J.Statistics
1995(1), 25-36

ASYMPTOTICALLY ROBUST HOMOGENEITY TESTS UNDER COMPLEX SURVEYS-II

BY

Muhammad Khalid Pervaiz

ABSTRACT

The superpopulation approach with unrestrictive assumptions is adopted. The asymptotically robust tests for homogeneity of variances are obtained for dependent cluster samples from finite populations. The test statistics are extended for stratified cluster samples as well.

Some Key Words: Superpopulation, robust, complex survey.

1 INTRODUCTION

Pervaiz (1995) obtained asymptotically robust homogeneity tests under complex surveys. The finite populations were supposed to consist on separate clusters. Therefore the samples were considered independent. But the finite populations may cut across the clusters. That is the clusters consist of elements from both finite populations. Under this situation one has to select a cluster sample from the union of finite populations and separate each cluster of sample according to the respective finite population. Consequently the cluster samples achieved from the finite populations are not independent. Therefore the homogeneity tests, obtained by Pervaiz (1995) cannot be applied. So it becomes necessary to modify these tests accordingly and this is the object of this paper.

* *Department of Statistics, Government College, Lahore*

PERVAIZ, M.K. (1989). *Asymptotically robust tests for the equality of two convariance matrices in complex surveys.* Proc. ICCS-1 Vol. II P.861-876

PERVAIZ, M.K. (1989 a). *Covariance Matrix Estimation in Complex Surveys.* Journal of Natural Sciences and Mathematics. Vol. 29, No.2, P177-185

PLACKETT, R.L. (1946). *Literature on testing the equality of variances and covariances in normal populations.* J.R. Statist. Soc. 109, 457-68.

PLACKETT, R. L. (1960). *Principles of Regression Analysis.* Oxford University Press.

WOLTER, K.M. (1985). *Introduction to variance estimation.* New York.

$$S_i^2 = \frac{1}{N_{oi}} \sum_{c=1}^{N} \sum_{e=1}^{m_{ic}} (x_{ice} - \overline{X}_{i..})^2$$ Variance of finite population.

$$s_i^2 = \frac{1}{n_{oi}} \sum_{c=1}^{n} \sum_{e=1}^{m_{ic}} (x_{ice} - \bar{x}_{i..})^2$$ Sample variance.

2.2 Sampling design

The n clusters are selected form N clusters of U by simple random sampling. The each cluster of n is separated for the respective finite population to obtain the sample form the finite populations. Within each selected cluster all subunits are included in a sample.

2.3 Problem under study

The superpopulation approach with unrestrictive assumptions is adopted; e.g. Pervaiz (1989). The x_{ice} are random variables which implies that S_i^2's are also random variables. It is also assumed that S_i^2's coverge to σ_i^2's as N_{oi} becomes larger. Then the finite populations with variances S_i^2's may be viewed as samples from superpopulations with variance σ_i^2's. The hypothesis of interest is:

$$H_{o1}: \sigma_1^2 = \sigma_2^2 \quad vs \quad H_{A1}: \sigma_1^2 \neq \sigma_2^2 .$$

2.4 Properties of sample variance

Fuller (1975) and Skinner (1986) give central limit law for complex samples with large sample sizes; i.e.

We omit the finite population corrections from the estimators of the variance of sample variances, (e.g. Cochran, 1977 p.39).

2 TESTS UNDER CLUSTER SAMPLING DESIGN

2.1 Notation

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

U	Union of finite populations. (It is assumed that each cluster of U contains more than one element of the i-th finite population).
N	Number of clusters in the union of finite populations.
n	Number of clusters in sample. c=1,2,......n.
m_{ic}	Number of observations in c-th cluster.
$N_{oi} = \sum_{c=1}^{N} m_{ic}$	Finite population size.
$n_{oi} = \sum_{c=1}^{n} m_{ic}$	Sample size.
x_{ice}	e-th observation of c-th cluster. $e = 1,2,\ldots, m_{ic}$
$\overline{X}_{i..} = \frac{1}{N_{oi}} \sum_{c=1}^{N} \sum_{e=1}^{m_{ic}} x_{ice}$	Mean of finite population.
$\overline{x}_{i..} = \frac{1}{n_{oi}} \sum_{c=1}^{n} \sum_{e=1}^{m_{ic}} x_{ice}$	Sample mean.

$$\bar{z}_c = \frac{n}{n_{01}}\sum_{e=1}^{m_{1c}}(x_{1ce} - \bar{x}_{1..})^2 - \frac{n}{n_{02}}\sum_{e=1}^{m_{2c}}(x_{2ce} - \bar{x}_{2..})^2 - n\left(\frac{m_{1c}}{n_{01}}\bar{x}_{1..}^2 - \frac{m_{2c}}{n_{02}}\bar{x}_{2..}^2\right)$$

For proof see Pervaiz (1990). If $m_{ic} = m_i$ *then* $n_{0i} = nm_i$ and the constant term of $\bar{z}_c$ is canceled out in the formula of $\hat{V}$.

