Jour. Ind. Soc. Ag. Statistics
44 (3), 1992 : 199–215

Comparing Coefficients of Variation in Income Distributions

M. Singh and R.P. Singh
International Crops Research Institute for the Semi-Arid Tropics, Patancheru, P.O. 502324, A.P.
(Received : March, 1986)

Summary

The likelihood ratio test is used to compare the coefficients of variation from univariate and multivariate log-normal and Pareto distributions. Income stability measured as coefficient of variation in several populations and income sources have been compared with the help of income data from the farmers of the Semi-Arid Tropics of India.
Keywords: Coefficient of variation, Income inequality, log-normal and Pareto distributions, maximum likelihood estimates.

Introduction

Economists often use coefficients of variation (CV) to measure the fluctuation, stability and inequality in incomes of people classified by different agro-climatic regions and socio-economic groups. The coefficient of variation, many times may be very close to each other or differ widely. The comparision made only on the basis of values of CV without considering its sampling distribution may not clearly indicate whether the groups differ significantly in terms of inequality, stability or fluctuations. Hence it seems to be worthwhile to develop a statistical procedure to test the differences in the variability in income as measured by coefficient of variation. Furthermore, a household receives income from different sources and it becomes necessary to see which source of income is more stable than other for designing a policy for stabilizing the income.

The distribution of income is frequently assumed to be log-normal and sometimes Pareto distribution (see Bresciani-Turroni, [1], Malik, [5], Miyoji, [7], Moothathu, [8]. Similarly for several sources of income, the marginal distribution of income may be log-normal or Pareto and their joint distributions can be approximated by multivariate log-normal or multivariate Pareto.

In section 2 the likelihood ratio test (LRT) is used to test equality of coefficients of variation of income for several (k) independent samples, one from each of k populations of incomes assuming the univariate log-normal and univariate Pareto distributions for income. For several sources the results have been given in Section 3 under the assumption of multivariate log- normal and pareto distribution as joint distribution of income due to different sources. Coefficients of variation in two component incomes across three regions of Semi-Arid Tropics (SAT) of India are presented in section 4.

The two problems as above can be parameterised as follows.

I : We have k independent samples with values

$$x_{11}, \ldots, x_{1n_1}; x_{21}, \ldots, x_{2n_2}; x_{k1}, \ldots, x_{kn_k}$$

and of respective sizes $n_1, \ldots n_k$ drawn from populations denoted by $\Pi_1, \ldots \Pi_k$ respectively where x_{ij} $(i=1 \ldots k, j=1 \ldots n_i)$ is the income value of j-th sample household in i-th population. Let the probability density function of a variable (income) from the i-th population be expressed by $f_{\Pi_i}(C_i, \underline{\theta}_i)$, where C_i is the coefficient of variation of income and $\underline{\theta}_i$ some nuisance parameters. $(C_i > 0)$

The hypothesis to be tested is

$$H_0 : C_1 = C_2 = \ldots = C_k$$

that is, the hypothesis of equality of coefficient of variation against the alternative hypothesis that at least two CV's are different.

II : A single p-multivariate sample of size n and p income sources.

$$(\underline{x}_1, \ldots, \underline{x}_n) = \begin{bmatrix} x_{11} & . & . & x_{1n} \\ . & . & . & . \\ . & . & . & . \\ x_{p1} & . & . & x_{pn} \end{bmatrix}$$

from a population Π $(\underline{C}, \underline{\theta})$,

where $\underline{C} = (C_1, \ldots, C_p)$ is the parameter vector of CV's of the marginal distribution of individual sources and $\underline{\theta}$ a vector of nuisance parameters. The hypothesis structure is the same as in Part I.

The random sample considered here is simple random sample with replacement for the theoretical development of the statistics in this paper. These results will be applied to the sample drawn from finite but large population, ignoring the finite population corrections in section 4.

2. *LRT for Comparing C.V.'S from Univariate Distributions*

2.1 Likelihood Ratio Test

For the sake of completeness LRT will be briefly mentioned. LRT proposed by Neyman and Pearson [10] has been explained by Kendall and Stuart [3], Rao [11] among others. Let the likelihood function of the parameter θ given the observations x from the distribution f(x, θ) be L(θ/x), $\theta \varepsilon\ \Omega$, the parameter space for θ. Let the hypothesis H_0 regarding the parameter θ confines θ such that $\theta \varepsilon\ \omega \subset \Omega$. Neyman and Pearson [10] proposed the likelihood ratio criterion

$$\lambda = \max_{\theta \varepsilon \omega} L(\theta/x) / \max_{\theta \varepsilon \Omega} L(\theta/x) \tag{1}$$

for testing the hyposesis H_o.

