EXTENDED RANK AND MATCHED-PAIR ANALYSIS OF COVARIANCE

by

CHU-CHIH CHEN

Department of Biostatistics
University of North Carolina

Institute of Statistics
Mimeo Series No. 2119T

September 1993
EXTENDED RANK AND MATCHED-PAIR ANALYSIS OF COVARIANCE

by

Chu-Chih Chen

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements of the degree of Doctor of Philosophy in the Department of Biostatistics

Chapel Hill

1993

Approved by:

Dane Shade  Advisor

Reader

M.D. Symons  Reader
ACKNOWLEDGMENTS

I sincerely thank my advisor, Professor Dana Quade, for his suggestion of this topic, and most of all, for his tremendous patience in correcting my thoughts in this research as well as my English writing. His full support and encouragement through the development of this thesis is most appreciated. Of course, the thesis could not have been feasible without his guidance and original work. I enjoy this topic very much.

I deeply thank Professor P. K. Sen for his many invaluable comments and insights of this research. The current version could not have been accomplished without his guidance and corrections.

Also, I want to thank Professor Michael Symons. His warm support through the years of my graduate study in this department both in spirit and in finance is greatly appreciated. I also have had a great time working for him as a research assistant.

I would like to thank Drs. Harry Guess and Imogene McCanless for their kind support as my doctoral committee members and reviewing of the manuscript. I am grateful to Dr. Guess for his suggestions, which help clarify a lot some of the concepts related to epidemiology studies.

Finally, I would like to thank my parents and my family as well as my friends for their constant support and encouragement, without whom my graduate study could be an endless struggle.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION AND LITERATURE REVIEW</td>
<td></td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Literature Review</td>
<td>2</td>
</tr>
<tr>
<td>1.2.1 Analysis of Covariance by Adjustment</td>
<td>2</td>
</tr>
<tr>
<td>1.2.2 Analysis of Covariance by Matching</td>
<td>6</td>
</tr>
<tr>
<td>II. DEFINITIONS AND BASIC FRAMEWORK</td>
<td></td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>14</td>
</tr>
<tr>
<td>2.2 The Assumptions</td>
<td>14</td>
</tr>
<tr>
<td>2.3 Definitions and Basic Framework</td>
<td>15</td>
</tr>
<tr>
<td>2.4 Extended Ranks and the Relationship Between the Generalized Scores and the Matched Pairs</td>
<td>22</td>
</tr>
<tr>
<td>2.5 Relationship Between the Coefficients of the Rank Adjusted Scores and the Weights of Pairs</td>
<td>24</td>
</tr>
<tr>
<td>2.6 Summary</td>
<td>28</td>
</tr>
<tr>
<td>III. ASYMPTOTIC RELATIVE EFFICIENCY PROPERTIES OF CALIPER MATCHING IN THE UNIVARIATE TWO-SAMPLE CASE</td>
<td></td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>29</td>
</tr>
<tr>
<td>3.2 Related Theorems</td>
<td>30</td>
</tr>
<tr>
<td>3.3 Tolerance ( \varepsilon ) of the Statistic ( T_1(\varepsilon) )</td>
<td>32</td>
</tr>
<tr>
<td>3.4 Asymptotic Normality of the Statistic ( T_1(\varepsilon) )</td>
<td>33</td>
</tr>
<tr>
<td>3.5 The Asymptotic Relative Efficiency of ( T_1(\varepsilon) ) With Respect to the Mann-Whitney Statistic</td>
<td>35</td>
</tr>
<tr>
<td>3.6 Category Matching</td>
<td>40</td>
</tr>
<tr>
<td>3.7 Discussion</td>
<td>46</td>
</tr>
<tr>
<td>IV. EXTENDED RANK AND MATCHED-PAIR ANALYSIS AND ITS EFFICIENCY PROPERTIES</td>
<td></td>
</tr>
</tbody>
</table>

iv
4.1 Introduction ........................................................................................................ 47

4.2 Asymptotic Distribution of the VR of the Rank Adjusted Scores and
   Its Efficiency Properties ............................................................................. 47

4.3 The Relationship Between the Spearman Correlation Coefficient $\rho_9$ and $\rho_9(\epsilon)$ Given $\epsilon$ 53

4.4 Estimation of the Optimal Weight $\tau$ ......................................................... 61

4.5 ARE of the VR of $Z^*_ij(\omega_1, \omega_2; \epsilon)$ With Respect to the VR of $Z^*_ij(\lambda; \epsilon)$ ............... 62

4.6 Category Matching Revisited ................................................................. 65

4.7 When the Response Variable is Dichotomous ......................................... 69

4.8 Discussion ...................................................................................................... 71

V. TEST STATISTIC OF MATCHING TOLERANCE

5.1 Introduction .................................................................................................... 73

5.2 Local Lack of Correlation in the Parametric ANOCOVA Setting ............. 74

5.3 Test Statistic of Tolerance $\epsilon$ .................................................................. 75

5.4 Discussion ...................................................................................................... 77

VI. EXAMPLES AND APPLICATIONS

6.1 Introduction .................................................................................................... 79

6.2 Example from the Study of the Efficacy of Nosocomial Infection Control ........................................................................................................... 79

6.3 Example of a Rheumatoid Arthritis Clinical Trial .................................... 84

6.4 Discussion ...................................................................................................... 88

VII. SUGGESTIONS FOR FUTURE RESEARCH

7.1 Introduction .................................................................................................... 89

7.2 Possible Research Work ............................................................................. 89

    7.2.1 Optimal Weights of Caliper Matching ................................................. 89

    7.2.2 When the Response is Discrete (at Least Ordinal) ......................... 90

    7.2.3 Adjustable Tolerance .......................................................................... 90

    7.2.4 Extensions to Ordered Alternatives .................................................. 91

    7.2.5 Decompositions of Sums of Squares With Tolerance ...................... 91

    7.2.6 Removal of the Concomitant Assumption ......................................... 91

    7.2.7 Actual Values Instead of Ranks .......................................................... 92
LIST OF TABLES

Table                      Page  
2.4.1 Small Example to Illustrate The Extended Ranks Given Tolerance $\epsilon$ .......................... 23  
2.4.2 Kernels of Two-sample U-statistics in Section 2.3 as Combinations of $M(\epsilon)$, $D(\epsilon)$, 
and $C(\epsilon)$ .................................................................................................................. 24 
6.2.1 Data from the SENIC Project ................................................................. 80  
6.2.2 Nonparametric ANOCOVA Test Statistics and Tolerance for Matching ............ 83  
6.3.1 Data Listing for Rheumatoid Arthritis Clinical Trial ................................. 85  
6.3.2 2 x 2 Tables of Dichotomized PRS (1=excellent or good, 0=all other) by Treatment 
(Active vs Placebo) by Age Categories ................................................................. 86
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.1</td>
<td>Average Length of Stay vs Hospital Service Index with Fitted Parallel Regression&lt;br&gt;Lines for North Central (0) and South (1) Regions</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Proportion Excellent or Good PRS by Treatment (Active vs Placebo) and Age&lt;br&gt;Category</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

It was Fisher's (1932) scheme of design of experiments that by completely random assignment of treatments to experimental units one can induce equivalent comparison groups. Standard statistical tools such as ANOVA can then be applied to test the treatment differences parametrically under normality assumptions or the Kruskal-Wallis test nonparametrically. However, due to the limitations of available sample sizes or experimental conditions, complete randomization cannot always be achieved. Often, there are some variables which are closely related to the response variable that cause imbalances in group comparisons. By controlling or adjusting for such variables, one can expect to offset the problem described above. The reasons for covariance analysis are summarized as follows (Koch et al, 1982; also, Cox and McCullagh, 1982; Cochran, 1957; Snedecor and Cochran, 1980):

(i) to adjust for inherent differences among comparison groups so that bias may be reduced;
(ii) to generate more powerful statistical tests through variance reduction if an appropriate covariable explains some of the variation present;
(iii) to induce equivalence of comparison groups that are generated by randomization;
(iv) to clarify the degree to which treatment effects are explained by other factors;
(v) to study the degree to which findings are uniform across subpopulations.

Typically, there are two distinct versions of analysis of covariance. One concept is based on adjustment. Let \((Y_{ij}, X_{ij})\) \(#1, 0\) be the j-th observation in the i-th group for response \(Y\) and covariable vector \(X\) respectively, where \(#1\) for the treatment group and \(#0\) for the control group. Suppose there is a mathematical structure that can be imposed on the relationship between \(Y\) and \(X\), say, \(E(Y) = \alpha + \beta(X)\). Then after "adjusting" for the effect of \(X\), one can go further to do a standard ANOVA test in the parametric setting or a Kruskal-Wallis test (for example) in the nonparametric setting under the null hypothesis of no treatment differences. But the linearity and normality as well as homogeneity assumptions of error terms of the parametric analysis of covariance (ANOCOVA) are hard to justify. Another approach (which is commonly used in epidemiology or social studies), is based on control by matching. A pair of observations
(Y_{1j}, X_{1j}) and (Y_{0j}, X_{0j}) is called matched, and thus Y_{1j} and Y_{0j} can be compared directly, if their covariates X_{1j} and X_{0j} are close to each other. But perfect matching — holding the covariable X to be constant — can only be approximately achieved when X is continuous.

The homogeneity of matched subjects has been assumed by almost all matching designs and analysis without much justification. Furthermore, inference from the analysis is only valid in matched subpopulations rather than the whole population. In addition, restriction to only the matched pairs may lose considerable information. The goal of this thesis is to develop a general nonparametric methodology that will incorporate both the concepts of adjustment and control, with the purpose of enhancing the efficiency of conventional matching analysis, while imposing less stringent assumptions on the relationship between Y and X than those of the regression setting of analysis of covariance.

1.2 Literature Review

In the following, we will use the terms analysis of covariance by adjustment and analysis of covariance by matching alternatively to mean the regression setting of analysis of covariance and matching analysis separately. Since the two methods have been developed separately along two different lines, we will discuss them individually in Sections 1.2.1 and 1.2.2. The advantages and limitations of each will also be discussed from existing results. We propose a unified approach to both the methods in Chapter 2.

1.2.1 Analysis of covariance by adjustment

The first illustration of the covariance method in the literature (Fisher, 1932) was to use the measurement of the covariable X on each experimental unit before the treatments are applied to predict the response Y of that unit to some degree (Cochran, 1957; Cox and McCullagh, 1982). The gain in precision from the covariance adjustment of the response Y through regression on X then depends on the size of the correlation coefficient \( \rho \) between Y and X. Let \( \sigma_y^2 \) be the experimental error variance when no covariance is employed; then the adjusted variance given X is \( \sigma_y^2 (1 - \rho^2) \left( 1 + \frac{1}{f_e} \right) \), where \( f_e \) is the number of error degrees of freedom (Cochran, 1957). In randomized experiments, the distributions of the covariable X in the different groups are approximately the same. For analysis of covariance in this setting randomization will generate equivalent comparison groups to compensate for the imbalance of any covariable X left out. Cochran pointed out that in observational studies, the analysis of covariance in the regression setting performs two distinct functions. One is to remove the bias. Let \((Y_{ij}, X_{ij})\) be the observed
response Y and covariate X on the j-th unit in the i-th population, i=1, 0. Suppose there is a mathematical structure such that

\[ Y_{ij} = \tau_i + \beta(X_{ij} - \bar{X}) + \epsilon_{ij}, \]

where \( \tau_i \) is the i-th treatment effect and \( \epsilon_{ij} \) is the error term of \( Y_{ij} \); then

\[ E(Y_{i.}) - E(Y_{i'.}) = \tau_i - \tau_{i'} + \beta(\bar{X}_{i.} - \bar{X}_{i'.}). \]

The amount of bias \( \beta(\bar{X}_{i.} - \bar{X}_{i'}). \) thus will render tests of significance and confidence limits invalid if not removed. Secondly, even if there is no problem of bias, analysis of covariance by adjustment can increase the precision of the comparison just as in randomized experiments. However, observational studies are subject to difficulties of interpretation from which randomized experiments are free. For example, when the covariable X is distributed differently among comparison groups, the covariance adjustment function then involves a greater or lesser degree of extrapolation. In this situation, covariance adjustment will not remove all the bias unless the regression equation holds in the region in which observations are lacking. Also, the adjusted differences may become insignificant statistically because of the enlarged standard errors.

Once the bias is removed by adjustment of the covariable X, the comparison of treatment effects then becomes clear if the regression equation is correct. Analysis of covariance by adjustment thus serves as a tool for performing analysis of covariance for the reasons listed in Section 1.1. Different situations of applying regression to one group only and non-parallel regression lines in the case of non-concomitant covariable X have also been discussed by Belson (1956), Cochran (1969), and Rubin (1973a).

Typically, one will perform parametric ANOCOVA under the assumption of linearity and parallelism. In addition, in the above model, one further assumes that the treatment effects are additive and the residuals \( \epsilon_{ij} \) are independently identically distributed as \( N(0, \sigma^2) \). ANOCOVA under the violation of normality was found to be much less robust than the corresponding analysis of variance, and the linearity assumption is critical (Atiquallah, 1964; Quade, 1967). Cox (1957b), on the other hand, examined the case where the relationship is not linear but is smooth, by comparing the method of analysis of covariance with a blocking method using blocks of the covariable X in increasing order, and concluded that blocking remains effective in efficiency while the covariance method does not.

Aside from the parametric approach, which treats the covariate X's as measured constants and fits a regression model assuming the same slope among different treatments, some nonparametric approaches are available. Quade (1967) proposed a rank analysis of covariance...
which can be summarized as follows:

"If the hypothesis $H_0$ is true that the conditional distribution of $Y$ given $X$ is the same for each population, then the populations are all identical and the samples can be pooled. So use the pooled sample to determine a relationship through which $Y$ can be predicted from $X$. Then compare each observed response $Y_{ij}$ with the value which would be predicted for it from the corresponding $X_{ij}$, and assign it a score $Z_{ij}$, positive if $Y_{ij}$ is greater than predicted, and negative if smaller. Finally, compare the populations by performing an ordinary one-way analysis of variance of the scores."

Let $R_{ij}$ be the corrected rank of $Y_{ij}$ among all $N = \sum n_i$ observed values of $Y$, meaning that $\lceil R_{ij} + (N+1)/2 \rceil$ is the ordinary rank, and let $\tilde{R}$ be the vector of ranks of $Y$ predicted from $X=(X^{(1)}, \ldots, X^{(s)})$ by performing an ordinary multiple linear regression of $R$ on $C^{(1)}, \ldots, C^{(s)}$, where $C_{ij}^{(u)}$ is the corrected rank of $X_{ij}^{(u)}$ among the $N$ observations on $X^{(u)}$, $1 \leq i \leq k$, $1 \leq j \leq n_i$. Then one can perform a one-way ANOVA on the residuals $Z_{ij} = R_{ij} - \tilde{R}_{ij}$. In the special case when $i=2$ and $p=1$,

\begin{equation}
Z_{ij} = R_{ij} - t_s C_{ij},
\end{equation}

where $t_s = \sum \sum R_{ij}C_{ij} / \sum \sum C_{ij}^2$ is the observed Spearman regression coefficient for the response $Y$ on the concomitant variable $X$. A more general score can be written as

\begin{equation}
Z_{ij}(\lambda) = R_{ij} - \lambda C_{ij},
\end{equation}

which is a function of the weight $\lambda$. It is then discussed when the variance ratio following the weighted score will achieve maximum asymptotic relative efficiency (ARE) with respect to the Kruskal-Wallis test, a nonparametric version of ANOVA, and the ARE with respect to the parametric ANOCOVA. It is shown that the AREs are $1/(1-\rho_s^2)$ and $3(1-\rho_p^2)/\pi(1-\rho_s^2)$ respectively, where $\rho_p$ and $\rho_s$ are the Pearson and Spearman correlations between $Y$ and $X$. Rank analysis of covariance avoids the assumptions of normality, linearity and homogeneity of variance while still fairly efficient in large samples. One disadvantage is the crucial assumption of concomitance of the covariables $X$ which is not required by the parametric approach. However, in randomized experiments, this assumption can be fairly satisfied.

Puri and Sen (1969) generalized rank analysis of covariance replaces the ranks $R_{ij}$ and $C_{ij}$ by rank scores $E_{N,R_{i\alpha}}^{(k)}$, $0 \leq i \leq p$, $1 \leq k \leq c$, $1 \leq \alpha \leq N$. They formed a conditional
distribution-free statistic given the permutation set. They then proposed a quadratic form of test statistic

$$\mathcal{L}_N = \left\{ \frac{V_{N,00}}{|V_N|} \right\} \sum_{k=1}^{c} n_k (T_{N,k}^*)^2$$

for the hypothesis that all treatments are equal, where the $T_{N,k}^*$ are the adjusted rank scores after fitting the regressions of rank scores of response variable on the scores of the $p$ covariates. Under the special case where the scores are ranks themselves, the proposed statistic $\mathcal{L}_N$ is a strictly monotonic function of VR($\lambda$) proposed by Quade (1967), and this technique is the same as Quade's approach. Furthermore, the permutation distribution of $\mathcal{L}_N$ will not depend on the unknown joint distribution $G$ of $Y$ and $X$, and hence an exact (conditional) test of size $\alpha$ can be obtained. In large samples, the exact permutation distribution of $\mathcal{L}_N$ can be approximated by a chi-squared distribution with $c-1$ degrees of freedom. The ARE properties were also studied and the following conclusions were drawn:

(i) When the scores are ranks, the ARE of $\mathcal{L}_N$ to the parametric test is bounded above by $(3\pi)^{-1}$ when $p \geq 1$ and the value can be very low when the underlying distribution is normal.

(ii) The generalized normal scores test is asymptotically as efficient as the classical test based on the variance ratio criterion when the underlying distribution is normal.

In the case of ordered alternatives in ANOCOVA, Boyd and Sen (1986) applied the union-intersection (UI) principle to form some rank tests and concluded that the UI test has asymptotically greater power than the global test based on $\mathcal{L}_N$ described before. On the other hand, Marcus and Genizi (1987) generalized caliper matching in the analysis of covariance (Quade, 1982) to the $k$-sample case and compared it to the Terpstra-Jonckheere statistic generalized for grouped data in the ordered alternatives case. They did some small sample simulation on different distributions of $Y$ and $X$ and discussed the power properties with respect to the tolerance $\epsilon$. They then proposed a general method of devising stepwise multiple testing procedures with fixed experiment-wise error.

Conover and Iman (1982) considered the rank transformation in ANOCOVA (also, McSweeney and Porter, 1971) and studied its robustness and power properties. They evaluated it by Monte Carlo methods and concluded that this approach appears to be robust and powerful and in certain cases gave similar results as Quade's approach (1967).

The above nonparametric approaches all require the assumption that the covariate $X$ is concomitant among treatment groups. However, when $X$ is nonconcomitant, test statistics based on the computed residuals would be affected by the estimate(s) of the unknown parameter(s), and hence are not distribution-free. Randles (1984) showed that, one can perform a "strongly
asymptotically distribution-free" analysis of covariance with a two-sample rank test for a location difference, provided that the distributions of the two covariate populations have the same means, which is required in the case of nonparallel lines. The statistic $S_n$ is said to be strongly asymptotically distribution-free if there exists a random variable $T_n$ such that $T_n$ is distribution-free and has a limiting distribution, and $(S_n - T_n) \to 0$ in probability as $n \to \infty$.

The discussions of the analysis of covariance by adjustment have mainly been based on assuming a continuous response $Y$ and covariate $X$. In the case of a dichotomous response, a $\chi^2$ test can be applied for comparing two treatments when no adjustment for a covariate is necessary. When there is a univariate continuous variable $X$ that is associated with the dichotomous response outcome $Y$, Bross (1964) considered pairwise comparisons of the "inversions": those pairs with response outcome and corresponding covariate going in the opposite directions (here assuming the correlation of $Y$ and $X$ is positive). Namely, one wishes to find the distribution of $(I_{10} - I_{01})$, where $I_{10}$ and $I_{01}$ are the numbers of inversions which favors the treatment and the control. The rationale is that this difference will provide clear-cut evidence of superiority. Ury (1966) considered conditioning on $r$, the proportion of comparisons potentially favoring the treatment, which is the degree of balance of distribution of the covariate $X$ among two treatment groups. Quade (1967) investigated the asymptotic distribution of $(I_{10} - I_{01})$ unconditionally through defining a function similar to that in defining rank scores. The variance ratio (VR) based on the resulting scores will then have asymptotically an $F$ distribution under the null hypothesis of no treatment effect. He also showed that the summation of statistic $Z_{ij}(\lambda)$ with weight $\lambda=1$ can be written in the form as $I_{10} - I_{01} + \frac{1}{2}(P_1 - P_2)$ where $P_1$ and $P_2$ are the number of "partial inversions" (those pairs which have higher values in either response group but are tied in their covariates) with greater response values (=1 in dichotomous case) of treatment group 1 and 0 respectively. Thus, one can get additional information from those pairs with the same response but varying in covariate $X$. Before we can go further into our approach, we need to review another mainstream of development, analysis of covariance by matching.

1.2.2 Analysis of Covariance by Matching

This arises very naturally in situations such as case-control and cohort follow-up studies in epidemiological research. In case-control studies subjects are selected separately from the case and control groups with matching on some covariates. They are then traced back for their exposure status and the odds ratios are then calculated for the strength of the exposure-disease relationship. In contrast, in follow-up studies, subjects are selected separately from the exposed and unexposed groups by matching on some variables. They are then followed up for a period
and the risk ratio of developing the disease of interest in exposed and unexposed groups is calculated within matched subgroups. It is assumed that subjects within the same matched group are homogeneous so that they can be compared with each other. Furthermore, the association across each subclassification (or matched group) is assumed to be consistent in the same direction so that it can be pooled together to form an overall test statistic for the difference in treatments. A well-known statistic following these assumptions is the Mantel-Haenszel (1959) statistic, a summation of the discrepancy between observations and expectations standardized by their variances across each strata. A test statistic for the homogeneity assumption across the strata was proposed by Breslow and Day (1982).

Similar to the role of removing the bias and enhancing efficiency of the study in the adjustment approach to analysis of covariance, matching also serves as a tool to solve these problems. Discussions of the role of matching have been mainly focused on the gain in efficiency by matching versus random sampling at the design stage.

In the case of dichotomous factors, Worcester (1964) discussed when the power of the McNemar $\chi^2$ statistic for pair-matching would exceed the ordinary crude $\chi^2$ for unmatched data. Bilewicz (1965) gave extensive simulation results comparing the variances of difference effect measures for pair-matched versus unmatched stratified analysis. Miettinen (1968, 1969) considered for follow-up studies the power efficiency of pair-matched analysis versus unmatched and unstratified analysis when validity issues can be ignored. He suggested that matching can lose efficiency when the matching factor is not strongly associated with the response (1970).

There is some degree of uncertainty as to whether one should adopt a matched design or simple random sample at the design stage. Kupper et al (1981) considered both the validity (lack of bias) and efficiency issues in comparing the use of matching and random sampling as alternative methods of subject selection in follow-up and case-control studies. They pointed out that pair-matching as used by the papers mentioned previously involves a random selection difficulty when the available subjects for pair selection are more than one. The method they used is thus category (frequency) matching by selecting matched subjects following a prespecified probability distribution which is the same as in exposed group or diseased group for different studies. They concluded that whether to match or not in follow-up studies should be based on efficiency considerations since a stratified analysis for the latter case can also serve to remove the bias (confounding factors). For case-control studies, however, matching can sometimes give misleading results. They investigated the gain in efficiency by comparing the widths of confidence intervals for some effect measures (risk ratios, risk differences in follow-up studies and odds-ratios in case-control studies) of the two corresponding methods. Their simulation shows that random sampling may be advantageous when matching causes substantial loss of available subjects, while there is an expected gain from matching when there is no
relationship between the method of sampling and the size of the comparison group, also, strong association between matching factor and disease status.

Walter (1989) evaluated the sampling effort involved if a matched design is used rather than a random unmatched design. He measured the effort by the number of candidates that must be sampled to identify the matches for analysis and the feasibility by the ratio of the expected value of N to its theoretical minimum and by its coefficient of variation. Greenland and Morgenstern (1990) pointed out that first, the decision whether to match or not must depend on whether one wishes to estimate relative or absolute effects. Second, the commonly drawn analogy between blocking in randomized trials and matching in cohort studies is misleading when one considers the impact of matching on covariate distributions. They concluded that, under an additive model, matching will usually increase the efficiency of both risk-difference and risk-ratio estimation, and the power of the Mantel-Haenszel test. Under a multiplicative model, the impact of matching is not as consistently beneficial. They illustrated by an example that, contrary to the conclusion by Kupper et al (1981), even when there is a strong confounder and matching produces no sample size reduction, a matched design can be less powerful and less efficient than an unmatched design.

Though in observational studies matching the treated and control units on the basis of observed covariates is an effort to compensate for the lack of randomization, there remains possible bias due to residual imbalances in other covariables. Rubin (1973b) applied regression adjustments after matching under different conditions for the matched variable and showed they perform better than either matching or regression separately through simulation. To take the possible confounders that are unmatched into account, one can stratify into a larger number of confounder categories in subsequent analysis. However, such an approach ignores matching and the resulting cell counts could be too small for asymptotic theory to hold. A seemingly satisfactory approach is by conditional likelihood analysis based on the logistic equation, which enables one to model the effects of covariates while preserving the original matching. A review of such approaches as well as categorical responses can be found in papers by Breslow (1982) and Koch et al (1982). See also McCullagh and Nelder (1990) for more recent results. Though logistic regression modeling is convenient and efficient in incorporating multiple unmatched variables into account, it inevitably shares the limitations of feasibility of complex model assumptions in the general regression setting.

For unobserved covariates in matched observational studies, Rosenbaum (1987) proposed a mechanism to perform a sensitivity analysis of permutation inferences under a range of assumptions on the unobserved covariate U after exact matching on a finite discrete covariate X. Given an assumed value of U, the conditional distribution of the test statistic and confidence limits for the treatment effect can be derived. The sensitivity analysis is applicable to
Wilcoxon's signed rank test and the McNemar-Cox test for paired binary responses. Similar permutation tests were also studied for observed covariates to adjust for the residual imbalance due to lack of homogeneity within pairs (Rosenbaum, 1988).

In the case of a continuous covariable X, there is no obvious criterion for defining a match. Several matching methods are available in literature. Greenberg (1953) proposed balancing the distributions of covariates among two compared samples by matching on their means (mean matching). The aim of this method is to make the compared groups have the same average value of the covariate, so that there is no need for adjustment of the covariables since at least the first moments are identical.

The method of pair-matching is widely used in epidemiology. Subjects are selected 1:1 or 1:M based on some matching criteria in follow-up or case-control studies. However, such a procedure requires a random selection if the choices are more than one. Rubin (1973a) drew the matched subjects from a larger control population by a nearest-neighbor criterion that the subject not yet matched with the lowest score on covariate X is drawn to be matched first (low-high) and high-low the opposite ordering.

The matching method that is most commonly in use is category matching: grouping into adjacent intervals. Observations on the covariate then fall within a specific interval following a multinominal distribution. Criticism has been made on the arbitrariness of cut points, and because two observations with very close values may fall into different categories (Greenberg, 1953; Quade, 1982). For example, two subjects with ages 39 and 41 may not be matched, while two subjects with ages 41 and 49 are matched, if the cut points are at 40 and 50.

Caliper matching (Quade, 1974, 1982) avoids this disadvantage. Here one allows a tolerance $\epsilon$ such that two observations $X_1$ and $X_0$ are matched if the distance between the two falls within $\epsilon$ (Quade, 1982). Thus, subjects being matched in caliper matching do not necessarily form disjoint matched subgroups, and loss of information for comparison of treatment difference due to matching may be less severe than with other matching methods. On the other hand, there is still an arbitrariness in the choice of $\epsilon$. It is argued that the decision of tolerance $\epsilon$ should be left to experts from the specific field of research rather than by statisticians (Quade, 1974), although statistically one can find the appropriate range of $\epsilon$ by examining the data at hand.

Furthermore, though it is desired to hold the covariate X constant within each matched pair or subgroup so that treatments can be compared directly, this can only be approximately achieved when X is continuous. For example, once two subjects are matched, then it is expected that the treatment difference due to the influence of the covariable X is negligible. As a result, for the matched observations, the fitted line is flat in parametric ANOCOVA and the corresponding ranks of the covariable are tied in rank analysis of covariance (Quade, 1967).

Different matching criteria all seem to share the same assumption that the matched subjects
are homogeneous with respect to the matching variable in analysis. It seems little can be found in the literature to justify this assumption except for the permutation adjustment of the residual imbalance (e.g. Rosenbaum, 1988). An attempt to test this assumption and to find a reasonable range of tolerance $\epsilon$ by caliper matching will be made in the present work.

Just as for analysis of covariance by adjustment, studies have been made on the role of removing bias and enhancing precision in the case where both covariate and response are continuous. Cochran (1968) considered the following three aspects of the effect of subclassification of continuous covariate X on the continuous response Y:

(i) the effect of removing bias when the covariate X has the same distribution with a location shift in two treatment groups;

(ii) the effect of subclassification on the precision of the comparison of the two population means;

(iii) the effect of measurement error on the performance of the subclassification.

Measurement error is a more complicated issue which will not be discussed here. Cochran evaluated bias removal in terms of the proportion of reduction in bias based on the linear regression model of Y on X. He showed that if the density functions of X in Groups 1 and 2 are $f(x)$ and $f(x-\theta)$ respectively, then the proportion of bias removed is $\sum_{i=1}^{c} M_i (f_{i-1} - f_i)$ if $c$ subclassifications are made, where $f_{i-1}$ and $f_i$ are the ordinates of $f(x)$ at the boundaries $x_{i-1}$ and $x_i$ and $M_i$ is the mean value of X in the $i$-th subclass. He also showed that, given normality, the proportion of reduction in bias by subclassification is the same as the proportion of reduction in variance given by Cox (1957a). Rubin (1973a) followed the same terminology of proportion of reduction in bias. He proposed a general matching scheme for choosing a matched sample from a larger control population and derived an index of bias reduction due to matched sampling under the linear regression model:

$$\theta = 100 \frac{\eta_{2*} - \eta_2}{\eta_1 - \eta_2},$$

(1.6)

where $\eta_i$ is the mean of X in population $i$ and $\eta_{2*}$ is the expected value of X in the matched sample. He then compared the performance of nearest-neighbor matching to mean matching and (random) pair-matching for their ability in removing bias through simulation. He concluded that a single mean matching can effectively remove the bias. However, in practice, one would choose pair matching rather than mean matching. One theoretical justification is that mean matching can have expected zero bias only when the response surface (functional form of X with respect to Y) is really linear while this is not necessarily the case for individual pair matching.

Billewicz (1965) studied extensively the efficiency of matching for the continuous variable case under various conditions. The underlying population he used for simulation is such that X
is $N(0, 1)$ and given $X, Y$ is $N(\mu_y|x^1, 1)$. He then grouped the samples into three to four categories for the properties of matching to compare with parametric ANOCOVA, and drew the following conclusions:

(i) Matching wastes information in the statistical sense.

(ii) When the number of matching criteria appears too large, it is difficult to match for many categories and it is time consuming.

(iii) Results from simulated "freak" samples show that matching is 20% more likely to miss a difference than using random samples.

(iv) Matching improves the precision of the experiment less than would be achieved by covariance analysis of adjustment for quantitative response variables.

Cochran (1968) considered the effect of the subclassification on the precision of the comparison of two population means by taking the ratio of the variance of the adjusted difference of the means to that of the initial difference. He examined the case when the marginal distributions of the covariable $X$ are concomitant and the regression of $Y$ on $X$ is linear. The net result is to reduce the variance of the difference in means $\bar{y}_2 - \bar{y}_1$ from $2\sigma_y^2/n$ to $2\sigma_y^2[1 - (1 - g)\rho^2]/n$, where $g = \sum p_i\sigma_1^2/\sigma_x^2$, $p_i$ is the proportion of the population and $\sigma_1^2$ the variance of $x$ within the $i$-th subclass, $\sigma_y^2$ and $\sigma_x^2$ are the variances of $Y$ and $X$, and $\rho$ is the correlation between $Y$ and $X$. General conclusions he reached for the percentage reduction in variance were:

(i) The percentage increases when the number $c$ of subclasses increases and $\rho$ is high.

(ii) In the case of $\rho < 0.7$, $c \geq 4$ is enough.

(iii) The linear covariance adjustment provides greater gain than stratified matching, which is the same result as obtained by Billewicz.

The simulations made by Billewicz and Cochran both take advantage of the fact that the underlying distribution is normal and regression is linear. Hence, it might be expected that regression would perform better than subclassification.

Raynor and Kupper (1981) indicated that, rather than concomitance of the covariate $X$, in case-control studies the matched variable usually has greater variance in the control group than in the case group. Given that $X$ is $N(\mu_0, \sigma_0^2)$ in the control group and $N(0, 1)$ in the case group, they showed that the relative efficiency of category-matching to the parametric analysis of covariance increases as $\mu_0$ increases and always passes one. They then concluded that the category-matched mean difference estimator can be a more efficient estimator than the analysis of covariance estimator in certain situations. To take the bias into account, they also investigated the relative mean squared error of the two estimators and concluded that the
category-matched mean difference estimator is a better estimator when there are a large mean difference, moderate sample sizes, and low to moderate correlation between the case group matching and response variables.

