Mathematics and Statistics 10(2): 419-430, 2022
DOI: 10.13189/ms.2022.100217 http://www.hrpub.org

Principal Canonical Correlation Analysis with Missing Data in Small Samples

Toru Ogura[1,*], **Shin-ichi Tsukada**[2]

[1]Clinical Research Support Center, Mie University Hospital, Mie, Japan

[2]Department of Education, Meisei University, Tokyo, Japan

[*]Corresponding Author: t-ogura@clin.medic.mie-u.ac.jp

Received December 18, 2021; Revised February 14, 2022; Accepted March 7, 2022

Cite This Paper in the following Citation Styles

***(a)**: [1] Toru Ogura, Shin-ichi Tsukada, "Principal Canonical Correlation Analysis with Missing Data in Small Samples," Mathematics and Statistics, Vol.10, No.2, pp. 419-430, 2022. DOI: 10.13189/ms.2022.100217*

***(b)**: Toru Ogura, Shin-ichi Tsukada, (2022). Principal Canonical Correlation Analysis with Missing Data in Small Samples. Mathematics and Statistics, 10(2), 419-430. DOI: 10.13189/ms.2022.100217*

Abstract Missing data occur in various fields, such as clinical trials and social science. Canonical correlation analysis often used to analyze the correlation between two random vectors, cannot be performed on a dataset with missing data. Canonical correlation coefficients (CCCs) can also be calculated from a covariance matrix. When the covariance matrix can be estimated by excluding (complete-case and available-case analyses) or imputing (multivariate imputation by chained equations, k-nearest neighbor (kNN), and iterative robust model-based imputation) missing data, CCCs are estimated from this covariance matrix. CCCs have bias even with all observation data. Usually, estimated CCCs are even larger than population CCCs when a covariance matrix estimated from a dataset with missing data is used. The purpose is to bring the CCCs estimated from the dataset with missing data close to the population CCCs. The procedure involves three steps. First, principal component analysis is performed on the covariance matrix from the dataset with missing data to obtain the eigenvectors. Second, the covariance matrix is transformed using first to fourth eigenvectors. Finally, the CCCs are calculated from the transformed covariance matrix. CCCs derived using with this procedure are called the principal CCCs (PCCCs), and simulation studies and numerical examples confirmed the effectiveness of the PCCCs estimated from the dataset with missing data. There were many cases in the simulation results where the bias and root-mean-squared error of the PCCC estimated from the missing data based on kNN were the smallest. In the numerical example results, the first PCCC estimated from the missing data based on kNN is close to the first CCC estimated from the dataset comprising all observation data when the correlation between two vectors is low. Therefore, PCCCs based on kNN were recommended.

Keywords Canonical Correlation Analysis, Missing Data, Principal Component Analysis, Small Samples

1 Introduction

A dataset from various fields such as clinical trials and social sciences, often contains missing data [1, 2, 3]. Many researchers have studied the estimation of covariance matrices from missing data, considering that multivariate analyses are based on a covariance matrix [4, 5, 6, 7, 8, 9]. Some methods estimate the covariance matrix by excluding missing data. Examples include complete-case (CC) and available-case (AC) analyses [10]. Other methods, such as multivariate imputation by chained equations (MICE) [11], k-nearest neighbor (kNN) [12], and iterative robust model-based imputation (IRMI) [12, 13], estimate covariance matrix by imputing missing data. An alternative method is to estimate covariance matrices using the maximum likelihood estimation (MLE) of multivariate normal data [14].

Canonical correlation analysis (CCA) examines the relationship between two random vectors [15, 16, 17]. However, it cannot be performed using a dataset with missing data, and researchers have studied various methods to perform CCA, even on datasets with missing data. Van de Velden and Takane [18] discussed two methods for performing CCA from a dataset with missing data to minimize the generalized CCA objective function. They include CC analysis and imputation. Yamada [19] developed CCA in two-step monotone incomplete data. Estimated canonical correlation coefficients (CCCs) have bias

even with all observation data [20]. The bias of the CCCs is large for small samples and small for large samples. Therefore, in small samples, even when the population CCCs are low, the relationship between two vectors can be misunderstood as being high due to the large bias of the CCCs. Additionally, estimated CCCs are much greater than population CCCs because they are based on a covariance matrix estimated from the dataset with missing data. The purpose is to estimate CCCs close to population CCCs from a dataset with missing data. For example, one method used to reduce bias from estimated CCCs is applying fewer variables for the CCA. Principal component analysis (PCA) transforms a given dataset into a new dataset of uncorrelated variables, or principal component (PC) scores [21]. PC scores are ranked in descending order according to the amount of information they contain. Principal CCA (PCCA), which uses CCA between two sets of PC scores, reduces the number of variables while retaining significant information [22, 23, 24, 25]. Furthermore, principal CCCs (PCCCs) can also be calculated directly from a covariance matrix without PC scores. These three steps estimate PCCCs: first, the eigenvectors are obtained by performing PCA on the covariance matrix from the dataset with missing data. Second, the covariance matrix is transformed using a few eigenvectors from the top. Finally, CCCs are calculated from the transformed covariance matrix.

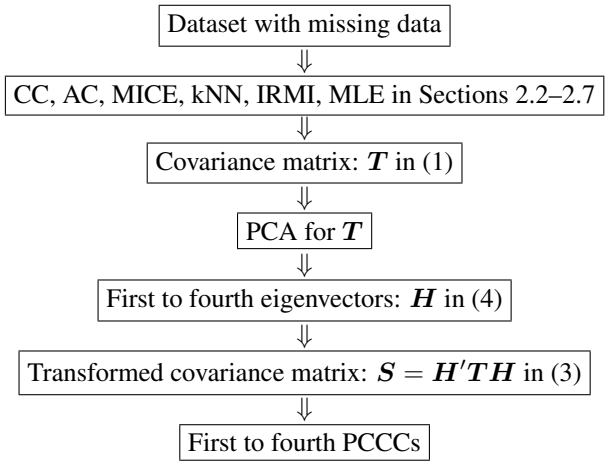
Considering the standard recommendation of a sample size at least ten times the number of variables in multivariate analysis [26], small samples in this manuscript are defined as the case where the sample size is less than ten times the number of each random vector.

In Section 2, the sequence of steps to estimate PCCCs from a dataset with missing data is explained. In Section 3, the effectiveness of PCCCs estimated from a dataset with missing data is verified using simulation studies. Section 4 attempts to calculate CCCs and PCCCs using numerical examples of the dataset in which observed data are replaced with missing data. Finally, in Section 5, conclusions of this study are presented.

2 Derivation of PCCCs from Dataset with Missing Data

2.1 Notation

Let $\boldsymbol{z} = (\boldsymbol{x}', \boldsymbol{y}') = (x_1, \ldots, x_p, y_1, \ldots, y_q)'$ be a $(p+q)$-dimensional random vector with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. Without loss of generality, $p \leq q$ is assumed. When the sample size is n, the size of the dataset is $n(p+q)$. Let $\tau n(p+q)$ be observation data and $(1-\tau)n(p+q)$ be missing data, where $0 \leq \tau \leq 1$. Sections 2.2–2.7 show six methods for estimating the covariance matrix from a dataset with missing data. A flowchart to calculate PCCCs from a dataset with missing data is shown below.



2.2 CC Analysis

CC analysis calculates the covariance matrix by excluding samples with at least one missing data. If there are many missing data, the covariance matrix is estimated from very small samples. CCCs have a bias of $O(1/n)$ [20]. When the very small samples are used, CCCs have a large bias because the denominator of $O(1/n)$ is small. Therefore, even if the population CCCs are small, the estimated first CCC may be obtained close to 1.

2.3 AC Analysis

The used sample is different for each variance and each covariance because the AC analysis uses all available data. The covariance matrix is estimated by combining those variances and covariances. Therefore, this covariance matrix may not have the characteristics of a general covariance matrix. In that case, it is possible that the CCCs are greater than one or cannot be calculated.

2.4 MICE

MICE is a multiple imputation method used in various research areas that employs multiple rather than single imputation values [11]. MICE is implemented in various statistical software. In this manuscript, the covariance matrix is calculated using the mice function in the mice package included in the R software [27].

2.5 kNN

The kNN estimates the value of the missing data using similar or close in the space. Distance computation for defining the kNN is based on an extension of the Gower distance [28], which can handle distance variables of the type binary, categorical, ordered, continuous and semi-continuous [12]. Each missing value is imputed using the mean value of the k-neighbors found in the dataset. In this manuscript, the covariance matrix is calculated using the kNN function in the VIM package included in the R software.

2.6 IRMI

The IRMI mimics the functionality of imputation and variance estimation software (IVEWARE) [29]. However, several improvements concerning the initialized values' stability or the imputed values' robustness have been made [12, 13]. The algorithm does not require at least one fully observed variable. One variable is used as a response variable in each iteration step, and the remaining variables serve as the regressors. In this manuscript, the covariance matrix is calculated using the irmi function in the VIM package included in the R software.

2.7 MLE

MLE finds the mean vector and variance-covariance matrix for multivariate normal data with missing data [14]. In this manuscript, the covariance matrix is calculated using the mlest function in the mvnmle package version 0.1-11 included in the R software.

