.0$ under $CN(0.20,9)$ Errors

<table>
<thead>
<tr>
<th>Method</th>
<th>Sample Sizes : $(n_1, n_2)$</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(20,40)$</td>
<td></td>
<td></td>
<td>$(40,40)$</td>
<td></td>
</tr>
<tr>
<td>IRHET</td>
<td>3.186</td>
<td>0.967</td>
<td>3.060</td>
<td>0.769</td>
<td></td>
</tr>
<tr>
<td>RHETloc</td>
<td>3.160</td>
<td>0.938</td>
<td>3.090</td>
<td>0.743</td>
<td></td>
</tr>
<tr>
<td>LSVHETtr1</td>
<td>3.066</td>
<td>0.966</td>
<td>3.097</td>
<td>0.771</td>
<td></td>
</tr>
<tr>
<td>LSVHETtr2</td>
<td>3.066</td>
<td>0.966</td>
<td>3.097</td>
<td>0.771</td>
<td></td>
</tr>
<tr>
<td>LSVHET</td>
<td>3.306</td>
<td>1.150</td>
<td>3.122</td>
<td>0.903</td>
<td></td>
</tr>
<tr>
<td>TrSDHET</td>
<td>3.299</td>
<td>1.180</td>
<td>3.130</td>
<td>0.904</td>
<td></td>
</tr>
</tbody>
</table>

1=Iterative, R=Rank, LSV=Least Squares Variance, HET=heteroscedastic, HOM=homoscedastic, footnotesize loc=with location specified.

LSVHETtr1 is computed using 10% trimming rate,
LSVHETtr2 is computed using 20% trimming rate.
TrSDHET= trimmed Standard Deviation is computed using 10% trimming rate.

General Remarks

It is worth noting in the results obtained above the IRHET method performed reasonably well when responses that were drawn from the non-normal distribution. It has already been demonstrated in Conover and Johnson (1976) that RHETloc performed well under both types of distributions: normal and non-normal. Thus, the result obtained above shows that the proposed method also draws on that same robust property of the latter method.

7.6 Coverage for 90% and 95% Confidence Intervals for Ratio of Scale

In this section, the results of empirical levels of the 95% confidence intervals are analyzed. Based on 10,000 trials, the proportion of intervals that contained the true parameter value of 3.0 was computed for every method and used as the corresponding
empirical coverage probability.

The interval estimates for $\eta$ that can be classified into two categories: (A) Rank based, (B) trimmed samples.

(A) The 90% and 95% confidence interval rank estimates for $\eta$ were computed using the formula,

$$
\left( e^{\hat{\theta}_i}, e^{\hat{\theta}_u} \right), \quad \text{for any } \beta,
$$

(7.6.1)

where $\hat{\theta}_i = \hat{\theta} - t_{(1 - \frac{\alpha}{2}), df} \hat{\gamma}_\phi \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, and $\hat{\theta}_u = \hat{\theta} + t_{(1 - \frac{\alpha}{2}), df} \hat{\gamma}_\phi \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

$t_{(1 - \frac{\alpha}{2}), df}$ is the 95th and 97.5th percentile of $t$ with $df = n_1 + n_2 - 2$,

$\hat{\gamma}_\phi = (\hat{\gamma}_2)^{-1}$, $\hat{\gamma}_2$ is the estimate of $\gamma_2$ which is based on the residuals $\log |y_i - x_i' \beta_0| - x_i' \hat{\theta}$, for $i = 1, \ldots, N$.

Recall that $\gamma_2$ is the dispersion parameter for the scale problem discussed in chapter five.

Next, interval estimates of the desired ratio based on trimmed samples were considered.

(B) The 90% and 95% confidence intervals for the ratios of standard deviations corresponding to these samples were calculated using the formula

$$
\left( \sqrt{\frac{S_{M2}^2}{S_{M1}^2} F_{\left( \frac{\alpha}{2}, df_1, df_2 \right)}} \sqrt{\frac{S_{M2}^2}{S_{M1}^2} F_{\left( 1 - \frac{\alpha}{2}, df_1, df_2 \right)}} \right),
$$

(7.6.2)
where $S_{tj}^2$ is the $\lambda$-trimmed variance for the $j$th sample, $j = 1, 2$, and $F_{(1-\frac{\lambda}{2}), df_1, df_2}$ is the 95th and 97.5th percentile of the $F$ distribution with denominator and numerator degrees of freedom, $df_1$ and $df_2$ respectively, where $df_j = h_j - 1$, $h_j = n_j - 2g_j$, $g_j = \lfloor \lambda n_j \rfloor$. The trimming rate is fixed at $\lambda$.

In search of interval estimates of the ratio of scale, the following table briefly describes how this is realized under each the method.

Table 7

Summary of Confidence Interval Estimation for the Ratio of Spread for the Two Sample Case

<table>
<thead>
<tr>
<th>Method (Category)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET (A)</td>
<td>The estimate was obtained using formula (7.6.1) with $\hat{\beta}^{(k-1)}$ produced iteratively. Under this estimate, a pooled estimate of dispersion $\hat{\tau}_2$ was computed from the residuals defined in (7.6.1)</td>
</tr>
<tr>
<td>RHETLoc (A)</td>
<td>The estimate was obtained using formula (7.6.1) with $\hat{\beta} = \beta$, the true value of $\beta$. Under this specified value, a pooled estimate of dispersion $\hat{\tau}_2$ was computed from the residuals defined in (7.6.1)</td>
</tr>
<tr>
<td>LSVHETtr1 (B)</td>
<td>Estimated ratio based on trimmed standard deviations was computed using formula (7.6.2) with $\lambda = 0.10$</td>
</tr>
<tr>
<td>LSVHETtr2 (B)</td>
<td>Estimated ratio based on trimmed standard deviations was computed using formula (7.6.2) with $\lambda = 0.20$</td>
</tr>
<tr>
<td>LSVHET (B)</td>
<td>Estimated ratio based on the classic standard deviations was computed using formula (7.6.2) with $\lambda = 0.0$</td>
</tr>
<tr>
<td>TrSDHET (B)</td>
<td>Estimated ratio based on the trimmed standard deviations was computed using formula (7.6.2) with $\lambda = 0.10$ trimmed mean.</td>
</tr>
</tbody>
</table>
7.6.1 Coverage Probability of Interval Estimates for $\vartheta_2/\vartheta_1$

In this section, we consider the empirical level obtained using all the methods for constructing confidence intervals for the ratio of scale at both the 90% and 95% nominal levels. The results for the cases in which the responses were generated from the $CN(0.20, 9)$ are analyzed.

Observe in Table 8 that the methods considered in the experiment yielded estimates that were reasonably close to the nominal levels, 90% and 95%, in all the designs.

We turn the empirical coverage probabilities given in Table 8. Although the two rank based method attain neither of the nominal confidence levels considered, their corresponding coverage probabilities were much closer to the desired 90% and 95% nominal levels than those of the other methods. Under such a relatively high level of contamination, the results for RHETloc are consistent with the findings by Conover and Johnson.

<table>
<thead>
<tr>
<th>Method</th>
<th>Sample Sizes : $(n_1, n_2)$</th>
<th>(20,40)</th>
<th>(40,40)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90%</td>
<td>95%</td>
<td>90%</td>
</tr>
<tr>
<td>IRHET</td>
<td>0.865</td>
<td>0.923</td>
<td>0.865</td>
</tr>
<tr>
<td>RHETloc</td>
<td>0.880</td>
<td>0.931</td>
<td>0.875</td>
</tr>
<tr>
<td>LSVHETtr1</td>
<td>0.774</td>
<td>0.850</td>
<td>0.780</td>
</tr>
<tr>
<td>LSVHETtr2</td>
<td>0.842</td>
<td>0.906</td>
<td>0.846</td>
</tr>
<tr>
<td>LSVHET</td>
<td>0.660</td>
<td>0.748</td>
<td>0.640</td>
</tr>
<tr>
<td>TrSDHET</td>
<td>0.654</td>
<td>0.739</td>
<td>0.649</td>
</tr>
</tbody>
</table>

l=Iterated, R=Rank, LSV=Least Squares Variance, HET=Heteroscedastic, HOM=Homoscedastic, loc=with location specified.

LSVHETtr1 is computed using 10% trimming rate, LSVHETtr2 is computed using 20% trimming rate.

TrSDHET =trimmed Standard Deviation is computed using 10% trimming rate.
191

(1981). The proposed method IRHET, reported empirical coverage probabilities that were close to those obtained from the RHETloc method. Except for the 20% trimmed standard deviations, all other methods not using ranks were more affected by the outliers as evidenced by much lower coverage values.

7.7 The Iterative Methods for the $p$ Group Problem

In the simulation experiments for the general $p$ group and the multiple regression we compare the performance of the proposed iterative method (IRHET) and two special cases of the method: where the scale constants are specified (RHETsca) and location coefficients are specified (RHETloc). Further, these methods are compared to the iterative LS analogue (ILSHET), the lower tail trimmed modification of ILSHET that was recommended by Davidian and Carroll (1987), following results obtained by Harvey (1976). In addition, the analysis included Huber's $M$-estimate analogue (IHMHET). This is the $M$-estimation given in Carroll and Ruppert (1982a) with $M$-estimation employed on the transformed variance function under the assumption of this study.

The recommendation of applying a trim of a few smallest absolute residuals was posed Davidian and Carroll as a strategy to circumvent the problem of inliers when the logarithm transformation is employed to estimation of scale coefficients.

Our motivation to compare to $M$-estimate emanates from the strong result on the equivalence of $R$-estimators and $M$-estimators by Jurečková (1977). Since it was shown by Davidian and Carroll (1982a) that due to the log transformation, the error terms of the new variance function are homoscedastic, Jurečková's result should also hold.
7.7.1 Estimation of Shift in Location for $p$ Groups

In this section, the iterative methods for estimating the shift in location for the general, $p$ group problem are presented. As a natural extension of the two sample group discussed above, the shifts $\mu_2 - \mu_1, \mu_3 - \mu_1, \ldots, \mu_p - \mu_1$ are of interest. The model given in (7.3.1) can be extended to

$$
y_i / \sigma_i = (x_{1i}^T / \sigma_i) \beta + \epsilon_i, \quad \sigma_i = e^{x_i^T \theta}, \text{ for } i = 1, \ldots, N,
$$

(7.7.1)

where $\beta = (\beta_0, \beta_1, \ldots, \beta_{p-1})^T = (\mu_1, \mu_2 - \mu_1, \ldots, \mu_p - \mu_1)^T$, and $\theta = (\log(\vartheta_2/\vartheta_1), \log(\vartheta_3/\vartheta_1), \ldots, \log(\vartheta_p/\vartheta_1))^T$, $\vartheta_j^2$ is the variance of the $j$th group, $j = 1, \ldots, p$. Note that by this definition, the ratio of the standard deviation of each group to that of group 1 is given by $e^{\vartheta_1}, e^{\vartheta_2}, \ldots, e^{\vartheta_{p-1}}$. Let $\theta^o = \hat{\theta}^{(r-1)}$, where $\hat{\theta}^{(r-1)}$ is the estimate of $\theta$, is the $(r-1)$th iteration. Suppose $b \in \mathcal{R}^{p+1}$, so that for the residuals $(y_i - x_{1i}^T b)/\sigma_i^o$, the rank estimate of $\beta$ in the $r$th iteration, $\hat{\beta}_N^{(r)}$, minimizes the dispersion function

$$
D_{1N}(b, \theta^o) := \sum_{i=1}^N \phi_i^+(R|y_i - x_{1i}^T b)/\sigma_i^o|) ||y_i - x_{1i}^T b)/\sigma_i^o||.
$$

(7.7.2)

Similar to the two sample problem, this objective function can be formulated as an IRLWS problem so that

$$
D_{1N}(b, \theta^o) := \sum_{i=1}^N w_i^+(b, \theta^o)(|y_i - x_{1i}^T b)/\sigma_i^o|)^2,
$$

(7.7.3)

where

$$
w_i^+(b, \theta^o) = \frac{\phi_i^+(R|y_i - x_{1i}^T b)/\sigma_i^o|/(N+1))}{||y_i - x_{1i}^T b)/\sigma_i^o||} \text{ for } ||y_i - x_{1i}^T b)/\sigma_i^o|| \neq 0
$$

$$
= 0 \text{ elsewhere,}
$$
Then, given the $k$th step in the $r$th iteration, $k_r$, the $(k_r + 1)$ step estimate of $\beta$, $\hat{\beta}_N^{(k_r)}$, minimizes the $k_r$th step dispersion given by

$$D_{1N}^*(b | b_{kr}, \theta^o) := \sum_{i=1}^{N} w_i^+(b_{kr}, \theta^o)[(y_i - x_{i1}^T b)/\sigma_{ij}^o]^2, \quad k_r = 0, 1, \ldots \quad (7.7.4)$$

Recall that $\phi^+_1(u) = \sqrt{3}(u), (0 < u < 1)$, is employed for the location problem. Table 9 gives a summary of all the methods that were used in estimation shifts in location.

