JOURNAL OF MULTIVARIATE ANALYSIS **56**, 1–19 (1996)
ARTICLE NO. 0001

Asymptotic Distribution of Restricted Canonical Correlations and Relevant Resampling Methods

SHUBHABRATA DAS

University of Nebraska at Lincoln

AND

PRANAB KUMAR SEN

University of North Carolina at Chapel Hill

As restricted canonical correlation with a nonnegativity condition on the coefficients depend only on the covariance matrix, their sample counterparts can be obtained from the sample covariance matrix. For such estimators, asymptotic normality results are established, and the role of resampling methods in this context is critically examined. The effectiveness of the usual jackknife and bootstrap methods is studied analytically, and the findings are supplemented by numerical studies.

1. INTRODUCTION AND PRELIMINARY NOTIONS

For a random vector $\mathbf{Y}' = (\mathbf{Y}^{(1)'}, \mathbf{Y}^{(1)'})$, with a finite covariance matrix $\boldsymbol{\Sigma} = ((\boldsymbol{\Sigma}_{ij}))_{i,j=1,2}$, the first canonical correlation (CC) between $\mathbf{Y}^{(1)}$ (p-variate) and $\mathbf{Y}^{(2)}$ (q-variate) is defined as

$$\max\{\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\beta} : \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} = 1 = \boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}\},$$

or, it is the maximum correlation between any *representative* linear combinations $\boldsymbol{\alpha}'\mathbf{Y}^{(1)}$ and $\boldsymbol{\beta}'\mathbf{Y}^{(2)}$, when the domains of the coefficients or the weights $\boldsymbol{\alpha} = ((\alpha_i))$ and $\boldsymbol{\beta} = ((\beta_j))$ are the entire $\mathbb{R}_p$ (p-dimensional Euclidean space) and $\mathbb{R}_q$, respectively. However, in many practical situations (e.g., educational testing problems, neural networks) (Das and Sen [3, 4]), some natural restrictions on the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ may arise which should be incorporated in this maximization procedure. The maximum correlation

Received July 14, 1993; revised March 1995.
AMS 1990 subject classification: 62H20, 62E20, 62G09.
Key words and phrases: asymptotic normality; bias; bootstrap; canonical coefficients; consistency; inequality restrictions; jackknife; nonnegativity restrictions; variance; simulation.

0047-259X/96 $12.00

subject to such constraints is referred to as the first restricted canonical correlation (RCC). In many cases some inequality restrictions can be reduced by simple transformations to a nonnegativity restriction on the coefficients, which we define as follows.

DEFINITION. The RCC between random vectors $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ is defined to be the maximum correlation between nonnegative linear combinations of the two sets of component variables ($\boldsymbol{\alpha}'\mathbf{Y}^{(1)}$ and $\boldsymbol{\beta}'\mathbf{Y}^{(2)}$); i.e.,

$$\text{RCC} = \max\{\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{12}\boldsymbol{\beta} : \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} = 1 = \boldsymbol{\beta}'\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}, \boldsymbol{\alpha} \in \mathbb{R}_p^+, \boldsymbol{\beta} \in \mathbb{R}_q^+\},$$

where $\mathbb{R}_p^+$ denotes the nonnegative orthant of $\mathbb{R}_p$.

The nonnegativity restriction is important primarily because it enables one to obtain both canonical variables as *convex* combinations of original random variables, and we shall mostly follow this case in detail.

It may be emphasized here that the interpretation of the coefficients is a chief motivation for the general study of RCC. In the usual CC analysis, often one can obtain very different sets of coefficients which lead to (at least approximately) maximal correlations. This is much less likely to happen with RCC. Also, if it does, it should not bother the experimenter, because, all the candidate coefficients must be satisfying the *reasonable* constraints that have been built into the problem. Toward this end, we consider couple of illustrative examples.

EXAMPLE 1. Suppose we are interested in studying the relationship/dependence of a group of students' performances in two different subjects, say mathematics and French. One solution is to assign *reasonable* weights (as is usually done) to the different examinations/assignments and compute the composite scores in the two subjects. Then the correlation between the mathematics and French composite scores would be a measure of such dependence. However, such an approach can be questioned because of the arbitrariness of the weights, and alternatively one may look for those weights which *maximize* the correlation between the two composite scores. The latter is precisely the traditional CC approach. RCC fits somewhere in the middle. It also maximizes the correlation between possible composite scores; but instead of allowing these weights to be totally flexible, one may force them to be more reflective of their individual importance. For instance, if for the mathematics course, the students had scores (say, for convenience, out of 100) in each of the categories: homework (hw), quiz, midterm (mt) and final, then one may like to place the restriction

$$0 \leqslant \alpha_{\text{quiz}} \leqslant \alpha_{\text{hw}} \leqslant \alpha_{\text{mt}} \leqslant \alpha_{\text{final}}. \tag{1}$$

If the French course had only three tests, then a reasonable restriction on the weights may be that they all be nonnegative, i.e.,

$$\beta_i \geqslant 0, \qquad i = 1, 2, 3. \tag{2}$$

Since correlation is scale-invariant, the optimal coefficients in each group can be rescaled to enforce the sum to be 1 (so that the composite score is out of 100). It has been shown in Das and Sen [3], how a RCC problem with restrictions (1) and (2) (or, a more general mixture of equality and inequality restrictions) can be transformed to a RCC problem with non-negative restriction.

EXAMPLE 2. Timm [10] has a data set for two dependent variables—time on target in seconds and the number of hits on target (average outcomes using two hands). For each variable, five repeated measures were obtained using 16 subjects. Naturally, CC analysis can be done to study the dependence between the two (groups of) variables. But, it is possible to obtain convex linear combinations of the five measurements (in each variable) which are almost as much dependent on each other as the canonical variables, and they are more intuitively appealing (because of the convexity). We defer numerical illustrations to Section 4.