The test statistic given in (1) can be written for a composite hypothesis in terms of maximum likelihood estimators $\hat{\theta}_\omega$ (under hypothesis H_0) and $\hat{\theta}_\Omega$ (under no restriction on parameter) of θ as

$$Q = -2 \ln(\lambda) = 2 \left\{ 1(\hat{\theta}_\Omega) - 1(\hat{\theta}_\omega) \right\} \tag{2}$$

Where $1(\hat{\theta}_\Omega)$ = max ln L(θ/x) and similarly $1(\hat{\theta}_\omega)$ under H_o. Hereafter we shall denote $1(\hat{\theta}_\Omega)$ by $1(\Omega)$ and $1(\hat{\theta}_\omega)$ by $1(\omega)$. The exact distribution of Q may be complicated in general situation but it asymptotically follows a chi-square distribution under H_0.

2.2 Log-normal Distributions

The probability density function of a univariate log-normal distribution (see Kendall and Stuart, [3], is

$$f(x,\theta) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right), \quad x>0,\ -\infty< \mu,< \infty\ ,\ \sigma> 0 \tag{3}$$

The mean, variance and CV are

$$E(x) = \exp(\mu + \sigma^2/2)$$
$$\text{Var}(x) = (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)$$
$$CV(x) = (\exp(\sigma^2) - 1)^{1/2} \qquad (4)$$

Thus the CV of a log-normal distribution depends on only one parameter σ. Further, we note that Y=log X follows a normal distribution with mean μ and variance σ^2. Therefore, the problem of testing the equality of CV of X is equivalent to testing the homogeneity of variance (σ^2) of Y. Test for homogeneity of variances, due to Bartlett (see Snedecor and Cochran, [12], p.252), is well known. Using the transformed variate values y = ln x, where x is the sample observation in j-th unit in i-th group i=1,...,k; j=1,...,n, the Bartlett's test is based on the statistic

$$Q_1 = \frac{\nu \ln s^2 - \sum_{i=1}^{k} \nu_i \ln s_i^2}{C}$$

where

$$C = 1 + \frac{\sum_{i=1}^{k} \frac{1}{\nu_i} - \frac{1}{\nu}}{3(k-1)}$$

$$s_i^2 = \sum \frac{(y_{ij} - \bar{y}_i)^2}{\nu_i}\ ;\ \bar{y}_i = \sum \frac{y_{ij}}{n_i}, \qquad y_{ij} = \ln x_{ij}$$

$$\nu_i = n_i - 1\ ;\ \nu = \sum \nu_i$$

$$s^2 = \sum \nu_i \frac{s_i^2}{\nu}$$

Q_1 has approximately a chi-square distribution with k–1 d.f. under the null hypothesis. For two sample (population) situation, we have an exact test

$$F = \frac{S_1^2}{S_2^2} \qquad (S_1^2 > S_2^2)$$

where statistic F has an F-distribution with v_1 and v_2 d.f.

2.3 Pareto Distribution

The probability density function of the univariate Patero distribution is (see Johnson and Kotz, [2])

$$f(x, \alpha, x_o) = \alpha x_o^{\alpha} x^{-(\alpha+1)} \qquad x > x_o > 0 \qquad (6)$$

The mean, variance and CV are

$$E(X) = \frac{\alpha x_o}{\alpha - 1} \qquad \alpha > 1$$

$$V(X) = \frac{\alpha x_o^2}{(\alpha-1)^2(\alpha-2)} \qquad \alpha > 2 \qquad (7)$$

$$CV(X) = \frac{1}{\sqrt{(\alpha(\alpha-2))}}$$

Thus the CV of Pareto distribution is a function of α. Therefore, in the following we shall develop the LRT for hypothesis