As was pointed out by Quade (1967), matching has the advantage of allowing the covariates to be completely arbitrary in form, even purely nominal, and also allowing any arbitrary relationships among covariates and response. Analysis of covariance by matching remains a convenient tool when the model assumptions are not justifiable.

Schoenfelder (1981; also, Schoenfelder and Quade, 1984) considered matching when the response Y is continuous and the covariate X is discrete, in the concomitant case. He showed that the ARE of a weighted linear combination of the Mann-Whitney statistics computed at the observed values of X for all matched pairs (AMPs) in an analysis of covariance is 1.5 with respect to the sign test of individual matched pairs (IMPs). The optimal weight obtained is the reciprocal of the corresponding probability density at the observed value of X. Thus, in the continuous covariate case, one can categorize X into intervals and estimate the corresponding density by their frequencies under this framework. This approach avoids any particular form of the conditional distributional distribution of Y given X, and is still highly efficient relative to individual pair-matching. Schoenfelder and Quade (1983) also investigated the case when X has finite support, to find the ARE of the optimal test statistic of AMPs with respect to the parametric ANOCOVA when the assumptions of the latter are satisfied. The ARE is $\frac{3}{8}$, which is the same as that of the Mann-Whitney statistic with respect to the parametric t-test. Hence, if all parametric assumptions are valid, the loss of efficiency is the same in both situations.

Quade (1982) proposed caliper matching for analysis of covariance when X is continuous and concomitant. Let $(Y_{ij}, X_{ij})$ be the corresponding response variable Y and covariate X in treatment group i and subject j; then the “matched difference in means” he proposed is

$$\bar{D} = \frac{\sum_j \sum_{j'} (Y_{1j} - Y_{0j'}) I\{D(X_{1j}, X_{0j'}) \leq \epsilon\}}{\sum_j \sum_{j'} I\{D(X_{1j}, X_{0j'}) \leq \epsilon\}}.$$  \hspace{1cm} (1.7)

Also, an alternative measure is the “matched difference in probability”

$$T = \frac{\sum_j \sum_{j'} \text{sgn}(Y_{1j} - Y_{0j'}) I\{D(X_{1j}, X_{0j'}) \leq \epsilon\}}{\sum_j \sum_{j'} I\{D(X_{1j}, X_{0j'}) \leq \epsilon\}}.$$  \hspace{1cm} (1.8)

Since the two matched difference statistics are both ratios of U-statistics, it can be shown that by standard U-statistic theory they are asymptotically normally distributed. He derived the standard errors of the corresponding statistics, which can be used to construct confidence
intervals for the population matched difference in mean and difference in probability. However, the power and efficiency properties of this method have not yet been established.

A notion in nonparametric smoothing similar to the matched difference D (1.7) is the “local average” of the response variable y at a point x (see, e.g. Härdle, 1990), which can be defined as

\[
\bar{m}(x) = n^{-1} \sum_{i=1}^{n} W_{ni}(x) Y_i ,
\]

where \( W_{ni}(x) \) denotes a sequence of weights such that

\[
n^{-1} \sum_{i=1}^{n} W_{ni}(x) = 1 .
\]

Let

\[
d_{ij'} = y_{ij} - y_{0j'} ,
\]

and

\[
W_{jj'} = \frac{1}{\sum_j \sum_{j'} I\{D(X_{1j}, X_{0j'}) \leq \epsilon\}} \sum_j \sum_{j'} I\{D(X_{1j}, X_{0j'}) \leq \epsilon\} .
\]

Then, clearly,

\[
\bar{D} = \sum_j \sum_{j'} W_{jj'} d_{jj'} ; \sum_j \sum_{j'} W_{jj'} = 1 .
\]

Thus, rather than smoothing at the point \((Y_i, X_i)\) as in nonparametric regression, the statistic \(\bar{D}\) is the weighted average difference of the response of the treatment group and the control group when their covariates are within the distance \(\epsilon\) for all the \(n_1 n_0\) possible comparisons. The weight \(W_{jj'}\) is similar to the normalized kernel estimator, and the tolerance \(\epsilon\) is equivalent to the smoothing parameter \(h\) (bandwidth) in smoothing technique. However, the matching interval \(\epsilon\) here is a fixed constant rather than dependent on the sample size \(n\), which is the case for selection of the bandwidth \(h\). Other forms of \(W_{jj'}\) are also possible following different features of the corresponding kernel estimators.

The matching method, though requiring less stringent structure assumptions than the analysis of covariance by adjustment, appears to lack power relative to the regression setting if some structure can be imposed on the data. Often, there exists a positive or negative correlation between the response variable Y and the covariate X. Unlike the parametric (nonparametric) ANOCOVA approach, matching methods fail to take this information into account. The present work will try to compensate for this disadvantage by incorporating the “inversion pairs” proposed by Bross (1964) together with the matched pairs to form a new test statistic. We will illustrate this approach together with the research plan in the next chapter.
CHAPTER 2
DEFINITIONS AND BASIC FRAMEWORK

2.1 Introduction

We define Extended Rank and Matched-Pair (ERMP) analysis of covariance in the k-sample and p-covariate case in this chapter. By matched pairs, we mean caliper matching, that is, pairs of observations with their covariates within a tolerance $\epsilon$. In Section 2.2, we give assumptions needed throughout the work. The basic framework of research and definition of the ERMP test is given in Section 2.3. By parametrizing a general score for the kernels of the proposed U-statistics, we show in Section 2.4 that the present work is a unified approach to both rank analysis of covariance (Quade, 1967) and "difference in probability" (Quade, 1982). We explore the relationship between the coefficients of extended ranks and the weights of pairs in Section 2.5. We can therefore examine the relationship between the response and the covariates based on either regression or matching. We present an outline for further development in Section 2.6.

2.2 The Assumptions

We present the assumptions needed for the two-sample case. The multi-sample case can be extended directly by replacing the treatment group (denoted by 1) and the control group (denoted by 0) with $i$-th and $i'$-th sample respectively. Suppose we have the following assumptions:

A1. The responses in the treatment and control groups have absolutely continuous distribution functions given $x$

\[
F_1(y \mid x) \text{ and } F_0(y \mid x)
\]

respectively, and hence there exist density functions $f_1(y \mid x), f_0(y \mid x)$.

A2. The marginal distribution of $X$ is the same, with distribution function $G(x)$, in both groups; i.e., they are concomitant.
A3. The Y's are mutually independent given the X's.

When there are k-treatment groups, we may replace Assumption A1 by A1':

A1'. The response in the i-th sample has absolutely continuous distribution function \( F_i(y|x) \) given \( x \).

In Chapter 3, we are mainly interested in testing the hypothesis that there is a location shift between the treatment group and the control group given \( x \), i.e.

\[
F_1(y - \theta|x) = F_0(y|x); \quad H_0 : \theta = 0 \text{ vs. } H_\alpha : \theta > 0.
\]

In Chapter 4, the location shift assumption is relaxed further where we consider only whether there is stochastic order between the treatment groups given \( x \). Under a more general set-up, the continuity assumption of the response variable in A1 is dropped to incorporate the (ordinal) discrete response variable case.

2.3 Definitions and Basic Framework

Our first task is to find the efficiency of the "matched difference in probability" (caliper matching) compared with the Mann-Whitney statistic U (simple random sampling). The main idea is that by incorporating the covariate \( X \) into our analysis, rather than ignoring it as the Mann-Whitney statistic U does, one could improve efficiency if the response \( Y \) is correlated with \( X \).

Suppose one is interested in whether \( Y_1 \) is stochastically larger than \( Y_0 \) given \( X \). More specifically, the null and the alternative hypothesis are

\[
H_0 : \quad F_1(y|x) = F_0(y|x) \quad \text{vs.} \quad H_\alpha : \quad F_1(y - \theta|x) = F_0(y|x), \quad \theta > 0.
\]

An intuitive and reasonable estimate of \( \theta \) would be

\[
\hat{\theta} = \frac{1}{M} \sum_{j=1}^{n_1} \sum_{j' = 1}^{n_0} (y_{1j} - y_{0j'}) I \{ |x_{1j} - x_{0j'}| \leq \epsilon \},
\]

where \( M = \sum_{j=1}^{n_1} \sum_{j' = 1}^{n_0} I \{ |x_{1j} - x_{0j'}| \leq \epsilon \} \), which is the statistic \( D \) in Quade (1982). One can
then test for $H_0$ following the asymptotic normality of $\tilde{\nu}$.

A different approach, which stresses only the stochastic order of the response of the comparison groups, is by looking at the “difference in probability”. The statistic proposed in that paper is:

$$T = \frac{1}{M} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_0} \text{sgn}(Y_{1j} - Y_{0j'}) I(|X_{1j} - X_{0j'}| \leq \epsilon).$$

For convenience in deriving the asymptotic distribution and ARE with respect to the Mann-Whitney statistic $U$, we use the statistic $T_1$ instead of $T$: $T_1 = \frac{MT}{n_1 n_0}$. Note that $T_1$ is a U-statistic of degree $(1, 1)$ with kernel

$$\text{sgn}(y_1 - y_0) I(|x_1 - x_0| \leq \epsilon),$$

which is an estimate of

$$\theta_{10} = P \{ Y_1 > Y_0, |X_1 - X_0| \leq \epsilon \} - P \{ Y_1 < Y_0, |X_1 - X_0| \leq \epsilon \}.$$

$\theta_{10}$ is zero under the null hypothesis and greater than zero under the alternative hypothesis.

Matched-pair analysis collects information on the treatment difference by controlling for the covariates to be equal or approximately equal. However, nothing can be done with the unmatched observations and hence they are discarded. Because of this inefficiency, one often needs to sample more observations which, of course, is more costly and time consuming. For example, in case-control studies, more controls for each case often need to be sampled. It has also been pointed out that there is a built-in regression effect left in matched pair analysis which costs efficiency. The real situation could be that the matched-pair analysis fails to take fully into account the positive or negative correlation which often exists between response and covariates.

Bross (1964) discussed the analysis of covariance by using the “inversions” when the response is a dichotomy. The reason might be that these pairs carry more information than those where the covariates are equal (matched). It is thus quite reasonable to include the inversions and assign even more weight to them than to the matched pairs. Let us reconsider the randomization test after adjustment in the following case:

$$E(Y_1) = \alpha_1 + \beta(X_1); \quad E(Y_0) = \alpha_0 + \beta(X_0).$$

Here, rather than comparing $Y_1$ and $Y_0$ directly, one compares $Y_1 - \beta(X_1)$ and $Y_0 - \beta(X_0)$ for the implicit difference $\alpha_1 - \alpha_0$ after adjusting for the effect of covariate through the regression function $\beta(X)$. Suppose $Y$ and $X$ are positively correlated and thus $\beta(X)$ is monotone increasing. If $Y_1 > Y_0$ and $X_1 < X_0$, or $Y_1 < Y_0$ and $X_1 > X_0$ (an inversion), then caliper matching (and
all other matching methods) does not include this pair unless \( |x_1 - x_0| \leq \epsilon \). Matching methods avoid the structure assumption, which is often hard to justify. The relative gain is, of course, the much looser assumption of relationship between Y and X rather than explicit functional form. However, the correlation between Y and X is still not fully taken into account by matching. By including the inversions, we can expect to get a better performance which is closer to the regression approach than conventional matched pair analysis.

We give definitions which will be needed throughout the development.

**Definition 2.3.1**

Let the kernels

\[
M((y_1, x_1), (y_0, x_0); \epsilon) = \text{sgn}(y_1 - y_0) \cdot I\{|x_1 - x_0| \leq \epsilon\},
\]

(2.12)

\[
D((y_1, x_1), (y_0, x_0); \epsilon) = \frac{\text{sgn}(y_1 - y_0) - \text{sgn}(x_1 - x_0)}{2} \cdot I\{|x_1 - x_0| > \epsilon\},
\]

(2.13)

\[
C((y_1, x_1), (y_0, x_0); \epsilon) = \frac{\text{sgn}(y_1 - y_0) + \text{sgn}(x_1 - x_0)}{2} \cdot I\{|x_1 - x_0| > \epsilon\}.
\]

(2.14)

Then we have three two-sample U-statistics based on these kernels.

Standardized number of matched pairs:

\[
M(\epsilon) = \frac{1}{n_1 n_0} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_0} M((y_{1j}, x_{1j}), (y_{0j'}, x_{0j'}); \epsilon) = T_1(\epsilon).
\]

(2.15)

Standardized number of inversions (discordant unmatched pairs):

\[
D(\epsilon) = \frac{1}{n_1 n_0} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_0} D((y_{1j}, x_{1j}), (y_{0j'}, x_{0j'}); \epsilon).
\]

(2.16)

Standardized number of concordant unmatched pairs:

\[
C(\epsilon) = \frac{1}{n_1 n_0} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_0} C((y_{1j}, x_{1j}), (y_{0j'}, x_{0j'}); \epsilon).
\]

(2.17)

We consider the statistic \( T_1(\epsilon) \) where all pairs are matched within a positive tolerance \( \epsilon \). When \( \epsilon = 0 \), \( T_1(0) = 0 \) (assuming no ties on X), which is a trivial case. And as \( \epsilon \to 0 \), the number of matched pairs is almost zero when X is continuous, which is expected to have low efficiency. Therefore, we exclude these cases for the statistic \( T_1(\epsilon) \).

Matching methods have been used mainly in epidemiology studies for specific considerations.
As pointed out, such methods lose efficiency when compared with classical analysis of covariance. In the following, we incorporate the correlation structure between \( Y \) and \( X \) and weigh \( D(\epsilon) \) and \( C(\epsilon) \) differently together with \( T_1(\epsilon) \) to form new statistics with a fixed \( \epsilon > 0 \).

Let us assume \( Y \) and \( X \) are positively correlated and weight the "standardized number of inversions" twice (say) as much as the "standardized number of matched pairs" to form a new statistic \( S(\epsilon) \):

\[
S(\epsilon) = T_1(\epsilon) + 2 \ D(\epsilon)
\]

\[
= \frac{1}{n_1 n_0} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_0} \left( \text{sgn}(Y_{1j} - Y_{0j'}) I\{|X_{1j} - X_{0j'}| \leq \epsilon\} + [\text{sgn}(Y_{1j} - Y_{0j'}) - \text{sgn}(X_{1j} - X_{0j'})] I\{|X_{1j} - X_{0j'}| > \epsilon\} \right)
\]

\[
= \frac{1}{n_1 n_0} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_0} (\text{sgn}(Y_{1j} - Y_{0j'}) - \text{sgn}(X_{1j} - X_{0j'}) I\{|X_{1j} - X_{0j'}| > \epsilon\})
\]

Define

\[
\phi (u; \epsilon) = \frac{1}{2} \text{sgn}(u) I\{|u| > \epsilon\},
\]

then

\[
S(\epsilon) = \frac{2}{n_1 n_0} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_0} \{ \phi(Y_{1j} - Y_{0j'}; 0) - \phi(X_{1j} - X_{0j'}; \epsilon) \}.
\]

Note that when \( \epsilon = 0 \),

\[
S(0) = \frac{2}{n_1 n_0} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_0} \{ \phi(Y_{1j} - Y_{0j'}; 0) - \phi(X_{1j} - X_{0j'}; 0) \} = 2 \sum_{j=1}^{n_1} (R_{1j} - C_{1j}),
\]

where

\[
R_{1j} = \sum_{j'=1}^{n_0} \phi(Y_{1j} - Y_{0j'}; 0) + \sum_{k=1}^{n_1} \phi(Y_{1j} - Y_{1k}; 0),
\]

and

\[
C_{1j} = \sum_{j'=1}^{n_0} \phi(X_{1j} - X_{0j'}; 0) + \sum_{k=1}^{n_1} \phi(X_{1j} - X_{1k}; 0),
\]

are the ranks (corrected for the mean) of \( Y_{1j} \) and \( X_{1j} \) respectively (\( \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \phi(Y_{1j} - Y_{1k}; 0) = \sum_{j=1}^{n_1} \sum_{k=1}^{n_1} \phi(X_{1j} - X_{1k}; 0) = 0 \)). In this situation \( n_1 n_0 S(\epsilon) \) is exactly equal to twice the statistic \( Z(1) \) in "rank analysis of covariance" (Quade, 1967). Thus, with tolerance \( \epsilon > 0 \), one can write
\[ S(\varepsilon) = \frac{2}{n_1 n_0} \sum_{j=1}^{n_1} (Rr_{1j} - Cr_{1j}(\varepsilon)) \],
where \( Cr_{1j}(\varepsilon) = \sum_{j'=1}^{n_0} \phi(X_{1j} - X_{0j'}; \varepsilon) \). The statistic \( Cr_{1j}(\varepsilon) \) can therefore be treated as an extended rank with those pairs of \( X \)'s falling within \( \varepsilon \) as ties. We discuss the extended ranks and other properties in the next section. Similarly to that paper, which stated that optimal efficiency can be achieved by appropriate \( \lambda \), where

\[ Z(\lambda) = \sum_{j=1}^{n_1} (Rr_{1j} - \lambda Cr_{1j}) \],

we can follow the same technique to find the optimal weight \( \omega \) in the statistic

\[ S(\omega; \varepsilon) = M(\varepsilon) + \omega D(\varepsilon); \quad (2.24) \]

note that the statistic \( S(\varepsilon) \) already discussed is now the special case \( S(2; \varepsilon) \). Alternatively, consider

\[ Z(\omega; \varepsilon) = S(1+\omega; \varepsilon) + (1-\omega) C(\varepsilon) \quad (2.25) \]

\[ = \frac{1}{n_1 n_0} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_0} (\text{sgn}(Y_{1j} - Y_{0j'}) - \omega \text{sgn}(X_{1j} - X_{0j'})) I\{|X_{1j} - X_{0j'}| > \varepsilon\}) \]

\[ = \frac{2}{n_1 n_0} \sum_{j=1}^{n_1} (Rr_{1j} - \omega Cr_{1j}(\varepsilon)); \]

note that \( Z(1; \varepsilon) \) is also \( S(\varepsilon) \). Thus, by further adding the "standardized number of concordant pairs" to \( S(\omega; \varepsilon) \), one can treat \( Z(\omega; \varepsilon) \) as a statistic of generalized rank analysis of covariance with weight \( \omega \). Partitioning the sums of squares as in the analysis of variance is then possible after "adjustment" of the matched pairs. For research interest, one can further generalize by finding a second weight parameter \( \omega_2 \) so that

\[ Z(\omega_1, \omega_2; \varepsilon) = S(\omega_1; \varepsilon) + \omega_2 C(\varepsilon) = M(\varepsilon) + \omega_1 D(\varepsilon) + \omega_2 C(\varepsilon) \quad (2.26) \]

might have even better asymptotic efficiency than \( Z(\omega; \varepsilon) \) does.

For the k-sample case, we can define a general score that incorporates the comparison of observations from the k samples at the same time. First we need the following definition of extended ranks given tolerance \( \varepsilon \).
Definition 2.3.2

Let \((Y_{ij}, X_{ij})\) \(i = 1, ..., k; j = 1, ..., n_i\) be the \(j\)-th observation in the \(i\)-th group for the response \(Y\) and the covariate \(X\). Then we say \(Cr_{ij}(\epsilon)\) is an extended rank (mean adjusted) of \(X_{ij}\) corresponding to tolerance \(\epsilon\) if

\[
Cr_{ij}(\epsilon) = \frac{1}{2} \sum_{i'} \sum_{j'} \text{sgn}(X_{ij} - X_{i'j'}) I(\{\mid X_{ij} - X_{i'j'} \mid > \epsilon\})
= \sum_{i'} \sum_{j'} \phi(X_{ij} - X_{i'j}; \epsilon) .
\]

One can then define the score corresponding to \((Y_{ij}, X_{ij})\) to be

\[
Z_{ij}(\omega; \epsilon) = Rr_{ij} - \omega Cr_{ij}(\epsilon)
= \sum_{i'} \sum_{j'} \psi(\{ Y_{ij}, X_{ij} \}, \{ Y_{i'j'}, X_{i'j'} \}, \omega; \epsilon )
= \sum_{i'} \sum_{j'} \{ \phi(Y_{ij} - Y_{i'j}; 0) - \omega \phi(X_{ij} - X_{i'j}; \epsilon) \}
= \frac{1}{2} \sum_{i'} \sum_{j'} \{ \text{sgn}(Y_{ij} - Y_{i'j'}) - \text{sgn}(X_{ij} - X_{i'j'}) \} I(\{\mid X_{ij} - X_{i'j'} \mid > \epsilon\})
+ (1 + \omega) \frac{1}{2} [ \text{sgn}(Y_{ij} - Y_{i'j'}) - \text{sgn}(X_{ij} - X_{i'j'}) ] I(\{\mid X_{ij} - X_{i'j'} \mid > \epsilon\})
+ (1 - \omega) \frac{1}{2} [ \text{sgn}(Y_{ij} - Y_{i'j'}) + \text{sgn}(X_{ij} - X_{i'j'}) ] I(\{\mid X_{ij} - X_{i'j'} \mid > \epsilon\}) ,
\]

where \(Rr_{ij}\) and \(Cr_{ij}(\epsilon)\) are the corrected ranks of the response variable \(Y\) and the covariate \(X\) adjusted for the tolerance \(\epsilon\) respectively for the \(j\)-th subject in the \(i\)-th treatment group. The summation includes all possible combinations of \((i', j')\) other than \((i, j)\). Similarly, one can define

\[
Z_{ij}^*(\omega_1, \omega_2; \epsilon) = \frac{1}{2} \sum_{i'} \sum_{j'} \{ \text{sgn}(Y_{ij} - Y_{i'j'}) \} I(\{\mid X_{ij} - X_{i'j'} \mid \leq \epsilon\})
+ \omega_1 \frac{1}{2} [ \text{sgn}(Y_{ij} - Y_{i'j'}) - \text{sgn}(X_{ij} - X_{i'j'}) ] I(\{\mid X_{ij} - X_{i'j'} \mid > \epsilon\})
+ \omega_2 \frac{1}{2} [ \text{sgn}(Y_{ij} - Y_{i'j'}) + \text{sgn}(X_{ij} - X_{i'j'}) ] I(\{\mid X_{ij} - X_{i'j'} \mid > \epsilon\}) \right) .
\]

Definition 2.3.3

We define the variance ratio (VR) of general scores \(Z_{ij}\) corresponding to the \(j\)-th observation in the \(i\)-th group to be
\begin{equation}
VR = \frac{(N - k) \sum_i (\sum_j Z_{ij})^2 / n_i}{(k - 1) \left[ \sum_i \sum_j Z_{ij}^2 - \sum_i (\sum_j Z_{ij})^2 / n_i \right]},
\end{equation}

where \( N = \sum_i n_i \).

When there are \( p \) covariates \( X = (X_1, X_2, \ldots, X_p)' \), we can define a general score adjusting for these covariates and the corresponding variance ratio as follows:

**Definition 2.3.4**

Let \( X = (X_1, X_2, \ldots, X_p)' \) be the vector of \( p \)-covariates with \( \epsilon_l \) the tolerance corresponding to the \( l \)-th covariate \( X_l \), \( 1 \leq l \leq p \). Then we define scores

\begin{equation}
Z_{ij}(\lambda; \epsilon) = Rr_{ij} - \sum_{l=1}^p \lambda_l Cr_{ij}(\epsilon_l),
\end{equation}

where

\[ \lambda = (\lambda_1, \ldots, \lambda_p)' \], \[ \epsilon = (\epsilon_1, \ldots, \epsilon_p)' \],

\( Rr_{ij} \) = mean adjusted rank of the response \( Y \) for the \( j \)-th subject in the \( i \)-th sample, and

\( Cr_{ij}(\epsilon_l) \) = mean adjusted rank of the covariable \( X_l \) corresponding to tolerance \( \epsilon_l \) for the \( j \)-th subject of the \( i \)-th sample.

The corresponding variance ratio of \( Z_{ij}(\lambda; \epsilon) \) is therefore

\begin{equation}
VR(\lambda; \epsilon) = \frac{(N - k) \sum_i (\sum_j Z_{ij}(\lambda; \epsilon))^2 / n_i}{(k - 1) \left[ \sum_i \sum_j Z_{ij}^2(\lambda; \epsilon) - \sum_i (\sum_j Z_{ij}(\lambda; \epsilon))^2 / n_i \right]}.
\end{equation}

**Definition 2.3.5**

We define the extended rank and matched-pair (ERMP) analysis of covariance based on the variance ratio \( VR(\lambda; \epsilon) \) to be \( ERMP(\lambda; \epsilon) \).

In general, a \( \chi^2 \)-statistic for treatment differences would be sufficient instead of the variance ratio statistic defined previously, which is easier to get and has the same efficiency as the \( VR \) test statistic. We derive our results, however, based on the \( VR \) statistic since the current work can be treated as a direct extension of that by Quade (1966, 1967a). Therefore, results already developed there can be applied directly with slight modifications.

Both \( Z_{ij}(\omega; \epsilon) \) and \( Z_{ij}^*(\omega_1, \omega_2; \epsilon) \) are interchangeable random variables, hence the results on
variance ratios based on general scores in Quade (1966) can be applied to show that both the variance ratios of the scores defined above have asymptotic F distributions. We then investigate the ARE of the ERMP(λ; ε) with respect to the Kruskal-Wallis k-sample statistic; and explore the ARE of the ERMP(λ; ε) with respect to category matching. Also, the ARE of VR(ω; ε) and VR(ω₁, ω₂; ε) to find the optimal weight(s) is of interest.

Though regression curves to describe the relationship between Y and X are justifiable most of the time, the strictly monotonic assumption does not hold in general. In Chapter 5, we propose a notion of "locally uncorrelated" when the latter condition not necessarily hold, and find a reasonable tolerance for matching.

We show the superiority of the ERMP(λ; ε) compared to the existing matched analysis methods by examples and make suggestions for further research in the last chapter. Computer programs for calculating the statistic are attached as an appendix.

2.4 Extended Ranks and the Relationship Between the Generalized Scores and the Matched Pairs

We illustrate the extended ranks defined in the last section by a small example shown in Table 1. Suppose there is one treatment, with n₁=4 observations, and there are n₀=4 controls. We show the ranks Rᵢⱼ (adjusted by mean) of the response Y and the extended ranks Crᵢⱼ(ε) (adjusted by mean) of the covariate X as a function of given ε in the combined sample. Note that in this example, all the extended ranks of the covariates X become 0 when tolerance is 10. This says that with tolerance this scale there is essentially no difference for the existence of the covariates X to the response Y since all the measurements of X fall within this category. Also, the variation of the extended ranks decreases when tolerance increases.
Table 2.4.1
Small Example to Illustrate The Extended Ranks Given Tolerance \( \epsilon \)

<table>
<thead>
<tr>
<th>( j )</th>
<th>( Y_{ij} )</th>
<th>( X_{ij} )</th>
<th>( R_{rij} )</th>
<th>( Cr_{ij}(0) )</th>
<th>( Cr_{ij}(2) )</th>
<th>( Cr_{ij}(5) )</th>
<th>( Cr_{ij}(10) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control obs. (i=0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>7</td>
<td>-7</td>
<td>-5</td>
<td>-4</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>10</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>9</td>
<td>-5</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>11</td>
<td>-1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Treated obs. (i=1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>14</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>8</td>
<td>-3</td>
<td>-3</td>
<td>-3</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>13</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>6</td>
<td>3</td>
<td>-7</td>
<td>-5</td>
<td>-2</td>
<td>0</td>
</tr>
</tbody>
</table>

* All the ranks displayed in the table are multiplied by 2.

To see that the ERMP ANOCOVA is really an extension of both the rank ANOCOVA (Quade 1967) and the matched pair ANOCOVA (Quade 1982), we show that the kernels of the U-statistics proposed in the last section can all be represented by some combinations of \( M(\epsilon) \), \( D(\epsilon) \), and \( C(\epsilon) \) defined previously. Note that all the kernels are of the form

\[
(2.33) \quad M(\epsilon) + \lambda_1 D(\epsilon) + \lambda_2 C(\epsilon) .
\]

One can therefore treat all of these U-statistics as members of a family. We will examine the ARE properties, however, only for the statistic \( S(\epsilon) \) for the two sample case, and the statistics \( Z_{ij}(\omega; \lambda) \) and \( Z_{ij}^*(\omega_1; \omega_2; \epsilon) \) for the k-sample and p-covariates case in the following chapters when convenient. We summarize the results in Table 2.4.2.
Table 2.4.2
Kernels of two-sample U-statistics in Section 2.3 as combinations of $M(\epsilon)$, $D(\epsilon)$, and $C(\epsilon)$

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U</td>
<td>$M(\epsilon) + D(\epsilon) + C(\epsilon)$</td>
</tr>
<tr>
<td>$S(\epsilon)$</td>
<td>$M(\epsilon) + 2\ D(\epsilon)$</td>
</tr>
<tr>
<td>$S(\omega; \epsilon)$</td>
<td>$M(\epsilon) + \omega\ D(\epsilon)$</td>
</tr>
<tr>
<td>$Z(\omega; \epsilon)$</td>
<td>$M(\epsilon) + (1+\omega)\ D(\epsilon) + (1-\omega)\ C(\epsilon)$</td>
</tr>
<tr>
<td>$Z(\omega_1, \omega_2; \epsilon)$</td>
<td>$M(\epsilon) + \omega_1\ D(\epsilon) + \omega_2\ C(\epsilon)$</td>
</tr>
</tbody>
</table>

2.5 Relationship Between the Coefficients of the Rank Adjusted Scores and the Weights of Pairs

We give the relationship between the coefficients of the extended ranks of the covariates and the weights of the pairs (matched or unmatched) in the following proposition.

Proposition 2.5.1

(2.34)

$$Z_{ij}(\lambda; \epsilon) = R_{ij} - \sum_{l=1}^{p} \lambda_l C_{i,j}(\epsilon_l)$$

$$= \sum_{i'} \sum_{j'} \left\{ \phi(Y_{ij} - Y_{i'j'} ; 0) - \sum_{l=1}^{p} \lambda_l \phi(X_{ij}^{(l)} - X_{i'j'}^{(l)} ; \epsilon_l) \right\}$$

$$= \frac{1}{2} \sum_{i'} \sum_{j'} \sum_{l=1}^{p} (1 - \sum_{l=1}^{p} \lambda_l s_l) \left[ I\{Y_{ij} > Y_{i'j'}\} \prod_{l=1}^{p} \psi(X_{ij}^{(l)}, X_{i'j'}^{(l)}; +s_l) - I\{Y_{ij} < Y_{i'j'}\} \prod_{l=1}^{p} \psi(X_{ij}^{(l)}, X_{i'j'}^{(l)}; -s_l) \right],$$

where

$$\sum_{p} s_p = -1, s_2 = -1, \ldots, s_p = -1,$$

and

24
\[ \psi(X_{ij}^{(l)}, X_{ij'}^{(l)}, s_l) = \begin{cases} 
I( | X_{ij}^{(l)} - X_{ij'}^{(l)} | \leq \epsilon ) & \text{if } s_l = 0 \\
I( X_{ij}^{(l)} > X_{ij'}^{(l)} + \epsilon ) & \text{if } s_l = 1 \\
I( X_{ij}^{(l)} < X_{ij'}^{(l)} - \epsilon ) & \text{if } s_l = -1 
\end{cases} \]

Proof:

One can prove this by induction.

