Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

STEPDOWN LIKELIHOOD RATIO TEST ON EACH PARAMETER COMPONENT IN TESTING EQUALITY OF COVARIANCE MATRICES

BY AKIMICHI TAKEMURA
University of Tokyo

We consider the likelihood ratio test for testing equality of covariance matrices of k multivariate normal populations $N_p(\mu_h, \Sigma_h)$, $h = 1, \dots, k$. The null hypothesis is $H_0 : \Sigma_1 = \cdots = \Sigma_k$. The likelihood ratio test is well known and the stepdown test procedure for the case $k = 2$ was given by J. Roy (1958). See also Sec.10.4 of Anderson (1984). The stepdown procedure can be regarded as a decomposition of likelihood ratio statistic. Here we demonstrate how this decomposition can be carried out to test each component of the covariance matrix Σ for the k sample problem.

1. Overview of the Stepdown Likelihood Ratio Test. Consider a general hypothesis testing problem

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad K : \theta \in \Theta. \tag{1}$$

For simplicity of notation we write $K : \theta \in \Theta$ instead of more usual $K : \theta \in \Theta - \Theta_0$ throughout this paper. Often we want to test an intermediate hypothesis or partial null hypothesis $H_1 : \theta \in \Theta_1$, where

$$\Theta_0 \subset \Theta_1 \subset \Theta. \tag{2}$$

Let $\lambda = \max_{\theta \in \Theta_0} f(x, \theta) / \max_{\theta \in \Theta} f(x, \theta)$ be the likelihood ratio statistic for (1) and similarly let $\lambda_{01}, \lambda_{12}$ be the likelihood ratio statistic for testing H_0 vs. H_1 and H_1 vs. K respectively. Then the overall likelihood ratio statistic λ can be decomposed as $\lambda = \lambda_{01}\lambda_{12}$.

Instead of testing H_0 vs. K, we could test each of the partial testing problems H_0 vs. H_1, H_1 vs. K in turn, using the component likelihood ratio statistics λ_{01} and λ_{12}. Usually the intermediate hypothesis H_1 is taken to be a hypothesis on some subvector of θ and then the above decomposition of likelihood ratio test is called stepdown procedure.

AMS 1980 Subject Classifications: 62H10, 62H15.
Key words: Stepdown test, Wishart distribution, covariance matrix.

With appropriately chosen intermediate hypothesis H_1, it often happens that λ_{01} and λ_{12} are mutually independently distributed under H_0. In this case the overall significance level of stepdown procedure can be easily computed. Therefore it is advantageous to take H_1 to achieve this independence under H_0.

When stepdown procedure is used, the order of tests has to be considered. The general principle seems to be that we first test the outer problem H_1 vs. K and then test the inner problem H_0 vs. H_1. More precisely:

1. Test H_1 vs. K using λ_{12}. If H_1 is rejected, then H_0 is rejected as well and we stop.

2. If H_1 is accepted, then we continue to test H_0 vs. H_1 using λ_{01}.

For the case of determining the order of polynomial regression, the optimality of the above "backward" order of testing is proved in Sec 3.2 of Anderson (1971).

We now describe stepdown procedure for testing equality of covariance matrices. Decomposition of the overall likelihood ratio statistic will be given in terms of submatrices of the covariance matrix Σ for simplicity. Let the covariance matrix Σ_h for the h-th population be partitioned as

$$\Sigma_h = \begin{pmatrix} \Sigma_{11,h} & \Sigma_{12,h} \\ \Sigma_{21,h} & \Sigma_{22,h} \end{pmatrix}, \qquad h = 1, \cdots, k. \tag{3}$$

Let

$$B_h = \Sigma_{11,h}^{-1}\Sigma_{12,h}, \qquad \Sigma_{22\cdot 1,h} = \Sigma_{22,h} - \Sigma_{21,h}\Sigma_{11,h}^{-1}\Sigma_{12,h}$$

be the regression coefficient matrix and the residual covariance matrix. Since $(\Sigma_{11,h}, \Sigma_{21,h}, \Sigma_{22,h})$ and $(\Sigma_{11,h}, B_h, \Sigma_{22\cdot 1,h})$ are in 1-to-1 relation, we can use the parametrization $(\Sigma_{11,h}, B_h, \Sigma_{22\cdot 1,h})$. We remark here that this reparametrization is advantageous in achieving independence of component likelihood ratio statistics, but the physical interpretation of new parameters are not necessarily simple. Consider the following 3 hypotheses:

$$\begin{aligned} H_{(11)} &: \ \Sigma_{11,1} = \cdots = \Sigma_{11,k}, \\ H_{(21)} &: \ B_1 = \cdots = B_k, \\ H_{(22)} &: \ \Sigma_{22\cdot 1,1} = \cdots = \Sigma_{22\cdot 1,k}. \end{aligned} \tag{4}$$

Then the null hypothesis H_0 is the intersection of the above 3 hypotheses:

$$H_0 = H_{(11)} \cap H_{(21)} \cap H_{(22)},$$

where $\cap$ denotes logical intersection. Let $\Theta_0, \Theta_{(11)}$, $\Theta_{(21)}, \Theta_{(22)}$ be the restricted parameter space corresponding to H_0, $H_{(11)}$, $H_{(21)}$,and $H_{(22)}$, respectively. Then $H_0 : \theta \in \Theta_0 = \Theta_{(11)} \cap \Theta_{(21)} \cap \Theta_{(22)}$. Now we form the following nested sequence of hypotheses

$$\Theta_0 \subset \Theta_{(11)} \cap \Theta_{(22)} \subset \Theta_{(11)} \subset \Theta, \tag{5}$$

where Θ is the whole parameter space. This nesting of partial hypotheses corresponds to the following ordering of testing:

1. Test $H_{(11)}$.
2. If $H_{(11)}$ is accepted then test $H_{(22)}$.
3. If $H_{(22)}$ is accepted then test $H_{(21)}$.