To present the solution under the nonnegativity restriction on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we need to introduce some notations. Let p and q be fixed integers as denoted previously. Also, define

$$\mathbb{N}_k = \{1, 2, ..., k\}, \quad \mathbb{W}_k = \{\mathbf{a} : \phi \neq \mathbf{a} \subseteq \mathbb{N}_k\}, \qquad k \geqslant 1;$$

with elements of $\mathbf{a}$ ($\in \mathbb{W}_k$) written in natural order. Further, let $|\mathbf{a}|$ denote the cardinality of $\mathbf{a}$. For a p-component vector $\mathbf{X}$, and $\mathbf{a} \in \mathbb{W}_p$, let ${}_{\mathbf{a}}\mathbf{X}$ stand for the $|\mathbf{a}|$-component vector consisting of those components of $\mathbf{X}$ whose indices belong to $\mathbf{a}$. Similarly, for a $(p \times q)$-dimensional matrix $\mathbf{S}$, $\mathbf{a} \in \mathbb{W}_p$, $\mathbf{b} \in \mathbb{W}_q$, let ${}_{\mathbf{a}:\mathbf{b}}\mathbf{S}$ represent the $(|\mathbf{a}| \times |\mathbf{b}|)$-dimensional submatrix of $\mathbf{S}$, consisting of those rows whose indices are in $\mathbf{a}$ and those collumns whose indices are in $\mathbf{b}$. So if

$$\mathbf{X} = (X_1, ..., X_p)', \qquad \mathbf{S} = ((s_{ij})), \qquad \mathbf{a} = (i_1, ..., i_l), \qquad \mathbf{b} = (j_1, ..., j_k),$$

then

$${}_{\mathbf{a}}\mathbf{X} = \begin{pmatrix} X_{i_1} \\ \vdots \\ X_{i_l} \end{pmatrix}; \qquad {}_{\mathbf{a}:\mathbf{b}}\mathbf{S} = \begin{bmatrix} s_{i_1 j_1} & \cdots & s_{i_1 j_k} \\ \vdots & \vdots & \vdots \\ s_{i_l j_1} & \cdots & s_{i_l j_k} \end{bmatrix}.$$

Then Das and Sen [3] have proved the following.

LEMMA 1. *If the set of squared CC's corresponding to* $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$ *is represented by* $\mathrm{setcc}^2(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{22})$, *then*

$$\mathrm{RCC}^2(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{22}) \in \bigcup_{\mathbf{a} \in \mathbb{W}_p,\, \mathbf{b} \in \mathbb{W}_q} \{\mathrm{setcc}^2({}_{\mathbf{a}:\mathbf{a}}\boldsymbol{\Sigma}_{11}, {}_{\mathbf{a}:\mathbf{b}}\boldsymbol{\Sigma}_{12}, {}_{\mathbf{b}:\mathbf{b}}\boldsymbol{\Sigma}_{22})\}.$$

Or, equivalently, the squared RCC between $\mathbf{Y}^{(1)}$ *and* $\mathbf{Y}^{(2)}$ *is equal to one of the squared CC's between* ${}_{\mathbf{a}}\mathbf{Y}^{(1)}$ *and* ${}_{\mathbf{b}}\mathbf{Y}^{(2)}$ *for some* $\mathbf{a} \in \mathbb{W}_p$, $\mathbf{b} \in \mathbb{W}_q$.

More precisely, the squared RCC is the *largest* squared CC (first or higher) between subvectors of the two groups, for which the canonical coefficients satisfy the nonnegativity condition.

Let us illustrate Lemma 1 with a simple numerical example. Suppose,

$$\boldsymbol{\Sigma}_{11} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} = \boldsymbol{\Sigma}_{22}, \qquad \boldsymbol{\Sigma}_{12} = \begin{pmatrix} 0.1 & 0 \\ 0 & -0.05 \end{pmatrix}.$$

It can be easily verified that, the first CC and the canonical coefficients are given by:

$$\mathrm{CC} \approx 0.1215, \qquad \boldsymbol{\alpha} = \boldsymbol{\beta} \approx \begin{pmatrix} 0.9426 \\ -0.3340 \end{pmatrix};$$

the second CC is ≈ 0.0548. Since the canonical coefficient corresponding to first CC does not satisfy the nonnegativity restriction, it does not give RCC. Clearly, if we drop a variable from one or both the two groups, then the CC resulting from the subvectors are going to be 0.1, 0.05, or 0; and the corresponding coefficients for each of these cases are nonnegative (0 or 1). Hence, by Lemma 1, the RCC $= 0.1$, which is the correlation between the first variables of the two groups.

Das and Sen [3] have considered more general form of inequality restraints (including the case where such restrictions apply only on some of the coefficients. Similar restricted principal components and part, partial, and bipartial canonical correlations have also been studied by them. Here, the main emphasis is on the sampling distribution of RCC with the nonnegativity restriction, as the other cases follow very similar tracks.

Sample restricted canonical correlation (SRCC) is obtained when in the calculation of RCC, the population covariance matrix $\boldsymbol{\Sigma}$ is replaced by the sample covariance matrix $\mathbf{S}_n$. The intuitive justification is the same as in the case of traditional canonical correlation; i.e., it provides the maximum likelihood estimate of population RCC when the underlying distribution is a multivariate normal one. (Note that it does not change the value of SRCC, if we use $((n-1)/n)\,\mathbf{S}_n$, instead of $\mathbf{S}_n$.) Otherwise, $\mathbf{S}_n$ is, at least, a method of moment estimator of $\mathbf{\Sigma}$, and this provides some justification. It would have been nice to have an alternative candidate which is an unbiased estimator of population RCC. But, as in the case of usual CC, this is not generally available.

In Section 2, a representation of SRCC in terms of different sample CC's along with the asymptotic normality results for the latter is incorporated in the derivation of the asymptotic normality for the SRCC. Relevant resampling methods are discussed in Section 3, and some related numerical studies are made in the last section.

2. Asymptotic Normality of SRCC

It follows from Lemma 1 that the squared RCC equals the square of the first or higher CC of some *proper* submatrix. There are in $\mathscr{P}\mathscr{Q}$ such submatrices, where $\mathscr{P}=2^p-1$ and $\mathscr{Q}=2^q-1$. Further, there are $\binom{p}{i}\binom{q}{j}$ proper submatrices of order $(i+j)\times(i+j)$, having $(i\wedge j)$ CC's. Hence, there are

$$\mathscr{K}=\sum_{i=1}^{p}\sum_{j=1}^{q}\binom{p}{i}\binom{q}{j}(i\wedge j)$$

squared CC's which are possible candidates of the squared RCC. Now index these $\mathscr{P}\mathscr{Q}$ proper submatrices in any fixed order. Let $\rho^2_{r,s}$ denote the squared rth CC of the sth proper submatrix of $\mathbf{\Sigma}$, where $s=1, ..., \mathscr{P}\mathscr{Q}$, and $r=1, ..., \min(i,j),$) if the sth proper submatrix is of dimension $(i+j)\times(i+j)$. Similarly, let $X_{(r,s),n}$ denote its sample version, i.e., the squared rth canonical correlation of the sth proper submatrix of $\mathbf{S}_n$. While such double-indexing describes the situation appropriately, the notation becomes clumsy. Hence, for the sake of notational convenience, $\rho^2_{r,s}$ and $X_{(r,s),n}$ will be denoted in the sequel by ρ^2_k and $X_{k,n}$, respectively, with the index k running from 1 to $\mathscr{K}$.

