ESTIMATION OF STRUCTURAL EQUATION MODELS THROUGH K- MEANS CLUSTER APPROACH - AN APPLICATION FOR ASSESSING SOCIO - ECONOMIC DEVELOPMENT IN HARYANA.

O.P. Sheoran, Lajpat Rai and K.K.Saxena*

Department of Mathematics and Statistics
CCS Haryana Agricultural University, Hisar -125 004 (India)
*Department of Statistics, P.O.Box 338, UDOM, Tanzania

ABSTRACT

Structural equation models have been estimated with maximum likelihood technique assuming three dimensions for socio-economic sectors. Scores for latent variables from the best fitted model were computed by principal component analysis and K-means cluster analysis. These models have been fitted for the purpose of studying levels of socio economic development in Haryana State.

Keywords: Structural equation models, Maximum Likelihood Estimation, Cluster Analysis, Factor Analysis, Latent variables

Introduction

Socio-economic development in a region is related to the facilities of education, transport, communication, electricity, financial institutions and medical facilities etc. The pace of socio-economic development among various regions in a State can never be uniform over the years owing to historic, geographic, ecological and climatic conditions and some other unknown factors. Wide disparities may exist in the level of development in different regions of the state.

The motivation to research workers has always been concerned to find out the extent of these disparities with regard to agricultural, infrastructural, socio-economic and industrial development parameters. Policy makers for the development of a State need such systematic information to frame schemes, policies and programmes.

Statistically the problem is to identify the factors (latent variables) of development by estimating the structural equation model by using principal component analysis and k-means cluster analysis.

In the present paper, Principal component analysis has been employed to identify the factors responsible for the development. Structural Equation Modelling (SEM) technique has been used to develop the model of socio-economic development in the state of Haryana and classify the tehsils according to their levels of socio-economic development. The assessment of the overall fit of the model to data has been judged by Chi-square test statistic, Goodness of fit index (GFI), Adjusted goodness of fit index (AGFI), Root mean square residual (RMR) and Model modification Index.

Various authors have studied Multivariate techniques to distinguish regional disparities for future planning in socio- economic aspects. Narain et al.(1991) used multivariate technique for the first time to evaluate the economic development in Orissa . Then in a series of papers, Narain et al. (1994,2001,2007 ,2012) studied the evaluation and identification of development in various states of India. Lipshitz and Raveh,(1994,1998) used the co-plot technique and Multi-dimentional scaling method for the socio economic differences among various localities. Upadhyay (2001) used spatial and temporal analysis to assess the development in Rajasthan. Soares *et al.*, 2003, used factor and cluster analysis to uncover regional disparities in the European Union. Sheoran *et. al.* (2012) estimated the parameters of structural equation model with latent variables and used them for the assessment of the development of Haryana State.

2. Methodology

The measurement model for each dimension in the form of standard factor analytical model is given by

$$\mathbf{y} = \mathbf{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad (1)$$

for latent endogenous variables with $E(\boldsymbol{\varepsilon\varepsilon}') = \mathbf{\Theta}_\varepsilon$ and

$$\mathbf{x} = \mathbf{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta} \qquad (2)$$

for latent exogenous variables with $E(\boldsymbol{\delta\delta}') = \mathbf{\Theta}_\delta$

We also define $E(\boldsymbol{\varepsilon\delta}') = \mathbf{\Theta}_{\delta\varepsilon}$ and $E(\boldsymbol{\xi\xi}') = \mathbf{\Phi}$, where

y is a *p* x *1* vector of observed indicators of the dependent (endogenous) latent variable $\boldsymbol{\eta}$

x is a *q* x *1* vector of observed indicators of the independent (exogenous) latent variables ξ

$\boldsymbol{\eta}$ is a *m* x *1* random vector of latent dependent or endogenous variables

ξ is a *n* x *1* random vector of latent independent or exogenous variables

$\boldsymbol{\varepsilon}$ is a *p* x *1* vector of measurement error in **y**

$\boldsymbol{\delta}$ is a *q* x *1* vector of measurement error in **x**

$\boldsymbol{\Lambda}_y$ is a *p* x *m* matrix of coefficients of regression of **y** on $\boldsymbol{\eta}$ and