Case p=2:

\[
Z_{ij}(\lambda_1, \lambda_2; \epsilon_1, \epsilon_2) = \sum_{i'} \sum_{j'} \phi(Y_{ij} - Y_{i'j'}; 0) - \lambda_1 \phi(X_{ij}^{(1)} - X_{ij'}^{(1)}; \epsilon_1) - \lambda_2 \phi(X_{ij}^{(2)} - X_{ij'}^{(2)}; \epsilon_2)
\]

\[
= \frac{1}{2} \sum_{i'} \sum_{j'} ( \text{sgn}(Y_{ij} - Y_{i'j'}) - \lambda_1 \text{sgn}(X_{ij}^{(1)} - X_{ij'}^{(1)})[1 - I(| X_{ij}^{(1)} - X_{ij'}^{(1)} | \leq \epsilon_1)] \\
- \lambda_2 \text{sgn}(X_{ij}^{(2)} - X_{ij'}^{(2)})[1 - I(| X_{ij}^{(2)} - X_{ij'}^{(2)} | \leq \epsilon_2)])
\]

\[
= \frac{1}{2} \sum_{i'} \sum_{j'} \left\{ (1 - \lambda_1 - \lambda_2)[ I(Y_{ij} > Y_{i'j'}) \cdot I(X_{ij}^{(1)} > X_{ij'}^{(1)} + \epsilon_1) \cdot I(X_{ij}^{(2)} > X_{ij'}^{(2)} + \epsilon_2) \\
- I(Y_{ij} < Y_{i'j'}) \cdot I(X_{ij}^{(1)} < X_{ij'}^{(1)} - \epsilon_1) \cdot I(X_{ij}^{(2)} < X_{ij'}^{(2)} - \epsilon_2)] \\
+ (1 - \lambda_1) \cdot I(Y_{ij} > Y_{i'j'}) \cdot I(X_{ij}^{(1)} > X_{ij'}^{(1)} + \epsilon_1) \cdot I(| X_{ij}^{(2)} - X_{ij'}^{(2)} | \leq \epsilon_2) \\
- I(Y_{ij} < Y_{i'j'}) \cdot I(X_{ij}^{(1)} < X_{ij'}^{(1)} - \epsilon_1) \cdot I(| X_{ij}^{(2)} - X_{ij'}^{(2)} | \leq \epsilon_2)] \\
+ (1 - \lambda_1 + \lambda_2) \cdot I(Y_{ij} > Y_{i'j'}) \cdot I(X_{ij}^{(1)} > X_{ij'}^{(1)} + \epsilon_1) \cdot I(X_{ij}^{(2)} < X_{ij'}^{(2)} - \epsilon_2) \\
- I(Y_{ij} < Y_{i'j'}) \cdot I(X_{ij}^{(1)} < X_{ij'}^{(1)} - \epsilon_1) \cdot I(X_{ij}^{(2)} > X_{ij'}^{(2)} + \epsilon_2)] \\
\right\}
\]

25
\[+(1+\lambda_2) \left[ I\{(Y_{ij}>Y_{ij'}) \mid X_{ij}(1)-X_{ij'}(1) \leq \epsilon_1 \} I\{X_{ij}(2)<X_{ij'}(2)-\epsilon_2\}ight. \\
\left. - I\{Y_{ij}<Y_{ij'}\} I\{X_{ij}(1)-X_{ij'}(1) \leq \epsilon_1 \} I\{X_{ij}(2)>X_{ij'}(2)+\epsilon_2\}\right] \\
+(1+\lambda_1 - \lambda_2) \left[ I\{(Y_{ij}>Y_{ij'}) \mid X_{ij}(1)<X_{ij'}(1)-\epsilon_1 \} I\{X_{ij}(2)>X_{ij'}(2)+\epsilon_2\}\right. \\
\left. - I\{Y_{ij}<Y_{ij'}\} I\{X_{ij}(1)>X_{ij'}(1)+\epsilon_1 \} I\{X_{ij}(2)<X_{ij'}(2)-\epsilon_2\}\right] \\
+(1+\lambda_1) \left[ I\{(Y_{ij}>Y_{ij'}) \mid X_{ij}(1)<X_{ij'}(1)-\epsilon_1 \} I\{X_{ij}(2)-X_{ij'}(2) \leq \epsilon_2\}\right. \\
\left. - I\{Y_{ij}<Y_{ij'}\} I\{X_{ij}(1)>X_{ij'}(1)+\epsilon_1 \} I\{X_{ij}(2)-X_{ij'}(2) \leq \epsilon_2\}\right] \\
+(1+\lambda_1+\lambda_2) \left[ I\{(Y_{ij}>Y_{ij'}) \mid X_{ij}(1)<X_{ij'}(1)-\epsilon_1 \} I\{X_{ij}(2)<X_{ij'}(2)-\epsilon_2\}\right. \\
\left. - I\{Y_{ij}<Y_{ij'}\} I\{X_{ij}(1)>X_{ij'}(1)+\epsilon_1 \} I\{X_{ij}(2)>X_{ij'}(2)+\epsilon_2\}\right]\]

\[= \frac{1}{2} \sum_{i' j'} \sum_{j'} \sum_{l=1}^{2^*} \left(1-\sum_{l=1}^{\lambda} \lambda_{1} s_{l}\right) \left[ I\{Y_{ij}>Y_{ij'}\} \prod_{l=1}^{2} \psi(X_{ij}(l), X_{ij'}(l); +s_{l}) \right. \\
\left. - I\{Y_{ij}<Y_{ij'}\} \prod_{l=1}^{2} \psi(X_{ij}(l), X_{ij'}(l); -s_{l}) \right].\]

Suppose this is true for \( p-1 \) covariables, where \( p \geq 3 \); then

\[Z_{ij}(\lambda_1, \ldots, \lambda_{p-1}, \lambda_p; \epsilon_1, \ldots, \epsilon_{p-1}, \epsilon_p) = \sum_{i' j'} \left\{ \phi(Y_{ij}-Y_{i'j'}; 0) - \sum_{l=1}^{p-1} \lambda_{l} \phi(X_{ij}(l)-X_{i'j'}(l); \epsilon_l) \right\} - \lambda_p \phi(X_{ij}(p)-X_{i'j'}(p); \epsilon_p) \}

\[= \frac{1}{2} \sum_{i' j'} \sum_{j'} \sum_{l=1}^{p-1} \lambda_{l} s_{l} \left[ I\{Y_{ij}>Y_{ij'}\} \prod_{l=1}^{p-1} \psi(X_{ij}(l), X_{ij'}(l); +s_{l}) \right. \\
\left. - I\{Y_{ij}<Y_{ij'}\} \prod_{l=1}^{p-1} \psi(X_{ij}(l), X_{ij'}(l); -s_{l}) \right] \\
- \frac{1}{2} \sum_{i'} \sum_{j'} \lambda_p \left[ I\{X_{ij}(p)>X_{i'j'}(p)+\epsilon_p\} - I\{X_{ij}(p)<X_{i'j'}(p)-\epsilon_p\} \right] \\
\left. \frac{1}{2} \sum_{i' j'} \sum_{j'} \sum_{l=1}^{p-1} \lambda_{l} s_{l} - \lambda_p \left[ I\{Y_{ij}>Y_{ij'}\} \prod_{l=1}^{p-1} \psi(X_{ij}(l), X_{ij'}(l); +s_{l}) \right. \\
\left. - I\{X_{ij}(p)>X_{i'j'}(p)+\epsilon_p\} \right] \\
+ \frac{1}{2} \sum_{i'} \sum_{j'} \sum_{l=1}^{p-1} \lambda_{l} s_{l} + \lambda_p \left[ I\{X_{ij}(p)>X_{i'j'}(p)+\epsilon_p\} \right].\]
\[ -I\{Y_{ij} < Y_{i'j'}\} \prod_{l=1}^{p-1} \psi(X_{ij}^{(l)}, X_{i'j'}^{(l)}; -s_l) I\{X_{ij}^{(p)} < X_{i'j'}^{(p)} - \epsilon_p\} \]
\[ + \frac{1}{2} \sum_{i'} \sum_{j'} \sum_{p=1}^{(p-1)^*} (1 - \sum_{l=1}^{p-1} \lambda_l s_l) [I\{Y_{ij} > Y_{i'j'}\} \prod_{l=1}^{p-1} \psi(X_{ij}^{(l)}, X_{i'j'}^{(l)}; +s_l) \]
\[ I\{|X_{ij}^{(p)} - X_{i'j'}^{(p)}| \leq \epsilon_p\} \]
\[ -I\{Y_{ij} < Y_{i'j'}\} \prod_{l=1}^{p-1} \psi(X_{ij}^{(l)}, X_{i'j'}^{(l)}; -s_l) I\{|X_{ij}^{(p)} - X_{i'j'}^{(p)}| \leq \epsilon_p\} \]
\[ + \frac{1}{2} \sum_{i'} \sum_{j'} \sum_{p=1}^{(p-1)^*} (1 - \sum_{l=1}^{p-1} \lambda_l s_l + \lambda_p) [I\{Y_{ij} > Y_{i'j'}\} \prod_{l=1}^{p-1} \psi(X_{ij}^{(l)}, X_{i'j'}^{(l)}; +s_l) \]
\[ I\{X_{ij}^{(p)} < X_{i'j'}^{(p)} - \epsilon_p\} \]
\[ -I\{Y_{ij} < Y_{i'j'}\} \prod_{l=1}^{p-1} \psi(X_{ij}^{(l)}, X_{i'j'}^{(l)}; -s_l) I\{X_{ij}^{(p)} > X_{i'j'}^{(p)} + \epsilon_p\} \]
\[ = \frac{1}{2} \sum_{i'} \sum_{j'} \sum_{p=1}^{p} (1 - \sum_{l=1}^{p} \lambda_l s_l) [I\{Y_{ij} > Y_{i'j'}\} \prod_{l=1}^{p} \psi(X_{ij}^{(l)}, X_{i'j'}^{(l)}; +s_l) - \]
\[ I\{Y_{ij} < Y_{i'j'}\} \prod_{l=1}^{p} \psi(X_{ij}^{(l)}, X_{i'j'}^{(l)}; -s_l) \].

Hence, the proposition holds. \(\square\)

As an example of this proposition, let \(p=1\) and \(\lambda_1=0.5\). Then for a given tolerance \(\epsilon\), we have

\[ Z_{ij}(\lambda_1; \epsilon) = R_{ij} - 0.5 C_{ij}(\epsilon) \]
\[ = \frac{1}{2} \sum_{i'} \sum_{j'} \sum_{l=1}^{1} (1 - 0.5 s_l) \{I\{Y_{ij} > Y_{i'j'}\} \psi(X_{ij}^{(1)}, X_{i'j'}^{(1)}; +s_l) - \]
\[ I\{Y_{ij} < Y_{i'j'}\} \psi(X_{ij}^{(1)}, X_{i'j'}^{(1)}; -s_l)\} \]
\[ = \frac{1}{2} \sum_{i'} \sum_{j'} (1.5 [I\{Y_{ij} > Y_{i'j'}, X_{ij}^{(1)} < X_{i'j'}^{(1)} - \epsilon\} - I\{Y_{ij} < Y_{i'j'}, X_{ij}^{(1)} > X_{i'j'}^{(1)} + \epsilon\} ] - \]
\[ [I\{Y_{ij} > Y_{i'j'}, X_{ij}^{(1)} - X_{i'j'}^{(1)} \leq \epsilon\} - I\{Y_{ij} < Y_{i'j'}, X_{ij}^{(1)} - X_{i'j'}^{(1)} \leq \epsilon\} ] + \]
\[ 0.5 [I\{Y_{ij} > Y_{i'j'}, X_{ij}^{(1)} > X_{i'j'}^{(1)} + \epsilon\} - I\{Y_{ij} < Y_{i'j'}, X_{ij}^{(1)} < X_{i'j'}^{(1)} - \epsilon\} ] \].

Hence, in addition to putting weight=1 on the "weighted number of matched pairs", we put weight=1.5 on the "weighted number of inversions" and weight=0.5 on the "weighted number of
concordant pairs" together to form the general score \( Z_{ij}(0.5; \epsilon) \) for the ERMP(0.5; \( \epsilon \)) ANOCOVA given \( \epsilon \). That is, we weigh the inversions three times as much as the concordant pairs, and 1.5 times as much as the matched pairs. We show later in Chapter 4, this is one typical case when \( Y \) and \( X \) are positively correlated though the weights may vary with the correlation between \( Y \) and \( X \). The case when \( p \geq 2 \) is similar but more complicated in form.

One important feature of this proposition is that it demonstrates the superiority of the ERMP to the conventional category matching. With \( p \) covariables to control at the same time, it is inevitable that there will be some empty cells for certain combinations of the covariates, which makes sampling process much more difficult and subsequent analysis complicated with inadequate observations. However, this is not a problem with the ERMP test, since it uses all possible combinations of pairs available and does not reduce the sample size. We show in Chapter 4 that the results derived there hold true when the total sample size is large enough.

2.6 Summary

In brief, in this thesis we extend the existing matched-pair analysis to a status more comparable with the standard ANOCOVA (parametric or nonparametric) under Assumptions A1 - A3 through completing the following steps:

1. Find the ARE of the "difference in probability" (caliper matching) to the Mann-Whitney statistic \( U \); and its ARE with respect to the standard category matching.
2. In the case of \( k \) samples, find the optimal weight \( \lambda \) for the ARE of the ERMP(\( \lambda; \epsilon \)) with respect to the \( k \)-sample Kruskal-Wallis statistic; and the optimal weights \( \omega_1 \) and \( \omega_2 \) of the scores \( Z_{ij}^*(\omega_1, \omega_2; \epsilon) \) for the ARE of the VR of \( Z_{ij}^*(\omega_1, \omega_2; \epsilon) \) to the ERMP(\( \omega; \epsilon \)). Also, we will derive the ARE of the ERMP(\( \tau; \epsilon \)) to the standard category matching, where \( \tau \) is the optimal weight with respect to simple random sampling in the two-sample case.
3. Develop test statistic for "local lack of correlation"; and find a reasonable matching interval.
4. Applications, examples, and computer programs.
5. Further research (\( k \)-sample ordered alternatives, dichotomous response, etc ...).

We develop each of the proposed steps into a subsequent chapter. However, each chapter is self-contained though they are closely related to each other.
CHAPTER 3
ASYMPTOTIC RELATIVE EFFICIENCIES OF CALIPER MATCHING
IN THE UNIVARIATE TWO-SAMPLE CASE

3.1 Introduction

In this chapter we establish the asymptotic relative efficiency of the matched pair analysis, using caliper matching, with respect to the Mann-Whitney statistic $U$ when comparing two completely randomized samples. The asymptotic relative efficiency with respect to category matching with matching intervals less than or equal to a given tolerance under Assumptions A1-A3 is also investigated. Unlike the Mann-Whitney statistic $U$, the matched-pair analysis compares experimental units in one sample to those in another sample only when they are matched. In the case of caliper matching, two subjects are called matched if the distance between their covariates is less than or equal to a tolerance $\epsilon$. The relative efficiency thus depends on the strength of the correlation between $Y$ and $X$ through the tolerance $\epsilon$ and the distribution of $X$.

It is assumed throughout that the marginal distribution of $X$ is the same in the two samples (i.e. they are concomitant), as in Quade (1982). Thus, rather than serving as a tool for removing bias due to different distributions of $X$ as in most observational studies, our primary purpose will be to increase the precision of treatment differences by controlling for $X$. The main application would be in carefully planned experiments where the distributions of possible confounding factors are as balanced as possible. This can be true in observational studies only when there is no question of bias resulting from the covariable $X$.

We extend the work of Schoenfelder (1981) when the covariable $X$ is discrete to the case of continuous $X$. Note that under Schoenfelder's framework, one can divide the support of a continuous covariable $X$ into adjacent intervals and assign a weight the reciprocal of the corresponding probability of $X$ within an interval to get an optimal AMP (All Matched Pair) analysis. We will, however, derive our results based on the caliper matching method of Quade (1982) without weighting schemes:

$$T_1(\epsilon) = \frac{1}{n_1n_0} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_0} \text{sgn}(Y_{1j} - Y_{0j'}) I\{|X_{1j} - X_{0j'}| \leq \epsilon\}.$$

We write $T_1$ as a function of tolerance $\epsilon$ since it depends on the choice of $\epsilon$. Note that $T_1(\infty) =$
\[
\frac{U_1 - U_0}{n_1^{1/2} n_0^{1/2}},
\]
where \(U_1\) and \(U_0\) are the ordinary Mann-Whitney statistics for comparing the treatment observations with the controls. Thus, \(T_1(\epsilon)\) is a natural extension of the Mann-Whitney statistic \(U\). The ARE of \(T_1(\epsilon)\) with respect to category matching, which is the AMP analysis with observations within the same interval considered matched, is given in Section 3.5.

As stated in Section 2.3, \(T_1(\epsilon)\) is a U-statistic with kernel

\[
\text{sgn}(y_1 - y_0) I\{ |x_1 - x_0| \leq \epsilon\},
\]

and degree \((1, 1)\). Its asymptotic normality for a given \(\epsilon\) can be derived by standard U-statistics theory. The following theorems stated in Section 2.2 will be needed in our approach.

### 3.2 Related Theorems

The asymptotic normality of U-statistics for the two sample case was established by the following theorem due to Lehmann (1951), an extension of Hoeffding's (1948) U-statistic theorem to two-sample statistics.

**Theorem 3.2.1 (Two-Sample U-statistic Theorem)**

Let \(X_1, \ldots, X_m\) and \(Y_1, \ldots, Y_n\) denote independent random samples from two populations. Let \(\gamma\) be an estimable parameter of degree \((r, s)\) with symmetric kernel \(h(.)\). If \(E[ h^2(X_1, \ldots, X_r; Y_1, \ldots, Y_s)] < \infty\), then

\[
\sqrt{N} [U(X_1, \ldots, X_m; Y_1, \ldots, Y_n) - \gamma]
\]

has a limiting normal distribution with mean zero and variance \(r^2 \zeta_{1,0}/\lambda + s^2 \zeta_{0,1}/(1 - \lambda)\), provided the variance is positive, where

\[
U(X_1, \ldots, X_m; Y_1, \ldots, Y_n) = \frac{1}{(m!)(n!)} \sum_{\alpha \in A} \sum_{\beta \in B} h(X_{\alpha_1}, \ldots, X_{\alpha_r}; Y_{\beta_1}, \ldots, Y_{\beta_s}),
\]

\(A(B)\) is the collection of all subsets of \((r,s)\) integers chosen without replacement from the integers \(\{1, \ldots, m\}(\{1, \ldots, n\})\). Also,

\[
\zeta_{1,0} = \text{Cov}(h(X_1, \ldots, X_r; Y_1, \ldots, Y_s), h(X_1, X_{r+1}, \ldots, X_{2r-1}; Y_{s+1}, \ldots, Y_{2s})),
\]

\[
\zeta_{0,1} = \text{Cov}(h(X_1, \ldots, X_r; Y_1, \ldots, Y_s), h(X_{r+1}, \ldots, X_{2r}; Y_1, Y_{s+1}, \ldots, Y_{2s-1})),
\]

\text{Cov}(h(X_1, \ldots, X_r; Y_1, \ldots, Y_s), h(X_1, X_{r+1}, \ldots, X_{2r-1}; Y_{s+1}, \ldots, Y_{2s})).
and \( \lambda = \lim_{N \to \infty} m/N \), \( N = m + n \).


After we have established the asymptotic normality of the two-sample U-statistics, the following theorems reported by Puri and Sen (1971) can then be applied to derive the ARE of the two respective U-statistics under investigation. We state the theorems in most practical applications when \( m = 1 \) and \( \delta = 1/2 \) there. Let \( T_N \) be a test for the hypothesis \( H_0: \theta = \theta_0 \) against the alternative \( H_1: \theta > \theta_0 \) based on the first \( N \) observations, and let the critical region be \( T_N \geq \lambda_N, \alpha \). Let further \( \theta_N \) be a sequence of alternatives such that

\[
\theta_N = \theta_0 + k/\sqrt{N}
\]

where \( k \) is some finite positive constant independent of \( N \). Then the following theorem gives the limiting power of the test statistic \( T_N \).

**Theorem 3.2.2** (Puri and Sen, 1971)

Suppose

(a) \( \lim_{N \to \infty} P_{\theta_0}(T_N \geq \lambda_N, \alpha) = \alpha \), where \( \alpha (0 < \alpha < 1) \) is a fixed value;

(b) there exist functions \( \mu_N(\theta) \) and \( \sigma_N(\theta) \) such that

\[
\lim_{N \to \infty} P\left\{ \frac{T_N - \mu_N(\theta)}{\sigma_N(\theta)} \leq x \right\} = \Psi(x)
\]

uniformly in \( \theta, \theta_0 \leq \theta \leq \theta_0 + \eta \) where \( \eta \) is any positive number and \( \Psi(x) \) is any distribution function;

(c) \( \mu_N'(\theta_0) > 0 \);

(d) \( \lim_{N \to \infty} \frac{\mu_N'(\theta_N)}{\mu_N'(\theta_0)} = 1 \), \( \lim_{N \to \infty} \frac{\sigma_N(\theta_N)}{\sigma_N(\theta_0)} = 1 \);

(e) \( \lim_{N \to \infty} N^{-1/2} \frac{\mu_N'(\theta_0)}{\sigma_N(\theta_0)} = c > 0 \).

Then, the limiting power of the sequence \( T_N \) is \( 1 - \Psi(\lambda_{\alpha} - kc) \).


The constant \( c \) defined by (e) of Theorem 3.2.2 is called the efficacy of the sequence \( T_N \), and abbreviated as \( \text{eff}(T(.)) \). The following theorem gives the asymptotic relative efficiency of two
tests as the ratio of their efficacies.

Theorem 3.2.3 (Puri and Sen, 1971)

If \( \{T^{(1)}\} \) and \( \{T^{(2)}\} \) satisfy the assumptions of Theorem 3.2.2 with the same distribution function \( \Psi(x) \) in each case, then the asymptotic relative efficiency of \( \{T^{(2)}\} \) with respect to \( \{T^{(1)}\} \) is

\[
\text{ARE}(T^{(2)}, T^{(1)}) = \lim_{N \to \infty} \left( \frac{\mu_2 N \theta_0^{(i)}}{\mu_1 N \theta_0^{(i)}} \right)^2 \left( \frac{\sigma_2 N \theta_0^{(i)}}{\sigma_1 N \theta_0^{(i)}} \right)^2 = \left( \frac{\text{eff}(T^{(2)})}{\text{eff}(T^{(1)})} \right)^2.
\]


3.3 Tolerance \( \epsilon \) of the Statistic \( T_1(\epsilon) \)

In matched analysis, we compare the responses of subjects which are matched with a certain tolerance, on the covariable \( X \). Under the null hypothesis,

\[
H_0 : F_1(y|x) = F_0(y|x),
\]

where \( F_1(y|x) \) and \( F_0(y|x) \) are the conditional distribution functions of the response \( Y \) given the covariable \( X \) of the treatment group and the control group respectively. We have

\[
P\{Y_1 > Y_0, |X_1 - X_0| \leq \epsilon\} = P\{Y_1 < Y_0, |X_1 - X_0| \leq \epsilon\}.
\]

Thus, since

\[
P\{Y_1 > Y_0, |X_1 - X_0| \leq \epsilon\} + P\{Y_1 < Y_0, |X_1 - X_0| \leq \epsilon\} = P\{|X_1 - X_0| \leq \epsilon\},
\]

we have

\[
P\{Y_1 > Y_0, |X_1 - X_0| \leq \epsilon\} = \frac{1}{2} P\{|X_1 - X_0| \leq \epsilon\},
\]

and

\[
P\{Y_1 > Y_0| |X_1 - X_0| \leq \epsilon\} = \frac{1}{2}.
\]

On the other hand, under \( H_0 \),

\[
P\{Y_1 > Y_0| X_1 = X_0\} = \int F_0(y|x) \, dF_1(y|x) = \int F_1(y|x) \, dF_1(y|x) = \frac{1}{2}.
\]

Therefore,

\[
P\{Y_1 > Y_0| |X_1 - X_0| \leq \epsilon\} = P\{Y_1 > Y_0| X_1 = X_0\} (= \frac{1}{2})
\]

under \( H_0 \). The scale of \( \epsilon \) is irrelevant here. Following the same argument under \( H_0 \), we have
(3.9) \[ P \{ Y_{11} > Y_{01}, Y_{11} > Y_{02} \mid |X_{11} - X_{01}| \leq \epsilon, |X_{11} - X_{02}| \leq \epsilon \} \]

\( = P \{ Y_{11} < Y_{01}, Y_{11} < Y_{02} \mid |X_{11} - X_{01}| \leq \epsilon, |X_{11} - X_{02}| \leq \epsilon \} \]

\( = P \{ Y_{11} > Y_{01}, Y_{11} > Y_{02} \mid X_{11} = X_{01} = X_{02} \} = \frac{1}{3} \]

(3.10) \[ P \{ Y_{11} > Y_{01}, Y_{11} < Y_{02} \mid |X_{11} - X_{01}| \leq \epsilon, |X_{11} - X_{02}| \leq \epsilon \} \]

\( = P \{ Y_{11} < Y_{01}, Y_{11} > Y_{02} \mid |X_{11} - X_{01}| \leq \epsilon, |X_{11} - X_{02}| \leq \epsilon \} \]

\( = P \{ Y_{11} > Y_{01}, Y_{11} < Y_{02} \mid X_{11} = X_{01} = X_{02} \} = \frac{1}{6} \]

Hence, whatever the \( \epsilon \) is, the result will not be biased applying \( T_1(\epsilon) \) in testing the null hypothesis. The main interest in choosing \( \epsilon \) will therefore be in the test efficiency.

3.4 Asymptotic Normality of the Statistic \( T_1(\epsilon) \)

We establish the asymptotic normality of \( T_1(\epsilon) \) given \( \epsilon \) in this section. Let

(3.11) \[ \pi_{10} = P \{ Y_1 > Y_0 \mid X_1 = X_0 \} - P \{ Y_1 < Y_0 \mid X_1 = X_0 \} . \]

Then under the null hypothesis \( H_0 \), \( \pi_{10} = 0 \). Suppose for a positive constant \( \theta > 0 \), the alternative hypothesis is true,

(3.12) \[ H_a : F_1(y - \theta \mid x) \equiv F_0(y \mid x) \equiv F(y \mid x) . \]

Denote the kernel of \( T_{1,N}(\epsilon) \) by

(3.13) \[ h_{T_{1,N}}((y_1, x_1), (y_0, x_0); \epsilon) = \text{sgn}(y_1 - y_0) I\{ |x_1 - x_0| \leq \epsilon \}. \]

Then

(3.14) \[ E(h_{T_{1,N}}((Y_{11}, X_{11}), (Y_{01}, X_{01}); \epsilon)) = E(|\text{sgn}(Y_{11} - Y_{01}) I\{ |X_{11} - X_{01}| \leq \epsilon \}|) < 1, \]

where the subscript \( N \) denotes that the underlying distribution may also depend on \( N \). Let

(3.15) \[ \gamma_{T_{1,N}}(\theta; \epsilon) = E(T_{1,N}(\epsilon)) = E(h_{T_{1,N}}((Y_1, X_1), (Y_0, X_0); \epsilon)), \]
where $\theta$ and $\epsilon$ denote that the underlying distribution depends on these two parameters. Denote by
\[
\psi_{1, T_1, N}((y_{1}, x_{1}); \epsilon) = E[ h_{T_1, N}((y_{1}, x_{1}), (Y_{0}, X_{0}); \epsilon) ] - \gamma_{T_1, N}(\theta; \epsilon),
\]
and
\[
\psi_{0, T_1, N}((y_{0}, x_{0}); \epsilon) = E[ h_{T_1, N}((Y_{1}, X_{1}), (y_{0}, x_{0}); \epsilon) ] - \gamma_{T_1, N}(\theta; \epsilon).
\]
Also,
\[
V_{T_1, N}^*(\epsilon) = \frac{1}{n_{1}} \sum_{j=1}^{n_{1}} \psi_{1, T_1, N}((y_{1j}, x_{1j}); \epsilon) + \frac{1}{n_{0}} \sum_{j'=1}^{n_{0}} \psi_{0, T_1, N}((y_{0j'}, x_{0j'}); \epsilon),
\]
\[
\xi_{1, T_1, N}(\theta; \epsilon) = E( \psi_{1, T_1, N}^2((Y_{1}, X_{1}); \epsilon) ),
\]
and
\[
\xi_{0, T_1, N}(\theta; \epsilon) = E( \psi_{0, T_1, N}^2((Y_{0}, X_{0}); \epsilon) ).
\]
Then we have
\[
\text{Var}(V_{T_1, N}^*(\epsilon)) = \frac{1}{n_{1}} \xi_{1, T_1, N}(\theta; \epsilon) + \frac{1}{n_{0}} \xi_{0, T_1, N}(\theta; \epsilon),
\]
since $(Y_{11}, X_{11}), \ldots, (Y_{1n_1}, X_{1n_1})$ and $(Y_{01}, X_{01}), \ldots, (Y_{0n_0}, X_{0n_0})$ are independent samples from $(Y_{1}, X_{1})$ and $(Y_{0}, X_{0})$ respectively. Let
\[
\tau_{T_1, N}(\theta; \epsilon) = N \text{Var}(V_{T_1, N}^*(\epsilon))
\]
we have the following result.

**Theorem 3.4.1**

Given a fixed $\epsilon > 0$, if the first derivative of the density function $g(x)$ of $X$ exists and is square integrable in a set of probability greater than zero, and the correlation between $Y$ and $X$ is less than one, then the sequence of test statistics
\[
\frac{T_{1, N}(\epsilon) - \gamma_{T_1, N}(\theta; \epsilon)}{\sqrt{\tau_{T_1, N}(\theta; \epsilon)/N}} \rightarrow N(0, 1)
\]
uniformly in $\theta$, $0 = \theta_0 \leq \theta \leq \theta_0 + \eta$, $\eta$ is a positive constant, as $N \rightarrow \infty$.

**Proof**: Since for any $\theta$ in a neighborhood of zero,
\[
|h_{T_1, N}((y_{1}, x_{1}), (y_{0}, x_{0}); \epsilon)| = |\text{sgn}(y_{1} - y_{0}) I\{ |x_{1} - x_{0}| \leq \epsilon \}| = |h_{T_1, N}^2((y_{1}, x_{1}), (y_{0}, x_{0}); \epsilon)| \leq 1,
\]
\[
\xi_{1, T_1, N}(\theta; \epsilon) = E[h_{T_1, N}((Y_{11}, X_{11}), (Y_{01}, X_{01}); \epsilon)h_{T_1, N}((Y_{11}, X_{11}), (Y_{02}, X_{02}); \epsilon)]
\]
\[
- E[h_{T_1, N}((Y_{11}, X_{11}), (Y_{01}, X_{01}); \epsilon)]^2
\]
\[
34
\]
\[
= \int (2F(y+\theta) - 1)^2 (G(x+\epsilon) + G(x-\epsilon) - 2G(x))^2 \, dH(y, x) \\
- \left( \int (2F(y+\theta) - 1) (G(x+\epsilon) + G(x-\epsilon) - 2G(x)) \, dH(y, x) \right)^2 \\
= \int (2F(y+\theta) - 1)^2 \left( \frac{\epsilon^2}{2} (g'(x^*) + g'(x^{**})) \right)^2 \, dH(y, x) \\
- \left( \int (2F(y+\theta) - 1) \left( \frac{\epsilon^2}{2} (g'(x^*) + g'(x^{**})) \right) \, dH(y, x) \right)^2 \\
= \frac{\epsilon^4}{4} V_1 > 0,
\]

where

\[V_1 = \int (2F(y+\theta) - 1)^2 (g'(x^*) + g'(x^{**})) \, dH(y, x) - \left( \int (2F(y+\theta) - 1)(g'(x^*) + g'(x^{**})) \, dH(y, x) \right)^2,
\]

\[H(y, x) \text{ is the joint distribution function of } Y \text{ and } X, \text{ and } x^* \text{ and } x^{**} \text{ are intermediate points between } (x, x+\epsilon) \text{ and } (x-\epsilon, x) \text{ respectively. Therefore, we can have } 0 < c_1(\epsilon) < \xi_{1,T_{1,N}}(\theta; \epsilon) < 4,
\]

where \(c_1(\epsilon) = \frac{1}{8} \epsilon^4 V_1\). Similarly, \(0 < c_2(\epsilon) < \xi_{0,1,T_{1,N}}(\theta; \epsilon) < 4\), we can have \(c_2(\epsilon) = \frac{1}{8} \epsilon^4 V_2\), where

\[V_2 = \int (2F(y-\theta) - 1)^2 (g'(x^*) + g'(x^{**})) \, dH(y, x) - \left( \int (2F(y-\theta) - 1)(g'(x^*) + g'(x^{**})) \, dH(y, x) \right)^2.
\]

Also,

\[E(|\psi_{1,T_{1,N}}((y_1, x_1); \epsilon)|^3) = E(||h_{T_{1,N}}((y_1, x_1), (Y_0, X_0); \epsilon)|| - \gamma_{T_{1,N}}(\theta; \epsilon)|^3 < 2^3 = 8,
\]

and

\[E(|\psi_{0,T_{1,N}}((y_1, x_1); \epsilon)|^3) < 8.
\]

We have by Hoeffding's result (see e.g. Puri and Sen, Theorem 3.2.2), (3.23) holds uniformly in \(\theta, 0 \leq \theta \leq \eta\). \(\square\)

3.5 The Asymptotic Relative Efficiency of \(T_1(\epsilon)\) With Respect to the Mann-Whitney Statistic

We are now ready to derive the main result of this chapter. We compare the relative efficiency of \(T_1(\epsilon)\) with respect to the Mann-Whitney statistic \(U\) through a sequence of local alternatives converging to the null in the manner of Pitman (1948). By fixing the same limiting significance level \(\alpha\), one can compare the two tests with respect to the sample sizes required to achieve a certain power. The smaller the sample size required, the higher the efficiency. Hence,
the efficiency index is the reciprocal of the ratio of the two required sample sizes. The main technique in deriving this index is Noether's (1955) theorem. We restate it as Theorems 3.2.2, 3.2.3 as from Puri and Sen (1971).

Recall Assumption A1: given the covariate $X$, we have the cumulative distribution functions of the control group and the treatment group $F_0(y|x)$ and $F_1(y|x)$ respectively. We are interested in testing the null hypothesis

$$H_0 : F_1(y|x) = F_0(y|x) = F(y|x)$$

versus the alternative hypothesis

$$H_a : F_1(y|x) < F_0(y|x) .$$

Assume that for each $N$ the alternative $H_a, N$ is true, where under $H_a, N$

$$F_1(y - \theta_N|x) = F_0(y|x) = F(y|x), \theta_N = \frac{\kappa}{N} + o\left(\frac{1}{\sqrt{N}}\right),$$

where $\kappa$ is a positive constant. The subscript $N$ denotes the dependence on the sample size $N$. We have the following result.

Lemma 3.5.1

Given a fixed $\epsilon > 0$, then the sequence of statistics $T_{1, N}(\epsilon)$ satisfies the conditions (a) -- (e) of Theorem 3.2.2 under Assumptions A1 -- A3, and the efficacy of $T_1(\epsilon)$ is

$$\text{eff}(T_1(\epsilon)) = \sqrt{12\lambda(1 - \lambda)} \int \int f(y|X_0) f(y|x_1) dG(x_1) dG(x_0) / \sqrt{P_2(\epsilon)} ,$$

where $P_2(\epsilon) = P\{ |X_1 - X_2| \leq \epsilon, |X_1 - X_3| \leq \epsilon\}, G(x)$ is the distribution function of $X$.