$$H: \alpha_1 = \alpha_2 \ldots = \alpha_k \; (= \alpha_o) \text{ say}$$

for comparing the CV for k-populations with probability density functions

$$f(x, \alpha_i, x_{oi}) = \alpha_i x_{oi}^{\alpha i} \quad x^{-(\alpha_i+1)}$$

$$x > x_{oi} > 0; \qquad \alpha_i > 2$$

for i-th population Π_i. Let the sample $x_{i1}, \ldots, x_{in_i}$, i=1, . . ., k be random sample from Π_i. Then the likelihood function for the 2k parameters α_i, x_{oi} i=1, . . ., k is given by

$$L(\underline{\alpha}, \underline{x}_o) = \prod_{i=1}^{k} \prod_{j=1}^{n_i} \alpha_i \, x_{oi}^{\alpha_i} \, x_{ij}^{-(\alpha_i+1)}$$

where $\underline{\alpha} = (\alpha_1, \ldots, \alpha_k)'$ and $\underline{x}_o = (x_{01}, \ldots, x_{ok})'$. The log likelihood

$$l(\underline{\alpha}, \underline{x}_o) = \sum_{i=1}^{k} n_i (\ln \alpha_i + \alpha_i \ln x_{oi}) - \sum_{i=1}^{k} (\alpha_i + 1) \sum_{j=1}^{n} \ln x_{ij}$$

The maximum likelihood estimates $\hat{\underline{\alpha}}$ and $\hat{\underline{x}}_o$ of $\underline{\alpha}$ and $\underline{x}_o$ are given by

$$\hat{\underline{\alpha}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_k)' ; \quad \hat{\underline{x}}_o = (\hat{\underline{x}}_{o1}, \ldots, \hat{\underline{x}}_{ok})'$$

where

$\hat{x}_{oi} = \min\limits_j \{x_{ij}\} = x_i(1)$ say, the smallest observation in the i-th sample;

$$\hat{\alpha}_i = \frac{n_i}{\sum\limits_{j=1} \ln x_{ij} - n_i \ln \hat{x}_{oi}} = \frac{n_i}{\sum\limits_{j=2}^{n_i} \ln\left(\frac{x_{ij}}{x_i(1)}\right)}$$

Thus the maximum log likelihood

$$l(\hat{\underline{\alpha}}, \hat{\underline{x}}) = \sum n_i (\ln \hat{\alpha}_i + \hat{\alpha}_i \ln \hat{x}_{oi}) - \sum (\hat{\alpha}_i + 1) \sum_{j=1} \ln x_{ij} = l(\Omega)$$

Under hypothesis H_0, we have m.l. estimate of common α_o

$$\hat{\alpha}_o = \frac{n}{\sum\limits_i \sum\limits_j \ln x_{ij} - \sum\limits_i n_i \ln \hat{x}_{oi}} = \frac{n}{\sum\limits_i \sum\limits_j \ln \frac{x_{ij}}{x_i(1)}}$$

where $n = \sum n_i$ and $\hat{\underline{x}}_{oi} = \hat{\underline{x}}_o$.

Maximum log likelihood is

$$l(\hat{\alpha}_o, \hat{x}_o) = \sum n_i (\ln \hat{\alpha}_o + \hat{\alpha}_o \ln \hat{x}_{oi}) - (\hat{\alpha}_o + 1) \sum_i \sum_j \ln x_{ij}$$

$$= l(\omega)$$

The LRT statistic for H_o is

$$Q_2 = 2[1(\Omega) - 1(\omega)] = 2\left[\sum (\hat{\alpha}_o - \hat{\alpha}_i) \sum_{j=1}^{n_i} \ln\frac{x_{ij}}{x_i(1)} - \sum_{i=1}^{k} n_i \ln\frac{\hat{\alpha}_o}{\hat{\alpha}_i}\right] \quad (9)$$

Q_2 has approximate distribution in χ^2_{k-1} d.f. when H_o is true.

3. *LRT for Comparing CV's of Marginal Components in Multivariate Distributions*

3.1 Multivariate Log-normal Distributions

A p-component multivariate log-normal is defined as the distribution, of $\underline{X} = (X_1, \ldots, X_p)'$ where $\underline{Z} = (Z_1, \ldots, Z_p)$, with $Z_j = \ln X_j$ $j=1,\ldots,p$, has p-variate distribution (see Johnson and Kotz [2]). Thus, if $\underline{Z}$ follows $N_p(\underline{\mu}, \underline{\Sigma})$, then the moments of $\underline{X}$ are given by

$$\mu_{r1, r2, \ldots, rp}(\underline{X}) = E\left(\prod_{j=1}^{p} X_j^{r_j}\right)$$

$$= \exp(\underline{r}'\underline{\mu} + \underline{r}\ \underset{\sim}{\Sigma} r/ / 2)$$

where $\underline{r} = (r_1, \ldots, r_p)'$.