$\boldsymbol{\Lambda}_x$ is a *q* x *n* matrix of coefficients of regression of x on ξ

The implied covariance/correlation matrix $\boldsymbol{\Sigma(\theta)}$ is given by

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_x \boldsymbol{\Phi} \boldsymbol{\Lambda}'_x + \boldsymbol{\Theta}_\delta \tag{3}$$

with the assumptions E(**x**) = E(**δ**) = 0 and E(**ξδ'**) = E(**δξ'**) = 0,

Then the structural part of the model is given by

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\xi + \zeta \tag{4}$$

We also define $E(\zeta\zeta') = \boldsymbol{\Psi}$, where

B is a *m* x *m* coefficient matrix that relates endogenous variables to each other

$\boldsymbol{\Gamma}$ is a *m* x *n* coefficient matrix that relates endogenous variables to exogenous variables and ζ is a *m* x *1* vector of errors (residuals)

The correlation matrix implied by the model is comprised of three separate correlation matrices, i.e. correlation matrix of the observed indicators of the latent endogenous variables $\boldsymbol{\Sigma}_{yy}$,the correlations between the indicators of latent endogenous and indicators of latent exogenous variables is $\boldsymbol{\Sigma}_{yx}$ and the correlation matrix of the indicators of the latent exogenous variables $\boldsymbol{\Sigma}_{xx}$.Arranging the above three matrices together, we get the joint covariance matrix implied by the model, ie. $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{bmatrix}$.

The implied covariance/correlation matrix can be written in terms of model parameters as

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Lambda}_{\mathbf{y}}(\mathbf{I}-\mathbf{B})^{-1}(\boldsymbol{\Gamma\Phi\Gamma}' + \boldsymbol{\Psi})[(\mathbf{I}-\mathbf{B})^{-1}]'\boldsymbol{\Lambda}'_{\mathbf{y}} + \boldsymbol{\Theta}_{\varepsilon} & \boldsymbol{\Lambda}_{\mathbf{y}}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Gamma\Phi\Lambda}'_{\mathbf{x}} + \boldsymbol{\Theta}'_{\boldsymbol{\delta\varepsilon}} \\ \boldsymbol{\Lambda}_{x}\boldsymbol{\Phi\Gamma}'[(\mathbf{I}-\mathbf{B})^{-1}]'\boldsymbol{\Lambda}'_{y} + \boldsymbol{\Theta}_{\boldsymbol{\delta\varepsilon}} & \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\Phi\Lambda}'_{\mathbf{x}} + \boldsymbol{\Theta}_{\boldsymbol{\delta}} \end{bmatrix} \quad (5)$$

After estimating an econometric model by maximum likelihood estimation procedure, testing of the specific model (structural) formation, the latent scores for each of the modelled latent variables for each unit (tehsil) were computed using the approach of Jöreskog (2000). Consider the equations (1) and (2) for measurement models and writing these equations in a system as given below

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}_{\mathbf{y}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{\mathbf{x}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix}$$

and using the following notation

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{\mathbf{y}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{\mathbf{x}} \end{pmatrix}, \ \boldsymbol{\xi}_{\mathbf{a}} = \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix}, \boldsymbol{\delta}_{\mathbf{a}} = \begin{pmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \end{pmatrix}, \mathbf{x}_{\mathbf{a}} = \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} \quad ,$$

The latent scores for the latent variables in the model can be computed with formula

$$\hat{\boldsymbol{\xi}}_{ai} = \mathbf{U}\mathbf{D}^{\frac{1}{2}}\mathbf{V}\mathbf{L}^{-\frac{1}{2}}\mathbf{V}'\mathbf{D}^{\frac{1}{2}}\mathbf{U}'\boldsymbol{\Lambda}'\boldsymbol{\Theta}^{-1}\mathbf{x}_{\mathbf{a}} \quad (6)$$

