Research Report

Strategies for Testing Slope Differences

Tim Moses and Alan Klockars

October 2009

ETS RR-09-32

Strategies for Testing Slope Differences

Tim Moses
ETS, Princeton, New Jersey

Alan Klockars
University of Washington, Seattle

October 2009

Abstract

The robustness and power of 9 strategies for testing the differences in groups' regression slopes were assessed under nonnormality and residual variance heterogeneity. For the conditions considered, the most robust strategies were the trimmed and Winsorized slope estimates used with the James second-order test, the Theil-Sen slope estimates used with James, and the Theil-Sen estimates used with percentile bootstrapping. The use of Theil-Sen slope estimates produced more powerful tests than the use of trimmed and Winsorized slopes.

Key words: Slopes, least squares, Theil-Sen, robust regression, James second-order, nonnormality, residual variance heterogeneity

Acknowledgments

The authors thank Frank Rijmen, Russell Almond, and Dan Eignor for helpful suggestions and Kim Fryer for editorial work.

Table of Contents

List of Tables

List of Figures

Page

Introduction

The issue of group differences across an individual difference variable (*X*) has been considered in many fields of social science (Aguinis & Pierce, 1998; Cronbach & Snow, 1977; Dance & Neufeld, 1988; Hunter, Schmidt, & Hunter, 1979; Seligman, 2002). The typical approach of strategies evaluating group differences across *X* involves fitting regression lines that predict outcome *Y* from *X* in separate treatment groups and then conducting a significance test for the homogeneity of the groups' regression slopes. In this study, strategies for comparing slopes are evaluated under conditions of nonnormality and residual variance heterogeneity.

The slope test strategies considered in this study are approaches to estimating the following model,

$$\boldsymbol{Y_{ij}} = \boldsymbol{\beta_{0j}} + \boldsymbol{\beta_{1j}} \boldsymbol{X_{ij}} + \varepsilon_{ij}, \quad (1)$$

where outcome *Y* for individual i (= 1 to *N*) in group *j* (= 1 to *J*) is a linear function of a continuous *X*, β_{0j} and β_{1j} are the population intercepts and slopes of the regression line for each of *J* groups, and the ε_{ij} are the residuals. The strategies for assessing differences in the β_{1j}'s reviewed below are most easily understood in terms of alternative expressions of (1). When *J* = 2, (1) can be expressed as,

$$\mathbf{Y_{ij}} = \beta_0 + \beta_1 \mathbf{X_{ij}} + \beta_2 \mathbf{G_{ij}} + \beta_3 \mathbf{X_{ij}} \mathbf{G_{ij}} + \varepsilon_{ij}, \quad (2)$$

where G_{ij} is a dichotomously-coded group membership variable. A more general matrix version of (1) and (2) is,

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}, \quad (3)$$

where $\mathbf{Y}$ is an *N*-by-1 column vector, $\mathbf{X}$ is an *N*-by-*K* *design matrix* corresponding to the *K* β's (including a column of 1's for estimating β_0), β is a *K*-by-1 column vector of β's, and *e* is an *N*-by-1 column vector of residuals.

Standard Slope Estimation and Slope Test

This section presents the standard least squares slope test and its technical details. The technical details provide a basis for understanding the alternative and robust versions presented in the later sections. The standard slope test uses least squares estimates of the β 's (i.e., $\hat{\beta}$'s) that minimize the sum of the squared residuals, $\hat{\mathbf{e}}^t\hat{\mathbf{e}} = \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)^t \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$. Because $\hat{\mathbf{e}}^t\hat{\mathbf{e}}$ is a convex function of $\hat{\boldsymbol{\beta}}$, it can be minimized by differentiating with respect to $\hat{\boldsymbol{\beta}}$, setting this derivative to zero, and solving for $\hat{\boldsymbol{\beta}}$, resulting in the closed form solution,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} \tag{4}$$

Equivalently, group *j*'s slope can be estimated as, $\hat{\beta}_{1j} = \dfrac{\sum_{\text{i in j}}(x_{ij} - \bar{x}_{.j})(y_{ij} - \bar{y}_{.j})}{\sum_{\text{i in j}}(x_{ij} - \bar{x}_{.j})^2}$, (5)

where $\bar{x}_{.j}$ and $\bar{y}_{.j}$ are the means of *X* and *Y* in group *j*.

The standard test for assessing the differences of *J* slopes is an *F* test,

$$F_{\text{SlopesStandard}} = \frac{\left(\dfrac{1}{J-1}\right)\sum_j\left(\left(\hat{\beta}_{1j}^2 - \hat{\beta}_{1.\text{Standard}}^2\right)\sum_{\text{i in j}}(x_{ij} - \bar{x}_{.j})^2\right)}{\left(\dfrac{1}{N-2J}\right)\sum_j(N_j - 2)\hat{\sigma}_{ej}^2}, \tag{6}$$

where $\hat{\sigma}_{ej}^2 = \dfrac{\sum_{\text{i in j}}\left(\hat{\varepsilon}_{ij}\right)^2}{N_j - 2}$ and $\hat{\beta}_{1.\text{Standard}} = \dfrac{\sum_j \hat{\beta}_{1j}\sum_{\text{i in j}}(x_{ij} - \bar{x}_{.j})^2}{\sum_j\sum_{\text{i in j}}(x_{ij} - \bar{x}_{.j})^2}$ is the variance-weighted common slope (Myers & Well, 1995, p. 421-422). (6) is evaluated on an *F* distribution with *J*-1 and *N*-2*J* degrees of freedom. With *J* = 2, a t-test of $\hat{\beta}_3$ in (2) that is equivalent to the *F* test in (6) can be conducted by obtaining the standard error of $\hat{\beta}_3$ as the square root of one of the diagonal elements in the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$, $\dfrac{\hat{\mathbf{e}}^t\hat{\mathbf{e}}}{N-K}(\mathbf{X}^t\mathbf{X})^{-1}$, and evaluating $\dfrac{\hat{\beta}_3}{SE(\hat{\beta}_3)}$ on a *t* distribution with $N - K = N - 4$ degrees of freedom. The referencing of the standard test statistics to *F* and *t* distributions is justified when the data meet particular assumptions, namely that the ε_{ij} are normally and independently distributed with equal variances across the *J* groups.

The standard methods for estimating and testing slopes are problematic when data are nonnormal and residual variances are heterogeneous (Alexander & Deshon, 1994; Conerly & Mansfield, 1988; Conover & Iman, 1982; Deshon & Alexander, 1996; Dretzke, Levin & Serlin, 1982; Headrick & Sawilowsky, 2000; Klockars & Moses, 2002; Overton, 2001). When distributions exhibit heavy-tailed nonnormality, extreme scores occur more often than when distributions are normal, increasing the variability of the estimated slopes, reducing the estimated standard errors, and making the standard test excessively liberal. When groups' residual variances and sample sizes differ, the standard test's pooling of groups' residual variances, $\left(\frac{1}{N-2J}\right)\sum_{j}(N_j-2)\hat{\sigma}_{ej}^2$, is problematic, making the standard slope test either liberal or conservative, depending on which of the larger and smaller groups has the larger or smaller residual variance. The inaccuracy of the standard test is disturbing, given that nonnormality and residual variance heterogeneity appear to be common in actual data (Aguinis, Peterson, & Pierce, 1999; Micceri, 1989). What follows are detailed definitions of slope test strategies that may outperform the standard test when distributions are nonnormal and residual variances are heterogeneous.

Slope Tests for Nonnormal Data: Central Tendency Strategies

Two approaches to slope estimation view group *j*'s slope in (5), $\hat{\beta}_{1j}=\frac{\sum_{i\ in\ j}(x_{ij}-\bar{x}_{.j})(y_{ij}-\bar{y}_{.j})}{\sum_{i\ in\ j}(x_{ij}-\bar{x}_{.j})^2}$, as a central value of the slopes that can be created from pairs of observations in the data, $b_{1,ij,i'j}=\frac{(y_{ij}-y_{i'j})}{(x_{ij}-x_{i'j})}, i\neq i', x_{ij}\neq x_{i'j}$, and then try reduce the influence of the extreme observations on the central value. These *central tendency* approaches define extreme observations in terms of both *X* and *Y*, so that the screening of extreme observations caused by nonnormality could potentially address slope estimation problems such as leverage (observations that are extreme on *X*), discrepancy (observations that are extreme with respect to the regression line), and outliers on *Y*. One popular strategy is the Theil-Sen slope estimator (Ebrahem & Al-Nasser, 2005; Sen, 1968; Theil, 1950; Wang, 2005; Wilcox, 2004; Wilcox & Keselman, 2004). The Theil-Sen estimate is the median of the slopes that can be computed from the $N_j(N_j-1)/2$ pairs of observations in the data. Percentile bootstrapping methods can be used to test for differences between groups' Theil-Sen slopes (i.e., draw 599 random samples with replacement from the $J = 2$ datasets, compute the differences in Theil-Sen slopes in each of these datasets, and determine if the middle (1-α)% of the 599 slope differences contain zero; Wilcox, 2005).