The sample covariance matrix $\mathbf{S}_n$ can be partitioned into $\mathscr{K}$ classes, and $\mathbf{S}_n$ belongs to the kth class if the squared RCC of $\mathbf{S}_n$ is its k-squared CC in the mode of ordering as described above.

Also let T_n and θ denote the squared sample and population RCC, respectively. By Lemma 1, $\theta=\rho^2_{\check{k}}$ for some $\check{k}$. The following assumption is crucial for most of the statistical properties discussed in this work:

$$\check{k} \text{ is unique, i.e., } \rho^2_k\neq\theta \qquad \text{for} \quad k\neq\check{k}. \tag{3}$$

Also, for notational convenience, we assume without loss of generality, that $\check{k}=1$.

Representing SRCC in Terms of Sample CC's

From Section 1, it is easy to see (replacing $\mathbf{\Sigma}$ by $\mathbf{S}_n$ in Lemma 1) that sample squared RCC is always equal to sample squared for some "subvector"; i.e.,

$$T_n(\omega)=X_{k,n}(\omega) \quad \text{for some } k=k(\omega,n)\in\{1,2,...,\mathscr{K}\}, \quad \forall n\geqslant 1, \forall \omega\in\Omega, \tag{4}$$

where Ω is the probability space on which all the random variables are defined. So if

$$A^0_{k,n}=\{\omega: T_n(\omega)=X_{k,n}(\omega)\}, \qquad k=1,2,...,\mathscr{K}, \tag{5}$$

then $\bigcup_{k=1}^{\mathscr{K}} A^0_{k,n}=\Omega$. Further, for each n, one can define a partition $\{A_{k,n}, k=1,2,...,\mathscr{K}\}$ of Ω by

$$A_{1,n}=A^0_{1,n}, \quad A_{k,n}=A^0_{k,n}\Big\backslash \bigcup_{i=1}^{k-1} A_{i,n}, \qquad k=2,...,n. \tag{6}$$

Thus, we obtain the following representation of SRCC,

$$T_n=\sum_{k=1}^{\mathscr{K}} \mathbb{I}_{A_{k,n}}\times X_{k,n}, \tag{7}$$

where $\mathbb{I}_A$ stands for the indicator function of the set A. Muirhead and Waternaux [7] showed that if ρ_k is a distinct population CC then

$$\sqrt{n}(X_{k,n}-\rho_k^2)\xrightarrow{\mathscr{L}} N(0,\eta_k^2), \tag{8}$$

where η_i^2 is of the form

$$\begin{aligned}4\rho_i^2(1-\rho_i^2)^2+\rho_i^4(\kappa_{i:4}+\kappa_{i+p:4})+2\rho_i^2(\rho_i^2+2)\,\kappa_{i,p+i:2,2}\\ -4\rho_i^3(\kappa_{i,p+i:3,1}+\kappa_{i,p+i:1,3}),\end{aligned} \tag{9}$$

with κ's being different fourth-order cumulants of the population. Also, since $X_{k,n}$ is a continuous function of the sample covariance matrix,

$$X_{k,n}\xrightarrow{\text{a.s.}}\rho_k^2, \qquad k=1,...,\mathscr{K}. \tag{10}$$

The following lemma, a key to the proof of main results, relates to different modes of convergence of indicator variables. The proof is fairly straightforward (Das [2]) and, hence, is omitted here.

LEMMA 2. *Given a sequence of events A_n, the following are equivalent*:

(i) $\mathbb{I}_{A_n} \xrightarrow{\mathscr{P}} 0$.

(ii) $\mathbb{I}_{A_n} = o_p(n^\alpha)$ *for some* $\alpha < 0$.

(iii) $\mathbb{I}_{A_n} \xrightarrow{L^1} 0$, *i.e.*, $P(A_n) \to 0$, *and are implied by*

(iv) $\mathbb{I}_{A_n} \xrightarrow{\text{a.s.}} 0$.

COROLLARY 1. *This lemma remains true if the limiting constant is* 1 *instead of* 0.

Since, (S)RCC is a continuous function of the (sample) covariance matrix and $\mathbf{S}_n \xrightarrow{\text{a.s.}} \mathbf{\Sigma}$ as $n \to \infty$, we immediately obtain that

$$T_n \xrightarrow{\text{a.s.}} \theta. \tag{11}$$

We may note that (11) holds even when Assumption (3) may not hold. Since the SRCC is an implicit function of $\mathbf{S}_n$, estimation of η_k^2 is somewhat involved, and an alternative method is pursued here. First, consider the following.

LEMMA 3. *Under Assumption* (3), $\mathbb{I}_{A_{1,n}} \xrightarrow{\text{a.s.}} 1$, *where* $A_{1,n}$ *is defined as in* (6).

Proof. Let $\Omega_0 = \{\omega : T_n(\omega) \to \theta;\ X_{k,n} \to \rho_k^2;\ \forall k = 1, 2, ..., \mathscr{K}\}$. By (10) and (11),

$$P(\Omega_0) = 1. \tag{12}$$

We claim that (since $\mathscr{K}$ is finite), $\forall \omega \in \Omega_0$,

$$\exists N(\omega) \qquad \text{such that} \quad n \geqslant N(\omega) \Rightarrow T_n(\omega) = X_{1,n}(\omega). \tag{13}$$

Otherwise, by (4), there must exist an index $k_1 \geqslant 2$ and a subsequence $\{n_l\}$, such that

$$T_{n_l}(\omega) = X_{k_1, n_l}(\omega) \qquad \forall l \geqslant 1.$$

Hence, we would have

$$\lim_{l \to \infty} T_{n_l}(\omega) = \lim_{l \to \infty} X_{k_1, n_l}(\omega) = \rho_{k_1}^2 \neq \theta \qquad \text{(by Assumption 3)}.$$

But that would imply

$$\lim_{n \to \infty} T_n(\omega) \neq \theta,$$

which is a contradiction to ω belonging to Ω_0.