Where $\mathbf{UDU}'$ is singular value decomposition of $\boldsymbol{\Phi} \equiv \mathbf{E}[\boldsymbol{\xi}_{\mathbf{a}}\boldsymbol{\xi}'_{\mathbf{a}}]$ and $\mathbf{VLV}'$ is the value of singular value decomposition of the matrix $\mathbf{D}^{\frac{1}{2}}\mathbf{UTBUVD}^{\frac{1}{2}}$, while $\boldsymbol{\Theta}_{\mathbf{a}}$ is the error covariance/correlation matrix of the observed variables. The latent scores $\boldsymbol{\xi}_{\mathbf{ai}}$ were computed for each observation x_{ij}in the $[(p+q)\times N]$ sample matrix whose rows are observations on each of observed variables and N is number of tehsils. These latent scores ($\boldsymbol{\xi}_{\mathbf{ai}}$) were used as an input for cluster analysis for the purpose of grouping the tehsils into several groups with similar characteristics. At the first step, Ward (1963) hierarchical procedure has been used to define the number of clusters and the group centroids. At second step, K-means method has been applied by taking the centroids from the Ward (hierarchical) method as initial seed-points. These steps were repeated until any re-assignment of cases does not make the clusters more internally cohesive (homogeneous) and more clearly separated from each other.

3. Results and Discussion

For the purpose of developing structural equation models, we identified 9 indicators for the year 2007-08, which have been included in the study. The codes and description of the variables have been presented in Table 1.

Table1: Codes and description of the variables in socio-economic sector

Code	Description	Symbol
LITERACY	Percentage literacy	y_1
LIT_MALE	Male literacy rate	y_2
LIT_FEMALE	Female literacy rate	y_3
BANK	Average population per bank (in '000)	y_4
POP_DEN	Density of population per square km of area	x_1
URBAN_POP	Percentage of urban population	x_2
VEH_REG	Different type of vehicles registered	x_3
M_WORKER	Percentage of main workers to total population	x_4
COOP_SOC	Number of co-operative societies per lakh of population	x_5

The factor (latent variables) responsible for socio-economic development in the State have been extracted by using Principal component analysis. The suitability of the factor analysis has been tested by visual inspection of the correlation matrices (Table 2) , by the Bartlett test of sphericity and by Keiser-Meyeer-Olkin (KMO) measure of sampling adequacy. The Bartlett test produced the Chi-square values as 385.42 ($P < 0.001$) which suggested that the original variables are significantly inter-correlated. The KMO statistic has been obtained as 0.68 which suggests that there are sufficient numbers of indicators for each factor. The Cattell's scree plots (Fig 1) level out after third eigen value and the variance contribution diminish after third component for all the periods, hence factor analysis has been carried out retaining these three factors.

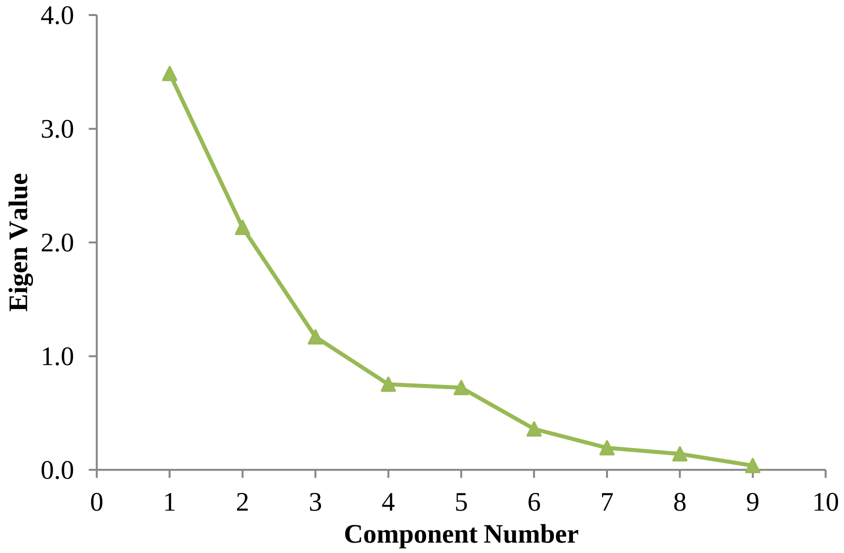

Fig 1: Scree plot for the number of factors to be retained

Table 2: Correlation matrix of indicators variables for socio-economic sector.