A less-familiar alternative to the Theil-Sen slope estimate is the application of the trimming and Winsorizing strategies that are typically proposed in tests of mean differences to $\frac{(y_{ij} - y_{i'j})}{(x_{ij} - x_{i'j})}$ (Guo, 1996; Luh & Guo, 2000). To obtain trimmed and Winsorized estimates of slopes and their variances, rank order the x's in each of the J groups, $x_{1j} < x_{2j} \ldots < x_{Nj}$. When the number of observations in group j is even ($N_j = 2m_j$), one can consider m_j independent slope estimates,

$$b_{1,i+m_j j,ij} = \frac{(y_{i+m_j j} - y_{ij})}{(x_{i+m_j j} - x_{ij})}. \quad (7)$$

When the number of observations in j is odd ($N_j = 2m_j + 1$), a pooling is done so that observations $y_{2m_j j}$ and y_{2m_j+1j} are pooled, $x_{2m_j j}$ and x_{2m_j+1j} are pooled, and $y_{2m_j j}$ and $x_{2m_j j}$ are replaced by $y_{2m_j j} = (y_{2m_j j} + y_{2m_j+1j})/2$ and $x_{2m_j j} = (x_{2m_j j} + x_{2m_j+1j})/2$.

The trimming and Winsorizing are done for each of the j slopes and standard errors. Let $g_j = \gamma m_j$ where γ represents the proportion of observations to be trimmed from each tail of the ordered distribution of $b_{1,lj}, l = 1 \text{ to } m_j, b_{1,lj} \le b_{1,2j} \le \ldots b_{1,m_j j}$. Let $h_j = m_j - 2g_j$ be the effective sample size after trimming.

The trimmed mean slope is computed as $\bar{b}_{1.j} = \frac{\sum_{l=g_j+1}^{m_j-g_j} b_{1,lj}}{h_j}$. (8)

Winsorized slope observations are obtained by,

$$bw_{1,lj} = \left\{\begin{array}{lll} b_{1,(g_j+1)j} & \text{if } b_{1,lj} \le b_{1,(g_j+1)j} & \\ b_{1,lj} & \text{if } b_{1,(g_j+1)j} < b_{1,lj} & < b_{1,(m_j-g_j)j} \\ b_{1,(m_j-g_j)j} & \text{if } b_{1,lj} \ge b_{1,(m_j-g_j)j} & \end{array}\right\}. \quad (9)$$

The variance of the trimmed mean slope is computed as a function of the Winsorized variance,

$$\sigma^2_{bwj} = \frac{1}{h_j(h_j - 1)} \sum_{l=1}^{m_j} (bw_{1,lj} - \frac{\sum_l bw_{1,lj}}{m_j})^2 . \tag{10}$$

To assess the differences in trimmed slopes, replace the $\hat{\beta}_{1j}$ in (6) with $\bar{b}_{1.j}$, the $\left(\frac{1}{N-2J}\right)\sum_j (N_j - 2)\hat{\sigma}^2_{ej}$ with $\left(\frac{1}{\sum_j h_j - J}\right)\sum_j h_j(h_j - 1)\sigma^2_{bwj}$, and the $\sum_{i \text{ in } j}(X_{ij} - \bar{X}_{.j})^2$ with h_j. These replacements to (6) cause the standard test of slope differences to resolve into the *F* test for independent trimmed means with *J*-1 and $\sum_j h_j - J$ degrees of freedom,

$$F_{SlopesTrimmed} = \frac{\left(\frac{1}{J-1}\right)\sum_j h_j \left(\bar{b}^2_{1.j} - \left(\frac{\sum_j h_j \bar{b}_{1.j}}{\sum_j h_j}\right)^2\right)}{\left(\frac{1}{\sum_j h_j - J}\right)\sum_j h_j(h_j - 1)\sigma^2_{bwj}} . \tag{11}$$

Slope Tests for Nonnormal Data: Minimum Maximum Likelihood Type (MM) Regression

In minimum maximum likelihood type (MM) regression (Yohai, 1987), extreme observations are addressed in the minimization process used to estimate the regression line. While the standard slope estimation process is based on minimizing the sum of all squared residuals, the robust regression paradigm views the least squares approach as one of several possible functions, ξ, of the scaled residuals that could be minimized,

$$\sum_j^J \sum_{i \text{ in } j}^{N_j} \xi\left(\frac{y_{ij} - \beta_{0j}x_{ij1} - x_{ij2}\beta_{1j}}{\sigma}\right) = \sum_j^J \sum_{i \text{ in } j}^{N_j} \xi\left(\frac{\varepsilon_{ij}}{\sigma}\right). \tag{12}$$

Some choices of ξ can produce β estimates that outperform the standard test's β's in terms of their breakdown rates (i.e., the smallest percentage of contaminated observations needed to render $\hat{\beta}$ useless). One popular ξ (SAS Institute, 2003) is the Tukey weight function,

$$\xi(s) = \begin{cases} 3\left(\frac{s}{\kappa}\right)^2 - 3\left(\frac{s}{\kappa}\right)^4 + \left(\frac{s}{\kappa}\right)^6 & \text{if } |s| \le \kappa, \\ 1 & \text{otherwise.} \end{cases} \quad (13)$$

In (13), κ is a constant selected to obtain desirable properties. A κ value of 3.44 results in parameter estimates that are 85% as efficient as least squares estimates when the data are normal (Holland & Welsh, 1977). When data contain outliers that are discrepant with respect to the regression line, κ defines a range around which the observations outside of the range have reduced contribution to the slope estimates.

The search for β's that minimize (12) is similar to the standard test, in that β_k's are found such that the derivatives of (12) with respect to the β_k's are zero, $\sum_{j}^{J}\sum_{i\,in\,j}^{N_j}\frac{\partial\xi}{\partial s}\left(s_{ijk}\right)x_{ijk} = 0, k = 1 \text{ to } K$. Unlike the least squares estimation used with the standard test, with MM regression there are no closed-form solutions to minimizing (12). Appendix A presents an outline of the three-stage MM algorithm for estimating the β_k's.

The kinds of nonnormality for which MM might be especially useful are probably situations with outliers that do not mask themselves by exerting heavy influence on the regression line. Many of the steps of the MM estimation process are analogues to the standard test's estimation, including the use of least squares estimation used in the least trimmed squares (LTS) starting values, the computation of the β_k's (Equation A4 in Appendix A is a weighted version of Equation 4), and the computation of the MM standard errors ($\mathbf{W}$ in Equation A5 is a weighted version of $(\mathbf{X}^t\mathbf{X})$ in $\frac{\hat{\mathbf{e}}^t\hat{\mathbf{e}}}{N-K}(\mathbf{X}^t\mathbf{X})^{-1}$). The relatedness of MM computations to the standard test's computations suggest that both procedures would do well with normal populations, while MM should outperform the standard test when there are outliers on *Y* (Anderson & Schumacker, 2003).

Slope Tests for Heterogeneous Residual Variances

Alternative parametric significance tests have been developed by Welch (1938), James (1951), and Deshon and Alexander (1994) to test for slope differences when residual variances are unequal. All three tests avoid the standard test's pooling of groups' residual variances in (6). Comparative research has shown that the three parametric alternative tests perform similarly in terms of robustness and power (Deshon & Alexander, 1996; Luh & Guo, 2000, 2002), so this study focuses solely on the James second-order test, which is slightly better than the Welch and Deshon and Alexander tests in terms of power and robustness to nonnormality.

The steps of the James second-order test are as follows:

1. Define a James weight, wj_j, based on each group slope's standard error,

$$\mathrm{wj}_\mathrm{j} = \frac{1/\sigma^2_{\beta_{1\mathrm{j}}}}{\sum_\mathrm{j} 1/\sigma^2_{\beta_{1\mathrm{j}}}}. \qquad (14)$$

2. Define a variance-weighted common slope as,

$$\beta^+ = \sum_\mathrm{j} \mathrm{wj}_\mathrm{j}\beta_\mathrm{j}. \qquad (15)$$

3. Define the James' test statistic as,

$$\mathrm{James} = \sum_\mathrm{j} \frac{\left(\beta_\mathrm{j} - \beta^+\right)^2}{\sigma^2_{\beta_{1\mathrm{j}}}}. \qquad (16)$$

4. Evaluate the significance of the James' test statistic by determining if it exceeds a critical value that is based on a mixture of chi-square distributions (Appendix B).

Hybrid Slope Tests for Nonnormal Data and Heterogeneous Residual Variances

Slope test strategies are not necessarily robust to problems for which they were not directly designed. The parametric alternative strategies that were designed to address residual variance heterogeneity have documented problems with nonnormal data (Deshon & Alexander, 1996). The slope tests that have been proposed for nonnormal data do not directly address residual variance heterogeneity. An important area of research assesses so-called *hybrid slope tests* that may be robust to several assumption violations by use of nonnormality-robust group slopes and standard errors with parametric alternative tests that avoid the pooling of heterogeneous residual variances.

