Communications of the Korean Statistical Society
2012, Vol. 19, No. 3, 471–478

DOI: http://dx.doi.org/10.5351/CKSS.2012.19.3.471

Canonical Correlation: Permutation Tests and Regression

Jae Keun Yoo[1,a], Hee Youn Kim[a], Hye Yeon Um[a]

[a]Department of Statistics, Ewha Womans University

Abstract

In this paper, we present a permutation test to select the number of pairs of canonical variates in canonical correlation analysis. The existing chi-squared test is known to be limited to normality in use. We compare the existing test with the proposed permutation test and study their asymptotic behaviors through numerical studies. In addition, we connect canonical correlation analysis to regression and we we show that certain inferences in regression can be done through canonical correlation analysis. A regression analysis of real data through canonical correlation analysis is illustrated.

Keywords: Canonical correlation, multivariate analysis, permutation test, regression.

1. Introduction

Principal component analysis should be one of the most favorable statistical methods for dimension reduction in high-dimensional data analysis. When a relationship between two sets of high-dimensional variables is of interest, principal component analysis may not be useful, because it does conduct a marginal dimension reduction of each set without considering any association between the two sets.

Canonical correlation analysis(CCA) replaces the original two sets of variables with pairs of linear combinations from two sets of variables, acquired through the maximization of the Pearson correlation between the two sets. The linear combination pairs and their correlations are called *canonical variates* and *canonical correlations*, respectively. A few pairs of canonical variates are expected to represent the original sets of variables to explain their relation and variabilities. Therefore, when we need to do simultaneous dimension reduction for two sets of variables, CCA can achieve a potentially better and more reliable dimension reduction than principal component analysis.

This paper develops a permutation test and compares it with the existing chi-squared test in the determination of the number of pairs of canonical variates. Since the chi-squared test assumes normality, it may mislead the determination under its violation. Since the permutation test does not require normality, it should be a possible alternative to the chi-squared test against non-normality. By the comparison of the two tests, we will investigate the robustness of the chi-squared test to non-normality. In addition, we study the relation between CCA and regression. One special case to study the association between two sets of variables should be to study changes of univariate or multivariate responses in distribution as predictors vary. In such case univariate or multivariate regression should be a popular statistical tools. We will show that certain inferences in regression can be done through CCA.

For the corresponding author Jae Keun Yoo, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (KRF) funded by the Ministry of Education, Science and Technology (2011-0005581).

[1] Corresponding author: Assistant Professor, Department of Statistics, Ewha Womans University, 11-1 Daehyun-Dong Seodaemun-Gu, Seoul 120-750, Korea. E-mail: peter.yoo@ewha.ac.kr

The organization of the paper is as follows. In Section 2 we review canonical correlation analysis. Section 3 is devoted to the development of a permutation test to select the number of pairs of canonical variates and compare the performance of the chi-squared tests via numerical studies. We will investigate a relation between canonical correlation analysis and regression in Section 4. Section 5 contains real data application. In Section 6, we summarize our work.

2. Canonical Correlation Analysis

2.1. Canonical correlation analysis

Suppose that our interest is placed onto studying an association of two sets of variables of $\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{Y} \in \mathbb{R}^r$. Define that $\text{cov}(\mathbf{X}) = \mathbf{\Sigma}_X > 0$, $\text{cov}(\mathbf{Y}) = \mathbf{\Sigma}_Y > 0$, $\mathbf{\Sigma}_{XY} = \text{cov}(\mathbf{X}, \mathbf{Y})$ and $\mathbf{\Sigma}_{XY}^{\mathrm{T}} = \text{cov}(\mathbf{Y}, \mathbf{X})$. Considering two linear combinations of $\mathbf{X}$ and $\mathbf{Y}$ of $U = \mathbf{a}^{\mathrm{T}}\mathbf{X}$ and $V = \mathbf{b}^{\mathrm{T}}\mathbf{Y}$, we have $\text{var}(U) = \mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}_X\mathbf{a}$, $\text{var}(V) = \mathbf{b}^{\mathrm{T}}\mathbf{\Sigma}_Y\mathbf{b}$, and $\text{cov}(U, V) = \mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}_{XY}\mathbf{b}$, where $\mathbf{a} \in \mathbb{R}^{p\times 1}$ and $\mathbf{b} \in \mathbb{R}^{r\times 1}$. We pursue to construct $\mathbf{a}$ and $\mathbf{b}$ to maximize Pearson-correlation between U and V:

$$\text{corr}(U, V) = \frac{\mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}_{XY}\mathbf{b}}{\sqrt{\mathbf{a}^{\mathrm{T}}\mathbf{\Sigma}_X\mathbf{a}}\sqrt{\mathbf{b}^{\mathrm{T}}\mathbf{\Sigma}_Y\mathbf{b}}}. \tag{2.1}$$