Proof: Following (3.23), we have

$$\frac{T_{1,N}(\epsilon) - \gamma_{T_{1,N}(\epsilon)} - \epsilon}{\sqrt{\gamma_{T_{1,N}(\epsilon)/N}}} \xrightarrow{N(0, 1)}$$

uniformly in $0 \leq \theta \leq \eta$ as $N \to \infty$, for a fixed $\epsilon > 0$. Thus, Conditions (a), (b) of Theorem 3.2.2 are satisfied.

$$\gamma_{T_{1,N}(\epsilon)} = P_N\{Y_1 > Y_0, |X_1 - X_0| \leq \epsilon\} - P_N\{Y_1 < Y_0, |X_1 - X_0| \leq \epsilon\}$$

$$= \int \int F(y + \theta_N|x_0) dF(y|x_1) dG(x_1) dG(x_0) -$$

$$\int \int F(y|x_0) dF(y|x_1) dG(x_1) dG(x_0)$$
\[
\int \int f(y - \theta_N | x_1) \, dF(y | x_0) \, dG(x_1) \, dG(x_0). 
\]

\[
\gamma_{T_1, N}^{\prime}(\theta_N; \epsilon) = \int \int f(y + \theta_N | x_0) \, f(y | x_1) \, d(y | x_1) \, dG(x_1) \, dG(x_0) + 
\int \int f(y - \theta_N | x_1) \, f(y | x_0) \, d(y | x_0) \, dG(x_1) \, dG(x_0).
\]

Therefore,

\[
(3.26) \quad \gamma_{T_1, N}^{\prime}(0; \epsilon) = 2 \int \int f(y | x_0) \, f(y | x_1) \, d(y | x_1) \, dG(x_1) \, dG(x_0) > 0,
\]

and

\[
(3.27) \quad \lim_{N \to \infty} \frac{\gamma_{T_1, N}^{\prime}(\theta_N; \epsilon)}{\gamma_{T_1, N}^{\prime}(0; \epsilon)} = 1.
\]

Following (3.21) and (3.22), we have

\[
(3.28) \quad \gamma_{T_1, N}(\theta_N; \epsilon) = N \text{Var}(V_{T_1, N}(\epsilon)) = N \left( \frac{1}{\nu_1} \xi_{1,T_1,N}(\theta_N; \epsilon) + \frac{1}{\nu_0} \xi_{0,T_1,N}(\theta_N; \epsilon) \right)
\]

\[
= N \left\{ \frac{1}{\nu_1} P_{(2)}(\epsilon) \{ P_N\{Y_{11}>Y_01, Y_{11}>Y_{02}, |X_{11}-X_{01}| \leq \epsilon, |X_{11}-X_{02}| \leq \epsilon \} - 
2 P_N\{Y_{11}>Y_01, Y_{11}<Y_{02}, |X_{11}-X_{01}| \leq \epsilon, |X_{11}-X_{02}| \leq \epsilon \} + 
P_N\{Y_{11}<Y_01, Y_{11}<Y_{02}, |X_{11}-X_{01}| \leq \epsilon, |X_{11}-X_{02}| \leq \epsilon \} \right\} + 
\frac{1}{\nu_0} P_{(2)}(\epsilon) \{ P_N\{Y_{11}>Y_01, Y_{12}>Y_{01}, |X_{11}-X_{01}| \leq \epsilon, |X_{12}-X_{01}| \leq \epsilon \} - 
2 P_N\{Y_{11}>Y_01, Y_{12}<Y_{01}, |X_{11}-X_{01}| \leq \epsilon, |X_{12}-X_{01}| \leq \epsilon \} + 
P_N\{Y_{11}<Y_01, Y_{12}<Y_{01}, |X_{11}-X_{01}| \leq \epsilon, |X_{12}-X_{01}| \leq \epsilon \}) - 
(\frac{1}{\nu_1} + \frac{1}{\nu_0}) \gamma_{T_1, N}^2(\theta_N; \epsilon) \right\}
\]

Since

\[
(3.29) \quad \lim_{N \to \infty} P_N\{Y_{11}>Y_01, Y_{11}>Y_{02}, |X_{11}-X_{01}| \leq \epsilon, |X_{11}-X_{02}| \leq \epsilon \} = 
\]

37
\[ P_0(Y_{11} > Y_{01}, Y_{11} > Y_{02} | |X_{11} - X_{01}| \leq \epsilon, |X_{11} - X_{02}| \leq \epsilon) = \frac{1}{3}. \]

Also,

\[ \lim_{N \to \infty} P_N(Y_{11} > Y_{01}, Y_{11} > Y_{02} | X_{11} = X_{01} = X_{02}) = \]

\[ P_0(Y_{11} > Y_{01}, Y_{11} > Y_{02} | X_{11} = X_{01} = X_{02}) = \frac{1}{3}. \]

Hence,

\[ \lim_{N \to \infty} \left( \frac{\tau_{T_1,N}(\theta_N; \epsilon)/N}{\tau_{T_1,N}(0; \epsilon)/N} \right)^{1/2} = 1, \]

and

\[ \tau_{T_1,N}(0; \epsilon) = \frac{N}{3} P_0(\epsilon) \left( \frac{1}{n_1} + \frac{1}{n_0} \right). \]

Finally,

\[ \lim_{N \to \infty} \frac{\tau_{T_1,N}'(\theta_N; \epsilon)}{\sqrt{\tau_{T_1,N}(0; \epsilon)}} = \lim_{N \to \infty} \frac{\sqrt{\frac{N}{3} P_0(\epsilon) \left( \frac{1}{n_1} + \frac{1}{n_0} \right)}}{\sqrt{\frac{2}{n_1 n_0}}} \]

\[ = \frac{2 \int \int_{\{ |X_1 - X_0| \leq \epsilon \}} f(y|x_0) f(y|x_1) \, dG(x_1) \, dG(x_0)}{\sqrt{3\lambda(1-\lambda) P_0(\epsilon)}} \]

\[ = \int \int_{\{ |X_1 - X_0| \leq \epsilon \}} f(y|x_0) f(y|x_1) \, dG(x_1) \, dG(x_0) \]

\[ \sqrt{12\lambda(1-\lambda)/2} \int \int_{\{ |X_1 - X_0| \leq \epsilon \}} f(y|x_0) f(y|x_1) \, dG(x_1) \, dG(x_0) / \sqrt{P_0(\epsilon)} = \text{eff}(T_1(\epsilon)) > 0, \]

and the result follows. \( \Box \)

Theorem 3.5.2

If \((Y_{11}, X_{11}), \ldots, (Y_{1n_1}, X_{1n_1}), (Y_{01}, X_{01}), \ldots, (Y_{0n_0}, X_{0n_0})\) are independent random samples satisfying Assumptions A1 – A3, and the conditions of Theorem 3.4.1, then for a given \( \epsilon > 0 \), the sequences of test statistics \( T_1, N(\theta_N; \epsilon) \) and \( U_N \) satisfy Conditions (a) – (e) of Theorem 3.2.2, and the asymptotic relative efficiency of \( T_1 \) with respect to \( U \) is

\[ 38 \]
(3.33) \[
\text{ARE}(T_1(\epsilon), U) = \frac{1}{P(2)(\epsilon)} \cdot \left( \frac{\int \int f(y|x_0) f(y|x_1) \, d(y|x_1) \, dG(x_1) \, dG(x_0)}{\int f^2(y) \, dy} \right)^2.
\]

Proof: Since \(\text{eff}(U) = \sqrt{12\lambda(1-\lambda)} \int f^2(y) \, dy\) (Randles and Wolfe, 1979, p 171), we have by Theorem 3.2.3 and Lemma 3.5.1,

\[
\text{ARE}(T_1(\epsilon), U) = \left( \frac{\text{eff}(T_1(\epsilon))}{\text{eff}(U)} \right)^2
\]

\[
= \left( \frac{\sqrt{12\lambda(1-\lambda)} \int \int f(y|x_0) f(y|x_1) \, d(y|x_1) \, dG(x_1) \, dG(x_0)/\sqrt{P(2)(\epsilon)}}{\sqrt{12\lambda(1-\lambda)} \int f^2(y) \, dy} \right)^2
\]

\[
= \frac{1}{P(2)(\epsilon)} \cdot \left( \frac{\int \int f(y|x_0) f(y|x_1) \, d(y|x_1) \, dG(x_1) \, dG(x_0)}{\int f^2(y) \, dy} \right)^2.
\]

When \(\epsilon = \infty\), \(T_1(\infty) \equiv U\), and (3.33) is equal to one, which is as desired.

In the case of \(X\) being discrete, i.e., \(X\) takes values at \(x_i\) with probabilities \(p_i\), \(i=1, 2, \ldots, k\), then for any \(\epsilon > 0\) such that \(\epsilon < \min |x_i - x_j|, i, j=1, \ldots, k\), (3.33) is equivalent to

(3.34) \[
\frac{P(1)(\epsilon)^2}{P(2)(\epsilon)} \cdot \left( \frac{\sum_i p_i \int f^2(y|x_i) \, d(y|x_i)}{\int f^2(y) \, dy} \right)^2,
\]

where \(P(1)(\epsilon) = P(|X_1 - X_0| \leq \epsilon) = \sum p_i^2\), and \(P(2)(\epsilon) = \sum p_i^3\). By the Schwartz inequality, \((\sum p_i^2)^2 \leq \sum p_i^3\). The equality holds only when \(p_i = 1/k\), \(i=1, \ldots, k\) (i.e., \(X\) is uniformly distributed on elements \(x_i\) with probabilities all equal to \(1/k\), \(i=1, \ldots, k\)).

In the case of standard parametric ANOCOVA, without loss of generality, suppose \(Y\) is normally distributed with mean zero and variance \(\sigma_y^2\) and \(Y \mid X\) is normally distributed with mean zero and variance \((1-\rho^2)\sigma_y^2 = \sigma^2\), we have

(3.35) \[
\int f^2(y|x) \, d(y|x) = \frac{1}{\sqrt{1-\rho^2}} \int f^2(y) \, dy = \frac{1}{2\sqrt{\pi(1-\rho^2)} \sigma_y}.
\]

39
Thus,

\[(3.36) \quad \text{ARE}(T_1(\epsilon), U) = \frac{1}{1 - \rho^2} \cdot \frac{P(1)(\epsilon)^2}{P(2)(\epsilon)}.\]

When \(Y\) and \(X\) are independent, \(\rho = 0\), and \(\text{ARE}(T_1(\epsilon), U) = \frac{P(1)(\epsilon)^2}{P(2)(\epsilon)} \leq 1\). This result is less desirable than that for AMP's given by Schoenfelder(1981), since here we are using an unweighted matching scheme rather than the optimal weights he found.

In general, we have that when \(X_1, X_2,\) and \(X_3\) are i.i.d. random variables, then

\[(3.37) \quad \text{P}(\epsilon) - \text{P}(\epsilon)^2 = \text{P}\{|X_1 - X_2| \leq \epsilon, |X_2 - X_3| \leq \epsilon\} - \text{P}^2\{|X_1 - X_2| \leq \epsilon\}
\]

\[= \int (F(x+\epsilon) - F(x-\epsilon))^2 dF(x) - (\int (F(x+\epsilon) - F(x-\epsilon))dF(x))^2 \]

\[= \text{Var}(h(X)) = \text{Var}(Y) \geq 0, \]

where \(Y = h(X) = F(X+\epsilon) - F(X-\epsilon)\), and \(F(.)\) is the distribution function of \(X\).

### 3.6 Category Matching

As an alternative matching method, we may also apply category matching rather than caliper matching. Suppose one can fairly divide the support of \(X\) into disjoint intervals \((x_0, x_1], (x_1, x_2], \ldots, (x_{m-1}, x_m]\), and the probability that \(x\) falls outside the range \((x_0, x_m]\) is exactly zero. The case when the range of the covariable is infinite can be derived similarly, and can be left for further theoretical interest. In most practical situations, however, the range of the covariable is bounded. Similar to (3.7), we have under the null hypothesis

\[(3.38) \quad \text{P}_0\{Y_1 > Y_0 \mid X_1 \in (x_l, x_{l+1}]\} = \text{P}_0\{Y_1 > Y_0 \mid X_1 = X_0\} = 1/2, 1 \leq l \leq m - 1.\]

We can therefore use the same set of Assumptions A1-A3 for both the matching schemes in deriving their ARE value. A weighted sum of the Mann-Whitney statistics \(U_i\) for each interval, similar to the AMP statistic of Schoenfelder (1981), is as follows:

\[(3.39) \quad Q = \sum_{l=1}^{m} \omega_l U_l,\]

where

\[(3.40) \quad U_l = \frac{1}{n_1 n_0} \sum_{j=1}^{n_1} \sum_{j'=1}^{n_0} \text{sgn}(Y_{1j} - Y_{0j'}) I\{X_{1j}, X_{0j'} \in (x_{l-1}, x_l]\}.\]

Also,
\begin{equation}
Q = \frac{1}{n_1 n_0} \sum_{j}^{n_1} \sum_{j'}^{n_0} h_Q[(Y_{1j}, X_{1j}), (Y_{0j'}, X_{0j'})],
\end{equation}

where

\begin{equation}
h_Q[(Y_{1j}, X_{1j}), (Y_{0j'}, X_{0j'})] = \sum_{l=1}^{m} \omega_l \text{sgn}(Y_{1j} - Y_{0j'}) I\{X_{1j}, X_{0j'} \in (x_{l-1}, x_l]\}.
\end{equation}

Thus, Q is a U-statistic of degree (1, 1) and kernel \( h_Q[(y_{1j}, x_{1j}), (y_{0j'}, x_{0j'})] \). For a sequence of alternatives, denote the kernel of \( Q_N \) by

\begin{equation}
h_{Q_N}((y_1, x_1), (y_0, x_0)).
\end{equation}

Then

\begin{equation}
E(\sum_{l=1}^{m} \omega_l \text{sgn}(Y_{11} - Y_{01}) I\{X_{11}, X_{01} \in (x_{l-1}, x_l]\}) < \sum_{l=1}^{m} \omega_l,
\end{equation}

where the subscript N denotes that the underlying distribution may also depend on N. Let

\begin{equation}
\gamma_{Q_N}(\theta) = E(Q_N) = E(h_{Q_N}((Y_1, X_1), (Y_0, X_0))),
\end{equation}

where \( \theta \) denotes that the underlying distribution depends on the parameter. Denote by

\begin{equation}
\psi_{1,Q_N}((y_1, x_1)) = E[ h_{Q_N}((y_1, x_1), (Y_0, X_0)] - \gamma_{Q_N}(\theta),
\end{equation}

and

\begin{equation}
\psi_{0,Q_N}((y_0, x_0)) = E[ h_{Q_N}((Y_1, X_1), (y_0, x_0))] - \gamma_{Q_N}(\theta).
\end{equation}

Also,

\begin{equation}
V_{Q_N} = \frac{1}{n_1} \sum_{j=1}^{n_1} \psi_{1,Q_N}((y_{1j}, x_{1j})) + \frac{1}{n_0} \sum_{j'=1}^{n_0} \psi_{0,Q_N}((y_{0j'}, x_{0j'})),
\end{equation}

\begin{equation}
\xi_{1,Q_N}(\theta) = E( \psi_{1,Q_N}^2((Y_1, X_1)) ),
\end{equation}

and

\begin{equation}
\xi_{0,Q_N}(\theta) = E( \psi_{0,Q_N}^2((Y_0, X_0)) ).
\end{equation}

Then we have

\begin{equation}
\text{Var}(V_{Q_N}) = \frac{1}{n_1} \xi_{1,Q_N}(\theta) + \frac{1}{n_0} \xi_{0,Q_N}(\theta),
\end{equation}

since \((Y_{11}, X_{11}), ..., (Y_{1n_1}, X_{1n_1})\) and \((Y_{01}, X_{01}), ..., (Y_{0n_0}, X_{0n_0})\) are independent samples from \((Y_1, X_1)\) and \((Y_0, X_0)\) respectively. Let

\begin{equation}
\tau_{Q_N}(\theta) = N \text{Var}(V_{Q_N}),
\end{equation}
Suppose $P\{ X \in (x_{l-1}, x_l] \} = q_l$, $\sum_{l=1}^{L} q_l = 1$ and Assumptions A1-A3 hold, then we have the following result.

Theorem 3.6.1

If the first derivative of the density function $g(x)$ of $X$ exists and is square integrable in a set of probability greater than zero, and the correlation between $Y$ and $X$ is less than one, then the sequence of test statistics

\[ \frac{Q_N - \gamma Q_N(\theta)}{\sqrt{\tau Q_N(\theta)/N}} \rightarrow N(0, 1) \]

(3.53)

uniformly in $\theta$, $0 = \theta_0 \leq \theta \leq \theta_0 + \eta$, $\eta$ is a positive constant, as $N \rightarrow \infty$.

Proof: We can proof this similarly to that in Theorem 3.4.1. □

Theorem 3.6.2

If $(Y_{11}, X_{11})$, ..., $(Y_{1n_1}, X_{1n_1})$ and $(Y_{01}, X_{01})$, ..., $(Y_{0n_0}, X_{0n_0})$ are independent random samples satisfying Assumptions A1 - A3, and the conditions of Theorems 3.4.1 and 3.6.1, then the efficacy of the statistic $Q$ is

\[ \sqrt{12\lambda(1-\lambda)} \sum_l \omega_l \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} f(y|x_0) f(y|x_1) \, d(y|x_1) \, dG(x_1) \, dG(x_0) \]

(3.54) $\text{eff}(Q) = \frac{\sum_l \omega_l^2 q_l^3}{\sqrt{\sum_l \omega_l^2 q_l^3}}$,

and

\[ \text{ARE}(T_1(\epsilon), Q) = \frac{\sum_l \omega_l^2 q_l^3}{P_2(\epsilon)} \left( \frac{\int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} f(y|x_0) f(y|x_1) \, d(y|x_1) \, dG(x_1) \, dG(x_0)}{(x_{l-1} - x_0) \leq \epsilon} \right)^2 \]

(3.55)

Proof: By Theorem 3.6.1, the sequence of statistics $Q_N$ satisfies Conditions (a), (b) of Theorem 3.2.2. The expectation of $Q_N$ is

(3.56)

\[ \gamma_{T_1, N}(\theta_N; \epsilon) = P_N\{Y_1 > Y_0, X_1, X_0 \in (x_{l-1}, x_l]\} - P_N\{Y_1 < Y_0, X_1, X_0 \in (x_{l-1}, x_l]\} \]
\[ \begin{align*}
&= \sum_l \omega_l \left( \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} F(y + \theta_N \mid x_0) \, dF(y \mid x_1) \, dG(x_1) \, dG(x_0) - \\
&\quad \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} F(y - \theta_N \mid x_1) \, dF(y \mid x_0) \, dG(x_1) \, dG(x_0) \right). \\
\gamma_{Q_N}(\theta_N; \epsilon) &= \sum_l \omega_l \left( \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} f(y + \theta_N \mid x_0) \, f(y \mid x_1) \, d(y \mid x_1) \, dG(x_1) \, dG(x_0) + \\
&\quad \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} f(y - \theta_N \mid x_1) \, f(y \mid x_0) \, d(y \mid x_0) \, dG(x_1) \, dG(x_0) \right),
\end{align*} \]

Therefore,
\[ \gamma_{Q_N}'(0; \epsilon) = 2 \sum_l \omega_l \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} f(y \mid x_0) \, f(y \mid x_1) \, d(y \mid x_1) \, dG(x_1) \, dG(x_0) > 0, \]

and
\[ \lim_{N \to \infty} \frac{\gamma_{Q_N}'(\theta_N; \epsilon)}{\gamma_{Q_N}(0; \epsilon)} = 1. \]

Now,
\[ \xi_{1,Q_N}(\theta_N) = E( \psi_{1,Q_N}^2((Y_1, X_1))) = \]
\[ = E[\sum_l \omega_l \text{sgn}(Y_{11} - Y_{01}) I\{X_{11}, X_{01} \in (x_{l-1}, x_l)\}] + \\
\sum_l \omega_l' \text{sgn}(Y_{11} - Y_{01}) I\{X_{11}, X_{01} \in (x_{l'-1}, x_{l'})\} - \gamma_{Q_N}^2(\theta_N) \]
\[ = \sum_l \omega_l^2 q^3 \left( P_N\{Y_{11} > Y_{01}, Y_{11} > Y_{02} \mid X_{11}, X_{01}, X_{02} \in (x_{l-1}, x_l)\} - 2 P_N\{Y_{11} > Y_{01}, Y_{11} < Y_{02} \mid X_{11}, X_{01}, X_{02} \in (x_{l-1}, x_l)\} + \\
P_N\{Y_{11} < Y_{01}, Y_{11} < Y_{02} \mid X_{11}, X_{01}, X_{02} \in (x_{l-1}, x_l)\}\right) - \gamma_{Q_N}^2(\theta_N), \]

and
\[ \xi_{0,Q_N}(\theta_N) = E( \psi_{0,Q_N}^2((Y_0, X_0))) = \]
\[ = E[\sum_l \omega_l \text{sgn}(Y_{11} - Y_{01}) I\{X_{11}, X_{01} \in (x_{l-1}, x_l)\}] + \]
\[ \sum_{\ell} \omega_{\ell} \text{sgn}(Y_{12} - Y_{01}) \mathbb{I}(x_{12}, x_{01} \in (x_{\ell - 1}, x_{\ell}]) - \gamma_{Q_N}^2(\theta_N) \]

\[ = \sum_{\ell} \omega_{\ell}^2 q_{\ell}^3 (P_N(Y_{11} > Y_{01}, Y_{12} > Y_{01} \mid X_{11}, X_{12}, X_{01} \in (x_{\ell - 1}, x_{\ell}]) - \]

\[ 2 \ P_N(Y_{11} > Y_{01}, Y_{12} < Y_{01} \mid X_{11}, X_{12}, X_{01} \in (x_{\ell - 1}, x_{\ell}]) + \]

\[ P_N(Y_{11} < Y_{01}, Y_{12} < Y_{01} \mid X_{11}, X_{01}, X_{02} \in (x_{\ell - 1}, x_{\ell}]) - \gamma_{Q_N}^2(\theta_N) \]

(since \((x_{\ell - 1}, x_{\ell}),(x_{\ell' - 1}, x_{\ell'})\) are disjoint intervals and \(q_\ell, q_{\ell'}\) follow a multinominal distribution). Thus,

\[ \tau_{Q_N}(\theta_N) = N \text{ Var}(V_{Q_N} \ast) = N \left( \frac{1}{n_1} \epsilon_{1,Q_N}(\theta_N) + \frac{1}{n_0} \epsilon_{0,Q_N}(\theta_N) \right) \]

\[ = N \sum_{\ell} \omega_{\ell}^2 q_{\ell}^3 \left( \frac{1}{n_1} (P_N(Y_{11} > Y_{01}, Y_{11} > Y_{02} \mid X_{11}, X_{01}, X_{02} \in (x_{\ell - 1}, x_{\ell}]) - \right. \]

\[ 2 \ P_N(Y_{11} > Y_{01}, Y_{11} < Y_{02} \mid X_{11}, X_{01}, X_{02} \in (x_{\ell - 1}, x_{\ell}]) + \]

\[ P_N(Y_{11} < Y_{01}, Y_{11} < Y_{02} \mid X_{11}, X_{01}, X_{02} \in (x_{\ell - 1}, x_{\ell}]) \]

\[ \frac{1}{n_0} (P_N(Y_{11} > Y_{01}, Y_{12} > Y_{01} \mid X_{11}, X_{12}, X_{01} \in (x_{\ell - 1}, x_{\ell}]) - \]

\[ 2 \ P_N(Y_{11} > Y_{01}, Y_{12} < Y_{01} \mid X_{11}, X_{12}, X_{01} \in (x_{\ell - 1}, x_{\ell}]) + \]

\[ P_N(Y_{11} < Y_{01}, Y_{12} < Y_{01} \mid X_{11}, X_{12}, X_{01} \in (x_{\ell - 1}, x_{\ell}])) - \]

\[ N \left( \frac{1}{n_1} + \frac{1}{n_0} \right) \gamma_{Q_N}^2(\theta_N) , \]

and

\[ \lim_{N \to \infty} \left( \frac{\tau_{Q_N}(\theta_N)/N}{\tau_{Q_N}(0)/N} \right)^{1/2} = 1, \]

since

\[ \lim_{N \to \infty} P_N(Y_{11} > Y_{01}, Y_{11} > Y_{02} \mid X_{11}, X_{01}, X_{02} \in (x_{\ell - 1}, x_{\ell}]) = \]

\[ P_0(Y_{11} > Y_{01}, Y_{11} > Y_{02} \mid X_{11}, X_{01}, X_{02} \in (x_{\ell - 1}, x_{\ell}]) = \frac{1}{3} , \]

and

\[ \lim_{N \to \infty} P_N(Y_{11} > Y_{01}, Y_{11} > Y_{02} \mid X_{11} = X_{01} = X_{02}) = \]

\[ P_0(Y_{11} > Y_{01}, Y_{11} > Y_{02} \mid X_{11} = X_{01} = X_{02}) = \frac{1}{3} . \]
Also,\n\[ \tau_{Q_N}(0) = \frac{\sum I_i^2 q_i^3}{3\lambda(1-\lambda)}. \]

Hence,\n\[
\text{eff}(Q) = \lim_{N \to \infty} \frac{\tau_{Q_N}'(\theta_N)}{\sqrt{\tau_{Q_N}(0)}} \approx \frac{\sqrt{12\lambda(1-\lambda) \sum I_i^2 q_i^3}}{\sqrt{\sum I_i^2 q_i^3}} \int_{x_l-1}^{x_l} \int_{x_l-1}^{x_l} \int f(y|x_0) f(y|x_1) d(y|x_1) dG(x_1) dG(x_0)
\]

Finally,\n\[
\text{ARE}(T_1(\epsilon), Q) = \left( \frac{\text{eff}(T_1(\epsilon))}{\text{eff}(Q)} \right)^2
\]

\[
= \left( \frac{\sqrt{12\lambda(1-\lambda) \sum I_i^2 q_i^3}}{\sqrt{12\lambda(1-\lambda)P(2)(\epsilon) \sum I_i^2 q_i^3}} \int_{x_l-1}^{x_l} \int_{x_l-1}^{x_l} \int f(y|x_0) f(y|x_1) d(y|x_1) dG(x_1) dG(x_0) \right)^2
\]

Again, in the case of $X$ being discrete ($X$ takes values at $x_i$ with probabilities $p_i$, $i=1, 2, ..., k$), then for any $\epsilon>0$ such that $\epsilon<\min |x_i-x_j|$, $i, j=1, ..., k$, and $x_i$ lies in either one of intervals $(x_l, x_{l+1})$, $l=1, ..., m$, (3.52) is equivalent to\n\[ \left( \frac{\sum I_i^2 q_i^3}{P(2)(\epsilon)} \right)^2 \]
\[
(3.65) \quad \text{ARE}(T_1(\epsilon), Q^*) = \frac{P_1(\epsilon) \cdot (\sum p_i^2)^2}{P_2(\epsilon) \cdot \sum p_i^3} \leq 1.
\]

In general, for a continuous covariable $X$, if the length of each interval is approximately $\epsilon$, then the inner integrants of the numerator and the denominator of (3.55) are approximately the same, thus
\[
(3.66) \quad \text{ARE}(T_1(\epsilon), Q^*) \approx \frac{P_1(\epsilon)^2}{P_2(\epsilon)}.
\]

In which case, the efficiency of $T_1(\epsilon)$ is less or equal to $Q^*$. This may be because the statistic $T_1(\epsilon)$ lacks an appropriate weighting scheme for the matched pairs.

### 3.7 Discussion

The unweighted caliper matching statistic $T_1(\epsilon)$, under the assumption of concomitance, may be more efficient or less efficient than simple random sampling ((3.33), (3.36)), and is in general less efficient than the optimal category matching statistic $Q^*$, when using the same scale of matching interval ((3.65), (3.66)). Thus, one might need to apply a certain weighting scheme to $T_1(\epsilon)$ to achieve better efficiency.

The test statistic $\bar{T}(\epsilon)$ for “matched difference in probability” (Quade, 1982), uses a natural weighting scheme based on the ratio of the number of pairs matched to a certain observation to the number of overall possible matched pairs, which is similar to the weighting scheme of the optimal category matching. Therefore, the statistic $\bar{T}(\epsilon)$ may be expected to have better efficiency than the unweighed statistic $T_1(\epsilon)$.

It seems that neither of the two matching schemes (caliper matching and category matching) takes into account the regression effect left out by the unmatched pairs. In the next chapter we turn to the ERMP test to investigate its efficiency properties when additional information is added to the matched pairs.
CHAPTER 4
EXTENDED RANK AND MATCHED-PAIR ANALYSIS
AND ITS EFFICIENCY PROPERTIES

4.1 Introduction

In the last chapter, we discussed efficiency properties of caliper matching for the two-sample case and gave its asymptotic relative efficiencies with respect to the Mann-Whitney U-statistic as well as category matching. We extend our discussion to the more general k-sample case, and give efficiency properties of the tests ERMP(ω; ε) and ERMP(λ; ε) (Definition 2.3.5) in this chapter.

It was shown in Section 2.3 that the use of scores Z_i^j(ω; ε) provides an extended version of "rank analysis of covariance" (Quade, 1967). On the other hand, the scores Z_i^j(ω; ε) combine the matched and unmatched pairs by assigning them specific weights. Thus, the approach by considering these scores also extends the method of analysis of covariance by matching (Quade, 1982). Since all the scores defined in the sequel are interchangeable random variables as well as U-statistics, the asymptotic F-distributions of their variance ratios follow directly from the results of Quade (1966). We use these results to examine the mutual asymptotic relative efficiencies of these scores by comparing the noncentrality parameters of their asymptotic F-distributions.

In Section 4.2, we give the asymptotic distribution of the variance ratio VR(λ; ε) (2.32), and discuss the ARE of ERMP(λ; ε) with respect to the Kruskal-Wallis k-sample statistic. Also, estimates of the optimal weights are given in Section 4.3. In Section 4.4, we discuss the ARE of the VR of the scores Z_i^j*(ω_1, ω_2; ε) with respect to the VR of the Z_i^j(ω; ε) in two-sample cases. We explore the relationship between the Spearman ρ and ρ(ε) (defined in Section 4.4) given tolerance ε in Section 4.5. The ARE of ERMP(λ; ε) with respect to the category matching (stratified matching) for univariate covariable case is given in Section 4.6. Finally, in Section 4.7, we present a summary discussion.

4.2 Asymptotic Distribution of the VR of the Rank Adjusted Scores And Its Efficiency Properties
Similar to the notation in Quade (1967), let

\begin{equation}
\theta_{ii'} = E(\phi(Y_{ij} - Y_{i'j'} ; 0)) = \frac{1}{2} E(\text{sgn}(Y_{ij} - Y_{i'j'})) = \frac{1}{2} (P(Y_i > Y_{i'}) - P(Y_i < Y_{i'})),
\end{equation}

\begin{equation}
\theta_i = \sum_{i'} \alpha_{i'} \theta_{ii'},
\end{equation}

where \(Y_i\) and \(Y_{i'}\) are independent random variables from the \(i\)-th and \(i'\)-th populations, \(\alpha_{i'} = n_{i'}/N\), \(1 \leq i, i' \leq k\). As in that paper, we may interpret \(\theta_i\) as the probability that a response chosen at random from the \(i\)-th population will exceed one from the "average" population, less the probability of the opposite ordering. Note that \(\theta_{ii'} + \theta_{i'i} = 0\), and \(\theta_{ii} = 0\). Now,

\[
E(\phi(X_{ij}^{(l)} - X_{i'j'}^{(l)} ; \epsilon)) = \frac{1}{2} E(I\{X_{ij}^{(l)} > X_{i'j'}^{(l)} + \epsilon\} - I\{X_{ij}^{(l)} < X_{i'j'}^{(l)} - \epsilon\}) = 0,
\]

since \(X_{ij}^{(l)}\) and \(X_{i'j'}^{(l)}\) are identically distributed for any \((i,j) \neq (i', j')\), \(1 \leq l \leq p\). Let \(\lambda = (\lambda_1, ..., \lambda_p)\) and \(\epsilon = (\epsilon_1, ..., \epsilon_p)\), and

\begin{equation}
\psi((Y_{ij}, Y_{i'j'}), (X_{ij}, X_{i'j'}); \lambda; \epsilon) = \phi(Y_{ij} - Y_{i'j'} ; 0) - \sum_{l=1}^{p} \lambda_l \phi(X_{ij}^{(l)} - X_{i'j'}^{(l)} ; \epsilon_l),
\end{equation}

then

\begin{equation}
E( Z_{ij}(\lambda; \epsilon) ) = E( \sum_{i'} \sum_{j'} [\phi(Y_{ij} - Y_{i'j'} ; 0) - \sum_{l=1}^{p} \lambda_l \phi(X_{ij}^{(l)} - X_{i'j'}^{(l)} ; \epsilon_l)])
\end{equation}

\[
= \sum_{i'} \sum_{j'} [E(\phi(Y_{ij} - Y_{i'j'} ; 0)) - \sum_{l=1}^{p} \lambda_l E(\phi(X_{ij}^{(l)} - X_{i'j'}^{(l)} ; \epsilon_l))] = N \sum_{i'} \alpha_{i'} \theta_{ii'} = N \theta_i.
\]

Under the null hypothesis \(H_0\) that the conditional distribution of \(Y\) given \(X\) is the same for each population, we have \(\theta_i = 0\) for \(1 \leq i \leq k\). Similarly, we have

\begin{equation}
\text{Var}[ Z_{ij}(\lambda; \epsilon) ] = \frac{N^2 - 1}{12} \left[ \sigma_i^2 - 2 \lambda' \eta_\epsilon(\epsilon) + \lambda' \Sigma(\epsilon) \lambda \right] = (N^2 - 1) \xi_i(\lambda; \epsilon),
\end{equation}

where \(\frac{N^2 - 1}{12} \sigma_i^2\) is the variance of the rank of a response from the \(i\)-th population, \(\frac{N^2 - 1}{12} \eta_\epsilon(\epsilon)\) are the covariances of the ranks of the response with the generalized ranks of the concomitant variates given the tolerance vector \(\epsilon\), and \(\frac{N^2 - 1}{12} \Sigma(\epsilon)\) is the \(p \times p\) variance matrix of the generalized ranks of the concomitant variates. Let
(4.5) \[ \xi(\lambda ; \epsilon) = \sum_i a_i \xi_i(\lambda ; \epsilon) . \]

Under \( H_0 \),

(4.6) \[ \text{Var}[Z_{ij}(\lambda ; \epsilon)] = N_1^2 \left[ \frac{N-1}{12} \sigma^2 - 2 \lambda' \eta(\epsilon) + \lambda' \Sigma(\epsilon) \lambda \right] = \sigma_N^2(\lambda ; \epsilon) \]

for all \( i, j \), where \( \sigma^2 \) and \( \eta(\epsilon) \) are common values of \( \sigma_i^2 \) and \( \eta_i(\epsilon) \) under \( H_0 \) given \( \epsilon \). Also,

(4.7) \[ \text{Cov}[Z_{ij}(\lambda ; \epsilon), Z_{i'j'}(\lambda ; \epsilon)] = \rho_N(\lambda ; \epsilon) \sigma_N^2(\lambda ; \epsilon), \quad (i, j) \neq (i', j'). \]

We have the following theorem which is a direct extension of the theorem of Quade (1967a). The sketch of the proof is mainly from Quade (1966).