Let λ_{11} , λ_{22}, and λ_{21} be the likelihood ratio statistics for these intermediate testing problems and $\lambda = \lambda_{11}\lambda_{22}\lambda_{21}$ be the overall likelihood ratio statistic. With this choice of intermediate hypotheses, the component likelihood ratio statistics $\lambda_{11}, \lambda_{21}, \lambda_{22}$ are mutually independently distributed under the null hypothesis (see Lemma 1 below). Testing $H_{(11)}$ first seems to be natural, if the components of the first subvector are considered to be more important.

In the sequel we adopt the above ordering (5) of intermediate subhypotheses. However we remark here that there are other possible orderings to achieve independence of component likelihood ratio statistics under H_0. They are

$$\Theta_0 \subset \Theta_{(11)} \cap \Theta_{(22)} \subset \Theta_{(22)} \subset \Theta, \tag{6}$$

$$\Theta_0 \subset \Theta_{(21)} \cap \Theta_{(22)} \subset \Theta_{(22)} \subset \Theta. \tag{7}$$

The component likelihood ratio statistics $\lambda_{11}, \lambda_{22}$ and λ_{21} remain the same for these orderings. The only requirement for independence is that we have to test $H_{(22)}$ before $H_{(21)}$. This means that 1) $B_h, h = 1, \dots, k$ have to be free when $H_{(22)}$ is tested, and 2) $\Sigma_{22\cdot 1,h}$ $h = 1, \dots, k$ have to be equal, when $H_{(21)}$ is tested. We describe this situation by saying that λ_{11} is the likelihood ratio statistic for $H_{(11)}$, λ_{22} is the likelihood ratio statistic for $H_{(22)}$ not assuming $H_{(21)}$, and λ_{21} is the likelihood ratio test for $H_{(21)}$ assuming $H_{(22)}$.

The explicit form of $\lambda_{11}, \lambda_{22}, \lambda_{21}$ is well known. Let W_h $(h = 1, \dots, k)$ be the sample sum of squares matrices from the h-th population. W_h is distributed according to Wishart distribution: $W_h \sim W_p(n_h, \Sigma_h)$, where $n_h = N_h - 1$ is the degrees of freedom and N_h is the sample size from h-th population. Throughout this paper we assume $n_h \geq p$, $h = 1, \cdots, k$ and $|\Sigma_h| \neq 0$ for simplicity, although the restriction on n_h can be relaxed. Let W_h be partitioned as

$$W_h = \begin{pmatrix} W_{11,h} & W_{12,h} \\ W_{21,h} & W_{22,h} \end{pmatrix}, \qquad h = 1, \cdots, k. \tag{8}$$

and let $W_{22\cdot 1,h} = W_{22,h} - W_{21,h} W_{11,h}^{-1} W_{12,h}$, $h = 1, \cdots, k$. Denote the pooled sum of squares matrix by

$$W_T = W_1 + \cdots + W_k = \begin{pmatrix} W_{11,T} & W_{12,T} \\ W_{21,T} & W_{22,T} \end{pmatrix} \tag{9}$$

and let $W_{22\cdot 1,T} = W_{22,T} - W_{21,T} W_{11,T}^{-1} W_{12\cdot T}$. For a ready reference we give explicit expressions for λ, $\lambda_{11}, \lambda_{22}, \lambda_{21}$ in the following lemma.

LEMMA 1. *Let $N = N_1 + \cdots + N_k$. Assume $n_h = N_h - 1 \geq p, h = 1, \cdots, k$, and $|\Sigma_h| \neq 0, h = 1, \cdots, k$. Then*

$$\lambda = \frac{\prod_{h=1}^{k} |W_h|^{N_h/2}}{|W_T|^{N/2}}, \qquad \lambda_{11} = \frac{\prod_{h=1}^{k} |W_{11,h}|^{N_h/2}}{|W_{11,T}|^{N/2}},$$
$$\lambda_{22} = \frac{\prod_{h=1}^{k} |W_{22\cdot1,h}|^{N_h/2}}{\left|\sum_{h=1}^{k} W_{22\cdot1,h}\right|^{N/2}}, \qquad \lambda_{21} = \frac{\left|\sum_{h=1}^{k} W_{22\cdot1,h}\right|^{N/2}}{|W_{22\cdot1,T}|^{N/2}}. \tag{10}$$

$\lambda_{11}, \lambda_{22}, \lambda_{21}$ *are mutually independently distributed under H_0.*

Note that λ_{11} and λ_{22} are of the same form as λ. λ_{11} is based on the $(1,1)$ block of W_h, whereas λ_{22} is based on $(2,2)$ residual sum of squares block of W_h.