Variables	y_1	y_2	y_3	y_4	x_1	x_2	x_3	x_4	x_5
y_1	1.000								
y_2	0.841*	1.000							
y_3	0.950*	0.797**	1.000						
y_4	-0.422	-.349**	-.438**	1.000					
x_1	0.187	0.093	0.245*	-0.105	1.000				
x_2	0.366**	0.255*	0.449	-0.196	0.728**	1.000			
x_3	0.094	-0.059	0.104	0.000	0.799**	0.579**	1.000		
x_4	0.032	0.158	-0.017	-0.085	-0.140	-0.046	-0.149	1.000	
x_5	0.190	0.094	0.245*	-0.184	0.100	0.227	0.194	-0.021	1.000

** significant at the 0.01 level
* significant at the 0.05 level

The normal varimax rotated solution for 9 selected indicators of socio-economic sector have been given in Table 3. The Table 3 reveals that three components have eigen values above one contributing 73.19 per cent of the total variance.

The factor analysis suggested three factors have been found crucial for socio-economic development. The first factor has significantly high loading on the variables namely LITERACY, LIT_MALE, LIT_FEMALE and BANK. This factor could be taken as dimension of "*Literacy and Banking*". This factor explains 38.73 per cent of the total variance. The second factor has been identified as "*Population Distribution and Vehicles*" as it loads very heavily on the variables like POP_DEN, URBAN_POP and VEH_REG. This factor explains 23.69 per cent of the total variance. The third factor "*Work Force*" loads high on M_WORKER only. This factor explained 10.76 per cent of the total variance.

A perusal of communality values indicates that for 7 variables, the communalities exceed 70 per cent. Thus there is a fair degree of representation of all the 9 considered variables by the three factors i.e. "*Literacy and Banking*", "*Population distribution and vehicles*" and "*Work Force*" identified crucial for socio-economic development.

In order to normalize these variables using monotonic transformation technique , these variables have been transformed to Normal distribution by normal score technique (Jöreskog*et al.* 2000).

Table3: Normal varimax solution for variables of socio-economic Sector

Variables	Factor 1	Factor 2	Factor 3	h^2
LITERACY	**0.95**	0.13	-0.04	0.92
LIT_MALE	**0.90**	-0.01	0.10	0.81
LIT_FEMALE	**0.94**	0.19	-0.09	0.93
BANK	**-0.57**	-0.07	-0.15	0.35
POP_DEN	0.07	**0.93**	-0.05	0.87
URBAN_POP	0.31	**0.81**	0.02	0.77
VEH_REG	-0.08	**0.91**	-0.08	0.83
M_WORKER	0.07	-0.09	**0.98**	0.98
COOP_SOC	0.24	**0.25**	-0.05	0.13
Eigen Value	3.49	2.13	1.17	
Percentage variance	38.73	23.69	10.77	
Cumulative percentage	38.73	62.42	73.19	

The structural equation model of socio-economic sector of Haryana has been hypothesized on the basis of the three latent variables as suggested by the preliminary exploratory factor analysis and then further improved by freeing the elements of residual matrices and adding or deleting the indicator variables to the latent variables as suggested by the largest modification indices. The model parameters have been re-estimated after every improvement. Finally, the model which converged to the optimum solution with acceptable fit statistics has been obtained. The exogenous measurement models using (2) of socio - economic sector have been given in matrix equation (7) as given below:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(x)} & 0 \\ 0 & \lambda_{22}^{(x)} \\ \lambda_{31}^{(x)} & \lambda_{32}^{(x)} \\ \lambda_{41}^{(x)} & 0 \\ 0 & \lambda_{52}^{(x)} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{pmatrix} \qquad (7)$$

From the equations (7), it can be revealed that the exogenous latent variable ξ_1 has been measured by POP_DEN, VEH_REG with positive factor loadings and M_WORKER with negative non-significant loading. The second latent dimension $\xi_2$has positive significant loadings on indicator variables like URBAN_POP, VEH_REG and COOP_SOC.The endogenous measurement model using (2), has been formulated as