(b) GROUPING TEST

The parent sample is divided into number of random groups of size $n_i' = n_i/L$ under the assumption that n_i are divisible by L and $L \geq 2$. The variances for each group of clusters are computed and treated as asymptotically normal with equal means and variances. Thus under H_{01}

$$X_G = \frac{\bar{s}_1^2 - \bar{s}_2^2}{\sqrt{\hat{V}^{+}(s_1^2 - s_2^2)}}$$

is approximately distributed as $t_{(n'-1)}$. Where

$$\bar{s}_i^2 = \frac{1}{n'}\sum_{g=1}^{n'} s_{i,g}^2$$

The $\hat{V}^{+}(s_1^2 - s_2^2)$ can be obtained by using

$$\hat{\Gamma}_i^{+} = \frac{1}{n'-1}\sum_{g=1}^{n'}(s_{i,g}^2 - \bar{s}_i^2)^2, \quad \text{and}$$

$$\hat{\Gamma}_{12}^{+} = \frac{1}{n'-1}\sum_{g=1}^{n'}(s_{1,g}^2 - \bar{s}_1^2)(s_{2,g}^2 - \bar{s}_2^2).$$

The asymptotic null distribution of X_G^2 is χ_1^2

$$n^{\frac{1}{2}}\,(s_i^2-\sigma_i^2)\xrightarrow{dist.} N\,(0,\Gamma_i)\;as\;n\to\infty \tag{2.4.1}$$

But here s_1^2 and s_2^2 are not independent. Instead (2.4.1)

$$n^{\frac{1}{2}}\left[\begin{bmatrix}s_1^2\\ s_2^2\end{bmatrix}-\begin{bmatrix}\sigma_1^2\\ \sigma_2^2\end{bmatrix}\right]\xrightarrow{dist.} N\left[\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\Gamma_1 & \Gamma_{12}\\ \Gamma_{21} & \Gamma_2\end{bmatrix}\right] \tag{2.4.2}$$

$$\therefore\quad V=V(s_1^2-s_2^2)=\frac{1}{n}(\Gamma_1+\Gamma_2-2\Gamma_{12})\;because\;\Gamma_{12}=\Gamma_{21}$$

2.5 Description of tests

(a) STANDARD ERROR TEST

From (2.4.2) under H_{01} the test statistic:

$$X_{TE}=\frac{s_1^2-s_2^2}{\sqrt{\hat{V}(s_1^2-s_2^2)}}$$

is approximately distributed as $t_{(n-1)}$ provided $\hat{V}$ is consistent estimate of V. The asymptotic null distribution of X_{TE}^2 *is* χ_1^2.

The Theorem 1 provides the consistent estimate of V.

Theorem 1

The Taylor expansion estimate of V is:

$$\hat{V}(s_1^2-s_2^2)=\frac{1}{n(n-1)}\sum_{c=1}^{n}(\bar{z}_c-\bar{z})^2,$$

where

$$\bar{z}=\frac{1}{n}\sum_{c=1}^{n}\bar{z}_c \qquad \text{and}$$

H	Number of strata in U.
N_h	Total number of clusters in h-th stratum in U.
n_h	Number of clusters in h-th stratum of a sample.
m_{ihc}	Observations in c-th cluster of h-th stratum.
$N = \sum_{h=1}^{H} N_h$	Number of clusters in U.
$n = \sum_{h=1}^{H} n_h$	Number of clusters in a sample.
$N_{oih} = \sum_{c=1}^{N_h} m_{ihc}$	Number of observations in h-th stratum.
$n_{oih} = \sum_{c=1}^{n_h} m_{ihc}$	Number of observations in a sample from h-th stratum.
$N_{oi} = \sum_{h=1}^{H} N_{oih} = \sum_{h=1}^{H}\sum_{c=1}^{N_h} m_{ihc}$	Finite Population size.
$n_{oi} = \sum_{h=1}^{H} n_{oih} = \sum_{h=1}^{H}\sum_{c=1}^{n_h} m_{ihc}$	Sample size
$W_{ih} = \frac{N_{oih}}{N_{oi}}$	Stratum weight.
x_{ihce}	e-th observation of c-th cluster of h-th stratum. $e = 1,2,\ldots\ldots, m_{ihc}$
$\overline{X}_{ih..} = \frac{1}{N_{oih}} \sum_{c=1}^{N_h}\sum_{e=1}^{m_{ihc}} x_{ihce}$	Mean of h-th stratum.
$\overline{x}_{ih..} = \frac{1}{n_{oih}} \sum_{c=1}^{n_h}\sum_{e=1}^{m_{ihc}} x_{ihce}$	Mean of h-th stratum of a sample.