To complete the proof of the lemma, we note that (13) implies

$$\forall \omega \in \Omega_0 \ \exists N(\omega) \qquad \text{such that} \quad n \geqslant N(\omega) \Rightarrow \mathbb{I}_{A_{1,n}}(\omega) = 1,$$

or equivalently, $\mathbb{I}_{A_{1,n}}(\omega) \to 1 \ \forall \omega \in \Omega_0$. ∎

COROLLARY 2. *Under Assumption* (3), $\mathbb{I}_{A_{k,n}} \xrightarrow{\text{a.s.}} 0$ *for* $k \geqslant 2$.

THEOREM 1. *Under Assumption* (3), $\sqrt{n}(T_n - \theta) \xrightarrow{\mathscr{L}} N(0, \eta_1^2)$.

Proof. From (7) we get

$$\sqrt{n}(T_n - \theta) = \mathbb{I}_{A_{1,n}} \times \sqrt{n}(X_{1,n} - \theta) + \sum_{k=2}^{\mathscr{K}} \mathbb{I}_{1_{k,n}} \times \sqrt{n}(X_{k,n} - \theta).$$

Since by Assumption (3), $\rho_k^2 \neq \theta$ for $k > 1$, $X_{1,n}$ corresponds to a distinct population CC and, hence, by (8), $\sqrt{n}(X_{1,n} - \theta) \xrightarrow{\mathscr{L}} N(0, \eta_1^2)$. Further, all CC's being bounded by 1, $\sqrt{n}\,|X_{k,n} - \theta| \leqslant 2\sqrt{n}$ and, hence, by Lemma 2, Lemma 3, Corollary 1, and Corollary 2, $\mathbb{I}_{A_{1,n}} = 1 - o_p(1)$, and $\mathbb{I}_{A_{k,n}} = o_p(1/\sqrt{n})$ for $k \geqslant 2$. This completes the proof via Slutsky's theorem. ∎

It may be worthwhile to reiterate the importance of Assumption (3) in this context. It is clear that the proof presented here may not work in the absence of (3). For example, if $\theta = \rho_1^2 = \rho_2^2$, then parallel to Lemma 3, we have $\mathbb{I}_{A_{1,n} \cup A_{2,n}} \xrightarrow{\text{a.s.}} 1$, which may hold even with none of $\mathbb{I}_{A_{i,n}}$'s (for $i = 1, 2$) converging individually. Also, the implications of this assumption is beyond this. The present authors are of the opinion that when (3) fails to hold, the limiting distribution of SRCC, even if it exists, may be non-normal.

3. RESAMPLING METHODS FOR RCC

In Section 2, we observed that the asymptotic variance of SRCC involves unknown fourth-order moments of the population. Hence, for obtaining a large sample confidence interval for the RCC or to test for a suitable hypothesis on it, a good estimate of the sample variance is required. It is natural to think of *resampling methods* (viz., *jackknife* and *bootstrap*), and these will be pursued here. In this context, estimates of bias of SRCC can be used to improve the estimates. The methods for proving effectiveness of jackknife and bootstrap in the context of RCC are similar in the sense that both use the representation (7) and parallel results for CC.

Jackknifing RCC

In this resampling method, the replications are obtained by deleting one observation at each time. For $k=1, 2, ..., \mathscr{K}$; $j=1, ..., n$, let $X_{k, n, -j}$ and $T_{n, -j}$ denote the kth (in the sense described in the beginning of Section 2) squared sample CC and SRCC (respectively), based on $\{\mathbf{Y}_t : 1 \leqslant t \leqslant n, t \neq j\}$. Further, as in (5) and (6), let $A_{k, n, -j} = \{\omega : T_{n, -j}(\omega) = X_{k, n, -j}(\omega)\}$, except for appropriate disjointifications. Also define

$$B_n = A_{1, n} \bigcap_{i=1}^{n} A_{1, n, -i}. \tag{14}$$

Then we have the following lemma which provides the key to the proof of the main results in this section.

LEMMA 4. *Under Assumption* (3), $\mathbb{I}_{B_n} \xrightarrow{\text{a.s.}} 1$, *i.e.*, $\mathbb{P}[\liminf(B_n)] = 1$.

Proof. First observe that, by Corollary B.2.3 of Das [2],

$$\sup_{0 \leqslant j \leqslant n} |T_{n, -j} - \theta| \xrightarrow{\text{a.s.}} 0, \qquad \text{as} \quad n \to \infty. \tag{15}$$

Now, for any fixed $i=1, ..., \mathscr{K}$,

$$\sup_{0 \leqslant j \leqslant n} |X_{i, n, -j} - \rho_i^2| \xrightarrow{\text{a.s.}} 0, \qquad \text{as} \quad n \to \infty. \tag{16}$$

Let $\Omega_1 = \{\omega : \sup_{0 \leqslant j \leqslant n} |T_{n, -j}(\omega) - \theta| \to 0, \sup_{0 \leqslant j \leqslant n} |X_{i, n, -j}(\omega) - \rho_i^2| \to 0 \ \forall i\}$. By (15) and (16), $\mathbb{P}(\Omega_1) = 1$. Suppose, if possible, there exists $\omega_1 \in \Omega_1$, such that $\mathbb{I}_{B_n}(\omega_1) \nrightarrow 1$, as $n \to \infty$. Then for each n, there exists i_n such that $0 \leqslant i_n \leqslant n$ and $T_{n, -i_n}(\omega_1) \neq X_{1, n, -i_n}(\omega_1)$. Since K is finite, this implies the existence of a subsequence $\{n'\}$ and index $k_1 > 1$, with

$$T_{n', -i_{n'}}(\omega_1) = X_{k_1, n', -i_{n'}}(\omega_1).$$

But since $\omega_1 \in \Omega_1$, the left-hand side converges to θ as $n \to \infty$, and the right-hand side (RHS) converges to $\rho_{k_1}^2 \neq \theta$. This leads to a contradiction. ∎

The jackknife estimates of the bias and the variance of T_n are respectively given by

$$(n-1)\{\bar{T}_n^J - T_n\} = \hat{\mathbf{B}}^J(T_n), \qquad \frac{n-1}{n} \sum_{i=1}^{n} \{T_{n, -i} - \bar{T}_n^J\}^2 = \hat{V}_n^J(T_n),$$

where $\bar{T}_n^J = (1/n) \sum_{i=1}^{n} T_{n, -i}$. Similarly, define the corresponding quantities for sample CC's.

