MULTIVARIATE NONPARAMETRIC TREND ASSESSMENT WITH ENVIRONMENTAL APPLICATIONS

by

Sungsue Rheem

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

IN

STATISTICS

APPROVED:

Eric P. Smith, Chairman

Raymond H. Myers

Golde I. Holtzman

Marvin Lentner

Clint W. Coakley

July 1992

Blacksburg, Virginia

MULTIVARIATE NONPARAMETRIC TREND ASSESSMENT WITH ENVIRONMENTAL APPLICATIONS

by

Sungsue Rheem

Eric P. Smith, Committee Chairman

Statistics

(ABSTRACT)

A major goal of environmental monitoring is to determine whether the environment is improving or deteriorating. Questions about the health of the environment are usually questions about trends in environmental indicators, typically a number of chemical, physical, and biological variables. Because multiple indicators are required to characterize any but the most simple environment, the problem is statistically a multivariate problem. In this work, methods for analyzing multivariate environmental trends are presented and illustrated on 17 years of approximately monthly observations on 5 water quality variables from southwestern Virginia, USA. Multivariate methods can also be applied to analyze correlated univariate data collected on a seasonal or monthly basis. A variety of methods from the literature are discussed. A unified approach is described based on a general class of correlation measures to construct a general framework for the nonparametric analysis of multivariate trends.

To My Parents

ACKNOWLEDGEMENTS

The author wishes to express his deepest appreciation and gratitude to Dr. Eric Smith, his Ph.D. program advisor, for inspiring and supervising this research. His constant guidance has been invaluable.

Also, many thanks to other members of my Ph.D. committee, Dr. Golde Holtzman, Dr. Raymond Myers, Dr. Klaus Hinkelmann, Dr. Marvin Lentner, and Dr. Clint Coakley, and to the members of my M.S. committee, Dr. Jeffrey Birch, Dr. Marion Reynolds, and Dr. Jesse Arnolds.

The author also expresses his deepest gratitude and appreciation to the faculty members of the Department of Statistics at Korea University from whom he learned statistics as well as the way to live a life, former Professor Uhn Boong Paik, Professor Jae Chang Lee, Professor Insik Jang, Professor Han Poong Shin, former Professor Ki Mok Kim, Professor Keeyoung Kim, and Professor Myung Hoe Huh. To their instruction and concern, the author owes what he is now. He additionally does appreciate a special concern of Professor Jae Chang Lee, his advisor in the graduate program of Korea University Statistics Department.

The author is grateful to all faculty members of the Department of Statistics at Virginia Tech. His experiences with them in research, consulting, and teaching have all been rewarding and enjoyable.

Additional thanks to Dr. Smith, Dr. Myers, Dr. Hinkelmann, Dr. Holtzman, and Dr. Birch for their thoughful concern and support.

The author is also grateful to all secretaries in the Department of Statistics. They have been very helpful and kind. Special thanks to Linda Breeding, the kindest secretary ever.

This research was partially supported by a grant from the Virginia Water Resources Research Center, Virginia Tech. The author is grateful to the Department of Statistics and to the Virginia Water Resources Research Center for their joint support for this research. Thanks also go to the co-investigators of the grant, Dr. Eric Smith, Dr. Golde Holtzman, Dr. Carl Zippper, and Dr. Gregory Evanylo.

The author expresses his deepest gratitude and appreciation to his parents, Chang Hyuk Rheem and Sang Bin Park, for their encouragement and sacrificial support, to his wife, Younghee, for her spiritual support and work as a housewife, and to his daughter, Jean, for just being cute.

Contents

List of Tables

List of Figures

1 Introduction

Questions about the health of the environment tend to be simple but the answers may be quite complex. The questions are usually about change and the extent of change. Is an industry's effluent harming the environment? Does a legislative act result in an improvement in the environment? Is the water quality of the Chesapeake Bay improving? These questions are difficult because of the quantity of information that must be processed to produce an answer. Questions about trends in time require measurements over a reasonably long period, and environments are complex, so many components need to be measured.

These requirements lead to statistical difficulties. Environmental data sets usually have a large number of variables. Some of the values may be missing or beyond the detection limits of the measuring devices. Sampling programs may be irregular, data may be collected at odd intervals, and appropriate variables may have been overlooked. Environmental data are often correlated in time or space and/or highly skewed with extreme values. As an example, consider the data in Figure 1.1 on page 4, which were collected as part of a monitoring program in southwestern Virginia, USA (Zipper et al. 1992). A number of variables were monitored at 38 stations in a nine county region containing four watersheds. Five of the variables are displayed in the graphs. Some of the above-mentioned problems are apparent in the graphs. For example, many of the measurements on nonfilterable residue (NFR) are at or below the detection limit. The distribution of fecal coliforms (FC) is quite skewed. Also, one of the values of dissolved oxygen (DO) is quite small relative to others. Sampling was done on roughly a monthly basis, but, sampling was not done at the same time each month; in some months there are multiple samples, and for some variables there were some months with no samples.

Because of these limitations, standard statistical methods are often inappropriate for environmental data. Alternative methods are developing in the applied literature that are based on nonparametric statistics. The focus here is on methods for the nonparametric assessment of multivariate trend in environmental variables. These nonparametric methods are resistant to some of the problems mentioned above and provide useful tools for answering environmental questions.

The dissertation is organized as follows: Chapter 2 presents background for the research. Generalizations and extensions are made in the subsequent chapters. In Chapter 3, a class of multivariate nonparametric trend tests is constructed based on the generalized correlation coefficient proposed by Daniels (1944). As a further generalization, a generalized multivariate trend statistic is suggested in Chapter 4. In Chapter 5, a canonical analysis is proposed for interpretation of the relative importance of each variable. ANOVA-like analyses of multivariate trends are presented in Chapter 6. A water-quality data set collected from southwestern Virginia was used for examples of multivariate nonparametric trend analyses in Chapters 4 through 6. Chapter 7 summarizes this research and comments on the future research. Proofs of some results are provided in Appendix 1. A SAS program to calculate Spearman trend statistics and associated covariances is presented in Appendix 2.

Although this research is primarily for environmental studies, specifically, the detection of trends in water-quality constituents, the methodology developed from this research is general and can be applied to problems in other areas. Actually, for an example of Chapter 3 in which several correlation coefficients are discussed, a pharmaceutical data set was used.

This research was partially supported by a grant for the development of environmental trend assessment methodology and the trend analysis of southwestern Virginia water-quality data set from the Virginia Water Resources Research Center, Virginia Tech. Full analyses of water-quality data collected from 38 stations in southwestern Virginia for trend assessment will be presented in the reports and bulletin that will be published by the Virginia Water Resources Research Center, Virginia Tech. Coauthors of those publications will be Eric P. Smith, Golde I. Holtzman, Sungsue Rheem, Carl Zippper, and Gregory Evanylo.

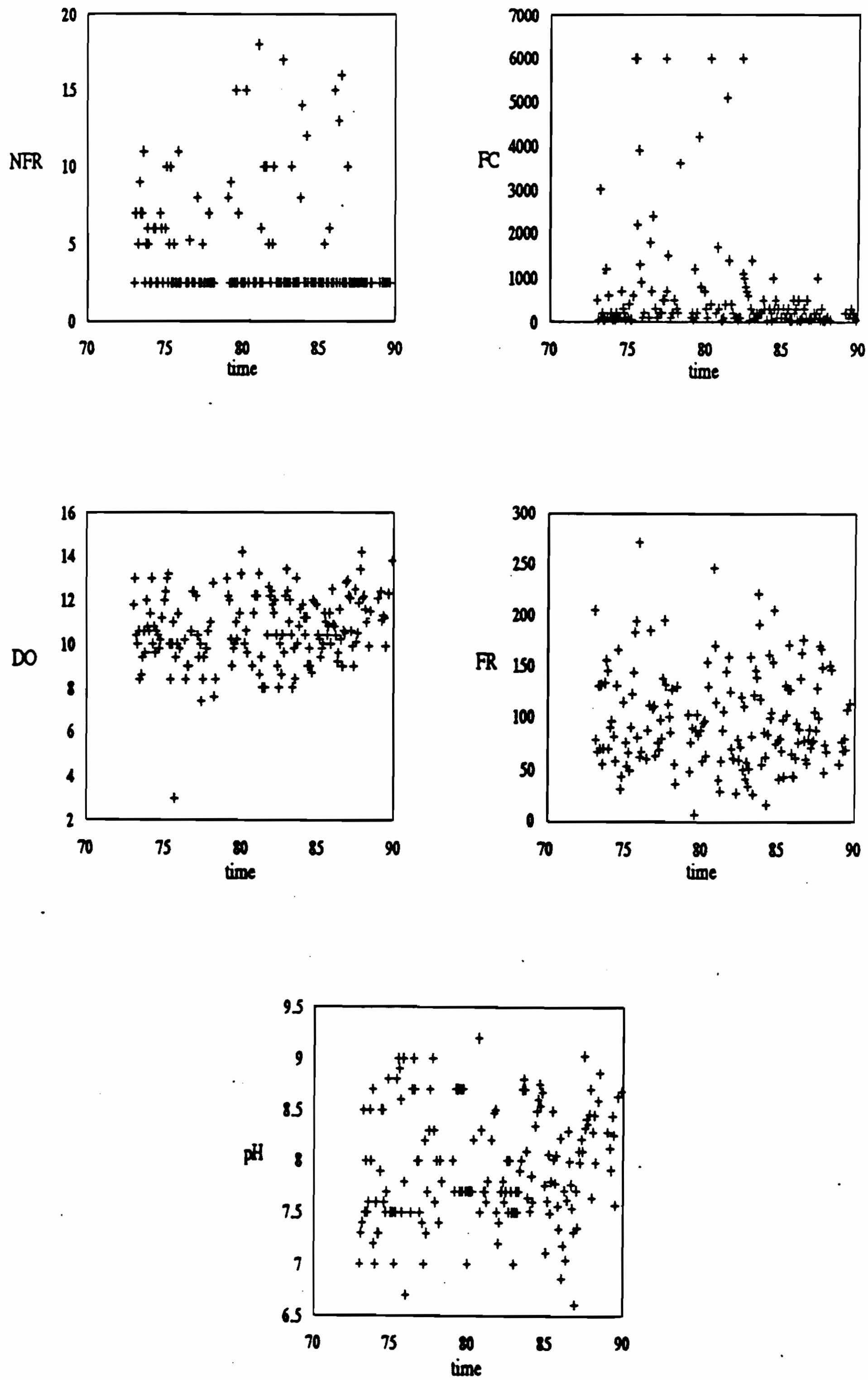

Figure 1.1: Graphs of NFR, FC, DO, FR, and pH versus time

2 Background

2.1 Case where p variables are measured at n time points

Initially the data described in Figure 1.1 will be ignored and we take the simpler view that there is only one measurement for each year. Thus the data can be viewed as observations on p variables measured at times 1, 2, ..., n. Let y_{iu} denote variable u at time x_i ($i = 1, 2, \ldots, n$ and $u = 1, 2, \ldots, p$) and let

$$\mathbf{Y} = \{y_{iu}\} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix}. \tag{2.1}$$

The question of trend can be viewed in terms of hypotheses about the relationships between variables and time. The null hypothesis (H_0) is that the n rows of $\mathbf{Y}$ are randomly ordered, i.e., each of the $n!$ permutations of the n rows of $\mathbf{Y}$ are equally likely. Under H_0, the n elements of each of the p columns are randomly ordered. This hypothesis is tested against the alternative that there exists a monotone trend in one or more of the variables.

The multivariate tests that are used here to test for trend can be viewed as non-parametric extensions of a parametric multivariate trend test. A parametric model for trend assessment is

$$\mathrm{E}(y_{iu}) = \alpha_u + \beta_u x_i, \tag{2.2}$$

where x_i is time for $i = 1, 2, \ldots, n$ and $u = 1, 2, \ldots, p$. This formulates a simple multivariate regression model. Let

$$\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$$

and

$$\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ip})'.$$

A test statistic for testing $\mathrm{H}_0 : \beta = \mathbf{0}$ against $\mathrm{H}_1 : \beta \neq \mathbf{0}$ is given by

$$A = \mathbf{b}' \mathbf{S}_{\mathbf{b}}^{-1} \mathbf{b} \tag{2.3}$$

where

$$\mathbf{b} = [1/\sum_{i=1}^{n}(x_i - \overline{x})^2] \sum_{i=1}^{n} \mathbf{y}_i (x_i - \overline{x}) \tag{2.4}$$

and

$$\mathbf{S}_{\mathbf{b}} = [1/\sum_{i=1}^{n}(x_i - \overline{x})^2] \sum_{i=1}^{n} (\mathbf{y}_i - \overline{\mathbf{y}}) (\mathbf{y}_i - \overline{\mathbf{y}})' - \mathbf{b}\,\mathbf{b}' \tag{2.5}$$

with

$$\overline{x} = \sum_{i=1}^{n} x_i / n, \quad \overline{\mathbf{y}} = \sum_{i=1}^{n} \mathbf{y}_i / n. \tag{2.6}$$

The quadratic form, A, in (2.3) is of the form of Hotelling's T^2 statistic and, under H_0, $(n-p-1)A/p$ is $\mathrm{F}_{p,\, n-p-1}$, assuming that $\mathbf{y}_i$'s are multivariate normal. For the case where time points, x_i's, are equally spaced and indexed as 1, 2, ..., n, $\mathbf{b}$ in (2.4) and $\mathbf{S}_{\mathbf{b}}$ in (2.5) are replaced by

$$\mathbf{b} = [12/(n^3 - n)] \sum_{i=1}^{n} \mathbf{y}_i [i - (n+1)/2] \tag{2.7}$$

and

$$\mathbf{S}_{\mathbf{b}} = [12/(n^3 - n)] \sum_{i=1}^{n} (\mathbf{y}_i - \overline{\mathbf{y}}) (\mathbf{y}_i - \overline{\mathbf{y}})' - \mathbf{b}\,\mathbf{b}', \tag{2.8}$$

respectively (Dietz 1983).

Nonparametric multivariate trend tests are similar, but they use nonparametric measures of trend. A univariate nonparametric test for trend is based on a trend statistic such as Kendall's or Spearman's rank correlation between time and variable. The Kendall trend statistic for variable u is

$$k_u = \sum_{i<j} \operatorname{sign}(y_{ju} - y_{iu}) \tag{2.9}$$

and the Spearman trend statistic for variable u is

$$s_u = \sum_i [i - (n+1)/2]\,[\operatorname{rank}(y_{iu}) - (n+1)/2]\,, \tag{2.10}$$

where $\operatorname{rank}(y_{iu})$ is the rank of y_{iu} ranked among $y_{1u}, y_{2u}, \ldots, y_{nu}$. For tied observations, signs are zero in (2.9) and ranks are replaced by midranks in (2.10). The Kendall sample rank correlation, $\hat{\tau}$, and the Spearman sample rank correlation, $\hat{\rho}$, between time and variable u, are defined as

$$\hat{\tau}_u = 2\,k_u/[n(n-1)] \tag{2.11}$$

and

$$\hat{\rho}_u = 12\,s_u/[n(n-1)(n+1)]\,, \tag{2.12}$$

respectively, provided that there are no ties. If there are ties, expressions (2.11) and (2.12) need some modifications (Kendall and Gibbons 1990).

Nonparametric multivariate trend tests having quadratic forms as test statistics were proposed by Bhattacharyya and Klotz (1966) using the Spearman statistic and by Dietz and Killeen (1981) using the Kendall statistic. Let

$$\mathbf{k} = (k_1, k_2, \ldots, k_p)'$$

be the vector of the Kendall trend statistics and

$$\mathbf{s} = (s_1, s_2, \ldots, s_p)'$$

the vector of the Spearman trend statistics. Also, let

$$k_{uv} = \sum_{i<j} \operatorname{sign}\left[(y_{ju} - y_{iu})(y_{jv} - y_{iv})\right] \tag{2.13}$$

and

$$s_{uv} = \sum_{i} [\operatorname{rank}(y_{iu}) - (n+1)/2]\,[\operatorname{rank}(y_{iv}) - (n+1)/2]. \tag{2.14}$$

Here, the results of Bhattacharyya and Klotz (1966) and Dietz and Killeen (1981) are reformulated in terms of $\mathbf{k}$, $\mathbf{s}$, k_{uv}, and s_{uv}. Null covariance matrices $\mathbf{S_k}$ and $\mathbf{S_s}$ are

$$\mathbf{S_k} = \{\operatorname{Cov}(k_u, k_v)\} = \{k_{uv}/3 + 4\,s_{uv}/3\} \tag{2.15}$$

and

$$\mathbf{S_s} = \{\operatorname{Cov}(s_u, s_v)\} = \{n(n+1)\,s_{uv}/12\}. \tag{2.16}$$

Let $\mathbf{S}^-$ denote the inverse of $\mathbf{S}$ when it is of full rank, or a generalized inverse of $\mathbf{S}$ when it is not of full rank. Then, under H_0, the test statistics

$$D = \mathbf{k}'\,\mathbf{S_k^-}\,\mathbf{k} \tag{2.17}$$

and

$$B = \mathbf{s}'\,\mathbf{S_s^-}\,\mathbf{s} \tag{2.18}$$

are asymptotically χ^2 with degrees of freedom equal to the ranks of $\mathbf{S_k}$ and $\mathbf{S_s}$, respectively. In the water resources literature, this test is referred to as the "covariance inversion" test (Lettenmaier 1988, Loftis et al. 1991).

Hirsch and Slack (1984) arrived at a similar test through different reasoning. They were interested in a test on a single variable that was measured over p seasons. Because measurements over time may be correlated, they treated the measurements in p

different seasons as p dependent variables. To test for trend, the Kendall trend statistic is computed as above for each season, and then the Kendall trend statistics are summed over the seasons. The standardized sum then becomes the test statistic for assessing trend. Their reasons for using the Kendall statistic include nonnormality, presence of missing values, and censoring of the data. However, the main reason for their choice of the Kendall statistic over the Spearman statistic would be the rapid covergence of the Kendall statistic to normality. Hirsch et al. (1982) demonstrated that the normal approximation for this statistic when there is between-season independence is quite accurate even for sample size as small as 2 years with 12 months. The test is seen to be a special case of the statistic in (2.17) by writing it in the form

$$H = (\mathbf{1}'\mathbf{k})'(\mathbf{1}'\mathbf{S_k}\,\mathbf{1})^{-1}(\mathbf{1}'\mathbf{k}) \tag{2.19}$$

This statistic is, under $\mathrm{H_0}$, asymptotically χ^2 with one degree of freedom and is sometimes called the "covariance sum" test statistic (Lettenmaier 1988, Loftis et al. 1991). This test works well if the signs of the seasonal trend statistics are the same. When the signs are different, the seasonal trend statistics cancel one another and this test will have low power. The test could be used when there are p variables rather than p seasons if information on the expected sign of the trend is available.