Classical CCA constructs such $\mathbf{a}$ and $\mathbf{b}$ based on the following criteria:

1. The first canonical variate pair ($U_1 = \mathbf{a}_1^{\mathrm{T}}\mathbf{X}, V_1 = \mathbf{b}_1^{\mathrm{T}}\mathbf{Y}$) is constructed from the maximization of (2.1).
2. The second canonical variate pair ($U_2 = \mathbf{a}_2^{\mathrm{T}}\mathbf{X}, V_2 = \mathbf{b}_2^{\mathrm{T}}\mathbf{Y}$) is constructed from the maximization of (2.1) with restriction that $\text{var}(U_2) = \text{var}(V_2) = 1$ and (U_1, V_1) and (U_2, V_2) are uncorrelated.
3. At the k step, the k^{th} canonical variate pair ($U_k = \mathbf{a}_k^{\mathrm{T}}\mathbf{X}, V_k = \mathbf{b}_k^{\mathrm{T}}\mathbf{Y}$) is obtained from the maximization of (2.1) with restriction that $\text{var}(U_k) = \text{var}(V_k) = 1$ and (U_k, V_k) are uncorrelated with the previous $(k-1)$ canonical variate pairs.
4. Repeat Step 1 and Step 3 until $q = \min(p, r)$.
5. Select the first d pairs of (U_k, V_k) to represent the relationship between $\mathbf{X}$ and $\mathbf{Y}$.

Then the according pairs $(\mathbf{a}_i, \mathbf{b}_i)$ are obtained as follows: $\mathbf{a}_i = \mathbf{\Sigma}_X^{-1/2}\boldsymbol{\psi}_i$ and $\mathbf{b}_i = \mathbf{\Sigma}_Y^{-1/2}\boldsymbol{\phi}_i$ for $i = 1, \ldots, q$, where $(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_q)$ and $(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_q)$ are the q eigenvectors of $\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1}\mathbf{\Sigma}_{XY}^{\mathrm{T}}\mathbf{\Sigma}_X^{-1/2}$ and $\mathbf{\Sigma}_Y^{-1/2}\mathbf{\Sigma}_{XY}^{\mathrm{T}}\mathbf{\Sigma}_X^{-1}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}$, respectively, with the corresponding common eigenvalues of $\rho_1^{*2} \geq \cdots \geq \rho_q^{*2} \geq 0$. Then matrices of $\mathbf{A}_d = (\mathbf{a}_1, \ldots, \mathbf{a}_d)$ and $\mathbf{B}_d = (\mathbf{b}_1, \ldots, \mathbf{b}_d)$ are called *canonical direction matrices* for $d = 1, \ldots, q$. For more details regarding the CCA, readers may refer to Johnson and Wichern (2007). In practice, $\mathbf{\Sigma}_X$, $\text{cov}(\mathbf{Y})$, $\mathbf{\Sigma}_{XY}$, and $\mathbf{\Sigma}_Y$ are replaced with their usual moment estimators of $\hat{\mathbf{\Sigma}}_X$, $\hat{\mathbf{\Sigma}}_Y$, $\hat{\mathbf{\Sigma}}_{XY}$, and $\hat{\mathbf{\Sigma}}_{YX}$. Throughout the rest of the paper, a notation of d will stand for the true number of pairs of canonical variates.

3. Permutation Test in Canonical Correlation Analysis

3.1. Permutation test

To determine how many canonical covariate pairs should be selected in CCA for two sets of variables of $\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{Y} \in \mathbb{R}^r$, large sample inferences by Bartlett (1938, 1939) are widely used. Recalling