The correlation coefficient between the components X_j and X_k is

$$\mathrm{Corr}(X_j, X_k) = \{\exp(\rho_{jk}\sigma_j\sigma_k) - 1\}\left[\{\exp(\sigma_j^2) - 1\}\{\exp(\sigma_k^2) - 1\}\right]^{-1/2}$$

and CV of X_j is

$$C_j = \mathrm{CV}(X_j) = \sqrt{\exp(\sigma_j^2) - 1}$$

where $$\underset{\sim}{\sum} = (\rho_{ij}\sigma_i\sigma_j)$$

Thus we require an LRT for the homogeneity of marginal variances σ_j^2. Suppose we have a p-variate sample

$$\underset{\sim}{X} = (\underline{X}_1, \ldots, \underline{X}_n) = \begin{bmatrix} X_{11} & . & X_{1n} \\ . & . & . \\ X_{p1} & . & X_{pn} \end{bmatrix}$$

of size n from a multivariate log-normal distribution of the above form, then use the transformation to get

$$Z = (\underline{Z}_1, \ldots, \underline{Z}_l) = \begin{bmatrix} Z_{11} & . & Z_{1n} \\ . & . & . \\ Z_{p1} & . & Z_{pn} \end{bmatrix}$$

where $Z_{ij} = \ln X_{ij}$; $j = 1, \ldots, n$, $k = 1, \ldots, p$.

Here $\underline{Z}_j$'s follow $N_p\left(\underline{\mu}, \underset{\sim}{\Sigma}\right)$. Hypothesis to be tested is

$$H_o : \sigma_1^2 = \sigma_2^2 = \ldots, \sigma_p^2 \ (= \sigma_0^2, \text{ say.})$$

The likelihood function for $\underline{\mu}, \underset{\sim}{\Sigma}$ using Z is

$$L\left(\underline{\mu}, \underset{\sim}{\Sigma}\right) = \left|\underset{\sim}{\Sigma}\right|^{-n/2} \frac{\exp\left[-\sum_{i=1}^{n} (\underline{Z}_j - \underline{\mu})' \underset{\sim}{\Sigma}^{-1} (\underline{Z}_j - \underline{\mu})\right]}{(2\pi)^{pn/2}}$$

The maximum likelihood estimates of $\underline{\mu}$ and $\underset{\sim}{\Sigma}$ are

$$\hat{\underline{\mu}} = \underline{\overline{Z}}$$

$$\hat{\underset{\sim}{\Sigma}} = n^{-1} \underset{\sim}{S}$$

where

$$\underline{\overline{Z}} = (\overline{Z}_1, \ldots, \overline{Z}_p)', \ \overline{Z}_i = \sum_{j=1}^{n} \frac{Z_{ij}}{n}, \qquad i = 1, \ldots, p$$

$$\underset{\sim}{S} = (S_{ii'}), \ S_{ii'} = \sum_{j=1}^{n} (Z_{ij} - \overline{Z}_i)(Z_{i'j} - \overline{Z}_{i'})$$

We get after some simplification, the value of likelihood (see section 2.2)

$$l\ (\Omega) = \text{const} - (n/2) \log \left|n^{-1} \underset{\sim}{S}\right| - \frac{np}{2}$$

where $\left|\underset{\sim}{A}\right|$ stands for the determinant of matrix $\underset{\sim}{A}$.

When the hypothesis H_0 is true, then we have the following maximum likelihood estimators of the parameters

$$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^{p} S_{ii}}{np}$$

and $$\hat{\rho}_{ij} = \frac{S_{ij}}{n\,\hat{\sigma}_0^2}$$

The maximum of log likelihood under H_0, $l(\omega)$ thus becomes

$$l(\omega) = \text{const.} - (n/2)\ln\left|\underset{\sim}{S}^*\right| - (n/2)\,\text{tr}(\underset{\sim}{I} + n\,\hat{\sigma}_0^2\,\underset{\sim}{S}^{-1} - \underset{\sim}{S}^{-1}\,\underset{\sim}{S}_d)^{-1}$$