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} \lambda_{11}^{(y)} \\ \lambda_{21}^{(y)} \\ \lambda_{31}^{(y)} \\ \lambda_{41}^{(y)} \end{pmatrix} (\eta_1) + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} \qquad (8)$$

The endogenous measurement model presented in (10) depict that the latent variable η_1 has been measured by the indicator variables namely LITERACY,LIT_MALE, LIT_FEMALE and BANK.The structural equation model with latent variables using (3), has been given by following equation

$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \zeta_1 \qquad (9)$$

The restricted model has been estimated by setting the off-diagonal elements of Θ_δ and Θ_ε to zero. The restricted model has a Chi-square value of 30.90 (d.f = 24) with GFI = 0.90 and SRMR = 0.07 showing that it does not fit well to the data. The restricted models have been fine tunedby relaxing the zero restrictions of off-diagonal elements in the residual matrices. The residual matrix of endogenous variables remained as diagonal matrix as given below in (10).

$$\Theta_\varepsilon = \begin{pmatrix} \theta_{11}^{\varepsilon} & & & \\ 0 & \theta_{22}^{\varepsilon} & & \\ 0 & 0 & \theta_{33}^{\varepsilon} & \\ 0 & 0 & 0 & \theta_{44}^{\varepsilon} \end{pmatrix} \tag{10}$$

In order to obtain the improved model which fit well to the data, certain elements of Θ_δ and $\Theta_{\delta\epsilon}$ have been kept free as given below in (11) and (12).The improved model has Chi-square value as 3.91 (d.f.=15), GFI as 0.99 and SRMR as 0.03 showing that model implied correlation matrix is equal to population correlation matrix.

$$\Theta_\delta = \begin{pmatrix} \theta_{11}^{\delta} & & & & \\ \theta_{21}^{\delta} & \theta_{22}^{\delta} & & & \\ 0 & 0 & \theta_{33}^{\delta} & & \\ 0 & 0 & 0 & \theta_{44}^{\delta} & \\ 0 & 0 & 0 & 0 & \theta_{55}^{\delta} \end{pmatrix} \tag{11}$$

$$\text{and } \Theta_{\delta\varepsilon} = \begin{pmatrix} 0 & \theta_{12}^{\delta\varepsilon} & \theta_{13}^{\delta\varepsilon} & 0 \\ 0 & 0 & \theta_{23}^{\delta\varepsilon} & 0 \\ 0 & \theta_{32}^{\delta\varepsilon} & 0 & 0 \\ 0 & \theta_{42}^{\delta\varepsilon} & \theta_{43}^{\delta\varepsilon} & 0 \\ \theta_{51}^{\delta\varepsilon} & \theta_{52}^{\delta\varepsilon} & 0 & 0 \end{pmatrix} \tag{12}$$

The path diagram of final models which fit well to data for socio-economic sector with estimated coefficients has been presented in Fig 2.

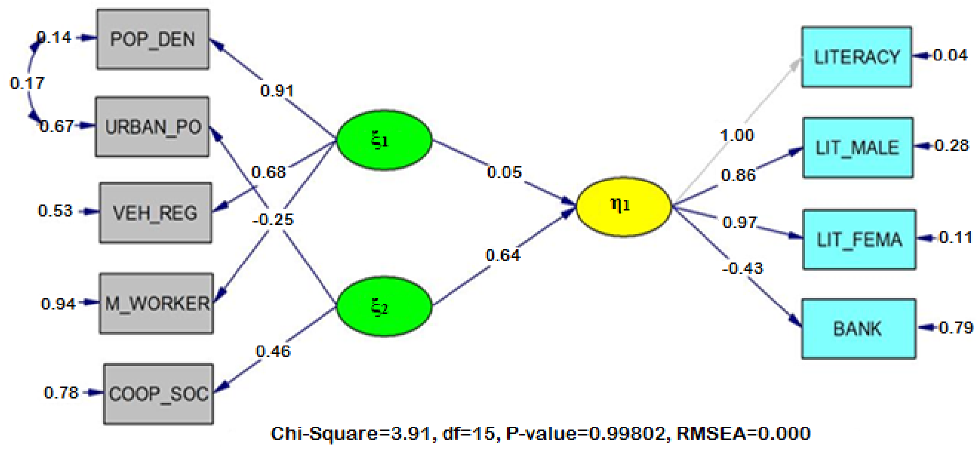

Fig 2: Path diagram for model of socio-economic sector with coefficient estimates