$q = \min(p, r)$, the inference procedure sequentially tests the following hypotheses:

$$H_0 : d = m \quad versus \quad H_1 : d > m, \quad m = 0, 1, \ldots, (q-1).$$

Beginning with $m = 0$, if $H_0 : d = m$ is rejected, increment m by 1 and redo the test. We stop the test the first time that H_0 is not rejected and setting $\hat{d} = m$. This test procedure requires a test statistic for $H_0 : d = m$, and, as the statistic, Bartlett (1938, 1939) proposed

$$\hat{\Lambda}_m^{\mathrm{BT}} = -\left\{n - 1 - \frac{1}{2}(p + r + 1)\right\} \sum_{i=m+1}^{q} \log\left(1 - \hat{\rho}_i^{*2}\right), \quad m = 0, 1, \ldots, (q-1),$$

where $\hat{\rho}_1^{*2} \geq \cdots \geq \hat{\rho}_q^{*2}$ are the ordered eigenvalues of $\hat{\mathbf{\Sigma}}_X^{-1/2}\hat{\mathbf{\Sigma}}_{XY}\hat{\mathbf{\Sigma}}_Y^{-1}\hat{\mathbf{\Sigma}}_{XY}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_X^{-1/2}$.

If $\mathbf{X}$ and $\mathbf{Y}$ are jointly normal, $\hat{\Lambda}_m^{\mathrm{BT}}$ tends to converge in distribution to $\chi^2_{(p-m)(r-m)}$ under $H_0 : d = m$. In practice, however, if the normality is a cause of concern, $\hat{\Lambda}_m^{\mathrm{BT}}$ may be problematic for the determination of d.

Here we propose a permutation test to estimate d as a possible alternative of the Barttlet test against non-normality. The proposed test procedure does not assume underlying distributions of $\mathbf{X}$ and $\mathbf{Y}$. Therefore, the permutation test may have potential advantages over the Bartlett test under violation of normality. We describe how to conduct the permutation test in CCA as follows:

(1) Under $H_0 : d = m$, using the original data of $\mathbf{X}$ and $\mathbf{Y}$, compute $\hat{\mathcal{T}}_m$ and partition eigenvectors of $\hat{\mathbf{\Sigma}}_X^{-1/2}\hat{\mathbf{\Sigma}}_{XY}\hat{\mathbf{\Sigma}}_Y^{-1}\hat{\mathbf{\Sigma}}_{YX}\hat{\mathbf{\Sigma}}_X^{-1/2}$ as follows:

$$\hat{\mathcal{T}}_m = \sum_{i=m+1}^{p} \hat{\rho}_i^{*2} \quad \text{and} \quad \hat{\mathbf{\Psi}}_1 = \left(\hat{\boldsymbol{\psi}}_1, \ldots, \hat{\boldsymbol{\psi}}_m\right), \;\; \hat{\mathbf{\Psi}}_2 = \left(\hat{\boldsymbol{\psi}}_{m+1}, \ldots, \hat{\boldsymbol{\psi}}_p\right),$$

where $\hat{\rho}_1^{*2} \geq \cdots \geq \hat{\rho}_q^{*2}$ and $\hat{\boldsymbol{\psi}}_i$ are the eigenvector corresponding to $\hat{\rho}_i^{*2}$. We will denote $\hat{\mathcal{T}}_m$ from the original data as $\hat{\mathcal{T}}_m^{\mathrm{ref}}$.

(2) Let $\hat{\mathbf{Z}}_{X_i} = \hat{\mathbf{\Sigma}}_X^{-1/2}(\mathbf{X}_i - \bar{\mathbf{X}})$, where $\mathbf{X}_i$ stands for the i^{th} observation of $\mathbf{X}$ and $\bar{\mathbf{X}}$ is the sample mean vector of $\mathbf{X}$, that is, $\bar{\mathbf{X}} = (1/n)\sum_{i=1}^{n} \mathbf{X}_i$. Construct $\hat{\mathbf{U}}_i = \hat{\mathbf{Z}}_{X_i}^{\mathrm{T}}\hat{\mathbf{\Psi}}_1$ and $\hat{\mathbf{W}}_i = \hat{\mathbf{Z}}_{X_i}^{\mathrm{T}}\hat{\mathbf{\Psi}}_2$.

(3) Randomly permute the indices i of the $\hat{\mathbf{W}}_i$ to obtain the permuted set $\hat{\mathbf{W}}_i^*$.

(4) Construct the test statistic $\hat{\mathcal{T}}_m$ by applying usual canonical correlation analysis to the permuted data of $(\hat{\mathbf{U}}, \hat{\mathbf{W}}^*)$ and $\mathbf{Y}$. We will denote $\hat{\mathcal{T}}_m$ from the permuted data as $\hat{\mathcal{T}}_m^{\mathrm{perm}}$.