### 7.7.2 Estimation of Ratio of Spread for $p$ Groups

Let us now consider the estimation of the ratios of spread for the general $p$ group problem. The linearized model used for the two sample problem can be extended to

$$\log |y_i - x_{i1}^T \theta^o| = x_i^T \theta + \log |c_i|, \quad i = 1, \ldots, N, \quad (7.7.5)$$

using $(\hat{\phi}_2/\hat{\phi}_1, \hat{\phi}_3/\hat{\phi}_1, \ldots, \hat{\phi}_p/\hat{\phi}_1) = (e^{\hat{\phi}_1}, e^{\hat{\phi}_2}, \ldots, e^{\hat{\phi}_{p-1}})$. Let $\beta^o = \hat{\beta}(r)$, where $\hat{\beta}(r)$ is the estimate of $\beta$ in the $r$th iteration. Let $t$ be an arbitrary vector $t \in \mathbb{R}^{p-1}$. Then, utilizing the residuals, $\log |y_i - x_{i1}^T b| - x_i^T t$, the rank estimate of $\theta$ in the $r$th iteration, $\hat{\theta}_N^{(r)}$, is the value $t$ such that

$$D_{2N}(\beta^o, t) := \sum_{i=1}^{N} \phi^+_2 \left( \frac{R[\log |y_i - x_{i1}^T \beta^o| - x_i^T t]}{N + 1} \right) (\log |y_i - x_{i1}^T b| - x_i^T t) \quad (7.7.6)$$

is minimized. This objective function for estimating $\theta$ can be rewritten with the IRWLS formulation

$$D_{2N}(\beta^o, t) := \sum_{i=1}^{N} w_i(\beta^o, t)[\log |y_i - x_{i1}^T b| - x_i^T t - m_v(\beta^o, t)]^2, \quad (7.7.7)$$
where the weights,

\[ w_i(\beta^o, t) = \frac{\phi_2^*(R(v_i(\beta^o, t))/(N + 1))}{[v(\beta^o, t) - m_v(\beta^o, t)]} \text{ for } v_i(\beta^o, t) \neq m_v(\beta^o, t) \]

\[ = 0, \text{ elsewhere} \]

\[ v_i(\beta^o, t) = \log|y_i - x_i^T \beta^o| - x_i^T t \text{ and } m_v(\beta^o, t) \text{ is the centering constant defined as } \nu \text{th quantile of the residuals } v_i(\beta^o, t). \text{ Note that } \nu \text{ is such that } \phi_2^*(u) < 0(\phi_2^*(u) > 0) \text{ if } u < v(u > v). \text{ Recall that in this study, we employ } \phi_2^*(u) = [\Phi^{-1}(\frac{1+u}{2})]^2, \text{ where } \Phi^{-1}(u), (0 < u < 1), \text{ is quantile of the standard normal distribution. Then, given the } l \text{th step in the } r \text{th iteration, } l_r, \text{ the } (l_r + 1) \text{-step estimate of } \theta, \hat{\theta}_N^{(l_r)} \text{ minimizes the } l_r \text{th dispersion given by} \]

\[ D_{2N}(\beta^o, t|t_{l_r}) := \sum_{i=1}^{N} w_i(\beta^o, t_{l_r})[v_i(\beta^o, t) - m_v(\beta^o, t_{l_r})]^2, \quad l_r = 0, 1, \ldots \quad (7.7.8) \]

Table 10 gives a summary of all the methods that were used in estimation of the ratio of the spread of each group with respect to group 1.
Table 9

Summary of the Methods Used in the p Group Shift in Location Problem

<table>
<thead>
<tr>
<th>Method</th>
<th>Dispersion Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET</td>
<td>The final rth iteration estimate of $\beta$, $\hat{\beta}$, is the value that minimizes the IRWLS formulation of the rank dispersion function in (7.7.4), with $\theta^o = \theta_{(r-1)}$, $\hat{\theta}_{(r-1)}$ is the rank estimate from the $(r-1)$th iteration, which is based on fitting model (7.7.5).</td>
</tr>
<tr>
<td>RHETsca</td>
<td>Let $\theta^o$ denote the true value of the scale coefficient. Then, the rank estimate of $\beta$, $\hat{\beta}$, is the value that minimizes the IRWLS formulation of the rank dispersion function in (7.7.4), with $\theta^o = \theta$.</td>
</tr>
<tr>
<td>ILSHETr</td>
<td>The final rth iteration estimate of $\beta$ is the value $\hat{\beta}$ that minimizes the LS dispersion function, $\sum_{i=1}^{N}[(y_i - x_i^T \beta)/\sigma_i]^2$, where $\sigma_i^o$ is taken to be $\hat{\sigma}<em>{i(r-1)} = \exp{x_i^T \hat{\theta}</em>{i(r-1)}}$, $\hat{\theta}_{i(r-1)}$ is the final $(r-1)$th iteration LS estimate of $\theta$ obtained from fitting model (7.7.5).</td>
</tr>
<tr>
<td>ILSHETr</td>
<td>The final rth iteration estimate of $\beta$ is the value $\hat{\beta}$ that minimizes the LS dispersion function, $\sum_{i=1}^{N}[(y_i - x_i^T \beta)/\sigma_i]^2$, where $\sigma_i^o$ is taken to be $\hat{\sigma}<em>{i(r-1)} = \exp{x_i^T \hat{\theta}</em>{i(r-1)}}$, $\hat{\theta}<em>{i(r-1)}$ is the final $(r-1)$th iteration LS estimate of $\theta$ obtained from fitting model (7.7.5) with $i = 1, \ldots, N</em>{tr}$, where 5% of the smallest absolute residuals have been trimmed.</td>
</tr>
<tr>
<td>IHMHETr</td>
<td>The final rth iteration M-estimate of $\beta$ is the value $\hat{\beta}$ that minimizes the dispersion function, $\sum_{i=1}^{N} \psi(u_i/\hat{\tau}<em>i)$, $u_i = (y_i - x_i^T \beta)/\sigma_i$, where $\psi(w) = \begin{cases} -m &amp; \text{for } w &lt; -m \ w &amp; \text{for } -m &lt; w &lt; +m \ +m &amp; \text{for } +m &lt; w \end{cases}$, $\sigma_i^o$ is taken to be $\hat{\sigma}</em>{mi(r-1)} = \exp{x_i^T \hat{\theta}<em>{i(r-1)}}$, $\hat{\theta}</em>{i(r-1)}$ is the $(r-1)$th iteration M-estimate of $\theta$ obtained from fitting model (7.7.5), $\hat{\tau}_i = 1.4826 \times \text{MAD}$, MAD is computed using the $(y_i - x_i^T \hat{\beta})/\sigma^o$.</td>
</tr>
</tbody>
</table>

I=Iterated, R=Rank, LS=Least Squares, HM=Huber’s M-estimation with $m=1.345$, HET=Heteroscedastic, HOM=Homoscedastic, sca=with scale coefficient specified, tr=trimmed absolute residuals, MAD=Median Absolute Deviation.
<table>
<thead>
<tr>
<th>Method</th>
<th>Dispersion Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET</td>
<td>The final ((r-1))th iteration estimate of (\theta) is the value (\hat{\theta}) that minimizes the IRWLS formulation of the rank dispersion function in (7.7.8), with (\beta^o = \hat{\beta}^{(r-1)}), (\hat{\beta}^{(r-1)}) is the ((r-1))th iteration rank estimate obtained by fitting model (7.7.1).</td>
</tr>
<tr>
<td>RHETloc</td>
<td>Set (\beta^o) to be the true value of the regression coefficient. Then the desired rank estimate of (\theta) is the value (\hat{\theta}) that minimizes the IRWLS formulation of the rank dispersion function in (7.7.8), with (\beta^o = \beta).</td>
</tr>
<tr>
<td>ILSHET</td>
<td>The final ((r-1))th iteration estimate of (\theta) is the value (\hat{\theta}) that minimizes the LS dispersion function, (\sum_{i=1}^{N} [(z_i(\beta^o) - x_i^T \hat{\theta})]^2), (\beta^o) is taken to be (\hat{\beta}^{(r-1)}), the ((r-1))th iteration LS estimate of (\beta) obtained from fitting model (7.7.1).</td>
</tr>
<tr>
<td>ILSHETtr</td>
<td>The final ((r-1))th iteration LS estimate of (\theta) is the value (\hat{\theta}) that minimizes the LS dispersion function, (\sum_{i=1}^{N_{tr}} [(z_i(\beta^o) - x_i^T \hat{\theta})]^2), where (z_i(\beta^o) = \log</td>
</tr>
</tbody>
</table>
| IHMHET       | The final \((r-1)\)th iteration M-estimate of \(\theta\) is the value \(\hat{\theta}\) that minimizes the dispersion function, \(\sum_{i=1}^{N} \psi(v_i/\hat{\tau}_2), v_i = (z_i(\beta^o) - x_i^T \hat{\theta})\), where \(\psi(v) = \begin{cases} -l & \text{for } v < -l \\ v & \text{for } -l < v < +l \\ +l & \text{for } +l < v \end{cases}\).  
\(z_i(\beta^o) = \log |y_i - x_i^T \beta^o|\), \(\beta^o\) is taken to be \(\hat{\beta}^{(r)}\), the \(r\)th iteration M-estimate of \(\beta\) obtained from fitting model (7.7.1), \(\hat{\tau}_2 = 1.4826 \times MAD\), \(MAD\) is computed using \(z_i(\beta^o) - x_i^T \theta\), \(i = 1, \ldots, N\). |

I=Iterated, R=Rank, LS=Least Squares, HM=Huber's M-estimation with \(l=1.345\),
HET=Heteroscedastic, HOM=Homoscedastic,
loc=with location coefficient specified,
tr=trimmed absolute residuals,
MAD=Median Absolute Deviation.
Next we consider the first experiment on the three group with unequal variances.

7.8 Example 2: Three Groups

In this experiment, the responses were drawn from three populations which had unequal dispersion. The question of estimating the shift in location was restricted to the cases from group 1 to group 2 and group 1 to group 3. It was assumed that the variance function is known up to scale, which is a fundamental condition to obtaining estimated shifts. We sought to demonstrate how the methods performed in the presence of a few extreme large outliers to contrast with the case of no outliers in the rarely satisfied standard normal $N(0, 1)$.

The experiment was performed in 10,000 simulation trials subject to the following conditions:

(i) Shifts in location: $\mu_2 - \mu_1 = 2.0; \mu_3 - \mu_1 = 4.0$.

(ii) Ratio of Spread: $\vartheta_2/\vartheta_1 = e^{1.5}, \vartheta_3/\vartheta_1 = e^{3.0}$ so that the logarithm of the ratios are $\log(\vartheta_2/\vartheta_1) = 1.5, \log(\vartheta_3/\vartheta_1) = 3.0$.

(iii) Sample sizes for groups: $n_1 = 40, n_1 = 60, n_1 = 80$.

(iv) Distributions: (1) standard normal $N(0, 1)$; (2) contaminated normal $CN(0.05, 100)$; (3) Slash $= normal(0, 1)/unif(0, 1)$ and (4) Laplace.

With a view to compare the responsiveness of the methods to a few extremely large outliers, the choice of $CN(0.05, 100)$ is similar to the selection of the so called "one-wild" in which, for example, a sample of size=20, one out of the 20 responses is drawn from a $N(0, 100)$, (Lax (1985)). In contrast, this study utilizes the percentage 5% so as to allow for the contamination level to be proportional to the different sample sizes for the different groups that were under consideration in this experiment.
7.8.1 Difference in Location with Respect to Group 1

Averages and standard deviations of the estimates obtained from the simulation trials are presented in the tables that follow.