2.8 Derivation of PCCCs

Using the method in Sections 2.2–2.7, let $\boldsymbol{T}$ be a covariance matrix estimated and it is partitioned into p and q as follows:

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{T}_{xx} & \boldsymbol{T}_{xy} \\ \boldsymbol{T}_{yx} & \boldsymbol{T}_{yy} \end{pmatrix}, \tag{1}$$

where $\boldsymbol{T}_{xx}$, $\boldsymbol{T}_{xy}$, $\boldsymbol{T}_{yx}$, and $\boldsymbol{T}_{yy}$ are the $p \times p$, $p \times q$, $q \times p$, and $q \times q$ matrices, respectively. The CCCs are $\tilde{r}_1 \geq \cdots \geq \tilde{r}_p$ and satisfy

$$| \boldsymbol{T}_{xy}\boldsymbol{T}_{yy}^{-1}\boldsymbol{T}_{yx} - \tilde{r}^2\boldsymbol{T}_{xx} | = 0. \tag{2}$$

Let $\ell_{1x} \geq \cdots \geq \ell_{px}$ be the ordered latent roots of $\boldsymbol{T}_{xx}$, and $\boldsymbol{h}_{1x}, \cdots, \boldsymbol{h}_{px}$ be the corresponding latent vectors. Similarly, let $\ell_{1y} \geq \cdots \geq \ell_{qy}$ be the ordered latent roots of $\boldsymbol{T}_{yy}$, and $\boldsymbol{h}_{1x}, \cdots, \boldsymbol{h}_{px}$ be the corresponding latent vectors. Then, $\boldsymbol{T}$ is transformed using the eigenvectors as follows:

$$\begin{aligned} \boldsymbol{S} = \boldsymbol{H}'\boldsymbol{T}\boldsymbol{H} &= \begin{pmatrix} \boldsymbol{H}_x'\boldsymbol{T}_{xx}\boldsymbol{H}_x & \boldsymbol{H}_x'\boldsymbol{T}_{xy}\boldsymbol{H}_y \\ \boldsymbol{H}_y'\boldsymbol{T}_{yx}\boldsymbol{H}_x & \boldsymbol{H}_y'\boldsymbol{T}_{yy}\boldsymbol{H}_y \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{D}_u & \boldsymbol{S}_{uv} \\ \boldsymbol{S}_{vu} & \boldsymbol{D}_v \end{pmatrix}, \end{aligned} \tag{3}$$

where

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_x & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H}_y \end{pmatrix}, \tag{4}$$

$\boldsymbol{H}_x = (\boldsymbol{h}_{1x}, \cdots, \boldsymbol{h}_{p_1x})$ is a $p \times p_1$ orthogonal matrix, $\boldsymbol{H}_y = (\boldsymbol{h}_{1y}, \cdots, \boldsymbol{h}_{q_1y})$ is a $q \times q_1$ orthogonal matrix, $\boldsymbol{D}_u = \mathrm{diag}(\ell_{1x}, \ldots, \ell_{p_1x})$ is a $p_1 \times p_1$ diagonal matrix, $\boldsymbol{D}_v = \mathrm{diag}(\ell_{1y}, \ldots, \ell_{q_1y})$ is a $q_1 \times q_1$ diagonal matrix, $\boldsymbol{S}_{uv}$ is a $p_1 \times q_1$ matrix, $\boldsymbol{S}_{vu}$ is a $q_1 \times p_1$ matrix, $p_1 \leq p$, and $q_1 \leq q$. As the latent roots descend in the order of the amount of information that they contain, $\boldsymbol{S}$ calculated using the first to p_1-th latent vectors and the first to q_1-th latent vectors have a sufficient amount of information. The PCCCs are $\hat{r}_1 \geq \cdots \geq \hat{r}_{\min(p_1,q_1)}$ and satisfy

$$| \boldsymbol{S}_{uv}\boldsymbol{D}_v^{-1}\boldsymbol{S}_{vu} - \hat{r}^2\boldsymbol{D}_u | = 0 \ \ (p_1 \leq q_1), \tag{5}$$

$$| \boldsymbol{S}_{vu}\boldsymbol{D}_u^{-1}\boldsymbol{S}_{uv} - \hat{r}^2\boldsymbol{D}_v | = 0 \ \ (p_1 > q_1). \tag{6}$$

In this manuscript, the case of $p_1 = q_1 = 4$ is used.

3 Simulation Studies

The performance of PCCCs and CCCs was compared using various population CCCs by excluding (CC and AC analyses) or imputing (MICE, kNN, and IRMI) missing data. In the simulation studies on CCA, the covariance matrix can be set as follows without loss of generality:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{P}' \\ \boldsymbol{P} & \boldsymbol{I}_q \end{pmatrix}, \tag{7}$$

where $\boldsymbol{P} = (\boldsymbol{P}_1, \boldsymbol{0})$, $\boldsymbol{P}_1 = \mathrm{diag}(\rho_1, \ldots, \rho_p)$, and $\rho_1, \ldots, \rho_p$ are the population CCCs [19]. This setting is optimal for CCA, but the same may not be true for PCCA. When PCCA is more effective than CCA under this setting, its effectiveness is clarified. The simulation is set to 16 cases, as shown in Table 1, and it is performed for all combinations of the 16 cases, $\tau = 0.9, 0.95$, and sample size $n = 30, 40$. Table 2 shows the number of observation and missing data in each setting.

In the simulation studies, $p_1 = q_1 = 4$ was set to retain sufficient information from the original dataset. The simulation studies were performed using the R software version 3.4.0. The simulation was run on an old R software version because MLE would not run on the new version. In this manuscript, the missing data mechanism was missing completely at random (MCAR) [10]. The simulation was repeated 10,000 times and evaluated using the bias and root-mean-squared error (RMSE $= \sqrt{\text{variance} + \text{bias}^2}$). The bias was calculated as the difference between population CCC and estimated CCC, or population CCC and estimated PCCC. The simulation study was conducted using the following procedure:

1. Generate a random vector $\boldsymbol{z}$ of $(p+q)$ components with a sample size of n, where $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$.
2. Replace $(1-\tau)n(p+q)$ of $n(p+q)$ observation data in $\boldsymbol{z}$ with missing data by MCAR.
3. Estimate the covariance matrix $\boldsymbol{T}$ by excluding or imputing missing data.
4. Convert $\boldsymbol{T}$ to $\boldsymbol{S}$ using each latent vector.
5. Estimate CCCs from $\boldsymbol{T}$ and PCCCs from $\boldsymbol{S}$.
6. Independently, repeat steps 1–5, 10,000 times.
7. Calculate the bias and RMSE of CCCs and PCCCs.

Tables 3–6 show the bias and RMSE. CCCs and PCCCs estimated from the missing data based on CC and AC analyses and MLE were sometimes incalculable or greater than or equal to 1. Thus, CC and AC analyses and MLE may be unsuitable for CCA in small samples. CCCs and PCCCs estimated from the missing data based on MICE, kNN, and IRMI ranged from 0 to 1 in all cases. The bias and RMSE were calculated, excluding the incalculable ones. As the CC analysis excluded many samples, CCCs and PCCCs could not be calculated in many settings, and they were close to 1 even if they could be calculated. Therefore, the simulation results were discussed, except for the CC analysis. In Cases 1, 5, 9, and 13, where the

Table 1. Population CCCs in Cases 1–16.

	(p, q)	$(\rho_1, \ldots, \rho_p)$
Case 1	(6,6)	$(0.4, 0.4/2, 0.4/2^2, 0.4/2^3, 0.4/2^4, 0.4/2^5)$
Case 2	(6,6)	$(0.5, 0.5/2, 0.5/2^2, 0.5/2^3, 0.5/2^4, 0.5/2^5)$
Case 3	(6,6)	$(0.6, 0.6/2, 0.6/2^2, 0.6/2^3, 0.6/2^4, 0.6/2^5)$
Case 4	(6,6)	$(0.7, 0.7/2, 0.7/2^2, 0.7/2^3, 0.7/2^4, 0.7/2^5)$
Case 5	(7,7)	$(0.4, 0.4/2, 0.4/2^2, 0.4/2^3, 0.4/2^4, 0.4/2^5, 0.4/2^6)$
Case 6	(7,7)	$(0.5, 0.5/2, 0.5/2^2, 0.5/2^3, 0.5/2^4, 0.5/2^5, 0.5/2^6)$
Case 7	(7,7)	$(0.6, 0.6/2, 0.6/2^2, 0.6/2^3, 0.6/2^4, 0.6/2^5, 0.6/2^6)$
Case 8	(7,7)	$(0.7, 0.7/2, 0.7/2^2, 0.7/2^3, 0.7/2^4, 0.7/2^5, 0.7/2^6)$
Case 9	(8,8)	$(0.4, 0.4/2, 0.4/2^2, 0.4/2^3, 0.4/2^4, 0.4/2^5, 0.4/2^6, 0.4/2^7)$
Case 10	(8,8)	$(0.5, 0.5/2, 0.5/2^2, 0.5/2^3, 0.5/2^4, 0.5/2^5, 0.5/2^6, 0.5/2^7)$
Case 11	(8,8)	$(0.6, 0.6/2, 0.6/2^2, 0.6/2^3, 0.6/2^4, 0.6/2^5, 0.6/2^6, 0.6/2^7)$
Case 12	(8,8)	$(0.7, 0.7/2, 0.7/2^2, 0.7/2^3, 0.7/2^4, 0.7/2^5, 0.7/2^6, 0.7/2^7)$
Case 13	(9,9)	$(0.4, 0.4/2, 0.4/2^2, 0.4/2^3, 0.4/2^4, 0.4/2^5, 0.4/2^6, 0.4/2^7, 0.4/2^8)$
Case 14	(9,9)	$(0.5, 0.5/2, 0.5/2^2, 0.5/2^3, 0.5/2^4, 0.5/2^5, 0.5/2^6, 0.5/2^7, 0.5/2^8)$
Case 15	(9,9)	$(0.6, 0.6/2, 0.6/2^2, 0.6/2^3, 0.6/2^4, 0.6/2^5, 0.6/2^6, 0.6/2^7, 0.6/2^8)$
Case 16	(9,9)	$(0.7, 0.7/2, 0.7/2^2, 0.7/2^3, 0.7/2^4, 0.7/2^5, 0.7/2^6, 0.7/2^7, 0.7/2^8)$

Table 2. The number of observation and missing data in simulation settings.