(5) Repeat Step (3)–Step (4) N times. The p-value of testing $H_0 : d = m$ is the fraction of $\hat{\mathcal{T}}_m^{\mathrm{perm}}$s that exceed $\hat{\mathcal{T}}_m^{\mathrm{ref}}$.

Since each of N permutations is not always the same, p-values should be different. Experience tells that $N = 500$ should be fine with most cases.

3.2. Comparison of the Bartllet chi-squared test via numerical studies

As numerical studies we constructed two sets of variables of $\mathbf{X}$ and $\mathbf{Y}$ as follows:

Model 1 $\mathbf{X} = (X_1, \ldots, X_{10}) \overset{iid}{\sim} N(0, 1)$ and $\mathbf{Y} = (Y_1, Y_2)$, where $Y_1 = \sin(X_1 + X_2) + \varepsilon_1$, $Y_2 = 2 * \sin(X_1 + X_2) + \varepsilon_2$.

Table 1: Simulation results for Model 1 in Section 3.2

		Permutation test			Barttlet test		
		$\hat{d}=0$	$\hat{d}=1$	$\hat{d}\geq 2$	$\hat{d}=0$	$\hat{d}=1$	$\hat{d}\geq 2$
$\boldsymbol{\varepsilon} \overset{iid}{\sim} N(0,1)$	$n=50$	1.5	90.5	8.0	1.5	92.0	6.5
	$n=100$	0.0	94.5	5.5	0.0	95.0	5.0
	$n=200$	0.0	95.0	5.0	0.0	95.0	5.0
$\boldsymbol{\varepsilon} \overset{iid}{\sim} t_3$	$n=50$	27.5	67.5	5.0	84.5	76.0	3.5
	$n=100$	3.0	92.0	5.0	87.8	94.5	5.0
	$n=200$	0.0	94.5	5.5	90.4	95.0	5.0
$\boldsymbol{\varepsilon} \overset{iid}{\sim} U(-2,2)$	$n=50$	11.0	85.5	5.0	12.0	81.5	6.5
	$n=100$	3.0	92.0	5.0	87.8	96.0	4.0
	$n=200$	0.0	95.0	5.0	90.4	93.5	6.5
$\varepsilon_i \overset{indep}{\sim} U(-i,i)$	$n=50$	7.0	89.0	4.0	6.0	90.5	3.5
	$n=100$	0.0	92.5	7.5	0.0	93.0	7.0
	$n=200$	0.0	96.0	4.0	0.0	95.5	4.5

In Model 1, we considered four types of the distributions for $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)$, which is independent of $\mathbf{X}$ in all cases: (1) $\boldsymbol{\varepsilon} \overset{iid}{\sim} N(0,1)$; (2) $\boldsymbol{\varepsilon} \overset{iid}{\sim} t_3$; (3) $\boldsymbol{\varepsilon} \overset{iid}{\sim} U(-2,2)$; (4) $\varepsilon_i \overset{indep}{\sim} U(-i,i)$, $i = 1,2$, where notations of t_k and $U(-k,k)$ stand for student t distribution with k degrees of freedom and continuous uniform distribution from $-k$ to k, respectively.

By the variable configurations, the joint distribution of $\mathbf{X}$ and $\mathbf{Y}$ is not normal for all types of $\boldsymbol{\varepsilon}$, and the relation between $\mathbf{X}$ and $\mathbf{Y}$ is not linear. In addition, one pair of canonical variates of $X_1 + X_2$ and $(Y_1 + Y_2)/2$ for $\mathbf{X}$ and $\mathbf{Y}$ should be adequate to summarize the association between $\mathbf{X}$ and $\mathbf{Y}$, and hence we have $d = 1$. For both tests, we considered $n = 50, 100$ and 200, and Model 1 was iterated 500 times for each case of $\boldsymbol{\varepsilon}$. In addition, the number of permutations was 500. The results of the sequential tests for $d = 0$ and $d = 1$ through the Barttlet test and the permutation test are summarized in Table 1.

These simulation results represent the characteristic behaviors in the estimation of d observed in other simulations. Regardless of the distributions of random errors of $\boldsymbol{\varepsilon}$, the two tests showed similar results. According to Table 1, (surprisingly and a bit disappointingly), the proposed permutation test for the estimation of d did not show clear dominance over the Barttlet test for most simulation models. The two tests provided similar results, and hence it can be concluded that the joint normality between $\mathbf{X}$ and $\mathbf{Y}$ is not problematic in the Barttlet test. In practice, however, the two tests can determine the estimate of d differently. If so, one can take advantage of the permutation test for the decision to select the number of pairs of canonical variates, and hence they can withdraw a more reliable conclusion for the estimation of d.