Table4: Maximum likelihood estimates of structural equation model for socio-economic sector

Parameter	Estimate (S.E)	Standardized Estimates	Parameter	Estimate (S.E)	Standardized Estimates
$\lambda_{11}^{(x)}$	0.91 (0.16)	0.92	θ_{44}^{δ}	0.94 (0.17)	0.94
$\lambda_{22}^{(x)}$	0.57 (0.16)	0.57	θ_{55}^{δ}	0.78 (0.18)	0.79
$\lambda_{31}^{(x)}$	0.68 (0.15)	0.68	θ_{21}^{δ}	0.17 (0.10)	0.18
$\lambda_{41}^{(x)}$	-0.25 (0.13)	-0.25	$\theta_{11}^{\varepsilon}$	0.04 (0.03)	0.04
$\lambda_{52}^{(x)}$	0.46 (0.17)	0.46	$\theta_{22}^{\varepsilon}$	0.28 (0.05)	0.29
$\lambda_{11}^{(y)}$	1.00	0.98	$\theta_{33}^{\varepsilon}$	0.11 (0.03)	0.11
$\lambda_{21}^{(y)}$	0.86 (0.07)	0.84	$\theta_{44}^{\varepsilon}$	0.79 (0.14)	0.79
$\lambda_{31}^{(y)}$	0.97 (0.05)	0.94	$\theta_{12}^{\delta\varepsilon}$	-0.12 (0.06)	-0.12
$\lambda_{41}^{(y)}$	-0.46 (0.11)	-0.45	$\theta_{13}^{\delta\varepsilon}$	0.06 (0.04)	0.06
ϕ_{21}	0.46 (0.19)	0.46	$\theta_{23}^{\delta\varepsilon}$	0.14 (0.05)	0.14
γ_{11}	0.05 (0.21)	0.05	$\theta_{32}^{\delta\varepsilon}$	-0.15 (0.07)	-0.15
γ_{12}	0.64 (0.24)	0.66	$\theta_{42}^{\delta\varepsilon}$	0.14 (0.07)	0.14
var(ζ_1)	0.51 (0.21)	0.53	$\theta_{43}^{\delta\varepsilon}$	-0.07 (0.04)	-0.07
θ_{11}^{δ}	0.14 (0.24)	0.15	$\theta_{51}^{\delta\varepsilon}$	-0.08 (0.05)	-0.08
θ_{22}^{δ}	0.67 (0.17)	0.68	$\theta_{52}^{\delta\varepsilon}$	-0.20 (0.07)	-0.20
θ_{33}^{δ}	0.53 (0.16)	0.53			
$\chi^2_{(df=15)}$			3.91		
GFI			0.99		
SRMR			0.03		

The un-standardized and standardized maximum likelihood estimates of the free parameters of models have been presented in Table 4. The estimates in Table 4 indicated that the exogenous variable ξ_1 has a positive non-significant influence on the endogenous latent variable η_1 whereas exogenous latent variables ξ_2 indicates a positive significant influence on η_1. Also the exogenous latent variables ξ_1 and ξ_2 have significant correlation. The estimates in the Table 4 also revealed that the off-diagonal elements θ_{21}^{δ} of the residual matrix Θ_{δ} have significant correlations. Certain error terms as mentioned in (12) of exogenous and endogenous variables have also been found significantly correlated .

Using (6), the latent scores ξ_{ai} were computed for each observation x_{ij} in the $[(p+q)\times N]$ sample matrix whose rows are observations on each of observed variables and N is number of tehsils.These latent scores (ξ_{ai}) were used for classifying the tehsils with cluster analysis. In order to search for the number of groups which shows similar socio-economic characteristics, Ward's hierarchical clustering method has been used on the latent scores of three socio-economic dimensions. The Ward's method based on squared Euclidean distances has been used to form the dendogram (Fig 3). From the analysis the dendogram, it has been concluded that 3-clusters solutions are appropriate for grouping of tehsils on the basis of socio-economic characteristics. The average scores on the latent variables η_1, ξ_1 and ξ_2 for the resulting three clusters are presented in Table 5.