Recent research on hybrid slope tests has considered using standard slope estimates and standard errors or trimmed slope estimates and Winsorized standard errors with skew-corrected versions of parametric alternative tests (Luh & Guo, 2000, 2002). The use of the trimmed slopes and Winsorized standard errors with parametric alternative tests like the James test is straightforward, with groups' degrees of freedom calculated as $v_j = h_j - 1$ rather than as $N_j - 2$. Luh and Guo also transformed the test statistics of the parametric alternatives to eliminate the effect of skewness (Hall, 1992; Johnson, 1978). For example, the proposed transformation for skewness for the James second-order test statistic from (16) is,

$$\text{James\_TT} = \sum_j \left(\sqrt{N_j} [\frac{(\beta_j - \beta^+)}{\sqrt{N_j}\sigma_{\beta_{1j}}} - \gamma^3_{x,j}\gamma^3_{\varepsilon,j} \frac{(\beta_j - \beta^+)^2}{N_j \sigma^2_{\beta_{1j}}} / 6 + \gamma^3_{x,j}\gamma^3_{\varepsilon,j} / (6N_j)] \right)^2, \quad (17)$$

where $\gamma^3_{x,j}$ and $\gamma^3_{\varepsilon,j}$ are the sample skews of *X* and ε in group *j*. Luh and Guo's studies showed that their hybrid strategies were robust to both nonnormality and residual variance heterogeneity.

This Study

This study directly compares the standard, MM, and Theil-Sen tests, extending the previous comparisons based on estimating one slope that have given recommendations for MM regression over the standard test (Anderson & Schumaker, 2003) and for Theil-Sen over MM regression and the standard test (Wilcox & Keselman, 2004). The comparison of the trimmed and Winsorized slope test with the Theil-Sen and MM tests has not been considered in previous studies, and it allows for an evaluation of some trimming (trimmed and Winsorized) with the most extreme trimming possible (Theil-Sen).

This study also extends Luh and Guo's (2000, 2002) work, first by separately evaluating the trimmed and Winsorized slope test and the skewness transformation of the James test statistic. Because the accuracy of slope estimation has more to do with the heaviness of the distribution's tails rather than its skew (Klockars & Moses, 2002), the test statistic transformation ought to have a smaller impact in correcting for nonnormality than the trimmed and Winsorized, Theil-Sen, and MM tests. Finally, Luh and Guo's efforts to form hybrid slope test strategies that are robust to both nonnormality and residual variance heterogeneity are extended to consider hybrid slope tests based not only on integrating the trimmed and

Winsorized tests and the skewness transformation with James second-order test, but also the MM and Theil-Sen tests.

Method

A simulation study was conducted to investigate the relative robustness and power of nine slope test strategies for comparing two groups' slopes. The robustness of the strategies' Type I error rates was assessed across 64 conditions where the population slope difference between the groups was zero. The 64 conditions featured a crossing of eight degrees of nonnormality in *Y* with four group sample size combinations and two degrees of residual variance heterogeneity. For each of the 64 conditions, the slope test strategies' rejection rates for the null hypothesis of population slope differences of zero were computed by conducting their significance tests in 10,000 randomly generated datasets. Strategies' power rates were assessed across the four group sample size combinations by computing their rejection rates in 10,000 random datasets generated with nonzero population slope differences, normal distributions, and equal residual variances.

Nine Slope Test Strategies

Five stand-alone slope test strategies and four hybrids of the five strategies were evaluated:

1. Standard: The standard F-test of slope differences in (6).
2. James: The James parametric alternative test in (16).
3. MM: Significance testing of the β_3 in (2) based on MM estimation with the default settings in SAS PROC ROBUSTREG (SAS Institute, 2003).
4. TW: The trimmed and Winsorized slope test in (11) using 10% trimming.
5. TS: The Theil-Sen estimator with percentile bootstrapping for the significance testing.

The following four hybrid strategies were also considered:

6. James-TT: The James procedure with the Johnson's one-sample *t*-statistic transformation for skewness in (17).
7. James-MM: The James procedure using MM slope estimates and standard errors.

8. James-TW: The James procedure using 10% trimmed slope estimates and Winsorized standard errors from Luh and Guo (2000).
9. James-TS: The James procedure using the Theil-Sen slope estimates and the standard deviations of 599 bootstrapped Theil-Sen estimates from Strategy 5 for the group slopes' standard errors.

Y's Distribution

Y's distribution was manipulated to produce eight shapes, including a normal shape (skew = 0, kurtosis = 0), and seven other shapes with various degrees of skews and kurtosis (Table 1). *Y*'s distribution was directly manipulated because this produced situations with high leverage points, that is, situations that affected the slope estimation strategies to a greater extent than other forms of nonnormality (i.e., in ε and/or *X*). The influence of these particular forms of nonnormality on the slope test strategies as compared to other forms of nonnormality is addressed in the discussion.

Table 1

Y's Generated Shapes and the Fleishman Constants Used to Generate Them

Y's skew	*Y*'s kurtosis	*A*	*b*	*c* (= -*a*)	*d*
0	-1.15	0	1.34	0	-0.132
0	0	0	1	0	0
1.2	1.11	-0.340774	1.095718	0.340774	-0.080735
1.6	2.86	-0.418206	0.975506	0.418206	-0.06113
0	3	0	0.782	0	0.068
0	6	0	0.66269	0	0.101888
0	9	0	0.573	0	0.126
0	12	0	0.498	0	0.145

Variance Heterogeneity

The two considered residual variance ratios for the groups were 1/1 and 3/1. For conditions of unequal sample size, the residual variances were directly and inversely paired with the treatment group sample sizes.

Sample Sizes

Treatment groups of 20 and 40 participants were used. The conditions of unequal sample size used 20 participants in one group and 40 in the other.

Data Generation Method: Robustness

The following data generation method was used to create *X* and *Y* variables of desired distributions and variances with equal slopes in the two groups.

1. *N* values of one standard normal variate, *Z*, were generated, where *N* is the total sample size in two groups.
2. *Y* was created as a transformation of *Z* using Fleishman's (1978) method for generating nonnormal variables:

$$Y = a + bZ + cZ^2 + dZ^3 \tag{18}$$

The constants (*a, b, c,* and *d*) and resulting distributions are listed in Table 1.

3. An error variable for *X* (ε) was generated as a standard normal variate. *X*'s degree of nonnormality was a compromise between *Y*'s nonnormality and ε's normality.
4. Desired numbers of Ys and εs were randomly assigned to treatment Groups 1 and 2.
5. *X* was created as a function of *Y* and ε:

$$X_{ij} = \rho_j Y_{ij} + \sqrt{(1 - \rho_j^2)}\varepsilon_{ij}, \tag{19}$$

where ρ_j is the desired *XY* correlation for treatment group *j*.

6. Y_{ij} was multiplied by a number, σ_{Yj}, that resulted in a desired standard deviation for *Y* in the *j*th treatment group and, in conjunction with ρ_j, a desired residual variance. The

values of σ_{Y_j} and ρ_j for the two groups achieved a particular residual variance ratio (Table 2), while keeping the slopes equal in the two groups.

Table 2

Correlations and Standard Deviations Used to Create Levels of Residual Variance Heterogeneity

Residual variance ratio	ρ_1	σ_{Y1}	ρ_2	σ_{Y2}
1/1	0.5	1.0	0.5	1.0
1/3	0.5	1.0	0.3162	1.5811

Data Generation Method: Power

The data generation process used to assess strategies' power was similar to the data generation process used to assess robustness. All variables' distributions were normal. One group's *XY* correlation and *Y* standard deviation were 0.5 and 1.0, respectively, while the second group's *XY* correlation and *Y* standard deviation were 0.0 and 0.866, respectively. The *XY* correlations and *Y* standard deviations across the groups resulted in a population slope difference of 0.5 while meeting the normality and equal residual variances assumptions of the standard test. The population slope difference of 0.5 was selected because when assessed with the four sample size combinations considered in this study, strategies' power rates were somewhat closer to 0.5 than they were to 0 or 1. Situations where strategies' power rates were not extremely high or low produced results that were useful for maximally differentiating the strategies.

Analysis Strategy

The assessment of strategies' robustness involved comparing their average rejection rates to the nominal 0.05 rate for conditions where no slope differences existed in the population. Deviations from the nominal 0.05 rate were determined to be excessively conservative or liberal when they were outside of the two-standard-error band reflective of the number of replications used in this study ($0.05 +/- 2\sqrt{\frac{(.05)(.95)}{10,000}} = 0.046 \text{ to } 0.054$). The standard error band roughly corresponded to Bradley's (1978) conservative range for robust Type I error rates, 0.045 to

0.055. The assessment of strategies' power involved comparing strategies' average rejection rates to each other for conditions where actual slope differences existed in the population.

Follow-up analyses were also conducted to gain further insight into how the slope estimation strategies were working in the conditions of this study. These follow-up analyses included assessments of averages and standard deviations of the strategies' slope estimates to indicate their bias and efficiency, and assessments of strategies' average standard errors to provide understanding of the accuracy of strategies' significance tests.

Results

Type I Error

Tables 3-9 present the considered strategies' empirical Type I error rates across 56 of the 64 combinations of nonnormality, residual variance heterogeneity, and sample size. The combination of group sample sizes of 40 and 20 and a residual variance ratio of 1/1 produced nearly identical results with the combination of group sample sizes of 20 and 40 and a residual variance ratio of 1/1, so the results with sample sizes of 40 and 20 are omitted. Nonnormality affected the Standard, James and MM tests similarly, creating liberal Type I error rates when the *Y* distributions were leptokurtic and conservative Type I error rates when the distributions were platykurtic. The TW test had Type I error rates that were close to the nominal rate across the conditions of nonnormality. The TS test had Type I error rates that were consistently conservative across the considered levels of nonnormality. In terms of the hybrid strategies, James-TT had Type I error rates that were almost indistinguishable from James, while the James-MM, James-TW and James-TS tests had Type I error rates reflective of the nonnormality strategy used, being excessively liberal for James-MM, near 0.05 for James-TW, and excessively conservative for James-TS.