4. Application of Canonical Correlation Analysis in Regression Analysis

4.1. Canonical direction matrix and ordinary least squares

Suppose that we consider regression of $\mathbf{Y} \in \mathbb{R}^r | \mathbf{X} \in \mathbb{R}^p$ with $r \geq 1$. Most of regression problems deal with parametric or non-parametric modeling through predictors $\mathbf{X}$. Therefore, to connect CCA and regression, we consider the canonical direction matrix $\mathbf{A}_d$ alone, which is related to $\mathbf{X}$. Also we define that $\mathcal{S}(\mathbf{M})$ represent a subspace spanned by the columns of a $p \times r$ matrix $\mathbf{M}$,

The canonical direction matrix $\mathbf{A}_d$ for $\mathbf{X}$ is equal to $\mathbf{A}_d = \boldsymbol{\Sigma}_X^{-1/2}(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_d)$, where vectors of $\boldsymbol{\psi}_i$, $i = 1, \ldots, d$, are the eigenvectors corresponding to its non-zero ordered eigenvalues $\boldsymbol{\Sigma}_X^{-1/2}\boldsymbol{\Sigma}_{XY}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\Sigma}_{XY}^{\mathrm{T}}$ $\boldsymbol{\Sigma}_X^{-1/2}$. Therefore, it can be easily noted that the set of the vectors of $(\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_d)$ forms an orthonormal basis matrix of $\mathcal{S}(\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1}\mathbf{\Sigma}_{XY}^{\mathrm{T}}\mathbf{\Sigma}_X^{-1/2})$. Then, by the following equivalence of

$$\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1}\mathbf{\Sigma}_{XY}^{tr}\mathbf{\Sigma}_X^{-1/2} = \mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}\mathbf{\Sigma}_Y^{-1/2}\mathbf{\Sigma}_{XY}^{\mathrm{T}}\mathbf{\Sigma}_X^{-1/2} = \left(\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}\right)\left(\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}\right)^{\mathrm{T}},$$

we easily have $\mathcal{S}(\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1}\mathbf{\Sigma}_{XY}^{\mathrm{T}}\mathbf{\Sigma}_X^{-1/2}) = \mathcal{S}(\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2})$. Therefore, we can establish the following relation from the above equivalence:

$$\mathcal{S}(A_d) = \mathbf{\Sigma}_X^{-1/2}\mathcal{S}\left(\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}\right) = \mathcal{S}\left(\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_X^{-1/2}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}\right) = \mathcal{S}\left(\mathbf{\Sigma}_X^{-1}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}\right). \tag{4.1}$$

Since, for a $p \times r$ matrix, post-multiplication of any $r \times r$ non-singular matrix does not change its rank and column space, we derive the following relation from the last equivalence of (4.1):

$$\mathcal{S}(\mathbf{A}_d) = \mathcal{S}\left(\mathbf{\Sigma}_X^{-1}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}\right) = \mathcal{S}\left(\mathbf{\Sigma}_X^{-1}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_Y^{-1/2}\mathbf{\Sigma}_Y^{1/2}\right) = \mathcal{S}\left(\mathbf{\Sigma}_X^{-1}\mathbf{\Sigma}_{XY}\right) = \mathcal{S}(\boldsymbol{\beta}). \tag{4.2}$$

The last equivalence of (4.2) directly implies that the columns of the canonical direction matrix for $\mathbf{X}$ and those of the OLS coefficient matrix span the same subspace. Therefore, whenever we have to do inference about $\mathcal{S}(\boldsymbol{\beta})$, for example its dimension or basis estimation, it can be done through CCA. In the next three subsections, we will show how CCA can be directly applied in regression inference problems by the relation of (4.2).

4.2. Linear regression: ANOVA F-test

In linear regression of $Y \in \mathbb{R}^1 \,|\, \mathbf{X} \in \mathbb{R}^p$, ANOVA F-test is to test the hypotheses of H_0 : all β_is are equal to zero against H_1 : at least one of β_is is not zero. Geometrically, the null hypothesis can be equivalently interpreted as $H_0 : \dim(\mathcal{S}(\boldsymbol{\beta})) = 0$. Then the ANOVA F-test is equivalent to test $H_0 : d = 0$ in CCA, or equivalently $\mathbf{\Sigma}_{z_x z_y} = 0$. Therefore, the F-test can be alternatively done via either the Barttlet test or the permutation test in CCA.

