Taxonomical analysis of regional development by outranking relations on multiple principal components

Mishra, SK

North-Eastern Hill University, Shillong (India)

10 February 1984

Taxonomical analysis of regional development by outranking relations on multiple principal components

SK Mishra
Department of Architecture
and Regional Planning
Indian Institute of Technology
Kharagpur

Introduction: For taxonomical analysis (TA) of levels of development of economies, use of the index of per capita income was almost unchallenged till the recent past. It was so due to several reasons. Historically, since development economics was an offshoot of the classical economics which flourished in developed economies of the West and adored competitive general equilibrium, it was not unexpected that the level of development of an economy was believed to be represented by per capita income. For institutional reasons, since development economics and planning had initially been a concern of central agencies and sovereign states and much was not taken care of the economies at regional levels, the question of representativeness of per capita income as a measure of level of development drew little attention. For operational reasons, social accounting at national and state levels was both practicable and practiced, and hence, there was no urge for any measure of development level other than per capita income.

With the recent emergence of interest in and need for regional planning, especially in under-developed economies, economists experienced uneasiness with regard to all the three norms mentioned above, and hence, there is observed a marked interest of the regional planner in searching for some new index or measure of development of regional economies. The quest is justified, but yet not achieved. The failure led the regional planner to seek help of some statistical-mathematical techniques by which a number of variables, each one representing level of development of some particular aspect of the economy, could somehow be aggregated to make a composite index, which, in turn, could be used for TA of overall level of development of regional economies. The principal component analysis was found quite promising in making such a composite index (3, 4) and hence is its popularity among regional planners.

No doubt, attempts to construct composite indices and using them for TA of regional level of development successfully solved the last two problems that one was faced with in using per capita income. However, the first problem remains intact. If regional economies are not well-evolved and integrated and do not characterize competitive general equilibrium such that price mechanism in product and factor markets generate income stream proportional to the level of development, there cannot be any possibility, by the same token, of satisfactorily representing several aspects of development by a single composite index. It has been frequently reported that correlations among different indicators of development across sectors (e.g. primary and secondary sectors) are weak and sometimes non-conformal (1, 5, 6, 11, 15). As a result, the recourse taken by the regional planner is to avoid construction of composite index of overall development using full information (i.e. full matrix of intersectoral and intrasectoral correlation coefficients) and instead, attempt is made to construct several sectoral indices (each one a first principal component derived from the submatrix of intrasectoral correlation coefficients). These sectoral indices are, at the second stage, subjected to principal component analysis once more and the index of overall development is derived, which, in turn is the first principal component of the sectoral indices. It is obvious that such an approach will lead to construction of highly inefficient index as it is based on partial information. It has been suggested by the author elsewhere (15) that while making the index of the level of overall development, the use of ordinary correlation coefficients (among sectoral composite indices) should be replaced by the use of canonical correlation coefficients (among several sets of variables) so that the index of the level of overall development is more efficient. However, the use of canonical correlation alleviates inefficiency but does not remove it.

There is yet another aspect of this problem of inefficiency. It has been reported by many researchers that they dropped a number of variables from further analysis after they found them to be non-conformal with the retained set of variables (1, 5, 6, 11). It is obvious that the purpose in doing so is to construct a composite index that is more representative. But it is illusive. The representativeness of an index derived from partial information is always less than that of another index which uses full information.

The problem of inefficiency would not have been acute, could the first principal component of each sector explain a very large portion of the total variance of the original set of variables, as we urged above. But we have also seen that in underdeveloped economies, due to reasons already discussed, any single index cannot represent numerous un-integrated forces of development. Hence TA of regional levels of development cannot be efficient unless we use many composite indices – i.e. multiple principal components together. It is remarkable, however, that most of the studies on taxonomical analysis of regional level of development have avoided using multiple principal components, relying on the first principal component, howsoever inefficient it might be.

No doubt, it has been partially due to the illusion about representativeness (based on several unwarranted manipulations) and partly due to lack of awareness regarding the possibilities of using multiple principal components for taxonomical analysis. The lack of awareness has been due to the inertia of technological culture. We have been tuned to cardinal mathematics and single criterion TA since ages, and hence, our readiness to accept multi-criteria analysis and non-cardinal mathematics must be low. This led to our restricted vision. A user of cardinal mathematics knows that two principal components (derived from the same set of original variables) cannot be subjected to ordinary mathematical operations and they cannot be used for making any composite index. Hence, better to be satisfied with the first principal component, howsoever inefficient it might be.