The effect of residual variance heterogeneity on Type I error differed for the equal and unequal sample size conditions. When sample sizes were equal (Tables 4 and 9), MM was the only strategy affected by residual variance heterogeneity, becoming excessively liberal. When sample sizes were unequal (Tables 6 and 7), the groups' sample size-residual variance pairing affected the Standard, MM and TW tests similarly, making them liberal with an inverse pairing and conservative with a direct pairing. The hybrid strategies were largely unaffected by the combination of unequal sample sizes and residual variances. James-TS produced conservative Type I error rates for most of the considered residual variance conditions.

Table 3

Empirical Type I Error Rates for Group Sample Sizes of 20 and 20 and a Residual Variance Ratio of 1/1

							Hybrid strategies			
Skew	Kurtosis	Standard	James	MM	TW	TS	James-TT	James-MM	James-TW	James-TS
0	-1.15	0.0256[a]	0.0257[a]	0.0260[a]	0.0479	0.0207[a]	0.0258[a]	0.0214[a]	0.0456	0.0204[a]
0	0	0.0486	0.0481	0.0610[a]	0.0497	0.0282[a]	0.0483	0.0486	0.0479	0.0262[a]
1.2	1.11	0.0668[a]	0.0676[a]	0.1099[a]	0.0508	0.0304[a]	0.0667[a]	0.0985[a]	0.0490	0.0358[a]
1.6	2.86	0.0941[a]	0.0961[a]	0.1383[a]	0.046[a]	0.0302[a]	0.0949[a]	0.1341[a]	0.0455[a]	0.0345[a]
0	3	0.0912[a]	0.0936[a]	0.0999[a]	0.0532	0.0307[a]	0.0935[a]	0.0874[a]	0.0508	0.0317[a]
0	6	0.1178[a]	0.1200[a]	0.1226[a]	0.0500	0.0341[a]	0.1198[a]	0.1091[a]	0.0482	0.0328[a]
0	9	0.1359[a]	0.1403[a]	0.1257[a]	0.0458	0.0308[a]	0.1395[a]	0.1193[a]	0.0429[a]	0.0288[a]
0	12	0.1645[a]	0.1727[a]	0.1347[a]	0.0542	0.0303[a]	0.1714[a]	0.1343[a]	0.0510	0.0281[a]

Note. MM = minimum maximum likelihood type, TW = trimmed and Winsorized slope test, TS = Theil-Sen estimator, TT = t-statistic transformation.

[a] Type I error rates outside +/- 2 standard errors of the nominal 0.0500 rate (0.0456 to 0.0544).

Table 4

Empirical Type I Error Rates for Group Sample Sizes of 20 and 20 and a Residual Variance Ratio of 3/1

							Hybrid strategies			
Skew	Kurtosis	Standard	James	MM	TW	TS	James-TT	James-MM	James-TW	James-TS
0	-1.15	0.0428[a]	0.0398[a]	0.0820[a]	0.0566[a]	0.0291[a]	0.0393[a]	0.0346[a]	0.0531	0.0281[a]
0	0	0.0514	0.0481	0.1020[a]	0.0511	0.0323[a]	0.0482	0.0438[a]	0.0486	0.0305[a]
1.2	1.11	0.0630[a]	0.0598[a]	0.1339[a]	0.0550[a]	0.0304[a]	0.0591[a]	0.0809[a]	0.0509	0.0292[a]
1.6	2.86	0.0792[a]	0.0774[a]	0.1525[a]	0.0517	0.0317[a]	0.0757[a]	0.1063[a]	0.0477	0.0278[a]
0	3	0.0689[a]	0.0682[a]	0.1141[a]	0.0523	0.0303[a]	0.0674[a]	0.0653[a]	0.0480	0.0302[a]
0	6	0.0931[a]	0.0946[a]	0.1286[a]	0.0517	0.0304[a]	0.0938[a]	0.0824[a]	0.0470	0.0251[a]
0	9	0.1106[a]	0.1096[a]	0.1368[a]	0.0481	0.0326[a]	0.1091[a]	0.0886[a]	0.0440[a]	0.0263[a]
0	12	0.1176[a]	0.1206[a]	0.1367[a]	0.0479	0.0323[a]	0.1202[a]	0.0904[a]	0.0437[a]	0.0235[a]

Note. MM = minimum maximum likelihood type, TW = trimmed and Winsorized slope test, TS = Theil-Sen estimator, TT = t-statistic transformation.

[a] Type I error rates outside +/- 2 standard errors of the nominal 0.0500 rate (0.0456 to 0.0544).

Table 5

Empirical Type I Error Rates for Group Sample Sizes of 20 and 40 and a Residual Variance Ratio of 1/1

							Hybrid strategies			
Skew	Kurtosis	Standard	James	MM	TW	TS	James-TT	James-MM	James-TW	James-TS
0	-1.15	0.0300[a]	0.0306[a]	0.0247[a]	0.0546[a]	0.0263[a]	0.0304[a]	0.0252[a]	0.0532	0.0259[a]
0	0	0.0496	0.0486	0.0581	0.0507	0.0314[a]	0.0482	0.0506	0.0468	0.0320[a]
1.2	1.11	0.0701[a]	0.0720[a]	0.1174[a]	0.0571[a]	0.0327[a]	0.0713[a]	0.1090[a]	0.0504	0.0407[a]
1.6	2.86	0.1040[a]	0.1054[a]	0.1451[a]	0.0556[a]	0.0321[a]	0.1049[a]	0.1530[a]	0.0526	0.0412[a]
0	3	0.0931[a]	0.0929[a]	0.0981[a]	0.0531	0.0327[a]	0.0927[a]	0.0870[a]	0.0482	0.0349[a]
0	6	0.1235[a]	0.1286[a]	0.1143[a]	0.0525	0.0325[a]	0.1269[a]	0.1080[a]	0.0482	0.0341[a]
0	9	0.1524[a]	0.1599[a]	0.1226[a]	0.0522	0.0371[a]	0.1593[a]	0.1225[a]	0.0485	0.0356[a]
0	12	0.1677[a]	0.1804[a]	0.1253[a]	0.0528	0.0349[a]	0.1799[a]	0.1293[a]	0.0508	0.0338[a]

Note. MM = minimum maximum likelihood type, TW = trimmed and Winsorized slope test, TS = Theil-Sen estimator, TT = t-statistic transformation.

[a] Type I error rates outside +/- 2 standard errors of the nominal 0.0500 rate (0.0456 to 0.0544).

Table 6

Empirical Type I Error Rates for Group Sample Sizes of 20 and 40 and a Residual Variance Ratio of 1/3 (Direct Pairing)

							Hybrid strategies			
Skew	Kurtosis	Standard	James	MM	TW	TS	James-TT	James-MM	James-TW	James-TS
0	-1.15	0.0134[a]	0.0353[a]	0.0251[a]	0.0273[a]	0.0312[a]	0.0352[a]	0.0299[a]	0.0492	0.0295[a]
0	0	0.0218[a]	0.0534	0.0406[a]	0.0258[a]	0.0313[a]	0.0535	0.0508	0.0505	0.0335[a]
1.2	1.11	0.0324[a]	0.0660[a]	0.0851[a]	0.0274[a]	0.0279[a]	0.0656[a]	0.0913[a]	0.0499	0.0358[a]
1.6	2.86	0.0443[a]	0.0905[a]	0.1171[a]	0.0269[a]	0.0345[a]	0.0893[a]	0.1203[a]	0.0488	0.0383[a]
0	3	0.0379[a]	0.0789[a]	0.0654[a]	0.0244[a]	0.0349[a]	0.0792[a]	0.0746[a]	0.0488	0.0352[a]
0	6	0.0612[a]	0.1087[a]	0.0883[a]	0.0283[a]	0.0348[a]	0.1088[a]	0.0953[a]	0.0493	0.0363[a]
0	9	0.0765[a]	0.1327[a]	0.0997[a]	0.0290[a]	0.0347[a]	0.1319[a]	0.1011[a]	0.0466	0.0344[a]
0	12	0.0900[a]	0.1516[a]	0.1172[a]	0.0294[a]	0.0342[a]	0.1503[a]	0.1151[a]	0.0486	0.0286[a]

Note. MM = minimum maximum likelihood type, TW = trimmed and Winsorized slope test, TS = Theil-Sen estimator, TT = t-statistic transformation.

[a] Type I error rates outside +/- 2 standard errors of the nominal 0.0500 rate (0.0456 to 0.0544).