2. The Main Result. In Lemma 1 stepdown procedure was described in terms of submatrices of the covariance matrix. In this section we carry out the decomposition down to each component of $\Sigma = (\sigma_{ij})$. The ordering of our stepdown procedure is to test the following elements of Σ in turn:

$$(1,1) \longrightarrow (2,2) \longrightarrow (2,1) \longrightarrow (3,3) \longrightarrow (3,2) \longrightarrow (3,1) \longrightarrow (4,4) \longrightarrow \cdots.$$

However as mentioned in the previous section, the only essential restriction on the ordering is that in each row we proceed as

$$(i,i) \longrightarrow (i,i-1) \longrightarrow \cdots \longrightarrow (i,1).$$

For the moment we omit the subscript h for notational simplicity. (We add "$,h$" to the subscript to denote quantities for h-th population.) Let

$$\Sigma_{ii} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1i} \\ \vdots & \cdots & \vdots \\ \sigma_{i1} & \cdots & \sigma_{ii} \end{pmatrix}, \qquad \sigma_{(i)} = (\sigma_{i1}, \cdots, \sigma_{i,i-1}),$$
$$\beta'_{(i)} = (\beta_{i1}, \cdots, \beta_{i,i-1}) = \sigma_{(i)} \Sigma_{i-1,i-1}^{-1},$$
$$\sigma_{ii\cdot1,\cdots,i-1} = \sigma_{ii\cdot} = \sigma_{ii} - \sigma_{(i)} \Sigma_{i-1,i-1}^{-1} \sigma'_{(i)}. \tag{11}$$

Then $\Sigma = (\sigma_{ij})$ and $(\sigma_{ii\cdot},\ 1 \leq i \leq p,\ \beta_{ij}, i > j)$ are in 1-to-1 relation and we use the latter parametrization. Consider the following set of hypotheses:

$$H_{ii}: \ \sigma_{ii\cdot,1} = \cdots = \sigma_{ii\cdot,k}, \qquad i = 1, \cdots, p, \tag{12}$$
$$H_{ij}: \ \beta_{ij,1} = \cdots = \beta_{ij,k}, \qquad 1 \leq j < i \leq p \tag{13}$$

and let Θ_{ij} be the parameter space corresponding to H_{ij}, $i \geq j$. Further write

$$\bar{H}_{ij} = H_{ij} \cap H_{i,j+1} \cap \cdots \cap H_{ii}, \qquad \bar{\Theta}_{ij} = \Theta_{ij} \cap \Theta_{i,j+1} \cap \cdots \cap \Theta_{ii}. \tag{14}$$

Let $W_h = (w_{ij,h})$ and define $w_{ii\cdot,h} = w_{ii\cdot 1,\cdots,i-1,h}$ as in (11). The likelihood ratio statistic for H_{ii} not assuming $H_{i1}, \cdots, H_{i,i-1}$ can be easily described in terms of $w_{ii\cdot,h}$, $h = 1, \cdots, k$. Now we consider testing $\bar{H}_{ij}$ vs. $\bar{H}_{i,j+1}$, i.e., testing H_{ij} assuming $H_{i,j+1}, \cdots, H_{ii}$ and not assuming $H_{i1}, \cdots, H_{i,j-1}$. Likelihood ratio statistic for this problem needs somewhat complicated notation. Let the $i \times i$ upper left block of W_h be partitioned as

$$\begin{array}{cc} & \begin{array}{cc} j & i-j \end{array} \\ \begin{array}{c} j \\ i-j \end{array} & \begin{pmatrix} W_{11,h} & W_{12,h} \\ W_{21,h} & W_{22,h} \end{pmatrix} \end{array} \tag{15}$$

and let

$$\begin{aligned} W_h(i \,|\, j) &= W_{22,h} - W_{21,h} W_{11,h}^{-1} W_{12,h}, \qquad 1 \leq j \leq i-1, \\ \widetilde{W}(i \,|\, j) &= \sum_h W_h(i \,|\, j). \end{aligned} \tag{16}$$

$\widetilde{W}(i \,|\, 0)$ is just the $i \times i$ upper left block of W_T. Note that $\widetilde{W}(i \,|\, j)$ is obtained by first subtracting off the regression onto first j elements in each of W_h and then pooling the residual sum of squares. Now regress the last ("i-th") element of $\widetilde{W}(i \,|\, j)$ onto other elements and let

$$\tilde{w}(i \,|\, j)_{ii\cdot} = \tilde{w}_{22} - \tilde{w}_{21} \widetilde{W}_{11}^{-1} \tilde{w}_{12}, \qquad 1 \leq j \leq i-2, \tag{17}$$

where

$$\widetilde{W}(i \,|\, j) = \begin{array}{c} \\ i-j-1 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} i-j-1 & 1 \end{array} \\ \begin{pmatrix} \widetilde{W}_{11} & \tilde{w}_{12} \\ \tilde{w}_{21} & \tilde{w}_{22} \end{pmatrix} \end{array}. \tag{18}$$