Table 11

Average and Standard Deviation of Shifts of Each Group from Group 1 Based on 10,000 Simulations for case $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\varphi_2/\varphi_1) = 1.5$, $\log(\varphi_3/\varphi_1) = 3.0$, under the Normal and Contaminated Normal Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>$N(0.00, 1.0)$</th>
<th>$CN(0.05, 100)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\mu}_2 - \hat{\mu}_1$</td>
<td>$\hat{\mu}_3 - \hat{\mu}_1$</td>
</tr>
<tr>
<td>Ave.</td>
<td>SD</td>
<td>Ave</td>
</tr>
<tr>
<td>IRHET</td>
<td>1.996</td>
<td>0.613</td>
</tr>
<tr>
<td>RHETsca</td>
<td>1.996</td>
<td>0.615</td>
</tr>
<tr>
<td>ILSHET</td>
<td>2.008</td>
<td>0.603</td>
</tr>
<tr>
<td>ILSHETtr</td>
<td>1.996</td>
<td>0.596</td>
</tr>
<tr>
<td>IHMHET</td>
<td>2.001</td>
<td>0.621</td>
</tr>
</tbody>
</table>

I= Iterative, R=Rank, L=Least Squares, M=M-estimator, tr=trimmed absolute residuals, HET=Heteroscedastic, sca=with scale specified.

It is seen from Table 11 that all the methods reported that average estimates that were close to 2.0 and 4.0 under both distributions. While all the methods reported larger standard deviations under the contaminated normal than in the standard normal, it is noted that the ILS methods reported the least variability under $N(0, 1)$ and the largest variability under $CN(0.05, 100)$. Thus, it is seen that the robust methods, IRHET, RHETsca, and IHMHET were much more resistant to the effect of outliers.

Consider results obtained under Slash and Laplace distribution given in Table 12. Observe that under the Slash distribution, only methods IRHET, RHETsca and IHMHET reported estimates that were close to 2.0 and 4.0. ILSHET and ILSHETtr, in contrast had broken down as would be anticipated under the thicker tails of Cauchy-like distribution. The variability reported by the method similarly exhibited the same pattern with ILS.
Table 12

Average and Standard Deviation of Shifts of Each Group from Group 1 Based on 10,000 Simulations for case $\mu_2 - \mu_1 = 2.0, \mu_3 - \mu_1 = 4.0, \log(\theta_2/\theta_1) = 1.5, \log(\theta_3/\theta_1) = 3.0$, under the Slash and Laplace Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>Slash Ave.</th>
<th>Slash SD</th>
<th>Laplace Ave.</th>
<th>Laplace SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET</td>
<td>2.000</td>
<td>1.636</td>
<td>2.002</td>
<td>0.703</td>
</tr>
<tr>
<td>RHETsca</td>
<td>2.003</td>
<td>1.617</td>
<td>1.999</td>
<td>0.705</td>
</tr>
<tr>
<td>ILSHET</td>
<td>-5.407</td>
<td>561.655</td>
<td>1.985</td>
<td>0.853</td>
</tr>
<tr>
<td>ILSHET_tr</td>
<td>-0.188</td>
<td>502.840</td>
<td>1.997</td>
<td>0.857</td>
</tr>
<tr>
<td>IHMHET</td>
<td>2.017</td>
<td>1.752</td>
<td>1.993</td>
<td>0.733</td>
</tr>
</tbody>
</table>

I= Iterative, R=Rank, L=Least Squares, M=M-estimator, tr=trimmed absolute residuals, HET=Heteroscedastic, sca=with scale specified.

method yielding inflated standard deviations. Although both the rank and M estimators gave much larger standard deviations than those seen under the normal distributions, the measures of variability were not extremely inflated. This indicates that methods were fairly stable under heavy contamination. Under the Laplace, all the methods yielded values that were close to the true values 2.0 and 4.0. It is noted that ILS methods still yielded larger standard deviations due to the lighter contamination of the distribution. The results obtained by IRHET and RHETsca had smaller variability than those reported by IHMHET under both distributions.

In the sequel, we utilize estimated relative efficiency to compare the performance of the methods. A brief definition of relative efficiency is furnished in what follows.
7.8.2 Estimated Relative Efficiency

Suppose for an arbitrary parameter we have two estimators \( T_1 \) and \( T_2 \). Then the relative efficiency of \( T_2 \) relative to \( T_1 \), \( \text{eff}(T_2, T_1) \) is given by

\[
\text{eff}(T_2, T_1) = c_2^2 / c_1^2,
\]

where \( c_j \) is the efficacy of the \( j \)th estimator which is defined by \( c_j = 1/\sigma_j \), \( \sigma_j^2 \) is the variance of the distribution. When the relative efficiency is equal to 1 then the estimators are equivalently efficient. If the relative efficiency is less than 1, then we have a measure of how much the sample size under \( T_2 \) should be increased to be as efficient as \( T_1 \). If the relative efficiency is greater than 1, then we have a measure of by how much the sample size under \( T_1 \) should be increased for the method to be as efficient as \( T_2 \).

For each method, an estimate of the dispersion of the distribution is available and it can be used to compute the estimated relative efficiency. For example, our estimate of relative efficiency of IRHET to RHETsca is given by

\[
\text{ere}(\text{IRHET}, \text{RHETsca}) = 1/\hat{\gamma}_{IR}^2 / 1/\hat{\gamma}_{\text{sca}}^2.
\]

It is observed in Table 13 that the proposed method IRHET is slightly less efficient than the RHETsca under \( N(0, 1) \). Recalling the well known fact that under homoscedasticity the Wilcoxon estimator is about 95% as efficient as the LS estimator under the standard normal, it is seen that this property is approximately confirmed by these results, if we take the relative efficiency of ILSHET with respect to RHETsca to be approximately 1.

It is seen that the relative efficiency of all the methods except IRHET deteriorate in the presence of outliers under the \( CN(0.05, 100) \). Observe that for IRHET, the decline in estimated relative efficiency as one increased contamination from \( N(0, 1) \) to \( CN(N0.05, 100) \) is negligible, only 0.958 to 0.944. Substantial deterioration of the esti-
Estimated Relative Efficiency of Shift in Location Based on 10,000 Simulations where $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\theta_2/\theta_1) = 1.5$, $\log(\theta_3/\theta_1) = 3.0$, under the Normal and Contaminated Normal Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>Distribution</th>
<th>$N(0.00, 1.0)$</th>
<th>$C\mathcal{N}(0.05, 100)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ERE</td>
<td>ERt</td>
<td>ERt</td>
</tr>
<tr>
<td>IRHET</td>
<td>0.958</td>
<td>0.944</td>
<td></td>
</tr>
<tr>
<td>RHETsca</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>ILSHET</td>
<td>1.021</td>
<td>0.138</td>
<td></td>
</tr>
<tr>
<td>ILSHETtr</td>
<td>0.921</td>
<td>0.129</td>
<td></td>
</tr>
<tr>
<td>IHMHET</td>
<td>1.020</td>
<td>0.689</td>
<td></td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, L=Least Squares, 
M=M-estimator, tr=trimmed trimmed residuals, HET=Heteroscedastic, 
residuals, sca=with scale specified, 
ERE= Estimated Relative Efficiency 
with respect to RHETsca.

mated relative efficiency was reported by ILSHET, ILSHETtr, and IHMHET as contamination was increased from 0 to 5% under the normal distribution.

For the non-normal distribution, Table 14 shows an even bigger decline in relative efficiency than that seen under $C\mathcal{N}(0.05, 100)$. However, IRHET incurred a 6% decline under the Slash distribution in comparison to the efficiency under the $N(0, 1)$. For the Laplace distribution, given its moderate contamination, the proposed method reported estimated relative efficiency that was close to that obtained in $N(0, 1)$.

### 7.8.3 Log of Ratio of Scale with Respect to Group 1

We now consider the estimates of $\log(\theta_j/\theta_1)$ for groups $j = 2, 3$. The results reported from all the other methods considered are compared to those obtained by the rank based estimator for the case in which the location is specified, RHETloc.

Consider Table 15. It is seen that the proposed estimator of $\theta$, IRHET yielded
Table 14

Estimated Relative Efficiency of Shift in Location Based on 10,000 Simulations where \( \mu_2 - \mu_1 = 2.0, \mu_3 - \mu_1 = 4.0, \log(\varphi_2/\varphi_1) = 1.5, \log(\varphi_3/\varphi_1) = 3.0 \), under the Slash and Laplace Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>Distribution</th>
<th>Slash</th>
<th>Laplace</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ERE</td>
<td>ERE</td>
</tr>
<tr>
<td>IRHET</td>
<td></td>
<td>0.896</td>
<td>0.936</td>
</tr>
<tr>
<td>RHETsca</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>ILSHET</td>
<td></td>
<td>0.000</td>
<td>0.369</td>
</tr>
<tr>
<td>ILSHETtr</td>
<td></td>
<td>0.000</td>
<td>0.330</td>
</tr>
<tr>
<td>IHMHET</td>
<td></td>
<td>0.017</td>
<td>0.573</td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, L=Least Squares, M=M-estimator, tr=trimmed trimmed residuals, HET=Heteroscedastic, sca=with scale specified, ERE=Estimated Relative Efficiency with respect to RHETsca.

Table 15

Averages and Standard Deviation of Logarithm of the Ratio of Scale for Each Group With Respect to Group 1, where \( \mu_2 - \mu_1 = 2.0, \mu_3 - \mu_1 = 4.0, \log(\varphi_2/\varphi_1) = 1.5, \log(\varphi_3/\varphi_1) = 3.0 \), under the Normal and Contaminated Normal Distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>[N(0.00, 1.0)]</th>
<th>[CN(0.05, 100)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>(\log(\varphi_2/\varphi_1))</td>
<td>(\log(\varphi_3/\varphi_1))</td>
</tr>
<tr>
<td>IRHET</td>
<td>1.509</td>
<td>0.150</td>
</tr>
<tr>
<td>RHETloc</td>
<td>1.504</td>
<td>0.148</td>
</tr>
<tr>
<td>ILSHET</td>
<td>1.507</td>
<td>0.228</td>
</tr>
<tr>
<td>ILSHETtr</td>
<td>1.458</td>
<td>0.191</td>
</tr>
<tr>
<td>IHMHET</td>
<td>1.504</td>
<td>0.225</td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, L=Least Squares, M=M-estimator, HET=Heteroscedastic, tr=trimmed absolute residuals, loc=with location specified.
averages and standard deviations that were close to those obtained under the more informed RHETloc, where location is specified. These two methods realized estimates with the least variability. IHMHET reported larger standard deviations under N(0, 1). The ILS methods yielded larger standard deviations in general. Note that ILSHETtr had smaller standard deviations than IHMHET but its estimate of $\log(\theta_2/\theta_1)$ was biased while the latter yielded estimates that were close to the true value under N(0, 1).

Table 16

Averages and Standard Deviation of Logarithm of the Ratio of Scale for Each Group With Respect to Group 1, where $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\theta_2/\theta_1) = 1.5$, $\log(\theta_3/\theta_1) = 3.0$, under the Slash and Laplace Distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Slash</th>
<th>Laplace</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ave.</td>
<td>SD</td>
</tr>
<tr>
<td>Method</td>
<td>log($\theta_2/\theta_1$)</td>
<td>log($\theta_3/\theta_1$)</td>
</tr>
<tr>
<td>IRHET</td>
<td>1.498</td>
<td>0.358</td>
</tr>
<tr>
<td>RHETloc</td>
<td>0.121</td>
<td>0.326</td>
</tr>
<tr>
<td>ILSHET</td>
<td>1.524</td>
<td>1.315</td>
</tr>
<tr>
<td>ILSHETtr</td>
<td>1.634</td>
<td>1.445</td>
</tr>
<tr>
<td>IHMHET</td>
<td>1.477</td>
<td>0.306</td>
</tr>
</tbody>
</table>

I= Iterative, R=Rank, L=Least Squares, M=M-estimator, HET=Heteroscedastic, tr=trimmed absolute residuals, loc=with location specified.