				Number of		
n	p	q	τ	observation	missing	Total
30	6	6	0.90	324	36	360
30	6	6	0.95	342	18	360
30	7	7	0.90	378	42	420
30	7	7	0.95	399	21	420
30	8	8	0.90	432	48	480
30	8	8	0.95	456	24	480
30	9	9	0.90	486	54	540
30	9	9	0.95	513	27	540
40	6	6	0.90	432	48	480
40	6	6	0.95	456	24	480
40	7	7	0.90	504	56	560
40	7	7	0.95	532	28	560
40	8	8	0.90	576	64	640
40	8	8	0.95	608	32	640
40	9	9	0.90	648	72	720
40	9	9	0.95	684	36	720

first population CCC was set to 0.4, the bias and RMSE of the CCCs estimated from the missing data based on all methods were large, but those of PCCCs were nearly half of those of CCCs. In these cases, the correlation between the two random vectors was low, but it might be misinterpreted as high because of the large bias of the CCCs. In contrast, the interpretation using the PCCCs might avoid major mistakes regarding the correlation between the two random vectors. As the population CCCs setting increased, the bias and RMSE of the CCCs and PCCCs decreased. Those of the PCCCs were less than half of those of the CCCs in many settings. The bias and RMSE of the second CCC were also large, whereas that of the second PCCC were often less than half of those of the second CCC. Comparing the simulation results between the methods indicated many cases wherein the bias and RMSE of the PCCC estimated from the missing data based on kNN were the smallest. Therefore, the PCCC based on kNN was recommended.

4 Numerical Examples

The attempt to estimate CCCs and PCCCs was presented using the dataset from a nutrition study [30]. The dataset consists of gene data with 120 variables and lipid data with 21 variables with a sample size of 40 in the CCA package of R software [31]. This dataset is all observation data. Datasets used in Examples 1 and 2 were chosen from 6 variables from gene data with 120 variables and 6 variables from lipid data with 21 variables. The variables are chosen to have low CCCs in Example 1 and high CCCs in Example 2. The size of these two datasets is $40 \times (6 + 6)$. Among the 480 observation data, $40 \times (6 + 6) \times (1 - 0.9) = 48$ is replaced with missing data by MCAR ($\tau = 0.90$). Therefore, the dataset with missing data comprises $40 \times (6 + 6) \times 0.9 = 432$ observation and 48 missing data. CCCs and PCCCs ($p_1 = q_1 = 4$) are estimated from the missing data based on CC and AC analyses, MICE, kNN, and IRMI. It is evaluated by the closeness of CCCs and PCCCs estimated from the dataset with missing data and CCCs estimated from the dataset comprising all observation data.

The recommended sample size for multivariate analysis is at least ten times the number of variables [26]. Therefore, the sample size in CCA may be recommended at least ten times the number of p and q. In numerical examples, because the sample size of at least 60 is recommended for $p = q = 6$, $n = 40$ can be considered small samples.

Table 3. Comparison of the bias and RMSE of CCCs and PCCCs from MCAR based on six methods given $n = 30$, $\tau = 0.90$ for Cases 1–16 ($p = q = 6, 7, 8, 9$, $p_1 = q_1 = 4$).

			1st						2nd					
Case			CC	AC	MICE	kNN	IRMI	MLE	CC	AC	MICE	kNN	IRMI	MLE
1	CCC	Bias	0.600	0.469	0.438	0.367	0.478	0.534	0.798	0.504	0.491	0.433	0.527	0.547
		RMSE	0.600	0.477	0.443	0.373	0.482	0.539	0.798	0.510	0.497	0.438	0.532	0.554
	PCCC	Bias	0.572	0.270	0.260	0.233	0.286	0.297	0.662	0.260	0.256	0.240	0.277	0.278
		RMSE	0.574	0.288	0.278	0.249	0.302	0.315	0.675	0.278	0.274	0.258	0.294	0.298
2	CCC	Bias	0.500	0.385	0.348	0.281	0.385	0.439	0.748	0.470	0.453	0.396	0.488	0.509
		RMSE	0.500	0.394	0.354	0.288	0.390	0.444	0.748	0.477	0.459	0.403	0.494	0.516
	PCCC	Bias	0.474	0.193	0.178	0.152	0.201	0.212	0.616	0.228	0.222	0.206	0.239	0.242
		RMSE	0.476	0.218	0.202	0.177	0.223	0.237	0.630	0.249	0.243	0.227	0.260	0.265
3	CCC	Bias	0.400	0.304	0.264	0.201	0.297	0.345	0.698	0.439	0.419	0.363	0.454	0.472
		RMSE	0.400	0.315	0.270	0.209	0.302	0.350	0.698	0.447	0.426	0.369	0.459	0.479
	PCCC	Bias	0.374	0.120	0.102	0.077	0.125	0.133	0.571	0.196	0.187	0.173	0.207	0.210
		RMSE	0.377	0.158	0.140	0.119	0.157	0.169	0.585	0.222	0.212	0.198	0.231	0.237
4	CCC	Bias	0.300	0.228	0.183	0.125	0.211	0.255	0.648	0.410	0.388	0.330	0.421	0.442
		RMSE	0.300	0.239	0.190	0.137	0.216	0.261	0.648	0.419	0.395	0.337	0.427	0.450
	PCCC	Bias	0.277	0.053	0.032	0.006	0.051	0.058	0.529	0.167	0.157	0.140	0.176	0.179
		RMSE	0.280	0.116	0.101	0.092	0.106	0.118	0.543	0.198	0.189	0.172	0.205	0.214
5	CCC	Bias	0.600	0.548	0.495	0.415	0.529	0.592	0.800	0.594	0.570	0.502	0.605	0.662
		RMSE	0.600	0.559	0.498	0.419	0.531	0.592	0.800	0.599	0.574	0.506	0.608	0.667
	PCCC	Bias	0.588	0.256	0.244	0.224	0.266	0.287	0.711	0.250	0.245	0.233	0.261	0.274
		RMSE	0.589	0.275	0.262	0.241	0.283	0.308	0.721	0.269	0.263	0.251	0.279	0.296
6	CCC	Bias	0.500	0.457	0.401	0.325	0.433	0.492	0.750	0.555	0.529	0.461	0.562	0.614
		RMSE	0.500	0.468	0.404	0.329	0.435	0.492	0.750	0.560	0.533	0.465	0.565	0.620
	PCCC	Bias	0.489	0.178	0.163	0.143	0.182	0.201	0.665	0.217	0.209	0.199	0.224	0.234
		RMSE	0.490	0.204	0.189	0.169	0.205	0.230	0.674	0.239	0.231	0.219	0.245	0.260
7	CCC	Bias	0.400	0.370	0.308	0.237	0.338	0.392	0.700	0.520	0.491	0.424	0.522	0.568
		RMSE	0.400	0.389	0.312	0.242	0.340	0.393	0.700	0.525	0.495	0.428	0.525	0.574
	PCCC	Bias	0.390	0.101	0.082	0.063	0.099	0.114	0.619	0.181	0.172	0.161	0.188	0.195
		RMSE	0.391	0.144	0.127	0.111	0.138	0.160	0.628	0.208	0.199	0.188	0.213	0.229
8	CCC	Bias	0.300	0.286	0.220	0.155	0.245	0.293	0.650	0.488	0.455	0.388	0.484	0.524
		RMSE	0.300	0.300	0.224	0.162	0.248	0.294	0.650	0.493	0.459	0.393	0.488	0.530
	PCCC	Bias	0.290	0.030	0.007	-0.011	0.023	0.036	0.571	0.148	0.137	0.127	0.153	0.165
		RMSE	0.291	0.108	0.098	0.093	0.099	0.121	0.580	0.182	0.171	0.161	0.185	0.207
9	CCC	Bias	0.600	0.630	0.538	0.456	0.564	0.598	0.800	0.678	0.637	0.562	0.665	0.713
		RMSE	0.600	0.650	0.539	0.459	0.565	0.598	0.800	0.681	0.640	0.564	0.667	0.717
	PCCC	Bias	0.591	0.248	0.234	0.219	0.251	0.286	0.679	0.246	0.236	0.232	0.251	0.273
		RMSE	0.591	0.267	0.253	0.237	0.269	0.315	0.687	0.265	0.255	0.249	0.269	0.302
10	CCC	Bias	0.500	0.535	0.441	0.363	0.466	0.498	0.750	0.636	0.594	0.519	0.620	0.664
		RMSE	0.500	0.549	0.443	0.365	0.466	0.499	0.750	0.640	0.597	0.522	0.622	0.669
	PCCC	Bias	0.490	0.164	0.147	0.134	0.163	0.196	0.632	0.205	0.196	0.190	0.210	0.232
		RMSE	0.491	0.193	0.176	0.162	0.190	0.237	0.640	0.227	0.218	0.211	0.231	0.268
11	CCC	Bias	0.400	0.445	0.345	0.272	0.367	0.398	0.700	0.598	0.551	0.479	0.577	0.615
		RMSE	0.400	0.473	0.347	0.275	0.368	0.398	0.700	0.601	0.553	0.482	0.579	0.620
	PCCC	Bias	0.391	0.087	0.065	0.054	0.080	0.110	0.586	0.170	0.159	0.154	0.172	0.193
		RMSE	0.392	0.134	0.116	0.106	0.124	0.171	0.594	0.198	0.186	0.181	0.199	0.237
12	CCC	Bias	0.300	0.354	0.251	0.184	0.271	0.299	0.650	0.560	0.511	0.440	0.536	0.568
		RMSE	0.300	0.408	0.253	0.188	0.272	0.299	0.650	0.564	0.514	0.443	0.538	0.573
	PCCC	Bias	0.291	0.012	-0.015	-0.024	0.002	0.029	0.539	0.135	0.122	0.118	0.137	0.161
		RMSE	0.292	0.105	0.099	0.097	0.097	0.140	0.547	0.171	0.158	0.153	0.171	0.217
13	CCC	Bias	0.600	0.732	0.568	0.493	0.585	0.600	0.800	0.758	0.691	0.614	0.711	0.742
		RMSE	0.600	0.780	0.569	0.494	0.585	0.600	0.800	0.760	0.693	0.616	0.712	0.745
	PCCC	Bias	-0.400	0.242	0.223	0.215	0.238	0.310	-0.200	0.240	0.229	0.228	0.242	0.297
		RMSE	0.400	0.261	0.242	0.233	0.256	0.349	0.200	0.259	0.247	0.245	0.260	0.339
14	CCC	Bias	0.500	0.636	0.470	0.397	0.486	0.500	0.750	0.714	0.645	0.570	0.663	0.694
		RMSE	0.500	0.713	0.471	0.399	0.486	0.500	0.750	0.716	0.646	0.571	0.664	0.697
	PCCC	Bias	0.463	0.157	0.136	0.130	0.150	0.219	0.564	0.200	0.187	0.188	0.201	0.255
		RMSE	0.467	0.186	0.166	0.158	0.178	0.271	0.575	0.222	0.210	0.209	0.222	0.304
15	CCC	Bias	0.400	0.535	0.371	0.303	0.386	0.400	0.700	0.670	0.600	0.526	0.617	0.648
		RMSE	0.400	0.582	0.372	0.305	0.387	0.400	0.700	0.673	0.601	0.528	0.618	0.651
	PCCC	Bias	0.366	0.075	0.048	0.045	0.064	0.130	0.520	0.164	0.148	0.151	0.162	0.216
		RMSE	0.368	0.126	0.108	0.102	0.114	0.204	0.531	0.192	0.177	0.178	0.189	0.273
16	CCC	Bias	0.300	0.441	0.274	0.211	0.287	0.300	0.650	0.629	0.555	0.485	0.572	0.605
		RMSE	0.300	0.544	0.274	0.213	0.287	0.300	0.650	0.632	0.556	0.487	0.573	0.607
	PCCC	Bias	0.277	-0.004	-0.033	-0.036	-0.018	0.046	0.502	0.126	0.107	0.110	0.122	0.180
		RMSE	0.279	0.104	0.104	0.100	0.098	0.165	0.509	0.162	0.146	0.146	0.157	0.249