**Theorem 4.2.1**

If \( H_0 \) is true and \( \xi(\lambda ; \epsilon) > 0 \), then as \( N \to \infty \) and \( n_i = Np_i + O(N^2) \), where \( p_i > 0 \), \( 1 \leq i \leq k \), the random variable VR(\( \lambda \); \( \epsilon \)) has asymptotically the F-distribution with \((k - 1, N - k)\) degrees of freedom; and VR(\( \lambda \); \( \epsilon \)) is consistent for testing \( H_0 \) against any alternative for which \( \sum_i a_i \theta_i^2 > 0 \).

**Proof:** Since

\[
|\phi(y_{ij} - y_{i'j'} ; 0) - \sum_{i=1}^k \lambda_i \phi(x_{ij}^{(i)} - x_{i'j'}^{(i)} ; \epsilon_i)| \leq \frac{1}{2} (1 + \sum |\lambda_i|),
\]

Assumption B2 of Quade (1966) is satisfied and hence Assumption B1 is satisfied. Let

(4.8) \[ \bar{Z}_i(\lambda ; \epsilon) = \frac{1}{n_i} \sum_{j=1}^{n_i} Z_{ij}(\lambda ; \epsilon) ; \quad \bar{Z}(\lambda ; \epsilon) = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} Z_{ij}(\lambda ; \epsilon) , \]

we have by Theorem 3 of that paper,

(4.9) \[ \sqrt{N} \left( \frac{1}{N - 1} \right) \bar{Z}_i(\lambda ; \epsilon) - \theta_i , \]

\( i = 1, \ldots, k \) have asymptotically a joint normal distribution with zero mean and finite variance matrix. Let

(4.10) \[ s^2(\lambda ; \epsilon) = (N - k)^{-1} \sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij}(\lambda ; \epsilon) - \bar{Z}_i(\lambda ; \epsilon))^2 , \]

(4.11) \[ \theta((Y_{ij}, X_{ij})) = \int \psi((Y_{ij}, X_{ij}), (Y_{i'j'}, X_{i'j'}); \lambda ; \epsilon) \, dH(Y_{i'j'}, X_{i'j'}) , \]

(4.12) \[ a_{ij} = \theta((Y_{ij}, X_{ij})) - \theta_i , \]

(4.13) \[ b_{ij} = \left( \frac{N - 1}{1} \right)^{-1} Z_{ij}(\lambda ; \epsilon) - \theta_i , \]

49
where $H(Y, X)$ is the average distribution function of the $k$ treatment groups. Then,

\begin{equation}
E[\theta((Y_{ij}, X_{ij}))] = \theta_i ; \ Var[\theta((Y_{ij}, X_{ij}))] = \xi_i(\lambda ; \epsilon).
\end{equation}

Also, by Khinchin's law of large numbers, we have

\begin{equation}
 n_i^{-1} \sum_j a_{ij}^2 \rightarrow \xi_i(\lambda ; \epsilon)
\end{equation}

in probability. Following Theorem 4 of the paper,

\begin{equation}
 n_i^{-1} \sum_j (a_{ij} - b_{ij})^2 = n_i^{-1} \sum_j \left[ \binom{N-1}{1} \left( \binom{N-1}{1} - 1 \right) Z_{ij}(\lambda ; \epsilon) - \theta_i((Y_{ij}, X_{ij})) \right]^2 \rightarrow 0
\end{equation}

in probability. Thus, by Lemma 1 of Sen (1960),

\begin{equation}
 n_i^{-1} \sum_j b_{ij}^2 \rightarrow \xi_i(\lambda ; \epsilon)
\end{equation}

in probability. Therefore,

\begin{equation}
 (n_i - 1)^{-1} \sum_j \left( \binom{N-1}{1} - 1 \right) Z_{ij}(\lambda ; \epsilon) - \theta_i \rightarrow \xi_i(\lambda ; \epsilon).
\end{equation}

Now,

\begin{equation}
 \binom{N-1}{1} (n_i - 1)^{-1} \sum_j Z_{ij}(\lambda ; \epsilon) - Z_i(\lambda ; \epsilon) = (n_i - 1)^{-1} \sum_j \left[ \binom{N-1}{1} - 1 \right] Z_{ij}(\lambda ; \epsilon) - \theta_i \rightarrow \xi_i(\lambda ; \epsilon) - \frac{n_i}{n_i - 1} \left( \binom{N-1}{1} - 1 \right) Z_i(\lambda ; \epsilon) - \theta_i.
\end{equation}

and the second of the two sums converges to zero in probability by Theorem 3 of the same paper. Hence,

\begin{equation}
 s^2(\lambda ; \epsilon) \binom{N-1}{1} - 2 = \binom{N-1}{1} - 2 (N - k) \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Z_{ij}(\lambda ; \epsilon) - Z_i(\lambda ; \epsilon))^2
\end{equation}

\begin{equation}
 \rightarrow \sum_i \alpha_i \xi_i(\lambda ; \epsilon) = \xi(\lambda ; \epsilon)
\end{equation}

in probability. Following the proof of Theorem 6 of the same paper, under the hypothesis, the $Z_{ij}(\lambda ; \epsilon)$ are interchangeable with variance of any score $\sigma_N^2(\lambda ; \epsilon)$ and correlation of any two scores $\rho_N(\lambda ; \epsilon)$. Let $Z = (Z_1, Z_2, ..., Z_k)'$, where

\begin{equation}
 Z_i = \sqrt{N} \left\{ \frac{1}{\binom{N-1}{1}} Z_i(\lambda ; \epsilon) - \theta_i \right\}.
\end{equation}

We have

50
(4.20) \[ \text{Var}(Z) = \Sigma_N = N \sigma_N^2(\lambda; \epsilon)(1 - \rho_N(\lambda; \epsilon)) D_N^{-1} + \rho_N(\lambda; \epsilon) J \] / \left( N - 1 \right)^2,

where \( D_N = \text{diag}(n_1, n_2, ..., n_k) \) and \( J \) is the matrix in which every element is unity. Let

(4.21) \[ A_N = \left( \begin{array}{c} N \end{array} \right)^2 \left( N D_N - n n' \right) / N^2 \sigma_N^2(\lambda; \epsilon)(1 - \rho_N(\lambda; \epsilon)), \]

where \( n = (n_1, n_2, ..., n_k)' \). Then \( A_N \Sigma_N = I - D_N J / N \) is an idempotent matrix of rank \((k - 1)\). Thus, following (4.8) and Cochran's Theorem, we have that the quadratic form

(4.22) \[ Z'A_N Z = \left[ \sum_i n_i (Z_i(\lambda; \epsilon) - \bar{Z}(\lambda; \epsilon))^2 \right] / \sigma_N^2(\lambda; \epsilon)(1 - \rho_N(\lambda; \epsilon)) \]

is asymptotically distributed as \( \chi^2(k - 1) \). Also, \( E(s^2(\lambda; \epsilon)) = \sigma_N^2(\lambda; \epsilon)(1 - \rho_N(\lambda; \epsilon)) \). Therefore, by (4.18) and Slutsky's Theorem, we have

(4.23) \[ (k - 1) \text{VR}(\lambda; \epsilon) = Z'A_N Z \sigma_N^2(\lambda; \epsilon)(1 - \rho_N(\lambda; \epsilon)) / s^2(\lambda; \epsilon) \]

is asymptotically distributed as \( \chi^2(k - 1) \). Following the same argument as in the paper, this is equivalent to saying that \( \text{VR}(\lambda; \epsilon) \) is asymptotically \( F(k - 1, N - k) \). \( \xi(\lambda; \epsilon) > 0 \) is equivalent to Assumption C of that paper, which says that the random variable \( Z_i(\lambda; \epsilon) \) should not equal a constant with probability one.

Under the alternative, we have by (4.18)

(4.24) \[ \left[ \sum_i n_i (Z_i(\lambda; \epsilon) - \bar{Z}(\lambda; \epsilon))^2 \right] / N \left( \begin{array}{c} N \end{array} - 1 \right)^2 = (k - 1) \text{VR}(\lambda; \epsilon) s^2(\lambda; \epsilon) / N \left( \begin{array}{c} N \end{array} - 1 \right)^2 \rightarrow \sum_{i=1}^{k} \alpha_i \theta_i^2 \]

in probability. Thus, by (4.8),

(4.25) \[ (k - 1) \text{VR}(\lambda; \epsilon) / N \rightarrow \sum_{i=1}^{k} \alpha_i \theta_i^2 / \xi(\lambda; \epsilon) \]

in probability, and the result follows. \( \Box \)

If \( \lambda = 0 \), then \( \text{VR}(0; \epsilon) \) is a modified version of the Kruskal-Wallis test (KW) as discussed in Quade (1967). Hence, one can find the ARE of the extended rank and matched-pair (ERM \( \text{MP}(\lambda; \epsilon) \)) analysis with respect to the modified Kruskal-Wallis test by comparing \( \text{VR}(\lambda; \epsilon) \) to \( \text{VR}(0; \epsilon) \), i.e., when the covariates are ignored. Under the alternative \( H_N \) for each \( N \),

(4.26) \[ \theta_{ii'} = \frac{\delta_{ii'}}{\sqrt{N}} + o\left(\frac{1}{\sqrt{N}}\right), \]

and

51
\begin{equation}
\text{Var}\left[ Z_{ij}(\lambda; \epsilon) \right] = \text{Var}\left[ R_{ij} - \sum_{l} \lambda_l C_{ij}^{(l)}(\epsilon_l) \right] \\
= \text{Var}\left[ \sum_{i'} \sum_{j'} (\phi(Y_{ij} - Y_{i'j'}) ; 0) - \sum_{l=1}^{p} \lambda_l \phi(X_{ij}^{(l)} - X_{i'j'}^{(l)} ; \epsilon_l) \right] \\
= \text{Var}\left[ \frac{1}{2} \sum_{i'} \sum_{j'} \text{sgn}(Y_{ij} - Y_{i'j'}) \right] - 2 \sum_{l=1}^{p} \lambda_l \text{Cov}\left[ \frac{1}{2} \sum_{i'} \sum_{j'} \text{sgn}(Y_{ij} - Y_{i'j'}) , \frac{1}{2} \sum_{i'} \sum_{j'} \text{sgn}(X_{ij}^{(l)} - X_{i'j'}^{(l)}) \right] \\
= \frac{1}{2} \sum_{i'} \sum_{j'} \text{sgn}(X_{ij}^{(l)} - X_{i'j'}^{(l)}) \mathbb{I}\{|X_{ij}^{(l)} - X_{i'j'}^{(l)}| > \epsilon_l\} \\
+ \text{Var}\left[ \frac{1}{2} \sum_{i'} \sum_{j'} \text{sgn}(X_{ij}^{(l)} - X_{i'j'}^{(l)}) \mathbb{I}\{|X_{ij}^{(l)} - X_{i'j'}^{(l)}| > \epsilon_l\} \right] \\
= (N^2 - 1) \left[ \xi(\lambda ; \epsilon) + O\left(\frac{1}{\sqrt{N}}\right) \right]
\end{equation}

by (4.26), where

\[ \xi(\lambda ; \epsilon) = \frac{1}{12} \left[ \sigma^2 - 2 \lambda' \eta(\epsilon) + \lambda' \Sigma(\epsilon) \lambda \right]. \]

Also, \((k - 1)\text{VR}(\lambda; \epsilon) \rightarrow \sum_{i} \alpha_i \delta_i^2 / \xi(\lambda ; \epsilon)\) by (4.26), where \(\delta_i = \sum_{i'} \alpha_{i'i'}\). Hence, asymptotically, \(\text{VR}(\lambda ; \epsilon)\) will have a noncentral F distribution with degrees of freedom \((k - 1, N - k)\) and noncentrality parameter

\begin{equation}
\Delta(\lambda ; \epsilon) = \sum_{i} \alpha_i \xi_i^{(l)}^2 / \xi(\lambda ; \epsilon)
\end{equation}

under \(H_N\). To achieve the same power for test statistics \(\text{VR}(\lambda; \epsilon)\) and \(\text{VR}(0 ; \epsilon)\), the required ratio of sample sizes of these two statistics would be the ratio of their noncentrality parameters.

Since the noncentrality parameter \(\Delta(\lambda; \epsilon)\) is a function of the weights \(\lambda\), one can achieve the optimal efficacy by minimizing \(\xi(\lambda ; \epsilon)\), and thus has optimal relative efficiency with respect to other tests either in analysis of variance or analysis of covariance setting. Following the discussion in Quade (1967), if \(\Sigma(\epsilon)^{-1}\) exists, then the optimal choice of the vector \(\lambda\) is \(\lambda = \tau\), where \(\tau = \Sigma(\epsilon)^{-1} \eta(\epsilon)\). Thus,

\begin{equation}
\min_{\lambda} \xi(\lambda ; \epsilon) = \xi(\tau ; \epsilon) = \sigma^2 - \eta'(\epsilon) \Sigma^{-1}(\epsilon) \eta(\epsilon).
\end{equation}

The optimal asymptotic relative efficiency of \(\text{ERMP}(\lambda; \epsilon)\) with respect to the rank analysis of covariance following Quade (1967) is

\begin{equation}
\text{ARE}(\text{ERMP}(\tau; \epsilon), \text{RankAC}(\tau')) = \frac{\Delta(\tau; \epsilon)}{\Delta(\tau')}
\end{equation}

\[ = \frac{1}{1 - \eta'(\epsilon) \Sigma^{-1}(\epsilon) \eta(\epsilon) / \sigma^2} \cdot \frac{1 - R^2_s}{1 - R^2_s(\epsilon)}, \]

52
where \( \Delta(\tau^r) = \sum_i \alpha_i \delta_i^2 / \xi(\tau^r) \) is the optimal efficacy of \( \text{RankAC}(\lambda^r) \), \( \xi(\tau^r) = \Sigma^{-1}\eta \), if \( \Sigma^{-1} \) exists, and

\[
R_{\theta}(\epsilon) = \pm \sqrt{\eta'(\epsilon)\Sigma^{-1}(\epsilon)\eta(\epsilon)/\sigma^2},
\]

(4.31)

may be called the Spearman multiple correlation coefficient between the response and the concomitant variates given the vector of tolerances \( \epsilon \). In the univariate case, \( R_{\theta} = \rho_{\theta} \), and \( R_{\theta}(\epsilon) = \rho_{\theta}(\epsilon) \). We show in the next section that, under certain condition, \( |\rho_{\theta}(\epsilon)| \leq |\rho_{\theta}| \) for all \( \epsilon \). Thus, (4.30) is less or equal to one for all \( \epsilon \) in such case, which is due to loss of information using the extended ranks.

In most practical situations, the relative efficiencies of matching statistics with respect to statistics of simple random sampling without matching are of interest. The ARE of \( \text{VR}(\tau; \epsilon) \) with respect to \( \text{VR}(0; \epsilon) \) (\( = \text{KW}(\epsilon) = \text{KW} \)) in the Pitman sense is

\[
\frac{\sigma^2}{\sigma^2 - \eta'(\epsilon) \Sigma^{-1}(\epsilon) \eta(\epsilon)} = \frac{1}{1 - R_{\theta}^2(\epsilon)}
\]

(4.32)

Similar to the derivation in Quade(1967), the optimal ratio of ARE of \( \text{ERMP}(\lambda; \epsilon) \) with respect to the classical parametric analysis of covariance (AC) when the conditions of the latter are appropriate is

\[
\text{ARE}(\text{ERMP}(\tau; \epsilon), \text{AC}) = \text{ARE}(\text{ERMP}(\tau; \epsilon), \text{KW}(\epsilon)) \ast \text{ARE}(\text{KW}(\epsilon), \text{AV}) \ast \text{ARE}(\text{AV}, \text{AC})
\]

\[
= \frac{1}{1 - R_{\theta}^2(\epsilon)} \ast \frac{3}{\pi} \ast (1 - R_p^2) = \frac{3(1 - R_p^2)}{\pi(1 - R_{\theta}^2(\epsilon))},
\]

(4.33)

where \( R_p \) is the Pearson correlation coefficient between the response and the covariates. We examine the bounds of this ARE, and find the relationship between \( \rho_{\theta} \) and \( \rho_{\theta}(\epsilon) \) given \( \epsilon \) for the univariate case in the next section.

4.3 The Relationship Between the Spearman Correlation Coefficient \( \rho_{\theta} \) and \( \rho_{\theta}(\epsilon) \) Given \( \epsilon \)

Suppose in the combined sample, the ranks of \( (Y_1, X_1), \ldots, (Y_N, X_N) \) are \( (R_1, C_1(\epsilon)), \ldots, (R_N, C_N(\epsilon)) \) (adjusted by means) given \( \epsilon \). Then we have

\[
\rho_{\theta}(\epsilon) = \frac{\text{Cov}(R_1, C_1(\epsilon))}{\sqrt{\text{Var}(R_1) \text{Var}(C_1(\epsilon))}}.
\]

(4.34)

On the other hand, \( \rho_{\theta} = \rho_{\theta}(0) \), hence
\[ \rho_s = \frac{\text{Cov}(R_i, C_i(0))}{\sqrt{\text{Var}(R_i) \text{Var}(C_i(0))}} = \frac{\text{Cov}(R_i, C_i)}{\sqrt{\text{Var}(R_i) \text{Var}(C_i)}} = \frac{\text{Cov}(R_i, C_i)}{12 (N^2 - 1)}. \]

Denote by
\[ u_1 = \frac{1}{N(N-1)} \sum_{i \neq j} \text{sgn}(Y_i - Y_j) \text{sgn}(X_i - X_j), \]
\[ u_1(\epsilon) = \frac{1}{N(N-1)} \sum_{i \neq j} \text{sgn}(Y_i - Y_j) \text{sgn}(X_i - X_j) I\{|X_i - X_j| \leq \epsilon\}, \]
\[ u_2(\epsilon) = \frac{1}{N(N-1)} \sum_{i \neq j} I\{|X_i - X_j| \leq \epsilon\}, \]
\[ v_1 = \frac{1}{N(N-1)(N-2)} \sum_{i \neq j \neq l} \text{sgn}(Y_i - Y_j) \text{sgn}(X_i - X_l), \]
\[ v_1(\epsilon) = \frac{1}{N(N-1)(N-2)} \sum_{i \neq j \neq l} \text{sgn}(Y_i - Y_j) \text{sgn}(X_i - X_l) I\{|X_i - X_l| \leq \epsilon\}, \]
\[ v_2(\epsilon) = \frac{1}{N(N-1)(N-2)} \sum_{i \neq j \neq l} \text{sgn}(X_i - X_j) \text{sgn}(X_i - X_l) I\{|X_i - X_l| \leq \epsilon\}, \]
and
\[ v_3(\epsilon) = \frac{1}{N(N-1)(N-2)} \sum_{i \neq j \neq l} \text{sgn}(X_i - X_j) \text{sgn}(X_i - X_l) I\{|X_i - X_j| \leq \epsilon\} I\{|X_i - X_l| \leq \epsilon\}. \]

Then \( u_1, u_1(\epsilon), u_2(\epsilon), v_1, v_1(\epsilon), v_2(\epsilon), \) and \( v_3(\epsilon) \) are U-statistics, which are unbiased estimates of the regular functionals

\[ \tau_1 = \int \int \text{sgn}(Y_1 - Y_2) \text{sgn}(X_1 - X_2) \text{d}H(Y_1, X_1) \text{d}H(Y_2, X_2), \]
\[ \tau_1(\epsilon) = \int \int \text{sgn}(Y_1 - Y_2) \text{sgn}(X_1 - X_2) I\{|X_1 - X_2| \leq \epsilon\} \text{d}H(Y_1, X_1) \text{d}H(Y_2, X_2), \]
\[ \tau_2(\epsilon) = \int \int I\{|X_1 - X_2| \leq \epsilon\} \text{d}H(Y_1, X_1) \text{d}H(Y_2, X_2) = P(1)(\epsilon), \]
\[ \kappa_1 = \int \int \int \text{sgn}(Y_1 - Y_2) \text{sgn}(X_1 - X_3) \text{d}H(Y_1, X_1) \text{d}H(Y_2, X_2) \text{d}H(Y_3, X_3), \]
\[ = \int (2 F(y) - 1)(2 G(x) - 1) \text{d}H(y, x), \]
\[ \kappa_1(\epsilon) = \int \int \int \text{sgn}(Y_1 - Y_2) \text{sgn}(X_1 - X_3) I\{|X_1 - X_3| \leq \epsilon\} \text{d}H(Y_1, X_1) \text{d}H(Y_2, X_2) \text{d}H(Y_3, X_3), \]
\[ = \int (2 F(y) - 1)(2 G(x) - G(x+\epsilon) - G(x-\epsilon)) \text{d}H(y, x), \]
\[ \kappa_2(\epsilon) = \int \int \int \text{sgn}(X_1 - X_2) \text{sgn}(X_1 - X_3) I\{|X_1 - X_3| \leq \epsilon\} \text{d}G(X_1) \text{d}G(X_2) \text{d}G(X_3) \]
\[ = \int (2 F(y) - 1)(2 G(x) - G(x+\epsilon) - G(x-\epsilon)) \text{d}H(y, x), \]
\[ \kappa_2(\epsilon) = \int \int \int \text{sgn}(X_1 - X_2) \text{sgn}(X_1 - X_3) I\{|X_1 - X_3| \leq \epsilon\} \text{d}G(X_1) \text{d}G(X_2) \text{d}G(X_3) \]
\[
= \int (2 \ G(x) - 1)(2 \ G(x) - G(x + \epsilon) - G(x - \epsilon)) \ dG(x),
\]

and

(4.49)

\[
\kappa_3(\epsilon) = \int \int \text{sgn}(X_1 - X_2) \ \text{sgn}(X_1 - X_3) \ I\{|X_1 - X_2| \leq \epsilon\} \ I\{|X_1 - X_3| \leq \epsilon\} \ dG(X_1) \ dG(X_2) \ dG(X_3)
\]

\[
= \int [G(x + \epsilon) + G(x - \epsilon) - 2 G(x)]^2 \ dG(x)
\]

of degrees 2, 2, 3, 3, 3, and 3 respectively, where H(y, x) is the joint distribution function of (Y, X) and F(y), G(x) are the distribution functions of Y and X respectively. Since

(4.50)

\[
\begin{aligned}
C_i(\epsilon) &= \frac{1}{2} \sum_j \text{sgn}(X_i - X_j) \ (1 - I\{|X_i - X_j| \leq \epsilon\}) \\
R_i &= \frac{1}{2} \sum_j \text{sgn}(Y_i - Y_j)
\end{aligned}
\]

we have

(4.51)

\[
\text{Cov}(R_i, C_i(\epsilon)) = \text{E}(R_i \ C_i(\epsilon)) - \text{E}(R_i) \ \text{E}(C_i(\epsilon)) = \text{E}(R_i \ C_i(\epsilon))
\]

\[
= \text{E}(\frac{1}{2} \sum_j \text{sgn}(Y_i - Y_j) \ \frac{1}{2} \sum_l \text{sgn}(X_i - X_l) \ (1 - I\{|X_i - X_l| \leq \epsilon\}))
\]

\[
= \text{E}(\frac{1}{2} \sum_j \text{sgn}(Y_i - Y_j) \ \frac{1}{2} \sum_l \text{sgn}(X_i - X_l))
\]

\[
- \text{E}(\frac{1}{2} \sum_j \text{sgn}(Y_i - Y_j) \ \frac{1}{2} \sum_l \text{sgn}(X_i - X_l) \ I\{|X_i - X_l| \leq \epsilon\})
\]

\[
= \text{E}(R_i C_i) - \frac{1}{4} \ (N - 1) \tau_1(\epsilon) - \frac{1}{4} \ (N - 1)(N - 2) \ \kappa_1(\epsilon)
\]

\[
= \frac{1}{12} \ (N^2 - 1) \ \rho_8 - \frac{1}{4} \ (N - 1) \ [(N - 2) \ \kappa_1(\epsilon) + \tau_1(\epsilon)].
\]

Also,

(4.52)

\[
\text{Var}(C_i(\epsilon)) = \text{E}(C_i^2(\epsilon)) - \text{E}^2(C_i(\epsilon)) = \text{E}(C_i^2(\epsilon))
\]

\[
= \text{E}(\frac{1}{2} \sum_j \text{sgn}(X_i - X_j) \ (1 - I\{|X_i - X_j| \leq \epsilon\}) \ \frac{1}{2} \sum_l \text{sgn}(X_i - X_l)(1 - I\{|X_i - X_l| \leq \epsilon\}))
\]

\[
= \text{E}(\frac{1}{2} \sum_j \text{sgn}(X_i - X_j) \ \frac{1}{2} \sum_l \text{sgn}(X_i - X_l))
\]

\[
- 2 \ \text{E}(\frac{1}{2} \sum_j \text{sgn}(X_i - X_j) \ \frac{1}{2} \sum_l \text{sgn}(X_i - X_l) \ I\{|X_i - X_l| \leq \epsilon\})
\]

55
\[ + E \left( \frac{1}{2} \sum_{i} \text{sgn} (X_i - X_j) \mathbb{I}\{ |X_i - X_j| \leq \varepsilon \} \right) \frac{1}{2} \sum_{i} \text{sgn} (X_i - X_j) \mathbb{I}\{ |X_i - X_j| \leq \varepsilon \} \]
\[ = \frac{1}{12} (N^2 - 1) - \frac{1}{4} (N - 1) \left[ (N - 2) (2 \kappa_2(\varepsilon) - \kappa_3(\varepsilon)) + P(1)(\varepsilon) \right]. \]

Hence, we have
\begin{equation}
\rho_\varepsilon = \frac{\text{Cov}(R_1, C_1(\varepsilon))}{\sqrt{\text{Var}(R_1) \text{Var}(C_1(\varepsilon))}}
\end{equation}
\begin{equation}
= - \frac{\frac{1}{12} (N^2 - 1) \rho_\varepsilon - \frac{1}{4} (N - 1) [(N - 2) \kappa_1(\varepsilon) + \tau_1(\varepsilon)]}{\sqrt{\frac{1}{12} (N^2 - 1) \left[ (N - 2) (2 \kappa_2(\varepsilon) - \kappa_3(\varepsilon)) + P(1)(\varepsilon) \right]}}
\end{equation}
\begin{equation}
= - \frac{\rho_\varepsilon - \frac{3}{N+1} [(N - 2) \kappa_1(\varepsilon) + \tau_1(\varepsilon)]}{\sqrt{1 - \frac{3}{N+1} [(N - 2) (2 \kappa_2(\varepsilon) - \kappa_3(\varepsilon)) + P(1)(\varepsilon)]}}.
\end{equation}

When \( Y \) and \( X \) are independent, \( \rho_\varepsilon = \kappa_1(\varepsilon) = \tau_1(\varepsilon) = 0 \). Thus, \( \rho_\varepsilon(\varepsilon) = 0 \). Also, \( \lim_{\varepsilon \downarrow 0} \tau_2(\varepsilon) = \lim_{\varepsilon \downarrow 0} \kappa_1(\varepsilon) = \lim_{\varepsilon \downarrow 0} \kappa_2(\varepsilon) = \lim_{\varepsilon \downarrow 0} \kappa_3(\varepsilon) = 0 \). Hence,
\begin{equation}
\lim_{\varepsilon \downarrow 0} \rho_\varepsilon(\varepsilon) = \rho_\varepsilon,
\end{equation}
which is as desired. When \( N \to \infty \),
\begin{equation}
\rho_\varepsilon(\varepsilon) \approx \rho_\varepsilon(\varepsilon^{*}(\varepsilon)) = \frac{\rho_\varepsilon - 3 \kappa_1(\varepsilon)}{\sqrt{1 - 6 \kappa_2(\varepsilon) + 3 \kappa_3(\varepsilon)}} = \frac{3(\kappa_1(\varepsilon) - \kappa_1(\varepsilon))}{\sqrt{1 - 6 \kappa_2(\varepsilon) + 3 \kappa_3(\varepsilon)}}.
\end{equation}

We have the following relationship between \( \rho_\varepsilon(\varepsilon) \) and \( \rho_\varepsilon \).

Proposition 4.3.1

If \( g(x) \) is a symmetric unimodal continuous function, then
\begin{equation}
|\rho_\varepsilon(\varepsilon)| \leq |\rho_\varepsilon|.
\end{equation}

Proof: We show that under the assumption, \( \rho_\varepsilon \geq \rho_\varepsilon(\varepsilon) \geq 0 \) when \( \rho_\varepsilon \geq 0 \). Since \( \rho_\varepsilon(0) = \rho_\varepsilon \). It suffices to show that \( \kappa_1 \geq \kappa_1(\varepsilon) \geq 0 \), and \( \rho_\varepsilon(\varepsilon) \) is decreasing in \( \varepsilon \), \( \forall \varepsilon \geq 0 \). Now, \( \kappa_1(0) = 0 \), \( \kappa_1(\infty) = \kappa_1 \), and
\[ \kappa_1'(\varepsilon) = \int \int (2F(y) - 1)(-g(x+\varepsilon) + g(x-\varepsilon)) \, dH(y, x) \geq 0 \]
under the assumption and \( E(\Psi(y)\gamma(x)) \geq 0 \), where \( \Psi(y) = 2F(y) - 1 \), and \( \gamma(x) = -g(x+\varepsilon) + \)
$g(x-\epsilon)$, since $\Psi(y)$ is increasing in $y$ and $Y$ and $X$ are positively correlated. Thus, $\kappa_1(\epsilon)$ increases from 0 to $\kappa_1$ as $\epsilon$ increases from 0 to $+\infty$. Also,

$$K(\epsilon) = 1 - 6\kappa_2(\epsilon) + 3\kappa_3(\epsilon) = 3 \int (G(x+\epsilon) + G(x-\epsilon) - 1)^2 \, dG(x).$$

Thus,

$$K'(\epsilon) = 6 \int (G(x+\epsilon) + G(x-\epsilon) - 1)(g(x+\epsilon) - g(x-\epsilon)) \, dG(x) \leq 0$$

under the assumption of symmetric unimodal of $g(x)$. Therefore,

$$\rho_S^{**}(\epsilon) = \frac{-6\kappa_1'(\epsilon) K(\epsilon) + (\kappa_1 - \kappa_1(\epsilon)) K'(\epsilon)}{K(\epsilon)^{3/2}} \leq 0.$$ 

Hence, $\rho_S^{**}(\epsilon)$ is decreasing in $\epsilon$, and $\rho_S^{**}(\epsilon) \geq 0$, $\forall \epsilon \geq 0$. The case when $\rho_S \leq 0$ follows similarly.