For $j = i-1$ or $j = 0$ define

$$\begin{aligned} \tilde{w}(i \,|\, i-1)_{ii\cdot} &= \sum_{h=1}^{k} w_{ii\cdot,h}, \\ \tilde{w}(i \,|\, 0)_{ii\cdot} &= w_{ii\cdot,T} = w_{ii\cdot 1,\cdots,i-1,T}. \end{aligned} \tag{19}$$

Finally consider the difference in residual sum of squares $\tilde{w}(i \,|\, j)_{ii\cdot}$ and $\tilde{w}(i \,|\, j-1)_{ii\cdot}$ and let

$$u_{ij} = \tilde{w}(i \,|\, j-1)_{ii\cdot} - \tilde{w}(i \,|\, j)_{ii\cdot}\,. \tag{20}$$

Note that

$$\begin{aligned} \tilde{w}(i \,|\, j)_{ii\cdot} &= u_{i,j+1} + \cdots + u_{i,i-1} + \tilde{w}(i \,|\, i-1)_{ii\cdot} \\ &= u_{i,j+1} + \cdots + u_{i,i-1} + \sum_{h=1}^{k} w_{ii\cdot,h}. \end{aligned} \tag{21}$$

The main result of this paper is the following theorem.

THEOREM 1. *Assume $n_h \geq p$, $h = 1, \cdots, n$, and $|\Sigma_h| \neq 0, h = 1, \cdots, k$. Let λ_{ii} be the likelihood ratio statistic for testing $\bar{H}_{ii}$ not assuming $H_{i1}, \cdots, H_{i,i-1}$ and let $\lambda_{ij}, i > j$, be the likelihood ratio statistic for testing $\bar{H}_{ij}$ vs. $\bar{H}_{i,j+1}$, namely testing H_{ij} assuming $H_{i,j+1}, \cdots, H_{ii}$ and not assuming $H_{i1}, \cdots, H_{i,j-1}$. Then*

$$\lambda_{ii} = \frac{\prod_{h=1}^{k} w_{ii\cdot,h}^{N_h/2}}{\left(\sum_{h=1}^{k} w_{ii\cdot,h}\right)^{N/2}}, \tag{22}$$

$$\begin{aligned}\lambda_{ij} &= \left(\frac{\tilde{w}(i\,|\,j)_{ii\cdot}}{\tilde{w}(i\,|\,j-1)_{ii\cdot}}\right)^{N/2} \\ &= \left(\frac{u_{i,j+1} + \cdots + u_{i,i-1} + \sum_{h=1}^{k} w_{ii\cdot,h}}{u_{ij} + u_{i,j+1} + \cdots + u_{i,i-1} + \sum_{h=1}^{k} w_{ii\cdot,h}}\right)^{N/2}.\end{aligned} \tag{23}$$

Under H_0, $\sum_{h=1}^{k} w_{ii\cdot,h}$, $1 \leq i \leq p$ and u_{ij}, $i < j$, are all mutually independently distributed and

$$\frac{w_{ii\cdot,h}}{\sigma_{ii\cdot}} \sim \chi^2(N_h - i), \qquad \frac{u_{ij}}{\sigma_{ii\cdot}} \sim \chi^2(k-1), \tag{24}$$

where $\chi^2(f)$ denotes chi-square distribution with f degrees of freedom.

COROLLARY 1. *Under H_0, λ_{ij}, $i \geq j$, are all mutually independently distributed. Furthermore for $i > j$*

$$\lambda_{ij}^{2/N} \sim \mathrm{Beta}\left([N - ki + (i-j-1)(k-1)]/2,\ (k-1)/2\right), \tag{25}$$

where Beta(a, b) *denotes beta distribution with parameter a, b.*

Proof of Theorem 1 will be given in the next section. Corollary 1 is an easy consequence of Theorem 1.

3. Lemmas and Proofs. Our Theorem 1 is a refinement of Lemma 1 and the method of proof for Lemma 1 and Theorem 1 are basically the same. For clarity of argument, we first prove Lemma 1 and then extend the proof to Theorem 1. Proof of Lemma 1 is based on the author's argument in Takemura(1991). We divide our proof into 2 parts: derivation of likelihood ratio statistics and derivation of distributional results under H_0.

We begin by deriving likelihood ratio tests given in Lemma 1.

3.1. Derivation of Likelihood Ratio Statistics in Lemma 1. Let $Z_h \sim N(1_{N_h}\mu_h', I_{N_h} \otimes \Sigma), h = 1, \cdots, k$, be $N_h \times p$ observation matrix from h-th population, where $1_n = (1, \cdots, 1)' \in R^n$, i.e., the rows of Z_h are independently distributed according to $N_p(\mu_h, \Sigma)$. Let $\bar{z}_h = (1/N_h)Z_h' 1_{N_h}$ be sample the mean vector. The sample sum of squares matrix W_h is $W_h = Z_h' Z_h - N_h \bar{z}_h \bar{z}_h'$. The maximum likelihood estimator for μ_h is $\bar{z}_h$ in any case and concentrated likelihood ignoring irrelevant constants is

$$L \propto \prod_{h=1}^{k} |\Sigma_h|^{-N_h/2} \exp\left(-\frac{1}{2} \sum_{h=1}^{k} \operatorname{tr} W_h \Sigma_h^{-1} \right). \tag{26}$$