Table 4. Comparison of the bias and RMSE of CCCs and PCCCs from MCAR based on six methods given $n = 30$, $\tau = 0.95$ for Cases 1–16 ($p = q = 6, 7, 8, 9$, $p_1 = q_1 = 4$).

			1st						2nd					
Case			CC	AC	MICE	kNN	IRMI	MLE	CC	AC	MICE	kNN	IRMI	MLE
1	CCC	Bias	0.540	0.406	0.397	0.363	0.414	0.425	0.639	0.459	0.454	0.428	0.471	0.474
		RMSE	0.542	0.413	0.403	0.369	0.419	0.431	0.643	0.465	0.460	0.434	0.477	0.480
	PCCC	Bias	0.393	0.242	0.237	0.226	0.250	0.253	0.385	0.243	0.241	0.235	0.251	0.251
		RMSE	0.404	0.260	0.255	0.243	0.268	0.270	0.401	0.261	0.259	0.252	0.268	0.268
2	CCC	Bias	0.443	0.324	0.313	0.280	0.328	0.338	0.597	0.426	0.421	0.394	0.436	0.441
		RMSE	0.445	0.332	0.319	0.287	0.334	0.344	0.602	0.433	0.427	0.400	0.443	0.447
	PCCC	Bias	0.306	0.166	0.161	0.148	0.172	0.175	0.351	0.209	0.206	0.200	0.216	0.217
		RMSE	0.318	0.192	0.186	0.174	0.196	0.199	0.369	0.231	0.227	0.222	0.238	0.238
3	CCC	Bias	0.349	0.247	0.232	0.203	0.246	0.255	0.556	0.396	0.389	0.362	0.403	0.409
		RMSE	0.351	0.255	0.240	0.211	0.252	0.262	0.560	0.403	0.395	0.369	0.410	0.416
	PCCC	Bias	0.221	0.096	0.089	0.078	0.100	0.103	0.318	0.178	0.175	0.170	0.186	0.187
		RMSE	0.237	0.137	0.130	0.120	0.136	0.140	0.338	0.205	0.201	0.196	0.211	0.213
4	CCC	Bias	0.255	0.177	0.159	0.133	0.170	0.179	0.516	0.366	0.358	0.331	0.373	0.379
		RMSE	0.257	0.187	0.167	0.143	0.177	0.187	0.520	0.374	0.366	0.339	0.380	0.386
	PCCC	Bias	0.139	0.032	0.022	0.012	0.032	0.035	0.285	0.149	0.144	0.139	0.155	0.156
		RMSE	0.161	0.103	0.097	0.092	0.099	0.101	0.306	0.181	0.177	0.171	0.186	0.188
5	CCC	Bias	0.592	0.467	0.452	0.412	0.468	0.498	0.759	0.538	0.531	0.498	0.547	0.562
		RMSE	0.593	0.471	0.456	0.416	0.471	0.501	0.761	0.542	0.535	0.502	0.551	0.566
	PCCC	Bias	0.424	0.233	0.229	0.220	0.240	0.247	0.419	0.234	0.233	0.229	0.242	0.244
		RMSE	0.434	0.251	0.247	0.238	0.258	0.265	0.436	0.253	0.251	0.246	0.260	0.263
6	CCC	Bias	0.493	0.376	0.360	0.323	0.375	0.402	0.710	0.499	0.490	0.458	0.507	0.520
		RMSE	0.493	0.381	0.364	0.327	0.379	0.406	0.712	0.504	0.494	0.462	0.511	0.524
	PCCC	Bias	0.332	0.152	0.147	0.139	0.158	0.164	0.380	0.199	0.196	0.192	0.205	0.207
		RMSE	0.344	0.181	0.175	0.166	0.184	0.191	0.398	0.221	0.218	0.214	0.227	0.230
7	CCC	Bias	0.393	0.292	0.273	0.239	0.287	0.312	0.663	0.466	0.456	0.424	0.471	0.485
		RMSE	0.394	0.298	0.278	0.244	0.291	0.317	0.665	0.471	0.460	0.429	0.475	0.489
	PCCC	Bias	0.245	0.081	0.072	0.065	0.083	0.087	0.345	0.166	0.162	0.159	0.171	0.174
		RMSE	0.259	0.127	0.120	0.113	0.126	0.131	0.364	0.193	0.189	0.186	0.198	0.202
8	CCC	Bias	0.294	0.213	0.191	0.161	0.202	0.225	0.616	0.433	0.422	0.390	0.437	0.451
		RMSE	0.294	0.220	0.196	0.167	0.207	0.230	0.617	0.438	0.427	0.395	0.441	0.455
	PCCC	Bias	0.159	0.009	-0.001	-0.008	0.008	0.013	0.311	0.132	0.128	0.124	0.136	0.140
		RMSE	0.177	0.099	0.097	0.094	0.096	0.100	0.332	0.167	0.162	0.158	0.170	0.175
9	CCC	Bias	0.600	0.521	0.497	0.454	0.511	0.562	0.799	0.608	0.596	0.558	0.612	0.641
		RMSE	0.600	0.524	0.499	0.456	0.513	0.563	0.799	0.611	0.598	0.561	0.615	0.643
	PCCC	Bias	0.460	0.224	0.219	0.214	0.231	0.243	0.461	0.229	0.227	0.225	0.236	0.242
		RMSE	0.468	0.244	0.238	0.232	0.249	0.264	0.477	0.247	0.245	0.243	0.253	0.262
10	CCC	Bias	0.500	0.427	0.403	0.362	0.416	0.464	0.749	0.567	0.553	0.516	0.569	0.597
		RMSE	0.500	0.431	0.405	0.364	0.418	0.465	0.749	0.570	0.556	0.519	0.571	0.600
	PCCC	Bias	0.365	0.143	0.134	0.131	0.146	0.156	0.421	0.192	0.187	0.187	0.196	0.202
		RMSE	0.375	0.173	0.165	0.160	0.174	0.188	0.439	0.214	0.209	0.209	0.218	0.227
11	CCC	Bias	0.400	0.337	0.311	0.273	0.322	0.366	0.699	0.529	0.514	0.478	0.528	0.554
		RMSE	0.400	0.342	0.313	0.276	0.324	0.368	0.699	0.532	0.516	0.480	0.531	0.557
	PCCC	Bias	0.273	0.065	0.056	0.051	0.066	0.076	0.378	0.155	0.149	0.149	0.159	0.165
		RMSE	0.286	0.117	0.110	0.105	0.115	0.129	0.398	0.183	0.177	0.177	0.186	0.196
12	CCC	Bias	0.300	0.251	0.221	0.187	0.231	0.270	0.649	0.492	0.476	0.440	0.490	0.515
		RMSE	0.300	0.256	0.224	0.191	0.233	0.272	0.649	0.495	0.479	0.442	0.493	0.519
	PCCC	Bias	0.184	-0.008	-0.019	-0.023	-0.011	-0.003	0.342	0.121	0.114	0.114	0.122	0.131
		RMSE	0.200	0.101	0.100	0.098	0.097	0.108	0.363	0.156	0.150	0.149	0.157	0.171
13	CCC	Bias	0.600	0.573	0.536	0.491	0.546	0.592	0.800	0.672	0.652	0.611	0.666	0.706
		RMSE	0.600	0.576	0.537	0.493	0.546	0.592	0.800	0.675	0.654	0.613	0.667	0.708
	PCCC	Bias	0.491	0.219	0.212	0.209	0.222	0.256	0.500	0.225	0.220	0.222	0.229	0.252
		RMSE	0.498	0.239	0.232	0.227	0.241	0.285	0.516	0.243	0.238	0.239	0.247	0.279
14	CCC	Bias	0.500	0.478	0.439	0.397	0.448	0.492	0.750	0.630	0.608	0.568	0.621	0.658
		RMSE	0.500	0.481	0.440	0.398	0.449	0.492	0.750	0.632	0.610	0.570	0.622	0.661
	PCCC	Bias	0.396	0.136	0.128	0.125	0.136	0.169	0.458	0.187	0.181	0.183	0.190	0.213
		RMSE	0.404	0.166	0.158	0.155	0.165	0.210	0.475	0.209	0.204	0.205	0.212	0.246
15	CCC	Bias	0.400	0.382	0.343	0.303	0.351	0.392	0.700	0.586	0.565	0.526	0.576	0.610
		RMSE	0.400	0.385	0.344	0.305	0.352	0.392	0.700	0.588	0.566	0.527	0.578	0.613
	PCCC	Bias	0.300	0.055	0.044	0.043	0.053	0.083	0.415	0.149	0.143	0.144	0.151	0.174
		RMSE	0.310	0.112	0.106	0.102	0.108	0.152	0.434	0.177	0.172	0.172	0.179	0.214
16	CCC	Bias	0.300	0.290	0.248	0.212	0.256	0.293	0.650	0.547	0.523	0.485	0.534	0.565
		RMSE	0.300	0.294	0.250	0.214	0.257	0.293	0.650	0.549	0.525	0.487	0.536	0.567
	PCCC	Bias	0.207	-0.021	-0.034	-0.034	-0.025	0.004	0.375	0.111	0.104	0.107	0.113	0.139
		RMSE	0.219	0.102	0.102	0.101	0.099	0.128	0.395	0.148	0.142	0.143	0.150	0.190