4.3. Linear regression: Brusch-Pagan test for heteroscedasticity

Breusch-Pagan test (1979) is one among many tests for heteroscedasticity in the linear regression of $Y \in \mathbb{R}^1 \,|\, \mathbf{X} \in \mathbb{R}^p$. To conduct the test, several assumptions are required: error terms are independently normal and their variance σ_i^2 has the following relation to $\mathbf{X} = (X_1, \ldots, X_p)$:

$$\log \sigma_j^2 = \gamma_0 + \sum_{i=1}^{p} \gamma_i X_{ij}, \quad j = 1, \ldots, n. \tag{4.3}$$

Under (4.3), if one of γ_is, $i = 1, \ldots, n$, is not zero, then the regression has heteroscedasticity. To test $H_0 : \gamma_1 = \cdots = \gamma_p = 0$ against H_1 : one of γ_is, at least, is not zero, the following procedure is taken:

(1) Compute the residuals e_i and usual sums of square of error, SSE from the OLS fit of $Y \,|\, \mathbf{X}$.

(2) Obtain usual sums of square of regression, SSR^*, from the OLS fit of $\mathrm{e}^2 \,|\, \mathbf{X}$.

(3) Compute $X^2 = (\mathrm{SSR}^*/2)/(\mathrm{SSE}/n)^2$.

(4) Then X^2 has the asymptotic χ^2 with p degrees of freedom.

Then the null hypothesis of the Brusch-Pagan test is equivalent to the hypothesis of $\dim(\mathcal{S}(\boldsymbol{\gamma})) = 0$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)$. That is, we must test that the space spanned by the OLS coefficient vector from the regression of $\mathrm{e}^2 \,|\, \mathbf{X}$ is the null space. Then it can be simply done through testing $H_0 : d = 0$ in usual CCA application of two sets of variables of $\mathbf{X}$ and e^2.

4.4. Reduced-rank regression

When the interest is placed on changes of multi-dimensional responses in distribution as predictors vary, multivariate linear regression should be one of the popular statistical tools. The classical multivariate linear regression of $\mathbf{Y} \in \mathbb{R}^r \,|\, \mathbf{X} \in \mathbb{R}^p$ with $r \geq 2$ is as follows:

$$\mathbf{Y} \,|\, \mathbf{X} = \boldsymbol{\alpha} + \mathbf{M}^{\mathrm{T}}\mathbf{X} + \boldsymbol{\varepsilon}, \tag{4.4}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^r$ is an intercept vector, $\mathbf{M} \in \mathbb{R}^{r\times p}$ is an unknown coefficient matrix, the error vector $\boldsymbol{\varepsilon} \in \mathbb{R}^r \sim MN(0, \boldsymbol{\Sigma} \geq 0)$ is independent of $\mathbf{X}$. A notation of MN stands for multivariate normal distribution. In addition, it is assumed that $\boldsymbol{\Sigma} > 0$ throughout the rest of the paper.

A classical reduced-rank regression under (4.4) is defined, if $\mathrm{rank}(\mathbf{M}) < \min(p, r)$. Therefore, the reduced-rank regression has been mostly used, when there is a necessity to reduce the number of parameters in (4.4). The reduced-rank regressions have a wide spectrum of applications in fields such as chemometrics, psychometrics, econometrics, and financial economics. The analysis of the reduced-rank regression is conducted under the assumption that the coefficient matrix $\mathbf{B}$ is not of full rank.

In the reduced-rand regression, the elements of $\mathbf{M}$ are subsequently estimated for a given value of the rank of $\mathbf{M}$, and $\mathbf{M}$ is be replaced by the ordinary least square, $\boldsymbol{\beta} = \boldsymbol{\Sigma}_X^{-1}\boldsymbol{\Sigma}_{XY}$, under (4.4).