The Kendall-based covariance sum test can be viewed as an extension of the seasonal Kendall test described by Hirsch et al. (1982) and a test proposed by Jonckheere (1954). Jonckheere's test is a rank order test for ordered alternatives in randomized blocks. In our case, seasons can be regarded as blocks. Jonckheere's test uses the sum of the intrablock Kendall rank correlations as in (2.11), i.e.,

$$\sum_u \hat{\tau}_u \tag{2.20}$$

under the assumption of interblock independence. The seasonal Kendall test of Hirsch et al. uses the sum of the within-season Kendall trend statistics as in (2.9), i.e.,

$$\sum_u k_u \tag{2.21}$$

under the assumption of between-season independence. The Kendall-based covariance sum test does not require the assumption of independence between seasons. Similarly, the Spearman-based covariance sum test can be viewed as such an extension of a test proposed by Page (1954), which is also a rank order test for ordered alternatives in randomized blocks. Page's test uses the sum of the intrablock Spearman rank correlations, as in(2.12), i.e.,

$$\sum_u \hat{\rho}_u \tag{2.22}$$

under the assumption of interblock independence. Holtzman (1985) described an algorithm for computation of the exact distribution of Page's criterion.

Lettenmaier (1988) suggested that the covariance inversion test has low power when the sample size is small and suggested an alternative test that would be a useful alternative to the Hirsch and Slack covariance sum test as well. Lettenmaier suggests the statistic

$$L = \mathbf{k}'\,\mathbf{k} \tag{2.23}$$

which is the sum of squared Kendall trend statistics. As the distribution of this statistic is not approximated by a χ^2 distribution, recalling results on the distribution of a normal-vector-based quadratic form presented in Johnson and Kotz (1970, Ch. 29), Lettenmaier suggests using either a scaled noncentral χ^2 distribution (parameters: scale parameter, noncentrality parameter, and degrees of freedom) or a scaled and shifted central χ^2 distribution (parameters: scale parameter, shift parameter, and degrees of freedom) to model the null distribution of $\mathbf{k}'\mathbf{k}$. Parameters of either distribution can be estimated from its first moments and, if $\mathbf{k}'$ is multivariate normal with mean $\mathbf{0}$, then the moments of $\mathbf{k}'\mathbf{k}$ can be expressed in terms of the eigenvalues of the covariance matrix of $\mathbf{k}$ (Johnson and Kotz 1970 Ch. 29). This test is sometimes called the "covariance eigenvalue" test (Lettenmaier 1988, Loftis et al. 1991).

These three tests all have the same basic form but approach the test for trend in different manners. The statistical properties of the tests are different. As shown by simulations in Lettenmaier (1988) and Loftis et al. (1991), the covariance inversion test has low power when the sample sizes are small. The covariance sum test is favorable unless the signs of the trend statistics are mixed. Thus the covariance sum test would not be recommended for a general multivariate procedure. Lettenmaier's covariance eigenvalue test performed well in simulation studies (Lettenmaier 1988, Loftis et al. 1991).

2.2 Case where p variables are measured over m blocks (seasons) at n time points (years)

Now, let us consider the case where there are p variables measured in each of m seasons. With the first subscript denoting variable and the second subscript denoting season, let

$$\mathbf{c}^* = (c_{11}, \ldots, c_{1m}, c_{21}, \ldots, c_{2m}, \ldots, c_{p1}, \ldots, c_{pm})'$$

be the vector of mp generalized trend statistics, and let $\mathbf{S}_{\mathbf{c}^*}$ be the associated covariance matrix. Then,

$$\mathbf{c}^{*\prime}\, \mathbf{S}_{\mathbf{c}^*}^{-}\, \mathbf{c}^* \qquad (2.24)$$

is asymptotically χ^2 with degrees of freedom equal to the rank of $\mathbf{S_{c^*}}$.

Further, let $\mathbf{M}$ be a $p \times pm$ matrix whose row u has 1's as elements $(g-1)m+1$ through gm and 0's as other elements, i.e.,

$$\mathbf{M} = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & 0 & \dots & \dots & 0 \\ & & & & & & \ddots & & & \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 & 1 & \dots & 1 \end{bmatrix}. \tag{2.25}$$

Then, extending the test of Hirsch and Slack (1984) to the p-variable case, the statistic

$$(\mathbf{Mc^*})'(\mathbf{MS_{c^*}M'})^-(\mathbf{Mc^*}) \tag{2.26}$$

is asymptotically χ^2 with degrees of freedom being the rank of $\mathbf{MS_{c^*}M'}$.

Alternatives to (2.24) and (2.26) are available based on the covariance eigenvalue method. An alternative to (2.24) is

$$\mathbf{c^{*\prime}c^*} \tag{2.27}$$

and an alternative to (2.26) is

$$(\mathbf{Mc^*})'(\mathbf{Mc^*}). \tag{2.28}$$

The null distributions of these statistics would be modeled by either a scaled noncentral χ^2 distribution or a three parameter gamma distribution (equivalent to a scaled and shifted central χ^2 distribution) with its parameters estimated by functions of the eigenvalues of the associated covariance matrix.

3 A class of multivariate nonparametric tests for monotone trend

In this chapter, a class of multivariate nonparametric tests for monotone trend will be presented that is based on the generalized correlation coefficient. So, the discussion starts with introduction of the generalized correlation coefficient proposed by Daniels (1944). A new interpretation of Daniels' results makes this class of multivariate nonparametric trend tests constructible. In addition, a new correlation measure with an asymptotically-distribution-free test, which is a particular case of the generalized correlation coefficient, is proposed by the author. The new correlation coefficient is designed to be more informative but less robust than Kendall's or Spearman's correlation coefficient and is compared with other correlations through a pharmaceutical example. A discussion about the choice of correlation for use in the assessment of trend in environmental variables, e.g, water quality constituents, is presented.

3.1 A generalized correlation coefficient

In an interesting paper on correlation, Daniels (1944) proposed a generalized correlation coefficient. Consider

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \tag{3.1}$$

where $y_1, y_2, \ldots, y_n$ are arranged in some given order relative to $x_1, x_2, \ldots, x_n$. Let us assign to each pair of x_i and x_j a score $a_{ij} = f(x_i, x_j)$ and to each pair of y_i and

y_j a score $b_{ij} = f(y_i, y_j)$, where f, a score function, is such that

$$f(h_i, h_j) = -f(h_j, h_i). \tag{3.2}$$

Daniels defined a generalized correlation coefficient, say r_g, as

$$r_g = \frac{\sum_{i,j} a_{ij} b_{ij}}{\sqrt{\sum_{i,j} a_{ij}^2 \sum_{i,j} b_{ij}^2}}, \tag{3.3}$$

the summation extending over all i and j from 1 to n. This measure can also be written as

$$r_g = \frac{\sum_{i<j} a_{ij} b_{ij}}{\sqrt{\sum_{i<j} a_{ij}^2 \sum_{i<j} b_{ij}^2}}. \tag{3.4}$$

As special cases of r_g, Kendall's rank correlation is obtained when

$$a_{ij} = \text{sign}\,(x_j - x_i), \quad b_{ij} = \text{sign}\,(y_j - y_i), \tag{3.5}$$

Spearman's rank correlation is given when

$$a_{ij} = \text{rank}\,(x_j) - \text{rank}\,(x_i), \quad b_{ij} = \text{rank}\,(y_j) - \text{rank}\,(y_i), \tag{3.6}$$

and Pearson's product-moment correlation is obtained when

$$a_{ij} = x_j - x_i, \quad b_{ij} = y_j - y_i \tag{3.7}$$

(Daniels 1944, Kendall and Gibbons 1990 pp. 25–27).

For discussion of the distribution of r_g over all $n!$ permutations of the y's relative to the x's, it is suffices to consider only the numerator of r_g, because the denominator of r_g has the same value for any permutation. Let c be

$$c = \sum_{i<j} a_{ij} b_{ij} \tag{3.8}$$

which is the numerator of r_g. Then, under the null hypothesis that all $n!$ possible permutaions are equally likely,

$$\mathrm{E}(c) = 0 \tag{3.9}$$

and

$$\begin{aligned}\mathrm{Var}(c) &= \frac{(\sum_{i,j,k} a_{ij}a_{ik} - \sum_{i,j} a_{ij}^2)(\sum_{i,j,k} b_{ij}b_{ik} - \sum_{i,j} b_{ij}^2)}{n(n-1)(n-2)} \\ &\quad + \frac{(\sum_{i,j} a_{ij}^2)(\sum_{i,j} b_{ij}^2)}{2n(n-1)}, \end{aligned} \tag{3.10}$$

the summation extending over all subscripts from 1 to n (Daniels 1944, Kendall and Gibbons 1990). Since c is a U-statistic, an asymptotically standard normal test statistic for testing $\mathrm{H}_0 : \mathrm{E}(c) = 0$ is

$$c/\sqrt{\mathrm{Var}(c)}. \tag{3.11}$$

Daniels also obtained the covariance between two correlation coefficients applied to the same variables x and y when all $n!$ permutations of the y's relative to the x's are equally likely. The purpose of his paper was to investigate the relationship between two different correlation measures, e.g., τ_{xy} and ρ_{xy}, estimated from the same data set, in the universe of sample permutations. Daniels found that under certain general conditions, the joint distribution of such two correlations tends to the bivariate normal with increasing sample size, which can now be showed by theorems on U-statistics of Hoeffding (1948).

3.2 A class of multvariate nonparametric trend tests

We now extend the results of Daniels by obtaining the covariance between two correlation coefficients of the same type measured on the same x variable but different y variables, e.g., τ_{xy_1} and τ_{xy_2}.

Consider

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \begin{bmatrix} \mathbf{y}_u & \mathbf{y}_v \end{bmatrix} = \begin{bmatrix} y_{1u} & y_{1v} \\ y_{2u} & y_{2v} \\ \vdots & \vdots \\ y_{nu} & y_{nv} \end{bmatrix}, \tag{3.12}$$

where x_i is time and y_{iu} and y_{iv} are environmental variables u and v, respectively, for $i = 1, 2, \ldots, n$. With f denoting a score function as in (3.2), let

$$a_{ij} = f(x_i, x_j), \; b_{ij}^{(u)} = f(y_{iu}, y_{ju}), \; b_{ij}^{(v)} = f(y_{iv}, y_{jv}). \tag{3.13}$$

Define c_u and c_v, generalized trend statistics for variables u and v, respectively, as

$$c_u = \sum_{i<j} a_{ij}\, b_{ij}^{(u)}, \quad c_v = \sum_{i<j} a_{ij}\, b_{ij}^{(v)}. \tag{3.14}$$

Then, the null covariance between the two measures c_u and c_v is given in the following theorem.

Theorem 3.1 *The covariance of c_u and c_v, when the joint distribution of $(c_u,\ c_v)$ is discrete uniform over the $n!$ permutations of the rows of $\begin{bmatrix} \mathbf{y}_u & \mathbf{y}_v \end{bmatrix}$ relative to the x's, is:*

$$\begin{aligned} \mathrm{Cov}(c_u, c_v) &= \frac{(\sum_{i,j,k} a_{ij}a_{ik} - \sum_{i,j} a_{ij}^2)(\sum_{i,j,k} b_{ij}^{(u)} b_{ik}^{(v)} - \sum_{i,j} b_{ij}^{(u)} b_{ij}^{(v)})}{n(n-1)(n-2)} \\ &\quad + \frac{(\sum_{i,j} a_{ij}^2)(\sum_{i,j} b_{ij}^{(u)} b_{ij}^{(v)})}{2n(n-1)}, \end{aligned} \tag{3.15}$$

the summation extending over all subscripts from 1 *to* n.

Proof of this theorem is presented in the Appendix. This theorem remains valid in the presence of ties. Results (2.15) and (2.16) can be obtained by plugging appropriate score functions into (3.15).

Corollary 3.1.1 *For the data matrix* $\mathbf{Y}$ *in (2.1), let the Kendall trend statistics for variables* u *and* v *be*

$$k_u = \sum_{i<j} \operatorname{sign}(y_{ju} - y_{iu})$$

and

$$k_v = \sum_{i<j} \operatorname{sign}(y_{jv} - y_{iv}),$$

respectively. Then, with

$$k_{uv} = \sum_{i<j} \operatorname{sign}[(y_{ju} - y_{iu})(y_{jv} - y_{iv})]$$

and

$$s_{uv} = \sum_{i} [\operatorname{rank}(y_{iu}) - (n+1)/2]\,[\operatorname{rank}(y_{iv}) - (n+1)/2],$$

the null covariance of k_u *and* k_v *is*

$$\operatorname{Cov}(k_u, k_v) = k_{uv}/3 + 4\,s_{uv}/3\,. \tag{3.16}$$

Proof of this corollary is given in the Appendix. This result is equivalent to that of Dietz and Killeen (1981), which is their main result.

Corollary 3.1.2 *For the data matrix* $\mathbf{Y}$ *in (2.1), let the Spearman trend statistics for variables* u *and* v *be*

$$s_u = \sum_{i} [i - (n+1)/2]\,[\operatorname{rank}(y_{iu}) - (n+1)/2]\,,$$

and

$$s_v = \sum_i [i - (n+1)/2]\,[\text{rank}\,(y_{iv}) - (n+1)/2]\,,$$

respectively. Then, with

$$s_{uv} = \sum_i [\text{rank}\,(y_{iu}) - (n+1)/2]\,[\text{rank}\,(y_{iv}) - (n+1)/2],$$

the null covariance of s_u and s_v is

$$\text{Cov}\,(s_u, s_v) = n(n+1)\,s_{uv}/12\,. \tag{3.17}$$

Proof of this corollary is presented in the Appendix. This result is also equivalent to that of Bhattacharyya and Klotz (1966). Corollaries 3.1.1 and 3.1.2 verify the correctness of the general approach.

Consider now

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix}, \tag{3.18}$$

where again x_i is time and y_{i1}, y_{i2}, $\ldots$, y_{ip} are environmental variables 1, 2, ..., p, respectively, for $i = 1, 2, \ldots, n$.

Theorem 3.2 *Let $\mathbf{c} = (c_1, c_2, \ldots, c_p)'$ be the vector of p trend statistics defined as in (3.14) and let*

$$\mathbf{S_c} = \{\text{Cov}(c_u, c_v)\} \tag{3.19}$$

be the covariance matrix of $\mathbf{c}$ with covariances defined as in (3.15). Then, under the null hypothesis of no trend, the statistic

$$\mathbf{c}'\,\mathbf{S_c^-}\,\mathbf{c} \tag{3.20}$$

is asymptotically χ^2 with degrees of freedom being the rank of $\mathbf{S_c}$.

Proof of this theorem is given in the Appendix.

Now, a class of multivariate nonparametric trend tests has been constructed. For any scoring system used for a correlation of the form of the generalized correlation coefficient, the corresponding asymtotically distribution-free multivariate test for monotone trend is constructible.

For any member of the class of tests proposed here, if the sample size is small, the randomization test can be used that exactly finds the smallest significant level at which the test rejects based on an observed test statistic by enumerating the statistic values corresponding to $n!$ equally-likely permutations of the observations. For moderate sample sizes, a random subset of the n! permutations can be used for estimation of the P-value. A multivariate nonparametric test based on the test statistic (3.20) can be viewed as an approximate version of the randomization test based on the $n!$ equiprobable permutations of the observations under H_0.

3. 3 A new correlation measure based on signed rank scores

We propose a new correlation measure, say w, as a special case of the generalized correlation coefficient by defining a new scoring system as

$$\begin{aligned} a_{ij} &= \text{sign}(x_j - x_i) \ \text{rank}\left(\mid x_j - x_i \mid\right), \\ b_{ij} &= \text{sign}(y_j - y_i) \ \text{rank}\left(\mid y_j - y_i \mid\right). \end{aligned}$$

The rank here takes a value between 1 and $\binom{n}{2}$.

A motivation for proposing w can be explained as follows: For Kendall's rank correlation, a_{ij} and b_{ij} are the scores we would assign to $(x_i\,,\, x_j)$ and $(y_i\,,\, y_j)$, respectively, if we were to calculate the *sign test statistic* from the $\binom{n}{2}$ pairs of $(x_i\,,\, x_j)$'s and the $\binom{n}{2}$ pairs of $(y_i\,,\, y_j)$'s. For the new correlation coefficient w, a_{ij} and b_{ij} are the scores we would assign to $(x_i\,,\, x_j)$ and $(y_i\,,\, y_j)$, respectively, if we were to calculate the *Wilcoxon signed-rank test statistic* from $\binom{n}{2}$ pairs of $(x_i\,,\, x_j)$'s and $\binom{n}{2}$ pairs of $(y_i\,,\, y_j)$'s.

A new correlation w is designed to contain more information on the association between x and y than Kendall's or Spearman's rank correlation. As for Kendall's rank correlation, 1 or -1 is assigned regardless of how near or far apart the data points are. Spearman's rank correlation statistic gives larger weight to the pairs $y_i < y_j$ for which rank(x_i) and rank(x_j) are further apart, since it is equivalent to the statistic

$$\sum_{i<j} [\text{rank}(x_j) - \text{rank}(x_i)] \ I\{y_j - y_i > 0\} \ ,$$

where

$$I\{\text{A}\} = 1 \ \text{ if A is true, or } 0 \ \text{ otherwise}$$

(Lehmann 1975, page 291). But, in the Spearman statistic, the absolute value of the score assigned to the pair of two adjacent ordered values is always 1 regardless of how large or small their actual difference is. The new correlation coefficient w assigns greater weight to the pair of data points for which the difference between the actual values is larger. As for Pearson's product-moment correlation, the score assigned to the pair of data points is their actual difference so the score may be affected by extreme values. For the new correlation w, the score for the Wilcoxon signed-rank statistic is used as the score for the data pairs so w becomes robust. Usually robustness is obtained at the price of information. The new correlation w is more robust but less informative than Pearson's correlation and more informative but less robust than Kendall's or Spearman's correlation.

In testing $H_0: E(w) = 0$, since w is a special case of the generalized correlation coefficient, the null distribution of the test statistic $w/\sqrt{\mathrm{Var}(w)}$ is asymptotically standard normal.

Following the procedure described in the preceding section, we can construct an asymptotically distribution-free multivariate test for trend based on the new correlation w.

3. 4 Example

Here, four correlation coefficients---Kendall's sample correlation $\hat{\tau}$, Spearman's sample correlation $\hat{\rho}$, Pearson's sample correlation r, and the new sample correlation w---in the universe of sample permutations have been discussed. For illustration of the methodology, all of these four coefficients are used for multivariate trend tests applied to a real data set. A blood constituent data set in a drug trial in Dietz and Killeen (1981) is to be used. A patient had four blood constituents measured at approximately monthly intervals over a period of 2 years. The experimenter wants to test whether there is a significant trend in any of these four blood constituents. The data set is in Table 3.1 on page 23 and the test results are presented in Tables 2 through 5 on pages 24 and 25.