Table 7

Empirical Type I Error Rates for Group Sample Sizes of 40 and 20 and a Residual Variance Ratio of 1/3 (Inverse Pairing)

							Hybrid strategies			
Skew	Kurtosis	Standard	James	MM	TW	TS	James-TT	James-MM	James-TW	James-TS
0	-1.15	0.0827[a]	0.0360[a]	0.1722[a]	0.1028[a]	0.0305[a]	0.0356[a]	0.0342[a]	0.0497	0.0337[a]
0	0	0.0986[a]	0.0504	0.1860[a]	0.0989[a]	0.0341[a]	0.0504	0.0497	0.0477	0.0320[a]
1.2	1.11	0.1165[a]	0.0624[a]	0.2039[a]	0.0981[a]	0.0347[a]	0.0615[a]	0.0860[a]	0.0509	0.0342[a]
1.6	2.86	0.1392[a]	0.0798[a]	0.2017[a]	0.0954[a]	0.0386[a]	0.0769[a]	0.0998[a]	0.0503	0.0312[a]
0	3	0.1359[a]	0.0760[a]	0.1880[a]	0.1007[a]	0.0370[a]	0.0760[a]	0.0682[a]	0.0510	0.0343[a]
0	6	0.1620[a]	0.0997[a]	0.1902[a]	0.1032[a]	0.0354[a]	0.0995[a]	0.0797[a]	0.0531	0.0311[a]
0	9	0.1812[a]	0.1164[a]	0.1879[a]	0.0936[a]	0.0390[a]	0.1148[a]	0.0863[a]	0.0467	0.0301[a]
0	12	0.1862[a]	0.1263[a]	0.1810[a]	0.0887[a]	0.0391[a]	0.1260[a]	0.0903[a]	0.0471	0.0291[a]

Note. MM = minimum maximum likelihood type, TW = trimmed and Winsorized slope test, TS = Theil-Sen estimator, TT = t-statistic transformation.

[a] Type I error rates outside +/- 2 standard errors of the nominal 0.0500 rate (0.0456 to 0.0544).

Table 8

Empirical Type I Error Rates for Group Sample Sizes of 40 and 40 and a Residual Variance Ratio of 1/1

							Hybrid strategies			
Skew	Kurtosis	Standard	James	MM	TW	TS	James-TT	James-MM	James-TW	James-TS
0	-1.15	0.0280[a]	0.0284[a]	0.0206[a]	0.0526	0.0289[a]	0.0283[a]	0.0200[a]	0.0526	0.0300[a]
0	0	0.0479	0.0477	0.0559[a]	0.0504	0.0321[a]	0.0479	0.0498	0.0501	0.0353[a]
1.2	1.11	0.0672[a]	0.0677[a]	0.1115[a]	0.0484	0.0388[a]	0.0678[a]	0.1101[a]	0.0478	0.0521
1.6	2.86	0.1028[a]	0.1042[a]	0.1435[a]	0.0495	0.0371[a]	0.1032[a]	0.1635[a]	0.0494	0.0483
0	3	0.1006[a]	0.1028[a]	0.0900[a]	0.0518	0.0391[a]	0.1030[a]	0.0878[a]	0.0518	0.0428[a]
0	6	0.1396[a]	0.1436[a]	0.1044[a]	0.0489	0.0352[a]	0.1432[a]	0.1057[a]	0.0483	0.0381[a]
0	9	0.1709[a]	0.1752[a]	0.1151[a]	0.0467	0.0389[a]	0.1743[a]	0.1235[a]	0.0458	0.0398[a]
0	12	0.1952[a]	0.2004[a]	0.1200[a]	0.0483	0.0396[a]	0.1990[a]	0.1347[a]	0.0476	0.0387[a]

Note. MM = minimum maximum likelihood type, TW = trimmed and Winsorized slope test, TS = Theil-Sen estimator, TT = t-statistic transformation.

[a] Type I error rates outside +/- 2 standard errors of the nominal 0.0500 rate (0.0456 to 0.0544).

Table 9

Empirical Type I Error Rates for Group Sample Sizes of 40 and 40 and a Residual Variance Ratio of 3/1

							Hybrid strategies			
Skew	Kurtosis	Standard	James	MM	TW	TS	James-TT	James-MM	James-TW	James-TS
0	-1.15	0.0360[a]	0.0354[a]	0.0809[a]	0.0506	0.0352[a]	0.0353[a]	0.0310[a]	0.0494	0.0388[a]
0	0	0.0494	0.0479	0.0884[a]	0.0506	0.0386[a]	0.0482	0.0470	0.0490	0.0423[a]
1.2	1.11	0.0662[a]	0.0636[a]	0.1511[a]	0.0499	0.0432[a]	0.0639[a]	0.0958[a]	0.0484	0.0503
1.6	2.86	0.0851[a]	0.0849[a]	0.1828[a]	0.0515	0.0390[a]	0.0841[a]	0.1236[a]	0.0506	0.0434[a]
0	3	0.0844[a]	0.0821[a]	0.1191[a]	0.0548	0.0386[a]	0.0820[a]	0.0733[a]	0.0520	0.0400[a]
0	6	0.1025[a]	0.1011[a]	0.1279[a]	0.0528	0.0386[a]	0.1008[a]	0.0820[a]	0.0513	0.0378[a]
0	9	0.1272[a]	0.1290[a]	0.1427[a]	0.0512	0.0408[a]	0.1278[a]	0.0945[a]	0.0482	0.0386[a]
0	12	0.1426[a]	0.1460[a]	0.1431[a]	0.0486	0.0367[a]	0.1439[a]	0.0998[a]	0.0474	0.0322[a]

Note. MM = minimum maximum likelihood type, TW = trimmed and Winsorized slope test, TS = Theil-Sen estimator, TT = t-statistic transformation.

[a] Type I error rates outside +/- 2 standard errors of the nominal 0.0500 rate (0.0456 to 0.0544).

The combination of nonnormality and residual variance heterogeneity (Tables 4, 6, 7, and 9) produced somewhat unique Type I error patterns for the nine strategies. For the Standard test, residual variance heterogeneity usually made the effect of nonnormality less extreme, except for when sample sizes were inversely paired with residual variances, in which case Type I error was made more extreme. For James and James-TT, residual variance heterogeneity made the effects of nonnormality less extreme, though James did not react as much to the combination of unequal sample sizes and residual variances as the standard test. The MM test often had the most problematic Type I error rates for combinations of nonnormality and residual variance heterogeneity. The TW and TS tests were not particularly affected by the combination of nonnormality and residual variance heterogeneity; the TW test was mainly impacted by the combination of unequal sample sizes and residual variances, while the TS test was largely uninfluenced by anything. The Type I errors of hybrid strategies were reflective of the nonnormality strategy on which they were based, being liberal for James-MM, conservative for James-TS, and staying close to the 0.05 level for James-TW.

Power

Table 10 compares the power of the nine strategies across three considered sample size conditions with normal distributions, equal residual variances, and a population slope difference of 0.5. The most powerful strategies were the Standard, James, and James-TT tests, among which there was no overwhelming winner. The MM test had smaller power rates than the Standard, James, and James-TT tests. The James-MM hybrid test had less power than the MM test. The TW and James-TW tests had the smallest power rates of the considered strategies. The James-TS and TS tests had higher power rates than the TW and James-TW tests and (mostly) lower power rates than the MM and James-MM tests. The use of TS as a hybrid with James (James-TS) increased its power relative to the TS test.

Table 10

Empirical Power Rates for Population Slope Differences of .5 and Normality and Residual Variance Assumptions Met

Sample sizes							Hybrid strategies			
Group 1	Group 2	Standard	James	MM	TW	TS	James-TT	James-MM	James-TW	James-TS
20	20	0.3910	0.3901	0.3735	0.2270	0.2655	0.3906	0.3372	0.2215	0.2891
20	40	0.5096	0.4989	0.4730	0.3082	0.3931	0.4994	0.4403	0.2930	0.4202
40	40	0.6909	0.6912	0.6359	0.4369	0.6179	0.6912	0.6180	0.4353	0.6483

Note. MM = minimum maximum likelihood type, TW = trimmed and Winsorized slope test, TS = Theil-Sen estimator, TT = t-statistic transformation.

Slope Estimation

To gain further insight into the four slope estimation strategies (standard, MM, TW, and TS), Table 11 summarizes each strategies' 10,000 estimates of one slope with population value 0.5 in samples of size 20. When distributions were normal, all four strategies gave average slope values close to 0.5. The strategies' standard deviations show that the standard strategy's estimates were least variable, followed by the MM estimates, the TS estimates, and finally the TW estimates (corresponding to TW's relatively low power). The strategies' average estimated standard errors correspond to the overall liberalness/conservativeness of the strategies' significance tests, and for normal distributions show that on average all methods except for TS have standard errors that closely approximate slope variability. TS's bootstrapped standard errors overestimated TS slope variability, corresponding to the conservativeness of its Type I error rates.

The slope estimation results in Table 11 for a leptokurtic *Y* (kurtosis = 12) differ from those for a normal *Y* (kurtosis = 0). For a leptokurtic *Y*, all estimation strategies underestimate the population slope value of 0.5, where the least biased estimator is the Standard strategy while the most biased is the MM estimate. The Standard strategy's slope estimates are the most variable, while the TS estimates are the least variable. The average standard errors of the standard test and MM underestimate slope variability, corresponding to the liberalness of the standard test's and MM's Type I error rates. The TW estimates have standard errors that slightly underestimate slope variability. The TS estimate has standard errors that overestimate slope variability, corresponding to the conservativeness of TS. The results in Table 11 support previous findings that the TS estimator is more stable than the MM and Standard estimates when distributions are nonnormal (Wilcox & Keselman, 2004). These results extend previous work by showing that with nonnormality, the standard test and MM underestimate slope variability (making the Type I error rates of the standard and MM tests liberal), the Winsorized standard errors provide relatively accurate estimates of the variability of the trimmed slopes, while the TS bootstrap strategy overestimates slope variability (making the Type I error rates of the TS slope test conservative).