Next consider the results obtained under non-normal distributions presented in Table 16. Observe that the estimates reported by RHETloc were not very close to the true values under Slash. For this distribution, IHMHET reported the least variability. The proposed method reported standard deviations that were smaller than those from IHMHET. In general, under these more heavily outlier populated distributions, the ILSHET and ILSHETtr methods reported the largest standard deviations. In addition, the latter method had biased estimates.
Table 17

Estimated Relative Efficiency of Scale Coefficient With Respect to Group 1 Based on 10,000 Simulations for $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\theta_2/\theta_1) = 1.5$ and $\log(\theta_3/\theta_1) = 3.0$, under the Normal and Contaminated Normal Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>$N(0.00, 1.0)$ ERE</th>
<th>$CN(0.05, 100)$ ERE</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET</td>
<td>1.255</td>
<td>3.199</td>
</tr>
<tr>
<td>RHETloc</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>ILSHET</td>
<td>1.018</td>
<td>1.002</td>
</tr>
<tr>
<td>ILSHETtr</td>
<td>3.046</td>
<td>3.452</td>
</tr>
<tr>
<td>IHMHET</td>
<td>1.241</td>
<td>3.195</td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, L=Least Squares, M=M-estimator, HET=Heteroscedastic, tr=trimmed absolute residuals, loc=with location specified, ERE= Estimated Relative Efficiency with respect to RHETloc.

7.8.4 Estimated Relative Efficiency

Analogous to the shift in location problem, we compute the estimated relative efficiency of each method with respect to RHETloc. Thus, for the proposed method the estimated relative efficiency given by

$$ere(IHRHET, RHETloc) = \frac{1}{\gamma^2_{IRHET}} / \frac{1}{\gamma^2_{RHETloc}}.$$  

It is seen in Table 17 that the proposed method IRHET is more efficient than the RHETloc under $N(0, 1)$. Recalling the well known fact that, under homoscedasticity, the Wilcoxon estimator is about 95% as efficient as the LS estimator under the standard normal, it is seen that this property is approximately confirmed by these results. Observe that the relative efficiency of ILSHET with respect to RHETloc, 1.018, is approximately 1.

It is seen that the relative efficiency of all the methods except IRHET deteriorate in
Table 18

Estimated Relative Efficiency of Scale Coefficient With Respect to Group 1 Based on 10,000 Simulations for $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\theta_2/\theta_1) = 1.5$ and $\log(\theta_3/\theta_1) = 3.0$, under the Slash and Laplace Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>Distribution</th>
<th>Slash</th>
<th>Laplace</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ERE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRHET</td>
<td>24.568</td>
<td>2.947</td>
<td></td>
</tr>
<tr>
<td>RHETloc</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>ILSHET</td>
<td>0.965</td>
<td>1.012</td>
<td></td>
</tr>
<tr>
<td>ILSHETtr</td>
<td>6.347</td>
<td>2.495</td>
<td></td>
</tr>
<tr>
<td>IHMHET</td>
<td>7.621</td>
<td>1.781</td>
<td></td>
</tr>
</tbody>
</table>

l=Iterative, R=Rank, L=Least Squares, M=M-estimator, HET=Heteroscedastic, tr=trimmed absolute residuals, loc=with location specified, ERE= Estimated Relative Efficiency with respect to RHETloc.

presence of outliers under the $CN(0.05, 100)$. IRHET gained in efficiency over the RHETloc under the contaminated normal distribution. From Table 18, we can see that IRHET was efficient when responses came from Slash and Laplace distribution.

7.9 Example 3: The Inlier Issue in 3 Group Problem

As pointed out in Davidian and Carroll (1987) that when the logarithm of absolute residuals strategy for fitting variance function model is employed, it induces extremely large values as result of the instability of the logarithm function when it is applied to values that are very close to zero. These induced extremely large values are the so called inliers. To curtail their effects, the authors recommended trimming a few smallest absolute residuals prior to applying the logarithm transformation. Since taking of logs is employed when obtaining estimates of scale, only the results of this estimation problem are presented, subsequent estimates of the locations were also indirectly affected by the
transformation.

In this experiment, the effect of the various levels of the trimming rate on the robustness and efficiency of the methods under consideration for small to moderate contamination was investigated. The experiment considered inlier contamination for cases in which outliers were absent as well as those in which outliers were present. Recall that, in this study, differences between estimates that are less than 0.0001 are considered to be zero. Hence, this tolerance was used as a cut off value in the generation of responses that produce absolute values that were close to 0. The variance of the inlier contamination portion of the mixture normal distribution was fixed at $1/10000^2 = 10^{-10}$, so that a 5% inlier-contaminated normal, for example, was generated by

$$CN(0.05, 10^{-10}) = 0.05 \times N(0.0, 10^{-10}) + 0.95 \times N(0.0, 1),$$

Further, for the responses drawn from the model in which both inliers and outliers are present, the inlier-outlier contaminated process under the normal can be denoted by $CN(\epsilon_1, \epsilon_2, \nu_1^2, \nu_2^2)$, where $\epsilon_1$ is the proportion of inliers, $\epsilon_2$ is the proportion of outliers, $\nu_1^2$ is variance of the inlier contamination population, and $\nu_2^2$ is variance of the outlier contamination population. Thus, for an inlier-outlier contaminated distribution with the proportions both set at 0.05 and variance of the outlier contamination portion of 9.0,

$$CN(0.05, 0.05, 10^{-10}, 9) = 0.05 \times N(0.0, 10^{-10}) + 0.90 \times N(0, 1) + 0.05 \times N(0, 9).$$

As Davidian and Carroll (1987) observed that under the logarithm absolute residuals strategy the inliers are a more serious issue than the outliers to the extent that the robustness of the method is of concern. Hence, this experiment considered inlier contamination to be no more than 10% while the outlier contamination was allowed to be no more than 5%. With regard to the trimming as an inlier effect curtailing device, our experience showed that rates over 5% were fairly counter productive, so the results are
limited to levels that were at most this rate.

The investigation focussed on the robustness and relative efficiency of all the methods subject to (1) trimming rates of 0%, 2.5% and 5% under the 5% inlier contamination levels in the absence of outliers; (2) trimming rates 0% and 5% when inlier-outlier contamination rates were equal, (5%, 5%) and when the inlier contamination was twice that of outliers (10%, 5%). The lower tail trimmed methods are essentially same methods that were considered in the previous example with the additional step in which trimmed absolute residual are used estimate the scale coefficients.

7.9.1 Estimates Obtained with Methods using Lower Tail Trim

Table 19

Averages and Standard Deviations of Estimates of Logarithm of Ratio of Spread Based on 10,000 Simulations where $\mu_2 - \mu_1 = 2.0, \mu_3 - \mu_1 = 4.0, \log(\hat{\theta}_2/\hat{\theta}_1) = 1.5, \log(\hat{\theta}_3/\hat{\theta}_1) = 3.0$, under Inlier Contaminated Distribution

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>Lower-tail Trimming Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0% trim</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>IRHET</td>
<td>$\log(\hat{\theta}_2/\hat{\theta}_1)$</td>
<td>1.509</td>
</tr>
<tr>
<td></td>
<td>$\log(\hat{\theta}_3/\hat{\theta}_1)$</td>
<td>3.012</td>
</tr>
<tr>
<td>RHETloc</td>
<td>$\log(\hat{\theta}_2/\hat{\theta}_1)$</td>
<td>1.505</td>
</tr>
<tr>
<td></td>
<td>$\log(\hat{\theta}_3/\hat{\theta}_1)$</td>
<td>3.006</td>
</tr>
<tr>
<td>ILSHET</td>
<td>$\log(\hat{\theta}_2/\hat{\theta}_1)$</td>
<td>1.458</td>
</tr>
<tr>
<td></td>
<td>$\log(\hat{\theta}_3/\hat{\theta}_1)$</td>
<td>2.911</td>
</tr>
<tr>
<td>IHMHET</td>
<td>$\log(\hat{\theta}_2/\hat{\theta}_1)$</td>
<td>1.504</td>
</tr>
<tr>
<td></td>
<td>$\log(\hat{\theta}_3/\hat{\theta}_1)$</td>
<td>3.001</td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, LS=Least Squares, HM=Huber’s M-estimation, HET=Heteroscedastic, loc=with location specified, tr=trimmed.

Consider the results given in Table 19. Note that except for RHETloc, observe that with the trimming rate increased from 0% to 5%, the resulting gain in reduction of variability reported by the estimates was also accompanied by an increase in bias with all
the methods yielding averages of \( \hat{\theta}_1 \log(\hat{\theta}_2/\hat{\theta}_1) \) that were 0.04 below the true value of 1.5. Note that for the 2.5% trimming rate, the IRHET and RHETloc methods reported averages that were approximately close to the true values. This may be due to the fact that the trimming was too small to have any notably significant effect.

We now consider the results of estimating the scale coefficients under the inlier-

Table 20

Averages and Standard Deviations of Estimates of Logarithm of Ratio of Spread Based on 10,000 Simulations where \( \mu_2 - \mu_1 = 2.0, \mu_3 - \mu_1 = 4.0, \log(\theta_2/\theta_1) = 1.5, \log(\theta_3/\theta_1) = 3.0 \), under Inlier-Outlier Contaminated Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>0% Trim Mean</th>
<th>0% Trim SD</th>
<th>5% Trim Mean</th>
<th>5% Trim SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET</td>
<td>( \log(\theta_2/\theta_1) )</td>
<td>1.507</td>
<td>0.179</td>
<td>1.509</td>
<td>0.150</td>
</tr>
<tr>
<td></td>
<td>( \log(\theta_3/\theta_1) )</td>
<td>3.009</td>
<td>0.169</td>
<td>3.012</td>
<td>0.144</td>
</tr>
<tr>
<td>RHETloc</td>
<td>( \log(\theta_2/\theta_1) )</td>
<td>1.504</td>
<td>0.186</td>
<td>1.505</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td>( \log(\theta_3/\theta_1) )</td>
<td>3.005</td>
<td>0.177</td>
<td>3.006</td>
<td>0.155</td>
</tr>
<tr>
<td>ILSHET</td>
<td>( \log(\theta_2/\theta_1) )</td>
<td>1.458</td>
<td>0.191</td>
<td>1.458</td>
<td>0.191</td>
</tr>
<tr>
<td></td>
<td>( \log(\theta_3/\theta_1) )</td>
<td>2.911</td>
<td>0.185</td>
<td>2.911</td>
<td>0.185</td>
</tr>
<tr>
<td>IHMHET</td>
<td>( \log(\theta_2/\theta_1) )</td>
<td>1.500</td>
<td>0.230</td>
<td>1.504</td>
<td>0.225</td>
</tr>
<tr>
<td></td>
<td>( \log(\theta_3/\theta_1) )</td>
<td>2.996</td>
<td>0.220</td>
<td>3.001</td>
<td>0.215</td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, LS=Least Squares, HM= Huber's M-estimation, HET=Heteroscedastic, loc=with location specified, tr=trimmed

outlier contaminated normal which are presented in Table 20. It is observed that under \( CN(0.05, 0.05, 10^{-10}, 9) \) and \( CN(0.10, 0.05, 10^{-10}, 9) \), the average estimates decreased in magnitude when the inlier contamination increased from 5% to 10% for the methods
RHETloc and IHMHET while IRHET estimates increased. Further, under the latter contaminated normal distribution, slightly smaller standard deviations were realized.

It is seen that, with the exception of results reported by RHETloc, considerably large negative bias was observed in the cases in which the inlier and outlier contamination levels were equal. In particular, under the 5% trimming rate, the problem of large negative bias was exacerbated. Observe that the average estimate of $\log(\theta_2/\theta_1)$ reported by IRHET was 1.428. The methods ILSHET and IHMHET reported values that were less than 1.463.

Consider the case where the inlier contamination is larger than the outlier contamination. Although the methods reported smaller negative bias than the case above, the proposed method had the largest bias, indicating that it had been most adversely affected by the lower tail trim.

7.9.2 Estimated Relative Efficiency of Methods

In this section, for the trimming rates under consideration, the estimated efficiency rates of each method under several combinations inlier-outlier contamination rates is presented. The analysis is focused on the results of estimated relative efficiency with respect to RHETloc and these are presented in Table 21. The results indicate that, in general, there was decreasing efficiency with increasing levels of trimming from 0% to 5% for all methods. There was a sharp increase in the efficiency of IHMHET from 0% to 2.5%.