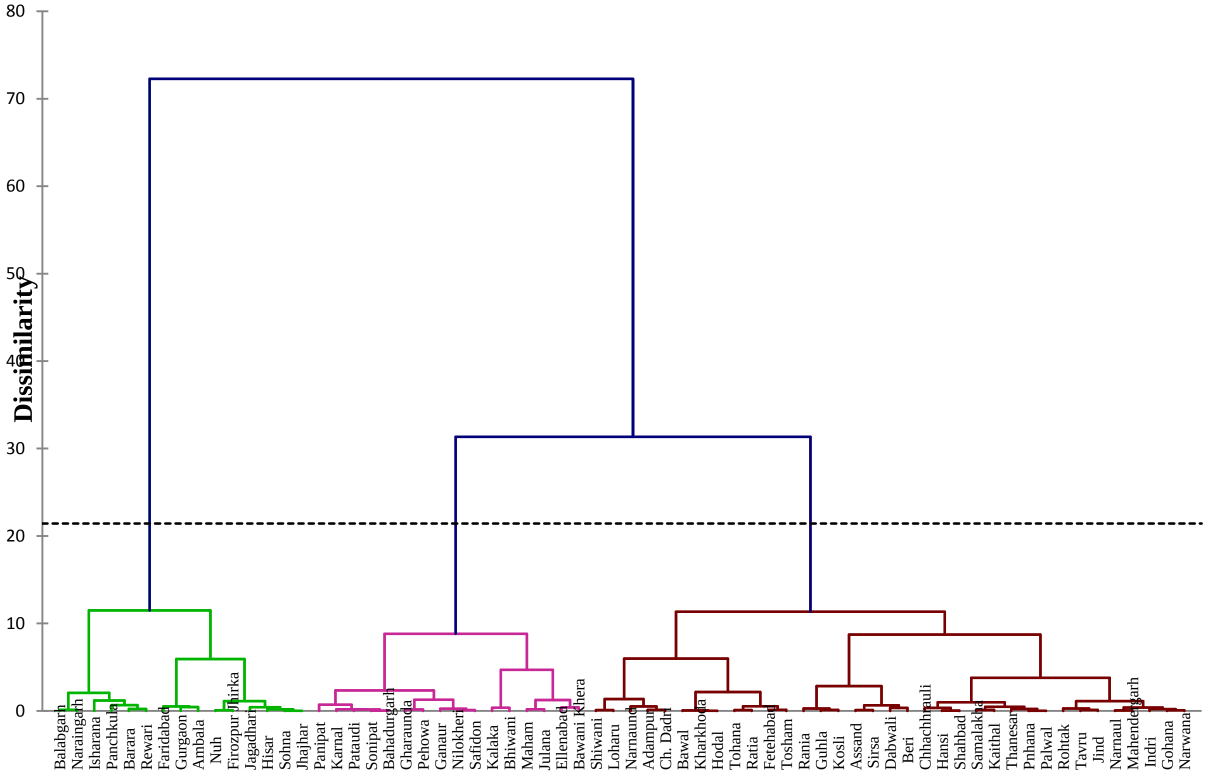

Fig 3: Dendogram for socio-economic sector

Table5: Cluster centroids

Cluster	η_1	ξ_1	ξ_2
1	1.209	-0.179	-0.742
2	0.682	0.223	-1.127

3	1.340	-0.129	-0.975

The non-hierarchical clustering procedure (K-means), using the cluster centres presented in Table 5 of hierarchical solution as the initial seed points has then been used to improve the results of best cluster solution derived by the Ward's method. The improved cluster solutions by K-mean procedure for the socio-economic sector have been presented in Table 6. The perusal of the Table 6 indicated that cluster 1 comprising of 18 tehsils as mentioned in Table 8 has been characterised by high scores on all the latent variables. The Cluster 2 has moderately higher scores on all the latent variables compared to cluster 3 and consisting of 28 tehsils as mentioned in Table 8. The remaining 20 tehsils form cluster 3 which has lower scores on all the latent dimensions. The largest distance of 3.212 between the cluster 1 and 3 has also supported this fact that these clusters differ significantly with respect to socio-economic characteristics. Further, the distance between cluster 1 and 2 as well as cluster 2 and 3 indicates that these clusters are also dissimilar in socio-economic characteristics.