Table 11

Descriptive Analyses for the Slope Estimation Strategies (Population Slope = 0.5)

Skew	Kurtosis	Sample size		Standard	MM	TW	TS
0	0	20	Mean of slope estimates	0.5043	0.5043	0.5079	0.4998
			Standard deviation of slope estimates	0.2080	0.2214	0.2819	0.2275
			Mean of slope standard error estimates	0.2046	0.2205	0.2721	0.2752
0	12	20	Mean of slope estimates	0.4598	0.2987	0.3538	0.3280
			Standard deviation of slope estimates	0.2866	0.2361	0.2528	0.2159
			Mean of slope standard error estimates	0.1907	0.1701	0.2383	0.2575

Note. MM = minimum maximum likelihood type, TW = trimmed and Winsorized slope test, TS = Theil-Sen estimator, TT = t-statistic transformation.

Discussion

In this study, some recently researched strategies for testing independent groups' regression slopes were compared. The standard test of slope differences had robustness problems with nonnormality and with the pairing of unequal sample sizes and residual variances. Alternative strategies proposed for addressing nonnormality and for addressing both nonnormality and residual variance heterogeneity were also assessed. The most robust and powerful alternative strategies were the Theil-Sen strategy and a hybrid of Theil-Sen and the James second-order parametric alternative test. These Theil-Sen strategies had somewhat conservative Type I error rates that were largely unaffected by nonnormality and residual variance heterogeneity, and slope estimates that were efficient even for nonnormal data. The hybrid strategy of trimming and Winsorizing slope estimates and using them with the James test had Type I error rates that were closest of all the considered strategies to the nominal 0.05 level, but trimming and Winsorizing also produced slope tests with the lowest power rates of the considered strategies. The results also suggest that other strategies are not recommended, including James with a test statistic transformation for skewness, MM regression, and the use of MM estimates with James.

To gain some final insight into the considered slope estimation tests, a representative sample of 20 observations was generated from this study's kurtosis = 12 condition. Figure 1 shows these *XY* data and plots the Standard least squares regression line. There is one very extreme *X* observation (almost 3 standard deviations from *X*'s population mean of zero) that is also very low on *Y* (i.e., a bad leverage point). This observation causes the standard slope estimation strategy to underestimate the population slope of 0.5 in its slope estimate of 0.421. Figures 2 and 3 plot the observations in the data that are not excluded in computing the trimmed slope (Figure 2) and the Theil-Sen slope (Figure 3). The trimmed and Theil-Sen strategies underestimate the population slope more than the Standard strategy, producing slope estimates of 0.393 and 0.231, respectively.

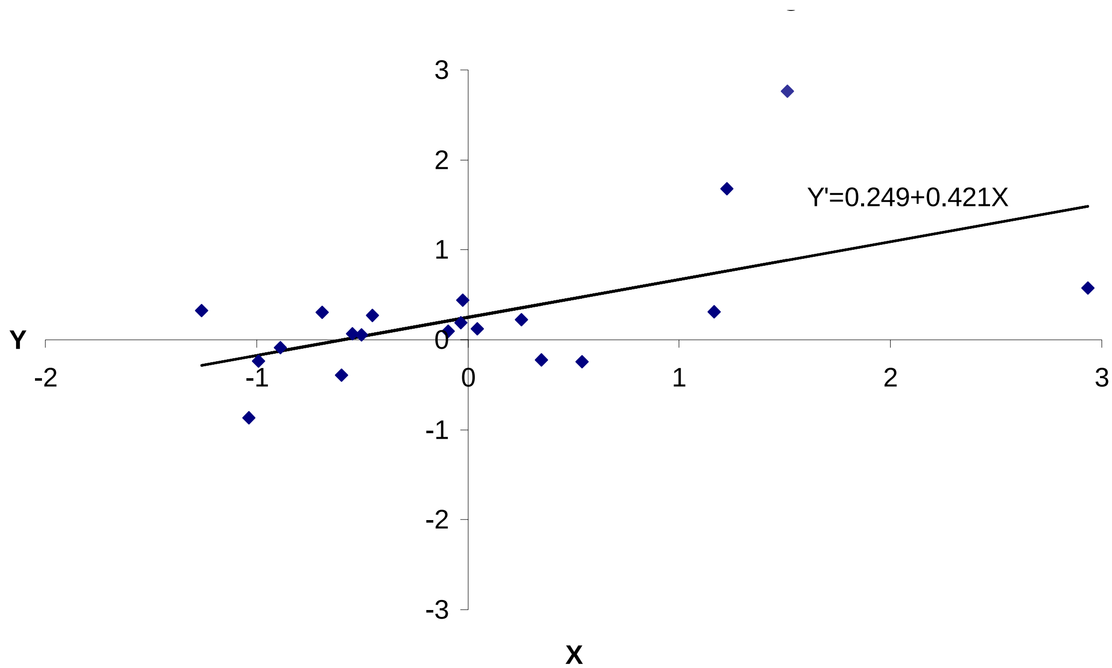

Figure 1. **Least squares regression with nonnormal Y. All 20 observations used to estimate the regression line.**

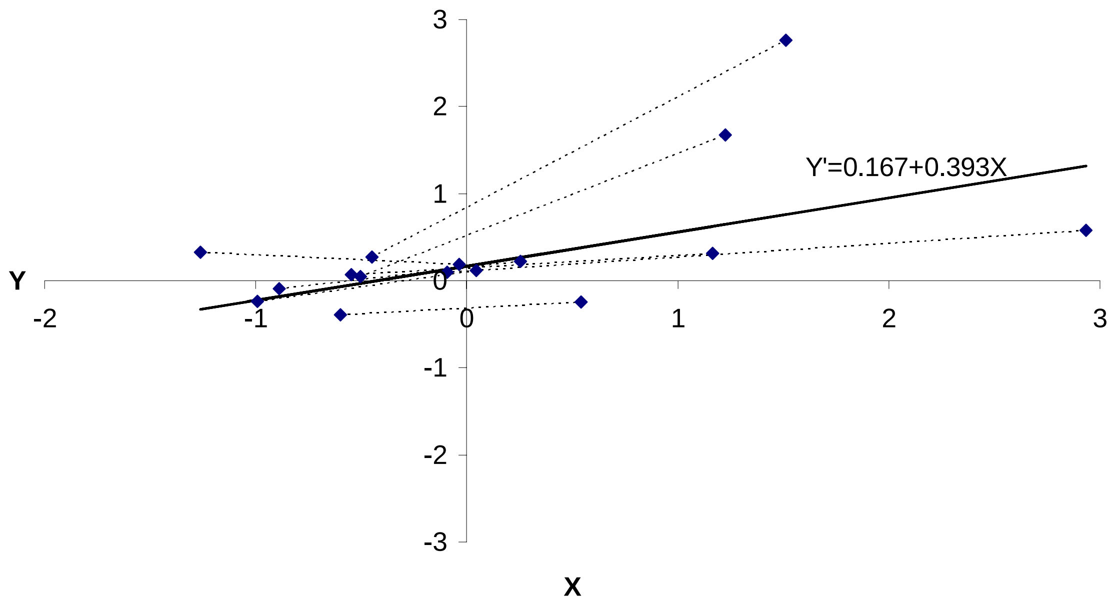

Figure 2. **Trimmed means regression with nonnormal Y. Untrimmed slopes used to estimate the regression line.**

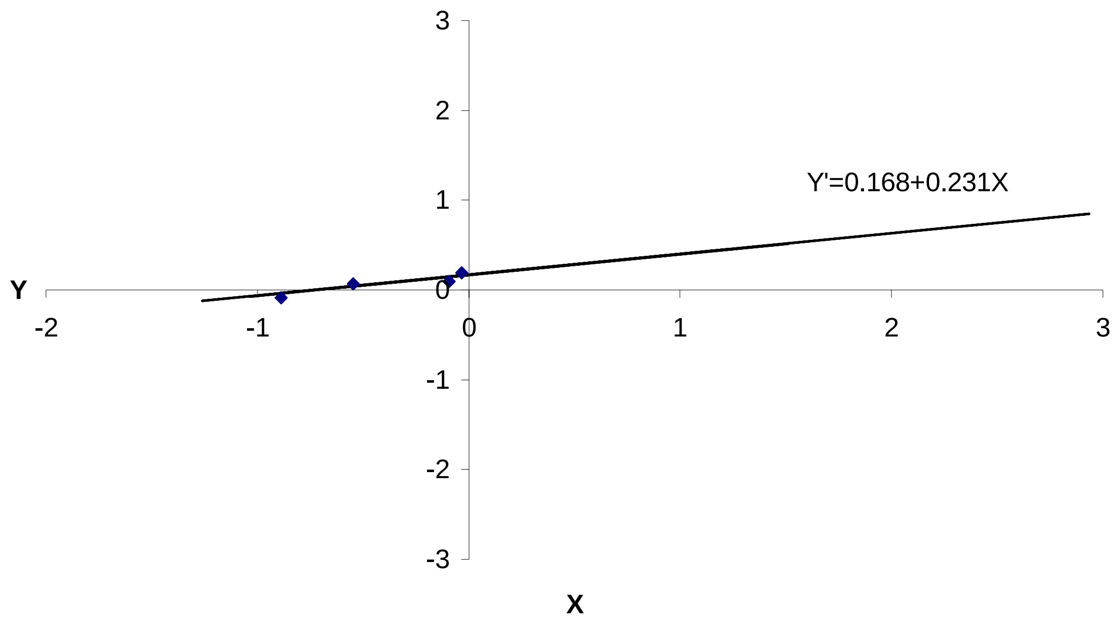

Figure 3. **Theil-Sen regression with nonnormal *Y*. Medians used to estimate the regression line.**

Figure 4 is especially useful for understanding the very complicated MM regression. All 20 of the original observations are used in MM regression, but contribute in weighted form to the final MM slope estimate. The observations' weights in Figure 4 show that the high-leverage observation is weighed very heavily by MM, causing the MM slope estimate to be relatively small (0.148). The observations that are far from the MM regression line are assigned small weights. Figure 4 shows that with MM, high leverage points can be weighted such that they influence the final slope estimate much more than the Standard least squares estimate. The large weights that are assigned to high leverage points in MM result in MM standard errors that underestimate slope variability (the $\mathbf{W}$ in Equation A5 is large) and inflate the Type I error of the MM strategy. Figure 4 makes it clear that the problems of the MM strategy with respect to high leverage points are not likely to be fixed by altering the weighting function, ξ, or the κ that determines how each of the scaled residuals are weighted. It may be possible to address MM's problems with high leverage data points through a wise choice of starting values that define the MM regression line and the residuals with respect to this line.