The effectiveness of a jackknife in the context of CC follows through the next two results (Das and Sen [4]).

THEOREM 2. *The bias-corrected jackknife estimate of the squared CC, $X^J_{i,n} = X_{i,n} - \hat{\mathbf{B}}^J(X_{i,n})$, is asymptotically normal with the same mean and asymptotic variance as the sample squared CC, i.e., $\sqrt{n}(X^J_{i,n} - \rho_i^2) \xrightarrow{\mathscr{L}} N(0, \eta_i^2)$.*

THEOREM 3. *A jackknife estimate of variance for the squared CC is strongly consistent, i.e., $n \times \hat{V}^J_n(X_{i,n}) \xrightarrow{\text{a.s.}} \eta_i^2$.*

Parallel to Theorems 2 and 3, the asymptotic normality of a bias-corrected jackknife estimate of RCC and the consistency of a jackknife estimate of variance of SRCC are established in the following theorems.

THEOREM 4. *The bias-corrected jackknife estimate of the squared RCC is asymptotically normal with the same mean and asymptotic variance as a usual sample squared RCC, i.e., $\sqrt{n}(T^J_n - \theta) \xrightarrow{L} N(0, \eta_1^2)$.*

Proof. Similar to the representation (7), we have

$$T_{n,-j} = \sum_{k=1}^{\mathscr{K}} \mathbb{I}_{A_{k,n,-j}} \times X_{k,n,-j}.$$

On $(A_{k,n} \bigcap_{i=1}^n A_{k_i,n,-i})$, $T_n = X_{k,n}$ and $T_{n,-i} = X_{k_i,n,-i}$. So

$$\begin{aligned}
\sqrt{n}\, T^J_n &= n^{-1/2} \sum_{i=1}^n \{nT_n - (n-1)\, T_{n,-i}\} \\
&= \mathbb{I}_{B_n} \times n^{-1/2} \sum_{i=1}^n \{nX_{1,n} - (n-1)\, X_{1,n,-i}\} + \mathbb{I}_{B_n^c} \times R_n,
\end{aligned}$$

where, noting that the CC's are bounded by 1,

$$\begin{aligned}
R_n = \sum_{\substack{k, k_1, ..., k_n \\ \text{not all 1's}}}^{\mathscr{K}} &\mathbb{I}_{[A_{k,n} \bigcap_{i=1}^n A_{k_i,n,-i}]} \\
&\times n^{-1/2} \sum_{i=1}^n \{nX_{k,n} - (n-1)\, X_{k_i,n,-i}\} = O_p(n^{5/2}).
\end{aligned}$$

Also, by Lemma 4 (and Lemma 2), $\mathbb{I}_{B_n} = 1 - o_p(1)$, and, hence, $\mathbb{I}_{B_n^c} = o_p(n^{-5/2})$ by Lemma 2. Thus, by Slutsky's theorem, the desired result follows from Theorem 2. ∎

THEOREM 5. *The jackknife estimate of variance of the squared RCC is strongly consistent, i. e.,* $n \times V_n^J \xrightarrow{\text{a.s.}} \eta_1^2$.

Proof. $n \times V_n^J = (n-1)\sum_{i=1}^n \{T_{n,-i} - (1/n)\sum_{j=1}^n T_{n,-j}\}^2 = \mathbb{I}_{B_n} \times nV_n^J(X_{1,n}) + \mathbb{I}_{B_n^c} \times R_n^*$, where $nV_n^J(X_1, n) = (n-1)\sum_{i=1}^n \{X_{1,n,-i} - (1/n)\sum_{j=1}^n X_{1,n,-j}\}^2$. And by the Theorem 3,

$$nV_n^J(X_{1,n}) \xrightarrow{\text{a.s.}} \eta_1^2. \tag{17}$$

Lemma 4 and (17) complete the proof. ∎

Bootstrap for RCC

To obtain appropriate confidence interval (C.I.) for RCC, we may note that by virtue of the implicit functional formulation of the RCC, there is no reliable and easy-to-compute approach to the estimation of the standard error of the SRCC, and hence, the standard bootstrap t-interval (see, for example, Efron [5] and Efron and Tibshirani [6]) may not be of much use. Of course, a two-level nested bootstrap as suggested by Efron and Tibshirani [6] can be implemented, but that will require a formidably large number of bootstrap replicates, and hence is not pursued here. Alternatively, in a computationally simpler way, C.I. can be obtained from the bootstrap percentiles, and we concentrate on this approach.

Given a set of sample observations $\mathbf{Y}(n) = \{\mathbf{Y}_1, ..., \mathbf{Y}_n\}$, the bootstrap resamples $\mathbf{Y}_1^*, ..., \mathbf{Y}_n^*$ are drawn i.i.d. with replacement from $\mathbf{Y}(n)$, or equivalently from the empirical distribution function F_n. The bootstrap replicates of SRCC or sample CC's (to be denoted respectively by, T_n^* and $X_{i,n}^*$'s) are simply the values of the relevant statistics based on $\mathbf{Y}_1^*, ..., \mathbf{Y}_n^*$. The above procedure can be repeated extensively to provide a large number (typically 1000 or higher) of replicates of the statistic, from which one can construct the bootstrap estimate of the entire distribution of the statistic. Appropriate (typically, equal exclusion probability on both tails) percentiles of this bootstrap distribution can serve as the upper and lower bound of the required C.I. In the following, a formal proof of this percentile method has been provided. Certain improvements of this method (namely, the BC_a and ABC methods of Efron and Tibshirani [6]) are available in the literature. But these modifications have not been dealt with in the current work, since the original percentile method seems to be quite succesful here.

Beran and Srivastava [1] showed that the bootstrapped sampling distribution of a smooth function (with continuous first derivative) of a sample covariance matrix converges to the right population distribution, i.e., if $J_n(F) = \mathbb{L}[\sqrt{n}\{g(\mathbf{S}_n) - g(\mathbf{\Sigma})\} \mid F]$ and g is continuously differentiable, then $J_n(F_n) \to \mathcal{N}(0, \Omega_F)$, which is the limit of $J_n(F)$ as $n \to \infty$. Hence it is reasonable to estimate the sampling distribution of $g(\mathbf{S}_n)$ by the appropriate bootstrapped distribution.