All of the four multivariate tests reject the null hypothesis of no trend at the 5% level, leading to the conclusion that there is a significant trend, which agrees with the univariate test results that found a significant trend in variables 2 and 4. As for the univariate results, we find the two groups of sample correlations in which the ranks of the absolute values of the sample correlations are the same. One group consists of Kendall's and Spearman's correlations and the other consists of the new and Pearson's correlations. Explaining in detail, when Kendall's or Spearman's correlation is used, the array of variables in the descending order of the absolute correlation is: variables 2, 4, 3, and 1, and when Pearson's or new correlation is used, the array of variables in the descending order of the absolute correlation is: variable 4, 2, 3, and 1. This result from this data set may suggest that the new correlation coefficient is relatively close to Pearson's correlation coefficient, as was discussed in the preceding section.

Table 3.1: Blood constituent data

DATE	BILIRUBIN	BLOOD UREA NITROGEN	CREATININE	ALKALINE PHOSPHATASE
01/08/74	0.4	13	0.7	96
02/05/74	0.4	14	0.7	89
02/19/74	0.3	15	0.9	67
03/12/74	0.4	13	0.9	83
04/09/74	0.6	12	0.9	78
05/14/74	0.3	14	0.8	70
06/11/74	0.4	15	0.8	65
07/09/74	0.3	16	0.7	70
08/13/74	0.3	16	0.8	67
09/10/74	0.4	16	0.9	64
10/08/74	0.3	14	0.9	74
11/12/74	0.3	15	0.8	81
12/10/74	0.4	17	0.9	86
01/14/75	0.5	12	0.8	69
02/11/75	0.4	13	0.8	65
03/11/75	0.4	16	0.7	63
04/08/75	0.3	17	0.5	69
05/13/75	0.4	20	0.8	54
06/10/75	0.3	16	0.8	65
07/08/75	0.3	15	0.7	62
08/12/75	0.4	14	0.7	63
09/09/75	0.4	16	0.8	64
10/14/75	0.5	17	0.8	68
11/11/75	0.4	14	0.8	72
12/09/75	0.5	16	0.8	77
01/13/76	0.4	16	0.7	71

Table 3.2: Multivariate trend test based on Kendall's correlation $\hat{\tau}$

VARIABLE	$\hat{\tau}$	P-VALUE	# OF OBS.
1	0.16154	0.3075	26
2	0.33511	0.0239	26
3	-0.21916	0.1612	26
4	-0.28974	0.0400	26

$S_{\hat{\tau}}$, NULL COVARIANCE MATRIX OF $\hat{\tau}$

0.0251	-0.0039	0.0037	0.0028
-0.0039	0.0220	-0.0033	-0.0070
0.0037	-0.0033	0.0245	0.0035
0.0028	-0.0070	0.0035	0.0199

$\hat{\tau}' S_{\hat{\tau}}^{-1} \hat{\tau}$	DF	P-VALUE OF MULTIVARIATE TEST
10.591	4	0.032

Table 3.3: Multivariate trend test based on Spearman's correlation $\hat{\rho}$

VARIABLE	$\hat{\rho}$	P-VALUE	# OF OBS.
1	0.20113	0.3146	26
2	0.45025	0.0244	26
3	-0.26190	0.1904	26
4	-0.39548	0.0480	26

$S_{\hat{\rho}}$, NULL COVARIANCE MATRIX OF $\hat{\rho}$

0.0400	-0.0067	0.0059	0.0051
-0.0067	0.0400	-0.0058	-0.0135
0.0059	-0.0058	0.0400	0.0064
0.0051	-0.0135	0.0064	0.0400

$\hat{\rho}' S_{\hat{\rho}}^{-1} \hat{\rho}$	DF	P-VALUE OF MULTIVARIATE TEST
10.073	4	0.039

Table 3.4: Multivariate trend test based on the new correlation w

VARIABLE	w	P-VALUE	# OF OBS.
1	0.19138	0.3296	26
2	0.41985	0.0316	26
3	-0.32289	0.0977	26
4	-0.42691	0.0283	26

$\mathbf{S_w}$, NULL COVARIANCE MATRIX OF $\mathbf{w}$

0.0385	-0.0081	0.0069	0.0061
-0.0081	0.0381	-0.0073	-0.0154
0.0069	-0.0073	0.0380	0.0060
0.0061	-0.0154	0.0060	0.0379

$\mathbf{w'S_w^{-1}w}$	DF	P-VALUE OF MULTIVARIATE TEST
11.119	4	0.025

Table 3.5: Multivariate trend test based on Pearson's correlation r

VARIABLE	r	P-VALUE	# OF OBS.
1	0.14413	0.4711	26
2	0.41586	0.0376	26
3	-0.26382	0.1871	26
4	-0.45972	0.0215	26

$\mathbf{S_r}$, NULL COVARIANCE MATRIX OF $\mathbf{r}$

0.0400	-0.0099	0.0097	0.0064
-0.0099	0.0400	-0.0074	-0.0172
0.0097	-0.0074	0.0400	0.0029
0.0064	-0.0172	0.0029	0.0400

$\mathbf{r'S_r^{-1}r}$	DF	P-VALUE OF MULTIVARIATE TEST
10.259	4	0.036

3.5 Choice of the correlation statistic for environmental trend assessment

Now, let us discuss what correlation we will choose for assessment of trend in environmental variables. Apparарent candidates are Pearson's product moment correlation, Kendall's rank correlation, Spearman's rank correlation, and a new correlation based on signed rank scores. As censoring frequently occurs in the environmental data, we eliminate Pearson's correlation and the new correlation, because they cannot be applied to the censored data. Now, between Kendall's and Spearman's correlation statistics, our preference is given to the Spearman trend statistic, which is of the form

$$c = \sum_{i<j} [\text{rank}\,(x_j) - \text{rank}\,(x_i)]\,[\text{rank}\,(y_j) - \text{rank}\,(y_i)]\,, \tag{3.21}$$

where x is time (year) and y is a water quality constituent, for the following reasons.

First, Spearman's rank correlation $\hat{\rho}$ has the desirable property that

$$\text{Corr}(\hat{\rho}_{xy},\ \hat{\rho}_{xz}) = \hat{\rho}_{yz}, \tag{3.22}$$

which follows from (2.16). Next, as the number of blocks (seasons) increases, the asymptotic relative efficiency (ARE) of the Kendall-based Jonckheere (1954) test with respect to the Spearman-based Page (1963) test is, for any underlying distribution,

$$\text{ARE}(\sum_u \hat{\tau}_u,\ \sum_u \hat{\rho}_u) = 2(n+1)^2/n(2n+5)\,, \tag{3.23}$$

where n is the number of treatments (years) (Puri and Sen 1971 pp. 324–325, van Belle and Hughes 1984). This is a comparison between the Kendall-based and the Spearman-based covariance sum tests for the case where there are no ties or missing values and seasons are independent. The covariance sum test based on the Spearman trend statistic is slightly more powerful because the ARE in (3.23) lies between 0.96 and 1 for $n \geq 2$. Kendall's rank correlation, $\hat{\tau}$, is known to have the advantage over Spearman's rank correlation, $\hat{\rho}$, of faster convergence to normality (Kendall and Gibbons 1990 p. 69, van Belle and Hughes 1984). For data spanning a short length of time, the asymptotic normal test based on $\hat{\tau}$ will be preferable because its P-value will be more accurate. In our example, however, the period of observation is $n = 17$ years, which is not a very short period: the normal approximation of the tail probabilities, e.g., 0.05 and 0.01, of the null distribution of $\hat{\rho}$ is good when $n = 14$ (Edgington 1980 pp. 214–215). Also, the statistic used to measure trend in each water quality constituent is the sum of 12 asymptotically normal monthly trend statistics. The convergence to normality of such a sum will be even more rapid.

From the next chapter, the Spearman trend statistic will be used in the example of each chapter.

4 A generalized multivariate trend statistic

Lettenmaier (1988) argued that the covariance eigenvalue approach was better than the covariance inversion test when the sample sizes were small because of problems in estimating and inverting the covariance matrix. Another possible cause of the low power of the test may be that the χ^2 approximation is poor because of the small sample size relative to the number of variables. Lettenmaier tried to improve the power of the test using an alternative test statistic, which is the sum of the squared trend statistics. Now, let us consider the generalization of such an alternative statistic.

4.1 A generalized multivariate trend statistic

Here, let us consider a general quadratic form in normal variables. Let $\mathbf{A}$ be a $p \times p$ matrix and let

$$Q = \mathbf{c}' \mathbf{A} \mathbf{c} \tag{4.1}$$

where $\mathbf{c}$ is a multivariate normal vector of size p with $\mathrm{E}(\mathbf{c}) = \mathbf{0}$ and the associate covariance matrix $\mathbf{S_c}$. Then, it is known that the distribution of $Q = \mathbf{c}'\mathbf{A}\mathbf{c}$ is the same as that of

$$\sum_u \lambda_u z_i^2 \tag{4.2}$$

where z_u are independent standard normal variables, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are eigenvalues of $\mathbf{S_c}\,\mathbf{A}$ (Johnson and Kotz 1970 Ch. 29). If $\lambda_u = 1$ for $u = 1, 2, \ldots, p$, then Q has a χ^2 distribution with p degrees of freedom. So, in general, the distribution of Q might be expected to be a distribution related to a χ^2 distribution—either central or noncentral.

While the exact distribution of Q is not known, its moments can be obtained from the following relationship between the s-th cumulant of Q, say κ_s, and eigenvalues of $\mathbf{S_c}\,\mathbf{A}$, λ's:

$$\kappa_s(Q) \;=\; 2^{s-1}\,(s-1)!\,\sum_{u=1}^{p}\lambda_u^s \;=\; 2^{s-1}\,(s-1)!\,\mathrm{trace}[(\mathbf{S_c}\,\mathrm{A})^s]. \tag{4.3}$$

(Johnson and Kotz 1970 Ch. 29). Based on the relationship between cumulants and moments and the result in (4.3), the first three moments of Q are obtained as

$$\kappa_1(Q) \;=\; \mu_1'(Q) \;=\; \mathrm{E(Q)} \;=\; \sum_{\mathrm{u}}\lambda_{\mathrm{u}} \;=\; \mathrm{trace}[\mathbf{S_c}\,\mathbf{A}], \tag{4.4}$$

$$\kappa_2(Q) \;=\; \mu_2(Q) \;=\; \mathrm{Var(Q)} \;=\; 2\sum_{\mathrm{u}}\lambda_{\mathrm{u}}^2 \;=\; 2\,\mathrm{trace}[(\mathbf{S_c}\,\mathbf{A})^2], \tag{4.5}$$

and

$$\kappa_3(Q) \;=\; \mu_3(Q) \;=\; \mathrm{E[(Q-E[Q])^3]} = 8\sum_{\mathrm{u}}\lambda_{\mathrm{u}}^3 \;=\; 8\,\mathrm{trace}[(\mathbf{S_c}\,\mathbf{A})^3]. \tag{4.6}$$

A section in Johnson and Kotz (1970 Ch. 29) on approximations of the distribution of Q includes discussion on approximations by either a scaled noncentral χ^2 distribution and a scaled and shifted central χ^2 distribution. In the former case, Q is approximated by

$$\beta\,{\chi'_{\nu,\,\alpha}}^2 \tag{4.7}$$

where α, β, and ν are positive numbers and ${\chi'_{\nu,\,\alpha}}^2$ denotes a variable having a noncentral χ^2 distribution with ν degrees of freedom and noncentrality α. In the latter case, Q is approximated by

$$\alpha + \beta\,\chi^2_\nu. \tag{4.8}$$

where α, β, and ν are also positive numbers and χ^2_ν denotes a variable having a central χ^2 distribution with ν degrees of freedom.

In either case, the values of parameters are chosen to make the first three moments as functions of parameters agree with those as functions of eigenvalues of $\mathbf{S_c}\,\mathbf{A}$.

Now, a question that may arise at this point is: Is there a criterion by which we determine what to choose between approximations (4.7) and (4.8)? The answer is yes. That criterion is a quantity

$$K \;=\; 2\,[\mu_2(Q)]^2 \;-\; \mu_1'(Q)\,\mu_3(Q) \tag{4.9}$$

and the rule is:

- Choose approximation (4.7) if $K \geq 0$;
- Choose approximation (4.8) if $K \leq 0$

(Johnson and Kotz 1970). If the distribution is a central χ^2 distribution with p degrees of freedom, K becomes 0, because the first three moments of a χ^2 distribution with p degrees of freedom are

$$\mu_1' \;=\; p, \tag{4.10}$$

$$\mu_2 \;=\; 2\,p, \tag{4.11}$$

and

$$\mu_3 \;=\; 8\,p. \tag{4.12}$$

The reason for this decision rule becomes evident if we express the first three moments in terms of appropriate parameters.

For approximation using a scaled noncentral χ^2 variable, the first three moments of Q, which is set to be equal to $\beta\,{\chi'_{\nu,\,\alpha}}^2$, are:

$$\mu_1' \;=\; \beta\,(\nu+\alpha), \tag{4.13}$$

$$\mu_2 = 2\beta^2(\nu + 2\alpha), \tag{4.14}$$

and

$$\mu_3 = 8\beta^3(\nu + 3\alpha). \tag{4.15}$$

In this case,

$$K = 2\mu_2^2 - \mu_1'\mu_3 = 8\beta^4\alpha^2 \tag{4.16}$$

which must be positive. The values of parameters satisfying equations (4.13)–(4.15) are:

$$\beta = \frac{\mu_3}{2}\left(2\mu_2 + \sqrt{4\mu_2^2 - 2\mu_1'\mu_3}\right)^{-1}, \tag{4.17}$$

$$\nu = \frac{2\mu_1'}{\beta} - \frac{\mu_2}{2\beta^2}, \tag{4.18}$$

and

$$\alpha = \frac{\mu_2}{2\beta^2} - \frac{\mu_1'}{\beta}. \tag{4.19}$$

[The values of parameters presented on page 166 in chapter 29 of Johnson and Kotz (1971) are not all correct. The solutions (4.17)–(4.19) given here are correct ones.]

For approximation using a shifted and scaled central χ^2 variable, the first three moments of Q, which is set to be equal to $\alpha + \beta\chi_\nu^2$, are:

$$\mu_1' = \alpha + \beta\nu, \tag{4.20}$$

$$\mu_2 = 2\beta^2\nu, \tag{4.21}$$

and

$$\mu_3 = 8\beta^3\nu. \tag{4.22}$$

In this case,

$$K = 2\mu_2^2 - \mu_1'\mu_3 = -8\alpha\beta^3\nu^2 \tag{4.23}$$

which must be negative. The values of parameters satisfying equations (4.20)–(4.22) are:

$$\alpha = \mu_1' - \frac{2\,\mu_2^2}{\mu_3}, \tag{4.24}$$

$$\beta = \frac{\mu_3}{4\,\mu_2}, \tag{4.25}$$

and

$$\nu = \frac{8\,\mu_2^3}{\mu_3^2}. \tag{4.26}$$

If the distribution of Q is a χ^2 distribution with p degrees of freedom, then its moments are as in (4.10)–(4.12) and in both approximations, the parameter values become

$$\alpha = 0, \;\; \beta = 1, \;\; \nu = p, \tag{4.27}$$

i.e., each of the both approximated distributions reduce to a χ^2 distribution with p degrees of freedom. Now, an approximation rule for the distribution of $Q = \mathbf{c}'\mathbf{A}\mathbf{c}$ has been established.

In multivariate trend assessment, $\mathbf{c}$ is a generalized trend statistic vector which is asymptotically multivariate normal. Here, we propose $Q = \mathbf{c}'\,\mathbf{A}\,\mathbf{c}$ as a *generalized multivariate trend statistic.* In traditional covariance inversion tests,

$$\mathbf{A} = \mathbf{S}_\mathbf{c}^{-1} \tag{4.28}$$

and so $\lambda_u = 1$ for all u as eigenvalues of $\mathbf{S}_\mathbf{c}\,\mathbf{S}_\mathbf{c}^{-1} = \mathbf{I}$, which reduces either approximated null distribution to a χ^2 distribution with p degrees of freedom. In Lettenmaier-suggested covariance eigenvalue tests,

$$\mathbf{A} = \mathbf{I} \tag{4.29}$$

and so λ_u's are eigenvalues of $\mathbf{S_c}\,\mathbf{I} = \mathbf{S_c}$, which determine the type of the approximated null distribution and the values of associated parameters according to the approximation rule described above.

Let us consider other forms of the generalized multivariate trend statistic. The basic form of the Lettenmaier-suggested covariance eigenvalue test statistic is

$$\mathbf{c}'\,\mathbf{c}\,, \tag{4.30}$$

where $\mathbf{c}$ is a generalized trend statistic vector. We can make some variations on the covariance eigenvalue test statistic such as

$$\mathbf{z}'\,\mathbf{z} \tag{4.31}$$

and

$$\mathbf{r}_{\mathbf{g}}'\,\mathbf{r}_{\mathbf{g}}\,, \tag{4.32}$$

where $\mathbf{z}$ is a standardized $\mathbf{c}$ and $\mathbf{r_g}$ is $\mathbf{c}$ converted to a generalized correlation coefficient vector. Statistic (4.31) is $Q = \mathbf{c}'\,\mathbf{A}\,\mathbf{c}$ with

$$\mathbf{A} = [\mathrm{Diag}(\mathbf{S_c})]^{-1} \tag{4.33}$$

and statistic (4.32) is $Q = \mathbf{c}'\,\mathbf{A}\,\mathbf{c}$ with

$$\mathbf{A} = [\mathrm{Diag}\{\sum_{i,j} a_{ij}^2 \sum_{i,j} b_{ij}^{(u)^2}\}]^{-1}\,, \tag{4.34}$$

where each diagonal term $\sum_{i,j} a_{ij}^2 \sum_{i,j} b_{ij}^{(u)^2}$ is the square of the denominator of the generalized correlation between variable u and time, which is defined in Chapter 3. If there are no ties or missing values and if Spearman's or Kendall's statistic is used, in each of the diagonal matrices (4.33) and (4.34), diagonal elements become all the same, i.e., each of the diagonal matrices (4.33) and (4.34) has the form of

$$\mathbf{A} = b\mathbf{I}, \tag{4.35}$$

where b is a constant. Then, each of the variations of the basic covariance eigenvalue test statistic is just a multiple of the basic form. This implies that only when there are no ties or missing values are these variations equivalent in testing to the basic form.