We have the present result. \qed

Under the normal case,

(4.57)

$$\rho_p = 2 \sin \frac{\pi}{6} \rho_S$$

$$= 2 \sin \frac{\pi}{6} \left( \sqrt{1 - \frac{3}{N+1} \left[(N-2)(2\kappa_2(\epsilon) - \kappa_3(\epsilon)) + P_{(1)}(\epsilon)\right]} \rho_S(\epsilon) + \frac{3}{N+1} \left[(N-2)\kappa_1(\epsilon) + \tau_1(\epsilon)\right] \right)$$

$$\simeq 2 \sin \frac{\pi}{6} \left( \sqrt{1 - 6\kappa_2(\epsilon) + 3\kappa_3(\epsilon)} \rho_S(\epsilon) + 3\kappa_1(\epsilon) \right).$$

Hence,

$$\text{ARE} (\text{ERMP} (\tau; \epsilon), \text{AC}) = \frac{3(1 - \rho_p^2)}{\pi(1 - \rho_S^{**}(\epsilon))}$$

can reach $3/\pi$ as $\rho_p = \rho_S(\epsilon) = 0$. When $\epsilon \to 0$, $\kappa_1(\epsilon) \to 0$, $i=1,2,3$, and $\rho_S(\epsilon) \to \rho_S$. The ARE has the same lower bound of $\sqrt{3}/2$ as that in Quade(1967) when $\rho_p$ and $\rho_S(\epsilon) \to \rho_S \pm 1$ in this case. However, the ARE could be very low when $\rho_p \to \pm 1$, while $\epsilon$ is a fixed constant, so that $\rho_S(\epsilon)$ does not go to $\pm 1$. This is because an inappropriate tolerance is used in the analysis in such case. A test statistic to find a reasonable tolerance is given in Chapter 5.

Similar to

(4.58)

$$r_S = \frac{\sum R_i C_i}{\sqrt{\sum R_i^2 \sum C_i^2}} = \frac{\sum R_i C_i}{\frac{1}{12} (N^3 - N)},$$

which is an unbiased estimate of $\rho_S$, we can let
(4.59) \[ r_s(\epsilon) = \frac{\sum_i R_i C_i(\epsilon)}{\sqrt{\sum_i R_i^2 \sum C_i^2(\epsilon)}} \]

be an estimate of \( \rho_s(\epsilon) \). Now,

\[ \sum_i R_i C_i(\epsilon) = \frac{1}{4} \sum_i \sum_j \sum_{l} \text{sgn}(Y_i - Y_j) \text{sgn}(X_i - X_j) (1 - I(|X_i - X_j| \leq \epsilon)) \]

\[ = \sum_i R_i C_i - \frac{1}{4} \sum_i \sum_j \sum_{l} \text{sgn}(Y_i - Y_j) \text{sgn}(X_i - X_l) I(|X_i - X_l| \leq \epsilon), \]

(4.60)

and

\[ \sum_i C_i^2(\epsilon) = \frac{1}{4} \sum_i \left( \sum_j \text{sgn}(X_i - X_j) (1 - I(|X_i - X_j| \leq \epsilon)) \right)^2 \]

\[ = \frac{1}{12} (N^3 - N) - \frac{1}{4} \sum_i \sum_j \sum_{l} \text{sgn}(X_i - X_j) \text{sgn}(X_i - X_l) I(|X_i - X_l| \leq \epsilon) \]

\[ + \frac{1}{4} \sum_i \sum_j \sum_{l} \text{sgn}(X_i - X_j) \text{sgn}(X_i - X_l) I(|X_i - X_l| \leq \epsilon) I(|X_i - X_l| \leq \epsilon). \]

Therefore, we can write

(4.62)

\[ r_s(\epsilon) = \frac{r_s - \frac{3}{N+1} \left( (N - 2) v_1(\epsilon) + u_1(\epsilon) \right)}{\sqrt{1 - \frac{3}{N+1} \left( (N - 2)(2v_2(\epsilon) - v_3(\epsilon)) + u_2(\epsilon) \right)}}. \]

\[ \approx \frac{r_s - 3v_1(\epsilon)}{\sqrt{1 - 6v_2(\epsilon) + 3v_3(\epsilon)}} \]

as \( N \to \infty \). Let

(4.64) \[ V_1(\epsilon) = r_s - \frac{3}{N+1} \left( (N - 2) v_1(\epsilon) + u_1(\epsilon) \right) \]

\[ = \frac{3}{N+1} \left( (N - 2) (v_1 - v_1(\epsilon)) + (u_1 - u_1(\epsilon)) \right), \]

(4.65) \[ \Theta_1(\epsilon) = 3(\kappa_1 - \kappa_1(\epsilon)), \]

(4.66) \[ V_2(\epsilon) = g(v_2(\epsilon), v_3(\epsilon), u_2(\epsilon)) = 1 - \frac{3}{N+1} \left( (N - 2)(2v_2(\epsilon) - v_3(\epsilon)) + u_2(\epsilon) \right), \]

and

(4.67) \[ \Theta_2(\epsilon) = 1 - 6\kappa_2(\epsilon) + 3\kappa_3(\epsilon). \]

Then,

(4.68) \[ r_s(\epsilon) = \frac{V_1(\epsilon)}{\sqrt{V_2(\epsilon)}}, \text{ and } \rho_s^*(\epsilon) = \frac{\Theta_1(\epsilon)}{\Theta_2(\epsilon)}. \]

58
We show, however, that \( r_8(\epsilon) \) is only asymptotically unbiased for \( \rho_{8}^*(\epsilon) \). Now,

\[
V_1(\epsilon) \simeq A(\epsilon) = 3(v_1 - v_1(\epsilon)),
\]

\[
V_2(\epsilon) \simeq B(\epsilon) = 1 - 6v_2(\epsilon) + 3v_3(\epsilon),
\]
as \( N \to \infty \), where \( A(\epsilon) \) and \( B(\epsilon) \) are U-statistics of degree 3. Thus,

\[
E(r_8(\epsilon)) = E\left( \frac{V_1(\epsilon)}{\sqrt{V_2(\epsilon)}} \right) = E\left( \frac{V_1(\epsilon)}{\sqrt{\Theta_2(\epsilon) + \mathcal{O}_p(N^{-1/2})}} \right) \simeq \frac{\Theta_1(\epsilon)}{\sqrt{\Theta_2(\epsilon)}} = \rho_{8}^*(\epsilon)
\]
as \( N \to \infty \).

It is known that \( \sqrt{N}(r_8 - \rho_8) \) has an asymptotic normal distribution (Hoeffding, 1948); we show a similar result for \( \sqrt{N}(r_8(\epsilon) - \rho_{8}(\epsilon)) \). To establish the asymptotic normality of \( r_8(\epsilon) \), first we need to investigate the variances and covariances of the \( A(\epsilon) \) and \( B(\epsilon) \). Now,

\[
\lim_{N \to \infty} N \text{Var}(v_1(\epsilon)) = 9 \zeta_1(\kappa_1(\epsilon)),
\]

where

\[
\zeta_1(\kappa_1(\epsilon)) = E(\Phi_1^2((Y, X); v_1(\epsilon))) - \kappa_1^2(\epsilon),
\]

\[
\Phi_1((y_1, x_1); v_1(\epsilon)) = E(\Phi((y_1, x_1), (Y_2, X_2), (Y_3, X_3); v_1(\epsilon))).
\]

Also,

\[
\lim_{N \to \infty} N \text{Cov}(v_1, v_1(\epsilon)) = 9 \zeta_1(\kappa_1, \kappa_1(\epsilon)),
\]

where

\[
\zeta_1(\kappa_1, \kappa_1(\epsilon)) = E(\Phi_1((Y_1, X_1); v_1)\Phi_1((Y_1, X_1); v_1(\epsilon))) - \kappa_1\kappa_1(\epsilon),
\]

\( \Phi_1((Y_1, X_1); v_1) \) is defined in a similar way as \( \Phi_1((Y_1, X_1); v_1(\epsilon)) \)((4.74)). Therefore,

\[
\lim_{N \to \infty} N \text{Var}(V_1(\epsilon)) = 81(\zeta_1(\kappa_1) + \zeta_1(\kappa_1(\epsilon)) - 2 \zeta_1(\kappa_1, \kappa_1(\epsilon))) .
\]

Since

\[
\sqrt{V_2(\epsilon)} = \sqrt{\Theta_2(\epsilon)} + \frac{3}{2\sqrt{\Theta_2(\epsilon)}} \frac{N-2}{N+1} (-2v_2(\epsilon) - \kappa_2(\epsilon)) + (v_3(\epsilon) - \kappa_3(\epsilon)) + \mathcal{O}_p(N^{-1}),
\]

we have

\[
\lim_{N \to \infty} N\text{Var}(\sqrt{V_2(\epsilon)}) = \lim_{N \to \infty} \frac{N}{4\Theta_2(\epsilon)} \text{Var}(V_2(\epsilon))
\]

\[
= \frac{81}{4\Theta_2(\epsilon)} (4\zeta_1(\kappa_2(\epsilon)) + \zeta_1(\kappa_3(\epsilon)) - 4\zeta_1(\kappa_2(\epsilon), \kappa_3(\epsilon))) ,
\]

59
where $\zeta_1(\kappa_2(\epsilon))$, $\zeta_1(\kappa_3(\epsilon))$, and $\zeta_1(\kappa_2(\epsilon), \kappa_3(\epsilon))$ are defined similar as $\zeta_1(\kappa_1(\epsilon))$ ((4.73)), and $\zeta_1(\kappa_1(\epsilon), \kappa_1(\epsilon))$ ((4.76)). Also,

\[
\lim_{N \to \infty} N \text{ Cov}(V_1(\epsilon), \sqrt{V_2(\epsilon)}) = \lim_{N \to \infty} N \text{ Cov}(V_1(\epsilon), \sqrt{\Theta_2(\epsilon)}) + \frac{3}{2\sqrt{\Theta_2(\epsilon)}} \frac{N-2}{N+1} (-2(\nu_2(\epsilon) - \kappa_2(\epsilon)) + (\nu_3(\epsilon) - \kappa_3(\epsilon))) + o_p(N^{-1})
\]

\[
= \lim_{N \to \infty} \frac{N}{2\sqrt{\Theta_2(\epsilon)}} \text{ Cov}(V_1(\epsilon), V_2(\epsilon))
\]

\[
= \frac{81}{2\sqrt{\Theta_2(\epsilon)}} (2\zeta_1(\kappa_1(\epsilon), \kappa_2(\epsilon)) - \zeta_1(\kappa_1(\epsilon), \kappa_3(\epsilon)) - 2\zeta_1(\kappa_1(\epsilon), \kappa_2(\epsilon)) + \zeta_1(\kappa_1(\epsilon), \kappa_3(\epsilon)))
\]

where $\zeta_1(\kappa_1(\epsilon), \kappa_2(\epsilon))$, $\zeta_1(\kappa_1(\epsilon), \kappa_3(\epsilon))$, $\zeta_1(\kappa_1(\epsilon), \kappa_2(\epsilon))$, and $\zeta_1(\kappa_1(\epsilon), \kappa_3(\epsilon))$ are defined similar as $\zeta_1(\kappa_1(\epsilon), \kappa_1(\epsilon))$ ((4.76)).

In general, we have the following result.

**Theorem 4.3.2**

The $r_6(\epsilon)$ defined in (4.63) has asymptotically a normal distribution. Namely, $\sqrt{N}(r_6(\epsilon) - \rho_6(\epsilon))$ is asymptotically normally distributed with mean zero and variance

\[
\sigma^2(\epsilon) = \frac{81}{\Theta_2(\epsilon)} (\zeta_1(\kappa_1(\epsilon)) + \zeta_1(\kappa_1(\epsilon), \kappa_1(\epsilon)) - 2 \zeta_1(\kappa_1(\epsilon), \kappa_1(\epsilon))) + \frac{81}{\Theta_2^3(\epsilon)} (4\zeta_1(\kappa_2(\epsilon)) + \zeta_1(\kappa_3(\epsilon)) - 4\zeta_1(\kappa_2(\epsilon), \kappa_3(\epsilon)) - 81\Theta_1(\epsilon) \frac{\Theta_2^2(\epsilon)}{\Theta_2^3(\epsilon)} (2\zeta_1(\kappa_1(\epsilon), \kappa_2(\epsilon)) - \zeta_1(\kappa_1(\epsilon), \kappa_3(\epsilon)) - 2\zeta_1(\kappa_1(\epsilon), \kappa_2(\epsilon)) + \zeta_1(\kappa_1(\epsilon), \kappa_3(\epsilon)))
\]

**Proof:**

\[
\sqrt{N}(r(\epsilon) - \rho_6(\epsilon)) = \sqrt{N} \left( \frac{V_1(\epsilon)}{\sqrt{V_2(\epsilon)}} - \frac{\Theta_1(\epsilon)}{\sqrt{\Theta_2(\epsilon)}} \right) = \sqrt{N} \left( \frac{V_1(\epsilon)\sqrt{\Theta_2(\epsilon)} - \Theta_1(\epsilon)\sqrt{V_2(\epsilon)}}{V_1(\epsilon)\sqrt{V_2(\epsilon)} - \Theta_1(\epsilon)\sqrt{V_2(\epsilon)}} \right)
\]

has the same asymptotic distribution as

\[
\sqrt{N} \left( \frac{V_1(\epsilon)\sqrt{\Theta_2(\epsilon)} - \Theta_1(\epsilon)\sqrt{V_2(\epsilon)}}{\Theta_2(\epsilon)} \right)
\]

by Slutsky's Theorem. But from (4.78) and the same theorem, the asymptotic distribution must be the same as

\[
\sqrt{N} \left( \frac{V_1(\epsilon)\sqrt{\Theta_2(\epsilon)} - \Theta_1(\epsilon)\sqrt{V_2(\epsilon)}}{2\sqrt{\Theta_2(\epsilon)} V_2(\epsilon)} \right),
\]

60
which has an asymptotic normal distribution with mean zero and variance

\[
\frac{81}{\Theta_2(\varepsilon)} (\zeta_1(\kappa_1) + \zeta_1(\kappa_1(\varepsilon)) - 2 \zeta_1(\kappa_1, \kappa_1(\varepsilon))) + \\
\frac{81\Theta_1(\varepsilon)}{4\Theta_2(\varepsilon)} (4\zeta_1(\kappa_2(\varepsilon)) + \zeta_1(\kappa_3(\varepsilon)) - 4\zeta_1(\kappa_2(\varepsilon), \kappa_3(\varepsilon))) - \\
\frac{81\Theta_1(\varepsilon)}{\Theta_2(\varepsilon)} (2\zeta_1(\kappa_1(\varepsilon), \kappa_2(\varepsilon)) - \zeta_1(\kappa_1(\varepsilon), \kappa_3(\varepsilon)) - 2\zeta_1(\kappa_1, \kappa_2(\varepsilon)) + \zeta_1(\kappa_1, \kappa_3(\varepsilon)))
\]

((4.77), (4.79), (4.80)) following standard U-statistic theorem; and hence the present theorem. □

4.4 Estimation of the Optimal Weight \( \tau \)

Let \( C(\varepsilon) = (C_1(\varepsilon_1), \ldots, C_p(\varepsilon_p)) \) be the \( N \times p \) matrix of extended ranks of the \( p \) covariates among the \( N \) observations given tolerance \( \varepsilon \), let also

\[
D(\ C(\varepsilon) \ ) = \begin{bmatrix}
||C_1(\varepsilon_1)|| & \ldots & 0 \\
0 & ||C_2(\varepsilon_2)|| & \ldots & 0 \\
0 & 0 & ||C_3(\varepsilon_3)|| & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \ldots & \ldots & \ldots & ||C_p(\varepsilon_p)||
\end{bmatrix}
\]

(4.82)

where \( ||C_l(\varepsilon_l)|| \) is the \( l_2 \)-norm ( \( ||C_l(\varepsilon_l)||^2 = C_l(\varepsilon_l)'C_l(\varepsilon_l) \) ) of \( C_l(\varepsilon_l) \), \( 1 \leq l \leq p \). Then the obvious estimates of \( \Sigma(\varepsilon) \) and \( \eta(\varepsilon) \) are

\[
\hat{\Sigma}(\varepsilon) = (C(\varepsilon) \ D^{-1}(C(\varepsilon)))' (C(\varepsilon) \ D^{-1}(C(\varepsilon)))
\]

(4.83)

\[
= D^{-1}(C(\varepsilon)) \ C'(\varepsilon) \ C(\varepsilon) \ D^{-1}(C(\varepsilon))
\]

and

\[
\hat{\eta}(\varepsilon) = \frac{1}{\sqrt{N(N^2 - 1)/12}} (C(\varepsilon) \ D^{-1}(C(\varepsilon)))'R
\]

(4.84)

\[
= \frac{1}{\sqrt{N(N^2 - 1)/12}} D^{-1}(C(\varepsilon)) \ C'(\varepsilon) \ R
\]

Hence, if \( C'(\varepsilon) \ C(\varepsilon) \) is nonsingular, one can take as the vector of optimal weights

\[
\hat{\tau} = \hat{\Sigma}^{-1}(\varepsilon) \hat{\eta}(\varepsilon) = \frac{1}{\sqrt{N(N^2 - 1)/12}} D(C(\varepsilon)) (C'(\varepsilon) \ C(\varepsilon))^{-1} C'(\varepsilon) \ R.
\]

(4.85)
Since $\tilde{\tau}$ is a continuous function of U-statistics $\tilde{X}(\epsilon)$ and $\tilde{\eta}(\epsilon)$ for each corresponding element which is symmetric in all the N observations, we have

$$E(\psi((Y_{ij}, X_{ij}), (Y_{ij'}, X_{ij'}); \tau) - \psi((Y_{ij}, X_{ij}), (Y_{ij'}, X_{ij'}); \tilde{\tau}))^2$$

$$= E(\sum_{l=1}^{p} (\tilde{\tau}_l - \tau_l) \phi(X_{ij}^{(l)} - X_{ij'}^{(l)}; \epsilon_l))^2$$

$$\leq \frac{1}{4} E(\sum_{l=1}^{p} (\tilde{\tau}_l - \tau_l))^2 = O(\frac{1}{N}).$$

Hence, Assumption D of Quade (1966) is satisfied and it follows by Theorem 8 of that paper that VR($\hat{\tau}; \epsilon$) has an asymptotic F-distribution with $(k - 1, N - k)$ degrees of freedom under the hypothesis.

In the univariate case,

$$\tilde{\tau} = \frac{1}{\sqrt{N(N^2 - 1)/12}} \left( \sum_{i} \sum_{j} C_{ij}^2(\epsilon_1) \right)^{-1/2} \sum_{i} \sum_{j} C_{ij}(\epsilon_1) R_{ij} = r_0(\epsilon_1).$$

Transformed back to weights of non-matched pairs, one should assign weight $1 + r_0(\epsilon_1)$ to the pairs $(y_{ij} > y_{ij'}, x_{ij} < x_{ij'} - \epsilon_1)$ and $1 - r_0(\epsilon_1)$ to the pairs $(y_{ij} < y_{ij'}, x_{ij} > x_{ij'} + \epsilon_1)$. Also, one should assign weight $1 - r_0(\epsilon_1)$ to the pairs $(y_{ij} > y_{ij'}, x_{ij} > x_{ij'} - \epsilon_1)$ and $r_0(\epsilon_1) - 1$ to the pairs $(y_{ij} < y_{ij'}, x_{ij} < x_{ij'} + \epsilon_1)$.

For multivariate covariates, we first estimate the vector of weights $\tau$ by $\tilde{\tau}$, then assign weights to the corresponding combination of pairs according to Proposition 2.2.1 which decomposes the scores $Z_{ij}(\lambda; \epsilon)$ into sums of all possible combinations of pairs.

4.5 ARE of the VR of $Z^*_{ij}(\omega_1, \omega_2; \epsilon)$ With Respect to the VR of $Z_{ij}(\lambda; \epsilon)$

We now turn to the statistic $Z^*_{ij}(\omega_1, \omega_2; \epsilon)$ for the optimal weights $\omega_1$ and $\omega_2$ in the univariate case. Note that $Z^*_{ij}(0, 0; \epsilon)$ is the ordinary matched pair statistic discussed in Chapter 2 when $k=2$ (two-sample cases). Also, $Z^*_{ij}(1+\lambda, 1-\lambda; \epsilon) = Z_{ij}(\lambda; \epsilon)$ discussed above in the univariate covariate case. Thus, $Z^*_{ij}(\omega_1, \omega_2; \epsilon)$ serves as a connection between ordinary and extended matched-pair analysis. Its ARE properties as a function of $\omega_1$ and $\omega_2$ are worth exploring.

$$Z^*_{ij}(\omega_1, \omega_2; \epsilon) = \sum_{i'} \sum_{j'} \psi((Y_{ij}, X_{ij}), (Y_{ij'}, X_{ij'}); \omega_1, \omega_2; \epsilon)$$

$$= \frac{1}{2} \sum_{i'} \sum_{j'} (\operatorname{sgn}(Y_{ij} - Y_{ij'}) I(\{X_{ij} - X_{ij'} \leq \epsilon\}}$$

$$= 62$$
\[ + \omega_1 [I(Y_{ij}>Y_{i'j'}, X_{ij}<X_{i'j'}-\epsilon) - I(Y_{ij}<Y_{i'j'}, X_{ij}>X_{i'j'}+\epsilon)] \]
\[ + \omega_2 [I(Y_{ij}>Y_{i'j'}, X_{ij}>X_{i'j'}+\epsilon) - I(Y_{ij}<Y_{i'j'}, X_{ij}<X_{i'j'}-\epsilon)] \]
\[ = \sum_{i'} \sum_{j'} [\phi(Y_{ij} - Y_{i'j';} 0) - (\omega_1 - 1)\phi(X_{ij} - X_{i'j';} \epsilon)] \]
\[ + \frac{1}{2} (\omega_2 + \omega_1 - 2) \sum_{i'} \sum_{j'} [(I(Y_{ij}>Y_{i'j'}, X_{ij}>X_{i'j'}+\epsilon) - I(Y_{ij}<Y_{i'j'}, X_{ij}<X_{i'j'}-\epsilon)] \]
\[ = Z_{ij} (\omega_1 - 1; \epsilon) + (\omega_2 + \omega_1 - 2) \sum_{i'} \sum_{j'} \varphi((Y_{ij}, X_{ij}), (Y_{i'j'}, X_{i'j'}); \epsilon), \]

where
\[ (4.88) \]
\[ \varphi((Y_{ij}, X_{ij}), (Y_{i'j'}, X_{i'j'}); \epsilon) = \frac{1}{2} [I(Y_{ij}>Y_{i'j'}, X_{ij}>X_{i'j'}+\epsilon) - I(Y_{ij}<Y_{i'j'}, X_{ij}<X_{i'j'}-\epsilon)]. \]

Let
\[ (4.89) \]
\[ \gamma_{ii'} = \frac{1}{2} \left[ P(Y_{i}>Y_{i'}, X_{i}>X_{i'}+\epsilon) - P(Y_{i}<Y_{i'}, X_{i}<X_{i'}-\epsilon) \right], \]
and
\[ (4.90) \]
\[ \gamma_i = \sum_{i'} \alpha_{ii'}, \gamma_{ii'}. \]

Then
\[ (4.91) \]
\[ E(Z_{ij}^*(\omega_1, \omega_2; \epsilon)) = E(Z_{ij}(\omega_1 - 1; \epsilon) + (\omega_2 + \omega_1 - 2) E(\varphi((Y_{ij}, X_{ij}), (Y_{i'j'}, X_{i'j'}); \epsilon)) \]
\[ = N[\theta + (\omega_2 + \omega_1 - 2)\gamma_i]. \]

Also,
\[ (4.92) \]
\[ \text{Var}(Z_{ij}^*(\omega_1, \omega_2; \epsilon)) = \text{Var}(Z_{ij}(\omega_1 - 1; \epsilon) + (\omega_2 + \omega_1 - 2) \sum_{i'} \sum_{j'} \varphi((Y_{ij}, X_{ij}), (Y_{i'j'}, X_{i'j'}); \epsilon)) \]
\[ = \frac{N^2 - 1}{12} \left[ \sigma_1^2 - 2(\omega_1 - 1) \eta_1(\epsilon) + (\omega_1 - 1)^2 \Sigma(\epsilon) \right] \]
\[ + 2 (\omega_2 + \omega_1 - 2) \text{Cov}(Z_{ij}(\omega_1 - 1; \epsilon), U_{ij}(\epsilon)) + (\omega_2 + \omega_1 - 2)^2 \text{Var}(U_{ij}(\epsilon)), \]

where
\[ U_{ij}(\epsilon) = \sum_{i'} \sum_{j'} \varphi((Y_{ij}, X_{ij}), (Y_{i'j'}, X_{i'j'}); \epsilon). \]

Thus,
\[ (4.93) \]
\[ \text{Var}(Z_{ij}^*(\omega_1 + 1, \omega_2; \epsilon)) \]
\[ = (N^2 - 1)\xi_i(\omega_1; \epsilon) + 2(\omega_2 + \omega_1 - 1) \text{Cov}(Z_{ij}(\omega_1; \epsilon), U_{ij}(\epsilon)) + (\omega_2 + \omega_1 - 1)^2 \text{Var}(U_{ij}(\epsilon)) \]

63
\[(N^2 - 1) \left( \xi_i(\omega_1; \epsilon) + 2\omega \nu_i(\omega_1; \epsilon) + \omega^2 \sigma_{U_i}^2(\epsilon) \right) \]

\[= (N^2 - 1) \xi^*(\omega_1, \omega_2; \epsilon),\]

where

\[\nu_i(\omega_1; \epsilon) = (N^2 - 1)^{-1}\text{Cov}(Z_{ij}(\omega_1; \epsilon), U_{ij}(\epsilon)),\]

\[\sigma_{U_i}^2(\epsilon) = (N^2 - 1)^{-1}\text{Var}(U_{ij}(\epsilon)),\]

and

\[\omega = \omega_2 + \omega_1 - 1.\]

Thus, under \(H_0\),

\[(4.94) \quad \text{Var}(Z_{ij}^*(\omega_1 + 1, \omega_2; \epsilon)) = (N^2 - 1) \left( \xi^*(\omega_1, \omega_2; \epsilon) + 2\omega \nu(\omega_1; \epsilon) + \omega^2 \sigma_{U_i}^2(\epsilon) \right) \]

\[= (N^2 - 1) \xi^*(\omega_1, \omega_2; \epsilon),\]

where \(\nu(\omega_1; \epsilon)\) and \(\sigma_{U_i}^2(\epsilon)\) are the common values of \(\nu_i(\omega_1; \epsilon)\) and \(\sigma_{U_i}^2(\epsilon)\) under \(H_0\). Also, under \(H_N\),

\[(4.95) \quad \text{Var}(Z_{ij}^*(\omega_1 + 1, \omega_2; \epsilon)) = (N^2 - 1) \left( \xi^*(\omega_1, \omega_2; \epsilon) + O\left(\frac{1}{\sqrt{N}}\right) \right),\]

and

\[\theta_{ii'}^* = \frac{\delta_{ii'}}{\sqrt{N}} + \omega \frac{\gamma_{ii'}}{\sqrt{N}} + o\left(\frac{1}{\sqrt{N}}\right).\]

Let

\[(4.96) \quad \text{VR}(\omega_1 + 1, \omega_2; \epsilon) = \frac{(N - k) \sum_i (\sum_j Z_{ij}^*(\omega_1 + 1, \omega_2; \epsilon))^2 / n_i}{(k - 1) \left[ \sum_i \sum_j Z_{ij}^2(\omega_1 + 1, \omega_2; \epsilon) - \sum_i (\sum_j Z_{ij}^*(\omega_1 + 1, \omega_2; \epsilon))^2 / n_i \right]} .\]

Then, since

\[| \psi( (Y_{ij}, X_{ij}), (Y_{ij}', X_{ij}'); \omega_1, \omega_2; \epsilon) | \leq \frac{1}{2} (1 + |\omega_1| + |\omega_2|) ,\]

Assumption B2 of Quade (1966) is satisfied, and hence Assumption B1 is satisfied. Also, \(\xi^*(\omega_1, \omega_2; \epsilon) > 0\) for a random observation from the average population. Thus, following Theorems 6 and 7 of that paper, \(\text{VR}(\omega_1 + 1, \omega_2; \epsilon)\) has an asymptotic F distribution with \((k - 1, N - k)\) degrees of freedom and noncentrality parameter

\[(4.97) \quad \Delta^*(\omega_1 + 1, \omega_2; \epsilon) = \sum_i \alpha_i \delta_i^* / \xi^*(\omega_1, \omega_2; \epsilon) ,\]

where

\[\delta_i^* = \sum_{i'} \alpha_{i'} (\delta_{ii'} + \omega \gamma_{ii'}).\]

Hence,
\[ \text{ARE}(\text{EMP}(\omega_1+1, \omega_2; \epsilon), \text{ERMP}(\omega_1; \epsilon)) = \frac{\Delta^*(\omega_1+1, \omega_2; \epsilon)}{\Delta(\omega_1; \epsilon)} \]

\[ = \frac{\sum \alpha_i \delta_{i}^* \xi(\omega_1; \epsilon)}{\sum \alpha_i \delta_{i}^2} \ast \frac{\xi^*(\omega_1, \omega_2; \epsilon)}{\xi(\omega_1; \epsilon)} \]

\[ = \frac{\sum \alpha_i (\sum \alpha_i (\delta_{ii} + \omega \gamma_{ii}))^2}{\sum \alpha_i \delta_{i}^2} \ast \frac{\xi(\omega_1; \epsilon)}{\xi(\omega_1; \epsilon) + 2\omega \nu(\omega_1; \epsilon) + \omega^2 \sigma_U^2(\epsilon)} \]

\[ = \frac{1 + 2\omega \sum \alpha_i \delta_{ii} + \omega^2 \sum \alpha_i \gamma_{i}^2}{1 + 2\omega \nu(\omega_1; \epsilon) + \omega^2 \sigma_U^2(\epsilon)} \]

\[ = 1 + a \omega + b \omega^2 \quad \text{where} \]

\[ a = \frac{\sum \alpha_i \gamma_{i}}{\sum \alpha_i \delta_{i}^2}, \quad b = \frac{\sum \alpha_i \gamma_{i}^2}{\sum \alpha_i \delta_{i}^2}, \quad \epsilon = \frac{\nu(\omega_1; \epsilon)}{\xi(\omega_1; \epsilon)} \quad \text{and} \quad d = \frac{\sigma_U^2(\epsilon)}{\xi(\omega_1; \epsilon)} \]

Hence, if \( \omega = 0 \), then \( A = 1 \), which is as expected. Also, \( A > 1 \) if and only if \( 1 + 2a\omega + b\omega^2 > 1 + 2c\omega + d\omega^2 \). After some algebra, one can get the result that, \( A > 1 \) if and only if \( (\omega > 0 \text{ and } \omega > 2 \frac{\epsilon - a}{b - d}) \) or \( (\omega < 0 \text{ and } \omega < 2 \frac{\epsilon - a}{b - d}) \). Further exploration of the optimal value of \( A \) is of interest.

### 4.6 Category Matching Revisited

We have compared the relative efficiency of caliper matching with respect to category matching for two-sample cases in Section 3.5. In this section, we compare the relative efficiency of extended rank-matched pair analysis with respect to category matching for more general k-sample cases. For simplicity, we consider only univariate covariables. The situation can be extended immediately to p covariables. Following the same setting as in Section 3.5, assume the covariable \( X \) takes values on the finite support \( (x_0, x_m] \) with probabilities \( P\{ X \in (x_{l-1}, x_l] \} = q_l, \sum q_l = 1 \). Define scores

\[ C_{ij}(\omega; \epsilon) = \sum_{i'} \sum_{j'} \sum_l \omega_l \text{sgn}(Y_{ij} - Y_{i'j'}) \mathbb{I}\{ X_{ij}, X_{i'j'} \in (x_{l-1}, x_l] \}, \]

where \( \omega = (\omega_1, \ldots, \omega_m)' \). Then, clearly the scores \( C_{ij}(\omega; \epsilon) \) are interchangeable random variables. Similar to (4.1), let

\[ \pi_{ii', l} = \frac{1}{2} E(\text{sgn}(Y_{ij} - Y_{i'j'})\mathbb{I}\{ X_{ij}, X_{i'j'} \in (x_{l-1}, x_l] \}) \]

\[ = \frac{1}{2} \text{sgn}(Y_{ij} - Y_{i'j'})\mathbb{I}\{ X_{ij}, X_{i'j'} \in (x_{l-1}, x_l] \}) \]
\[
\pi_{i,l} = \sum_{i'} \alpha_{i'} \pi_{i',l}.
\]

Then,\(\;\)
\[
E(C_{ij}(\omega ; \epsilon)) = \sum_{i'} \sum_{j'} \sum_{l} E(\omega_l \; \text{sgn}(Y_{ij} - Y_{i'j'})) I\{X_{ij}, X_{ij'} \in (x_{l-1}, x_l)\})
\]
\[
= \sum_{l} \omega_l E(U_{i,l})
\]
\[
= \sum_{i'} \sum_{j'} \sum_{l} \omega_l (P\{Y_{ij} > Y_{i'j'}, X_{ij}, X_{ij'} \in (x_{l-1}, x_l)\}) - P\{Y_{ij} < Y_{i'j'}, X_{ij}, X_{ij'} \in (x_{l-1}, x_l)\})
\]
\[
= 2 \sum_{i'} \sum_{j'} \sum_{l} \omega_l \pi_{i',l} = 2 N \sum_{l} \omega_l \pi_{i,l},
\]

where\(\;\)
\[
U_{i,l} = \sum_{i'} \sum_{j'} \text{sgn}(Y_{ij} - Y_{i'j'}) I\{X_{ij}, X_{ij'} \in (x_{l-1}, x_l)\}.
\]

Also,\(\;\)
\[
\text{Var}(C_{ij}(\omega ; \epsilon)) = \sum_{l} \omega_l^2 \text{Var}(U_{i,l}) + \sum_{l \neq l'} \omega_l \omega_{l'} \text{Cov}(U_{i,l}, U_{i,l'})
\]

Since \(X_{ij}\) cannot belong in \((x_{l-1}, x_l)\) and \((x_{l'-1}, x_{l'})\) at the same time, \(E(U_{i,l} \cdot U_{i,l'}) = 0\).