Recent developments in multi-criteria analysis has, however, opened before us the possibilities of using multiple mutually non-commensurable criteria for deciding dominance relations of paired comparisons (2, 7, 9, 12, 13, 14). In this context, techniques like concordance analysis (9) ELECTRE I, II and III (12, 13) have proven their mettle. These techniques can be applied for multi-criteria TA, in which multiple principal components might be used as criteria for deriving dominance relations among sub-regions of a region.

The methodology of TA by outranking relations on multiple principal components: We propose here a method of TA of regional levels of development by outranking relations on multiple principal components. We suppose that multiple principal components are, among themselves, non-commensurable, and in themselves, semi-cardinally measurable.

Let there be N sub-regions belonging to a region and let there be M variables each of which represents some particular aspect of the regional development. Thus, x_{ij} is the value of the variable j recorded in sub-region i. Given X(N,M), we could go in for multi-criteria analysis to establish dominance relation of sub-region i over sub-region k (i, k ϵ N). But since we do not know the relative weights and the one variable is not independent of the other ($r_{ij} \neq 0$; $j \neq i$; i,j ϵ M) and relative contribution of each principal component in explaining the total variance in X is obtained. We use these relative contributions as weights w(M). Now, how many principal components we should use is based on our judgment. We should, however, see to it that P_0 represents X as exhaustibly as possible ($P_0 \subseteq P$; $P_0(N, M_0)$).

With $P_0(N,M_0)$ and $w(M_0)$ we carry out multi-criteria outranking analysis. Using outranking relations we construct two matrices, C(N,N) and D(N,N), named as concordance and discordance matrices respectively. With C and D we construct E(N,N) by any suitable method detailed out in the appendix of this paper and use E for establishing dominance relations of i over j (i, j ϵ N). These dominance relations readily give the taxonomy of the levels of development of the sub-regions.

An illustrative application: We illustrate the use of the proposed method in classifying the districts of Bihar according to their levels of overall development. We have seventeen districts and five indicators of sectoral levels of development. (We treat these indicators as variables for purely illustrative purpose). Thus, X is for 17 regions and 5 variables, or X(17,5). We derive five principal components P(17,5), to exhaustively represent X(17,5). Explanatory contributions of P's are: P_1 (0.65527), P_2 (0.26696), P_3 (0.04174), P_4 (0.02685), P_5 (0.00918). We could use the first three P's also for an effective classification, but we preferred to use all the five.

Next, we construct C and D, and compare them with varying T_c and T_d to obtain logically stable classes. Our finding is that by using multiple principal components it is not possible to establish strong dominance relation which was possible to do by using the first principal component (single criterion) only. On the criterion of the first principal component our ordering would have been as follows: $16 _p 1 _p 17 _p 9 _p 8 _p 3 _p 7 _p 4 _p 6 _p 2 _p 5 _p 15 _p 11 _p 14 _p 10 _p 12 _p 13$, where $_p$ means 'dominates over' and numerals are the codes for the districts. But on all the five criteria our dominance relations are as follows (where $_q$ means 'does not dominate over'):
$(16 _q 1) _p (17 _q 9 _q 8 _q 3) _p (7 _q 4) _p (6 _q 2 _q 5) _p (15 _q 11) _p (14 _q 10) _p (12 _q 13)$.

Conclusion: It is remarkable that multi-criteria analysis simply denies the undue dominance relations established by a single-criterion analysis. Relevance of this denial is immense. To recall back, we urged that per capita income as a criterion of taxonomy may establish an unwarranted dominance of sub-region i over sub-region j ($i \neq j$), and hence, we searched for some other criterion. We found that the criterion of the 1st principal component, especially when its explanatory contribution is low (which is quite expected in less-developed economies), also established unwarranted dominance. It is true that the 1st principal component provides us with a numerical measure (and we have a bias for numerical measures howsoever illegitimate they might be), while outranking method proposed here provides us with logical relations of dominance and non-dominance and none of the numerical measures, yet the former is illusive and the latter is logically sound. The former is based on partial information and therefore apt to be inefficient, while the latter is based on full information and hence more efficient.