Since $\boldsymbol{\beta}$ is not full-column rank by the assumption of the reduced-rank regression, it can be expressed as $\boldsymbol{\beta} = \boldsymbol{\eta}\boldsymbol{\gamma}$, where $\boldsymbol{\eta} \in \mathbb{R}^{p\times d}$ and $\boldsymbol{\gamma} \in \mathbb{R}^{d\times r}$. Then, in (4.4), $\mathbf{B}$ can be replaced with $\boldsymbol{\eta}\boldsymbol{\gamma}$:

$$\mathbf{Y} \,|\, \mathbf{X} = \boldsymbol{\alpha} + \boldsymbol{\gamma}^{\mathrm{T}} \left(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}\right) + \boldsymbol{\varepsilon}.$$

In the regression model above, $\boldsymbol{\gamma}$ is nothing but the OLS coefficient matrix from the regression of $\mathbf{Y} \,|\, \boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$. Therefore, once $\boldsymbol{\eta}$ is known, $\boldsymbol{\gamma}$ can be easily constructed.

It can be easily noted that $\boldsymbol{\eta}$ forms a basis matrix of $\mathcal{S}(\boldsymbol{\beta})$. Then such $\boldsymbol{\eta}$ is clearly not unique. That is, for any other basis $\boldsymbol{\eta}_*$ of $\mathcal{S}(\boldsymbol{\beta})$, the relation of $\boldsymbol{\beta} = \boldsymbol{\eta}_*\boldsymbol{\gamma}_*$ and the following equation hold:

$$\mathbf{Y} \,|\, \mathbf{X} = \alpha + \boldsymbol{\gamma}_*^{\mathrm{T}} \left(\boldsymbol{\eta}_*^{\mathrm{T}}\mathbf{X}\right) + \delta.$$

In addition, the reduced-rank regression under (4.4) forces the following equivalences:

$$E(\mathbf{Y} \,|\, \mathbf{X}) = E\left(\mathbf{Y} \,|\, \mathbf{M}^{\mathrm{T}}\mathbf{X}\right) = E\left(\mathbf{Y} \,|\, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right) = E\left(\mathbf{Y} \,|\, \boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}\right) = E\left(\mathbf{Y} \,|\, \boldsymbol{\eta}_*^{\mathrm{T}}\mathbf{X}\right). \tag{4.5}$$

The equivalence (4.5) directly implies that lower-dimensional linearly transformed predictors of either $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ or $\boldsymbol{\eta}_*^{\mathrm{T}}\mathbf{X}$ can replace the original p-dimensional predictor $\mathbf{X}$ without information on $E(\mathbf{Y} \,|\, \mathbf{X})$. Therefore, the primary interest of the reduced-rank regression is placed onto the estimation of any orthonormal basis matrix $\boldsymbol{\eta}$ of $\mathcal{S}(\boldsymbol{\beta})$.

Then the key-equivalence of $\mathcal{S}(\mathbf{A}_d) = \mathcal{S}(\boldsymbol{\beta})$ in (4.2) directly implies that $\mathbf{A}_d$ should be one of possible choice of $\boldsymbol{\eta}$. Therefore, one can easily accomplish reduced-rank regression of $\mathbf{Y} \,|\, \mathbf{X}$ through a canonical direction matrix for $\mathbf{X}$ acquired by usual CCA application of $\mathbf{X}$ and $\mathbf{Y}$. That is, we can replace $\mathbf{X}$ with $\mathbf{A}_d^{\mathrm{T}}\mathbf{X}$ without loss of information on $E(\mathbf{Y} \,|\, \mathbf{X})$.

Since most statistical packages provide functions or procedures to conduct CCA, one can do reduced-rank regression, if necessary, although the packages do not provide those for reduced-rank regression.

Table 2: Determination of the number of pairs of canonical variates in Section 4

	Permutation test	Barttlet test
$H_0 : d = 0$ *vs.* $H_1 : d > 0$	0.000	0.000
$H_0 : d = 1$ *vs.* $H_1 : d > 1$	0.150	0.165

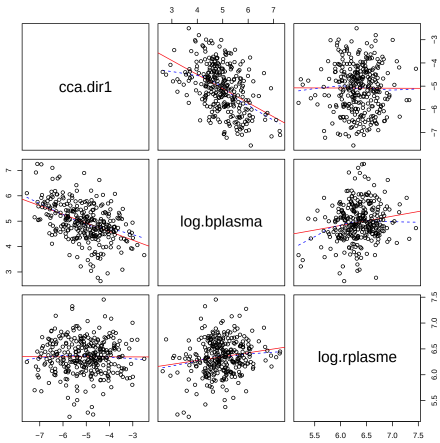

Figure 1: *A scatter plot matrix of the first estimated canonical variate for predictors and two responses in Section 4: cca.dir1, the first estimated canonical variate for predictors; log.bplasma, log-scale beta-carotene plasma concentration levels; log.rplasma, log-scale retinol plasma concentration levels*