Table 5. Comparison of the bias and RMSE of CCCs and PCCCs from MCAR based on six methods given $n = 40$, $\tau = 0.90$ for Cases 1–16 ($p = q = 6, 7, 8, 9$, $p_1 = q_1 = 4$).

			1st						2nd					
Case			CC	AC	MICE	kNN	IRMI	MLE	CC	AC	MICE	kNN	IRMI	MLE
1	CCC	Bias	0.594	0.365	0.359	0.292	0.399	0.396	0.765	0.410	0.412	0.359	0.444	0.434
		RMSE	0.594	0.373	0.366	0.300	0.404	0.404	0.767	0.417	0.418	0.365	0.451	0.441
	PCCC	Bias	0.505	0.196	0.195	0.166	0.216	0.211	0.528	0.203	0.204	0.187	0.221	0.214
		RMSE	0.512	0.217	0.217	0.187	0.236	0.232	0.546	0.222	0.223	0.205	0.239	0.233
2	CCC	Bias	0.494	0.288	0.278	0.214	0.313	0.311	0.717	0.379	0.377	0.326	0.410	0.401
		RMSE	0.494	0.298	0.286	0.223	0.320	0.320	0.719	0.386	0.384	0.332	0.417	0.408
	PCCC	Bias	0.410	0.124	0.119	0.092	0.142	0.136	0.484	0.169	0.169	0.152	0.187	0.181
		RMSE	0.417	0.158	0.153	0.128	0.171	0.166	0.503	0.192	0.192	0.175	0.209	0.204
3	CCC	Bias	0.394	0.221	0.204	0.143	0.235	0.236	0.668	0.351	0.347	0.294	0.380	0.372
		RMSE	0.395	0.233	0.213	0.157	0.243	0.246	0.670	0.359	0.354	0.302	0.387	0.380
	PCCC	Bias	0.316	0.061	0.051	0.024	0.071	0.070	0.447	0.140	0.138	0.120	0.156	0.150
		RMSE	0.325	0.118	0.108	0.094	0.119	0.120	0.467	0.170	0.168	0.151	0.183	0.180
4	CCC	Bias	0.295	0.159	0.135	0.079	0.161	0.164	0.620	0.325	0.320	0.265	0.351	0.346
		RMSE	0.295	0.174	0.147	0.098	0.170	0.175	0.622	0.335	0.328	0.274	0.359	0.355
	PCCC	Bias	0.224	0.001	-0.012	-0.040	0.006	0.005	0.407	0.110	0.108	0.089	0.125	0.120
		RMSE	0.234	0.104	0.099	0.102	0.096	0.098	0.428	0.149	0.147	0.130	0.160	0.157
5	CCC	Bias	0.594	0.426	0.415	0.337	0.453	0.482	0.800	0.486	0.485	0.420	0.519	0.524
		RMSE	0.599	0.432	0.419	0.342	0.457	0.489	0.800	0.491	0.489	0.424	0.523	0.528
	PCCC	Bias	0.554	0.185	0.182	0.158	0.203	0.202	0.616	0.194	0.196	0.180	0.210	0.207
		RMSE	0.557	0.208	0.205	0.180	0.224	0.224	0.632	0.213	0.214	0.198	0.228	0.226
6	CCC	Bias	0.500	0.341	0.326	0.251	0.362	0.390	0.750	0.450	0.446	0.382	0.480	0.485
		RMSE	0.500	0.348	0.331	0.257	0.367	0.397	0.750	0.455	0.451	0.386	0.484	0.490
	PCCC	Bias	0.455	0.108	0.104	0.079	0.121	0.121	0.570	0.159	0.158	0.144	0.173	0.169
		RMSE	0.460	0.147	0.140	0.120	0.155	0.156	0.587	0.182	0.182	0.168	0.196	0.193
7	CCC	Bias	0.400	0.264	0.244	0.173	0.276	0.303	0.699	0.419	0.412	0.349	0.444	0.450
		RMSE	0.400	0.273	0.250	0.182	0.281	0.311	0.700	0.425	0.417	0.354	0.449	0.456
	PCCC	Bias	0.359	0.037	0.028	0.004	0.044	0.044	0.531	0.124	0.121	0.108	0.136	0.133
		RMSE	0.364	0.108	0.101	0.093	0.107	0.111	0.547	0.155	0.153	0.140	0.166	0.164
8	CCC	Bias	0.300	0.196	0.167	0.104	0.194	0.220	0.649	0.390	0.381	0.316	0.414	0.421
		RMSE	0.300	0.206	0.174	0.116	0.200	0.228	0.649	0.397	0.386	0.321	0.419	0.427
	PCCC	Bias	0.262	-0.027	-0.040	-0.062	-0.024	-0.023	0.489	0.094	0.089	0.077	0.106	0.104
		RMSE	0.268	0.109	0.108	0.114	0.101	0.106	0.506	0.135	0.130	0.119	0.143	0.144
9	CCC	Bias	0.600	0.482	0.463	0.375	0.502	0.566	0.800	0.555	0.549	0.473	0.583	0.618
		RMSE	0.600	0.487	0.466	0.379	0.504	0.568	0.800	0.558	0.552	0.476	0.586	0.622
	PCCC	Bias	0.573	0.176	0.172	0.152	0.189	0.194	0.637	0.186	0.187	0.175	0.199	0.198
		RMSE	0.575	0.200	0.196	0.175	0.211	0.218	0.649	0.205	0.205	0.194	0.217	0.218
10	CCC	Bias	0.500	0.395	0.371	0.287	0.408	0.469	0.750	0.517	0.508	0.434	0.542	0.576
		RMSE	0.500	0.400	0.375	0.291	0.410	0.472	0.750	0.521	0.512	0.438	0.545	0.581
	PCCC	Bias	0.473	0.096	0.089	0.070	0.104	0.108	0.592	0.151	0.147	0.138	0.160	0.159
		RMSE	0.475	0.136	0.128	0.113	0.140	0.147	0.604	0.175	0.172	0.162	0.184	0.184
11	CCC	Bias	0.400	0.311	0.282	0.203	0.315	0.371	0.700	0.482	0.471	0.397	0.502	0.535
		RMSE	0.400	0.317	0.286	0.209	0.318	0.374	0.700	0.486	0.474	0.400	0.505	0.539
	PCCC	Bias	0.374	0.021	0.011	-0.007	0.024	0.026	0.548	0.113	0.109	0.100	0.123	0.121
		RMSE	0.377	0.102	0.096	0.092	0.099	0.106	0.560	0.146	0.142	0.132	0.153	0.154
12	CCC	Bias	0.300	0.234	0.198	0.127	0.225	0.275	0.650	0.450	0.436	0.361	0.467	0.495
		RMSE	0.300	0.242	0.203	0.135	0.229	0.278	0.650	0.455	0.439	0.365	0.470	0.500
	PCCC	Bias	0.276	-0.050	-0.064	-0.081	-0.050	-0.050	0.505	0.077	0.074	0.063	0.086	0.084
		RMSE	0.280	0.117	0.119	0.126	0.112	0.120	0.517	0.122	0.119	0.110	0.129	0.133
13	CCC	Bias	0.600	0.540	0.506	0.412	0.544	0.595	0.800	0.620	0.606	0.522	0.640	0.714
		RMSE	0.600	0.544	0.508	0.415	0.546	0.596	0.800	0.623	0.608	0.524	0.642	0.718
	PCCC	Bias	0.564	0.168	0.165	0.147	0.178	0.188	0.608	0.181	0.181	0.172	0.192	0.193
		RMSE	0.566	0.193	0.188	0.172	0.200	0.215	0.618	0.201	0.199	0.191	0.210	0.215
14	CCC	Bias	0.500	0.447	0.411	0.319	0.446	0.495	0.750	0.578	0.562	0.479	0.595	0.665
		RMSE	0.500	0.452	0.413	0.322	0.448	0.495	0.750	0.581	0.565	0.482	0.597	0.669
	PCCC	Bias	0.460	0.084	0.076	0.062	0.091	0.097	0.558	0.143	0.139	0.131	0.150	0.150
		RMSE	0.462	0.128	0.120	0.108	0.130	0.145	0.569	0.167	0.164	0.156	0.174	0.178
15	CCC	Bias	0.400	0.358	0.317	0.232	0.350	0.396	0.700	0.539	0.521	0.439	0.553	0.613
		RMSE	0.400	0.363	0.320	0.236	0.352	0.396	0.700	0.542	0.523	0.442	0.555	0.617
	PCCC	Bias	0.364	0.008	-0.004	-0.016	0.009	0.012	0.518	0.107	0.100	0.095	0.111	0.108
		RMSE	0.367	0.100	0.097	0.093	0.097	0.110	0.530	0.140	0.134	0.128	0.143	0.148
16	CCC	Bias	0.300	0.274	0.227	0.150	0.255	0.295	0.650	0.507	0.484	0.403	0.514	0.560
		RMSE	0.300	0.281	0.230	0.156	0.257	0.295	0.650	0.510	0.486	0.406	0.516	0.564
	PCCC	Bias	0.267	-0.067	-0.083	-0.094	-0.070	-0.071	0.475	0.069	0.063	0.057	0.073	0.071
		RMSE	0.270	0.123	0.128	0.133	0.120	0.136	0.488	0.116	0.110	0.105	0.117	0.128