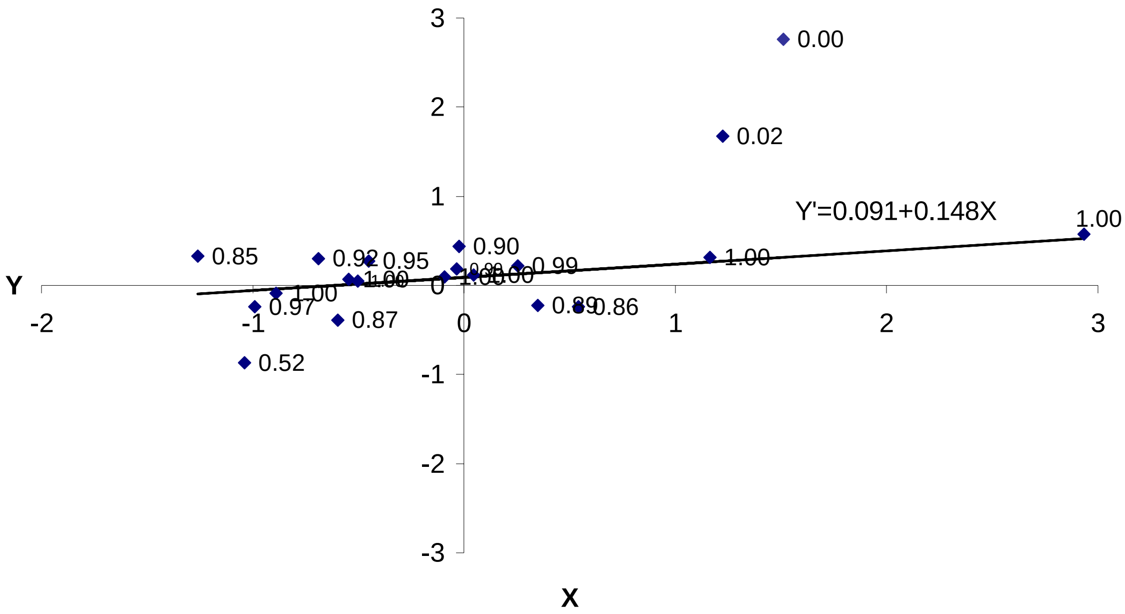

Figure 4. **Minimum maximum likelihood type (MM) regression. Weighted observations used to estimate the regression line.**

Implications

This paper considered only a small number of available slope test strategies. Some of the strategies not considered in this paper were excluded because of noted problems and criticisms, including nonparametric alternative tests (Deshon & Alexander, 1996; Dretzke et al., 1982; Marascuilo, 1966), residuals-based bootstrapping (Luh & Guo, 2000), ranked data (Headrick & Sawilowsky, 2000; Klockars & Moses, 2002), data transformations (Aguinis & Pierce, 1998; Glass, Peckham, & Saunders, 1972; Keselman, Carriere, & Lix, 1995; Wilcox & Keselman, 2004), several robust regression strategies (Anderson & Schumaker, 2003), and judgment-based elimination of outliers (He & Portnoy, 1992; Wilcox, 1996). This study suggests that a potentially promising slope test strategy would combine the best features of trimming and Winsorizing with Theil-Sen. By using the trimming and Winsorizing strategy on the $N(N-1)/2$ slopes that could be created out of all pairs of observations rather than only $N/2$ pairs, the final trimmed slope estimates should have stability levels that are similar to those of Theil-Sen, ultimately improving the power of the trimmed and Winsorized slope test. This proposed test would avoid the excessively time consuming bootstrapping associated with the Theil-Sen strategy, reduce the bias of the Theil-Sen estimates for nonnormal data, and provide the analyst some flexibility in terms of the extent of trimming used in the final slope estimates. Future considerations of how the number of slopes (Ebrahem & Al-Nasser, 2005) and the extent of trimming affect Type I error and power for nonnormal data would be useful for establishing this potential test of slope differences.

References

Aguinis, H., & Pierce, C. A. (1998). Heterogeneity of error variance and the assessment of moderating effects of categorical variables: A conceptual review. *Organizational Research Methods, 1,* 296-314.

Aguinis, H., Petersen, S. A., & Pierce, C. A. (1999). Appraisal of the homogeneity of error variance assumption and alternatives to multiple regression for estimating moderating effects of categorical variables. *Organizational Research Methods, 2*(4)*,* 315-339.

Alexander, R. A., & DeShon, R. P. (1994). Effect of error variance heterogeneity on the power of tests for regression slope differences. *Psychological Bulletin, 115*(2)*,* 308-314.

Anderson, C., & Schumaker, R. E. (2003). A comparison of five robust regression methods with ordinary least squares regression: Relative efficiency, bias, and test of the null hypothesis. *Understanding Statistics, 2*(2), 79-103.

Bradley, J. C. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31,* 144-152.

Conerly, M. D., & Mansfield, E. R. (1988). An approximate test for comparing heteroscedastic regression models. *Journal of the American Statistical Association, 83*(403), 811-817.

Conover, W. J., & Iman, R. L. (1982). Analysis of covariance using the rank transformation. *Biometrics, 38,* 715-724.

Cronbach, L. JJ. & Snow, R. E. (1977). Aptitude and instructional methods: A handbook for research on interactions. New York: Irvington.

Dance, K. A., & Neufeld, R. W. J. (1988). Aptitude-treatment interaction research in the clinical setting: A review of attempts to dispel the "Patient Uniformity" myth. *Psychological Bulletin, 104*(2)*,* 192-213.

Deshon, R. P., & Alexander, R. A. (1994). A generalization of James' second-order approximation to the test for regression slope equality. *Educational and Psychological Measurement, 54,* 328-335.

Deshon, R. P., & Alexander, R. A. (1996). Alternative procedures for testing regression slope homogeneity when group error variances are unequal. *Psychological Methods, 1*(3)*,* 261-277.

Dretzke, B. J., Levin, J. *R.,* & Serlin, R. C. (1982). Testing for regression homogeneity under variance heterogeneity. *Psychological Bulletin, 91,* 376-383.

Ebrahem, M. A., & Al-Nasser, A. D. (2005). Estimating the slope of simple linear regression in the presence of outliers. *Journal of Modern Applied Statistical Methods, 4*(2), 509-513.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*(4), 521-532.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research, 42,* 237-288.

Guo, J. H. (1996). Distribution-free test for slope in linear regression model with replications. *Communications in Statistics—Theory and Methods, 25,* 2069-2078.

Hall, P. (1992). On the removal of skewness by transformation. *Journal of Royal Statistical Society, B54,* 221-228.

He, X., & Portnoy, S. (1992). Reweighted LS estimators converge at the same rate as the initial estimator. *Annals of Statistics, 20,* 2161-2167.

Headrick, T. C., & Sawilowsky, S. S. (2000). Properties of the rank transformation in factorial analysis of covariance. *Communications in Statistics—Simulation and Computation, 29,* 1059-1087.

Holland, P., & Welsch, R. (1977). Robust regression using interactively reweighted least squares. *Communications in Statistics—Theory and Methods, 6,* 813-827.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin, 86*(4), 721-735.

James, G. S. (1951). The comparison of several groups of observations when the ratio of population variances is unknown. *Biometrika, 38,* 324-329.

Johnson, N. J. (1978). Modified *t* tests and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association, 73,* 536-544.

Keselman, H. J., Carriere, K. C., & Lix, L. M. (1995). Robust and powerful nonorthogonal analyses. *Psychometrika, 60*(3), 395-418.

Klockars, A., & Moses, T. (2002). Type I error rates for rank-based tests of homogeneity of slopes. *Journal of Modern Applied Statistical Methods, 1*(2), 452-460.

Luh, W., & Guo, J. (2000). Approximate transformation trimmed mean methods to the test of simple linear regression slope equality. *Journal of Applied Statistics, 27*(7), 843-857.

Luh, W., & Guo, J. (2002). Using Johnson's transformation with approximate test statistics for the simple regression slope homogeneity. *The Journal of Experimental Education, 71*(1), 69-81.