Observe that at 0% trimming rate, that ILSHET reported much larger value than 1 in contrast to the other methods. This is because without the curtailing device, the full effect of the inliers significantly reduced the variability of the method. However, this reduced variability, notwithstanding a desirable property, was a consequence of the inlier effect.

It should be noted that the efficiency of IRHET was not as responsive to the trimming rate increase as were the other two methods.
Table 21

Estimated Relative Efficiency of Logarithm of Ratio of Spread Based on 10,000 Simulations where $\mu_2 - \mu_1 = 2.0$, $\mu_3 - \mu_1 = 4.0$, $\log(\theta_2/\theta_1) = 1.5$, $\log(\theta_3/\theta_1) = 3.0$, under Inlier Contaminated Distribution

<table>
<thead>
<tr>
<th>Method</th>
<th>$CN(0.05, 10^{-10})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower-tailTrimming Rate</td>
</tr>
<tr>
<td></td>
<td>0%</td>
</tr>
<tr>
<td>ERE</td>
<td>ERE</td>
</tr>
<tr>
<td>IRHET</td>
<td>1.233</td>
</tr>
<tr>
<td>RHETloc</td>
<td>1.000</td>
</tr>
<tr>
<td>ILSHET</td>
<td>2.992</td>
</tr>
<tr>
<td>IHMHET</td>
<td>1.341</td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, LS=Least Squares, HM=Huber’s M-estimation, HET=Heteroscedastic, loc=with location specified, tr=trimmed.

We finally turn to the efficiency results of the methods for the inlier-outlier contaminated normal distributions that are displayed in Table 22. The table shows that using the 5% trimming rate decreased the relative efficiency of IRHET and IHMHET methods under both contaminated distributions $CN(0.05, 0.05, 10^{-10}, 9)$ and $CN(0.05, 0.05, 10^{-10}, 9)$. In contrast, ILSHET with 5% trim realized a net gain in efficiency under the same distributions. This is the “woe in disguise” caused by the bias that comes along with increased efficiency.

These results suggest that the trimming absolute residuals before taking the logarithms does not improve the performance of the estimators as had been suggested. This is largely due to the fact that for the rank methods, the logarithm function had already been accounted for via the scores that were employed. Consequently, the trimming was altering the structure of the data leading to poorer results.
Table 22

Estimated Relative Efficiency of Logarithm of Ratio of Spread Based on 10,000 Simulations where \(\mu_2 - \mu_1 = 2.0, \mu_3 - \mu_1 = 4.0, \log(\theta_2/\theta_1) = 1.5, \log(\theta_3/\theta_1) = 3.0\), under Inlier-Outlier Contaminated Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>0% Trim</th>
<th>Contaminated Normal Distributions</th>
<th>5% Trim</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(CN(0.05, 0.05, 10^{-10}, 9))</td>
<td>(CN(0.10, 0.05, 10^{-10}, 9))</td>
</tr>
<tr>
<td>IRHET</td>
<td>1.726</td>
<td>ERE 1.233</td>
<td>ERE 1.228</td>
</tr>
<tr>
<td>RHETloc</td>
<td>1.000</td>
<td>ERE 1.000</td>
<td>ERE 1.000</td>
</tr>
<tr>
<td>ILSHET</td>
<td>2.869</td>
<td>ERE 0.924</td>
<td>ERE 3.951</td>
</tr>
<tr>
<td>IHMHET</td>
<td>1.903</td>
<td>ERE 2.992</td>
<td>ERE 1.496</td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, LS=Least Squares, HM=Huber's M-estimation, HET=Heteroscedastic, loc=with location specified, tr=trimmed.

7.10 Example 4: Multiple Regression Model

In this experiment, the design consisted of 2 predictors which were used to fit a linear model as well as estimate the heteroscedastic structure via a variance function model that is being considered in this study. The focus of the experiment was to explore the performance of the methods under non-normal distribution in contrast to the normal distribution. In particular, the study considers the responsiveness of the methods to departures from normality as observed under thick-tailed and light-tailed distributions. The relative efficiency of the methods of estimation considered was also sought furnish a comparative analysis with respect to variability. In this experiment, we considered the
model

\[ y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sigma_i e_i, \quad i = 1, \ldots, N, \]  

(7.10.1)

where

\[ \sigma_i = e^{\theta_1 x_{1i} + \theta_2 x_{2i}}, \]  

(7.10.2)

and \( e_1, \ldots, e_N \) are independently and identically distributed errors with common distribution \( F \). Clearly, (7.10.1) is a case of the multiplicative heteroscedasticity that is being considered in the current study. We performed 10,000 simulation trials each with a sample of size 60 under the following conditions:

(i) Regression coefficients were such that \( \beta_0 = 2.0; \beta_1 = 3.0; \beta_2 = 4.0. \)

(ii) Scale coefficients were such that \( \theta_1 = 1.25 \) and \( \theta_2 = 1.5. \)

(iii) The design was constructed so that the predictors were obtained as follows:

(1) \( x_1 = (x_{11}, \ldots, x_{1N})^T \) was obtained by generating a single set from \( \text{unif}(-2, 2) \) that was used in all trials.

(2) \( x_2 = (x_{21}, \ldots, x_{2N})^T \) was obtained by generating a single set from \( \text{unif}(4, 10) \) that was used in all trials.

(iv) Further, to guarantee convergence of the estimation in each trial the design was orthogonalized using a Gram-Schmidt process.

(v) The errors, \( e_1, \ldots, e_N \) were drawn from two normal distributions and two non-normal distributions. These are described in greater detail below.

The experiment was conducted under four distributions with distinct tail behavior, a factor that is crucial to the extent that estimators of scale coefficients with good robustness properties are desired. The following distributions were considered:
(i) standard normal, $N(0, 1)$, which is the case without outliers.

(ii) contaminated normal, $CN(0.05, 100)$, which is similar to the so-called one - wild proposed by Tukey (see Hoaglin, et. al. (1983)). Our version of the "one-wild" ensured very few but extremely large outlying values.

(iii) Slash- $N(0, 1)/unif(0, 1)$, which has tail behavior that is similar to that of a Cauchy distribution which guarantees a substantially large presence of outlying values.

(iv) Laplace, due to lighter tail area, guarantees a moderate presence of outlying values.

Next, we discuss the estimation results under these conditions.

7.10.1 Estimates of Regression Coefficients

Based on the 10,000 simulations, averages and standard deviations of the estimates are presented in the tables that follow. Consider the results from normal distribution given in Table 23. It is seen that ILSHET was very sensitive to the presence of a few extremely large outlying values while the other three methods were resistant. Observe that the LS method had a slightly larger increase in standard deviation going from the standard normal to the contaminated normal. Under both distributions, all four methods reported averages that were very close if not equal to the true values of the regression coefficients.

Let us turn to the estimation results from the non-normal distributions given in Table 24. Except for ILSHET, all of the methods converged to the true values of the regression coefficients. ILSHET had broken down under the heavy contamination of the Slash distribution, as evidenced by the extremely inflated average and standard deviations of the estimates.

The outlier resistant methods are discussed next. As expected the standard deviations of the estimates increased with increase in the proportion of outliers. Neverthe-
Table 23

Averages and Standard Deviations of Estimates of Regression Coefficients Based on 10,000 Simulations where \( \beta_0 = 2.0, \beta_1 = 3.0, \beta_2 = 4.0, \theta_1 = 1.25, \theta_2 = 1.50 \), under Standard Normal and Contaminated Error Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>Normal Distribution</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( N(0.0,1) )</td>
<td>( CN(0.05,100) )</td>
<td>( N(0.0,1) )</td>
<td>( CN(0.05,100) )</td>
</tr>
<tr>
<td>IRHET</td>
<td>( \beta_0 )</td>
<td>2.000</td>
<td>0.013</td>
<td>2.000</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>3.000</td>
<td>0.002</td>
<td>3.000</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>( \beta_2 )</td>
<td>4.000</td>
<td>0.004</td>
<td>4.000</td>
<td>0.004</td>
</tr>
<tr>
<td>RHETsca</td>
<td>( \beta_0 )</td>
<td>2.000</td>
<td>0.005</td>
<td>2.000</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>3.000</td>
<td>0.001</td>
<td>3.000</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>( \beta_2 )</td>
<td>4.000</td>
<td>0.001</td>
<td>4.000</td>
<td>0.002</td>
</tr>
<tr>
<td>ILSHET</td>
<td>( \beta_0 )</td>
<td>2.000</td>
<td>0.008</td>
<td>1.993</td>
<td>0.589</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>3.000</td>
<td>0.001</td>
<td>2.997</td>
<td>0.298</td>
</tr>
<tr>
<td></td>
<td>( \beta_2 )</td>
<td>4.000</td>
<td>0.002</td>
<td>4.000</td>
<td>0.072</td>
</tr>
<tr>
<td>IHMHET</td>
<td>( \beta_0 )</td>
<td>2.000</td>
<td>0.006</td>
<td>2.000</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>3.000</td>
<td>0.001</td>
<td>3.000</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>( \beta_2 )</td>
<td>4.000</td>
<td>0.001</td>
<td>4.000</td>
<td>0.002</td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, LS=Least Squares, 
HM=Huber’s M-estimation, 
HET=Heteroscedastic, sca=with scale specified.

less, the effect of the extreme values on the estimates was curtailed. While RHETsca reported smallest deviations under every distribution, it was seconded by IRHET under the \textit{Slash} distribution. In contrast, under the \textit{Laplace} distribution, RHETsca was seconded by IHMHET.

The results in Table 24 can be attributed to under-performance of the signed-rank Wilcoxon scores when utilized in the \textit{Cauchy}-like distributions. However, the IRHET method still reported the smallest standard deviations. For the IHMHET, larger standard deviations may be attributed to using the default constant value 1.345 in the IHMHET method when other values of \( k \) could be more suitable. It is also known that Huber’s Proposal 2 produces more robust estimates of scale than the standard Huber’s estimate.
Table 24

Averages and Standard Deviations of Estimates of Regression Coefficients Based on 10,000 Simulations where $\beta_0 = 2.0$, $\beta_1 = 3.0$, $\beta_2 = 4.0$, $\theta_1 = 1.25$, $\theta_2 = 1.50$, under Slash, and Laplace Error Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>( \beta_0 )</th>
<th>Mean</th>
<th>SD</th>
<th>( \beta_1 )</th>
<th>Mean</th>
<th>SD</th>
<th>( \beta_2 )</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET</td>
<td>$\beta_0$</td>
<td>2.000</td>
<td>0.072</td>
<td>$\beta_1$</td>
<td>3.000</td>
<td>0.012</td>
<td>$\beta_2$</td>
<td>4.000</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>$\beta_0$</td>
<td>2.001</td>
<td>0.079</td>
<td>$\beta_1$</td>
<td>3.000</td>
<td>0.015</td>
<td>$\beta_2$</td>
<td>4.000</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>$\beta_0$</td>
<td>15.797</td>
<td>1097.651</td>
<td>$\beta_1$</td>
<td>7.436</td>
<td>394.663</td>
<td>$\beta_2$</td>
<td>4.682</td>
<td>106.957</td>
</tr>
<tr>
<td></td>
<td>$\beta_0$</td>
<td>1.977</td>
<td>2.115</td>
<td>$\beta_1$</td>
<td>2.997</td>
<td>0.227</td>
<td>$\beta_2$</td>
<td>3.993</td>
<td>0.717</td>
</tr>
</tbody>
</table>

Moreover, the Cauchy type M-estimate may be better suited for the responses obtained from the Slash distribution than Huber’s M-estimate. Note that sign scores for the linear model estimation problem would be more suitable for the Laplace errors, however, they do not satisfy the at least two distinct score values condition that is required by IRWLS rank formulation. Recall that this formulation is inherent in the IRHET and RHETsca methods.