Dropping subscript h for the moment, consider partitioning W and Σ as in (3) and (8). Let $B = \Sigma_{11}^{-1}\Sigma_{12}$. Denote the (i, j) block of Σ^{-1} by Σ^{ij}. Using the well known relation

$$\Sigma^{22} = \Sigma_{22\cdot 1}^{-1}, \qquad \Sigma^{21} = -\Sigma_{22\cdot 1}^{-1} B', \qquad \Sigma^{11} = \Sigma_{11}^{-1} + B\Sigma_{22\cdot 1}^{-1} B',$$

we can express $\operatorname{tr} W\Sigma^{-1}$ as

$$\begin{aligned}
\operatorname{tr} W\Sigma^{-1} =& \operatorname{tr} W_{11}\Sigma^{11} + 2\operatorname{tr} W_{12}\Sigma^{21} + \operatorname{tr} W_{22}\Sigma^{22} \\
=& \operatorname{tr} W_{11}\Sigma_{11}^{-1} + \operatorname{tr} W_{11} B \Sigma_{22\cdot 1}^{-1} B' \\
&- 2\operatorname{tr} W_{12}\Sigma_{22\cdot 1}^{-1} B' + \operatorname{tr} W_{22}\Sigma_{22\cdot 1}^{-1} \qquad (27)\\
=& \operatorname{tr} W_{11}\Sigma_{11}^{-1} + \operatorname{tr}(\widehat{B} - B)' W_{11}(\widehat{B} - B)\Sigma_{22\cdot 1}^{-1} + \operatorname{tr} W_{22\cdot 1}\Sigma_{22\cdot 1}^{-1} \qquad (28)
\end{aligned}$$

where $\widehat{B} = W_{11}^{-1} W_{12}$.

First consider $H_{(11)}$. From (27) or (28) it follows that the maximum likelihood estimator of $\Sigma_{11,h}$ is given by $\widehat{\Sigma}_{11,h} = W_{11,T}/N$ when $H_{(11)}$ is assumed and $\widehat{\Sigma}_{11,h} = W_{11,h}/N_h, h = 1, \cdots, k$, when $H_{(11)}$ is not assumed. Note that maximum likelihood estimators for $B_h, \Sigma_{22\cdot 1,h}$ remain the same whether $H_{(11)}$ is assumed or not. From this observation it is easy to derive the expression for λ_{11} in (10). Now consider $H_{(22)}$ assuming that $B_h, h = 1, \cdots, k$ are free. From (28) we see that the maximum likelihood estimator of B_h is just $\widehat{B}_h = W_{11,h}^{-1} W_{12,h}$. Then the maximum likelihood estimator of $\Sigma_{22\cdot 1}$ is given by $\sum_{h=1}^{k} W_{22\cdot 1,h}/N$ when $H_{(22)}$ is assumed and $W_{22\cdot 1,h}/N_h, h = 1, \ldots, k$, when $H_{(22)}$ is not assumed. Hence λ_{22} is easily derived as in (10).

Finally consider $H_{(21)}$ assuming $H_{(22)}$. We need to obtain maximum likelihood estimator of $B = B_1 = \cdots = B_k$ under $H_{(21)}$. From (27) we have

$$\begin{aligned}
\sum_{h=1}^{k} W_h \Sigma_h^{-1} =& \sum_h \operatorname{tr} W_{11,h}\Sigma_{11,h}^{-1} + \operatorname{tr} W_{11,T} B \Sigma_{22\cdot 1}^{-1} B' \\
&- 2\operatorname{tr} W_{12,T}\Sigma_{22\cdot 1}^{-1} B' + \operatorname{tr} W_{22,T}\Sigma_{22\cdot 1}^{-1} \\
=& \sum_h \operatorname{tr} W_{11,h}\Sigma_{11,h}^{-1} + \operatorname{tr} W_{11,T}(B - W_{11,T}^{-1} W_{12,T})\Sigma_{22\cdot 1}^{-1} \\
&(B - W_{11,T}^{-1} W_{12,T})' + \operatorname{tr} W_{22\cdot 1,T}\Sigma_{22\cdot 1}^{-1}. \qquad (29)
\end{aligned}$$

Therefore under $H_{(21)} \cap H_{(22)}$ the maximum likelihood estimator is

$$\widehat{B} = W_{11,T}^{-1} W_{12,T}, \qquad \widehat{\Sigma}_{22\cdot 1} = W_{22\cdot 1,T}/N. \tag{30}$$

Substituting this into (26) and comparing it to the maximized concentrated likelihood under $H_{(22)}$ (with $\widehat{\Sigma}_{22\cdot 1} = \sum_h W_{22\cdot 1,h}/N, \widehat{B}_h = W_{11,h}^{-1} W_{12,h}$) we easily obtain λ_{21} in (10). ∎