The required smoothness of CC as a function of sample covariance can be shown (viz., Das and Sen [4]). Hence we have

$$\mathbb{L}[\sqrt{n}(X^*_{i,n} - X_{i,n}) \mid \mathbf{Y}(n)] \xrightarrow{\text{a.s.}} \mathcal{N}(0, \eta_i^2). \tag{18}$$

Note that, the RHS of (18) is simply the limiting distribution of $\sqrt{n}(X_{i,n} - \rho_i^2)$. The next theorem shows the consistency of the bootstrapped sampling distribution for SRCC.

THEOREM 6. *Under assumption* (3), *the bootstrapped distribution of SRCC is strongly consistent, i.e.*,

$$\mathbb{L}[\sqrt{n}(T^*_n - T_n) \mid \mathbf{Y}(n)] \xrightarrow{\text{a.s.}} \mathcal{N}(0, \eta_1^2). \tag{19}$$

Proof. Denote the normal cumulative distribution function in the RHS of (19) by $G(\cdot)$. By (18), $\exists \Omega^{(1)} \subseteq \Omega$ s.t. $\forall \omega \in \Omega^{(1)}$, $\forall x$,

$$\mathbb{P}[\sqrt{n}(X^*_{1,n} - X_{1,n}) \leqslant x \mid \mathbf{Y}(n)](\omega) \to G(x) \qquad \text{as} \quad n \to \infty. \tag{20}$$

Let $\Omega^* = \Omega^{(1)} \cap \{\omega : \mathbb{I}_{A_{1,n}}(\omega) \to 1\}$. By Lemma 4 and (18), it follows that $\mathbb{P}(\Omega^*) = 1$. If $\omega \in \Omega^*$, then $T_n(\omega) = X_{1,n}(\omega)$ for sufficiently large n. By Theorem 1, under any i.i.d. sampling scheme, the asymptotic distribution of SRCC is the same as that of appropriate sample CC. Applying this on the bootstrap resampling procedure (where population RCC and the first CC, identify with T_n and $X_{1,n}$, respectively), it follows that

$$\begin{aligned}
&\lim_{n \to \infty} \mathbb{P}[\sqrt{n}(T^*_n - T_n) \leqslant x \mid \mathbf{Y}(n)](\omega) \\
&\qquad = \lim_{n \to \infty} \mathbb{P}[\sqrt{n}(X^*_{1,n} - X_{1,n}) \leqslant x \mid \mathbf{Y}(n)](\omega),
\end{aligned}$$

and the RHS is equal to $G(x)$ by (20). ∎

4. NUMERICAL AND SIMULATION STUDIES

In this section, numerical illustrations of Examples 1 and 2 are provided, and simulating observations from multivariate normal (MN) and multivariate Poisson (MP) distributions, the relative performances (with regard to variance estimation and reduction of bias) of jackknifing and bootstrapping, are studied.

EXAMPLE 1 (Continued). The scores in different exams of eight students who took both the French and mathematics courses are shown in Table I:

TABLE I

Example 1

	Mathematics				French		
Student	HW	Quiz	Midterm	Final	Test1	Test2	Test3
1	87	56	78	69	86	74	79
2	91	66	82	81	83	79	81
3	90	70	87	94	88	85	93
4	55	48	60	54	71	56	63
5	76	72	75	79	68	63	70
6	98	90	90	96	91	93	78
7	80	75	73	84	68	77	85
8	92	83	80	74	65	60	72

The first sample CC is 0.9985 while the SRCC with restrictions (1)–(2) is 0.8651. The rescaled coefficients for both these are listed in Table II. At first look, the zero weights for the SRCC, may seem counterintuitive. But, by Lemma 1, that is likely to happen with nonnegativity type restrictions. The jackknife estimate of the standard error of SRCC was found to be 0.0962. The bootstrap methods could not be implemented because small sample size leads to singularity of the covariance matrix.

Example 2 (Revisited). Table III shows the measurements of 16 subjects for the two variables. The sample CC turns out to be 0.9475 with canonical coefficients $(0.23, 0.08, 0.01, -0.09, -0.92)'$ for the time-on-target/variables and $(-0.09, -0.06, 0.09, 0.28, 0.17)'$ for the hits-on-target variables. Because the cross-covariance matrix ($\mathbf{S}_{12}$) has all entries negative, by Lemma 7 of Das and Sen [3], SRCC $= -0.6837$, which is the largest element of the cross-correlation matrix. But clearly in this case one should consider *minimum* RCC (i.e., the minimum correlation between non-negative linear combinations); in effect this amounts to replacing $\mathbf{S}_{12}$ by

TABLE II

Optimal Coefficients in Example 1

Subject	Test	Canonical coefficient	Restricted can. coef.
Mathematics	Homework	−2.2437	0
	Quiz	−0.6964	0
	Midterm	6.3089	0.4929
	final	−2.3688	0.5071
French	Test1	4.4133	0
	Test2	−3.8266	0.5714
	Test3	0.4133	0.4286

TABLE III

Example 2

Ind.	Time on target/s 1	2	3	4	5	Hits on target 1	2	3	4	5
1	13.95	12.00	14.20	14.40	13.00	31.50	37.50	36.50	35.50	34.00
2	18.15	22.60	19.30	18.25	20.45	22.50	12.00	17.50	19.00	16.50
3	19.65	21.60	19.70	19.55	21.00	18.50	18.00	21.50	18.50	14.50
4	20.80	21.35	21.25	21.25	20.90	20.50	18.50	17.00	16.50	16.50
5	17.80	20.05	20.35	19.80	18.30	29.00	21.00	19.00	23.00	21.00
6	17.35	20.85	20.95	20.30	20.70	22.00	15.50	18.00	18.00	22.50
7	16.50	16.70	19.25	16.25	18.55	36.00	29.50	22.00	26.00	25.50
8	19.10	18.35	22.95	22.70	22.65	18.00	9.50	10.50	10.50	14.50
9	12.05	15.40	14.75	13.45	11.60	28.00	30.50	37.50	31.50	28.00
10	8.55	9.00	9.10	10.50	9.55	36.00	37.00	36.00	36.00	33.00
11	7.35	5.85	6.20	7.05	9.15	33.50	32.00	33.00	32.50	36.50
12	17.85	17.95	19.05	18.40	16.85	23.00	26.00	20.00	21.50	30.00
13	14.50	17.70	16.00	17.40	17.10	31.00	31.50	33.00	26.00	29.50
14	22.30	22.30	21.90	21.65	21.45	16.00	14.00	16.00	19.50	18.00
15	19.70	19.25	19.85	18.00	17.80	32.00	22.50	24.00	30.00	26.50
16	13.25	17.40	18.75	18.40	18.80	23.50	24.00	22.00	20.50	21.50

$-\mathbf{S}_{12}$. The minimum SRCC turns out to be -0.9380, with the weights (standardized, to make the sum 1) are (0, 0, 0, 0, 1) for the time-on-target/ variables and (0, 0, 0.3435, 0.2365, 0.42) for the hits-on-target variables. The jackknife and bootstrap estimates of the standard error of SRCC are 0.0275 and 0.0220, respectively. The 90% as obtained from the bootstrap percentile method is found to be (0.9203, 0.9921).