Thus,\(\;\)
\[
\text{Cov}(U_{i,l}, U_{i,l'}) = -E(U_{i,l}) E(U_{i,l'}) = -N^2 \pi_{i,l} \pi_{i,l'}.
\]

Given \(X\), the marginal distribution of \(Y\) for the \(i\)-th treatment group is \(F_i(y|x)\), we have

\[
\text{Var}(C_{ij}(\omega ; \epsilon)) = \sum_{l} \omega_l^2 [E(U_{i,l}^2) - E(U_{i,l})^2] - \sum_{l \neq l'} \omega_l \omega_{l'} N^2 \pi_{i,l} \pi_{i,l'}
\]
\[
= \sum_{l} \omega_l^2 \left\{ \sum_{i_1, i_2, l_1, l_2} n_{i_1} n_{i_2} \left[ \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} F_{i_1}(y|x_1) F_{i_2}(y|x_2) dG(x_1) dG(x_2) dG(x_3) 
\right.
\]
\[
\left. - 2 \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} (1 - F_{i_1}(y|x_1)) F_{i_2}(y|x_2) dF_i(y|x_3) dG(x_1) dG(x_2) dG(x_3) 
\right]
\]
\[
+ \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} (1 - F_{i_1}(y|x_1)) (1 - F_{i_2}(y|x_2)) dF_i(y|x_3) dG(x_1) dG(x_2) dG(x_3) \right\}
\]

66
\[ + 2 \sum_{i=1}^{x_l} n_{i_2}(n_1 - 1) \left[ \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} F_1(y|x_1) F_2(y|x_2) dF_1(y|x_3) dG(x_1) dG(x_2) dG(x_3) \\
- 2 \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} (1 - F_1(y|x)) F_{i_2}(y|x) dF_1(y|x) dG(x_1) dG(x_2) dG(x_3) \\
+ \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} (1 - F_1(y|x_1)) (1 - F_{i_2}(y|x_2)) dF_1(y|x_3) dG(x_1) dG(x_2) dG(x_3) \right] \\
+ \sum_{i=1}^{x_l} n_1 q_i^2 + (n_1 - 1) q_i^2 \\
+ \sum_{i=1}^{x_l} n_1 (n_1 - 1) \left[ \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} F_1(y|x_1) F_1(y|x_2) dF_1(y|x_3) dG(x_1) dG(x_2) dG(x_3) \\
- 2 \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} (1 - F_1(y|x)) F_1(y|x) dF_1(y|x_2) dG(x_1) dG(x_2) dG(x_3) \\
+ \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} (1 - F_1(y|x_1)) (1 - F_1(y|x_2)) dF_1(y|x_3) dG(x_1) dG(x_2) dG(x_3) \right] \\
+ (n_1 - 1)(n_1 - 2) \left[ \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} F_1(y|x_1) F_1(y|x_2) dF_1(y|x_3) dG(x_1) dG(x_2) dG(x_3) \\
- 2 \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} (1 - F_1(y|x)) F_1(y|x) dF_1(y|x_2) dG(x_1) dG(x_2) dG(x_3) \\
+ \int_{x_{l-1}}^{x_l} \int_{x_{l-1}}^{x_l} (1 - F_1(y|x_1)) (1 - F_1(y|x_2)) dF_1(y|x_3) dG(x_1) dG(x_2) dG(x_3) \right] - (N \pi_{i,1})^2 \\
- \sum_{l \neq l'} \sum_{i} \omega_i \omega_{i'} N^2 \pi_{i,l} \pi_{i'l'} . \]

Under H_0,

\[ \text{Var}(C_{ij}(\omega \varepsilon)) = \sum_{l} \omega_l^2 \left( \frac{q_l^3}{3} (N^2 - 3N + 2) + (N - 1) q_l^2 \right) \]

\[ = N^2 \left( \frac{1}{3} \sum_{l} \omega_l^2 q_l^3 + O(N^{-1}) \right) . \]

67
Also, under the alternative hypothesis \( H_N \) for each \( N \)

\[
\pi_{ii',t} = \frac{\zeta_{ii',t}}{\sqrt{N}} + o\left(\frac{1}{\sqrt{N}}\right),
\]

we have

\[
\text{Var}(C_{ij}(\omega; \epsilon)) = N^2 \left( \frac{1}{3} \sum_I \omega_I^2 q_I^3 + O\left(\frac{1}{\sqrt{N}}\right) \right).
\]

Since

\[
|\sum_I \omega_I \text{sgn}(Y_{ij} - Y_{ij'}) \mathbb{1}\{X_{ij}, X_{ij'} \in (x_{i-1}, x_i)\}| \leq \sum_I \omega_I,
\]

Assumption B2 of Quade (1966) is satisfied, and thus B1 is satisfied. Also, \( \text{Var}(C_{ij}(\omega; \epsilon)) > 0 \), Assumption C of that paper is satisfied. Again, the variance ratio

\[
\text{VR}(Q(\omega; \epsilon)) = \frac{(N - k) \sum_i (\sum_j C_{ij}(\omega; \epsilon))^2 / n_i}{(k - 1) \left[ \sum_i \sum_j C_{ij}(\omega; \epsilon)^2 - \sum_i (\sum_j C_{ij}(\omega; \epsilon))^2 / n_i \right]}
\]

has an asymptotic \( F \)-distribution with \((k - 1, N - k)\) degrees of freedom, and the noncentrality parameter

\[
\Delta(Q(\omega; \epsilon)) = \frac{4 \sum_i \alpha_i (\sum_I \omega_I \zeta_{i,t})^2}{\frac{1}{3} \sum_I \omega_I^2 q_I^3},
\]

where \( \zeta_{i,t} = \sum_I \alpha_i \zeta_{i,t} \). Thus,

\[
\text{ARE(ERMP}(\tau; \epsilon), Q(\omega; \epsilon)) = \frac{\Delta(\text{ERMP}(\tau; \epsilon))}{\Delta(Q(\omega; \epsilon))}
\]

\[
= \frac{\sum_i \alpha_i \delta_i^2}{12 \left( \sigma^2 - \frac{\text{Cov}^2(R, C(\epsilon))}{\text{Var}(C(\epsilon))} \right)} \frac{\frac{1}{3} \sum_I \omega_I^2 q_I^3}{4 \sum_i \alpha_i (\sum_I \omega_I \zeta_{i,t})^2}
\]

\[
= \frac{1}{1 - \rho^2(\epsilon)} \frac{\sum_I \omega_I^2 q_I^3 \sum_i \alpha_i \delta_i^2}{\sum_i \alpha_i (\sum_I \omega_I \zeta_{i,t})^2}
\]

where \( R \) and \( C(\epsilon) \) are the ranks (corrected by mean) of the response \( Y \) and the covariable \( X \) respectively, and \( \sigma^2 = 1 \) when \( Y \) is continuous. Also, \( \rho^2(\epsilon) \left( = \text{Cov}^2(R, C(\epsilon)) / \text{Var}(C(\epsilon)) \right) \geq 0 \), and \( \sum_I \omega_I^2 q_I^3 \geq (\sum_I \omega_I q_I^2)^2 \) by the Schwartz Inequality. For the optimal \( Q^*(\omega; \epsilon) \) of Chapter 3, when \( \omega_I \) is proportional to \( 1/q_I \), we have

68
\begin{equation}
\text{ARE(ERMP}(\tau; \epsilon), Q^*(\omega; \epsilon)) = \frac{1}{1 - \rho^2(\epsilon)} \cdot \frac{\sum_i \alpha_i \delta_i^2}{\sum_i \alpha_i (\sum_l \zeta_{i,l}/q_l)^2}.
\end{equation}

If \( x_{i,l} \simeq q_l^2 \theta_i \) (\((4.1), (4.50)\)), we then have \( \zeta_{i,l} \simeq q_l^2 \delta_i \), and

\begin{equation}
\text{ARE(ERMP}(\tau; \epsilon), Q^*(\omega; \epsilon)) \simeq \frac{1}{1 - \rho^2(\epsilon)}.
\end{equation}

The ERMP test clearly has better efficiency than the optimal category matching in such case.

4.7 When the Response Variable Is Dichotomous

We show that, when the response variable is dichotomous, the category matching statistic \( Q_{n_1, n_0} \) is equivalent to the Mantel-Haenszel statistic \( Q_{M-H} \), given all the marginals of each stratum being fixed for the two-sample case. Under such conditions, we have

\begin{equation}
Q_{n_1, n_0} = \frac{1}{n_1 n_0} \sum_{l=1}^m \omega_l \sum_{j}^{n_1} \sum_{j'}^{n_0} \text{sgn}(Y_{1j} - Y_{0j'}) \ I\{X_{1j}, X_{0j'} \in (x_{l-1}, x_l)\}
\end{equation}

\begin{align*}
&= \frac{1}{n_1 n_0} \sum_{l=1}^m \omega_l \left( \sum_{j}^{n_1} \sum_{j'}^{n_0} (I\{Y_{1j}=1, Y_{0j'}=0; X_{1j}, X_{0j'} \in (x_{l-1}, x_l)\}
\right. \\
&\quad - I\{Y_{1j}=0, Y_{0j'}=1; X_{1j}, X_{0j'} \in (x_{l-1}, x_l)\}) \\
&= \frac{1}{n_1 n_0} \sum_{l=1}^m \omega_l (a(l) d(l) - b(l) c(l)),
\end{align*}

\begin{center}
\begin{tabular}{ccc}
\hline
\multicolumn{1}{c}{a(l)} & \multicolumn{1}{c}{b(l)} & n_{1l} \\
\hline
\multicolumn{1}{c}{c(l)} & \multicolumn{1}{c}{d(l)} & n_{0l} \\
\hline
m_{1l} & m_{0l} & n_l
\end{tabular}
\end{center}

given all the marginals for the \( l \)-th stratum \( a(l) + b(l) = n_{1l} \), \( c(l) + d(l) = n_{0l} \), \( a(l) + c(l) = m_{1l} \), and \( b(l) + d(l) = m_{0l} \) being fixed. Thus,
\[(4.115) \quad \text{Var}(a(t)d(t) - b(t)c(t)) = \text{Var}(a(t)d(t) - (n_{1l} - a(t))(n_{0l} - d(t))) \]
\[= \text{Var}(a(t)n_{0l} + d(t)n_{1l}) = \text{Var}(a(t)n_{0l} + (m_{0l} - n_{1l} + a(t))n_{1l}) \]
\[= n_l^2 \text{Var}(a(t)) = \frac{n_{1l} n_{0l} m_{1l} m_{0l}}{n_l - 1}, \]

where \(n_l = n_{1l} + n_{0l} = m_{1l} + m_{0l}\) is the total number of observations in the \(l\)-th stratum.
Substituting for the optimal weights \(\omega_l \sim 1/q_l\) their maximum likelihood estimates \(\hat{\omega}_l = n_l/N\), where \(N = \sum_{l=1}^{m} n_l\), we have that
\[\frac{n_{1l} n_{0l}}{N} Q_{n_1}, \quad n_0^* = \sum_{l=1}^{m} \left( \frac{a(t)d(t) - b(t)c(t)}{n_l} \right) / n_l \]
follows a normal distribution with variance \(\sum_{l=1}^{m} \frac{n_{1l} n_{0l} m_{1l} m_{0l}}{n_l^2(n_l - 1)}\). Hence,
\[(4.116) \quad \left( \frac{n_{1l} n_{0l}}{N} Q_{n_1}, n_0^* \right)^2 / \sum_{l=1}^{m} \frac{n_{1l} n_{0l} m_{1l} m_{0l}}{n_l^2(n_l - 1)} \]
\[= \frac{\left( \sum_{l=1}^{m} \left( \frac{a(t)d(t) - b(t)c(t)}{n_l} \right) / n_l \right)^2}{\sum_{l=1}^{m} \frac{n_{1l} n_{0l} m_{1l} m_{0l}}{n_l^2(n_l - 1)}} = \frac{\left( \sum_{l=1}^{m} \left( a(t) - E(a(t)) \right) \right)^2}{\sum_{l=1}^{m} \frac{n_{1l} n_{0l} m_{1l} m_{0l}}{n_l^2(n_l - 1)}} \]
\[= Q_{M-H}, \]

which is the Mantel-Haenszel \(X_1^2\) statistic.

In the case of dichotomous response \(Y\), say, \(u\) of the responses are zero, and \(N - u\) have
response equal to 1. Then the variance of \(R_{ij}\) of the ERMP test reduces to
\[(4.117) \quad \text{Var}(R_{ij}) = \frac{1}{12}(N^2 - 1) - \frac{1}{12N} [(u^3 - u) + (N - u)^3 - (N - u)] = \frac{1}{4} u(N - u) \]
under \(H_0\). A similar situation occurs for the variance of \(C_{ij}(\omega; \epsilon)\). Hence, \(\text{ARE(ERMP}(\tau; \epsilon), Q(\omega; \epsilon))\) remains unchanged.
4.8 Discussion

The test ERMP($\lambda; \epsilon$) which incorporates the "inversions" has an optimal ARE (4.32) equal to unity with respect to the Kruskal-Wallis test when the response $Y$ is uncorrelated with the covariates $X$. This clearly improves the relative efficiency of the unweighted matched pair analysis with respect to the Mann-Whitney statistic discussed in Chapter 3 (3.15). When $Y$ and $X$ are correlated, the relative efficiency always exceeds one. Hence, rather than the usual conclusion that whether to match or not should depend on the strength of the correlation between $Y$ and $X$ and other conditions (Kupper et al, 1981, Miettinen, 1968, 1969), one can always apply ERMP($\lambda; \epsilon$) for a more powerful test of the treatment difference as long as the covariates are concomitant among the treatment groups. In addition, the test ERMP($\lambda; \epsilon$) considers $p$ covariates simultaneously and does not reduce the sample size as conventional matched pair analysis does. Thus, the difficulty of wanting to control for as many variables as possible while lacking enough matching subjects is avoided.

As an extended version of rank analysis of covariance (Quade, 1967), ERMP($\lambda; \epsilon$) can also be treated as multiple regression in a nonparametric setting. Its ARE with respect to parametric ANACOVA (4.33) is similar to that of "rank analysis of covariance", differing only in that the Spearman $R^2_s$ is adjusted for the tolerance $\epsilon$ here. As pointed out by Puri and Sen (1969), the above ARE could be very low in the multiple covariates case when the underlying cdf is normal; and ERMP($\lambda; \epsilon$) could have the same disadvantage. However, there is no obvious way of transforming the extended ranks adjusted for the tolerance $\epsilon$ to generalized normal scores.

The ARE of EMP($\omega_1+1, \omega_2; \epsilon$) with respect to ERMP($\omega_1; \epsilon$) in the univariate case apparently could exceed one under certain circumstances. When $\omega_2=\omega_1-1$, EMP($\omega_1+1, \omega_2; \epsilon$) is equivalent to ERMP($\omega_1; \epsilon$). Further study of EMP($\omega_1+1, \omega_2; \epsilon$) may be required.

When the response is dichotomous, the test ERMP($\lambda; \epsilon$) is still valid with the ranks of the response $Y$ being $\frac{u-N}{2}$ and $\frac{N+u+1}{2}$ for the $u$ zeros and the $N-u$ ones respectively. One application of this is case-control or follow-up studies in epidemiology. In follow-up studies, for example, subjects selected for the exposed and unexposed groups are controlled for some potential confounders to be as balanced as possible. Hence, there is no need to adjust for these variables at the analysis stage on validity grounds.

Discussions on whether to match or not have been centered on the relative efficiency with respect to simple random samples. One typical matching method discussed in literature is category matching rather than caliper matching. We showed in (4.116) that the optimal category matching statistic we derived is equivalent to the familiar Mantel-Haenszel statistic given the marginals of all the strata when the response is dichotomous. Also, (4.113) shows that the test ERMP($\lambda; \epsilon$) tends to be more efficient than category matching. This is true when there
is only one covariable. However, when there are $p$ covariables, stratification starts to sacrifice available sample size, but the ERMP test avoids this difficulty. Thus, together with the findings of its ARE with respect to simple random samples discussed at the beginning of this section, we recommend the use of test $\text{ERMP}(\lambda; \epsilon)$ when issues of bias can be fairly ignored.

As a final comment, owing to the nontransitivity of caliper matching, matching criteria by the $\text{ERMP}(\lambda; \epsilon)$ test would be more complicated than other matching methods at the design stage. Also, when there is doubt concerning validity, stratified analysis may serve as a better tool to remove the bias while $\text{ERMP}(\lambda; \epsilon)$ may no longer be valid.
CHAPTER 5
TEST STATISTIC OF MATCHING TOLERANCE

5.1 Introduction

When the response variable Y and the covariable X are correlated, one may fit a set of parallel regression curves to increase efficiency by adjusting for X in comparing the response Y conditional on X in parametric ANOCOVA. Stratification or matching, on the other hand, provides an alternative way of looking at treatment difference while controlling X within a certain matching interval. As discussed in Section 3.3, changing the scale of matching interval in comparing the stochastic order of the response does not bias the result, but only affects efficiency. Also, though the ERMP test has relatively much better efficiency than category matching ((4.112), (4.113)), in general, it may not be not as efficient as parametric ANOCOVA ((4.31)) or rank ANOCOVA ((4.30), (4.54)). One natural question one might then ask is, “Why bother to do matched analysis when apparently it is inferior to the regression setting of ANOCOVA (parametric or nonparametric) in efficiency?”

One reason is that the model assumptions in ANOCOVA may not necessarily be true. Though one may fit regression curves when Y and X are correlated, in general, this principle is based on that the usual correlation index (product-moment correlation) is a measurement of the tendency of the sample to be concentrated on a straight line (Quade, 1974). Thus, the model assumptions will be violated if the relationship between Y and X is not necessarily monotonic. Certainly the regression setting is not valid when X is nominal. Another reason may be that, intuitively, matching is more acceptable and natural to researchers other than statisticians. In observational studies, other than controlled clinical trials, subjects are often matched with respect to their background characteristics (such as age, sex, historical exposure, etc...) to reduce the bias and to increase the efficiency in comparison of the outcome variable of interest.

While it is not explicitly specified, subjects with matched characteristics are assumed to be “homogeneous” with respect to the matching variables (Mantel and Haenszel, 1959). As a consequence, one may compare their outcomes without further adjusting for the covariables (even if they are continuous). Along this line, we may want to investigate statistically if there really is an interval of X such that the response Y is “insensitive” or “nearly insensitive” to the variation of X within that range.

In this chapter, we propose a notion of “locally uncorrelated” - locally there is no evidence
of correlation, while overall there is evidence of correlation. In Section 5.2, we examine the feasibility of this notion in the standard parametric ANOCOVA setting. In Section 5.3, we propose a statistic using the difference of concordant pairs less the discordant pairs as an index of correlation given tolerance \( \epsilon \), and provide an estimate of the variance of the statistic. A brief discussion is given in Section 5.5.

5.2 Local Lack of Correlation in the Parametric ANOCOVA Setting

We show in this section that the stochastic order of response \( Y \) of two samples can be insensitive to small variation of their covariable \( X \) in the standard parametric ANOCOVA setting under certain situations. Hence, locally the matched pairs given tolerance would show no or little evidence of correlation. Let \((Y_1, X_1)\) and \((Y_0, X_0)\) be observations from the treatment group and the control group respectively. Also, let \( Y_i = \mu + \nu + \beta X_i + E_i \), \( E_i \) i.i.d. \( \sim N(0, \sigma^2) \), \( i=1,0 \). Then, we have

\[
P\{Y_1 > Y_0 \mid \{X_1 - X_0\} \leq \epsilon\} = \frac{\Phi\left(\frac{\beta(x_1 - x_0) + \nu}{\sqrt{2} \sigma}\right) dF(\xi_1) dF(\xi_0)}{\int_{\{X_1 - X_0\} \leq \epsilon} dF(\xi_1) dF(\xi_0)}
\]

\[
= \frac{\int_{-\epsilon}^{\epsilon} \Phi\left(\frac{\beta u + \nu}{\sqrt{2} \sigma}\right) g(u) du}{\int_{-\epsilon}^{\epsilon} g(u) du}
\]

where \( g(u) \) is the density of \( U = X_1 - X_0 \), \( g(u) = g(-u) \). By Taylor’s expansion,

\[
\Phi\left(\frac{\beta u + \nu}{\sqrt{2} \sigma}\right) = \Phi\left(\frac{\nu}{\sqrt{2} \sigma}\right) + \Phi\left(\frac{\nu}{\sqrt{2} \sigma}\right) \cdot \frac{\beta u}{\sqrt{2} \sigma} + \Phi\left(\frac{\xi(u) + \nu}{\sqrt{2} \sigma}\right) \cdot \frac{\beta^2 u^2}{4 \sigma^2},
\]

where \( 0 < |\xi(u)| < |\beta u| \), \( \xi(u) = -\xi(-u) \). Since

\[
|\Phi\left(\frac{\xi(u) + \nu}{\sqrt{2} \sigma}\right) \cdot \frac{\beta^2 (x_1 - x_0)^2}{4 \sigma^2}| \leq \left| \frac{(|\beta| + \nu) \beta^2 \epsilon^2}{8 \sigma^3} \right|,
\]

we have

\[
\Phi\left(\frac{\nu}{\sqrt{2} \sigma}\right) \cdot \frac{\beta^2 (x_1 - x_0)^2}{4 \sigma^2} \leq \Phi\left(\frac{\nu}{\sqrt{2} \sigma}\right) \cdot \frac{\beta^2 \epsilon^2}{8 \sigma^3},
\]

where
\[ P\{Y_1 > Y_0 \mid |X_1 - X_0| \leq \epsilon \} \leq \Phi\left(\frac{\nu}{\sqrt{2} \sigma}\right) + \left| \frac{(|\beta|\epsilon + \nu)\beta^2 \epsilon^2}{8 \sigma^3} \right| = \Phi\left(\frac{\nu}{\sqrt{2} \sigma}\right) + O(\beta^2 \epsilon^2). \]

On the other hand,

\begin{equation}
(5.5) \quad P\{Y_1 > Y_0 \mid X_1 = X_0 \} = \Phi\left(\frac{\nu}{\sigma}\right). 
\end{equation}

Thus, under either of the following two conditions:

(i) if $|\beta|$ is significantly greater than zero, and $\epsilon$ is very small;

(ii) if $\epsilon$ is a fixed constant, and $|\beta|$ is very small,

the response $Y$ is not sensitive to the residual term $O(\beta^2 \epsilon^2)$ of (5.4). When $\epsilon \rightarrow \infty$, neither of the two conditions is satisfied, and $P\{Y_1 > Y_0 \mid X_1 = X_0\}$ may no longer be approximated by $P\{Y_1 > Y_0 \mid |X_1 - X_0| \leq \epsilon\}$ ((5.4) (5.5)).

Condition (i) is the usual case when one has the slope $\beta$ of the two parallel regression lines significantly greater than zero in parametric ANOCOVA. In which case, the response variable $Y$ is highly sensitive to small variation of the covariable $X$. In which case, one cannot assume subjects in matched analysis with large tolerance to be homogeneous with respect to their covariable $X$.

On the other hand, Condition (ii) is satisfied when the variance $\sigma_x^2$ of the covariable $X$ is relatively much larger than the the variance $\sigma_y^2$ of the response variable $Y$, or when the correlation coefficient $\rho$ between $Y$ and $X$ is small, the slope $\beta$ of the regression line is small ($\beta = \rho \sigma_y / \sigma_x$ when $Y$ and $X$ are bivariate normally distributed). This is the case when the scale of $X$ is relatively much larger than $Y$, or when the correlation between $Y$ and $X$ is not strong enough, and therefore $Y$ is not so sensitive to a small variation of $X$. A matched analysis with appropriate tolerance $\epsilon$ might be better than the standard analysis of covariance in this situation, since the latter may not be able to incorporate the covariable as a significant explanatory factor in the regression model. We will come back to this point in one of our examples (Chapter 6).

### 5.3 Test Statistic of Tolerance $\epsilon$

In this section, we develop a test statistic to find a legitimate tolerance directly from a given data set. Let

\begin{equation}
(5.6) \quad M_1(\epsilon) = \sum_{\substack{i=1 \atop j_1 \neq j_2}}^{n_1} I \{ |X_{ij_1} - X_{ij_2}| \leq \epsilon \}, 
\end{equation}

75
(5.7) \[ T_i(\epsilon) = \sum_{j_1 \neq j_2}^{n_i} \text{sgn}(Y_{ij_1} - Y_{ij_2}) \text{sgn}(X_{ij_1} - X_{ij_2}) I\{ |X_{ij_1} - X_{ij_2}| \leq \epsilon \} \]

\[ = C_i(\epsilon) - D_i(\epsilon) , \]

where

(5.8) \[ C_i(\epsilon) = \sum_{j_1 \neq j_2}^{n_i} (I\{Y_{ij_1} > Y_{ij_2}, X_{ij_1} > X_{ij_2}, |X_{ij_1} - X_{ij_2}| \leq \epsilon \} + I\{Y_{ij_1} < Y_{ij_2}, X_{ij_1} < X_{ij_2}, |X_{ij_1} - X_{ij_2}| \leq \epsilon \} , \]

and

(5.9) \[ D_i(\epsilon) = \sum_{j_1 \neq j_2}^{n_i} (I\{Y_{ij_1} > Y_{ij_2}, X_{ij_1} < X_{ij_2}, |X_{ij_1} - X_{ij_2}| \leq \epsilon \} + I\{Y_{ij_1} < Y_{ij_2}, X_{ij_1} > X_{ij_2}, |X_{ij_1} - X_{ij_2}| \leq \epsilon \} . \]

are the number of concordant pairs and discordant pairs of the i-th sample given \( \epsilon \) respectively, \( i=1, 0. \) Then, clearly, the i-th sample "local correlation" index given \( \epsilon \) defined by

(5.10) \[ R_i(\epsilon) = \frac{T_i(\epsilon)}{M_i(\epsilon)} \]

satisfies the desired characteristics of a correlation index: standardization, symmetric under interchange, symmetric under reversal, invariance under transformation (for discussion, see, e.g. Quade, 1974). Under the assumptions of parallelism and concomitance of the covariable, the association between the response \( Y \) and the covariable \( X \) is the same for both groups, so we can combine the correlations of the two groups together to form a more powerful test statistic. Let

(5.11) \[ R(\epsilon) = R_1(\epsilon) + R_0(\epsilon) . \]

Then, under the null hypothesis that \( Y \) and \( X \) are locally uncorrelated within tolerance \( \epsilon \), asymptotically \( R(\epsilon) \) is normally distributed with mean zero and variance \( \sigma^2(\epsilon) = \sigma_1^2(\epsilon) + \sigma_0^2(\epsilon) \), where \( \sigma_1^2(\epsilon) \) is the variance of \( R_i(\epsilon), i=1, 0 \) (since \( R_1(\epsilon) \) and \( R_0(\epsilon) \) are independent).

We can have an estimate of \( \sigma^2(\epsilon) \) following the formula given in Quade (1967b, 1974) directly

(5.12) \[ \text{Var} \left( \frac{T_1(\epsilon)}{M_1(\epsilon)} + \frac{T_0(\epsilon)}{M_0(\epsilon)} \right) = \sum_i \frac{4}{M_i^4(\epsilon)} \left\{ M_i^2(\epsilon) \sum_j T_{ij}^2(\epsilon) - 2 M_i(\epsilon) T_i(\epsilon) \sum_j T_{ij}(\epsilon) M_{ij}(\epsilon) + T_i^2(\epsilon) \sum_j M_{ij}^2(\epsilon) \right\} , \]
\[ T_{ij}(\epsilon) = \sum_{j' \neq j}^{n_i} \text{sgn}(Y_{ij} - Y_{ij'}) \text{sgn}(X_{ij} - X_{ij'}) I\{ |X_{ij} - X_{ij'}| \leq \epsilon \}, \]

and
\[ M_{ij}(\epsilon) = \sum_{j' \neq j}^{n_i} I \{ |X_{ij} - X_{ij'}| \leq \epsilon \}. \]

A second method of estimation of the variance is by jackknifing U-statistics. In general, jackknife estimators reduce the bias of estimation from order \( n^{-1} \) to order \( n^{-2} \), where \( n \) is the total sample size. We omit the details here (for reference, see Arvesen, 1969, and Sen, 1977).

5.5 Discussion

Strictly speaking, if \( Y \) and \( X \) are correlated, then they should also be correlated locally. Thus, the notion of “locally uncorrelated” is somewhat vague. We may therefore need to base our philosophy on Type I and Type II errors: the null hypothesis is accepted because there is no statistical evidence to reject it. Certainly, as \( \epsilon \) becomes larger and larger, the evidence of correlation becomes more and more significant. Hence, we should be able to find a certain range of \( \epsilon \) such that \( Y \) and \( X \) have “no correlation” within that range (unless \( Y \) and \( X \) are independent overall, in which case \( \epsilon \) is \( \infty \)).

When there are several covariates, say, \( p \) covariates \( X = (X_1, ..., X_p) \), let \( \epsilon = (\epsilon_1, ..., \epsilon_p) \) be the tolerance vector corresponding to \( X \). Since \( E(\epsilon) = \{ |X_{11} - X_{10}| \leq \epsilon_1, ..., |X_{p1} - X_{p0}| \leq \epsilon_p \} \subset E(\epsilon_i) = \{ |X_{11} - X_{10}| \leq \epsilon_1, ..., |X_{i1} - X_{i0}| \leq \epsilon_i, ..., |X_{p1} - X_{p0}| \leq \infty \} \), we have that if \( Y \) and \( X_i \) are locally uncorrelated for \( E(\epsilon_i) \), then they are for \( E(\epsilon) \) also. Hence, one can examine each covariable to find the corresponding largest tolerance (and name it by \( \epsilon_{sup_i} \), say) individually, and then put these tolerance values together to form \( \epsilon_{sup} = (\epsilon_{sup_1}, ..., \epsilon_{sup_p}) \). Adjusted ranks of the covariates \( X \) based on \( \epsilon_{sup} \) can then be calculated for the ERMP(\( \lambda; \epsilon \)) to test for the treatment difference. The tolerance \( \epsilon_{sup} \) could be a bit smaller than the true maximum tolerance vector since each element obtained individually could be smaller than the one obtained simultaneously. However, such an approach does provide the advantage that when one incorporates an inappropriate covariable into the analysis, the corresponding tolerance should tend to infinity, and hence the covariable should be deleted.

The incorporation of tolerance into the classical regression model has two major impacts. First, the sum of squares for treatments and the sum of squares for error are changed accordingly. Although the error sum of squares given tolerance \( \epsilon \) could increase since the local behavior of \( Y \) is left unexplained by \( X \) within \( \epsilon \), the treatment sum of squares could increase also when \( Y \) and \( X \) are “locally uncorrelated”. Thus, we may have a different result from the F-ratio.
statistic given tolerance \( \epsilon \). Second, owing to the invariance property for covariates within the tolerance \( \epsilon \), a small amount of measurement error (such as blood pressure measurements) in experimental studies or recall bias of continuous covariates in certain observational studies will not bias the result using caliper matching, and thus the ERMP(\( \lambda; \epsilon \)) proposed in Chapter 4. For further research interest, one might relate the concept of tolerance to techniques developed for measurement error as an input. We will discuss these further in Chapter 7 on suggestions for future research.

As a final comment, although there is a similarity between the tolerance \( \epsilon \) and the bandwidth \( h \) in nonparametric regression, they are different in that the former does not depend on the sample size \( N \) while the latter does. In smoothing technique, an unknown functional form \( y = f(x) \) is always assumed. Therefore, to increase the precision of the estimate, a decreasing bandwidth \( h \) is desirable when the sample size increases. But in the philosophy of “local lack of correlation”, a natural limitation between two variables always exists. Therefore, one can only describe the behavior of the response variable \( Y \) through the covariable \( X \) to a certain extent. This amount of “uncertainty” is reflected in the “locally uncorrelated” parameter \( \epsilon \), which is unaffected by the sample size \( N \).
CHAPTER 6
EXAMPLES AND APPLICATIONS

6.1 Introduction

We have developed a new nonparametric test (ERMP) for analysis of covariance, which is a unified approach to both "rank analysis of covariance" (Quade, 1967) and "matched difference in probability" (Quade, 1982), and has shown better efficiency properties as compared with existing matching methods. Also, a statistic for the optimal tolerance of a given pair of response variable and covariable is proposed. In this chapter, we demonstrate these techniques using two real examples.

The first example is from the Study on the Efficacy of Nosocomial Infection Control (SENIC Project), as reported by Neter, Wasserman, and Kutner (1985). Given a random sample of 113 hospitals located in four different geographical regions, one is interested in comparing the regions with respect to the average length of stay (ALOS) in hospital (in days) for all patients, controlling for the overall service index provided by the hospital (a continuous variable) as a covariable. We describe the details of this data set as well as its analysis in Section 6.2.

The second example is a randomized clinical trial on 59 female patients with rheumatoid arthritis (Koch et al, 1982). Patient response status (PRS, with codes 1=excellent through 5=poor) is the response, and patient age is controlled as a covariable; we compare treatments (active vs placebo). We also dichotomize PRS as zero (poor, fair, or moderate) and one (good or excellent), and compare the results of the Mantel-Haenszel statistics with and without stratification by age, with that of the ERMP test using 10 years as the tolerance for age differences. The details are given in Section 6.3.

We provide a general discussion in Section 6.4.