Now, let us apply these tests to a real data set and take a look at what happens.

4.2 Example

As an example for multivariate trend tests, consider the data displayed in Figure 1.1 on page 4, a data set on water quality collected from the Clinch River watershed in southwestern Virginia, USA (Zipper et al. 1992). The data set has 5 variables—dissolved oxygen (DO), pH, nonfilterable residue (NFR), filterable residue (FR), and fecal coliforms (FC)—taken on a monthly basis for 17 years (1973-1989).

Table 4.1 on page 37 gives some results of multivariate Spearman trend tests based on two views. First, view the five constituents and twelve months as defining 60 variables and 60 trend statistics. Then, obtain $\mathbf{c}^*$, a vector of 60 ($= 5 \times 12$) Spearman trend statistics, and $\mathbf{S}_{\mathbf{c}^*}$, the associated covariance matrix of dimension 60×60. The covariance inversion test, with test statistic

$$\mathbf{c}^{*\prime}\,\mathbf{S}_{\mathbf{c}^*}^{-1}\,\mathbf{c}^*, \tag{4.36}$$

indicated that the null hypothesis of no trend is not rejected; the P-value is almost 1. This test has low power because of the small number of observations relative to the large number of variables. The three versions of the covariance eigenvalue test statistic are

$$\mathbf{c}^{*\prime}\,\mathbf{c}^{*}, \tag{4.37}$$

$$\mathbf{z}^{*\prime}\,\mathbf{z}^{*}, \tag{4.38}$$

and

$$\hat{\rho}'\,\hat{\rho}, \tag{4.39}$$

where $\mathbf{z}^*$ is the standardized $\mathbf{c}^*$ and $\hat{\rho}$ is $\mathbf{c}^*$ converted to the correlation vector and is called the $\hat{\rho}$ vector. These three covariance eigenvalue tests produce quite different results. The P-values for the covariance eigenvalue tests are considerably smaller than the P-value of the covariance inversion test and two of them support rejection of the hypothesis of no trend at the 5% level.

In the second analysis, the data are viewed as consisting of five variables. Trend statistics are computed for each combination of variable and month. The trend statistics are then summed over the twelve months. The sum of 12 monthly trend statistics is called the *overall trend statistic.* The covariance inversion test statistic is then given by

$$(\mathbf{Mc}^*)'\,(\mathbf{MS}_{\mathbf{c}^*}\mathbf{M}')^{-1}(\mathbf{Mc}^*), \tag{4.40}$$

where $\mathbf{M}$ is a 5×60 matrix of the form of (2.25) that sums trend statistics across the months. Three versions of the covariance eigenvalue test statistic are

$$(\mathbf{Mc}^*)'\,(\mathbf{Mc}^*), \tag{4.41}$$

$$\mathbf{z}'_{\cdot}\mathbf{z}_{\cdot}, \tag{4.42}$$

and

$$\hat{\rho}'_{\mathrm{o}}\,\hat{\rho}_{\mathrm{o}}, \tag{4.43}$$

where $\mathbf{z}$ is the standardized $\mathbf{Mc}^*$ and $\hat{\rho}_o$ is $\mathbf{Mc}^*$ converted to the correlation vector and is called the *overall* $\hat{\rho}$ vector. From Table 4.1, we see that the covariance inversion test statistic is significant at the 5% level and that the covariance eigenvalue test statistics are significant at the 1% level.

What is clear from the example is that the covariance inversion test will have low power if the sample size is small relative to the number of variables. In a simulation study on multivariate trend tests based on the Kendall statistics with 10- and 20-year data records having 3 variables and 4 seasons, the covariance eigenvalue test was more powerful than the covariance inversion test for the 10-year period, but its power advantage over the covariance inversion test disappeared for the 20-year period (Loftis et al. 1991). There are variations on the covariance eigenvalue test and the relative merits of these alternative statistics need to be investigated.

Table 4.1: Results of multivariate trend tests based on Spearman's $\hat{\rho}$

View	Test Stat.	Observed Test Stat.	K in (4.9)	α	β	ν (d.f)	P-value
60	(4.36)	32.11	0	0.00	1.00	60.00	0.99881
trend	(4.37)	69775983.50	< 0	14993672.04	3295704.06	8.76	0.04937
stat.	(4.38)	88.19	< 0	18.99	3.13	12.80	0.04941
	(4.39)	7.00	< 0	1.49	0.24	13.15	0.05367
5	(4.40)	14.20	0	0.00	1.00	5.00	0.01438
overall	(4.41)	230918772.00	< 0	9873611.40	20504324.53	2.57	0.00856
trend	(4.42)	21.29	< 0	0.57	1.36	3.24	0.00217
stat.	(4.43)	0.22	< 0	0.01	0.02	3.14	0.00210

5 A canonical analysis of multivariate trends

If a multivariate trend test detects the presence of trend, one would want to identify the variables that are important contributors to the overall trend. For this purpose, a canonical analysis can be used with the covariance inversion test.

5.1 Estimation in canonical analysis

Given $\mathbf{c} = (c_1, c_2, \ldots, c_p)'$, the vector of p trend statistics, and $\mathbf{S_c}$, the covariance matrix of $\mathbf{c}$, we seek a linear combination of $\mathbf{c}$, say $\mathbf{q}'\mathbf{c}$, which is in some sense the best one-dimensional summary of the p trends. Under the null hypothesis of no trend, we want to find a vector of coefficients $\mathbf{q}$ that maximizes the absolute value of the standardized $\mathbf{q}'\,\mathbf{c}$. Because $\mathrm{E}(\mathbf{q}'\,\mathbf{c}) = 0$ and $\mathrm{Var}(\mathbf{q}'\,\mathbf{c}) = \mathbf{q}'\,\mathbf{S_c}\,\mathbf{q}$ under the null hypothesis, the standardized $\mathbf{q}'\,\mathbf{c}$ is $(\mathbf{q}'\,\mathbf{c})(\mathbf{q}'\,\mathbf{S_c}\,\mathbf{q})^{-1/2}$. Because the absolute value of the standardized $\mathbf{q}'\,\mathbf{c}$ is the same as the absolute value of the standardized $b\,\mathbf{q}'\,\mathbf{c}$ for any constant b, we avoid indeterminacy by imposing the restriction that $\mathbf{q}'\,\mathbf{S_c}\,\mathbf{q} = 1$. Now, maximizing the absolute value being equivalent to maximizing the squared value, we make a definition as follows:

Definition 5.1 *Let* $\mathbf{c}$ *be a vector with* $\mathrm{E}(\mathbf{c}) = 0$ *and* $\mathbf{S_c}$ *be the associated sample covariance matrix. For*

$$\frac{(\mathbf{q}'\,\mathbf{c})^2}{\mathbf{q}'\,\mathbf{S_c}\,\mathbf{q}} = \max_{\mathbf{q}_\circ} \left[\frac{(\mathbf{q}_\circ'\,\mathbf{c})^2}{\mathbf{q}_\circ'\,\mathbf{S_c}\,\mathbf{q}_\circ}\right], \tag{5.1}$$

where $\mathbf{q}_\circ$ *denotes any member in the set of* $p \times 1$ *vectors satisfying the restriction that* $\mathbf{q}_\circ'\,\mathbf{S_c}\,\mathbf{q}_\circ = 1$,

- $\mathbf{q}$ *is said to be the sample canonical coefficient vector for* $\mathbf{c}$.

- $\mathbf{q}'\mathbf{c}$ *is said to be the sample canonical variate for* $\mathbf{c}$.

We can make a similar definition using $\boldsymbol{\Sigma}_{\mathbf{c}}$, the population covariance matrix of $\mathbf{c}$.

Definition 5.2 *Let* $\mathbf{c}$ *be a vector with* $\mathrm{E}(\mathbf{c}) = 0$ *and* $\boldsymbol{\Sigma}_{\mathbf{c}}$ *be the associated population covariance matrix. For*

$$\frac{(\xi'\,\mathbf{c})^2}{\xi'\,\boldsymbol{\Sigma}_{\mathbf{c}}\,\xi} = \max_{\xi_o} \left[\frac{(\xi_o'\,\mathbf{c})^2}{\xi_o'\,\boldsymbol{\Sigma}_{\mathbf{c}}\,\xi_o}\right], \tag{5.2}$$

where ξ_o *denotes any member in the set of* $p \times 1$ *vectors satisfying the restriction that* $\xi_o'\,\boldsymbol{\Sigma}_{\mathbf{c}}\,\xi_o = 1$,

- ξ *is said to be the population canonical coefficient vector for* $\mathbf{c}$.

- $\xi'\mathbf{c}$ *is said to be the population canonical variate for* $\mathbf{c}$.

The sample canonical coefficients and the canonical variate are found to be related to the eigenstructure of $\mathbf{S}_{\mathbf{c}}^{-}\,\mathbf{c}\,\mathbf{c}'$, where $\mathbf{S}_{\mathbf{c}}^{-}$ is the inverse of $\mathbf{S}_{\mathbf{c}}$ if it is nonsingular, or a generalized inverse of $\mathbf{S}_{\mathbf{c}}$ if it is singular.

Theorem 5.1 *Let* $\mathbf{c}$ *be a vector with* $\mathrm{E}(\mathbf{c}) = 0$ *and* $\mathbf{S}_{\mathbf{c}}$ *be the associated sample covariance matrix. Let* $(\lambda,\, \mathbf{q})$ *denote the nonzero eigenvalue and the corresponding eigenvector of* $\mathbf{S}_{\mathbf{c}}^{-}\,\mathbf{c}\,\mathbf{c}'$. *Then, provided* $\mathbf{q}'\,\mathbf{S}_{\mathbf{c}}\,\mathbf{q} = 1$,

- $\mathbf{q}$ *is the sample canonical coefficient vector for* $\mathbf{c}$, (5.3)

- $\lambda = (\mathbf{q}'\,\mathbf{c})^2 = \mathbf{c}'\,\mathbf{S}_{\mathbf{c}}^{-}\,\mathbf{c}$. (5.4)

The proof is in the Appendix. We see now that λ, the nonzero eigenvalue of $\mathbf{S}_{\mathbf{c}}^{-}\,\mathbf{c}\,\mathbf{c}'$, is equal to the χ^2 statistic for the related multivariate trend test. Moreover, the corresponding eigenvector $\mathbf{q}$ assigns weights to the variables in the linear combination that summarizes the p trends in one dimension. The test for significance of the sample canonical variate is equivalent to the test for a monotone trend in one or more of the p variables. In a trend analysis for assessment of environmental impact, canonical analysis can produce a single measurement of hazard or stress that is a weighted combination of individual indicators.

A canonical coefficient q_u for the trend statistic c_u is closely related to the magnitude and variance of c_u and covariances among the c_u's. A canonical coefficient q_u can be regarded as an indicator of the contribution to the overall trend made by variable u in the presence of other variables. The overall trend $\mathbf{q}'\mathbf{c}$ is a linear combination of all trends in which trends with different directions may not be canceled. The interpretation of these canonical coefficients is similar to that of multiple regression coefficients.

When there are missing values and/or ties, the variances of nonparametric trend statistics become different from each other. If variances of trend statistics are different, we cannot compare corresponding canonical coefficients. For comparability among canonical coefficients, it is suggested that we use the standardized $\mathbf{c}$, say $\mathbf{z}$, that is obtained by premultiplying $\mathbf{c}$ by the inverse of the diagonal matrix consisting of the standard errors of c_u's. In this case, the corresponding covariance matrix is the correlation matrix of $\mathbf{c}$.

5.2 Testing in canonical analysis

Here, let us consider testing canonical coefficients. A researcher may think of a simpler form of a canonical variate as a represenative summary of p trends. For example, in the case where a single variable is measured in each of 4 seasons, a researcher may want to test whether the unweighted sum of 4 seasonal trend statistics, i.e.,

$$c_1 + c_2 + c_3 + c_4 \tag{5.5}$$

could be used as a measure of overall trend. In other words, he/she is interested in testing the following relationship between population canonical coefficients:

$$\xi_1 = \xi_2 = \xi_3 = \xi_4 \tag{5.6}$$

or

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5.7}$$

As a second example, in the case where there are 5 variables, the reasercher may want to test if the marginal contribution of trend statistic for variable 3 to the canonical variate, which is equivalent to the covariance inversion multivariate trend statistic, is significant. Then, the null hypothesis in terms of canonical coefficients is:

$$\xi_3 = 0 \tag{5.8}$$

or

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{bmatrix} = 0. \tag{5.9}$$

This is a test for the significance of a canonical coefficient.

For a third case, if a researcher wants to test whether the marginal contributions of trend statistics for variables 1 and 2, c_1 and c_2, and to the canonical variate are the same and the marginal contributions of upward trend statistic for variable 4 is the same as that of downward trend statistic for varaible 5 then his/her null hypothesis is:

$$\xi_1 = \xi_2, \;\; \xi_4 = -\xi_5 \tag{5.10}$$

or

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{5.11}$$

Now, we will develop a test of a general linear hypothesis for canonical coefficients, i.e.,

$$\mathrm{H}_0 : \; \mathbf{H}\,\xi \;=\; \mathbf{0}, \tag{5.12}$$

where $\mathbf{H}$ is an $m \times p$ matrix ($m \leq p$) and ξ is a population canonical coefficient vector of size p. First, we need to estimate canonical coefficients under the restriction that $\mathbf{H}\,\xi \;=\; \mathbf{0}$, which implies some relatioship among canonical coefficients.

Theorem 5.2 *Let* $\mathbf{c}$ *be a vector with* $\mathrm{E}(\mathbf{c}) = 0$ *and* $\mathbf{S_c}$ *be the associated sample covariance matrix. Let* $\mathbf{G} = \mathbf{I} - \mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}\mathbf{H}$. *Let* $(\lambda_o, \mathbf{q}_o)$ *denote the nonzero eigenvalue and the corresponding eigenvector of* $(\mathbf{G}\mathbf{S_c}\mathbf{G})^-(\mathbf{G}\mathbf{c})(\mathbf{G}\mathbf{c})'$, *where* $(\mathbf{G}\mathbf{S_c}\mathbf{G})^-$ *is a generalized inverse of* $\mathbf{G}\mathbf{S_c}\mathbf{G}$. *Then, provided* $\mathbf{q}_o{}'\,\mathbf{S_c}\,\mathbf{q}_o = 1$,

- $\mathbf{q}_o$ *is the sample canonical coefficient vector for* $\mathbf{c}$, (5.13)

  under the restriction that $\mathbf{H}\,\xi \;=\; \mathbf{0}$.

- $\lambda_o = (\mathbf{q}_o{}'\,\mathbf{c})^2 = (\mathbf{G}\mathbf{c})'(\mathbf{G}\mathbf{S_c}\mathbf{G})^-(\mathbf{G}\mathbf{c})$ (5.14)

  under the restriction that $\mathbf{H}\,\xi \;=\; \mathbf{0}$.

The proof of this theorem is given in Appendix. Here, the distribution of

$$(\mathbf{G}\mathbf{c})'(\mathbf{G}\mathbf{S_c}\mathbf{G})^-(\mathbf{G}\mathbf{c})$$

is asymptotically χ^2 with degrees of freedom equal to the rank of $\mathbf{G}\mathbf{S}_\mathbf{c}\mathbf{G}$.

Using these results, we can now contruct a test for $\mathrm{H}_0 : \mathbf{H}\boldsymbol{\xi} = \mathbf{0}$, using a "full model statistic minus reduced model statistic" technique, as follows:

- *The test statistic, T, is, with* $\mathbf{G} = \mathbf{I} - \mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}\mathbf{H}$,

$$\begin{aligned} T &= \chi^2(\text{full model}) - \chi^2(\text{model under } \mathrm{H}_0) \qquad (5.15) \\ &= \mathbf{c}'\,\mathbf{S}_\mathbf{c}^-\,\mathbf{c} - (\mathbf{G}\mathbf{c})'\,(\mathbf{G}\mathbf{S}_\mathbf{c}\mathbf{G})^-\,(\mathbf{G}\mathbf{c}). \end{aligned}$$

- *The null distribution of T is is asymptotically χ^2 with degrees of freedom equal to* $\mathrm{rank}(\mathbf{S}_\mathbf{c}) - \mathrm{rank}(\mathbf{G}\mathbf{S}_\mathbf{c}\mathbf{G})$.

Notice that a test for equality of the canonical coefficients, for exmaple, the hypothsis in (5.7), can be used for checking the validity of the covariance sum test.

5.3 Example

Table 5.1 on page 46 contains results of univariate trend and canonical analysis. For comparability among canonical coefficients, standardized overall $\hat{\rho}$'s are used in the analysis. The canonical coefficients corresponding to standardized trend statistics are called standardized canonical coefficients. The largest canonical coefficient corresponds to the largest standardized overall $\hat{\rho}$ and the signs of canonical coefficients and corresponding overall $\hat{\rho}$'s are the same. But, the rank of the absolute value of a canonical coefficient is not always the same as that of the corresponding standardized overall $\hat{\rho}$. This may be due to the correlation structure of overall $\hat{\rho}$'s (Table 5.2 on page 46).

From the results of tests for significance of canonical coefficients, we see that only NFR has a canonical coefficient significant at the 5% level. (NFR also has the largest correlation.) However, it seems that in canonical analysis, the relative magnitudes of canonical coefficients are more important in interpretation than P-values for individual canonical coefficients.

Although canonical coefficients in Table 5.1 are apparently different from each other, let us apply the test for equality of canonical coefficients to this data set and take a look at what happens. The null hypothesis here is:

$$\xi_1 = \xi_2 = \xi_3 = \xi_4 = \xi_5 \tag{5.16}$$

or

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5.17}$$

Canonocal coeficient estimation under this null hypothesis and the test result are presented in Tables 5.3 and 5.4 on page 47. As we guess before testing, the null hypothesis of equality between canonical coefficients is rejected at 5% level. Thus, the covariance sum test should not be used in this situation.

As another example, let us test the following null hypothesis using the same data set.

$$\mathrm{H}_0 : \ \xi_1 = -\xi_3, \ \xi_3 = \xi_5, \ \xi_2 = \xi_4 = 0 \tag{5.18}$$

or

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5.19}$$

Canonical coeficient estimation under this null hypothesis and the test result are presented in Tables 5.5 and 5.6 on page 47. The test result is the acceptance of this null hypothesis. Notice that χ^2 value of this parsimonious model is over 90% (=12.89/14.20) of that of the full model.