3.2. Derivation of Likelihood Ratio Statistics in Theorem 1. Since the derivation of likelihood ratio statistics in Theorem 1 is analogous to Lemma 1, we only sketch the proof. The likelihood ratio statistic λ_{ii} for $\bar{H}_{ii}$ in (22) is the same as in Lemma 1 and omitted. Consider H_{ij}, $i > j$. We first note that the likelihood ratio statistic depends only on $i \times i$ upper left block of $W_h, h = 1, \cdots, k$, just as $\lambda_{(11)}$ in Lemma 1 depends only on $W_{11,h}$, $h = 1, \cdots, k$. Hence we can assume $i = p$ without loss of generality. Then it suffices to consider the case where (dropping the subscript h) Σ and W are partitioned as

$$W = \begin{pmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix}. \tag{31}$$

Furthermore write

$$\begin{aligned} W &= \begin{pmatrix} W_{(11)} & W_{(13)} \\ W_{(31)} & W_{33} \end{pmatrix}, \qquad W_{(11)} = \begin{pmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{pmatrix}, \\ W_{33\cdot} &= W_{33\cdot 1,2} = W_{33} - W_{(31)} W_{(11)}^{-1} W_{(13)}, \\ W_{33\cdot 1} &= W_{33} - W_{31} W_{11}^{-1} W_{13}, \\ W_{23\cdot 1} &= W_{23} - W_{21} W_{11}^{-1} W_{13} = W_{32\cdot 1}', \\ W_{22\cdot 1} &= W_{22} - W_{21} W_{11}^{-1} W_{12}, \end{aligned} \tag{32}$$

and write similarly for Σ. Let $B' = \Sigma_{(31)} \Sigma_{(11)}^{-1}$ be partitioned as $B' = (B_1', B_2')$. Now assume

$$\Sigma_{33\cdot,1} = \cdots = \Sigma_{33\cdot,k} = \Sigma_{33\cdot,}, \qquad B_{2,1} = \cdots = B_{2,k} = B_2, \tag{33}$$

and $B_{1,h}, h = 1, \cdots, k$, are free parameters. The essential step in our proof is to obtain maximum likelihood estimates of these parameters. It is straightforward to derive

$$\begin{aligned} \operatorname{tr} W \Sigma^{-1} =& \operatorname{tr} W_{(11)} \Sigma_{(11)}^{-1} + \operatorname{tr} W_{(11)} B \Sigma_{33\cdot}^{-1} B' \\ &- 2 \operatorname{tr} W_{(13)} \Sigma_{33\cdot}^{-1} B' + \operatorname{tr} W_{33} \Sigma_{33\cdot}^{-1} \\ =& \operatorname{tr} W_{(11)} \Sigma_{(11)}^{-1} \\ &+ \left[\operatorname{tr} W_{33\cdot 1} \Sigma_{33\cdot}^{-1} - 2 \operatorname{tr} W_{23\cdot 1} \Sigma_{33\cdot}^{-1} B_2' + \operatorname{tr} W_{22\cdot 1} B_2 \Sigma_{33\cdot}^{-1} B_2'\right] \\ &+ \operatorname{tr} W_{11} (B_1 - W_{11}^{-1} W_{13} + W_{11}^{-1} W_{12} B_2) \Sigma_{33\cdot}^{-1} \\ &\cdot (B_1 - W_{11}^{-1} W_{13} + W_{11}^{-1} W_{12} B_2)'. \end{aligned} \tag{34}$$

Writing subscript h again and adding with respect to h, let

$$\widetilde{W}_{22\cdot 1} = \sum_h W_{22\cdot 1,h}, \qquad \widetilde{W}_{23\cdot 1} = \sum_h W_{23\cdot 1,h}, \qquad \widetilde{W}_{33\cdot 1} = \sum_h W_{33\cdot 1,h}. \tag{35}$$

From (34) it follows that

$$\begin{aligned}
&\sum_{h=1}^{k} \operatorname{tr} W_h \Sigma_h^{-1} \\
&= \sum_{h=1}^{k} \operatorname{tr} W_{(11),h} \Sigma_{(11),h}^{-1} + \operatorname{tr} \Sigma_{33\cdot}^{-1} (\widetilde{W}_{33\cdot 1} - \widetilde{W}_{32\cdot 1} \widetilde{W}_{22\cdot 1}^{-1} \widetilde{W}_{23\cdot 1}) \\
&\quad + \operatorname{tr} \widetilde{W}_{22\cdot 1} (B_2 - \widetilde{W}_{22\cdot 1}^{-1} \widetilde{W}_{23\cdot 1}) \Sigma_{33\cdot}^{-1} (B_2 - \widetilde{W}_{22\cdot 1}^{-1} \widetilde{W}_{23\cdot 1})' \\
&\quad + \sum_{h=1}^{k} \operatorname{tr} W_{11,h} (B_{1,h} - W_{11,h}^{-1} W_{13,h} + W_{11,h}^{-1} W_{12,h} B_2) \Sigma_{33\cdot}^{-1} \\
&\quad \cdot (B_{1,h} - W_{11,h}^{-1} W_{13,h} + W_{11,h}^{-1} W_{12,h} B_2)'.
\end{aligned} \tag{36}$$