5. Real Data Application: Beta-Carotene Plasma

For illustration purpose, we investigate a reduced-rank regression study of Beta-carotene and Retinol plasma concentration levels given the following dietary factors and smoking status: (1) dietary beta-carotene consumed (mcg per day, bet); (2) grams of fat consumed per day (fat); (3) number of calories consumed per day (cal); (4) grams of fiber consumed per day (fiber); (5) dietary retinol consumed (mcg per day) (ret); (6) quetelet (weight/height^2, quet); (7) vitamin usage (1 = often; 2 = sometimes; 3 = no usages, vit); (8) smoker (0 = non-smoker; 1 = former smoker; 2 = current smoker).

This study was originally done in Nierenberg *et al.* (1989). They found that dietary carotene was positively related to Beta-carotene levels, while Quetelet was negatively related. The data was obtained from *StatLib* webpage and used under permission. Since cases with numbers 257 was suspected as an outlier, they were deleted from the data set, and the total number of sample sizes were 314.

We log-transformed 5 continuous predictors other than quetelet, which was transformed to its inverse scale to reduce variabilities in predictors. Vitamin usages and smoker were re-coded with dummy variables such that vit2 (1, if vitamin usage = 2 and 0, otherwise), vit3 (1, if vitamin usage = 3 and 0, otherwise), smoker1 (1, if smoker = 1 and 0, otherwise), and smoker2 (1, if smoker = 2 and 0, otherwise). Therefore, the total numbers of predictors and responses were 10 and 2, respectively.

We conducted the the Barttlet and permutation tests for the data, and the test results are reported in Table 2. We used 500 permutations for the permutation test. According to Table 2, the results from the permutation and Barttlet tests are the same with level 5%, and both tests determine that $\hat{d} = 1$. Therefore, one linear combination of $\mathbf{X}$ can summarize the two-dimensional response regression summarized in Figure 1 with the log-transformed responses. The estimated canonical covariate (cca.dir1) for the predictors is defined as follows:

$$\text{cca.dir1} = -0.513\log(\text{bet}) + 0.553\log(\text{fat}) - 0.039\log(\text{cal}) - 0.627\log(\text{fiber}) + 0.052\log(\text{ret}) \\ - 68.619\,\text{quet}^{-1} + 0.122\,\text{smoke1} + 0.719\,\text{smoke2} + 0.501\,\text{vit2} + 1.048 * \text{vit3}.$$

In Figure 1, the red-colored line is the OLS fitted line and the blue-colored line stands for LOWESS smooths with smoothing parameter 0.7. For beta-carotene plasma concentration levels, the simple linear regression should be fine, while no relation is expected for retinol plasma concentration levels and predictors.

6. Discussion

In the paper, we propose a permutation test to determine the number of pairs of canonical variates. The permutation test can be considered as non-parametric and it can be a possible alternative of the existing chi-squared test developed by Barttlet (1938, 1939) against non-normality. Various numerical studies, however, show that the latter is robust to non-normality and provide about the same accuracy as the proposed permutation test. In practice, one can enjoy potential advantages to select the number of pairs of canonical variates more reliably by conducting both tests. In addition, we investigate relation between canonical correlation analysis and regression. We show that certain inferences of regression analysis, for example ANOVA F-test, tests for heteroscedasticity, and reduced-rank regression, can be done via canonical correlation analysis. We hope that this paper re-highlights methodological merits of canonical correlation analysis that is often forgotten in high-dimensional data analysis.

References

Bartlett, M. S. (1938). Further aspects of the theory of multiple regression, *Proceedings of the Cambridge Philosophical Society*, **34**, 33–40.

Bartlett, M. S. (1939). A note on tests of significance in multivariate analysis, *Proceedings of the Cambridge Philosophical Society*, **35**, 180–185.

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation, *Econometrika*, **47**, 1287–1294.

Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*, 6th Ed., Pearson Prentice Hall, New Jersey.

Nierenberg, W. D., Stukel, A. T., Baron, A. J., Dain, J. B. and Greenberg, E. R. (1989). The skin cancer prevention study group, determinants of plasma levels of beta-carotene and retinol, *American Journal of Epidemiology*, **130**, 511–521.

Received March 30, 2012; Revised April 20, 2012; Accepted April 23, 2012