Table 5.1: Overall $\hat{\rho}$ and canonical coefficient for each constituent

Constituent	Overall $\hat{\rho}$	P-value	Standardized Overall $\hat{\rho}$	Standardized Canonical Coefficient	P-value for Canonical Coefficient	Number of Obs.
DO	0.23	0.0179	2.37	0.27	0.3996	167
PH	0.13	0.2377	1.18	0.18	0.5614	167
NFR	-0.30	0.0037	-2.90	-0.64	0.0205	158
FR	-0.07	0.4094	-0.82	-0.20	0.4773	163
FC	-0.24	0.0226	-2.28	-0.39	0.1907	162

Table 5.2: Correlation between overall $\hat{\rho}$'s

Constituent	DO	pH	NFR	FR	FC
DO	1.00	0.36	-0.21	-0.05	-0.38
pH	0.36	1.00	-0.21	0.31	0.08
NFR	-0.21	-0.21	1.00	-0.05	0.12
FR	-0.05	0.31	-0.05	1.00	0.22
FC	-0.38	0.08	0.12	0.22	1.00

Table 5.3: Canonical coefficient estimation under H_0: $\xi_1 = \xi_2 = \xi_3 = \xi_4 = \xi_5$

Constituent	Canonical Coefficient under H_0	Standardized Trend Statistic
DO	0.43	2.37
pH	0.43	1.18
NFR	0.43	–2.90
FR	0.43	–0.82
FC	0.43	–2.28

Table 5.4: Test of H_0: $\xi_1 = \xi_2 = \xi_3 = \xi_4 = \xi_5$

	χ^2	d.f.	P-value
Full Model	14.20313	5	0.01437
Model under H_0	1.11571	1	0.29084
Remainder (Test Statistic)	13.08743	4	0.01086

Table 5.5: Canonical coefficient estimation under H_0: $\xi_1 = -\xi_3,\ \xi_3 = \xi_5,\ \xi_2 = \xi_4 = 0$

Constituent	Canonical Coefficient under H_0	Standardized Trend Statistic
DO	0.48	2.37
pH	0.00	1.18
NFR	-0.48	-2.90
FR	0.00	-0.82
FC	-0.48	-2.28

Table 5.6: Test of H_0: $\xi_1 = -\xi_3,\ \xi_3 = \xi_5,\ \xi_2 = \xi_4 = 0$

	χ^2	d.f.	P-value
Full Model	14.20313	5	0.01437
Model under H_0	12.89649	1	0.00033
Remainder (Test Statistic)	1.30664	4	0.86025

6 ANOVA-like analyses of multivariate trends

van Belle and Hughes (1984) describe an approach to testing for homogeneity of trend when there are several seasons and/or stations. For each of p seasons, a standardized trend statistic, say z, is computed. Then the z's are squared and summed. If the number of years is large enough and if the seasons are independent, the sum of the squared standardized statistics is approximately χ^2_p. This χ^2_p statistic can then be decomposed into a χ^2_1 statistic for testing trend assuming homogeneity in trend across seasons and a χ^2_{p-1} statistic for testing homogeneity in trend between seasons (van Belle and Hughes 1984). They also provided a table like an ANOVA table for testing heterogeneity in trend of a single variable between seasons and sites under the assumption that seasons are independent. In this chapter, extensions to the correlated and multivariate cases are made.

6.1 Case of a single variable with multiple seasons

As the approach of van Belle and Hughes can be expressed as tests involving contrasts, the extension to the case where seasons are dependent is straightforward. Suppose the data are collected for twelve seasons, i.e., $p = 12$. Let

$$\mathbf{z} = (z_1,\ z_2,\ ,z_3,\ z_4,\ z_5,\ z_6,\ ,z_7,\ z_8,\ z_9,\ z_{10},\ ,z_{11},\ z_{12})$$

be the vector of standardized generalized trend statistics and let $\mathbf{S_z}$ be the covariance matrix of $\mathbf{z}$ which, because of the standardization, is a correlation matrix. Consider matrices $\mathbf{C}_1$ and $\mathbf{C}_2$, where

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \qquad (6.1)$$

and

$$
\mathbf{C}_2 = \begin{bmatrix}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1
\end{bmatrix}. \tag{6.2}
$$

Then,

$$\chi^2(\mathbf{C}_j) = (\mathbf{C}_j\mathbf{z})'\,(\mathbf{C}_j\mathbf{S}_\mathbf{z}\mathbf{C}_j')^{-1}(\mathbf{C}_j\mathbf{z}) \tag{6.3}$$

has asymptotically a χ_1^2 distribution for $j = 1$ and a χ_{p-1}^2 distribution for $j = 2$. The statistic $\chi^2(\mathbf{C}_1)$ is used for a test of trend assuming homogeneity or a common direction in trend, which is sometimes called a covariance sum test, and the statistic $\chi^2(\mathbf{C}_2)$ is used for assessment of homogeneity in trend. The validity of the χ^2 approximation depends on the number of time points and the number of seasons (van Belle and Hughes 1984). The covariance eigenvalue approach can be used to approximate the distribution and percentage points of the statistic in cases where sample sizes are small. The covariance eigenvalue test statistic corresponding to (6.3) is

$$Q(\mathbf{C}_j) = (\mathbf{C}_j\mathbf{z})'\,(\mathbf{C}_j\mathbf{z}) \;=\; \mathbf{z}'\mathbf{C}_j'\mathbf{C}_j\mathbf{z} \tag{6.4}$$

and its approximated null distribution depends on the eigenvlaues of $\mathbf{S}_\mathbf{z}\mathbf{C}_j'\mathbf{C}_j$.

If seasons are uncorrelated, $\chi^2 = \mathbf{z}'\,\mathbf{S}_\mathbf{z}^{-1}\,\mathbf{z}$ can be decomposed into $\chi^2(\mathbf{C}_1)$ and $\chi^2(\mathbf{C}_2)$, that is, with $\mathbf{S}_\mathbf{z}$ being an identity matrix,

$$\chi^2 = \mathbf{z}'\,\mathbf{S}_\mathbf{z}^{-1}\,\mathbf{z} = \chi^2(\mathbf{C}_1) + \chi^2(\mathbf{C}_2). \tag{6.5}$$

Under dependence among seasons, the following decomposition is possible. With $\mathbf{S_z}$ being a correlation matrix,

$$\chi^2 = \mathbf{z}'\,\mathbf{S}_{\mathbf{z}}^{-1}\,\mathbf{z} = \chi^2(\mathbf{C}_1) + \text{remainder}, \tag{6.6}$$

where the remainder is considered what remains in all p trends after an average trend is removed. The insignificance of the remainder validates using $\chi^2(\mathbf{C}_1)$ as a test statistic for the test for an overall trend. The remainder is asymptotically χ^2_{p-1}, which is independent of $\chi^2(\mathbf{C}_1)$. Under dependence, $\chi^2(\mathbf{C}_1)$ and $\chi^2(\mathbf{C}_2)$ are not independent and the decomposition as in (6.5) does not hold.

6.1.1 Example

An example of trend analysis for a single variable, DO, is given in Tables 6.1—6.5 on pages 53 and 54 using the Spearman trend statistic. Table 4 gives the trend tests for each of 12 seasons. Trends in nine of twelve months are upward and sample sizes range from 11 to 16 (Table 6.1). Overall $\hat{\rho}$, which is an overall estimate of the correlation, is 0.23 (See Table 5.1 on 46). This estimate is computed as a weighted average of the 12 monthly Spearman $\hat{\rho}$'s.

To test for the presence of at least one monthly trend, the twelve months are treated as twelve variables and the covariance inversion test was applied. The test statistic is not significant (Table 6.2). This statistic was decomposed using

$$\chi^2 = \mathbf{c}'\,\mathbf{S}_{\mathbf{c}}^{-1}\,\mathbf{c} = (\mathbf{C}_1\mathbf{c})'\,(\mathbf{C}_1\mathbf{S}_{\mathbf{c}}\mathbf{C}_1')^{-1}(\mathbf{C}_1\mathbf{c}) + \text{remainder}, \tag{6.7}$$

where

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{6.8}$$

The test for overall $\hat{\rho}$ is based on $\chi^2(\mathbf{C}_1)$. The remainder indicates what remains in all 12 monthly trends after an average monthly trend is removed. If this remainder is not significant, overall $\hat{\rho}$ can be considered to represent the overall trend. [Notice that this test is equivalent to the test of equality of canonical coefficients discussed in the preceding chapter.] The remainder is insignificant ($\chi^2 = 12.14$ with 11 degrees of freedom). So, a test for overall test based on $\chi^2(\mathbf{C}_1)$ is considered valid and the test rejects the null hypothesis of no trend at the 5% level.

Here, the covariance inversion multivariate test for the presence of at least one monthly trend, whose test statistic is insignificant ($\chi^2 = 17.75$ with 12 degrees of freedom), seems to be too conservative. The three versions of the covariance eigenvalue multivariate test all resulted in significance at the 5% level (Table 6.3).

Table 6.4 contains the test results based on test statistics

$$\chi^2(\mathbf{C}_j) = (\mathbf{C}_j\mathbf{z})'\,(\mathbf{C}_j\mathbf{S}_\mathbf{z}\mathbf{C}_j')^{-1}(\mathbf{C}_j\mathbf{z}),\ \ j = 1,\,2,$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are given as in (6.1) and (6.2). The null hypothesis of trend homogeneity is accepted and the null hypothesis of no overall trend is rejected. Also, the covariance eigenvalue tests based on test statistics

$$Q(\mathbf{C}_j) = (\mathbf{C}_j\mathbf{z})'\,(\mathbf{C}_j\mathbf{z})\ =\ \mathbf{z}'\mathbf{C}_j'\mathbf{C}_j\mathbf{z},\ \ j = 1,\,2,$$

were conducted and their results were presented in Table 6.5. Statistical conclusions from theses tests are the same as those from covaiance inversion tests, though the covariance eigenvalue homogeneity test produced a smaller P-value. From Tables 6.4 and 6.5, we notice that the covariance inversion test with one degree of freedom is equivalent to the corresponding covariance eigenvalue test.

Table 6.1: Test of monthly trend in DO based on Spearman's $\hat{\rho}$

Month	$\hat{\rho}$	P-value	Number of Obs.	Standardized Trend Statistic, z	Trend Statistic, c
1	0.55	0.0546	13	1.92	1313.0
2	-0.38	0.1740	14	-1.36	-1197.0
3	-0.15	0.5831	15	-0.55	-615.0
4	0.25	0.3354	16	0.96	1344.0
5	-0.06	0.8151	16	-0.23	-328.0
6	0.57	0.0269	16	2.21	3072.0
7	0.54	0.0512	14	1.95	1715.0
8	0.07	0.8186	13	0.23	156.0
9	0.20	0.5271	11	0.63	242.0
10	0.70	0.0111	14	2.54	2240.0
11	0.08	0.7823	13	0.28	188.5
12	0.58	0.0525	12	1.94	996.0

Table 6.2: Decomposition of χ^2 statistic for any monthly trend in DO

What to be Assessed	χ^2	d.f.	P-value
Any Monthly Trend in DO	17.7476	12	0.12357
Overall Trend in DO (Test Based on the Sum of Trend Statistics)	5.6047	1	0.01791
Remainder	12.1429	11	0.35300

Table 6.3: Covariance eigenvalue tests for any monthly trend in DO

Statistics Used	P-value
Trend Statistics	0.03941
Standardized Trend Statistics	0.01692
$\hat{\rho}$'s	0.01808

Table 6.4: Covariance inversion tests for overall trend and trend homogeneity in DO

Null Hypothesis	[Contrast] Matrix	χ^2	d.f.	P-value
No overall trend	$\mathbf{C}_1$	7.30016	1	0.00689
Homogeneity in trend	$\mathbf{C}_2$	13.41789	11	0.26689

Table 6.5: Covariance eigenvalue tests for overall trend and trned homogeneity in DO

Null Hypothesis	[Contrast] Matrix	Q	K	α	β	d.f.	P-value
No overall trend	$\mathbf{C}_1$	110.6704	0	0.00	15.16	1.00	0.00689
Homogeneity in trend	$\mathbf{C}_2$	35.7616	< 0	3.93	3.46	4.38	0.07144

6.2 Case of multiple variables with multiple seasons

A generalization for assessing homogeneity in trend across variables and seasons can be made using appropriate matrices. For example, let us consider the case where $p = 3$ variables are measured in each of $m = 4$ seasons. A standardized trend statistic vector in this case is, with the first subscript denoting variable and the second subscript denoting season,

$$\mathbf{z} = (z_{11}, z_{12}, z_{13}, z_{14}, z_{21}, z_{22}, z_{23}, z_{24}, z_{31}, z_{32}, z_{33}, z_{34}).$$

Let $\mathbf{S_z}$ be the associated covariance matrix. For the tests for homegeneity in trend discussed below, the test statistic has a general form

$$\chi^2(\mathbf{C}) = (\mathbf{Cz})' (\mathbf{CS_zC'})^{-1}(\mathbf{Cz}), \tag{6.9}$$

where $\mathbf{C}$ is an appropriate contrast matrix. The distribution of statistic (6.9) is asymptotically χ^2 with degrees of freedom being the rank of $\mathbf{CS_zC'}$. Contrast matrices appropriate for assessing several types of homogeneity in trend across seasons and variables are listed below.

- For assessing homogeneity in trend between seasons:

$$\mathbf{C}_S = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 \end{bmatrix}. \tag{6.10}$$

- For assessing homogeneity in trend between variables:

$$\mathbf{C}_V = \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \end{bmatrix}. \tag{6.11}$$

- For assessing interaction between seasons and variables:

$$\mathbf{C}_{\mathbf{S*V}} = \begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}. \qquad (6.12)$$

The covariance eigenvalue method can also be used in testing these contrasts. In this case, the form of the test statistic is:

$$Q(\mathbf{C}) = (\mathbf{Cz})'(\mathbf{Cz}) = \mathbf{z}'\mathbf{C}'\mathbf{Cz} \qquad (6.13)$$

where $\mathbf{C}$ is an appropriate contrast matrix.

6.2.1 Example

This example is for the case of 3 variables—DO, NFR, and FC—measured in each of 4 seasons. A univariate trend analysis based on the Spearman statistic is summarized in Table 6.6.

ANOVA-like analysis based on the contrast matrices (6.10)–(6.12) using the covariance inversion method is presented in Table 6.7. In Table (6.7), only variables are a significant source of heterogeneity in trend; the differences in trend statistics between variables are apparent in Table 6.6.

The contrast in Table 6.7 were tested using the covariance eigenvalue method and the results were given in Table 6.8. Test results were similar to those from the covariance inversion tests and statistical conclusions are the same.

Table 6.6: Univariate trend analysis: 3 variables and 4 seasons

Variable	Season	$\hat{\rho}$	z-value	P-value
Do	Jan.	0.5529	1.9155	0.0554
	Apr.	0.2487	0.9632	0.3354
	Jul.	0.5408	1.9500	0.0511
	Oct.	0.7040	2.5385	0.0111
NFR	Jan.	-0.2446	-0.8473	0.3968
	Apr.	-0.3577	-1.2897	0.1971
	Jul.	-0.1153	-0.4160	0.6773
	Oct.	-0.2774	-1.0003	0.3171
FC	Jan.	-0.2290	-0.7596	0.4474
	Apr.	-0.1532	-0.5735	0.5662
	Jul.	-0.0027	-0.0096	0.9923
	Oct.	-0.4585	-1.7155	0.0862

Table 6.7: Covariance-inversion-based ANOVA-like analysis (3 variables and 4 seasons)

Source of Heterogeneity	Contrast Matrix	χ^2	d.f.	P-value
Seasons	$\mathbf{C_s}$	1.1215	3	0.77188
Variables	$\mathbf{C_v}$	14.9284	2	0.00057
Season-Variable Interaction	$\mathbf{C_{s*v}}$	1.56768	6	0.95487

Table 6.8: Covariance-eigenvalue-based ANOVA-like analysis (3 variables and 4 seasons)

Source of Heterogeneity	Contrast Matrix	Q	K	α	β	d.f.	P-value
Seasons	$\mathbf{C_s}$	3.175	< 0	2.81	9.69	1.32	0.91907
Variables	$\mathbf{C_v}$	227.962	< 0	2.18	15.02	1.16	0.00015
Season-Variable Interaction	$\mathbf{C_{s*v}}$	5.321	< 0	4.66	7.88	2.40	0.98020

6.3 Case of multiple stations with multiple variables and seasons

An extension to the case having several stations is also possible. It will be reasonable to assume that trends in different stations are uncorrelated. So, a block diagonal matrix may be used as a covariance matrix in this case. For example, if there are two stations, the associated covariance matrix may be

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{bmatrix}. \tag{6.14}$$

where $\mathbf{S}_j$ is the covariance matrix of the trend statistic vector at station j, $j = 1, 2$, and $\mathbf{0}$ is a matrix of zeros.

Let us consider, for example, the case where two variables were measured in each of four seasons at two stations. A standardized trend statistic vector in this case is, with the first subscript denoting station, the second subscript denoting variable, and the third subscript denoting season,

$$\begin{aligned} \mathbf{z} \;=\; & (z_{111},\ z_{112},\ z_{113},\ z_{114},\ z_{121},\ z_{122},\ z_{123},\ z_{124}, \\ & z_{211},\ z_{212},\ z_{213},\ z_{214},\ z_{221},\ z_{222},\ z_{223},\ z_{224}) \end{aligned}$$

Let $\mathbf{S_z}$ be the associated covariance matrix described in (6.14). For the tests for homegeneity in trend discussed below, the test statistic has a general form

$$\chi^2(\mathbf{C}) = (\mathbf{Cz})'\,(\mathbf{CS_zC}')^{-1}(\mathbf{Cz}), \tag{6.15}$$

where $\mathbf{C}$ is an appropriate contrast matrix. The distribution of statistic (6.9) is asymptotically χ^2 with degrees of freedom being the rank of $\mathbf{CS_zC}'$. Contrast matrices appropriate for assessing several types of homogeneity in trend across stations, variables, ans seasons are listed below.

- For assessing homogeneity in trend between stations:

$$\mathbf{C}_{\mathrm{St}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix}. \tag{6.16}$$

- For assessing homogeneity in trend between variables:

$$\mathbf{C}_{\mathrm{V}} = \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{bmatrix}. \tag{6.17}$$

- For assessing homogeneity in trend between seasons:

$$\mathbf{C}_{\mathrm{S}} = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 \end{bmatrix}. \tag{6.18}$$

- For assessing interaction between stations and variables:

$$\mathbf{C}_{\mathrm{St*V}} = \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{6.19}$$

- For assessing interaction between stations and seasons:

$$\mathbf{C}_{\mathrm{St*S}} = \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \end{bmatrix}. \tag{6.20}$$

- For assessing interaction between variables and seasons:

$$\mathbf{C}_{\mathrm{V*S}} = \begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \end{bmatrix}. \tag{6.21}$$

- For assessing three-way interaction between stations, variables and seasons:

$$\mathbf{C}_{\text{St*V*S}} = \begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 & -1 \end{bmatrix}. \tag{6.22}$$

The covariance eigenvalue method can also be used in testing these contrasts. In this case, the form of the test statistic is:

$$Q(\mathbf{C}) = (\mathbf{Cz})'(\mathbf{Cz}) = \mathbf{z}'\mathbf{C}'\mathbf{Cz} \tag{6.23}$$

where $\mathbf{C}$ is an appropriate contrast matrix.