where $\underset{\sim}{S}^* = S^*_{ij}$, $\quad S^*_{ii} = \hat{\sigma}_0^2$, $\quad S^*_{ij} = \hat{\rho}_{ij}\,\hat{\sigma}_0^2$

$$\underset{\sim}{S}_d = \text{diag}\,(S_{11}, S_{22}, \ldots, S_{pp})$$

and tr(A) represent the trace of matrix A. Hence the LRT statistics for testing H_0 is

$$Q_3 = n\ln\left|\underset{\sim}{G}\right| + n\,\text{tr}\,\underset{\sim}{G} - np$$

where

$$\underset{\sim}{G} = \underset{\sim}{I} + n\,\tilde{\sigma}_0^2\,\underset{\sim}{S}^{-1} - \underset{\sim}{S}^{-1}\,\underset{\sim}{S}_d$$

$$\hat{\sigma}_0^2 = \frac{\text{tr}\,\underset{\sim}{S}_d}{np},$$

and Q_3 has approximate null distribution as χ^2_{p-1}.

A Special Case (p=2) the expression for Q_3 reduces to

$$Q_3 = n\ln\frac{1-\hat{\hat{\rho}}^2}{1-\hat{\rho}^2} + 2n\ln\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1\,\hat{\sigma}_2} \qquad (11)$$

with

$$\hat{\sigma}_i^2 = \sum_{j=1}^{n} \frac{(z_{ij} - \bar{z}_i)^2}{n-1} \qquad i = 1, 2$$

$$\hat{\sigma}_0^2 = \frac{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}{2}$$

$$\hat{\rho} = \sum \frac{(z_{1j} - \bar{z}_1)(z_{2j} - \bar{z}_2)}{n\,\hat{\sigma}_1\,\hat{\sigma}_2}$$

$$\hat{\hat{\rho}} = \hat{\rho}\,\hat{\sigma}_1 \frac{\hat{\sigma}_2}{\hat{\sigma}_0^2} .$$

Tests are available for some hypotheses on the structures of the variance covariance matrix of multivariate normal vectors in Rao [11] and Lee, Chang and Krishnaiah [4]. Further, these hypotheses also restrict covariances along with variances. But in the present case only variances of correlated normal variables are subjected to homogeneity.

3.2 Multivariate Pareto Distribution

Multivariate Pareto distributions have been studied by Mardia [6] (see Johnson and Kotz, [2]) among others. In this section, we consider the multivariate Pareto distribution where the marginal components $X_1, \ldots, X_p$ have pareto distributions with following density function

$$P_{x_j}(x_j, a_j, \theta_j) = a_j\, \theta_j^{a_j}\, x_j^{-(a_j - 1)}$$

and the variables

$$Y_j = a_j \ln\left(\frac{X_j}{\theta_j}\right) \qquad (j = 1, \ldots, p)$$

have some form of standard multivariate exponential distributions.

The general case of p Pareto variate would be complicated, therefore, we confine ourselves to the case of bivariate Pareto distribution. Let X_1 and X_2 have above density function and Y_1 and Y_2 satisfy Morgenstern's [9] family of bivariate exponential distribution given by

$$P_{Y_1Y_2}(y_1, y_2) = \exp(-y_1 - y_2)\left\{1+\alpha(2\exp(-y_1)-1)(2\exp(-y_2)-1)\right\}$$

where parameter α brings the dependence between two components Y_1 and Y_2 and hence between X_1 and X_2. The probability density function of X_1 and X_2 can be written as

$$P_{X_1.X_2}(x_1, x_2) = a_1 a_2 \theta_1^{a_1} \theta_2^{a_2}\left[1+\alpha\left\{2\left(\frac{\theta_1}{x_1}\right)^{a_1} - 1\right\}\left\{2\left(\frac{\theta_2}{x_2}\right)^{a_2} - 1\right\}\right]$$

$$x_i > \theta_i > 0, \quad a_i > 2 \qquad (12)$$

The maximum log likelihood function value is

$$l(\Omega) = n\sum_{i=1}^{2}(\ln\hat{a}_i + \hat{a}_i \ln\hat{\theta}_i) - \sum_{i=1}^{2}(\hat{a}_i+1)\sum_{j=1}^{n}\ln x_{ij} + \sum_{j=1}^{n}\ln\left[1+\hat{a}\left\{2\left(\frac{\hat{\theta}_1}{x_{1j}}\right)^{\hat{a}_1} - 1\right\}\left\{2\left(\frac{\hat{\theta}_2}{x_{2j}}\right)^{\hat{a}_2} - 1\right\}\right]$$