(c) JACKKNIFE TEST

Let

$$s_{i,c}^2 = n_i s_i^2 - (n_i - 1)\, s_{i-c}^2$$

The s_{i-c}^2 are sample variances by using $n_i - 1$ clusters with c-th cluster omitted. The jackknife estimators are

$$s_i^{2*} = n s_i^2 - \frac{(n-1)}{n} \sum_{c=1}^{n} s_{i-c}^2$$

The $s_{i,c}^2$ are approximately independent and have asymptotically equal means and variances. That being so, under H_{01}, the test statistic

$$X_J = \frac{s_1^{2*} - s_2^{2*}}{\sqrt{\hat{V}^*(s_1^2 - s_2^2)}}$$

has approximately distributed as $t_{(n-1)}$, Wolter (1985). The $\hat{V}^*(s_1^2 - s_2^2)$ can be obtained by using

$$\hat{\Gamma}_i^* = \frac{1}{n-1} \sum_{c=1}^{n} (s_{i,c}^2 - s_i^{2*})^2 \qquad \text{and}$$

$$\hat{\Gamma}_{12}^* = \frac{1}{n-1} \sum_{c=1}^{n} (s_{1,c}^2 - s_1^{2*})\,(s_{2,c}^2 - s_2^{2*}).$$

The asymptotic null distribution of X_J^2 *is* χ_1^2.

3. TESTS UNDER STRATIFIED CLUSTER SAMPLING DESIGN

3.1 Notation

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

U : Union of finite populations. (It is assumed that each cluster of U contains more than one element of i-th finite population).

$$n^{\frac{1}{2}}\left[\begin{bmatrix} s_1^2 \\ s_2^2 \end{bmatrix} - \begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \end{bmatrix}\right] \xrightarrow{dist.} N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_1 & \psi_{12} \\ \psi_{21} & \psi_2 \end{bmatrix}\right] \qquad (3.4.1)$$

$$\therefore \quad V = V(s_1^2 - s_2^2) = \frac{1}{n}(\psi_1 + \psi_2 - 2\psi_{12}) \textit{ because of } \psi_{12} = \psi_{21}$$

$$= (V_1 + V_2 - 2V_{12})$$

3.5 Description of tests

(a) STANDARD ERROR TEST

From (3.4.1) the test statistic

$$X_{TE} = \frac{s_1^2 - s_2^2}{\sqrt{\hat{V}(s_1^2 - s_2^2)}}$$

is distributed approximately as $t_{(n-1)}$ under H_{02} if $\hat{V}$ is consistent estimate of V. The asymptotic null distribution of X_{TE}^2 is χ_1^2. To estimate V by Taylor expansion method applying Theorem 1 define

$$\bar{z}_{hc} = \frac{W_{1h}\, n_h}{n_{o1h}} \sum_{e=1}^{m_{1hc}} (x_{1hce} - \bar{x}_{1\ldots})^2 - \frac{W_{2h}}{n_{02h}} \sum_{e=1}^{m_{2hc}} (x_{2hce} - \bar{x}_{2\ldots})^2 - n_h \left(\frac{W_{1h}\, m_{1hc}}{n_{o1h}} \bar{x}_{1\ldots}^2 + \frac{W_{2h}\, m_{2hc}}{n_{o2h}} \bar{x}_{2\ldots}^2 \right) \text{an}$$

$$\text{d } \bar{z}_h = \frac{1}{n_h} \sum_{c=1}^{n_h} \bar{z}_{hc}$$

Then

$$\hat{V}(s_1^2 - s_2^2) = \sum_{h=1}^{H} \frac{1}{n_h(n_h - 1)} \sum_{c=1}^{n_h} (\bar{z}_{hc} - \bar{z}_h)^2$$

$$\overline{X}_{i..} = \frac{1}{N_{oi}} \sum_{h=1}^{H} \sum_{c=1}^{N_h} \sum_{e=1}^{m_{ihc}} x_{ihce} \quad \text{Mean of finite population.}$$

$$\overline{x}_{i..} = \sum_{h=1}^{H} \frac{W_{ih}}{n_{oih}} \sum_{c=1}^{n_h} \sum_{e=1}^{m_{ihc}} x_{ihce} \quad \text{Weighted mean of a sample.}$$

$$S_i = \frac{1}{N_{oi}} \sum_{h=1}^{H} \sum_{c=1}^{N_h} \sum_{e=1}^{m_{ihc}} (x_{ihce} - \overline{X}_{i..})^2 \quad \text{Variance of finite population}$$

$$s_i^2 = \sum_{h=1}^{H} \frac{W_{ih}}{n_{oih}} \sum_{c=1}^{n_h} \sum_{e=1}^{m_{ihc}} (x_{ihce} - \overline{x}_{i..})^2 \quad \text{Weighted sample variance.}$$

