J. Statistics

1996(1), 7-14

TESTING THE EQUALITY OF TWO COVARIANCE MATRICES UNDER COMPLEX SURVEYS

By

MUHAMMAD KHALID PERVAIZ

(Department of Statistics, Government College, Lahore.)

ABSTRACT

A test for testing the equality of two covariance matrices under complex survey designs is developed. The proposed test is a modification of T^2-test developed by Tiku and Balakrishnan (1985).

KEY WORDS

Complex survey, superpopulation, asymptotic.

1. INTRODUCTION

Tiku and Balakrishnan (1985) have given a T^2-test for testing the equality of two covariance matrices. They empirically evaluated the test for normal/non-normal distributions, and found that for small samples the test is not robust; in fact, its Type I error is usually smaller than its presumed level under normality. Pervaiz (1994) compared the size and power properties of the T^2-test with the tests given by Layard (1972) and confirmed the findings of Tiku and Balakrishan (1985). But in some situations its observed significance level performance is comparatively better than the tests proposed by Layard (1972).

The object of this paper is to modify T^2-test under complex survey designs. In section 2 cluster sampling design and in section 3 stratified cluster sampling design is considered.

2. TEST STATISTIC UNDER CLUSTER SAMPLING DESIGN

2.1 Notation

The suffix i denotes the finite population, i=1,2.

N_i	Number of clusters in finite population.
n_i	Number of clusters in sample.
m_{ic}	Number of observations in c-th cluster, $c = 1,\ldots, n_i$
$N_{oi} = \sum_{c=1}^{N_i} m_{ic}$	Finite population size.
$n_{oi} = \sum_{c=1}^{n_i} m_{ic}$	Sample size.
$\underline{x}_{ice} = \left(x_{ice1}, x_{ice2}\right)^T$	e-th vector observation of c-th cluster, $e = 1,\ldots, m_{ic}$
$\overline{\underline{X}}_{i..} = \frac{1}{N_{oi}} \sum_{c=1}^{N_i} \sum_{e=1}^{m_{ic}} \underline{x}_{ice}$	Mean vector of finite population.
$\bar{\underline{x}}_{i..} = \frac{1}{n_{oi}} \sum_{c=1}^{n_i} \sum_{e=1}^{m_{ic}} \underline{x}_{ice}$	Sample mean vector.

$$\underline{C}_i = \frac{1}{N_{oi}} \sum_{c=1}^{N_i} \sum_{e=1}^{m_{ic}} (\underline{x}_{ice} - \overline{\underline{X}}_{i..})(\underline{x}_{ice} - \overline{\underline{X}}_{i..})^T$$

Covariance matrix of finite population.

$$\underline{S}_i = \frac{1}{n_{oi}} \sum_{c=1}^{n_i} \sum_{e=1}^{m_{ic}} (\underline{x}_{ice} - \bar{\underline{x}}_{i..})(\underline{x}_{ice} - \bar{\underline{x}}_{i..})^T$$

Sample covariance matrix.

2.2 Sampling Design

A simple random sample of n_i clusters is selected from the N_i clusters of i-th finite population. The samples are independent. Within each selected cluster all subunits are included in a sample.

2.3 Null hypothesis to be tested

The model/super population approach with unrestrictive assumptions is adopted (Pervaiz, 1989). It is assumed the $\underline{x}_{ice}$ are random variables which implies that $\underline{C}_i$ are also random variables. It is also assumed that the $\underline{C}_i$ converge to $\underline{\Sigma}_i$ (covariance matrix of super population) as N_{oi}

"finite population size" go to infinity. Thus the finite populations with covariance matrices $\underline{C}_i$'s may be assumed to be random samples from infinite populations called super populations with covariance matrices $\underline{\Sigma}_i$'s. Then hypotheses of interest are:

$$H_{o1}: \underline{\Sigma}_1 = \underline{\Sigma}_2 \quad vs \quad H_{A1}: \underline{\Sigma}_1 \neq \underline{\Sigma}_2$$