where the sample consists of n pairs of observations

$$(x_{1j}, x_{2j}) \qquad j = 1, \ldots, n.$$

The maximum likelihood estimates $\hat{a}_1, \hat{a}_2, \hat{\theta}_1, \hat{\theta}_2, \hat{\alpha}$ satisfy the following equations: $\hat{\theta}_i = \min_j\{x_{ij}\} = x_i(1)$ (least value of the observed components) $i = 1, 2$

Equations for $\hat{a}_1, \hat{a}_2$, and $\hat{\alpha}$:

$$\frac{n}{a_i} + n\ln\hat{\theta}_i - \sum_{j=1}^{n}\ln x_{ij} + \sum_{j=1}^{n}\frac{\left\{2\hat{\alpha}\left(\frac{\hat{\theta}}{x_{ij}}\right)^{ai}\ln\left(\frac{\theta_i}{x_{ij}}\right)\right\}\left\{2\left(\frac{\theta_{i'}}{x_{i'j}}\right)^{a_{i'}} - 1\right\}}{\left[1+\alpha\left\{2\left(\frac{\hat{\theta}_1}{x_{1j}}\right)^{a_1} - 1\right\}\left\{2\left(\frac{\hat{\theta}_2}{x_{2j}}\right)^{a_2} - 1\right\}\right]} = 0$$

$$i = 1, 2 \quad \text{and } i' = 3 - i; \qquad (13)$$

$$\sum_{j=1}^{n}\frac{\left\{2\left(\frac{\theta_1}{X_{1j}}\right)^{\hat{a}_1} - 1\right\}\left\{2\left(\frac{\theta_2}{X_{2j}}\right)^{\hat{a}_2} - 1\right\}}{1+\alpha\left\{2\left(\frac{\hat{\theta}_1}{X_{1j}}\right)^{\hat{a}_1} - 1\right\}\left\{2\left(\frac{\hat{\theta}_2}{X_{2j}}\right)^{\hat{a}_2} - 1\right\}} = 0 \qquad (14)$$

The solution of the above equations can be done using iteration algorithm.

Under the hypothesis $H_0 : a_1 = a_2$ (= a), say, we have

$$1(\omega) = 2n \ln \hat{a} + n\hat{a} \ln (\hat{\theta}_1 \theta_2) - (\hat{a} + 1) \sum_{j=1}^{n} \ln (x_{1j} x_{2j}) +$$

$$\sum_{j=1}^{n} \ln \left[1 + \hat{\alpha} \left\{ 2\left(\frac{\hat{\theta}_1}{x_{1j}}\right)^{\hat{a}} - 1 \right\} \left\{ 2\left(\frac{\hat{\theta}_2}{x_{2j}}\right)^{\hat{a}} - 1 \right\} \right],$$

where $\hat{a}$ and $\hat{\alpha}$ can be obtained from the following.

For $\hat{a}$:

$$\frac{2n}{a} + n \ln(\hat{\theta}_1 \hat{\theta}_2) - \sum^{n} \ln (x_{1j} x_{2j}) +$$

$$\sum_{j=1}^{n} \frac{2\hat{\alpha} \left(\frac{\hat{\theta}_1}{x_{1j}}\right)^{\hat{a}} \ln\left(\frac{\hat{\theta}_1}{x_{1j}}\right) \left\{ 2\left(\frac{\hat{\theta}_2}{x_{2j}}\right)^{\hat{a}} - 1 \right\} + \left(\frac{\hat{\theta}_2}{x_{2j}}\right)^{\hat{a}} \ln\left(\frac{\hat{\theta}_2}{x_{2j}}\right) \left\{ 2\left(\frac{\hat{\theta}_1}{x_{1j}}\right)^{\hat{a}} - 1 \right\}}{1 + \hat{\alpha} \left\{ 2\left(\frac{\hat{\theta}_1}{x_{1j}}\right)^{\hat{a}} - 1 \right\} \left\{ 2\left(\frac{\hat{\theta}_2}{x_{2j}}\right)^{\hat{a}} - 1 \right\}} = 0 \qquad (15)$$