6.3.1 Example

This example is for the case of 2 variables—DO, NFR, and FC—measured in each of 4 seasons at 2 stations. A univariate trend analysis based on the Spearman statistic is summarized in Table 6.9 on page 61.

ANOVA-like analysis based on the contrast matrices (6.10)–(6.12) using the covariance inversion method is presented in Table 6.10 on page 61. In Table 6.10, the station-variable interaction and the station main effect are significant. From Table 6.9, we can see that the differences in trend between variables are not the same across the stations. Also, the overall shape of trends at station 1 seems to be different from that at station 2 (Table 6.9).

The contrast in Table 6.10 were tested using the covariance eigenvalue method and the results were given in Table 6.11. Test results were similar to those from the covariance inversion tests and statistical conclusions are the same. Here, we again see that the covariance inversion test with one degrees of freedom is equivalent to the corresponding covariance eigenvalue test.

Table 6.9: Univariate trend analysis: 2 stations, 2 variables and 4 seasons

Station	Variable	season	$\hat{\rho}$	z-value	P-value
1	DO	Jan.	0.553	1.916	0.05543
		Apr.	0.249	0.963	0.33542
		Jul.	0.541	1.950	0.05117
		Oct.	0.704	2.539	0.01113
	FC	Jan.	−0.229	−0.760	0.44747
		Apr.	−0.153	−0.574	0.56627
		Jul.	−0.003	−0.010	0.99233
		Oct.	−0.459	−1.716	0.08624
2	DO	Jan.	0.144	0.476	0.63387
		Apr.	−0.077	−0.319	0.74963
		Jul.	−0.548	−2.193	0.02831
		Oct.	−0.595	−1.975	0.04828
	FC	Jan.	0.106	0.335	0.73747
		Apr.	−0.448	−1.737	0.08246
		Jul.	−0.027	−0.095	0.92418
		Oct.	−0.082	−0.260	0.79493

Table 6.10: Covariance-inversion-based ANOVA-like analysis (2 stations, 2 variables and 4 seasons)

Source of Heterogeneity	Contrast Matrix	χ^2	d.f.	P-value
Stations	$\mathbf{C}_{\mathrm{St}}$	6.6721	1	0.00979
Variables	$\mathbf{C}_{\mathrm{V}}$	3.2102	1	0.07318
Seasons	$\mathbf{C}_{\mathrm{S}}$	2.2019	3	0.53157
Station-Variable Interaction	$\mathbf{C}_{\mathrm{St*V}}$	7.7299	1	0.00543
Station-Season Interaction	$\mathbf{C}_{\mathrm{St*S}}$	2.1228	3	0.54732
Variable-Season Interaction	$\mathbf{C}_{\mathrm{V*S}}$	1.6597	3	0.64592
Station-Variable-Season Interaction	$\mathbf{C}_{\mathrm{St*V*S}}$	3.4736	3	0.32421

Table 6.11: Covariance-eigenvalue-based ANOVA-like analysis (2 stations, 2 variables and 4 seasons)

Source of Heterogeneity	Contrast Matrix	Q	K	α	β	d.f.	P-value
Stations	$\mathbf{C}_{\mathrm{St}}$	101.528	0	0.00	15.22	1.00	0.00979
Variables	$\mathbf{C}_{\mathrm{V}}$	66.773	0	0.00	20.80	1.00	0.07318
Seasons	$\mathbf{C}_{\mathrm{S}}$	29.980	< 0	3.91	14.83	1.21	0.23237
Station-Variable Interaction	$\mathbf{C}_{\mathrm{St*V}}$	160.783	0	0.00	20.80	1.00	0.00543
Station-Season Interaction	$\mathbf{C}_{\mathrm{St*S}}$	26.864	< 0	3.91	14.83	1.21	0.26596
Variable-Season Interaction	$\mathbf{C}_{\mathrm{V*S}}$	8.824	< 0	4.40	14.70	1.41	0.72680
Station-Variable-Season Interaction	$\mathbf{C}_{\mathrm{St*V*S}}$	19.951	< 0	4.40	14.70	1.41	0.42806

7 Summary and Future Research

This major emphasis of this dissertation is the construction of the general framework of multivariate nonparametric trend analysis. First, a generalization was made by obtaining the covariance between two generalized trend statistics. Further generalization was achieved by suggesting a generalized multivariate trend statistic with an approximated null distribution. Canonical anaysis and ANOVA-like analysis make trend analysis more complete; in some situations, they are dual analyses.

For future research, the choice of the coefficient matrix in a generalized multivariate trend statistic needs to be investigated. Both theoretical and simulation research will be needed to shed light on this problem.

8 Appendix 1: Proofs

Appendix 1.1: Proof of Theorem 3.1.

Under H_0, all $n!$ orderings of the rows of $[\mathbf{y}_u \ \mathbf{y}_v]$ in (3.12) are equally likely, which implies each of $n!$ orderings of (y_{iu}, y_{iv}) is an equally-likely possibility. So, letting E denote the mean value on summation over all $n!$ possible permutations, the null covariance of c_u and c_v is:

$$\mathrm{Cov}\,(c_u\,,\,c_v) = \mathrm{E}\,(c_u\,c_v) \tag{A.1}$$

since $\mathrm{E}\,(c_u) = 0$ and $\mathrm{E}\,(c_v) = 0$ under H_0.

In order to find $\mathrm{E}\,(c_u\,c_v)$, we need to consider the summation of $c_u\,c_v$'s over all n! permutations of rows of the following $n \times 2$ matrix whose rows are the observed (y_{iu}, y_{iv})'s:

$$\begin{bmatrix} y_{1u} & y_{1v} \\ y_{2u} & y_{2v} \\ \vdots & \vdots \\ y_{nu} & y_{nv} \end{bmatrix}. \tag{A.2}$$

Fortunately, this situation is actually the same as the situation Daniels [1944] considered for the purpose of finding the covariance between the two different correlation coefficients obtained from the same data set using the two different score functions. He obtained $E(c_1 c_2)$ where c_1 and c_2 are:

$$c_1 = \sum_{i,j} a_{ij}^{(1)} b_{ij}^{(1)} \quad \text{and} \quad c_2 = \sum_{i,j} a_{ij}^{(2)} b_{ij}^{(2)} , \tag{A.3}$$

the superscripts (1) and (2) denoting the two different score functions applied to the same x's and y's. His c_1 and c_2 are corresponding to our c_u and c_v where

$$c_u = \sum_{i<j} a_{ij} b_{ij}^{(u)} \quad \text{and} \quad c_v = \sum_{i<j} a_{ij} b_{ij}^{(v)} . \tag{A.4}$$

The result (3.15) is obtained using Daniels' (1944, pp. 130-131) result.

Appendix 1.2: Proof of Corollary 3.1.1.

This is a special case of Theorem 3.1, where the score function $f(\cdot\,,\,\cdot)$ is

$$f(Z_i\,,\,Z_j) \;=\; \operatorname{sign}(Z_j - Z_i)\,. \tag{A.5}$$

With this score function, we obtain

$$\sum_{i,j,k} a_{ij} a_{ik} = \tfrac{1}{3}\, n\,(n^2 - 1)\ , \tag{A.6}$$

$$\sum_{i,j} a_{ij}{}^2 = n\,(n-1) \tag{A.7}$$

(Kendall and Gibbons 1990, page 93),

$$\sum_{i,j,k} b_{ij}{}^{(u)}\, b_{ik}{}^{(v)} = \sum_{i,j,k} \operatorname{sign}[(y_{ju} - y_{iu})(y_{kv} - y_{iv})]$$

$$= \;4\; S_{uv} \tag{A.8}$$

(Lehmann 1975, p. 370), and

$$\sum_{i,j} b_{ij}{}^{(u)}\, b_{ij}{}^{(v)} = \sum_{i,j} \operatorname{sign}[(y_{ju} - y_{iu})(y_{jv} - y_{iv})]$$

$$= 2\;\; k_{uv}\ . \tag{A.9}$$

Substituting in (3.15), we thus get (3.16).

Appendix 1.3: Proof of Corollary 3.1.2.

This is also a special case of Theorem 3.1, where the score function $f(\cdot\,,\,\cdot)$ is

$$f(Z_i\,,\,Z_j) \;=\; \text{rank}(Z_j) \;-\; \text{rank}(Z_i)\;. \tag{A.10}$$

With this score function, we obtain

$$\sum_{i,j,k} a_{ij}a_{ik} = \frac{1}{12}\,n^3\,(n^2-1)\;, \tag{A.11}$$

$$\sum_{i,j} a_{ij}{}^2 = \;\frac{1}{6}\,n^2\,(n^2-1) \tag{A.12}$$

(Kendall and Gibbons 1990, page 98). With R denoting the rank, we get

$$\sum_{i,j,k} b_{ij}{}^{(u)}\,b_{ik}{}^{(v)} = \sum_{i,j,k}\;[(R_{ju}-R_{iu})(R_{kv}-R_{iv})]$$

$$= \sum_{i,j,k} R_{iu}R_{iv} \;-\; \frac{1}{4}\,n^3\,(n+1)^2\;, \tag{A.13}$$

since $\sum_{i,j,k} R_{ju}R_{kv} \;=\; \sum_{i,j,k} R_{ju}R_{iv} \;=\; \sum_{i,j,k} R_{iu}R_{kv}$

$$=\; n\,\big[(1+\cdots+n)\,(1+\cdots+n)\big]\;,$$

and

$$\sum_{i,j} b_{ij}{}^{(u)}\,b_{ij}{}^{(v)} = \sum_{i,j}\;(R_{ju}-R_{iu})(R_{jv}-R_{iv})]$$

$$=\;2n\,\sum_{i} R_{iu}R_{iv} \;-\; \frac{1}{2}\,n^2\,(n+1)^2 \tag{A.14}$$

(Kendall and Gibbons 1990, page 26) .

Substituting in (3.15) , we get

$$Cov(c_u\,,\,c_v) = \frac{n\,(n+1)}{12}\left\{n^2 \sum_i R_{iu}\,R_{iv} \;-\; \tfrac{1}{4}\,n^3\,(n+1)^2\right\} \tag{A.15}$$

$$= \frac{n^3\,(n+1)}{12} \;\sum_i [R_{iu} - \tfrac{1}{2}\,(n+1)]\;[R_{iv} - \tfrac{1}{2}\,(n+1)]. \tag{A.16}$$

Since $c_u = \sum_{i<j} (j-i)\,(R_{ju} - R_{iu}) \;=\; n\,\sum_i [i - \tfrac{1}{2}\,(n+1)]\;[R_{iv} - \tfrac{1}{2}\,(n+1)]$

$$= \; n\; s_u \tag{A.17}$$

(Kendall and Gibbons 1990, page 27) ,

$$\mathrm{Cov}(s_u\,,\,s_v) = \frac{n\,(n+1)}{12}\; s_{uv}\,, \tag{A.18}$$

as stated in Corollary 3.1.2.

Appendix 1.4: Proof of Theorem 3.2.

Since the c_u, $g = 1, 2, \ldots, p$, are U-statistics, due to Theorem 7.1 of Hoeffding (1948), $\mathbf{c}$ has asymptotic multivariate normality, from which the result of the theorem follows.

Appendix 1.5: Proof of Theorem 5.1.

The first step of this proof is similar to that of the derivation of canonical correlations and variates in Anderson (1958). Let us use the Lagrange multiplier technique for finding the vector $\mathbf{q}$ that satisfies the following:

$$\frac{(\mathbf{q}'\mathbf{c})^2}{\mathbf{q}'\,\mathbf{S_C}\,\mathbf{q}} = \max_{\mathbf{q_o}}\left[\frac{(\mathbf{q_o}'\mathbf{c})^2}{\mathbf{q_o}'\,\mathbf{S_C}\,\mathbf{q_o}}\right], \tag{A.19}$$

where $\mathbf{q_o}$ denotes any member in the set of $p \times 1$ vectors such that $\mathbf{q_o}'\,\mathbf{S_C}\,\mathbf{q_o} = 1$. Write $T(\mathbf{q_o})$ as the function to be maximized under the constraint that $\mathbf{q_o}'\,\mathbf{S_C}\,\mathbf{q_o} = 1$:

$$\begin{aligned} T(\mathbf{q_o}) &= (\mathbf{q_o}'\mathbf{c})^2 - \lambda\,(\mathbf{q_o}'\,\mathbf{S_C}\,\mathbf{q_o} - 1) \\ &= \mathbf{q_o}'\mathbf{c}\,\mathbf{c}'\mathbf{q_o} - \lambda\,(\mathbf{q_o}'\,\mathbf{S_c}\,\mathbf{q_o} - 1), \end{aligned} \tag{A.20}$$

where λ is a Lagrange multiplier. Taking the partial derivative of T with respect to λ and equating to $\mathbf{0}$, we get

$$\frac{\partial T}{\partial \lambda} = 2\,\mathbf{c}\,\mathbf{c}'\mathbf{q_o} - 2\,\lambda\,\mathbf{S_c}\,\mathbf{q_o} = \mathbf{0}. \tag{A.21}$$

Let $\mathbf{q}$ be the value of $\mathbf{q_o}$ that satisfies (A.21). Then, $\mathbf{q}$ is the sample canonical coefficient vector. From (A.21), we have

$$\mathbf{c}\,\mathbf{c}'\mathbf{q} = \lambda\,\mathbf{S_c}\,\mathbf{q}. \tag{A.22}$$

Premultiplying both sides of (A.22) by $\mathbf{q}'$, we obtain

$$\mathbf{q}'\mathbf{c}\,\mathbf{c}'\mathbf{q} \;=\; \lambda\,\mathbf{q}'\mathbf{S_c}\,\mathbf{q}\,,$$

that is,

$$(\mathbf{q}'\mathbf{c})^2 \;=\; \lambda \tag{A.23}$$

since $\mathbf{q}'\mathbf{S_c}\,\mathbf{q} = 1$ by the constraint. (A.23) says that λ is equal to the squared value of the sample canonical variate.

Now, if $\mathbf{S_c}$ is nonsingular, premultiplying both sides of (A.22) by $\mathbf{S_c}^{-1}$, we obtain

$$\mathbf{S_c}^{-1}\,\mathbf{c}\,\mathbf{c}'\mathbf{q} \;=\; \lambda\,\mathbf{q}\,, \tag{A.24}$$

which implies that $(\lambda\,,\ \mathbf{q})$ is an eigenvalue-eigenvector pair of $\mathbf{S_c}^{-1}\,\mathbf{c}\,\mathbf{c}'$. $\mathbf{S_c}^{-1}\,\mathbf{c}\,\mathbf{c}'$ has only one nonzero positive eigenvalue, λ, since it is of rank 1 and λ is the squared value of the sample canonical variate.

If $\mathbf{S_c}$ is singular, premultiply both sides of (A.22) by $\mathbf{S_c}^{-}$, a generalized inverse of $\mathbf{S_c}$ satisfying

$$\mathbf{S_c}\,\mathbf{S_c}^{-}\,\mathbf{S_c} = \mathbf{S_c}\,. \tag{A.25}$$

Then, we have

$$\mathbf{S_c}^{-}\,\mathbf{c}\,\mathbf{c}'\mathbf{q} \;=\; \lambda\,\mathbf{S_c}^{-}\,\mathbf{S_c}\,\mathbf{q}\,. \tag{A.26}$$

Here, what we have to show is that $\mathbf{S_c}^{-}\,\mathbf{S_c}\,\mathbf{q} = \mathbf{q}$. Postmultiplying both sides of (A.25) by $\mathbf{q}$, we get

$$\mathbf{S_c}\,\mathbf{S_c}^{-}\,\mathbf{S_c}\,\mathbf{q} = \mathbf{S_c}\,\mathbf{q}\,, \tag{A.27}$$

which implies that $\mathbf{S_c}^{-}\,\mathbf{S_c}\,\mathbf{q} = \mathbf{q}$. So, we can write (A.26) as

$$\mathbf{S_C^-}\,\mathbf{c}\,\mathbf{c'q} = \lambda\,\mathbf{q}\,, \tag{A.28}$$

which says that (λ , $\mathbf{q}$) is an eigenvalue-eigenvector pair of $\mathbf{S_C^-}\,\mathbf{c}\,\mathbf{c'}$. Now, let $\mathbf{S_C^-}$ be the inverse of $\mathbf{S_C}$ if it is nonsingular, or a generalized inverse of $\mathbf{S_C}$ satisfying $\mathbf{S_C}\,\mathbf{S_C^-}\,\mathbf{S_C} = \mathbf{S_C}$ if it is singular.

We have seen that λ is the only one nonzero eigenvlaue of $\mathbf{S_C^-}\,\mathbf{c}\,\mathbf{c'}$. Recalling that the sum of eigenvalues is equal to the trace, we write

$$\lambda + 0 + \cdots + 0 = \mathrm{tr}(\mathbf{S_C^-}\,\mathbf{c}\,\mathbf{c'})\,. \tag{A.29}$$

By the property of the trace, we have

$$\lambda = \mathrm{tr}(\mathbf{S_C^-}\,\mathbf{c}\,\mathbf{c'}) = \mathrm{tr}(\mathbf{c'}\mathbf{S_C^-}\,\mathbf{c}) = \mathbf{c'}\mathbf{S_C^-}\,\mathbf{c}\,. \tag{A.30}$$

Then, combining (A.23) and (A.30), we obtain

$$\lambda = (\mathbf{q'c})^2 = \mathbf{c'}\,\mathbf{S_C^-}\,\mathbf{c}\,. \tag{A.31}$$

This completes the proof.

Appendix 1.6: Proof of Theorem 5.2.