Table 6. Comparison of the bias and RMSE of CCCs and PCCCs from MCAR based on six methods given $n = 40$, $\tau = 0.95$ for Cases 1–16 ($p = q = 6, 7, 8, 9$, $p_1 = q_1 = 4$).

			1st						2nd					
Case			CC	AC	MICE	kNN	IRMI	MLE	CC	AC	MICE	kNN	IRMI	MLE
1	CCC	Bias	0.454	0.319	0.319	0.288	0.337	0.328	0.526	0.374	0.376	0.353	0.392	0.383
		RMSE	0.458	0.327	0.326	0.296	0.344	0.335	0.532	0.380	0.383	0.359	0.398	0.390
	PCCC	Bias	0.303	0.174	0.174	0.162	0.187	0.180	0.301	0.186	0.187	0.181	0.196	0.192
		RMSE	0.318	0.196	0.196	0.183	0.208	0.201	0.318	0.204	0.206	0.198	0.214	0.210
2	CCC	Bias	0.364	0.245	0.242	0.214	0.258	0.251	0.488	0.343	0.344	0.321	0.361	0.352
		RMSE	0.368	0.255	0.251	0.223	0.267	0.260	0.494	0.351	0.351	0.328	0.368	0.360
	PCCC	Bias	0.221	0.102	0.101	0.089	0.112	0.107	0.266	0.152	0.153	0.146	0.162	0.157
		RMSE	0.241	0.138	0.137	0.126	0.144	0.142	0.285	0.175	0.177	0.169	0.185	0.180
3	CCC	Bias	0.279	0.183	0.176	0.150	0.190	0.185	0.454	0.319	0.319	0.296	0.335	0.328
		RMSE	0.284	0.195	0.187	0.162	0.199	0.195	0.460	0.328	0.327	0.304	0.343	0.336
	PCCC	Bias	0.149	0.043	0.039	0.028	0.049	0.046	0.236	0.127	0.127	0.121	0.135	0.132
		RMSE	0.174	0.106	0.103	0.097	0.106	0.106	0.259	0.158	0.157	0.151	0.165	0.161
4	CCC	Bias	0.196	0.127	0.117	0.093	0.127	0.125	0.420	0.290	0.290	0.265	0.305	0.298
		RMSE	0.202	0.142	0.129	0.109	0.139	0.136	0.426	0.300	0.299	0.274	0.314	0.308
	PCCC	Bias	0.076	-0.011	-0.019	-0.029	-0.009	-0.010	0.206	0.096	0.095	0.089	0.105	0.102
		RMSE	0.117	0.102	0.100	0.100	0.097	0.098	0.233	0.136	0.135	0.129	0.143	0.140
5	CCC	Bias	0.536	0.371	0.370	0.333	0.389	0.385	0.648	0.443	0.444	0.415	0.463	0.457
		RMSE	0.537	0.377	0.375	0.338	0.394	0.391	0.651	0.448	0.449	0.419	0.467	0.461
	PCCC	Bias	0.332	0.163	0.163	0.153	0.175	0.171	0.327	0.179	0.179	0.175	0.188	0.185
		RMSE	0.345	0.187	0.187	0.176	0.197	0.194	0.344	0.198	0.198	0.193	0.207	0.203
6	CCC	Bias	0.439	0.289	0.286	0.251	0.304	0.301	0.605	0.409	0.409	0.380	0.428	0.422
		RMSE	0.441	0.296	0.292	0.257	0.309	0.307	0.608	0.414	0.414	0.385	0.432	0.427
	PCCC	Bias	0.245	0.088	0.087	0.077	0.098	0.095	0.288	0.144	0.144	0.139	0.154	0.149
		RMSE	0.263	0.129	0.127	0.118	0.134	0.132	0.308	0.168	0.169	0.163	0.178	0.173
7	CCC	Bias	0.343	0.217	0.210	0.179	0.225	0.223	0.562	0.379	0.378	0.349	0.395	0.390
		RMSE	0.345	0.225	0.217	0.187	0.232	0.230	0.565	0.385	0.383	0.354	0.400	0.396
	PCCC	Bias	0.164	0.022	0.018	0.009	0.028	0.026	0.254	0.112	0.112	0.107	0.121	0.117
		RMSE	0.187	0.099	0.097	0.093	0.099	0.099	0.277	0.145	0.144	0.140	0.153	0.149
8	CCC	Bias	0.250	0.154	0.141	0.114	0.154	0.153	0.522	0.348	0.346	0.316	0.363	0.359
		RMSE	0.253	0.164	0.150	0.125	0.162	0.161	0.525	0.354	0.352	0.322	0.369	0.365
	PCCC	Bias	0.085	-0.042	-0.049	-0.057	-0.041	-0.040	0.221	0.078	0.077	0.072	0.085	0.084
		RMSE	0.123	0.112	0.112	0.113	0.108	0.108	0.248	0.123	0.121	0.117	0.128	0.127
9	CCC	Bias	0.587	0.417	0.414	0.372	0.435	0.438	0.750	0.503	0.504	0.468	0.523	0.522
		RMSE	0.587	0.420	0.417	0.376	0.438	0.442	0.751	0.506	0.507	0.471	0.526	0.525
	PCCC	Bias	0.360	0.154	0.154	0.145	0.164	0.163	0.354	0.172	0.174	0.168	0.181	0.180
		RMSE	0.373	0.178	0.178	0.169	0.187	0.186	0.372	0.192	0.193	0.187	0.199	0.198
10	CCC	Bias	0.487	0.332	0.327	0.287	0.345	0.349	0.702	0.467	0.465	0.432	0.484	0.483
		RMSE	0.488	0.337	0.331	0.291	0.349	0.354	0.703	0.471	0.469	0.435	0.487	0.487
	PCCC	Bias	0.272	0.077	0.074	0.067	0.084	0.083	0.313	0.136	0.137	0.132	0.145	0.142
		RMSE	0.288	0.120	0.117	0.111	0.124	0.124	0.333	0.161	0.162	0.157	0.169	0.167
11	CCC	Bias	0.388	0.252	0.243	0.207	0.260	0.264	0.655	0.433	0.430	0.396	0.449	0.449
		RMSE	0.389	0.258	0.248	0.213	0.264	0.269	0.656	0.436	0.434	0.400	0.452	0.452
	PCCC	Bias	0.186	0.005	0.001	-0.006	0.009	0.009	0.278	0.101	0.101	0.097	0.107	0.107
		RMSE	0.207	0.098	0.095	0.093	0.096	0.096	0.300	0.135	0.134	0.130	0.140	0.139
12	CCC	Bias	0.289	0.179	0.166	0.134	0.179	0.184	0.608	0.400	0.397	0.362	0.415	0.416
		RMSE	0.290	0.186	0.172	0.142	0.184	0.189	0.609	0.405	0.401	0.366	0.419	0.420
	PCCC	Bias	0.102	-0.064	-0.070	-0.077	-0.062	-0.061	0.243	0.065	0.063	0.060	0.072	0.071
		RMSE	0.137	0.121	0.123	0.125	0.117	0.117	0.269	0.113	0.111	0.109	0.118	0.117
13	CCC	Bias	0.599	0.461	0.454	0.408	0.474	0.494	0.796	0.560	0.558	0.518	0.578	0.585
		RMSE	0.600	0.464	0.457	0.411	0.477	0.497	0.796	0.563	0.560	0.520	0.580	0.587
	PCCC	Bias	0.394	0.147	0.146	0.140	0.156	0.156	0.388	0.167	0.168	0.165	0.175	0.175
		RMSE	0.406	0.172	0.171	0.164	0.180	0.181	0.405	0.187	0.187	0.184	0.194	0.194
14	CCC	Bias	0.499	0.371	0.362	0.319	0.381	0.400	0.746	0.521	0.517	0.478	0.536	0.543
		RMSE	0.500	0.375	0.365	0.322	0.384	0.404	0.746	0.523	0.519	0.480	0.538	0.546
	PCCC	Bias	0.302	0.066	0.064	0.058	0.073	0.075	0.347	0.130	0.130	0.128	0.136	0.136
		RMSE	0.317	0.114	0.112	0.106	0.119	0.120	0.367	0.155	0.155	0.153	0.160	0.162
15	CCC	Bias	0.400	0.285	0.274	0.234	0.291	0.308	0.696	0.483	0.478	0.440	0.497	0.504
		RMSE	0.400	0.289	0.277	0.238	0.294	0.312	0.696	0.486	0.480	0.442	0.499	0.506
	PCCC	Bias	0.211	-0.010	-0.015	-0.020	-0.006	-0.006	0.307	0.092	0.090	0.088	0.096	0.097
		RMSE	0.230	0.097	0.095	0.095	0.094	0.096	0.330	0.127	0.125	0.123	0.131	0.132
16	CCC	Bias	0.300	0.206	0.191	0.156	0.205	0.220	0.646	0.447	0.442	0.403	0.460	0.468
		RMSE	0.300	0.211	0.195	0.161	0.208	0.224	0.647	0.450	0.445	0.406	0.462	0.470
	PCCC	Bias	0.124	-0.083	-0.090	-0.093	-0.082	-0.081	0.268	0.056	0.053	0.052	0.061	0.061
		RMSE	0.153	0.132	0.135	0.135	0.129	0.130	0.294	0.106	0.104	0.103	0.110	0.111