7.10.2 Estimates of Scale Coefficients

We now turn to the estimation of the scale coefficients beginning with results from normal distributions presented in Table 25. It is seen that IRHET and RHETsca reported average estimates that were very close to the true value of the scale coefficients. In con-
Table 25

Averages and Standard Deviations of Estimates of Scale Coefficients Based on 10,000 Simulations where $\beta_0 = 2.0$, $\beta_1 = 3.0$, $\beta_2 = 4.0$, $\theta_1 = 1.25$, $\theta_2 = 1.50$, under Standard Normal, Contaminated Normal Error Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>Normal Distribution</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$N(0.0, 1)$</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>IRHET</td>
<td>$\theta_1$</td>
<td>1.256</td>
<td>0.081</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>1.504</td>
<td>0.074</td>
<td></td>
</tr>
<tr>
<td>RHETloc</td>
<td>$\theta_1$</td>
<td>1.252</td>
<td>0.055</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>1.498</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td>ILSHET</td>
<td>$\theta_1$</td>
<td>1.093</td>
<td>0.204</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>1.367</td>
<td>0.179</td>
<td></td>
</tr>
<tr>
<td>IHMHET</td>
<td>$\theta_1$</td>
<td>1.140</td>
<td>0.167</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>1.415</td>
<td>0.137</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$CN(0.05, 100)$</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.231</td>
<td>0.146</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.484</td>
<td>0.136</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.251</td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.499</td>
<td>0.063</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.098</td>
<td>0.191</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.367</td>
<td>0.167</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.137</td>
<td>0.168</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.410</td>
<td>0.137</td>
<td></td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, LS=Least Squares, HM=Huber’s M-estimation, HET=Heteroscedastic, loc=with scale specified.

Contrast, ILSHET and IHMHET reported average estimates that were much smaller than the true values. Furthermore, from the standard deviations, it is observed that higher variability was reported by the latter pair of methods than the rank methods. Thus, as anticipated, RHETsca had the smallest response in variability to the change from $N(0, 1)$ to $CN(0.05, 100)$. The proposed method, IRHET, yielded the next smallest response.

Finally, we consider the estimates of the scale coefficients that were obtained under the non-normal distributions given in Table 26. Under the so called Slash distribution, only the average estimates obtained from RHETloc were very close to the true value which is in agreement with the results by Fligner and Killeen (1976). IHMHET estimates were not very close to the true estimates but were still better than IRHET. This result may be attributed to fact that using signed-rank Wilcoxon scores to obtain residuals needed in this step had produced values $(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})$ that were less than optimal under the Cauchy-like behaved distribution.
Under the Laplace, IRHET and RHETsca reported average estimates that were very close to the true values, while ILSHET and IHMHET, respectively, yielded averages that were much smaller than 1.25 and 1.5. With respect to variability, the smallest standard deviations were realized by the RHETloc method, and immediately followed by IRHET. ILSHET yielded the largest standard deviations.

In conclusion, the results in Tables 23-26 indicate that the proposed method IRHET is very robust to departures from normality provided the most suitable scores are used to estimate the regression coefficients. Observe that IHMHET performed very well under the Slash yielding less bias as well as the smallest standard deviations.

7.10.3 Estimated Relative Efficiency of the Iterative Methods

In what follows, we examine the results of estimated relative efficiency of the methods with respect to RHETsca and RHETloc, respectively, for the regression and scale co-
efficient estimations. The estimated relative efficiency for the regression coefficients com-
puted under both the normal and non-normal distributions is presented in Table 27. It is observed that with respect to RHETsca, the proposed method, IRHET, reported estimated relative efficiency that was more or less the same across all the four distributions. Observe that both IHMHET and ILSHET methods incurred substantial loss in efficiency when applied under non-normal distributions with the least efficiency realized under the Slash distribution.

Consider the relative efficiency of the scale constants given in Table 28. With respect to RHETloc, both IRHET and IHMHET methods yielded considerable gains in efficiency under the non-normal when compared to analogous results under the standard normal. The highest gains were observed under the Slash distribution. In all of the cases, IRHET was higher than IHMHET with respect to efficiency. ILSHET performed poorly under all the distributions which is not surprising because even when the errors were drawn from a $N(0, 1)$, the method was still very sensitive to inliers when a curtailing mechanism was not applied to the method. It is worth recalling from the inlier and outlier analysis discussed above, that the improvement in efficiency owing to the introduction of trimming as a remedial device was achieved at the cost of a large negative bias.
Table 28

Estimated Relative Efficiency of Scale Coefficients Based on 10,000 Simulations where \( \beta_0 = 2.0, \beta_1 = 3.0, \beta_2 = 4.0, \theta_1 = 1.25, \theta_2 = 1.50 \), under standard normal = \( N(0, 1) \), contaminated normal = \( CN(0.05, 100) \), Slash and Laplace Error Distributions

<table>
<thead>
<tr>
<th>Method</th>
<th>Distribution</th>
<th>( N(0.0, 1) )</th>
<th>( CN(0.05, 100) )</th>
<th>Slash</th>
<th>Laplace</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET</td>
<td></td>
<td>1.054</td>
<td>1.904</td>
<td>8.380</td>
<td>2.141</td>
</tr>
<tr>
<td>RHETloc</td>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>ILSHET</td>
<td></td>
<td>0.120</td>
<td>0.103</td>
<td>0.105</td>
<td>0.121</td>
</tr>
<tr>
<td>IHMHET</td>
<td></td>
<td>0.994</td>
<td>1.809</td>
<td>7.779</td>
<td>2.080</td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, LS=Least Squares, HM=Huber’s M-estimation, HET=Heteroscedastic, loc=with location specified.

7.11 Concluding Remarks for the Simulation Trials

Recall the results from the two sample problem. Under the 20% contaminated normal data, the proposed method, IRHET, performed well under both balanced and unbalanced designs in which the larger variance was associated with the larger sample. The method reported smaller standard deviations than the other methods including LSHET and LSHOM which were less resistant to the outliers in the data. Further, the proposed method had better coverage results than SRHOM which indicates that it was more useful to model the underlying heteroscedasticity than to treat the resulting extreme values purely as outliers. Furthermore, in the difference in location model, IRHET yielded results that were close to those obtained from the difference in trimmed means, a method that does not account for the underlying variance function. The method does not perform well when one of the samples is of size 10, due to insufficient number of observations to model the heteroscedasticity.

Recall the results from the three group problem. For the shift in location estimation problem, the proposed method is 95% as efficient as the ILSHET which is consistent with the efficiency of signed-rank Wilcoxon estimator versus the LS estimator under ho-
mosedasticity. This condition is easily satisfied when fitting the location model that has been pre-multiplied by the inverse of scale constants. Finally, this example shows that IRHET is more efficient than the lower tail trimmed LSHET. In the $CN(0.05, 100)$ results, the efficiency of ILSHET and ILSHETtr is significantly diminished while the proposed method has the same efficiency as that it had realized at $N(0, 1)$.

Consider the estimation of the scale coefficients. Note that under the $N(0, 1)$, the IRHET method was more efficient than ILSHET. Although, the results showed that ILSHETtr was the most efficient, it should noted that the method achieved this super efficiency at the cost of high negative bias caused by the lower trimming.

Under both distributions, $N(0, 1)$ and $CN(0.05, 100)$, the methods IRHET and IHMHET reported approximately equal efficiency. ILSHETtr was highly efficient but following the note above, caution should be taken when interpreting this efficiency: the high efficiency does not come without the price of bias.

Recall that the cases in which the data come from a distribution with heavy contamination and light tail contamination were also considered. The study showed that for shift in location problem, the proposed method IRHET performed poorly, in contrast to RHETsca. However, it is noted that both methods utilized signed-rank Wilcoxon scores which are not optimal for Cauchy type data. It was observed that under the Cauchy-like data, both of the ILSHET methods broke down as evidenced by inflated estimates. For the data from the lighter tailed Laplace distribution, the proposed method, utilizing signed-rank Wilcoxon score still performed better than the other estimators, albeit, signed-rank sign score would be the most optimal. Recall that the latter scores could not be employed because they do not satisfy the at least two distinct score values condition of IRWLS.

It is seen in all four distributions that the proposed method yielded smaller standard deviations for the estimates of the log ratio of the scale than all the other methods expect for RHETloc. This confirms that the good robustness qualities of the RHETloc method do carry to the iterative scheme recommended by this study. It was observed
that LS iterative methods, ILSHET and ILSHETtr reported the largest standard deviations with the latter exhibiting an additional indicator of poor performance namely, negative bias. The standard deviations of the $M$-estimates were always slightly larger than those reported by the proposed method. Thus, the estimated efficiency results showed that the IRHET methods performed well under the normal distribution as well as non-normal distributions with varying levels of outlier presence.

We now recall the inlier analysis and the lower tail trimmed modified methods. At the 5% level of inlier contamination, estimates from the 5% lower tail trimming rate had large negative bias. Using 2.5% trimming rate, the proposed method had a negligible gain in variance reduction and slight negative bias. Introducing outlier contamination, the IRHET method incurred more negative bias than when this latter type of contamination was absent. It is clear that the rank methods already curtails extreme large values induced by residuals that are very close to zero. It was also observed that there was slight decrease in the efficiency of IRHET and IHMHET while a large increase was realized by ILSHET. However, increased efficiency of ILSHETtr was acquired at the cost of large negative bias. It is clear that trimming does not improve the overall performance of the methods.

Finally, some remarks on the multiple regression model are furnished. It is seen that the ILSHET was not resistant to outliers under $CN(0.05,100)$, Slash, and Laplace. The proposed method, IRHET is quite stable even under the heavily contaminated distribution, Slash. IHMHET yielded the smallest standard deviations when the data came from the Laplace distribution. For the scale parameter estimation under $N(0,1)$ and $CN(0.05,100)$ distributions, the IRHET reported smaller standard deviations than the other methods except for RHETloc. In contrast, under the Slash distribution, the proposed method had more negative bias where as IHMHET yielded less bias. As anticipated, ILSHET performed poorly under Slash distribution due to the presence of outliers.

In contrast to the computer generated examples discussed above, the next illustra-
tion uses real life data from a clinical trial conducted among patients with depression. The example seeks to demonstrate how the proposed method can be applied in basic-new treatment type setting with a known covariate. Thus, in this regard, the example combines the methods of the two sample problem in Example 1 and a special case of the multiple regression problem in Example 4 of this chapter.
CHAPTER VIII

CASE STUDY: BEATING THE BLUES DATA

8.1 A Psychiatric Clinical Trial on the Beating the Blues Intervention

The Beating the Blues (BtB) program is an intervention that was developed as an alternative to prescribing antidepressants for the alleviating the major problem of depression among patients. The program was analyzed using Least Squares by Everitt and Wessely (2008). The objective of the study was to obtain the treatment effect of the BtB using covariate analysis based on the outcome obtained before the treatment. Since it is known that patient response to treatment depends on the duration of the current episode of depression, a subgroup analysis was sought by the primary investigators of the study. The subgroups of interest were: (1) subjects for whom the length of current episode was greater than 6 months and (2) subjects for whom the length of current episode was less than 6 months.

In the study, outcomes were observed before the treatment was administered, and the main outcome of interest was observed at 2 months after treatment had began. It was of interest to employ covariate analysis using the outcome at baseline. The authors acknowledge that there was adequate evidence that suggested that the pre and post treatment observations are correlated. When treatment types have the unequal variance, any analysis that ignores the underlying structure could produce results that would not be satisfactory with respect to variability. It is also pointed out in Everitt and Wessely (2008) that the outcomes were not normally distributed. This and the possibility that the variances of the two treatment being unequal are a source of motivation for applying the proposed rank method for heteroscedastic cases (IRHET). The method is used to obtain an estimate of the treatment effect and its corresponding 95% confidence interval. In what
follows, the study and the nature of data are briefly discussed.

Only the estimation of the effect and interval estimates are considered. Although certain testing problems can still be addressed using the confidence interval, significant tests for appropriateness of reduced models require reduction in dispersion tests, which were excluded from this analysis. To this extent, this study did not include the evaluation of improvement between fitting various models as was done in Everitt and Wessely (2008). However, the gains from incorporating heteroscedasticity in the modeling problem are evident in this illustration.