2.4 Description of Test Statistic

Let

$$U_{1ce} = \left(x_{1ce1} - \bar{x}_{1..1}\right)^2 \quad and \quad U_{2ce} = \left(x_{1ce2} - \bar{x}_{1..2}\right)^2 \; ; \quad and$$
$$V_{1ce} = \left(x_{2ce1} - \bar{x}_{2..1}\right)^2 \quad and \quad V_{2ce} = \left(x_{2ce2} - \bar{x}_{2..2}\right)^2$$

Where

$$\underline{x}_{ice2} = x_{ice2} - \hat{b}x_{ice1},$$

$\bar{x}_{1..1}$ and $\bar{x}_{1..2}$ and usual sample means.

$$\hat{b} = \frac{\sum_{i=1}^{2}\sum_{c=1}^{n_i}\sum_{e=1}^{m_{ic}}\left(x_{ice1} - \bar{x}_{i..1}\right)\left(x_{ice2} - \bar{x}_{i..2}\right)}{\sum_{i=1}^{2}\sum_{c=1}^{n_i}\sum_{e=1}^{m_{ic}}\left(x_{ice1} - \bar{x}_{i..1}\right)^2}$$

is the pooled regression co-efficient.

Fuller (1975) and Skinner (1986) have given central limit laws under complex surveys. Thus the test statistic:

$$T_c^2 = \frac{n_1 n_2}{n_1 + n_2}\begin{bmatrix}\bar{w}_1 & \bar{w}_2\end{bmatrix}\begin{bmatrix}\hat{\phi}_1^2 & 0 \\ 0 & \hat{\phi}_2^2\end{bmatrix}^{-1}\begin{bmatrix}\bar{w}_1 \\ \bar{w}_2\end{bmatrix} \tag{2.4.1}$$

is distributed approximately as Hotelling's T^2 with 2 and (n_1+n_2-2) degrees of freedom under H_{01} Where

$$\bar{w}_1 = \bar{U}_1 - \bar{V}_1$$
$$\bar{w}_2 = \bar{U}_2 - \bar{V}_2$$

$$\hat{\phi}_1^2 = \frac{\sum_{c=1}^{n_1}\left(\overline{U}_{1c}-\overline{U}_1\right)^2+\sum_{c=1}^{n_2}\left(\overline{V}_{1c}-\overline{V}_1\right)^2}{n_1+n_2-2} \quad and$$

$$\hat{\phi}_2^2 = \frac{\sum_{c=1}^{n_1}\left(\overline{U}_{2c}-\overline{U}_2\right)^2+\sum_{c=1}^{n_2}\left(\overline{V}_{2c}-\overline{V}_2\right)^2}{n_1+n_2-2}$$

The T_c^2 is asymptotically distributed as χ_3^2 under H_{01}

3. TEST STATISTIC UNDER STRATIFIED CLUSTER SAMPLING DESIGN

3.1 Notation

The suffix i denotes the finite population, i=1,2.

H_i	Number of strata.
N_{ih}	Total number of clusters in h-th stratum.
n_{ih}	Number of clusters in h-th stratum of sample.
m_{ihc}	Observations in c-th cluster in h-th stratum.
$N_i = \sum_{h=1}^{H_i} N_{ih}$	Total number of clusters.
$n_i = \sum_{h=1}^{H_i} n_{ih}$	Number of clusters in a sample.
$N_{oih} = \sum_{c=1}^{N_{ih}} m_{ihc}$	Number of observations in h-th stratum.
$n_{oih} = \sum_{c=1}^{n_{ih}} m_{ihc}$	Number of observations in a sample from h-th stratum.
$N_{oi} = \sum_{h=1}^{H_i} N_{oih}$	Finite population size.

$n_{oi} = \sum_{h=1}^{H_i} n_{oih} = \sum_{h=1}^{H_i}\sum_{c=1}^{n_{ih}} m_{ihc}$ Sample size.

$W_{ih} = \frac{N_{oih}}{N_{oi}}$ Stratum weight.