Table6 : Final cluster centres of socio-economic sector

Latent Variable	**Cluster**		
	1	**2**	**3**
	N=18	**N=28**	**N=20**
η_1	1.075	-0.290	-0.561
ξ_1	0.705	0.229	-0.954
ξ_2	1.209	-0.062	-1.001

Table7: Distance between final cluster centres of socio-economic sector.

Cluster	**Cluster**		
	1	**2**	**3**
1	-	1.925	3.212
2	1.925	-	1.534
3	3.212	1.534	-

Table8 : Classification of tehsils for socio-economic development in Haryana State.

Cluster	Tehsils
1	Hisar, Panchkula, Asandh, Isharana, Gurgaon, Sohna, Nuh, FirozpurJhirka, Faridabad, Balabgarh, Jagadhari, Kaithal, Sirsa, Ambala, Barara, Narayangarh, Jhajhar, Rewari
2	Adampur, Hansi, Narnaund, Gohana, Fatehabad, Ratia, Narnaul, Mahendergarh, Indri, Smalakha, Taoru, Punhana, Palwal, Rohtak, Chhachhrauli, Jind, Narwana, Guhla, Thanesar, Shahbad, Dabwali, Rania, Tosham, Siwani, Loharu, CharkhiDadri, Beri, Kosli
3	Sonipat, Kharkhoda, Ganaur, Tohana, Kalaka, Karnal, Gharaunda, Nilokheri, Panipat, Pataudi, Hodal, Meham, Julana, Safidon, Pehowa, Ellenabad, BawaniKhera, Bhiwani, Bahadurgarh, Bawal

References

Joreskog, K.G. (2000). Latent variable scores and their uses. Scientific Software International, http://www.ssicentral.com/lisrel.

Joreskog, K.G., D. Sorbom, S.duToit and M. du Toit (2000).*LISREL 8:* New Statistical Features. Chicago: *Scientific Software International*.

Lipshitz, G. and A. Raveh (1994). Application the Co-Plot Method in the Study of Socio-economic differences among Cities: A Basis for a Differential Development Policy. *Urban Studies*, **31:** 123-135.

Lipshitz, G. and A. Raveh (1998). Socio-economic differences among localities: A new method of multivariate analysis. *Regional Studies*, **32:** 747–757.

Narain, P., S.C. Rai, and P. Sarup (1991).Statistical evaluation of development on socio-economic front.*Jour. Ind. Soc. Agric. Stat.***43(3):** 329-345.

Narain, P., S.D. Sharma, S.C. Rai and V.K. Bhatia (2001).Regional dimensions of disparities in crop productivity in Uttar Pradesh.*Jour. Ind. Soc. Agric. Stat.***54:** 62-79.

Narain, P., S.D. Sharma, S.C. Rai and V.K. Bhatia (2007).Statistical evaluation of social development at district level.*Jour. Ind. Soc. Agric. Stat.***61(2):** 216-226.

Sheoran, O.P; Rai, Lajpat and Saxena, K.K. (2012). Structural Equation Modelling with Latent variables for assessment of Agricultural Development of Haryana State. *Int. J. Agricult. Stat. Sci.*, **8(2):**415-431.

Soares, J.O.; M.M.L Marqus and C.M.F. Monteiro.(2003). A multivariate methodology to uncover regional disparities: A contribution to improve European Union and governmental decisions. *European Journal of Operational Research*, **145**: 121–135.

Upadhyay, B. 2001.Statistical Assessment of Development in Rajasthan – A Spatial and Temporal Analysis. Unpublished Ph.D. Thesis, M.P.U.A.T, Udaipur.

Ward, J. 1963. Hierarchical grouping to optimize and objective function. *Journal of the American Statistical Association*, **58:** 236-244.