8.2 Background of the BtB study

The BtB program is an interactive approach utilizing multimedia techniques to alleviate depression among patients. A more detailed description can be found in Proudfoot et. al (2004). The patients were randomly assigned to the standard method, 'treatment as usual' (TAU), or the BtB program. The patients on the BtB regiment received every service available to those receiving TAU. In this study, out of the several outcomes considered, the authors paid attention to readings from Beck Depression Inventory II (BDI, Beck et. al., (1996)) scores. BDI scores were obtained prior to treatment, (BDI-pre), and at two months after treatment, which is referred to as BDI hereafter. In addition, BDI scores were obtained in the follow-up visits at three, five and eight months after treatments. However, the BDI scores from these follow-up visits are not analyzed, since a longitudinal study of BDI scores is not the central for the purposes of illustrating the heteroscedastic method under consideration. Moreover, the longitudinal study is beyond the scope of the problem being considered.

The baseline variable length of current episode was classified as greater than 6 months or less than 6 months. In the analysis, this variable entered as \( > 6m \) or \( < 6m \) is transformed to 0's and 1's, respectively. The treatment variable entered as TAU or BtheB (shortened to BtB) are entered as 0's or 1's depending on which program the patient was
randomly assigned to. Out of the 100 patients enrolled 3 had missing observations at the 2 months visit. We use the 97 as our sample and exclude the problem of missing data in this analysis.

8.3 Model

In the discussion that follows, we limit our attention to the subgroup analysis so that for each group, we analyze the BDI scores based on the variables: BtB treatment, BDI-pre and the interaction term between treatment and BDI-pre. Thus, in each subgroup, the model for estimating the treatment effect adjusting for the covariate, BDI-pre, is given by

\[ b_i = \beta_0 + \beta_1 t_i + \beta_2 b_{pi} + \beta_3 t_i b_{pi} + \sigma_i e_i, \quad i = 1, \ldots, n_j, \quad j = 1, 2, \]  

(8.3.1)

where the scale constants were modeled by

\[ \sigma_i = \exp\{\theta_1 t_i + \theta_2 b_{pi} + \theta_3 t_i b_{pi}\}, \]  

(8.3.2)

\( b_i \) is the \( i \)th BDI score at 2 months after treatment, \( t_i \) is the \( i \)th treatment level that is 0 for TAU and 1 for BtB, \( b_{pi} \) is the \( i \)th BDI-pretreatment score, and \( t_i b_{pi} \) is the interaction of the \( i \)th treatment and BDI-pretreatment. Note that \( n_j \) is size of the \( j \)th subgroup, \( j = 1, 2 \), such that 1 corresponds to TAU and 2 corresponds to BtB.

Outline of the Model Fitting

In the sequel, the preliminary analysis based on boxplots and scatter plots is furnished to demonstrate whether or not the choice of the model was justifiable. The boxplots of BDI scores for each subgroup were included to examine the validity of the assumption that the variances between the treatment types were equal by graphical means. The group-wise scatter plots were utilized to ascertain if an interaction effect between the
treatment and BDI-pre was suggested by the data. Intersecting lines of best fit for BDI versus BDI-pre for each treatment type would indicate that there was a strong interaction effect.

For both problems involving the treatment effect and ratio of the dispersion between the treatment groups, estimates and 95% confidence intervals were obtained using the heteroscedastic methods IRHET and ILSHET. In addition, results from the homoscedastic methods RHOM and LSHOM were also included. The latter method provided results that were comparable to those in Everitt and Wessely (2008).

To validate the model fitting process, each analysis concluded with diagnostics. Residual plots from fits from IRHET and LSHOM were analyzed. Residual plots were used to assess the validity of the model with regard to variance structure. A random pattern in scatter plot of the studentized residuals indicated the good fit. Finally, the appropriateness of the choice of the scores employed in the IRHET method was assessed using $q - q$ plots. In the quantile-quantile plots, negligible or null departures from linearity indicated good choice of the score function.

8.4 Boxplots of the BDI Scores by Length of Current Episode

We begin this section with a graphical view of the BDI scores at 2 months for each treatment level under each subgroup using boxplots. This is presented in Figure 3. It is clear from the box that the subgroup of subjects whose current episode at baseline was less than 6 months depicted larger variability amongst those that were assigned to the BtB program than those in TAU. In contrast, except for the outlying values, there was an almost negligible difference in the spread for the treatment types among subjects whose current episode at baseline was greater than 6 months.
8.5 Scatterplots of BDI vs BDI-pre by Length of Current Episode

We now turn to scatter plot of BDI scores at two months versus BDI-pre to determine whether or not there is interaction between treatment and BDI-pre levels. Overlaid plots are given in Figure 4.
Figure 4. Scatter Plot BDI Scores at 2 Months Versus BDI-pre Scores for (a) Subjects Whose Current Episode Less Than 6 Months and (b) Subjects Whose Current Episode Greater Than 6 Months at Baseline.

T:IRHET=IRHET fit on TAU recipients; B:IRHET=IRHET fit on BtB recipients; T:LSHOM=LSHOM fit on TAU recipients; B:LSHOM=LSHOM fit on BtB recipients.
To obtain the lines of best fit for each subgroup and treatment type combination, we employed IRHET to estimate the coefficients in the model

\[ b_i = \beta_0 + \beta_1 p_i + \sigma_i e_i, \quad i = 1, \ldots, n_j, \quad (8.5.1) \]

\[ \sigma_i = \exp(\theta_i b p_i). \quad (8.5.2) \]

For the LSHOM fit we let \( \sigma_i = 1 \), for \( i = 1, \ldots, n_j \).

Overlaid plots of the IRHET and LSHOM fits for the subjects whose current episode was less than 6 months at the baseline given in Figure 4 (a), which indicates that there was interaction between treatment and BDI-pre. For the subjects whose current episode was greater than 6 months at the baseline, Figure 4 (b) clearly suggested that there was interaction between treatment and BDI-pre based on the IRHET while a possible interaction is suggested by LSHOM. Thus, this preliminary analysis of the scatter plots for each subgroup strongly supported the model with interaction term in treatment and BDI-pre score, and this was the justification for using a full model which contained the interaction term.

8.6 Subjects Whose Current Episode was Less Than 6 Months

We consider the subjects whose current episode was less than 6 months at the baseline. Table 29 contains the results of the estimation of both the effects on the BDI-scores with respect to decrease or increase (shift) and variability of the BDI scores. It is seen that the iterative methods reported treatment effect that was at least 1.0 larger than the non-iterative methods did. Further, the IRHET method yielded the smallest dispersion estimate amongst all the methods indicating that it is the most efficient. Observe that for IRHET, \( \tau_{\phi} = 4.455 \) while LSHOM yielded \( \sigma = 6.085 \). All the methods reported 95% confidence interval for the BDI-pre that did not include 0, indicating that they are all in agreement that the covariate was significant. None of the methods indicated that the in-
teraction term was significant.

We now turn the estimation of the scale coefficients and the effects are presented in Table 29.

Table 29
Length of Current Episode Less Than 6 Months; Fitted Location Model Includes Constant BDI Value ($\beta_0$), and Effects: Treatment ($\beta_1$), BDI-pretreatment ($\beta_2$) and Interaction between Treatment and BDI-pretreatment ($\beta_3$)

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET</td>
<td>$\beta_0$</td>
<td>-4.440</td>
<td>3.211</td>
<td>(-10.919, 2.040)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>5.114</td>
<td>3.573</td>
<td>(-2.097, 12.326)</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>0.937</td>
<td>0.142</td>
<td>(0.650, 1.224)</td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>-0.276</td>
<td>0.179</td>
<td>(-0.637, 0.085)</td>
</tr>
<tr>
<td></td>
<td>$\tau_0$</td>
<td>4.455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RHOM</td>
<td>$\beta_0$</td>
<td>-4.881</td>
<td>3.614</td>
<td>(-12.175, 2.413)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>4.118</td>
<td>4.200</td>
<td>(-4.359, 12.595)</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>0.960</td>
<td>0.154</td>
<td>(0.649, 1.271)</td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>-0.197</td>
<td>0.178</td>
<td>(-0.557, 0.162)</td>
</tr>
<tr>
<td></td>
<td>$\tau_0$</td>
<td>5.723</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILSHET</td>
<td>$\beta_0$</td>
<td>-5.224</td>
<td>4.159</td>
<td>(-13.618, 3.170)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>5.504</td>
<td>4.424</td>
<td>(-3.424, 14.431)</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>0.960</td>
<td>0.172</td>
<td>(0.612, 1.307)</td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>-0.274</td>
<td>0.201</td>
<td>(-0.679, 0.132)</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td>7.292</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSHOM</td>
<td>$\beta_0$</td>
<td>-4.557</td>
<td>3.843</td>
<td>(-12.313, 3.200)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>4.021</td>
<td>4.466</td>
<td>(-4.992, 13.035)</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>0.930</td>
<td>0.164</td>
<td>(0.599, 1.260)</td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>-0.187</td>
<td>0.190</td>
<td>(-0.570, 0.195)</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td>6.085</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, LS=Least Squares, HET=Heteroscedastic, HOM=Homoscedastic.

in Table 30. The dispersion parameter estimate that was obtained using IRHET is almost half the that reported by ILSHET, leading to the narrower intervals that were produced by the former method. It is worth noting that neither of the methods produced intervals that suggested significant treatment effects.
Table 30

Length of Current Episode Less Than 6 Months; Fitted Scale Model Includes Effects: Treatment ($\theta_1$), BDI-pretreatment ($\theta_2$), and Interaction between Treatment and BDI-pretreatment ($\theta_3$)

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET</td>
<td>$\theta_1$</td>
<td>-0.577</td>
<td>0.584</td>
<td>(-1.757, 0.602)</td>
</tr>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>0.006</td>
<td>0.021</td>
<td>(-0.037, 0.050)</td>
</tr>
<tr>
<td></td>
<td>$\theta_3$</td>
<td>0.032</td>
<td>0.025</td>
<td>(-0.018, 0.082)</td>
</tr>
<tr>
<td></td>
<td>$\tau_0$</td>
<td>0.796</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILSHET</td>
<td>$\theta_1$</td>
<td>-1.103</td>
<td>1.129</td>
<td>(-3.380, 1.175)</td>
</tr>
<tr>
<td></td>
<td>$\theta_2$</td>
<td>-0.005</td>
<td>0.041</td>
<td>(-0.088, 0.079)</td>
</tr>
<tr>
<td></td>
<td>$\theta_3$</td>
<td>0.043</td>
<td>0.048</td>
<td>(-0.054, 0.139)</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td>1.538</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, LS=Least Squares, HET=Heteroscedastic.

8.7 Subjects Whose Current Episode was Greater Than 6 Months

In the subjects whose current episode was greater than 6 months, given in Table 31, the standard errors for the effects reported by IRHET were smaller than the rest, in general. It is seen that for the treatment effect, both RHOM and LSHOM had a negative sign while IRHET and ILSHET reported positive coefficient. This is symptomatic of heteroscedasticity in the models in which it is present. From the 95% confidence intervals from all the methods reported, a significant BDI-pre effect. Further, only IRHET showed that there was a significant interaction between Treatment and BDI-pre effects. It is also worth noting that estimated dispersion reported by the IRHET and ILSHET methods were much smaller than those computed under RHOM and LSHOM. As Everitt and Wessely (2008) pointed out, subjects whose current episode is longer than 6 month tend to have higher BDI outcomes, which provides the rationale studying these subgroups separately. This fact evidently holds true for all the methods above.