From this the maximum likelihood estimators are obtained as

$$\begin{aligned}
\widehat{\Sigma}_{33\cdot} &= \frac{1}{N} (\widetilde{W}_{33\cdot 1} - \widetilde{W}_{32\cdot 1} \widetilde{W}_{22\cdot 1}^{-1} \widetilde{W}_{23\cdot 1}), \\
\widehat{B}_2 &= \widetilde{W}_{22\cdot 1}^{-1} \widetilde{W}_{23\cdot 1}, \\
\widehat{B}_{1,h} &= W_{11,h}^{-1} W_{13,h} - W_{11,h}^{-1} W_{12,h} \widehat{B}_2, \qquad h = 1, \cdots, k,
\end{aligned} \tag{37}$$

and the maximized likelihood function is, omitting irrelevant constants,

$$L \propto |\widehat{\Sigma}_{33\cdot}|^{-N/2} \propto |\widetilde{W}_{33\cdot 1} - \widetilde{W}_{32\cdot 1} \widetilde{W}_{22\cdot 1}^{-1} \widetilde{W}_{23\cdot 1}|^{-N/2}. \tag{38}$$

Now it is easy to see that λ_{ij} is given by (23). ∎

REMARK 3.1. $\widetilde{W}_{33\cdot 1} - \widetilde{W}_{32\cdot 1} \widetilde{W}_{22\cdot 1}^{-1} \widetilde{W}_{23\cdot 1}$ is obtained by 1) regressing out the first block in each W_h, 2) adding with respect to h, i.e., pooling the sum of squares, and 3) regressing out the second block. $\tilde{w}(i \mid j)_{ii\cdot}$ of (17) was defined by the same consideration.

REMARK 3.2. Derivation of maximum likelihood estimates for $B_{1,h}$ and B_2 is somewhat easier to see in conditional regression setup in canonical form. This approach was suggested by a referee. Let Σ_{33}, W_{33} be scalars in (32) and consider the following regreesion model:

$$X_{3h} \sim N_{n_h}(X_{1h}\beta_{1h} + X_{2\cdot 1,h}\beta_2, \sigma_{pp\cdot} I), \tag{39}$$

where $X'_{1h} X_{2\cdot 1,h} = 0$. Let

$$b_{1h} = (X'_{1h} X_{1h})^{-1} X'_{1h} X_{3h}, \qquad b_{2h} = (X'_{2\cdot 1,h} X_{2\cdot 1,h})^{-1} X'_{2\cdot 1,h} X_{3h}.$$

Then

$$\begin{aligned} Q =& \|X_{3h} - X_{1h}\beta_{1h} + X_{2\cdot 1,h}\beta_2\|^2 \\ =& \sum_{h=1}^{k}(b_{1h} - \beta_{1h})'(X_{1h}'X_{1h})(b_{1h} - \beta_{1h}) \\ &+ \sum_{h=1}^{k}(b_{2h} - \beta_2)'(X_{2\cdot 1,h}'X_{2\cdot 1,h})(b_{2h} - \beta_2) + \sum_{h=1}^{k} e_h' e_h, \end{aligned} \tag{40}$$

where $e_h = X_{3h} - X_{1h}b_{1h} - X_{2\cdot 1,h}b_{2h}$. Now it easily follows that the least square estimates of β_{1h} and β_2 are given as

$$\hat{\beta}_{1h} = b_{1h}, \qquad \hat{\beta}_2 = \left(\sum_{h=1}^{k} X_{2\cdot 1,h}'X_{2\cdot 1,h}\right)^{-1} \left(\sum_{h=1}^{k} X_{2\cdot 1,h}'X_{3h}\right). \tag{41}$$

This corresponds to $\widehat{B}_2, \widehat{B}_{1,h}$ in (37).

3.3. Derivation of Distributional Results. It remains to show the null distributional results in Lemma 1 and Theorem 1. They are the consequences of a version of Cochran's theorem (Lemma 2) and Lemma 3 below.

LEMMA 2. *Let $X : n \times p \sim N(0, I_n \otimes \Sigma)$. Let $V_h, h = 1, \cdots, k$, be mutually orthogonal subspaces of R^n. Let V be a subspace of R^n such that $V_h \subset V, h = 1, \cdots, k$. Denote the orthogonal projectors onto V_h and V by P_{V_h} and P_V. Define $W_h = X'P_{V_h}X$, $h = 1, \cdots, k$ and $V_{k+1} = X'P_VX - W_1 - \cdots - W_k$. Then $W_1, \cdots, W_{k+1}$ are mutually independently distributed according to Wishart distribution*

$$\begin{aligned} &W_h \sim W_p(\dim V_h, \Sigma), \qquad h = 1, \cdots, k, \\ &W_{k+1} \sim W_p(\dim V - \dim V_1 - \cdots - \dim V_k, \Sigma). \end{aligned}$$

Proof of this lemma is easy and omitted.

Based on Lemma 2 the following result can be established. Because of usefulness for the case of small degrees of freedom, we do not assume $n_h \geq p$ here.