6.2 Example from the Study on the Efficacy of Nosocomial Infection Control

The primary objective of the SENIC project was to determine whether infection surveillance and control programs have reduced the rates of nosocomial (hospital-acquired) infection in United States hospitals.

The data set consists of a random sample of 113 hospitals selected from the original 338 hospitals surveyed during the 1975-76 study period. Among the variables observed for each
<table>
<thead>
<tr>
<th>ID</th>
<th>ALOS</th>
<th>SERVICE</th>
<th>REGION*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.82</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>9.76</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>11.18</td>
<td>60</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>7.58</td>
<td>37.1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>9.06</td>
<td>37.1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>10.24</td>
<td>37.1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>8.28</td>
<td>57.1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>9.89</td>
<td>37.1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>9.84</td>
<td>54.3</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>9.74</td>
<td>54.3</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>9.97</td>
<td>34.3</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>10.47</td>
<td>54.3</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>8.37</td>
<td>34.3</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>10.90</td>
<td>51.4</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>9.23</td>
<td>71.4</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>12.07</td>
<td>31.4</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>10.73</td>
<td>51.4</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>11.18</td>
<td>68.6</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>9.66</td>
<td>28.6</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>10.02</td>
<td>48.6</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>7.39</td>
<td>28.6</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>8.45</td>
<td>48.6</td>
<td>0</td>
</tr>
<tr>
<td>23</td>
<td>10.23</td>
<td>68.6</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>8.88</td>
<td>48.6</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>10.30</td>
<td>45.7</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>10.79</td>
<td>65.7</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>7.94</td>
<td>45.7</td>
<td>0</td>
</tr>
<tr>
<td>28</td>
<td>10.39</td>
<td>65.7</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>8.92</td>
<td>5.7</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>9.77</td>
<td>25.7</td>
<td>0</td>
</tr>
<tr>
<td>31</td>
<td>8.02</td>
<td>22.9</td>
<td>0</td>
</tr>
<tr>
<td>32</td>
<td>11.80</td>
<td>62.9</td>
<td>0</td>
</tr>
<tr>
<td>33</td>
<td>8.34</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>9.68</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>8.67</td>
<td>40</td>
<td>1</td>
</tr>
</tbody>
</table>

* REGION: 0 - North Central, 1 - South.
Figure 6.2.1 Average length of stay vs hospital service index with fitted parallel regression lines for North Central (0) and South (1) regions.
single hospital, one is interested in comparing the ALOS of all patients in hospital (in days), among hospitals located in four different geographic regions (North East, North Central, South, and West), controlling for the overall service index provided by the hospital as a covariable.

Here we compare the 32 hospitals in North Central (denoted by 0) and the 37 hospitals in South (denoted by 1). We list the data in Table 6.2.1, and show the scatter plot in Figure 6.2.1.

Notice a tendency that the better the service provided by the hospital, the longer the ALOS in the hospital for both regions; however, the trend is not entirely clear. The difference in ALOS in the hospital between the two regions is only 0.5 (9.68 vs 9.19) without adjusting for the service index. Thus, this may not reflect any real difference in ALOS between hospitals located in these two geographic regions. The average service index provided by hospitals in North Central is 45.7, which is 6 points higher than that in South (mean=39.5). However, the difference may also not be significant, since the service index ranges from 5.7 to 77.1. The Wilcoxon test for equal mean service indices of the two regions and the Kolmogorov-Smirnov test for the equivalence of their distribution functions have p-values 0.082 and 0.530 respectively. Hence, the assumption of concomitance needed for "rank analysis of covariance" and the ERMP test seems satisfactory.

Fitting an ANOCOVA model with the covariable SERVICE and an interaction term of SERVICE×REGION, one finds that the interaction is not significant (p-value=0.483). Thus, a parametric ANOCOVA adjusting for the covariable SERVICE (p-value=0.0088) with parallel regression lines may be considered. From this, we could reach the conclusion that ALOS in these two regions is about the same (p-value=0.239). However, we may find that the fitted lines do not explain the data quite well (Figure 6.2.1), and may doubt the legitimacy of a strict linear relationship between these two variables ALOS and SERVICE. The result from ANOVA ignoring the covariable SERVICE, on the other hand, gives a p-value 0.097, which is barely significant.

We give results from alternative nonparametric approaches in Table 6.2.2. Using "rank analysis of covariance" ($\epsilon=0$ of ERMP($\tau; \epsilon$) in Table 6.2.2), one could reach the same conclusion that there is no region difference (p-value=0.216). However, when the level of tolerance increases from 0 to 40, using the ERMP test, the geographic region difference appears to be more evident (with p-value decreasing from 0.216 to 0.078). The p-value stays the same ($\epsilon=0.077$) when tolerance increases from 50 to 75, when eventually all pairs are matched (which is equivalent to the Wilcoxon two-sample test). With closer examination of the scatter plot, one may find that ALOS is not so sensitive to small variations in SERVICE. Thus, the model specification with a strict linear relationship between ALOS and SERVICE may not be so appropriate. Matching methods, alternatively, may be able to explore more the nature of the relationship between these two variables in this case.
Table 6.2.2 Nonparametric ANOCOVA test statistics and tolerance for matching

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>$\tau$</th>
<th>VR</th>
<th>p-value</th>
<th>$\bar{T}$</th>
<th>p-value</th>
<th>$\bar{D}$</th>
<th>p-value</th>
<th>R($\epsilon$)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.347</td>
<td>1.559</td>
<td>0.216</td>
<td>0.091</td>
<td>0.671</td>
<td>0.114</td>
<td>0.737</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.344</td>
<td>1.668</td>
<td>0.201</td>
<td>0.124</td>
<td>0.475</td>
<td>0.190</td>
<td>0.553</td>
<td>-0.162</td>
<td>0.839</td>
</tr>
<tr>
<td>10</td>
<td>0.307</td>
<td>1.951</td>
<td>0.167</td>
<td>0.170</td>
<td>0.263</td>
<td>0.241</td>
<td>0.416</td>
<td>0.110</td>
<td>0.200</td>
</tr>
<tr>
<td>20</td>
<td>0.213</td>
<td>2.627</td>
<td>0.110</td>
<td>0.197</td>
<td>0.174</td>
<td>0.339</td>
<td>0.252</td>
<td>0.308</td>
<td>0.005</td>
</tr>
<tr>
<td>30</td>
<td>0.127</td>
<td>3.018</td>
<td>0.087</td>
<td>0.210</td>
<td>0.137</td>
<td>0.377</td>
<td>0.192</td>
<td>0.382</td>
<td>0.002</td>
</tr>
<tr>
<td>40</td>
<td>0.049</td>
<td>3.204</td>
<td>0.078</td>
<td>0.225</td>
<td>0.104</td>
<td>0.439</td>
<td>0.128</td>
<td>0.468</td>
<td>0.001</td>
</tr>
<tr>
<td>50</td>
<td>0.017</td>
<td>3.230</td>
<td>0.077</td>
<td>0.243</td>
<td>0.075</td>
<td>0.477</td>
<td>0.096</td>
<td>0.476</td>
<td>0.001</td>
</tr>
<tr>
<td>60</td>
<td>0.002</td>
<td>3.232</td>
<td>0.077</td>
<td>0.249</td>
<td>0.067</td>
<td>0.493</td>
<td>0.086</td>
<td>0.480</td>
<td>0.001</td>
</tr>
<tr>
<td>70</td>
<td>0.000</td>
<td>3.232</td>
<td>0.077</td>
<td>0.249</td>
<td>0.067</td>
<td>0.493</td>
<td>0.086</td>
<td>0.488</td>
<td>0.001</td>
</tr>
<tr>
<td>75</td>
<td>0</td>
<td>3.232</td>
<td>0.077</td>
<td>0.248</td>
<td>0.068</td>
<td>0.492</td>
<td>0.087</td>
<td>0.488</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Results using “matched difference in probability” (denoted by $\bar{T}$) and “matched difference in mean” (denoted by $\bar{D}$) (Quade, 1982) are also listed in Table 6.2.2, for comparison. And the statistic R($\epsilon$) for a reasonable tolerance is given in the last two columns. Note that the ERMP test tends to be more powerful in detecting the region difference than both “matched difference in probability” and “matched difference in mean” as $\epsilon$ varies from 0 to 40. However, “matched difference in probability” gives slightly more significant results than those by other tests for $\epsilon$ greater than 40. The results of R($\epsilon$) as $\epsilon$ varies from 5 to 75 indicate that any tolerance which is greater than 10 would not be appropriate. Thus, though SERVICE does affect ALOS, it may not make any difference when the SERVICE index varies within tolerance 10.

Classical ANOCOVA or ANOVA methods may lead to inappropriate conclusions in either direction, since both assume either a strict monotonic relationship between ALOS and SERVICE or no relationship between them at all. The ERMP test, on the other hand, provides a way to compromise between these two approaches. Locally (within the given tolerance) the covariable plays no role in the ERMP test, while overall it is still an explanatory variable in the regression setting. Matching methods, though sharing similar advantages to those of the ERMP test, do not make full use of the correlation structure between the response variable and the covariable (if there is any).
The seemingly contradictory result of the ERMP test compared with the parametric ANOCOVA adjusting for the covariable SERVICE demonstrates the hidden danger of the classical parametric ANOCOVA. Although it is highly desirable to adjust for a covariable for testing the treatment difference, the result could be misleading, even when the covariable term is significant in the model (as in our example).

A more complicated issue is implicit in this example. Though the slope of the regression model is significantly different from zero, the correlation coefficient between ALOS and SERVICE is not significant for any of the correlation indices for the South region. Thus, we may not be able to fit a common slope to both of the regions. However, the test for interaction between SERVICE and REGION is nonsignificant also. The test statistic $R(c)$, though requiring the same association between ALOS and SERVICE for the two regions, may be used to the two regions individually. A separate analysis of these two regions individually indicates that a reasonable tolerance for the North Central region is about 12, while that of the South region is about 30 (not shown here).

We suggest, from this example, that the decision whether to apply an ANOCOVA or an ANOVA without covariate adjustment should be based on whether the optimal tolerance between the response and the covariable approaches zero or not. Unless the optimal tolerance is approximately zero, which assures the strict monotone relationship between the response and the covariable, the standard ANOCOVA procedure may not be valid. Most of the time, we might need to adjust for a covariable with tolerance as the ERMP does, which is somewhere between ANOVA and ANOCOVA (nonparametrically).

6.3 Example of a Rheumatoid Arthritis Clinical Trial

In this section we give an example of dichotomous response to demonstrate the better efficiency property of the ERMP test compared with the Mantel-Haenszel statistic.

Fifty-nine female rheumatoid arthritis patients participated in a randomized clinical trial with two treatment groups (treatment vs placebo). Patient age is an important covariable as arthritis is strongly related to age. We list the data in Table 6.3.1. For simplicity, we dichotomize further the response status (PRS) into excellent or good versus other. Figure 6.3.1 shows the proportions of excellent or good PRS in the four age groups ($\leq 44$ years, 45-54 years, 55-64 years, and $\geq 65$ years).

Similar to the previous example, the questions of interest for the data in Table 6.3.1 are
Table 6.3.1 Data listing for rheumatoid arthritis clinical trial

<table>
<thead>
<tr>
<th>TGRP</th>
<th>Patient number</th>
<th>Age (yrs)</th>
<th>PRS</th>
<th>TGRP</th>
<th>Patient number</th>
<th>Age (yrs)</th>
<th>PRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>23</td>
<td>5</td>
<td>2</td>
<td>31</td>
<td>31</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>32</td>
<td>5</td>
<td>2</td>
<td>32</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>37</td>
<td>3</td>
<td>2</td>
<td>33</td>
<td>33</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>41</td>
<td>2</td>
<td>2</td>
<td>34</td>
<td>37</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>41</td>
<td>4</td>
<td>2</td>
<td>35</td>
<td>44</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>48</td>
<td>2</td>
<td>2</td>
<td>36</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>48</td>
<td>5</td>
<td>2</td>
<td>37</td>
<td>46</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>55</td>
<td>1</td>
<td>2</td>
<td>38</td>
<td>48</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>55</td>
<td>2</td>
<td>2</td>
<td>39</td>
<td>49</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>56</td>
<td>2</td>
<td>2</td>
<td>40</td>
<td>51</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>57</td>
<td>2</td>
<td>2</td>
<td>41</td>
<td>53</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>57</td>
<td>2</td>
<td>2</td>
<td>42</td>
<td>54</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>57</td>
<td>2</td>
<td>2</td>
<td>43</td>
<td>54</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>58</td>
<td>5</td>
<td>2</td>
<td>44</td>
<td>54</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>59</td>
<td>2</td>
<td>2</td>
<td>45</td>
<td>55</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>59</td>
<td>1</td>
<td>2</td>
<td>46</td>
<td>57</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>60</td>
<td>1</td>
<td>2</td>
<td>47</td>
<td>57</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>61</td>
<td>2</td>
<td>2</td>
<td>48</td>
<td>58</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>19</td>
<td>62</td>
<td>2</td>
<td>2</td>
<td>49</td>
<td>59</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>62</td>
<td>3</td>
<td>2</td>
<td>50</td>
<td>59</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>66</td>
<td>1</td>
<td>2</td>
<td>51</td>
<td>61</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>22</td>
<td>67</td>
<td>2</td>
<td>2</td>
<td>52</td>
<td>63</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>68</td>
<td>3</td>
<td>2</td>
<td>53</td>
<td>64</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>24</td>
<td>68</td>
<td>1</td>
<td>2</td>
<td>54</td>
<td>65</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>69</td>
<td>3</td>
<td>2</td>
<td>55</td>
<td>66</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>69</td>
<td>5</td>
<td>2</td>
<td>56</td>
<td>66</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>27</td>
<td>70</td>
<td>3</td>
<td>2</td>
<td>57</td>
<td>66</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>23</td>
<td>4</td>
<td>2</td>
<td>58</td>
<td>68</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>30</td>
<td>4</td>
<td>2</td>
<td>59</td>
<td>74</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>30</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TGRP: 1, active treatment; 2, placebo treatment.

PRS (Patient response status): 1, Excellent; 2, Good; 3, Moderate; 4, Fair; 5, Poor.
Table 6.3.2 2 x 2 tables of dichotomized PRS (1=excellent or good, 0=all other) by treatment (active vs placebo) by age categories.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trt</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Con</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

RR = 1.6
(Age ≤ 44)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trt</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Con</td>
<td>1</td>
<td>8</td>
</tr>
</tbody>
</table>

RR = 4.5
(Age 45-54)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trt</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>Con</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

RR = 3.8
(Age 55-64)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trt</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Con</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

RR = 1.3
(Age ≥ 65)

Figure 6.3.1 Proportion excellent or good PRS by treatment (active vs placebo) and age category.
(i) The extent to which patients on active treatment had more favorable response than those on placebo;
(ii) the extent to which a treatment difference was maintained after adjustment for age; and
(iii) the extent to which a treatment difference was homogeneous across patient ages and so generalizable to all ages (Koch et al., 1982).

Different points of view toward the sampling processes generating the data and methods of covariance analysis are discussed by Koch et al. Presuming them representative of a larger population of patients in each treatment group would lead to log-linear and logistic model approaches. Here we view the subjects under study as a finite population which has been randomized into two groups prior to treatment. For this sampling framework, hypergeometric models are used as the basis for unadjusted and covariance-adjusted randomization tests of no difference between treatment effects.

From Figure 6.3.1, we notice that the treatment group tends to have a higher proportion of excellent or good response across all age groups. Also, the proportions are positively correlated with the age groups. The classical stratification analysis, however, does not reflect this relationship in the analysis scheme. The distribution of patient responses in the two treatment groups within each age category is shown in the 2 x 2 tables (Table 6.3.1). Though the active treatment group has higher proportions of excellent or good response than those of the placebo group, the relative rates are not approximately the same across the age strata. The Breslow-Day test of homogeneity of odds ratio does not reject (p-value=0.36). However, since there are not enough subjects within each stratum, the test may not be valid.

The Mantel-Haenszel statistics without and with age stratification are both highly significant (unstratified: $\chi^2 = 10.1$, p-value=0.00148; stratified: $\chi^2 = 7.017$, p-value=0.008). The less significant result of the Mantel-Haenszel statistic with stratification can be explained by the fact that the age distribution is slightly unbalanced between the two treatment groups. However, the Wilcoxon test of equal age distribution is not rejected (p-value=0.15). The ERMP test, using tolerance ten years of age, gives an even more significant result ($\chi^2 = 10.554$, p-value=0.00116) than the Mantel-Haenszel statistic without stratification, when approximated by a chi-squared distribution.

The result may not be surprising, since it is expected from the theoretical results that the ERMP test would be more efficient than simple random sampling if the covariable is appropriate, and more efficient than the optimal category matching, which is equivalent to the Mantel-Haenszel statistic given all the marginals of the strata being fixed (see Chapter 4).

The analysis of this example is far from complete in that we have considered only the dichotomous case and followed the same age interval (ten years) given in Koch et al. (1982) as our tolerance in the ERMP test. More thorough discussion and analysis of this data set can be
found in their paper. The relative merits of the ERMP test compared with other nonparametric covariance analysis as well as model approaches are unknown. Further theoretical work may be necessary for these comparisons.

6.4 Discussion

The advantages of the ERMP test in analysis of covariance seem clear here for both continuous (Example 1) and dichotomous response (Example 2).

Although the ERMP test is less efficient than the parametric ANOCOVA when the conditions of the latter are satisfied, the assumptions of parametric ANOCOVA may be too strict to be true in most cases. In Example 1, we showed their possible falsity using the standard approach in ANOCOVA without further adjustment for an appropriate tolerance ($\epsilon=10$ in the example). It seems, thus, that the concept of tolerance should be brought into data analysis when one builds a "regression" model for a given data set.

Matching methods, although they provide an alternative view in analysis, fail to take into account the correlation between the response and the covariable and thus lose efficiency. In both our examples, we can see that the ERMP test does better in detecting group differences than alternative matching methods do.

A methodology for finding the optimal tolerance in ERMP analysis of covariance when the response is discrete has not been developed. The choice of a ten year age interval in Example 2 may be somewhat arbitrary, and we may need to rely on other sources to obtain this information without justification. It is promising that a parallel technique can be developed following that in the continuous response case.

When the response is (ordered) categorical, though the ERMP test is still more efficient than the optimal category matching, the latter is no longer equivalent to the generalized Mantel-Haenszel statistic. Some comparisons of the ERMP test with the generalized Mantel-Haenszel statistic would be of interest.
CHAPTER 7
SUGGESTIONS FOR FUTURE RESEARCH

7.1 Introduction

Analysis of covariance is obviously too broad a topic to cover in every aspect. The present work, though providing nice efficiency properties and other theoretical merits, is of course far from complete. Extensions in some directions to cope with existing methodologies in related fields are possible. We list some of them and give each of them a brief discussion in the following sub-sections. The purpose, rather than to give solutions, is to provide a checklist for future research interest.

(i) The optimal weights for caliper matching to have optimal ARE with respect to category matching;
(ii) the properties of $\rho_s(\epsilon)$ when the response is (ordered) dichotomous or polytomous;
(iii) the ARE of ERMP($\lambda; \epsilon$) with respect to logistic regression when the response is (ordered) dichotomous or polytomous;
(iv) adjustable tolerance between the response variable and the covariable;
(v) extensions to ordered alternatives;
(vi) decompositions of sums of squares of extended ranks of covariables given vector of tolerance;
(vii) removal of the assumption of concomitance in the ERMP test;
(viii) using actual measurement values instead of ranks;
(ix) measurement errors and repeated measurements applications;
(x) more powerful test for matching tolerance.

7.2 Possible Research Work

7.2.1 Optimal Weights for Caliper Matching

We have compared the asymptotic relative efficiency of the unweighted caliper matching with respect to category matching as well as simple random sampling in Chapter 3. It was shown
that the unweighted caliper matching is, in general, less efficient than the optimal category matching, and may be less efficient than simple random sampling.

The statistic for "matched difference in probability" (Quade, 1982), which uses a natural weight by taking the ratio of the number of pairs matched to a certain covariate to the overall number of matched pairs given a tolerance, provides an alternative way in raising the efficiencies. Comparing to the weighting scheme of the optimal category matching, the weights proposed there may already possess certain optimal properties. Further investigation in this direction is needed.

7.2.2 When the Response Variable is Discrete (at Least Ordinal)

We showed that, given all the marginals of the strata in a stratified analysis, the ERMP test is better than stratification in efficiency by an approximate factor $1/(1 - \rho_s^2(\epsilon))$, when the response is continuous. When the response is dichotomous, the optimal category matching statistic we use is equivalent to the familiar Mantel-Haenszel statistic. However, no direct theoretical justification other than continuous response variable is made. A more general setting to include both the case when the response variable is continuous and (ordered) categorical is desirable. Also, the properties of $\rho_s^2(\epsilon)$ are unknown when the response is (ordered) dichotomous or polytomous. One should explore these properties further for completion.

When the response is polytomous, the optimal category matching statistic we used is no longer equivalent to the generalized Mantel-Haenszel statistic conditional on all the marginal being fixed. Thus, the ARE of the ERMP test with respect to the generalized Mantel-Haenszel statistic is unknown.

Also, the ARE of the ERMP test with respect to logistic regression when the response is polytomous or dichotomous is unknown. Most likely, the result will be similar to that of the comparison to the standard ANOCOVA.

7.2.3 Adjustable Tolerance

Though a constant tolerance is satisfactory in that, when a given tolerance is too large to be legitimate, one can always choose a smaller one, the situation could not hold if the relationship between the response and the covariable is changing across the range of the covariable. For example, the growth curve of children is changing at different rates by age. One may therefore, use adjustable levels of tolerance instead of a constant in dealing with this issue. However, the technique required may become very complicated trying to find different levels of tolerance from a given data set.
7.2.4 Extensions to Ordered Alternatives

Two methods in nonparametric analysis of covariance with ordered alternatives are the “union-intersection rank test” (Boyd and Sen, 1986), which extends “rank analysis of covariance” (Quade, 1967) and the “general rank scores analysis of covariance” (Puri and Sen, 1969) based on the union-intersection principle, and “nonparametric analysis of covariance with ordered alternatives” (Marcus and Ginizi, 1987) which extends “matched difference in probability” (Quade, 1982) to k-sample cases with ordered alternatives.

Since the ERMP test is an extension to both “rank analysis of covariance” and “matched difference in probability”, extensions to the case of ordered alternatives which unify further the above two methods are possible.

7.2.5 Decompositions of Sums of Squares With Tolerance

We have seen in Example 1 of Chapter 6, using tolerance 30 or above (a “legitimate” tolerance is 10 using R statistic in Chapter 5) gives a more “significant” p-value than using tolerance 0 (“rank analysis of covariance”). A possible explanation to this is that the tolerance in rank adjusted scores increases further the component of sum of squares of treatment, thus gives a more significant result in that example. Geometrically, using a positive tolerance (>0) might fit the data better than using a strict linear relationship (tolerance=0) when the pattern of the latter is not clear, since it explains the local behavior better in such a case.

In general, the concept of tolerance could be introduced into linear regression models to increase the sum of squares of treatment further while the sum of squares of error increases relatively a small amount, especially when the test statistic R does not reject the tolerance of interest. Most important, the methodology of fitting regression curve might need to be modified accordingly, since “locally lack of correlation” between the response and the explanatory variable would make the regression model unjustifiable.

7.2.6 Removal of the Assumption of Concomitance

For the ERMP test, we assume that the marginal distribution of the covariate is the same among the comparison groups. Although this can be achieved in a completely randomized design, this does not always hold in observational studies. One cannot thus remove the bias using the ERMP test in testing treatment difference as the parametric ANOCOVA does.

Often in case-control studies, the variance of the covariate of the control group is larger than the case group (Raynor and Kupper, 1981). While the Mantel-Haenszel statistic is still
applicable in such situation, the ERMP test can no longer be valid. A modification of the ERMP test which allows location and scale difference of the covariable is therefore necessary.

7.2.7 Actual Values Instead of Ranks

Two parallel methods using caliper matching in analysis of covariance are “matched difference in probability” and “matched difference in mean” (Quade, 1982). While the ERMP test is an extension of the “matched difference in probability”, using ranks and extended-ranks of the response and the covariable, a parallel technique using actual values may be possible, which is the approach in “matched difference in mean”.

7.2.8 Measurement Errors and Repeated Measurements Applications

While the ERMP test is invariant when the covariable varies within a given tolerance, this certainly has some implications in measurement errors and repeated measurements study.

In many studies, one does not obtain precise measurements of the covariable, and thus there is a certain amount of measurement error in a data set. Rather than treating the covariable(s) as stochastic as the ERMP does (also, “rank analysis of covariance”, and “matched difference in probability”), standard regression models assume the covariable(s) as known constants. The possible sources of error are therefore twofold, (1) misspecification of the model structure; (2) measurement errors of the covariables.

A related technique in taking into account the variation of the covariables is repeated measurements for each subject in the study. Since the response variable is measured repeatedly as well, the structure of the analysis could be a bit more complicated than the previous case.

7.2.9 More powerful test in finding tolerance for matching

In Chapter 5, we proposed a concept of “locally uncorrelated” to justify the purpose of matching as well as to find a corresponding tolerance using Kendall’s τ statistic. While the test statistic is asymptotically normally distributed, it may not have good enough power to reject the null hypothesis. Some modification of the proposed statistic or other index of correlation in finding a reasonable tolerance for matching is of interest.
APPENDIX A. MACROS FOR ERMP(\tau; \epsilon)

%macro grank(dset,outset,var1,var2,var3,outx,N,epi1);

* Calculation of the extended ranks given tolerance \epsilon; *
******************************************************************************
* dset - dataset name;
* outset - output dataset;
* var1 - input variable name for extended ranks;
* var2 - the group variable;
* var3 - identification number;
* outx - name of the output array of extended ranks;
* N - total sample size;
* epi1 - tolerance for extended ranks;
******************************************************************************

proc sort data=&dset out=a1;
by &var1;

proc sort data=&dset out=b1;
by &var2;

proc means n data=b1;
var &var1;
by &var2;
output out=c1 n=n1 n2;

data a2;
set a1;
array y{*} y1-y&N;
array r{*} r1-r&N;
array t{*} t1-t&N;
array idn{*} idn1-idn&N;
retain i 0;
i=i+1;
```plaintext
y(i)=&var1;
r(i)=i;
t(i)=&var2;
idn(i)=&var3;
run;

proc means data=a2 noprint;
var y1-y&N r1-r&N t1-t&N idn1-idn&N;
output out=a2
   sum=y1-y&N r1-r&N t1-t&N idn1-idn&N;

data rank;
set a2;
array y(*) y1-y&N ;
array r(*) r1-r&N ;
array t(*) t1-t&N ;
array idn(*) idn1-idn&N;
do j=1 to &N-1;
ul=y(j)+&epi1;
ll=y(j)-&epi1;
do i=j+1 to &N;
if ll<=y(i) <= ul then do;
r(j)=r(j)+0.5;
r(i)=r(i)-0.5;
end;
end;
end;
run;

data &outset;
set rank;
array r(*) r1-r&N;
array t(*) t1-t&N;
array idn(*) idn1-idn&N;
do i=1 to &N;
idx=idn(i);
```

94
&outx=r{i}-(&N+1)/2;
trtx=t{i};
output;
end;
keep idx trtx &outx;

%mend grank;

%macro coeff(din1,din2,N,gp1,gp2);

* Calculation of the optimal weight ρ; *
******************************************************************************
* din1 - dataset 1 for the array of extended ranks of the covariable; *
* din2 - dataset 2 for the array of extended ranks of the response variable; *
* N    - total sample size; *
* gp1  - index of group 1; *
* gp2  - index of group 2; *
******************************************************************************

proc sort data=&din1 out=b1;
by idx;
run;

proc sort data=&din2 out=b2;
by idx;
run;

data cof;
merge b1 b2;
by idx;
retain n csq cr 0;
n=n+1;
csq=csq+ranky*ranky;
cr=cr+ranky*rankx;
if n=&N then do;
tau=cr/(sqrt(csq*n+(n*n-1)/12));
output;
end;
keep tau ;

proc print;
run;

%mend coeff;

%macro pval(din1,din2,tau,N,gp1,gp2);
  * Calculation of the variance ratio VR(τ; ε) statistic and the corresponding p-value; *

******************************************************************************
* din1 - dataset 1;
* din2 - dataset 2;
* tau  - estimated optimal weight from the previous macro %coeff;
* N    - total sample size;
* gp1  - index of group 1;
* gp2  - index of group 2;
******************************************************************************

data all;
merge b1 b2;
by idx;
retain n s1 s2 sq1 sq2 n1 n2 0;
n=n+1;
z=ranky-&tau*rankx;
if trtx=&gp1 then do;
s1=s1+z;sq1=sq1+z*z;n1=n1+1;end;
if trtx=&gp2 then do;
s2=s2+z;sq2=sq2+z*z;n2=n2+1;end;
if n=&N then do;
  vr=(n-2)*((s1*s1/n1+s2*s2/n2))/(sq1+sq2-(s1*s1/n1+s2*s2/n2));
output;
end;

96
data q;set all;
p=1-prob((vr,1,&N-2);

proc print;
run;

%mend pval;
APPENDIX B. MACROS FOR R(ε)

%macro varn(dset, var1, var2, var3, N, n1, n0, epi);

****************************************************************************
*  dset   - data set name;
*  var1   - the response variable;
*  var2   - the covariable;
*  var3   - group variable;
*  N      - sample size;
*  n1     - sample size of the first group;
*  n0     - sample size of the second group;
*  epi    - test tolerance;
*****************************************************************************

proc sort data=&dset out=a1;
by &var3 &var1;

data a2;
set a1;
array y{*} y1-y&N;
array x{*} x1-x&N;
retain i 0;
i=i+1;
y{i}=&var1;
x{i}=&var2;
run;

proc means data=a2 noprint;
var y1-y&N x1-x&N ;
output out=a2
    sum=y1-y&N x1-x&N;

* estimation of the variance by delta method; *
data varian;
set a2;
array y{*} y1-y&N;
array x{*} x1-x&N;
array t1{*} t11-t1&n1;
array t0{*} t01-t0&n0;
array c1{*} c11-c1&n1;
array c0{*} c01-c2&n0;
array d1{*} d11-d1&n1;
array d0{*} d01-d0&n0;
array m1{*} m11-m1&n1;
array m0{*} m01-m0&n0;
retain t11-t1&n1 t01-t0&n0 m11-m1&n1 m01-m0&n0
     c11-c1&n1 c01-c0&n0 d11-d1&n1 d01-d0&n0
     m1  m0  t1  t0  t1m1 t0m0 t1sq t0eq m1sq m0sq 0;
do i=1 to &n1;
   ul=x[i]+&epi;
   ll=x[i]-&epi;
do j=1 to &n1;
   if y[i]>y[j] and x[i]>x[j] and ll<=x[j] <= ul then c1{i}=c1{i}+1;
   if y[i]<y[j] and x[i]<x[j] and ll<=x[j] <= ul then c1{i}=c1{i}+1;
   if y[i]>y[j] and x[i]<x[j] and ll<=x[j] <= ul then d1{i}=d1{i}+1;
   if y[i]<y[j] and x[i]<x[j] and ll<=x[j] <= ul then d1{i}=d1{i}+1;
   t1{i} = c1{i} - d1{i};
   if i=j and ll<=x[j] <= ul then m1{i}=m1{i}+1;
end;
do i=&n1+1 to &N;
   ul=x[i]+&epi;
   ll=x[i]-&epi;
do j=&n1+1 to &N;
   k=j-n1;
   if y[i]>y[j] and x[i]>x[j] and ll<=x[j] <= ul then c0{k}=c0{k}+1;
   if y[i]<y[j] and x[i]<x[j] and ll<=x[j] <= ul then c0{k}=c0{k}+1;
   if y[i]>y[j] and x[i]<x[j] and ll<=x[j] <= ul then d0{k}=d0{k}+1;
   if y[i]<y[j] and x[i]<x[j] and ll<=x[j] <= ul then d0{k}=d0{k}+1;
   t0{k} = c0{k} - d0{k};
if i ^= j and ll<=x(j) <= ul then m0(k)=m0(k)+1;
end;
end;

drop y1-y&N x1-x&N;

do k=1 to &n1;
    mel=me1+m1{k};
    tel=te1+t1{k};
    cl=c1+c1{k};
    d1=d1+d1{k};
    t1sq=t1sq+t1{k}*t1{k};
    m1sq=m1sq+m1{k}*m1{k};
    t1m1=t1m1+t1{k}*m1{k};
end;

do k=1 to &n0;
    me0=me0+m0{k};
    te0=te0+t0{k};
    c0=c0+c0{k};
    d0=d0+d0{k};
    t0sq=t0sq+t0{k}*t0{k};
    m0sq=m0sq+m0{k}*m0{k};
    t0m0=t0m0+t0{k}*m0{k};
end;

var1=4*(mel*me1*t1sq+tel*te1*m1sq-2*me1*te1*t1m1)/(me1**4);
var0=4*(me0*me0*t0sq+te0*te0*m0sq-2*me0*te0*t0m0)/(me0**4);
var=var1+var0;
r=te1/me1+te0/me0;

z=r/sqrt(var);

proc print;
run;
%
mend;
REFERENCES


Mann, H. B. and Whitney, D. R. (1947), "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other," *Annals of Mathematical Statistics*, 18, 50-60.


— (1969), "Individual Matching with Multiple Controls in the Case of All or None Response," *Biometrics*, 25, 339-354.


University Press.