Finally, we consider the scale coefficients for the model fitted on BDI scores from
Table 31

Length of Current Episode Greater Than 6 Months; Fitted Location Model Includes Constant BDI Value ($\beta_0$), and Effects: Treatment ($\beta_1$), BDI-pretreatment ($\beta_2$), and Interaction between Treatment and BDI-pretreatment ($\beta_3$)

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRHET</td>
<td>$\beta_0$</td>
<td>4.806</td>
<td>2.597</td>
<td>(-0.419, 10.030)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>5.136</td>
<td>4.409</td>
<td>(-3.734, 14.007)</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>0.755</td>
<td>0.149</td>
<td>(0.455, 1.055)</td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>-0.570</td>
<td>0.219</td>
<td>(-1.011, -0.128)</td>
</tr>
<tr>
<td></td>
<td>$\tau_\phi$</td>
<td>2.132</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RHOM</td>
<td>$\beta_0$</td>
<td>8.431</td>
<td>4.374</td>
<td>(-0.369, 17.231)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>-1.297</td>
<td>6.053</td>
<td>(-13.474, 10.879)</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>0.575</td>
<td>0.160</td>
<td>(0.253, 0.896)</td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>-0.262</td>
<td>0.224</td>
<td>(-0.712, 0.188)</td>
</tr>
<tr>
<td></td>
<td>$\tau_\phi$</td>
<td>8.122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILSHET</td>
<td>$\beta_0$</td>
<td>7.495</td>
<td>3.446</td>
<td>(0.563, 14.428)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>2.451</td>
<td>6.415</td>
<td>(-10.454, 15.356)</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>0.592</td>
<td>0.148</td>
<td>(0.294, 0.890)</td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>-0.401</td>
<td>0.283</td>
<td>(-0.970, 0.167)</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td>4.537</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSHOM</td>
<td>$\beta_0$</td>
<td>9.075</td>
<td>5.108</td>
<td>(-1.200, 19.351)</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>-1.373</td>
<td>7.068</td>
<td>(-15.591, 12.845)</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>0.523</td>
<td>0.187</td>
<td>(0.147, 0.899)</td>
</tr>
<tr>
<td></td>
<td>$\beta_3$</td>
<td>-0.233</td>
<td>0.261</td>
<td>(-0.758, 0.292)</td>
</tr>
<tr>
<td></td>
<td>$\sigma$</td>
<td>9.484</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I=Iterative, R=Rank, LS=Least Squares, HET=Heteroscedastic, HOM=Homoscedastic

subjects for whom the duration of the current episode at baseline was greater than 6 months. The results are given in Table 32. It is seen that the variability in IRHET was almost a third of that reported by ILSHET. Further, the proposed method yielded a 95% confidence interval that indicated that the BDI-pre was significant. The estimates of $\theta_1$ and $\theta_2$ reported by IRHET were almost twice as large as those yielded by ILSHET. The two methods reported different sign for the interaction effect.
8.8 Residual Analysis by Length of Current Episode

In this section, the subgroup residual plots from the IRHET fit are discussed. Consider Figure 5 which contains residual plots for subjects whose current episode at baseline was less than 6 months. Observe that there was no specific pattern exhibited in Figure 5 (a) and Figure 5 (b) corresponding to the treatment and covariate, BDI-pre. This suggests that after fitting the IRHET the were no more symptoms of non-constant variability.

Similarly, consider Figure 6 which contains residual plots for subjects whose current episode was greater than 6 months. A random scatter is depicted in both Figure 6 (a) and Figure 6 (b).

8.9 Score Function Validity by Length of Current Episode

In this section, we are interested in assessing how well the proposed method performed by utilizing signed-rank Wilcoxon scores to estimate the regression coefficients. This is done using q-q plots given in Figure 7. In the q-q plot for the rank fit on subject
Figure 5. IRHET Fit Residual Plot of BDI for Subjects Whose Current Episode Less Than 6 Months at Baseline Against (a) BtB and (b) BDI-pre Less Than 6 Months at Baseline

whose current episode was greater than 6 month, Figure 7 (b), a very slight departure from linearity was observed. In contrast, a much more pronounced one was seen in Figure 7 (a). This suggests that the selected signed-rank Wilcoxon scores were more suitable for this group than that containing subjects whose current episode was less than 6 months at baseline.
8.10 Concluding Remarks on the BtB Study

It has been demonstrated in the analysis that the proposed iterative rank method, IRHET reported much smaller standard errors than the classic least squares LSHOM method. It is well known that when heteroscedasticity is present in data, the method fails to capture it and the resultant large standard errors lead to wider confidence intervals. Thus, wider 95% confidence intervals provided less accurate ranges of true effects on BDI-scores. Even when the LSHOM results were compared to the homoscedastic rank method SRHOM, it is evident that departures from the normality of the underlying distri-
Figure 7. q-q Plot for Validation of Score Function Selection Fitting Model 1 Using IRHET for Subjects Whose (a) Current Episode Less Than 6 Months; (b) Current Episode Greater Than 6 Months at Baseline

Distribution of the BDI scores proved costly with respect to efficiency. Note that the estimates of spread for the model under SRHOM were much lower than those reported by LSHOM in all the cases considered above.

It is worth noting that the boxplots treatment levels for the group with the duration of current episode that was less than 6 months depicted BDI scores that were extremely right skewed. In contrast, the boxplots treatment levels for the group with the duration of current episode that was greater than 6 months were much closer to symmetrical type. Since the proposed method is dependent on the symmetry of the underlying data being satisfied, the former pair of boxplots would suggest that employing the proposed method
which rightly utilizes signed-rank Wilcoxon scores was not suitable for this data. This may explain why for the subjects whose duration of current episode was greater than 6 months, nonsignificant interaction effect was obtained from the interval estimation. In contrast, for the subjects whose duration of current episode was greater than 6 months, the confidence interval estimate of the interaction term did in fact support the effects' statistical significance.

The result above demonstrates further that determining the suitable scores for the responses is such an important validation check in so far as a researcher is seeking to employ the rank method to estimate heteroscedastic model. Clearly, both scatter plots strongly supported the assertion that there was significant interaction effect between the BDI-pre and the BtB, a fact well established by expert knowledge and the preliminary plots analysis.
CHAPTER IX

CONCLUSION

9.1 Introduction

This chapter includes a summary of all the main results and conclusions of the various sections of this investigation. In addition, in light of limitations of the study as well as strong indications that have been noted, some suggestions on avenues of research for future pursuit are recommended.

9.2 Concluding Remarks

This investigation sought to model the heteroscedastic linear model,

\[ y_i = \beta_0 + x_i^T \beta_1 + \sigma_i e_i, \quad i = 1, \ldots, N, \]

(9.2.1)

where \( y_1, \ldots, y_N \) are responses, \( \beta_0 \in \mathbb{R} \) and \( \beta_1 \in \mathbb{R}^p \) are unknown regression parameters, \( x_i^T \) is the \( i^{th} \) row of the \( N \times p \) centered matrix \( X \). The variables \( e_1, \ldots, e_N \) are random errors assumed to be independently and identically distributed (iid) with a common cdf \( F \). Here, \( \sigma_1, \ldots, \sigma_N \), are scale constants that express heteroscedasticity through the relationship

\[ \sigma_i = \exp(x_i^T \theta), \quad i = 1, \ldots, N, \]

(9.2.2)

where \( \theta \) is a \( p \times 1 \) vector of unknown scale parameters.

In this study, the logarithmic transformation was employed to obtain a linearized variance function model. The main objective of the investigation was to develop unified
asymptotic theory for the rank estimators of $\beta$ and $\theta$.

In chapter two, we developed the theory for the general linear model where at least 2 independent variables were considered.

In chapter three, the estimator for the location or regression coefficient $\beta$ was demonstrated to follow an asymptotically normal distribution, with mean $\beta$ and covariance $\gamma_1^{-2}(X_1^T X_1)^{-1}$. Consequently, this result proved that the estimator is consistent and efficient under any fixed value of the scale coefficient. Further, it was demonstrated that when $\theta$ is replaced by a robust estimate, the minimization of the dispersion function for the problem of estimating $\beta$ yields an estimate of the desired parameter that is robust and efficient.

In chapter four, the estimator for the scale coefficient, $\theta$, was shown to have limiting distribution which was normal, with mean $\theta$ and covariance $\gamma_2^{-2}(X^T X)^{-1}$. This proved that the proposed estimator was consistent and efficient under any fixed regression coefficient, as would be anticipated with scale estimation problem. Further, it was shown that when estimating $\theta$ with $\beta$ replaced by its robust estimate, the minimization of the dispersion function for this problem yielded an estimate that was both robust and efficient.

In chapter five, a brief review of the score functions that were under consideration in the study was given. In addition, the specification of the scale parameters $\gamma_1$ and $\gamma_2$ that were based on the method by Koul, Sievers, and McKean (1987) was presented. Following the strong results from chapters three and four, the IRWLS formulations of dispersion functions for each problem were provided in the spirit of Sievers and Abebe (2004). Further, due to the last theoretical result, the formulations of the dispersion functions that updated the estimates in an iterative fashion were specified. Finally, algorithm for the iterative estimation of $\beta$ and $\theta$ was furnished.

In chapters six, seven and eight, the analyzes of simulation experiments and real life data were presented. The implementation of the IRHET method in two sample set-
ting is presented. To compare the proposed IRHET to several non-robust as well as robust methods in a two sample problem, a description of the all of the other methods including M-estimates considered in the experiment was also provided.

In chapter seven, we presented the analysis of the results from the simulation trials performed in this study. This experiment yielded results that supported the good robustness and efficiency of IRHET that were given by the theoretical development above. The a comparative analysis of coverage of the interval estimates from the various methods was analyzed directly. Overall, the IRHET performed well under the moderate contaminated normal responses.

In a simple linear model setting, the proposed methods proved to be non-resistant to effects of inliers that evidently adversely affected the performance of an LS type analog for the heteroscedastic case, in terms of efficiency. It was seen that the lower-tail trim modification was not well suited for the current ranks method as the modification introduces large negative bias, efficiency of the method in contrast to its LS analog, and standard LS and signed-rank estimator for the homoscedastic cases, were further studied in the multiple regression model. In summary, the IRHET method yielded results that were superior to those reported by the other methods, with the exception of the benchmark methods RHETsca and RHETloc, in some cases.

In chapter eight, is application of the method to data from Psychiatric Clinical Trial, ”Beating the Blues”, a computerized behavioral therapy conducted by Proudfoot, et. al, (2004). It was seen that, provided the underlying distribution of a data was symmetric or approximately symmetric, IRHET revealed more meaningful differences due to its high efficiency in the subgroup analysis than that obtained by LS.

9.3 Future Research

In this section, a discussion of the problems that were not considered in this study is presented. This includes applications that would extend the usefulness to many levels
of the decision making whose success wholly depends on how robust the method is. In this regard, knowing how much contamination the methods can withstand before breaking down will be useful. The most important next step is the inclusion of the testing problem and extension to a method that curtails outlying values in the design space.

(1) Extend to Testing Problem Using Reduction in Dispersion

When there are more than 2 independent variables, only a few of them may best explain the variability in the responses. Since the solution to both location and scale estimation was the minimization of the dispersion function, fitting the full model and reduced model leads to the reduction in dispersion test of Hettmansperger and McKean (1998), naturally. In the IRWLS approach, this has already been employed by Sievers and Abebe (2004) in their treatise on rank estimation of homoscedastic linear models. Thus, an extension to the current problem is appealing so as to furnish subsequent inference by increasing our knowledge on the underlying heteroscedasticity in the general linear model.

(2) Extend to the Case of Outlying Values in Design

So far, the method employed in this study has only addressed the question of how to curtail the outlying values that occurred in the responses. However, it is clear that when the variance function depends on the design, outlying values in the design space can affect the stability and usefulness of the fitted heteroscedasticity. A considerable amount of research, including bounded influence GR estimation by Naranjo and Hettmansperger (1994), has been done since weighted rank estimates were introduced by Sievers (1983). In their work, the authors recommended estimators that were resistant to outlying values on both y-space and design space. Then, the next step would be to extend by first developing the influence function and sec-
ondly, extending the GR estimates to the current heteroscedastic linear problem.

(3) Explore Methods for Increasing Coverage of The Scale Coefficient Estimators

It was seen in the constructed interval estimates for the scale coefficients that the simulation experiment indicated that coverage values that were slightly lower than the nominal levels were reported by the proposed method. In particular, modifying the IRWLS or scores so that responses from Cauchy and Laplace distribution can be more efficiently modeled by the iterative rank procedure proposed in this study. Consequently, such modifications would lead to increased coverage.

(4) Explore How Sensitive the Method is to Mis-specification of the Variance Function

The method performs well when the responses indeed contained the form of heteroscedasticity that was assumed in the scaling constants. It would be of interest to investigate how well the method performs when underlying function is different from that which is currently prescribed by constants. Subsequently, a comparison with the performance of the estimators that do not assume that the form of variance function is known can give more insights on the proposed method.

(5) Develop More Diagnostic Tools

It will be of interest to explore how the method performs under mis-specification of the nature of relationships between the responses and their variability, respectively, and the design. This will also utilize the existing residual plots and partial residual plots.
After developing the components of the heteroscedastic linear model outlined above, the final step will be to package these tools in R so that it is easily accessible for most queries for making inferences on the parameters of the general linear model.
REFERENCES