and for $\hat{\hat{a}}$

$$\sum_{j=1}^{n} \frac{\left\{ 2\left(\frac{\hat{\theta}_1}{x_{1j}}\right)^{\hat{a}} - 1 \right\} \left\{ 2\left(\frac{\hat{\theta}_2}{x_{2j}}\right)^{\hat{a}} - 1 \right\}}{1 + \hat{\hat{a}} \left\{ 2\left(\frac{\hat{\theta}_1}{x_{1j}}\right)^{\hat{a}} - 1 \right\} \left\{ 2\left(\frac{\hat{\theta}_2}{x_{2j}}\right)^{\hat{a}} - 1 \right\}} = 0 \qquad (16)$$

After solving these equations and substituting back into expressions for $1(\Omega)$ and $1(\omega)$, we get

$$Q_4 = 2[1(\Omega) - 1(\omega)] \qquad (17)$$

and Q_4 will be approximately distributed as chi square distribution with 1 d.f.

4. *Results and Discussion*

The test statistic discussed in the proceeding sections can be explained by using the actual data. For comparing the coefficients of variation we have used income data collected from a set of 40 rural households in three agroclimatic regions in semi-arid tropical areas of central peninsular India. The data used for analysis refer to average of five cropping years (1975-76 to 1979-80). The income used was real income suitably adjusted for the price variation over time.

Table 1 shows per household and per capita mean income and coefficient of variation across three regions. The table clearly indicates that mean income in Mahabubnagar and Sholapur region is relatively low compared to Akola region. For the sake of convenience total income has been grouped into two major components labour and others (including crops, livestock, rental, handicraft and trade, and transfers). Variability (measured in terms of coefficient of variation) in labour income is quite high in Mahabubnagar compared to Sholapur and Akola. This is mainly

Table 1. Means and coefficients of variation⁺ in per household (HH) and percapita income (net returns to family owned resources) by regions (1975-76 to 1979-80)

Regions	Income sources: Labor		Other		Total	
	Per HH	Per capita	Per HH	Per capita	Per HH	Per capita
Mahabubnagar	494 (109)	87 (101)	2964 (111)	563 (102)	3440 (86)	650 (80)
Sholapur	1821 (51)	319 (58)	2107 (77)	348 (77)	3928 (50)	667 (55)
Akola	1905 (55)	375 (48)	3329 (140)	616 (145)	5234 (92)	991 (90)
All	1404 (77)	260 (77)	2812 (123)	513 (125)	4215 (84)	773 (84)

⁺Figures in parenthesis are coefficient of variations in %

because of active labour market and more employment opportunities in the latter regions than in Mahabubnagar where employment opportunities outside the farms are meagre. Similarly, income from sources other than labour is also highly variable in resource rich areas where cotton cash crops are largely planted by resource rich farmers.

The variability in income expressed in terms of coefficient of variation widely differs under different distributional forms. This suggests that assumptions of a common distributional form always may not be appropriate under all situations. It is clear from Table 2 that estimated coefficient of variations in the total income do not change substantially while assuming log- normal distribution but if CVs are estimated for different components of income under log-normal distribution it varies quite substantially. This suggests that choice of distributional form has a strong implications in describing the variability in income. Hence, for comparing the variability in any population one may think of different appropriate

Table 2. Estimated coefficient of variation (%) in income* under log-normal distribution by region.

Income sources						
Regions	Labor		Others		Total	
	Per HH	Per capita	Per HH	Per capita	Per HH	Per capita
Mahabub nagar	10282	1382	213	222	79	76
Sholapur	46	51	116	96	57	46
Akola	71	59	229	184	89	66
All	1584	550	181	173	77	67

* Denotes net return to family owned resources; HH=household.

distributional forms and a suitable parametric or nonparametric measures of variation.

The coefficient of variation in the incomes across three regions has been compared assuming log-normal distribution in Table 3.

Table 3. Values of "Test Statistics" (Q1) and associated probabilities (Pr)[a] for comparing coefficients of variation across regions under log-normal distribution.

Test statistics and probability				
Sources of income	Per household		Per capita	
	Q_1	P_r	Q_1	Pr
Mahabubnagar	135.00	0.00	104.30	0.00
Sholapur	5.50	0.06	8.70	0.01
Total	4.00	0.09	6.80	0.03

a P_r = probability $\left[X_2^2 > Q_1\right]$

There seems to be no significant differences in variability in total income across regions if the household is considered as a unit of observation. But when the variability in per capita income is compared, differences are sharp. This raises the question that while analysing the different aspects of income distribution or inequality whether household or a member should be chosen as the basis of analysis.