Let us again use the Lagrange multiplier technique for finding the vector $\mathbf{q}$ that satisfies the following:

$$\frac{(\mathbf{q}'\mathbf{c})^2}{\mathbf{q}'\,\mathbf{S_C}\,\mathbf{q}} = \max_{\mathbf{q}_*}\left[\frac{(\mathbf{q}_*'\mathbf{c})^2}{\mathbf{q}_*'\,\mathbf{S_C}\,\mathbf{q}_*}\right], \tag{A.32}$$

where $\mathbf{q}_*$ denotes any member in the set of $p\times 1$ vectors such that $\mathbf{q}_*'\,\mathbf{S_C}\,\mathbf{q}_* = 1$ and $\mathbf{H}\,\mathbf{q}_* = \mathbf{0}$. Write $T(\mathbf{q}_*)$ as the function to be maximized under the constraint that $\mathbf{q}_*'\,\mathbf{S_C}\,\mathbf{q}_* = 1$ and $\mathbf{H}\,\mathbf{q}_* = \mathbf{0}$,

$$T(\mathbf{q}_*) = (\mathbf{q}_*'\mathbf{c})^2 - \lambda\,(\mathbf{q}_*'\,\mathbf{S_C}\,\mathbf{q}_* - 1) - 2\boldsymbol{\mu}'\,(\mathbf{H}\,\mathbf{q}_* - \mathbf{0}) \tag{A.33}$$

where λ and $2\,\boldsymbol{\mu}'$ are Lagrange multipliers. Taking the partial derivative of T with respect to λ and equating to $\mathbf{0}$, we get

$$\frac{\partial T}{\partial \lambda} = 2\,\mathbf{c}\,\mathbf{c}'\mathbf{q}_* - 2\,\lambda\,\mathbf{S_C}\,\mathbf{q}_* - 2\,\mathbf{H}'\boldsymbol{\mu} = 0\,. \tag{A.34}$$

Let $\mathbf{q}_o$ and λ_o be the values of $\mathbf{q}_*$ and λ that satisfy (A.34). Then, $\mathbf{q}_o$ is the sample canonical coefficient vector under the restriction that $\mathbf{H}\,\mathbf{q}_* = \mathbf{0}$. From (A.34), we have

$$\mathbf{H}'\boldsymbol{\mu} = \mathbf{c}\,\mathbf{c}'\mathbf{q}_o - \lambda_o\,\mathbf{S_C}\,\mathbf{q}_o\,. \tag{A.35}$$

Premultiplying both sides of (A.35) by $\mathbf{H}$, we obtain

$$\mathbf{H}\,\mathbf{H}'\boldsymbol{\mu} = \mathbf{H}\,(\,\mathbf{c}\,\mathbf{c}'\mathbf{q}_o - \lambda_o\,\mathbf{S_c}\,\mathbf{q}_o\,)\,. \tag{A.36}$$

Then,

$$\boldsymbol{\mu} = (\mathbf{H}\,\mathbf{H}')^{-1}\mathbf{H}\,(\,\mathbf{c}\,\mathbf{c}'\mathbf{q}_o - \lambda_o\,\mathbf{S_c}\,\mathbf{q}_o\,)\,. \tag{A.37}$$

From (A.35) and (A.37),

$$\mathbf{H}'(\mathbf{H}\,\mathbf{H}')^{-1}\mathbf{H}\,(\,\mathbf{c}\,\mathbf{c}'\mathbf{q}_o - \lambda_o\,\mathbf{S_c}\,\mathbf{q}_o\,) = \mathbf{c}\,\mathbf{c}'\mathbf{q}_o - \lambda_o\,\mathbf{S_c}\,\mathbf{q}_o\,. \tag{A.38}$$

Rewriting (A.38),

$$[\mathbf{I} - \mathbf{H}'(\mathbf{H}\,\mathbf{H}')^{-1}\mathbf{H}]\,(\,\mathbf{c}\,\mathbf{c}' - \lambda_o\,\mathbf{S_c}\,)\,\mathbf{q}_o = \mathbf{0} \tag{A.39}$$

Since $\mathbf{H}\,\mathbf{q}_o = \mathbf{0}$ from the restriction, we can write $\mathbf{q}_o$ as

$$\mathbf{q}_o = [\mathbf{I} - \mathbf{H}'(\mathbf{H}\,\mathbf{H}')^{-1}\mathbf{H}]\,\mathbf{q}_o \tag{A.40}$$

From (A.39) and (A.40), letting $\mathbf{G} = \mathbf{I} - \mathbf{H}'(\mathbf{H}\,\mathbf{H}')^{-1}\mathbf{H}$, that is symmetric idempotent, we get

$$\mathbf{G}\,(\,\mathbf{c}\,\mathbf{c}' - \lambda_o\,\mathbf{S_c}\,)\,\mathbf{G}\,\mathbf{q}_o = \mathbf{0} \tag{A.41}$$

That is,

$$(\mathbf{G}\,\mathbf{c}\,\mathbf{c}'\,\mathbf{G})\,\mathbf{q}_o = \lambda_o\,\mathbf{G}\,\mathbf{S_c}\,\mathbf{G}\,\mathbf{q}_o\,. \tag{A.42}$$

Multiplying $(\mathbf{G}\,\mathbf{S_c}\,\mathbf{G})^-$, a generalized inverse of $\mathbf{G}\,\mathbf{S_c}\,\mathbf{G}$, we have

$$(\mathbf{G}\,\mathbf{S_C}\,\mathbf{G})^{-}(\mathbf{G}\,\mathbf{c}\,\mathbf{c}'\,\mathbf{G})\,\mathbf{q_o} \;=\; \lambda_o\;\mathbf{q_o}\,, \tag{A.43}$$

which says that λ and $\mathbf{q_o}$ are an eigenvalue and an eigenvector of

$$(\mathbf{G}\,\mathbf{S_C}\,\mathbf{G})^{-}(\mathbf{G}\,\mathbf{c}\,\mathbf{c}'\,\mathbf{G}).$$

Now, multiplying $\mathbf{q_o}'$ by both sides of (A.42), we obtain

$$\mathbf{q_o}'(\mathbf{G}\,\mathbf{c}\,\mathbf{c}'\,\mathbf{G})\,\mathbf{q_o} \;=\; \lambda_o\,\mathbf{q_o}'\mathbf{G}\,\mathbf{S_C}\,\mathbf{G}\,\mathbf{q_o}\,. \tag{A.42}$$

Since $\mathbf{G}\,\mathbf{q_o} \;=\; \mathbf{q_o}$ in (A.40) and we set $\mathbf{q_o}'\,\mathbf{S_C}\,\mathbf{q_o} = 1$, we have

$$(\mathbf{q_o}'\,\mathbf{c})^2 = \lambda_o\,. \tag{A.43}$$

Since $(\mathbf{G}\,\mathbf{S_C}\,\mathbf{G})^{-}(\mathbf{G}\,\mathbf{c}\,\mathbf{c}'\,\mathbf{G})$ is of rank 1, it has only one nonzero eigenvalue λ_o. Also, recalling that the sum of eigenvalue is equal to the trace,

$$\begin{aligned}
\lambda_o \;=\; \lambda_o + 0 + \cdots + 0 \;&=\; \mathrm{tr}[(\mathbf{G}\,\mathbf{S_C}\,\mathbf{G})^{-}(\mathbf{G}\,\mathbf{c})\,(\mathbf{G}\,\mathbf{c})'] \\
&=\; \mathrm{tr}[(\mathbf{G}\,\mathbf{c})'(\mathbf{G}\,\mathbf{S_C}\,\mathbf{G})^{-}(\mathbf{G}\,\mathbf{c})] \\
&=\; (\mathbf{G}\,\mathbf{c})'(\mathbf{G}\,\mathbf{S_C}\,\mathbf{G})^{-}(\mathbf{G}\,\mathbf{c})\,.
\end{aligned}$$

Now, the proof is complete.

9 Appendix 2: A SAS program to calculate Spearman trend statistics and associated covariances based on the "generalized trend statistic" approach

```
**** Data Input;

OPTIONS NODATE LS=79;
DATA ST16;  ** This data set has been through data screening.;
     INPUT YY MM DO PH NFR FR FC;
     YYMM=YY*100+MM;
LINES;
    73     1    11.790    7.000     2.50    206     500
    73     2    13.000    7.299     7.00     80      50
    73     3    10.390    7.399      .       68    3000
    73     4    10.000    8.500     5.00    132     100
    73     5    10.590    7.500     9.00     71     200
    73     6     8.399    8.000     7.00    133      50
    73     7     8.599    7.500     7.00     56      50
    73     8     9.399    7.599    11.00     71    1200
    73     9    10.590    8.500     2.50    135     100
    73    10     9.599    8.000     5.00    157     600
    73    11    12.000    8.699     6.00    147     100
    73    12    10.790    7.199     5.00     71     200
    74     1    10.590    7.000     2.50     92      50
    74     2    11.390    7.599     2.50     98     100
    74     3    13.000    7.299      .        .      50
    74     4    10.000    7.299     6.00     83     200
    74     5    10.790    7.899     6.00     59     100
    74     6     9.599    8.500     2.50    132     100
    74     7    10.590    8.500     2.50    167     100
    74     8    10.390    7.599      .        .     700
    74     9     9.799    7.500     7.00     32     300
    74    10    10.190    7.699     6.00     44     200
    74    11    11.190    8.799     2.50    116      50
    74    12      .         .        .        .       .
    75     1    12.000    7.500     6.00     77     100
    75     2    12.390    7.500    10.00     54     400
    75     3    13.000    7.500     2.50     67      50
    75     4    13.190    7.000     5.00     50      50
    75     5    10.000    7.500    10.00     92     600
    75     6     8.399    8.799     2.50    124    6000
    75     7    10.000    9.000     2.50    145    6000
    75     8    11.000    8.899     5.00    184    2200
    75     9     3.000    8.599     2.50    195    3900
    75    10     9.399    7.500     2.50     82    1300
    75    11    10.000    9.000    11.00    272     900
    75    12    11.390    7.799     2.50     63     100
    76     1     9.799    6.699     2.50     68     200
    76     5    10.190    7.500     2.50     61     100
```

76	6	8.399	8.699	2.50	89	1800
76	7	9.000	9.000	2.50	113	700
76	8	8.999	8.699	5.25	186	2400
76	10	10.590	8.000	2.50	110	300
76	11	12.390	8.000	2.50	112	100
76	12	12.390	7.500	2.50	64	100
77	2	12.190	7.399	8.00	77	200
77	3	10.390	7.000	2.50	70	200
77	4	9.399	8.199	2.50	99	500
77	5	10.000	7.299	2.50	81	500
77	6	7.399	7.699	5.00	139	6000
77	7	8.399	8.299	2.50	196	700
77	8	9.399	8.699	2.50	133	1500
77	10	9.799	9.000	2.50	114	100
77	11	10.590	8.299	7.00	102	200
77	12	.	7.599	2.50	87	200
78	1	11.000	8.000	2.50	128	500
78	3	12.790	7.399	2.50	56	300
78	4	7.599	8.000	.	37	200
78	5	8.399	7.799	.	131	3600
78	6	.	.	.	.	.
78	7	.	.	.	.	.
78	8	.	.	.	.	.
78	9	.	.	.	.	.
78	10	.	.	.	.	.
78	11	.	.	.	.	.
78	12	.	.	.	.	.
79	1	.	.	.	.	.
79	2	13.000	8.000	8.0	104	100
79	3	12.200	7.700	2.5	49	200
79	4	12.000	8.700	9.0	77	50
79	5	10.200	8.700	2.5	91	1200
79	6	9.000	8.700	.	.	100
79	7	9.800	7.700	2.5	7	200
79	8	10.000	8.700	15.0	89	4200
79	9	11.000	7.700	2.5	104	.
79	10	10.200	8.700	7.0	84	800
79	12	11.400	7.700	2.5	89	.
80	1	13.200	7.000	2.5	59	700
80	2	14.200	7.700	2.5	96	300
80	3	.	7.700	2.5	98	100
80	4	10.000	7.700	15.0	64	.
80	5	10.600	7.700	.	155	6000
80	6	9.600	8.200	2.5	131	400
80	7	.	.	.	.	.
80	10	9.000	9.200	2.5	247	200
80	11	11.400	7.500	2.5	171	1700
80	12	12.200	8.300	2.5	116	300
81	1	.	.	.	.	.
81	2	12.200	7.700	18.0	41	50
81	3	13.200	7.700	2.5	30	50
81	4	8.600	7.600	6.0	59	100