LEMMA 3. *Let $W_h \sim W_p(n_h, \Sigma)$ be independently distributed. Let $W_T = W_1 + \cdots + W_k$ and $n = n_1 + \cdots + n_k$. Assume $|\Sigma| \neq 0$. Let W_h and W_T be partitioned as*

$$W_h = \begin{pmatrix} W_{11,h} & W_{12,h} \\ W_{21,h} & W_{22,h} \end{pmatrix}, \qquad W_T = \begin{pmatrix} W_{11,T} & W_{12,T} \\ W_{21,T} & W_{22,T} \end{pmatrix},$$

where $W_{11,h}, W_{11,T}$ *are* $q \times q$. *Let*

$$W_{22\cdot 1,h} = \begin{cases} W_{22,h} - W_{21,h}W_{11,h}^{-1}W_{12,h}, & \text{if } n_h \geq q; \\ 0, & \text{otherwise,} \end{cases} \tag{42}$$

and define $W_{22\cdot 1,T}$ *similarly. Then* $W_{22\cdot 1,1}, \cdots, W_{22\cdot 1,k}, W_{22\cdot 1,T} - W_{22\cdot 1,1} - \cdots - W_{22\cdot 1,k}$ *are mutually independently distributed according to*

$$\begin{aligned} &W_{22\cdot 1,h} \sim W_{p-q}(\max(n_h - q, 0), \Sigma_{22\cdot 1}) \\ &W_{22\cdot 1,T} - W_{22\cdot 1,1} - \cdots - W_{22\cdot 1,k} \\ &\sim \; W_{p-q}(\max(n-q,0) - \sum_{h=1}^{k} \max(n_h - q, 0), \Sigma_{22\cdot 1}), \end{aligned} \tag{43}$$

where Wishart distribution with 0 degree of freedom is degenerate at 0 *matrix.*

PROOF. Let $X : n \times p \sim N(0, I_n \otimes \Sigma)$ be partitioned as

$$X = (X_{(1)}, X_{(2)}) = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ \vdots & \vdots \\ X_{k1} & X_{k2}, \end{pmatrix}$$

where $X_{(1)} : n \times q$, $X_h : n_h \times p$. Then $W_T = X'X, W_h = X_h'X_h$. It suffices to show that the results hold conditionally when all elements of $X_{(1)}$ are fixed and the results do not depend on $X_{(1)}$. Let $M_1 = \operatorname{span} X_{(1)}$ be the subspace of R^n spanned by the columns of $X_{(1)}$ and let $V = M_1^{\perp}$ be the orthogonal complement of M_1. Let $U_h \subset R^n$ be the n_h dimensional subspace of vectors of the following form

$$\Big(\underbrace{0, \cdots, 0}_{n_1 + \cdots + n_{h-1}}, x_{n_1 + \cdots + n_{h-1} + 1}, \cdots, x_{n_1 + \cdots + n_h}, \underbrace{0, \cdots, 0}_{n_{h+1} + \cdots + n_k} \Big)$$

Now let $\widetilde{U}_h = \operatorname{span}(0, \cdots, 0, X_{h1}', 0, \cdots, 0)' = \operatorname{span} P_{U_h} X_{(1)}$ be the subset of U_h spanned by columns of X_{h1}. Then $\dim \widetilde{U}_h = \min(n_h, q)$. Furthermore define $V_h = U_h \cap \widetilde{U}_h^{\perp}$. Note that V_h is the orthogonal complement of $\widetilde{U}_h$ in U_h. This corresponds to subtracting off the regression on X_{h1}. Note $\dim V_h = \max(n_h - q, 0)$. With these definitions we have

$$\begin{aligned} W_{22\cdot 1,T} &= X_{(2)}' P_V X_{(2)}, \\ W_{22\cdot 1,h} &= X_{(2)}' P_{V_h} X_{(2)}. \end{aligned} \tag{44}$$

Now $V_h \subset U_h$ and U_h's are mutually orthogonal subspaces. Hence V_h's are mutually orthogonal subspaces. Therefore by Lemma 2 it suffices to show $V_h \subset V, h = 1, \cdots, k$. Noting $V = M_1^{\perp}$ and

$$V_h \subset M_1^{\perp} \iff M_1 \subset V_h^{\perp},$$

it suffices to show that the orthogonal projection of M_1 onto V_h is $\{0\}$ or

$$0 = P_{V_h} X_{(1)} = P_{U_h} X_{(1)} - P_{\widetilde{U}_h} X_{(1)}. \tag{45}$$

Since $\widetilde{U}_h = \operatorname{span} P_{U_h} X_{(1)}$, $P_{\widetilde{U}_h}$ can be written as

$$P_{\widetilde{U}_h} = P_{U_h} X_{(1)} [X'_{(1)} P_{U_h} X_{(1)}]^{-1} X'_{(1)} P_{U_h}.$$

Hence $P_{\widetilde{U}_h} X_{(1)} = P_{U_h} X_{(1)}$ and this proves the lemma. ▮

Based on Lemma 3 it is easy to prove distributional results in Lemma 1 and Theorem 1.

Acknowledgement. The author would like to thank a referee for comments leading to Remark 3.2.

REFERENCES

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley,New York.

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd. ed. Wiley, New York.

ROY, J. (1958). Step-down procedure in multivariate analysis. *Ann. Math. Statist.* **29**, 1177–1187.

TAKEMURA, A. (1991). *An Introduction to Multivariate Statistical Inference* (in Japanese). Kyoritsu Shuppan, Tokyo.

FACULTY OF ECONOMICS
UNIVERSITY OF TOKYO, 7-3-1 HONGO
BUNKYO-KU TOKYO 113
JAPAN