The variability in different components of income also shows the same trend. There are significant differences in the variability in labour income across three regions but in other sources of income the differences are not significant if household is considered as a unit of observation. But when it is estimated on per capita basis it shows significant differences in the income across three regions. This further supports the explanation given in Table 2.

Moreover, in each region variability in different sources of income can also be compared with the help of an appropriate test statistics (Q_3) using a bivariate log-normal distribution (see section 3.1). Table 4 indicates that there are highly significant differences in CV's among different components on income in all the regions whether household or per capita is considered as the unit of analysis. The estimated value of $\hat{\rho}$ shows that except in Mahabubnagar region, where the two sources of income are negatively but highly correlated, there is poor correlation between the two sources of income in the regions.

Table 4. Test Statistics values (Q_3), associated probabilities $(Pr)^a$, $\hat{\rho}$ and $\hat{\hat{\rho}}$ for comparing coefficients of variation in labor and other sources of income across three regions (under Bivariate lognormal distributions)

Test statistics (Q_3) probablities and correlation coefficents								
Regions	Per Household				Per Capita			
	Q_3	P_r	$\hat{\rho}$	$\hat{\hat{\rho}}$	Q_3	Pr	$\hat{\rho}$	$\hat{\hat{\rho}}$
Mahabubnagar	99.00	0.00	–0.60	–0.19	109.91	0.00	–0.67	–0.20
Sholapur	10.55	0.00	0.12	0.11	5.16	0.02	0.26	0.24
	64.30	0.00	0.02	0.01	72.03	0.00	–0.10	–0.04
regions	125.00	0.00	–0.08	–0.05	126.20	0.00	–0.11	–0.06

a Pr = probability $(\chi_1^2 > Q_3)$

b $\hat{\rho}$ and $\hat{\hat{\rho}}$ can be seen in section (3.1)

ACKNOWLEDGMENTS

Authors express their thanks to a referee for his valuable suggestions leading to an improvement in the paper and Mrs. Mayada Shehabi for her assistance on word processing.

REFERENCES

[1] Bresciani-Turroni, C., 1939. Annual survey of statistical data: Pareto's law and index of inequality of incomes. *Econometrica*, **7**, 107-133.

[2] Johnson, N.L., and Kotz, S. 1972. *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley and Sons, Inc., New York.

[3] Kendall, M.G., and Stuart, A. 1979. *The advanced Theory of Statistics. Vol. II, 4th Ed.* (Charles Griffin and Company Ltd., London).

[4] Lee, J.C., Chang, T.C. and Krishnaiah, P.R., 1977. Approximation to the distributions of the likelihood ratio for testing certain structures on the covariance matrices of real multivariate normal populations. *Multivariate Analysis - IV.* pp. 105-118. North- Holland Publishing Company; Amesterdam.

[5] Malik, H.J., 1970. Estimation of the parameters of the Pareto distribution.

Metrika, **15**, 126-132.

[6] Mardia, K.V., 1962. Multivariate Pareto distributions. *Annals of Mathematical Statistics* **33**, 1008-1015.

[7] Miyoji Hayakawa, 1951. The application of Pareto's law of income to Japanese data. *Econometrica*, **19**, 174-183.

[8] Moothathu, T.S.K., 1985. Sampling distributions of Lorenz Curve and Gini index of the Pareto distribution. *Sankhya*. **47**, Series B, 247-258.

[9] Morgenstern, D. 1956. Einfache Beispiele Zweidimensionaler Verteilungen. *Mitteilungsblatt fur mathmatische Statistik*, **8**, 234-235.

[10] Neyman, J., and Pearson, E.S., 1928. On the use and interpretation of certain test criteria for the purpose of statistical inference. *Biometrika*, **20A**, 175-240 and 263-294.

[11] Rao. C.R., 1973, *Linar Statistical Inference and its Applications*. 2nd Ed. (Wiley Eastern Pvt. Ltd., New Delhi).

[12] Snedecor, G. W. And Cochran, W.G., 1980. *Statistical Methods* (seventh Ed.) The Iowa State University Press, Ames, Iowa.