81	5	9.400	7.800	2.5	89	400
81	6	8.000	.	10.0	107	5100
81	8	8.000	8.200	10.0	146	1400
81	10	10.400	8.470	5.0	160	400
81	11	12.600	8.500	2.5	126	200
81	12	12.200	7.500	2.5	71	100
82	1	12.400	7.200	5.0	62	100
82	2	11.800	7.400	10.0	61	100
82	3	11.400	7.700	.	.	50
82	4	12.000	7.700	2.5	28	100
82	5	10.400	7.800	2.5	80	100
82	6	9.000	7.600	2.5	60	6000
82	7	8.000	7.700	2.5	73	1100
82	8	10.000	8.000	2.5	121	1000
82	9	8.600	7.500	17.0	51	800
82	10	10.400	8.000	2.5	112	700
82	11	9.600	7.700	2.5	42	600
82	12	12.200	7.500	2.5	58	50
83	1	13.400	7.000	2.5	35	300
83	2	12.400	7.500	2.5	52	1400
83	3	11.000	7.700	2.5	160	200
83	4	12.000	7.500	10.0	83	100
83	5	10.4	7.70	2.5	27	100
83	6	8.0	7.90	2.5	123	200
83	7	9.8	8.00	2.5	147	100
83	8	8.4	8.70	2.5	140	200
83	9	13.0	8.80	2.5	222	200
83	10	10.0	8.70	2.5	192	200
83	11	10.8	8.09	8.0	119	500
83	12	11.6	7.64	14.0	56	300
84	2	11.8	7.51	2.5	87	50
84	3	11.2	7.85	2.5	63	.
84	4	10.4	7.60	12.0	17	300
84	5	11.2	.	2.5	85	200
84	6	9.0	8.34	.	162	50
84	7	9.8	8.49	2.5	101	1000
84	8	9.0	8.60	2.5	106	300
84	9	8.7	8.75	2.5	155	500
84	10	12.0	8.54	2.5	206	100
84	11	11.9	8.67	2.5	76	300
85	1	11.8	7.76	2.5	79	100
85	2	10.4	7.11	2.5	42	100
85	3	10.2	7.61	2.5	81	300
85	4	9.4	8.06	2.5	69	100
85	5	9.8	7.49	2.5	99	100
85	6	10.0	7.80	5.0	44	200
85	7	11.4	8.49	2.5	132	300
85	8	11.0	8.01	2.5	104	50
85	9	10.4	7.78	2.5	129	50
85	10	10.8	8.05	6.0	172	50
85	11	11.4	7.56	2.5	128	500
85	12	10.0	7.34	2.5	66	200

```
     86      1     12.5     8.22      2.5        45      300
     86      2     10.9     6.86     15.0        44       50
     86      3     10.8     7.18      2.5        62      500
     86      4     10.4     7.71      2.5        96       50
     86      5      9.2     7.04     13.0        77       50
     86      6      9.6     7.62      2.5        90      100
     86      7     11.6     8.29     16.0       139      300
     86      8     10.2     7.99      2.5       164      300
     86      9      9.0     7.77      2.5       177      100
     86     10     10.6     7.54      2.5        79      500
     86     11     10.5     7.31      2.5        61       50
     86     12     12.8     6.61     10.0        57      100
     87      1     12.9     7.71      2.5        80       50
     87      2     12.2     7.35      2.5        90        .
     87      3     12.1     8.09      2.5        72      200
     87      4     10.6     7.98      2.5        77      100
     87      5      9.9     8.21      2.5        79      200
     87      6      9.0     8.09      2.5       107     1000
     87      7     12.5     9.03      2.5        89      100
     87      8     10.1     8.32      2.5       130       50
     87      9     10.5     8.41      2.5       101      300
     87     10     11.9     8.36      2.5       171       50
     87     11     13.4     8.46      2.5       168       50
     87     12     14.2     8.70      2.5       150       50
     88      1     12.1     7.64      2.5        48      100
     88      2     12.2     8.28      2.5        75      100
     88      3     11.6     8.45      2.5        68        .
     88      4     11.0     7.98       .          .       50
     88      6      9.9     8.59      2.5       152        .
     88      7     11.5     8.86      2.5       148        .
     89      1     12.1     8.28      2.5        56        .
     89      2       .        .        .          .        .
     89      3     12.4     8.12      2.5        78      200
     89      4     11.1     7.91      2.5        69      200
     89      5     11.3     8.44      2.5        81        .
     89      6     11.2     8.25      2.5        70      100
     89      7      9.9     7.57      2.5       109      200
     89      8       .        .        .          .      300
     89      9     12.3     8.63      2.5       115      100
     89     10       .        .        .          .      200
     89     11       .        .        .          .      100
     89     12     13.8     8.68       .          .       50
;

**** Converting the Univariate Form of Data to the Multivariate form;

%MACRO TRANS(CNSTTNT,C1,C2,C3,C4,C5,C6,C7,C8,C9,C10,C11,C12,TRANSD);

DATA A; SET ST16; T=YY; KEEP T MM &CNSTTNT;
PROC SORT; BY T MM;
PROC UNIVARIATE NOPRINT; VAR &CNSTTNT; BY T MM;
     OUTPUT OUT=B MEDIAN=&CNSTTNT;
```

```
PROC TRANSPOSE
  OUT=C(RENAME=( _1=&C1   _2=&C2   _3=&C3
                 _4=&C4   _5=&C5   _6=&C6
                 _7=&C7   _8=&C8   _9=&C9
                _10=&C10 _11=&C11 _12=&C12));
  BY T; ID MM;
DATA &TRANSD; SET C; DROP _NAME_;

%MEND TRANS;

%TRANS(CNSTTNT=DO,
       C1=DO1, C2=DO2, C3=DO3, C4 =DO4,  C5 =DO5,  C6 =DO6,
       C7=DO7, C8=DO8, C9=DO9, C10=DO10, C11=DO11, C12=DO12,
       TRANSD=DO_TR );
%TRANS(CNSTTNT=PH,
       C1=PH1, C2=PH2, C3=PH3, C4 =PH4,  C5 =PH5,  C6 =PH6,
       C7=PH7, C8=PH8, C9=PH9, C10=PH10, C11=PH11, C12=PH12,
       TRANSD=PH_TR );
%TRANS(CNSTTNT=NFR,
       C1=NFR1, C2=NFR2, C3=NFR3, C4 =NFR4,  C5 =NFR5,  C6 =NFR6,
       C7=NFR7, C8=NFR8, C9=NFR9, C10=NFR10, C11=NFR11, C12=NFR12,
       TRANSD=NFR_TR);
%TRANS(CNSTTNT=FR,
       C1=FR1, C2=FR2, C3=FR3, C4 =FR4,  C5 =FR5,  C6 =FR6,
       C7=FR7, C8=FR8, C9=FR9, C10=FR10, C11=FR11, C12=FR12,
       TRANSD=FR_TR );
%TRANS(CNSTTNT=FC,
       C1=FC1, C2=FC2, C3=FC3, C4 =FC4,  C5 =FC5,  C6 =FC6,
       C7=FC7, C8=FC8, C9=FC9, C10=FC10, C11=FC11, C12=FC12,
       TRANSD=FC_TR );

DATA MULTDATA;
     MERGE DO_TR PH_TR NFR_TR FR_TR FC_TR;
     BY T;

**** Getting mp Trend Statistics & mp by mp Covariance Matrix;
   * m=12 (# of Seasons);
   * p= 5 (# of Constituents);

DATA T;     SET MULTDATA;  KEEP T;
DATA YS;    SET MULTDATA;  DROP T;

PROC IML;
          USE T;     READ ALL INTO T;
          USE YS;    READ ALL INTO YS
VAR { DO1  DO2  DO3  DO4  DO5  DO6  DO7  DO8  DO9  DO10  DO11  DO12
      PH1  PH2  PH3  PH4  PH5  PH6  PH7  PH8  PH9  PH10  PH11  PH12
     NFR1 NFR2 NFR3 NFR4 NFR5 NFR6 NFR7 NFR8 NFR9 NFR10 NFR11 NFR12
      FR1  FR2  FR3  FR4  FR5  FR6  FR7  FR8  FR9  FR10  FR11  FR12
      FC1  FC2  FC3  FC4  FC5  FC6  FC7  FC8  FC9  FC10  FC11  FC12};

   PM=NCOL(YS);  CREATE PM FROM PM; APPEND FROM PM;
```

```
*** Getting mp Trend Statistics & Associated Variances;

START KVAR;
      DO Q=1 TO PM;
          Y=YS(|,Q|);

          X=T#(Y ^= .);
       IF SUM(X)=0 THEN DO; N=0; END;
       ELSE DO;
          X=X(|LOC(X ^= 0),|);
          Y=Y(|LOC(Y ^= .),|);
          N=NROW(Y); END;

       IF N<=2 THEN DO;
                    DEN_K_=0;
                    CORR_=0; P_VALUE_=1;
                    K_=0; VARK_=0; VG_=0;
                    END;
       ELSE DO;

********************;
X=RANKTIE(X);
Y=RANKTIE(Y);
********************;

             DO I=1 TO N-1;
               DO J=I+1 TO N;
                  XD=X(|J,|)-X(|I,|);
                  YD=Y(|J,|)-Y(|I,|);
                  IF I=1 & J=2 THEN DO;
                     II=I; JJ=J; XDXD=XD; YDYD=YD;
                      END;
                  ELSE DO;
                     II=II//I; JJ=JJ//J; XDXD=XDXD//XD; YDYD=YDYD//YD;
                      END;
               END;
             END;

 /*
  ** Kendall score function;
  ******************;
     XDXD=SIGN(XDXD);
     YDYD=SIGN(YDYD);
  ******************;
 */

     A=J(N,N,0);
     B=J(N,N,0);
     NCOMP=N#(N-1)/2;
```

```
               DO L=1 TO NCOMP;
                 I=II(|L,|); J=JJ(|L,|);
                 A(|I,J|)=XDXD(|L,|);
                 B(|I,J|)=YDYD(|L,|);
               END;

      A=A-A`;
      B=B-B`;
      AB=TRACE(A*B`);
      K_=AB/2;
      ASQ=SUM(A##2);
      BSQ=SUM(B##2);
      VG_=SQRT(ASQ#BSQ);
      DEN_K_=SQRT(ASQ#BSQ)/2;

  IF K_=0 & DEN_K_=0 THEN DO; CORR_=0; P_VALUE_=1; VARK_=0; END;

  ELSE DO;
      CCK_=1/DEN_K_;
      CORR_=CCK_*K_;
     *CORR_=AB/SQRT(ASQ#BSQ);

                DO I=1 TO N;
                DO J=1 TO N;
                   ATERM=SUM(A(|I,J|)#A(|I,|));
                   BTERM=SUM(B(|I,J|)#B(|I,|));
                   IF I=1 & J=1 THEN DO; AT=ATERM; BT=BTERM; END;
                   ELSE DO; AT=AT+ATERM; BT=BT+BTERM; END;
                END;
                END;

    VARAB=4/(N#(N-1)#(N-2))#(AT-ASQ)#(BT-BSQ)
         +2/(N#(N-1))#ASQ#BSQ;

  * Z=AB/SQRT(VARAB);
    VARK_=VARAB/4;
    Z=K_/SQRT(VARK_);

   /* This part is only for the Kendall trend statistic.;
   ******************************************;
         IF K_>0 THEN DO; KK_=K_-1; END;
    ELSE IF K_=0 THEN DO; KK_=K_;   END;
    ELSE IF K_<0 THEN DO; KK_=K_+1; END;

    Z=KK_/SQRT(VARK_);
   ******************************************;
  */

    P_VALUE_=2#(1-PROBNORM(ABS(Z)));
  END;
END;
```

```
  IF Q=1 THEN DO; VARN=Q; NT=N; DEN_K=DEN_K_;
                  RHO=CORR_; P_VALUE=P_VALUE_;
                  K=K_; VAR=VARK_; VG=VG_; END;
  ELSE DO; VARN=VARN//Q; NT=NT//N; DEN_K=DEN_K//DEN_K_;
                  RHO=RHO//CORR_; P_VALUE=P_VALUE//P_VALUE_;
                  K=K//K_; VAR=VAR//VARK_; VG=VG//VG_; END;

  IF Q=PM THEN DO;

  NAME1={RHO}; NAME2={P_VALUE}; NAME3={N_OBS}; NAME4={K};
  NAME5={VAR}; NAME6={I}; NAME7={J}; NAME8={DEN_K}; NAME9={VG};

  CREATE RHO     FROM RHO    (|COLNAME=NAME1|); APPEND FROM RHO;
  CREATE P_VALUE FROM P_VALUE(|COLNAME=NAME2|); APPEND FROM P_VALUE;
  CREATE N_OBS   FROM NT     (|COLNAME=NAME3|); APPEND FROM NT;
  CREATE K       FROM K      (|COLNAME=NAME4|); APPEND FROM K;
  CREATE VAR   FROM VAR  (|COLNAME=NAME5|); APPEND FROM VAR;
  CREATE I       FROM VARN   (|COLNAME=NAME6|); APPEND FROM VARN;
  CREATE J       FROM VARN   (|COLNAME=NAME7|); APPEND FROM VARN;
  CREATE DEN_K FROM DEN_K(|COLNAME=NAME8|); APPEND FROM DEN_K;
  CREATE VG FROM VG(|COLNAME=NAME9|); APPEND FROM VG;
 END;

END;
FINISH;
RUN KVAR;

*** Getting mp(mp-1)/2 Covariances between mp Trend Statistics;

START KCOV;
    DO P=1 TO PM-1;
       DO Q=P+1 TO PM;
          COLS=P||Q;
          YY=YS(|,COLS|);
          X=T#( YY(|,1|) ^= . & YY(|,2|) ^= .);

 IF SUM(X)=0 THEN DO; COVBC=0; COVBB=0; COVCC=0; N=0; END;

 ELSE DO;
          YY=YY(|LOC(X ^= 0),|);
          X=X(|LOC(X ^= 0),|);

          Y=YY(|,1|); Z=YY(|,2|);
          N=NROW(X);
  IF N<=2 THEN DO; COVBC=0; COVBB=0; COVCC=0; END;
  ELSE DO;

*******************;
X=RANKTIE(X);
Y=RANKTIE(Y);
Z=RANKTIE(Z);
*******************;
```

```
          DO I=1 TO N-1;
             DO J=I+1 TO N;
                XD=X(|J,|)-X(|I,|);
                YD=Y(|J,|)-Y(|I,|);
                ZD=Z(|J,|)-Z(|I,|);
                IF I=1 & J=2 THEN DO;
                    II=I; JJ=J;
                    XDXD=XD; YDYD=YD; ZDZD=ZD;
                    END;
                ELSE DO;
                    II=II//I; JJ=JJ//J;
                    XDXD=XDXD//XD; YDYD=YDYD//YD; ZDZD=ZDZD//ZD;
                  END;
           END;
         END;

/*
 ** Kendall score function;
 *****************;
    XDXD=SIGN(XDXD);
    YDYD=SIGN(YDYD);
    ZDZD=SIGN(ZDZD);
 *****************;
*/

    A=J(N,N,0);
    B=J(N,N,0);
    C=J(N,N,0);
    NCOMP=N#(N-1)/2;

    DO L=1 TO NCOMP;
             I=II(|L,|); J=JJ(|L,|);
             A(|I,J|)=XDXD(|L,|);
             B(|I,J|)=YDYD(|L,|);
             C(|I,J|)=ZDZD(|L,|);
    END;

    A=A-A`;
    B=B-B`;
    C=C-C`;

    SUMAA=SUM(A##2);
    SUMBC=SUM(B#C);
    SUMBB=SUM(B##2);
    SUMCC=SUM(C##2);

       DO I=1 TO N;
             DO J=1 TO N;
                AATERM=SUM(A(|I,J|)#A(|I,|));
                BCTERM=SUM(B(|I,J|)#C(|I,|));
                BBTERM=SUM(B(|I,J|)#B(|I,|));
                CCTERM=SUM(C(|I,J|)#C(|I,|));
```

```
                IF I=1 & J=1 THEN DO;
                       AAT=AATERM; BCT=BCTERM;
                       BBT=BBTERM; CCT=CCTERM; END;
                ELSE DO;
                       AAT=AAT+AATERM; BCT=BCT+BCTERM;
                       BBT=BBT+BBTERM; CCT=CCT+CCTERM; END;
             END;
      END;

  COVBC=(4/(N#(N-1)#(N-2))#(AAT-SUMAA)#(BCT-SUMBC)
       +2/(N#(N-1))#SUMAA#SUMBC)/4;
  COVBB=(4/(N#(N-1)#(N-2))#(AAT-SUMAA)#(BBT-SUMBB)
       +2/(N#(N-1))#SUMAA#SUMBB)/4;
  COVCC=(4/(N#(N-1)#(N-2))#(AAT-SUMAA)#(CCT-SUMCC)
       +2/(N#(N-1))#SUMAA#SUMCC)/4;
 END;
END;

  IF COVBC=0 | COVBB=0 | COVCC=0 THEN DO; WGHT=0; CORBC=0; END;

  ELSE DO; WGHT=SQRT(COVBB#COVCC); CORBC=COVBC/WGHT; END;

  IF P=1 & Q=2 THEN DO; III=P; JJJ=Q; NN=N;
                        COR=CORBC; COV=COVBC; WGT=WGHT; END;
  ELSE DO; III=III//P; JJJ=JJJ//Q; NN=NN//N;
           COR=COR//CORBC; COV=COV//COVBC; WGT=WGT//WGHT; END;

  IF P=PM-1 & Q=PM THEN DO;

     NAME1={I}; NAME2={J}; NAME3={COR}; NAME4={COV}; NAME5={WGT};
     NAME6={NN};

     CREATE II  FROM III(|COLNAME=NAME1|); APPEND FROM III;
     CREATE JJ  FROM JJJ(|COLNAME=NAME2|); APPEND FROM JJJ;
     CREATE COR FROM COR(|COLNAME=NAME3|); APPEND FROM COR;
     CREATE COV FROM COV(|COLNAME=NAME4|); APPEND FROM COV;
     CREATE WGT FROM WGT(|COLNAME=NAME5|); APPEND FROM WGT;
     CREATE NN  FROM  NN(|COLNAME=NAME2|); APPEND FROM  NN; END;

END;
END;
FINISH;

RUN KCOV;

DATA RHO_; MERGE I RHO P_VALUE N_OBS; LABEL I='VARIABLE';
DATA K; MERGE I K;
DATA DEN_K; MERGE I DEN_K;
DATA VAR; MERGE I J VAR;
DATA COR; MERGE II JJ COR;
DATA K;   SET K;   KEEP K;
DATA RHO; SET RHO_; KEEP RHO;
```

```
DATA N_OBS; SET N_OBS; KEEP N_OBS;
DATA P_VALUE; SET RHO_; KEEP P_VALUE;
DATA DEN_K; SET DEN_K; KEEP DEN_K;
DATA VAR; SET VAR; KEEP VAR;
DATA COR; SET COR; KEEP COR;

*** Forming mp by mp Covariance Matrix of Trend Statistics;

PROC IML;
     USE K;          READ ALL INTO K;
     USE RHO;        READ ALL INTO RHO;
     USE N_OBS;      READ ALL INTO N_OBS;
     USE P_VALUE;    READ ALL INTO PVALUE;
     USE DEN_K;        READ ALL INTO DEN_K;
     USE VAR;        READ ALL INTO VAR;
     USE COR;        READ ALL INTO COR;
     USE COV;        READ ALL INTO COV;
     USE WGT;        READ ALL INTO WGT;
     USE NN;         READ ALL INTO NN;
     USE II;         READ ALL INTO II;
     USE JJ;         READ ALL INTO JJ;

     N=NROW(K);
     NCOMP=N#(N-1)/2;

     COV_K=J(N,N,0);
     START;  DO L=1 TO NCOMP;
                I=II(|L,|); J=JJ(|L,|); COV_K(|I,J|)=COV(|L,|); END;
     FINISH; RUN;

     COV_K=DIAG(VAR)+COV_K+COV_K`;
     CREATE COV_K  FROM COV_K; APPEND FROM COV_K;

  FREE COV_K COV;

*** Getting p Overall Trend Statistics for Constituents and
    Forming p by p Covariance Matrix of Overall Trend Statistics;

PROC IML;    USE K;          READ ALL INTO K;
             USE COV_K;      READ ALL INTO COV_K;
             USE DEN_K;      READ ALL INTO DEN_K;
             USE N_OBS;      READ ALL INTO N_OBS;

     M=12;
     P=NROW(K)/M;
     C=J(P, M*P, 0);
     START;
     DO G=1 TO P; DO H=1 TO M*P;
        IF H >= (G-1)*M+1 & H <= G*M THEN DO; C(|G,H|)=1; END;
                                     ELSE DO; C(|G,H|)=0; END;
     END; END;
     FINISH; RUN;
```

```
 W=C*K;
 COV_W=C*COV_K*C`;
 NOBS=C*N_OBS;
 DEN_W=C*DEN_K;
 OVERCOR=DIAG(DEN_W##(-1))*W;
 COV_OC=DIAG(DEN_W##(-1))*COV_W*(DIAG(DEN_W##(-1)))`;

 FREE K COV_K DEN_K N_OBS;

 VNAMES={DO PH NFR FR FC};

CREATE W       FROM W;        APPEND FROM W;
CREATE COV_W   FROM COV_W(|COLNAME=VNAMES|);   APPEND FROM COV_W;
CREATE NOBS    FROM NOBS;     APPEND FROM NOBS;
CREATE OVERCOR FROM OVERCOR; APPEND FROM OVERCOR;
CREATE COV_OC  FROM COV_OC(|COLNAME=VNAMES|);  APPEND FROM COV_OC;
```

10 References

Bhattacharyya, G. K. and J. H. Klotz (1966). The bivariate trend of Lake Mendota, *Technical Report No. 98*, Department of Statistics, University of Wisconsin, Madison, Wisconsin, USA.

Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations, *Biometrika*, **33**, 129-135.

Dietz, E. J. (1983). Small sample power of multivariate nonparametric tests for monotone trend, *Journal of Statistical Computation and Simulation*, **18**, 159-177.

Dietz, E. J. and T. J. Killeen (1981). A nonparametric multivariate test for monotone trend with pharmaceutical applications, *Journal of the American Statistical Association*, **76**, 169-174.

Edgington, E. S. (1980). *Randomization Tests*, New York: Marcel Dekker.

Hirsch, R. M. and J. R. Slack (1984). A nonparametric test for seasonal data with serial dependence, *Water Resources Research*, **20**(6), 727-732.

Hirsch, R. M. J. R. Slack, and R. A. Smith (1982). Techniques of trend analysis for monthly water quality data, *Water Resources Research*, **18**(1), 107-121.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution, *Annals of Mathematical Statistics* , **19**, 293-325.

Holtzman, G. I. (1985). Computation of the distribution of Page's criterion, given any configuration of ties, by a fast Fourier transform, *Journal of Statistical*

Computation and Simulation, **23**(1 & 2), 149-154.

Johnson, N. L. and S. Kotz (1970). *Distributions in Statistics: Continuous Univariate Distributions-2*, Boston: Houghton Mifflin Company.

Jonckheere, A. R. (1954). A distribution-free k sample test against ordered alternatives, *Biometrika*, **41**, 135-145.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Baed on Ranks*, San Francisco: Holden-Day.

Lettenmaier, D. O. (1988). Multivariate nonparametric tests for trend in water quality, *Water Resources Bulletin*, **24**(3), 505-512.

Loftis, J. C. C. H. Taylor, and P. L. Chapman (1991). Multivariate tests for trend in water quality, *Water Resources Research*, **27**(7), 1421-1429.

Kendall, M. G. and J. D. Gibbons (1990). *Rank Correlation Methods*, 5th edition, New York: Oxford University Press, London: Edward Arnold.

Page, E. B. (1963). Ordered hypothesis for multiple treatments; A significant test for linear ranks, *Journal of the American Statistical Association*, **58**, 216-230.

Puri, M. L. and P. K. Sen (1971). *Nonparametric Methods in Multivariate Analysis*, New York: John Wiley.

van Belle, G. and J. P. Hughes (1984). Nonparametric tests for trend in water quality, *Water Resources Research*, **20**(1), 127-136.

Zipper, C. E. G. I. Holtzman, S. Rheem, and G. K. Evanylo (1992). Surface water quality trends in southwestern Virginia, 1970-1989: I. Seasonal Kendall analysis, *Bulletin 173*, Virginia Water Resource Research Center, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA.

11 Vita

Sungsue Rheem was born on August 26, 1957 in Seoul, South Korea to Changhyuk Rheem and Sangbin Park. He graduated from Bosung High School in Seoul in 1976. Sungsue continued his education as a statistics major at Korea University, Seoul, Korea, where he received a bachelor's degree (1981) and a master's degree (1983) in statistics. After being discharged from the 14-month military service in 1985, he taught introductory courses in statistics at Korea University. He then entered the Ph.D. program in Statistics at Virginia Polytechnic Institute and State University (Virginia Tech) in 1987. He received his M.S. degree in statistics in 1989 and was awarded his Ph.D. degree in statistics in 1992 at Virginia Tech.

Sungsue has actively partipated in applied statistical activities as a research assistant and a statistical consultant. He topped the list among applicants for eight departments in three colleges in the master's program entrance examination at Korea University in 1981. Sungsue received a Student Travel Award from the Biometric Society in 1992 for his paper presented at the Spring 1992 East Northern American Region (ENAR) Meeting. He married Younghee Kim in 1987 and his daughter, Jean, was born in 1989.

He was the 1991-92 President of the Korean Graduate Student Association at Virginia Tech. He likes to sing and make music; so far, he has composed over twenty songs including two hymns.

Sungsue Rheem