Next, we discuss simulation results regarding the distribution of SRCC and effectiveness of the jackknife and bootstrap estimates of the bias and variance in finite-sample cases. The asymptotic distribution of SRCC seems to be very close to the normal distribution for sample sizes of 500 or higher in most cases (in some cases, even for moderate sample sizes like 200). The following two correlation matrices were among those which were considered for simulation purpose:

$$\Sigma_1 = \left[\begin{array}{cc|ccc} 1 & & & & \\ 1/2 & 1 & & & \\ \hline 1/6 & 1/6 & 1 & & \\ 1/6 & 1/6 & 1/3 & 1 & \\ 1/6 & 1/6 & 1/3 & 1/3 & 1 \end{array}\right], \quad \Sigma_2 = \left[\begin{array}{cc|ccc} 1 & & & & \\ 0.4 & 1 & & & \\ \hline 0.3 & 0 & 1 & & \\ 0.15 & 0.35 & 0.3 & 1 & \\ 0.15 & 0.25 & 0.4 & 0.2 & 1 \end{array}\right].$$

$\boldsymbol{\Sigma}_1$ is a typical case, where the population RCC and the first population CC are the same (here, equal to $1/\sqrt{15} \approx 0.2582$); $\boldsymbol{\Sigma}_2$ provides an example where that is not the case (RCC ≈ 0.3968, whereas first CC is ≈ 0.4685). Experiments were performed with continuous as well as discrete data. In the figures, we summarized the performances of the resampling estimators when simulated data was generated from (i) MN distribution with covariance matrix $\boldsymbol{\Sigma}_1$ (ii) MP distribution with covariance matrix $\boldsymbol{\Sigma}_2$. A $(p+q)$-variate MR distribution (see Teicher [9]) is characterized by $2^{p+q}-1$ parameters and to retain generality the data from such a distribution should be generated from the same number of independent univariate Poisson random variables. However, a simpler model is considered here for generating the MP random vectors by using only $(p+q)(p+q-1)/2$ univariate Poisson variables which account for the second moments of the distribution.

Comparing Bias Reduction

In Fig. 1 and 2 the performances of the bias-corrected jackknife and bootstrap estimates of RCC are compared with the original estimator

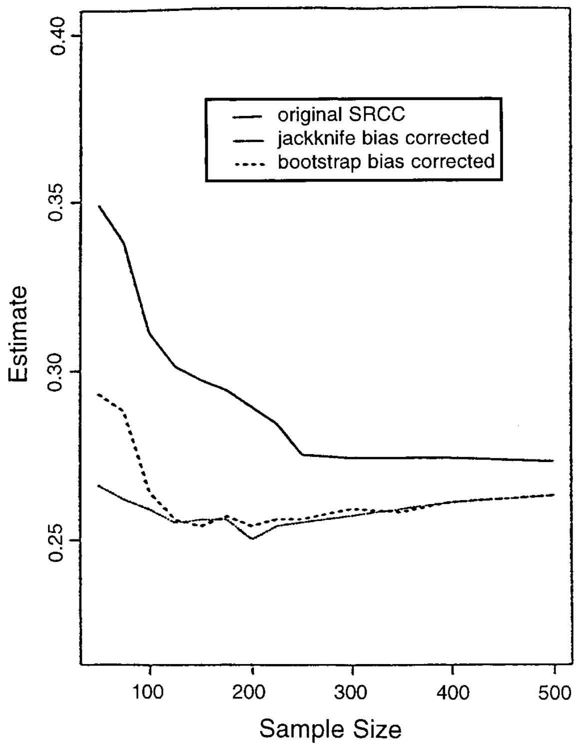

FIG. 1. Estimates of RCC; MN–Σ_1

(SRCC). The formula of jackknife of bias is described in Section 3. The bootstrap estimate of bias is given by $\hat{\mathbb{E}}_{F_n}(T_n^*) - \theta(F_n)$. Hence the bias corrected bootstrap estimate of RCC is $2T_n - \hat{\mathbb{E}}_{F_n}(T_n^*)$, and the second term is approximated by $\bar{T}_n^* = (1/B)\sum_{i=1}^{B} T_{n,i}^*$, where $T_{n,i}^*$'s are the bootstrap replicates. These simulations show that the SRCC has significant positive bias for smaller sample sizes. The curve corresponding to "original SRCC" is actually obtained by taking averages over large (5000) simulated realizations of SRCC, for different sample sizes. Of course, the bias goes down with increasing sample size. Jackknife does an excellent job of estimating this bias and, hence, the bias-corrected jackknife estimates are remarkably closer to the target even for small sample sizes. The performance of the bootstrap is quite comparable for larger sample sizes, although for small samples jackknife seems to be slightly more effective.

Comparing Performances of the Resampling Methods in Estimating Sample Variance and MSE.

Next, the performances of the jackknife and the bootstrap estimate of variance for SRCC and sample CC are compared for finite sample sizes.

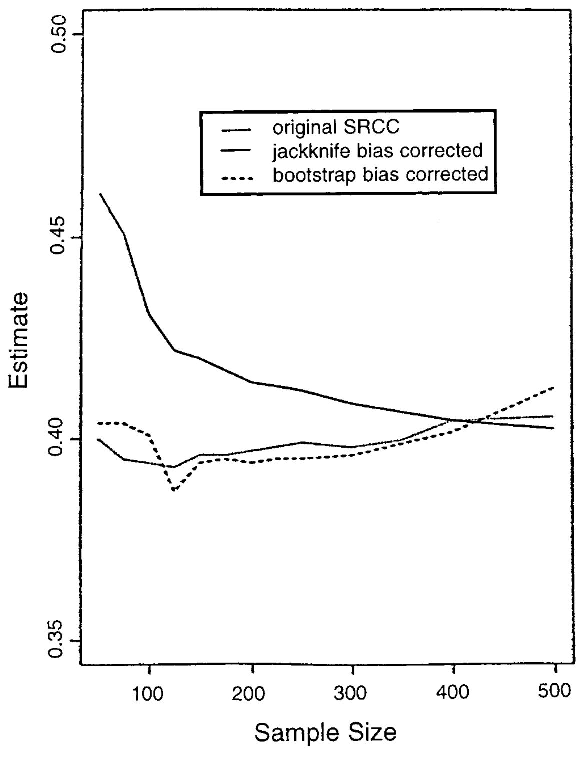

FIG. 2. Estimates of RCC; MP–Σ_2

The bootstrap estimate of variance is taken as $(1/(B-1))\sum_{i=1}^{B}(T_{n,i}^{*}-\bar{T}_{n}^{*})^{2}$.