Marascuilo, L. A. (1966). Large-sample multiple comparisons. *Psychological Bulletin, 65,* 280-290.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105,* 156-166.

Myers, J. L., & Well, A. D. (1995). *Research design and statistical analysis.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Overton, R. C. (2001). Moderated multiple regression for interactions involving categorical variables: A statistical control for heterogeneous variance across two groups. *Psychological Methods, 6*(3), 218-233.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association, 79,* 871-880.

Rousseeuw, P. J., & Van Driessen, K. (2000). An algorithm for positive-breakdown regression based on concentration steps. In W. Gaul, O. Optiz, & M. Schader (Eds.), *Data analysis: Scientific modeling and practical application* (pp. 335-346). New York: Springer-Verlag.

SAS Institute. (2003). SAS/STAT user's guide, version 9.1: The ROBUSTREG Procedure [Software manual]. Cary, NC: Author.

Seligman, M. E. P. (2002). Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment. New York: Free Press.

Sen, P. K. (1968). Estimate of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association, 63,* 1379-1389.

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indagationes Mathematicae, 12,* 85-91.

Wang, X. (2005). Asymptotics of the Theil-Sen estimator in the simple linear regression model with a random covariate. *Nonparametric Statistics, 17*(1), 107-120.

Welch. B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika, 29,* 350-362.

Wilcox, R. R. (1996). A review of some recent developments in robust regression. *British Journal of Mathematical and Statistical Psychology, 49,* 253-274.

Wilcox, R. R. (2004). Some results on extensions and modifications of the Theil-Sen regression estimator. *British Journal of Mathematical and Statistical Psychology, 57,* 265-280.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). San Diego, CA: Academic Press.

Wilcox, R. R., & Keselman, H. J. (2004). Robust regression methods: Achieving small standard errors when there is heteroscedasticity. *Understanding Statistics, 3*(4), 349-364.

Yohai, V. J. (1987). High breakown point and high efficiency robust estimates for regression. *Annals of Statistics, 15,* 642-656.

Appendix A

The Minimum Maximum Likelihood Type (MM) Regression Algorithm

The first step of MM regression is to obtain robust starting values for the β_k's and σ. The current SAS procedure for MM uses LTS estimates as starting values (Rousseeuw, 1984; SAS Institute, 2003). The basic idea of LTS estimation is to draw samples of K observations from the *N* total observations in the data set. In each sample, obtain least squares estimates of the β_k's and find the ones that minimize $\sum_i^h (\varepsilon_i)^2$, where $h = \frac{3N + K + 1}{4}$ and observations i through h reference the h smallest squared residuals. Additional features of the LTS algorithm involve intercept adjustments that reduce $\sum_i^h (\varepsilon_i)^2$ and computational search processes designed to find final β_k estimates quickly in extremely large datasets (Rousseeuw & Van Driessen, 2000). One preliminary estimate of σ is computed as,

$$s_{LTS} = d\sqrt{\frac{1}{h}\sum_i^h (\varepsilon_i)^2}\ , \qquad \text{(A1)}$$

where $d = 1/\sqrt{1 - \frac{2N}{hc}\phi(1/c)}$, $c = 1/\Phi(\frac{h+N}{2N})$, and Φ and ϕ are the cumulative and probability density functions of the standard normal distribution.

A more efficient estimate of σ than s_{LTS} can also be computed,

$$\text{Wscale} = \sqrt{\frac{\sum_i^N w_i (\varepsilon_i)^2}{\sum_i w_i - K}}\ , \qquad \text{(A2)}$$

where $w_i = \begin{cases} 0 & \text{if} \quad |\varepsilon_i|/s_{LTS} > 3 \\ 1 & \text{otherwise} \end{cases}$.

With initial estimates of the β_k 's and σ, the second step is to conduct iterative calculations to produce a converged σ value,

$$\left(\sigma^{m+1}\right)^2 = \frac{1}{(N - K)(\int \xi(s)\partial\Phi(s))}\sum_{i}^{N}\xi(\frac{\varepsilon_i}{\sigma^m})\left(\sigma^m\right)^2, \tag{A3}$$

where $\int \xi(s)\partial\Phi(s)$ denotes an expected value of $\xi(s)$ when the s are from a normal distribution (about .25 for the Tukey bisquare $\xi(s)$ with $\kappa = 2.9366$). In (A3), setting $\kappa = 2.9366$ results in the σ having a breakdown rate of 25% (SAS Institute, 2003).

The third step is to conduct an iterative search for a final solution of the β_k 's with a fixed σ value,

$$\boldsymbol{\beta}^{\mathbf{m+1}} = \left(\mathbf{X}^{\mathbf{t}}\boldsymbol{\Omega}\mathbf{X}\right)^{\mathbf{-1}}\mathbf{X}^{\mathbf{t}}\boldsymbol{\Omega}\mathbf{Y}, \tag{A4}$$

where $\boldsymbol{\Omega}$ is an *N* by *N* matrix with diagonal entries $\frac{\partial\xi(s)}{\partial s}\frac{1}{s}$ where the s are the scaled residuals from the mth iteration step and $\kappa = 3.44$ in default SAS routines (SAS Institute, 2003). The entries for $\boldsymbol{\Omega}$ are the "reweighted" part of MM's iteratively reweighted least squares algorithm, and for the Tukey ξ (s) given in (13) are known as the Tukey bisquare weight function.

At convergence, there are several estimates of the asymptotic variance-covariance matrix of $\boldsymbol{\beta}$ (SAS, 2003). One version is,

$$\left(1 + \frac{K}{N}\frac{\sigma^2\left(\partial^2\xi(\varepsilon)/\partial^2\varepsilon\right)}{\left((1/N)\sum_{\boldsymbol{i}}\left(\partial^2\xi(\varepsilon_i)/\partial^2\varepsilon\right)\right)^2}\right)\frac{(1/(N - K)\sum_{i}\left(\partial\xi(\varepsilon_i)/\partial\varepsilon\right)^2}{\left((1/N)\sum_{i}\partial^2\xi(\varepsilon_i)/\partial^2\varepsilon\right)}\mathbf{W}^{-1}, \tag{A5}$$

where $\left(1+\frac{K}{N}\frac{\sigma^2\left(\partial^2\xi(\varepsilon)/\partial^2\varepsilon\right)}{\left((1/N)\sum_{i}\left(\partial^2\xi(\varepsilon_i)/\partial^2\varepsilon\right)\right)^2}\right)$ is a correction factor, $\partial^2\xi(\varepsilon)/\partial^2\varepsilon$ is the second derivative of ξ with respect to the residuals, and $\mathbf{W}$ is a K by K matrix with entries

$$\mathbf{W}_{kk'}=\sum_{i}\left(\partial^2\xi(\varepsilon_i)/\partial^2\varepsilon_i\right)x_{ik}x_{ik'}\ .$$

Appendix B

The Critical Value of the James Second-Order Test Statistics

The significance of the James' test statistic depends on whether it exceeds the following critical value,

$$\begin{aligned}
&\mathrm{James}_{crit} = \\
&c + (1/2)(3\chi_4 + \chi_2)\sum_j [(1 - wj_j)^2 / v_j] \\
&+(1/16)(3\chi_4 + \chi_2)^2[1 - (J-3)/c]\sum_j [(1 - wj_j)^2 / v_j]^2 \\
&+(1/2)(3\chi_4 + \chi_2)[(8R_{23} - 10R_{22} + 4R_{21} - 6R_{12}^2 + 8R_{12}R_{11} \\
&-4R_{11}^2) + (\chi_2 - 1)(2R_{23} - 4R_{22} + 2R_{21} - 2R_{12}^2 + 4R_{12}R_{11} - 2R_{11}^2) \\
&+(1/4)(3\chi_4 - 2\chi_2 - 1)(4R_{12}R_{11} - R_{12}^2 - 2R_{12}R_{10} - 4R_{11}^2 \\
&+4R_{11}R_{10} - R_{10}^2)] + (5\chi_6 + 2\chi_4 + \chi_2)(R_{23} - 3R_{22} + 3R_{21} - R_{20}) \\
&+(3/16)(35\chi_8 + 15\chi_6 + 9\chi_4 + 5\chi_2)(R_{12}^2 - 4R_{23} + 6R_{22} - 4R_{21} \\
&+R_{20}) + (1/16)(9\chi_8 - 3\chi_6 - 5\chi_4 - \chi_2)(4R_{21} - 2R_{22} - R_{20} + 2R_{12}R_{10} \\
&-4R_{11}R_{10} + R_{10}^2) + (1/4)(27\chi_8 + 3\chi_6 + \chi_4 + \chi_2)(R_{11}^2 - R_{22}) \\
&+(1/4)(45\chi_8 + 9\chi_6 + 7\chi_4 + 3\chi_2)(R_{23} - R_{12}R_{11})
\end{aligned} \tag{B1}$$

where $v_j = N_j - 2$, c is the 1-α quantile of the central chi-square distribution with *J*-1 degrees of freedom, $R_{ut} = \sum_j \frac{wj_j^t}{v_j^u}$ and $\chi_{2s} = \frac{c^s}{\prod_{q=1}^{s}(J + 2q - 3)}$ (for χ_2, χ_4, χ_6, and χ_8, *s* is 1, 2, 3, and 4, respectively).