4.1 Example 1

The pair of low CCCs is the gene with six variables (X36b4, ACAT1, ADISP, AM2R, Bcl.3, and C16SR) and the lipid with six variables (C14.0, C16.0, C16.1n.7, C18.1n.9, C20.3n.9, and C18.2n.6). CCCs and PCCCs are summarized in Table 7. MLE was excluded because it was not a valid method in the simulation results and it can only be run with older R software version. The first CCC estimated from the dataset comprising all observation data is 0.454, and the correlation between the gene and the lipid is low. However, the first CCCs estimated from the missing data based on the five methods are greater than 0.6. Therefore, they may be interpreted with high correlation. The first PCCCs estimated from the missing data based on AC analysis, MICE, kNN, and IRMI are 0.540, 0.561, 0.518, and 0.586, respectively. These values do not make a big misinterpretation because their values are not far from 0.454. The first PCCC estimated from the missing data based on kNN is closest to the first CCC estimated from the dataset comprising all observation data. PCCC estimated from the missing data based on kNN, which is the same method recommended in simulation results, is recommended.

The sample code used for the calculation is presented in the Appendix. Note that because observations are replaced with missing data by MCAR, the results change each time the calculation is performed. CCCs and PCCCs based on CC and AC analyses may be incalculable or greater than or equal to 1. Another practical example with missing data can be analyzed by setting the dataset of zmis and variables of p1 and q1 appropriately, and then executing the ninth and subsequent lines.

4.2 Example 2

The pair of high CCCs is the gene with six variables (X36b4, ACAT1, ACAT2, ACBP, ACC1, and ACC2) and the lipid with six variables (C14.0, C16.0, C18.0, C16.1n.9, C16.1n.7, and C18.1n.9). CCCs and PCCCs are summarized in Table 8. The first and second CCCs estimated from the dataset comprising all observation data are 0.849 and 0.742, respectively. CCCs and PCCCs estimated from the missing data based on the five methods are slightly different from the CCCs estimated from the dataset comprising all observation data. Therefore, the interpretation of them would be approximately similar to that based on CCCs estimated from the dataset comprising all observation data. When the correlation between two vectors is high, both CCCs and PCCCs are effective even for dataset with missing data.

5 Conclusions

The relationship between two random vectors was studied using missing data in small samples. CCCs were calculated using each estimated covariance matrix by excluding or imputing the missing data. In the simulation results, CCCs were greater than population CCCs. Therefore, even when the correlation of population CCCs is low, they can be misunderstood as being high. This problem was solved by calculating PCCCs following three steps. First, PCA was performed on the covariance matrix from the dataset with missing data to obtain the eigenvectors. Second, the covariance matrix was transformed using the first to fourth eigenvectors. Finally, CCCs were calculated from the transformed covariance matrix. In the simulation results, the bias and RMSE of PCCCs estimated from the missing data were smaller than those of the CCCs estimated from the missing data. Simulation results also showed that PCCCs estimated from the missing data based on kNN had the smallest RMSE in many cases. The numerical examples showed that the first to third PCCCs estimated from the missing data based on kNN were similar to the first to third CCCs estimated from the dataset comprising all observation data. This result was not dependent on the degree of correlation. As PCCCs estimated from the missing data based on kNN had a small bias, they could accurately interpret the correlation between the two random vectors. Therefore, PCCCs based on kNN were recommended. This manuscript targeted small samples, and their effectiveness was verified when the proportion of missing data was small. The proportion of missing data may be high for large samples. Therefore, future studies are to verify the effectiveness of the research method in this manuscript in large samples when the proportion of missing data is set variously. As a limit of PCCCs, when the correlation between each random vector is low, the contribution rate obtained by PCA does not have a large value. In this case, if the number of eigenvectors used for PCCA is small, a large amount of information may be lost. One of the countermeasures is to determine the number of eigenvectors to be used from the contribution rate.

Acknowledgements

The authors are deeply grateful to the referees for useful comments.

Appendix

The sample code of the R software for estimating the CCCs and PCCCs from missing data by MCAR is presented below. The code presents the setting used for the calculation shown in Example 1. Another practical example with missing data can be analyzed by setting the dataset of zmis and variables of p1 and q1 appropriately, and then executing the ninth and subsequent lines.

```
#Replace observation data with missing data
library(CCA); data(nutrimouse)
x<-as.matrix(nutrimouse$gene[,c("X36b4",
"ACAT1","ADISP","AM2R","Bcl.3","C16SR")])
y<-as.matrix(nutrimouse$lipid[,c("C14.0",
"C16.0","C16.1n.7","C18.1n.9","C20.3n.9",
"C18.2n.6")]); p<-ncol(x); q<-ncol(y); n<-
nrow(x); tau<-0.90; z<-cbind(x,y); zmis<-z
M<-sample(n*(p+q),(1-tau)*n*(p+q))-1; Mx<-
M%/%(p+q)+1; My<-M%%(p+q)+1; for(i in 1:
((1-tau)*n*(p+q))){zmis[Mx[i],My[i]]<-NA}

#CCCs and PCCCs for missing data
```

Table 7. CCCs and PCCCs based on five methods between the gene and lipid in Example 1.

		1st	2nd	3rd	4th	5th	6th
All observation	CCC	0.454	0.413	0.316	0.264	0.217	0.111
CC	CCC	0.993	0.909	0.872	0.665	0.285	0.113
	PCCC	0.967	0.489	0.414	0.030		
AC	CCC	0.832	0.711	0.447	0.311	0.193	0.020
	PCCC	0.540	0.361	0.321	0.158		
MICE	CCC	0.690	0.523	0.356	0.302	0.203	0.073
	PCCC	0.561	0.347	0.228	0.051		
kNN	CCC	0.621	0.521	0.366	0.324	0.227	0.050
	PCCC	0.518	0.316	0.266	0.016		
IRMI	CCC	0.751	0.507	0.353	0.267	0.258	0.072
	PCCC	0.586	0.264	0.182	0.074		
0.340	0.175						

PCCCs are calculated from first to fourth because $p_1 = q_1 = 4$.

Table 8. CCCs and PCCCs based on five methods between the gene and lipid in Example 2.