Since in most cases MSE, rather than the sample variance are of primary importance, MSE (instead of variance) estimates are shown in Fig. 3 and 4. (It may be added here that looking solely at variance figures, the performances of both jackknife and bootstrap estimators are found to be very good, especially for sample sizes of 500 and higher. For smaller sample sizes, bootstrap does a somewhat better job, although possibly at somewhat higher computational cost.) The MS estimators via jackknife and bootstrap are calculated from the respective bias and variance estimates provided earlier. For the jackknife and bootstrap estimates, the points represent the relevant values averaged over 50 simulations. Bootstrap resample size is taken to be 1000. The simulated MSE is obtained from 5000 simulated values of the statistic. These figures show that the performance of the jackknife and the bootstrap are quite comparable in this context.

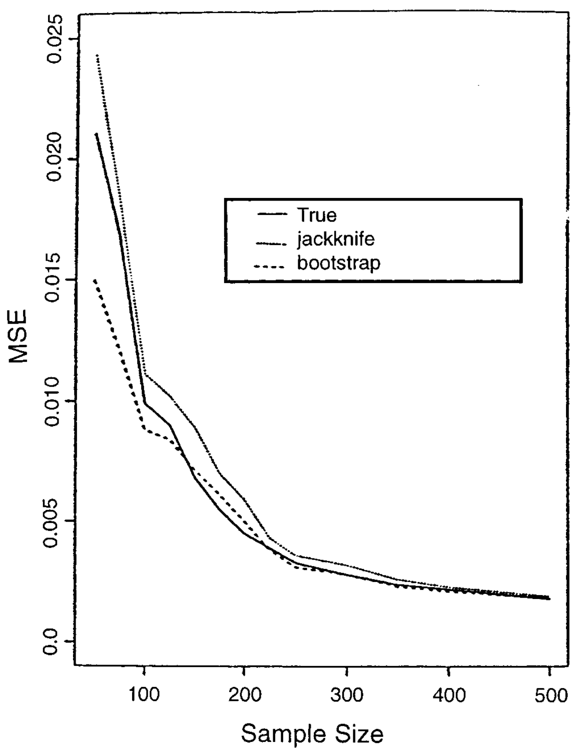

FIG. 3. Estimates of MSE; MN–Σ_1

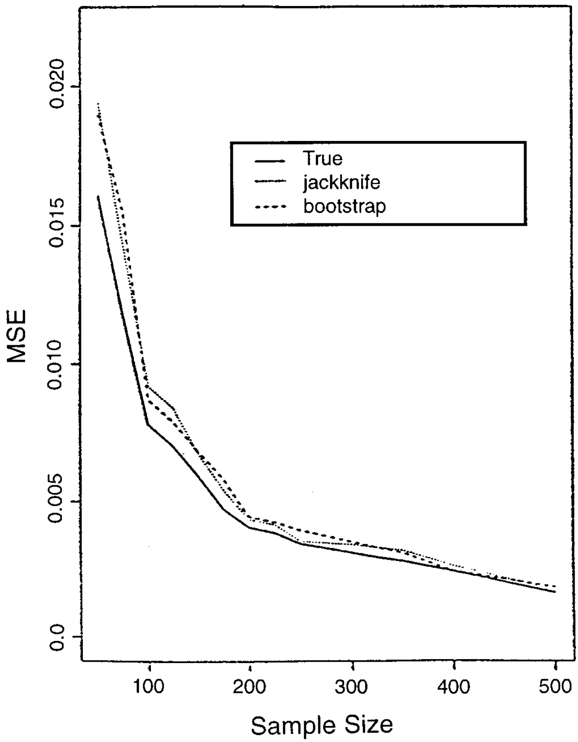

FIG. 4. Estimates of MSE; MP–Σ_2

Concluding Remarks

Apparently, it may seem that the resampling plans may not be necessary for RCC since the asymptotic variance for SECC is the same as that for an appropriate sample CC, and the latter is of known form (viz., (9)). But, unfortunately there are hidden impasses in such a direct approach. First, the expression in (9) is quite complicated and simpler estimates are not that apparent. Secondly appropriate moment estimates are needed to be plugged into (9) in order to have such a natural estimate. Finally, the most important complication arises from the fact that $\check{k}$ is itself an unknown parameter, and, hence, it needs to be estimated from the sample. Therefore, such a direct estimation scheme seems to be very "unstable." On all counts, the proposed resampling methods are more convenient.

ACKNOWLEDGMENT

The authors are grateful to both the referees for their careful reading and constructive suggestions.

References

[1] Beran, R., and Srivastava, M. S. (1985). Bootstrap tests and covariance regions for functions of a covariance matrix. *Ann. Statist.* **13** 95–115.

[2] Das, S. (1993). *Restricted Canonical Correlations*. Doctoral dissertation, Mimeo Series 2095, The Institute of Statistics, University of North Carolina at Chapel Hill.

[3] Das, S., and Sen, P. K. (1994). Restricted canonical correlations. *Linear Algebra Appl.* **210** 29–47.

[4] Das, S., and Sen, P. K. (1995). Simultaneous spike-trains and stochastic dependence. *Sankhya Ser. B* **57**, in press.

[5] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

[6] Efron, B., and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

[7] Muirhead, R. J., and Waternaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for non-normal populations. *Biometrika* **67** 31–43.

[8] Quenouille, M. (1949). Approximate tests of correlation in time series. *J. Roy. Statist. Soc. Ser. B* **11** 68–84.

[9] Teicher, H. (1975). On the multivariate Poisson distribution. *Skand. Aktua.* **37** 1–9.

[10] Timm, N. H. (1975). *Multivariate Analysis with Applications in Education and Psychology*. Brooks–Cole, Monterey, CA. 1975.