3.2 Sampling design

The n_h clusters are selected from the N_h clusters of U by simple random sampling. The each cluster of n_h is separated for the respective finite population to obtain the sample from the h-th stratum of the finite populations. The drawings are independent in different strata. The clusters obtained for each stratum of finite populations together comprise the full sample.

3.3 Problem under study

It is assumed that x_{ihce} are random variables, which implies that $S_{i.}^2$'s are also random variables. Furthermore it is assumed that S_i^2's converge to σ_i^2's (variance of superpopulation) as N_{oi} becomes larger. The hypothesis of interest is:

$$H_{02}: \sigma_1^2 = \sigma_2^2 \quad vs \quad H_{A2}: \sigma_1^2 \neq \sigma_2^2$$

3.4 Properties of Sample Variances

From (2.4.1) and (2.4.2)

$$s_{ih}^{2*} = \frac{1}{n_h}\sum_{c=1}^{n_h} s_{ih,c}^2$$

and s_i^{2*} are jackknife estimators based on full sample. The $s_{ih,c}^2$'s are approximately independent and have asymptotically equal means and variances. Thus

$$X_J = \frac{s_1^{2*} - s_2^{2*}}{\sqrt{\hat{V}^*(s_1^2 - s_2^2)}}$$

is approximately distributed as $t_{(n-1)}$ under H_{02}. The $\hat{V}^*(s_1^2 - s_2^2)$ can be determined by using

$$\hat{V}_i^* = \sum_{h=1}^{H}\frac{1}{n_h(n_h-1)}\sum_{c=1}^{n_h}(s_{ih,c}^2 - s_{ih}^{2*})^2 \quad \text{and}$$

$$\hat{V}_{12}^* = \sum_{h=1}^{H}\frac{1}{n_h(n_h-1)}\sum_{c=1}^{n_h}(s_{1h,c}^2 - s_{1h}^{2*})(s_{2h,c}^2 - s_{2h}^{2*})$$

The asymptotic null distribution of X_J^2 *is* χ_1^2

4. CONCLUDING REMARKS

These tests can be extended to obtain the confidence intervals for two variances when finite populations cut across the clusters.

(b) GROUPING TEST

The parent sample from h-th stratum is randomly divided into groups of size L, i.e.

$$n_{ih} = Ln'_{ih} \;\; for \;\; L \geq 2 \; (n_{ih} \; are \; divisible \; by \; L).$$

It is assumed that variance for each group of clusters of h-th stratum, is asymptotically normal with equal means and variances. Thus

$$X_G = \frac{\bar{s}_1^2 - \bar{s}_2^2}{\sqrt{\hat{V}^+(s_1^2 - s_2^2)}}$$

is approximately distributed as $t_{(n'-1)}$ under H_{02}. Where

$$\bar{s}_i^2 = \sum_{h=1}^{H} \bar{s}_{ih}^2$$

and $\hat{V}^+(s_1^2 - s_2^2)$ can be obtained by using

$$\hat{V}_i^+ = \sum_{h=1}^{H} \frac{1}{n'_h(n'_h - 1)} \sum_{g=1}^{n_h} (s_{ih,g}^2 - \bar{s}_{ih}^2)^2 \qquad \text{and}$$

$$\hat{V}_{12}^+ = \sum_{h=1}^{H} \frac{1}{n'_h(n'_h - 1)} \sum_{g=1}^{n_h} (s_{1h,g}^2 - \bar{s}_{1h}^2)\,(s_{2h,g}^2 - \bar{s}_{2h}^2)$$

The asymptotic null distribution of X_G^2 *is* χ_1^2

(c) JACKKNIFE TEST

Following Wolter (1985) let

$$s_{ih,c}^2 = \{H(n_h - 1) + 1\} s_i^2 - H(n_h - 1)\, s_{ih-c}^2$$

where s_{ih-c}^2's are variances after deleting (h,c)th cluster. The jackknif estimators are the average of $s_{ih,c}^2$, i.e.