		1st	2nd	3rd	4th	5th	6th
All observation	CCC	0.849	0.742	0.396	0.231	0.105	0.073
CC	CCC	0.991	0.951	0.903	0.658	0.351	0.147
	PCCC	0.966	0.890	0.779	0.046		
AC	CCC	1.032	0.915	0.759	0.619	0.421	0.190
	PCCC	0.881	0.800	0.216	0.150		
MICE	CCC	0.911	0.877	0.609	0.456	0.228	0.151
	PCCC	0.875	0.835	0.223	0.200		
kNN	CCC	0.901	0.737	0.521	0.446	0.235	0.081
	PCCC	0.883	0.722	0.302	0.029		
IRMI	CCC	0.923	0.876	0.642	0.394	0.340	0.175
	PCCC	0.872	0.795	0.255	0.137		

PCCCs are calculated from first to fourth because $p_1 = q_1 = 4$.

```
library(mice) #MICE
library(VIM)  #kNN, IRMI
p1<-4; q1<-4
S_CC<-cov(zmis,use="complete"); S_AC<-cov(
zmis,use="pairwise"); S_MICE<-cov(complete(
mice(zmis,method="pmm",print=F)))
S_kNN<-cov(kNN(data.frame(zmis))[,1:(p+q)])
S_IRMI<-cov(irmi(zmis))
CCC_PCCC<-function(S){Sxx<-S[1:p,1:p]; Sxy<-
S[1:p,(p+1):(p+q)]; Syx<-S[(p+1):(p+q),1:p]
Syy<-S[(p+1):(p+q),(p+1):(p+q)]; CCC<-sqrt(
eigen(solve(Sxx)%*%Sxy%*%solve(Syy)%*%Syx)$
values); Hx<-eigen(Sxx)$vectors[,1:p1]; Hy<-
eigen(Syy)$vectors[,1:q1]; Suu<-t(Hx)%*%
Sxx%*%Hx; Suv<-t(Hx)%*%Sxy%*%Hy; Svu<-t(Hy
)%*%Syx%*%Hx; Svv<-t(Hy)%*%Syy%*%Hy; PCCC<-
sqrt(eigen(solve(Suu)%*%Suv%*%solve(Svv)%*%
Svu)$values); return(rbind(CCC,PCCC=c(PCCC,
rep(NA,p-min(p1,q1)))))}
cancor(x,y)$cor
CCC_PCCC(S_CC)
CCC_PCCC(S_AC)
CCC_PCCC(S_MICE)
CCC_PCCC(S_kNN)
CCC_PCCC(S_IRMI)
```

REFERENCES

[1] C. K. Enders, Applied Missing Data Analysis, Guilford press, New York, 2010.

[2] J. Ma, R. R. Wilcox, Robust within groups anova: dealing with missing values, Mathematics and Statistics, vol. 1, no. 1, pp. 1–4, 2013. DOI: 10.13189/ms.2013.010101

[3] M. O'Kelly, B. Ratitch. Clinical Trials with Missing Data: A Guide for Practitioners, John Wiley & Sons, West Sussex, 2014.

[4] J. Hao, K. Krishnamoorthy. Inferences on a normal covariance matrix and generalized variance with monotone missing data, Journal of Multivariate Analysis, vol. 78, no. 1, pp. 62–82, 2001. DOI: 10.1006/jmva.2000.1939

[5] M. Jamshidian, P. M. Bentler. ML estimation of mean and covariance structures with missing data using complete data routines, Journal of Educational and behavioral Statistics, vol. 24, no. 1, pp. 21–24, 1999. DOI: 10.3102/10769986024001021

[6] J. Liu, D. P. Palomar. Regularized robust estimation of mean and covariance matrix for incomplete data, Signal Processing, vol. 165, pp. 278–291, 2019. DOI: 10.1016/j.sigpro.2019.07.009

[7] S. Tsukada. Unbiased estimator for a covariance matrix under two-step monotone incomplete sample, Communications in Statistics - Theory and Methods, vol. 43, no. 8, pp. 1613–1629,

2014a. DOI: 10.1080/03610926.2012.671881

[8] S. Tsukada. Asymptotic expansion for distribution of the trace of a covariance matrix under a two-step monotone incomplete sample, Journal of Multivariate Analysis, vol. 129, pp. 206–219, 2014b. DOI: 10.1016/j.jmva.2014.04.019

[9] S. Tsukada. Unbiased estimator for a covariance matrix in a three-step monotone incomplete sample, Linear Algebra and its Applications, vol. 579, pp. 169–205, 2019. DOI: 10.1016/j.laa.2019.05.023

[10] R. J. Little, D. B. Rubin. Statistical Analysis with Missing Data, 3rd ed., Wiley Series in Probability and Statistics, New Jersey, 2019.

[11] S. V. Buuren, K. Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in R, Journal of statistical software, vol. 45, no. 3, pp. 1–68, 2011. DOI: 10.18637/jss.v045.i03

[12] A. Kowarik, M. Templ. Imputation with the R package VIM. Journal of Statistical Software, vol. 74, no. 7, pp. 1–16, 2016. DOI: 10.18637/jss.v074.i07

[13] M. Templ, A. Kowarik, P. Filzmoser. Iterative stepwise regression imputation using standard and robust methods, Computational Statistics & Data Analysis, vol. 55, no. 10, pp. 2793–2806, 2011. DOI: 10.1016/j.csda.2011.04.012

[14] K. Gross, D. Bates. mvnmle: ML estimation for multivariate normal data with missing values, R package version 0.1–10, 2011, https://rdrr.io/cran/mvnmle/ (accessed Dec. 15, 2021).

[15] H. Hotelling. The most predictable criterion, Journal of Educational Psychology, vol. 26, no. 2, pp. 139–142, 1935. DOI: 10.1037/h0058165

[16] H. Hotelling. Relations between two sets of variates, Biometrika, vol. 28, no. 3/4, pp. 321–377, 1936. DOI: 10.1093/biomet/28.3-4.321

[17] T. W. Anderson. An Introduction to Multivariate Statistical Analysis, 3rd ed., John Wiley & Sons, New Jersey, 2003.

[18] M. Van de Velden, Y. Takane, Generalized canonical correlation analysis with missing values, Computational Statistics, vol. 27, no. 3, pp. 551–571, 2012. DOI: 10.1007/s00180-011-0276-y

[19] T. Yamada. Asymptotic properties of canonical correlation analysis for one group with additional observations, Journal of Multivariate Analysis, vol. 114, pp. 389–401, 2013. DOI: 10.1016/j.jmva.2012.08.001

[20] D. N. Lawley. Tests of significance in canonical analysis, Biometrika, vol. 46, no. 1/2, pp. 59–66, 1959. DOI: 10.1093/biomet/46.1-2.59

[21] R. A. Cahyoningtyas, A. A. R. F. Solimun. The implementation of first order and second order with mixed measurement to identify farmers satisfaction. Mathematics and Statistics, vol. 8, no. 6, pp. 671–682, 2020. DOI: 10.13189/ms.2020.080607

[22] T. Sugiyama, T. Ogura, F. Sakaori, T. Yamada, Distributions and bootstrap method of some statistics in principal canonical correlation analysis, Journal of the Japan Statistical Society, vol. 37, no. 2, pp. 239–251, 2007.

[23] M. Yamamoto, T. Sugiyama, H. Murakami, F. Sakaori. Correlation analysis of principal components from two populations, Computational Statistics & Data Analysis, vol. 51, no. 9, pp. 4707–4716, 2007. DOI: 10.1016/j.csda.2006.08.034

[24] T. Ogura. A variable selection method in principal canonical correlation analysis, Computational Statistics & Data Analysis, vol. 54, no. 4, pp. 1117–1123, 2010. DOI: 10.1016/j.csda.2009.09.008

[25] T. Ogura, Y. Fujikoshi, T. Sugiyama. A variable selection criterion for two sets of principal component scores in principal canonical correlation analysis, Communications in Statistics - Theory and Methods, vol. 42, no. 12, pp. 2118–2135, 2013. DOI: 10.1080/03610926.2011.605235

[26] J. T. Roscoe. Fundamental Research Statistics for the Behavioral Science, International Series in Decision Process, 2nd ed., Holt, Rinehart and Winston, New York, 1975.

[27] R Core Team. R: a language and environment for statistical computing (Version 3.4.0), Vienna, Austria: R Foundation for Statistical Computing, Vienna, 2017. https://www.R-project.org/ (accessed Dec. 15, 2021).

[28] J. C. Gower. A general coefficient of similarity and some of its properties, Biometrics, vol. 27, no. 4, pp. 857–871, 1971. DOI: 10.2307/2528823

[29] T. E. Raghunathan, J. M. Lepkowski, J. Van Hoewyk, P. Solenberger. A multivariate technique for multiply imputing missing values using a sequence of regression models, Survey methodology, vol. 27, no. 1, pp. 85–96, 2001.

[30] P. G. Martin, H. Guillou, F. Lasserre, S. Déjean, A. Lan, J. M. Pascussi, M. Sancristobal, P. Legrand, P Besse, T. Pineau. Novel aspects of PPARα-mediated regulation of lipid and xenobiotic metabolism revealed through a nutrigenomic study, Hepatology, vol. 45, no. 3, pp. 767–777, 2007. DOI: doi.org/10.1002/hep.21510

[31] I. González, S. Déjean, P. Martin, A. Baccini. CCA: An R package to extend canonical correlation analysis, Journal of Statistical Software, vol. 23, no. 12, pp. 1–14, 2008. DOI: 10.18637/jss.v023.i12