$\underline{x}_{ihce} = \left(x_{ihce1}, x_{ihce2}\right)^T$ e-th vector observation of ç-th çluster of h-th stratum. $e = 1,2,\ldots\ldots, m_{ihc}$

$\overline{\underline{X}}_{i\ldots} = \frac{1}{N_{oi}} \sum_{h=1}^{H_i}\sum_{c=1}^{N_{ih}}\sum_{e=1}^{m_{ihc}} \underline{x}_{ihce}$ Mean vector of finite population.

$\overline{\underline{x}}_{i\ldots} = \sum_{h=1}^{H_i} \frac{W_{ih}}{n_{oih}} \sum_{c=1}^{n_{ih}}\sum_{e=1}^{m_{ihc}} \underline{x}_{ihce}$ Weighted mean vector of sample.

$\underline{C}_i = \frac{1}{N_{oi}} \sum_{h=1}^{H_i}\sum_{c=1}^{N_{ih}}\sum_{e=1}^{m_{ihc}} (\underline{x}_{ihce} - \overline{\underline{X}}_{i\ldots})(\underline{x}_{ihce} - \overline{\underline{X}}_{i\ldots})^T$

Covariance matrix of a finite population.

$\underline{S}_i = \sum_{h=1}^{H_i} \frac{W_{ih}}{n_{oih}} \sum_{c=1}^{n_{ih}}\sum_{e=1}^{m_{ihc}} (\underline{x}_{ihce} - \overline{\underline{x}}_{i\ldots})(\underline{x}_{ihce} - \overline{\underline{x}}_{i\ldots})^T$

Weighted sample covariance matrix.

3.2 Sampling Design

The n_{ih} clusters are drawn by simple random sampling from N_{ih} clusters. Within each selected cluster all subunits are included in a sample. The samples are independent.

3.3 Null hypothesis to be tested

It assumed that $\underline{x}_{ihce}$ are random vectors, which implies that $\underline{C}_i$ are also random variables. Furthermore it is assumed that $\underline{C}_i$ converge to $\underline{\Sigma}_i$ and N_{oi} become larger. Thus, the finite populations with covariance matrices $\underline{C}_i$'s may be viewed as sample from infinite populations called superpopulation with covariance matrices $\underline{\Sigma}_i$'s. The hypotheses of interest are:

$$H_{02}: \underline{\Sigma}_1 = \underline{\Sigma}_2 \quad vs \quad H_{A2}: \underline{\Sigma}_1 \neq \underline{\Sigma}_2$$

3.4 Description of test statistic

Let

$$U_{1hce} = \left(x_{1hce1} - \bar{x}_{1...1}\right)^2 \text{ and } U_{2hce} = \left(x_{1hce2} - \bar{x}_{1...2}\right)^2 \text{ and}$$

$$V_{1hce} = \left(x_{2hce1} - \bar{x}_{2...1}\right)^2 \text{ and } V_{2hce} = \left(x_{2hce2} - \bar{x}_{2...2}\right)^2$$

Where

$$x_{ihce2-} = x_{ihce2} - \hat{b}x_{ihce1}$$

$\bar{x}_{i...1}$ and $\bar{x}_{i...2-}$ are usual sample means.

$$\hat{b} = \frac{\sum_{i=1}^{2}\sum_{h=1}^{H_i}\frac{W_{ih}}{n_{oih}}\sum_{c=1}^{n_{ih}}\sum_{e=1}^{m_{ihc}}\left(x_{ihce1} - \bar{x}_{i...1}\right)\left(x_{ihce2} - \bar{x}_{i...2}\right)}{\sum_{i=1}^{2}\sum_{h=1}^{H_i}\frac{W_{ih}}{n_{oih}}\sum_{c=1}^{n_{ih}}\sum_{e=1}^{m_{ihc}}\left(x_{ihce1} - \bar{x}_{i...1}\right)^2}$$

is the pooled regression co-efficient. The test statistic:

$$T_{sc}^2 = \begin{bmatrix}\bar{w}_1 & \bar{w}_2\end{bmatrix}\begin{bmatrix}\hat{\psi}_1^2 & 0 \\ 0 & \hat{\psi}_2^2\end{bmatrix}^{-1}\begin{bmatrix}\bar{w}_1 \\ \bar{w}_2\end{bmatrix}$$

is distributed approximately as Hotelling's T^2 with 2 and n_1+n_2-2 degree of freedom under H_{02}. Where

$$\bar{U}_1 = \sum_{h=1}^{H_1}\frac{W_{1h}}{n_{o1h}}\sum_{c=1}^{n_{1h}}\sum_{e=1}^{m_{1hc}}U_{ihce},$$

$$\bar{U}_2 = \sum_{h=1}^{H_1}\frac{W_{1h}}{n_{o1h}}\sum_{c=1}^{n_{1h}}\sum_{e=1}^{m_{1hc}}U_{2hce},$$

$$\bar{V}_1 = \sum_{h=1}^{H_2}\frac{W_{2h}}{n_{o2h}}\sum_{c=1}^{n_{2h}}\sum_{e=1}^{m_{2hc}}V_{1hce},$$

$$\overline{V}_2 = \sum_{h=1}^{H_2} \frac{W_{2h}}{n_{o2h}} \sum_{c=1}^{n_{2h}} \sum_{e=1}^{m_{2hc}} V_{2hce},$$

$$\hat{\psi}_1^2 = \sum_{h=1}^{H} \frac{W_{1h}^2}{n_{1h}(n_{1h}-1)} \sum_{c=1}^{n_{1h}} \left(\overline{U}_{1hc} - \overline{U}_{1h}\right)^2 + \sum_{h=1}^{H} \frac{W_{2h}^2}{n_{2h}(n_{2h}-1)} \sum_{c=1}^{n_{2h}} \left(\overline{V}_{1hc} - \overline{V}_{1h}\right)^2$$

and

$$\hat{\psi}_2^2 = \sum_{h=1}^{H} \frac{W_{2h}^2}{n_{1h}(n_{1h}-1)} \sum_{c=1}^{n_{1h}} \left(\overline{U}_{2hc} - \overline{U}_{2h}\right)^2 + \sum_{h=1}^{H} \frac{W_{2h}^2}{n_{2h}(n_{2h}-1)} \sum_{c=1}^{n_{2h}} \left(\overline{V}_{2hc} - \overline{V}_{2h}\right)^2$$

REFERENCES

[1] **FULLER, W.A. (1975)** *REGRESSION ANALYSIS FOR SAMPLE SURVEYS.* SANKHYA, THE INDIAN JOURNAL OF STATISTICS. 37 C 117-132.

[2] **LAYARD, M.W.J. (1972).** *LARGE SAMPLE TESTS FOR THE EQUALITY OF TWO COVARIANCE MATRICES.* ANN MAT. STATIST. 43. 123-141.

[3] **PERVAIZ, M. K. (1989).** *ASYMPTOTICALLY ROBUST RESTS FOR THE EQUALITY OF TWO CAVARIANCE MATRICES IN COMPLEX SURVEYS-I.* PROC. ICCS - I VOL. II P.861-876.

[4] **PERVAIZ, M. K. (1994).** *SIZE AND POWER PROPERTIES OF ASYMPTOTICALLY ROBUST TESTS FOR EQUALITY OF TWO COVARIANCE MATRICES.* PAKISTAN JOURNAL OF STATISTICS. VOL. 2. P. 1-14

[5] **SKINNER, C.J. (1986).** *DESIGN EFFECTS OF TWO STAGE SAMPLING,* JOURNAL OF THE ROYAL STATISTICAL SOCIETY. B.PT. 1.

[6] **TIKU M.L. & BALAKRISHANAN, N. (1985).** *TESTING THE EQUALITY OF VARIANCE-COVARIANCE MATRICES THE ROBUST WAY, COMM. STATIST. - THEOR. METH.* 11(9) 985-1001.