Table-1. Principal Components of X						
Sl No.	Districts	P_1	P_2	P_3	P_4	P_5
1	Patna	2.168	0.983	2.313	1.417	0.405
2	Gaya	-0.436	0.517	1.071	1.088	0.570
3	Sahabad	0.176	1.046	1.952	1.543	0.664
4	Saran	-0.060	0.621	1.814	1.344	0.569
5	Champaran	-0.678	0.670	1.784	1.490	0.724
6	Muzaffarpur	-0.134	-0.045	0.079	0.133	0.025
7	Darbhanga	-0.001	0.108	0.680	0.501	0.135
8	Monghyr	0.571	0.402	1.315	0.901	0.171
9	Bhagalpur	0.707	0.393	1.182	0.536	0.196
10	Saharsa	-1.167	-0.111	0.432	0.143	0.437
11	Purnea	-0.819	-0.081	-0.877	0.344	0.214
12	S. Pargana	-1.187	-0.802	-1.871	-1.044	-0.338
13	Palamau	2.636	-0.409	-2.090	-1.111	0.308
14	Hazaribagh	-1.129	-0.419	-1.832	-0.956	-0.203
15	Ranchi	-0.713	-0.938	-2.908	-2.102	-0.843
17	Dhanbad	4.134	-0.664	-1.049	-1.659	-1.446
17	Sighbhum	1.203	-1.269	-2.787	-2.568	-1.589

P as weighted aggregation of X

$P_1=0.971x_2+0.933x_4+0.863x_3+0.846x_4-0.055x_1$

$P_2=0.976x_1+0.428x_3+0.052x_4+0.012x_2-0.444x_5$

$P_3=0.281x_4+0.154x_1+0.141x_5-0.176x_2-0.235x_3$

$P_4=0.257x_5+0.141x_1+0.020x_2-0.029x_3-0.218x_4$

$P_5=0.161x_2+0.033x_1-0.002x_4-0.048x_5-0.130x_3$

For identification of x ref (1), Chattopadhyay and Mishra.

Table-2. Concordance and Discordance Matrices (CD)																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	-	B	C	B	B	O	O	O	O	A	O	O	O	O	O	N	O
2	M	-	N	N	I	I	N	N	N	O	O	O	O	O	O	N	M
3	J	O	-	O	B	O	O	J	B	O	O	O	O	O	O	N	H
4	M	O	N	-	N	O	E	N	N	O	O	O	O	O	O	N	K
5	M	J	M	J	-	L	N	N	N	O	O	O	O	O	O	N	N
6	N	G	N	N	G	-	N	N	N	E	E	O	O	O	O	N	N
7	N	G	N	G	G	O	-	N	N	C	B	O	B	O	O	N	M
8	N	F	G	G	G	O	O	-	N	B	B	O	B	O	O	N	G
9	N	F	G	G	G	O	O	G	-	B	A	O	B	O	O	N	G
10	M	N	N	N	N	M	M	M	M	-	N	O	O	E	I	N	N
11	N	N	N	N	N	M	M	M	M	C	-	O	B	O	E	N	N
12	N	N	N	N	N	N	N	N	N	N	N	-	F	N	N	N	N
13	N	N	N	N	N	N	M	M	M	N	M	K	-	N	N	N	N
14	N	N	N	N	N	N	N	N	N	G	N	O	F	-	L	N	N
15	N	N	N	N	N	N	N	N	N	G	G	G	G	G	-	N	N
16	G	G	G	G	G	G	G	G	G	G	G	C	F	F	C	-	O
17	N	G	G	G	G	G	G	G	G	G	G	G	G	G	F	N	-

The matrix CD contains Concordance and Discordance matrices. If i<j, $CD_{ij}=D_{ij}$ and if i>j $CD_{ij}=C_{ij}$. For compactness we have used symbols for numerical values. Hence, in the cells, the letters represent numerical in the following scheme (exclusive of the lower class limits). O=0.0; N=1.0; A=(0.000-0.0005); B=(0.0005 – 0.005); C=(0.005-0.05); D=(0.05 – 0.1); E=(0.1-0.2); F=(0.2-0.3); G=0.3-0.4); H=)0.4-0.5); I=(0.5-0.6); J=(0.6-0.7), K=(0.7-0.8) L=(0.8-0.9); M=(0.9-0.99999)

Appendix

Given $x_{ij}; i=1,N; \quad j=1,M\,(N\ sub-regions; M\ \mathrm{var}iables)$

$x_{ij} \Leftarrow (x_{ij}-\bar{x})/\sigma_j; \ \ i=1,N; \quad j=1,M$ where $\bar{x}=(1/N)\sum_{i=1}^{N}x_{ij}; \ \ \sigma_j^2=(1/N)\sum_{i=1}^{N}x_{ij}^2-\bar{x}_j^2$

$R_{jk}=(1/N)\sum_{i=1}^{N}x_{ij}x_{ik}; \quad j,k=1,M$

Set $l=0$

1. $p_j=1; \quad j=1,M$

   $l \Leftarrow l+1$

2. $q_k=\sum_{j=1}^{M}R_{jk}p_j; \ \ k=1,M$

   $\lambda_l=\max(q_k); k=1,M$

   $q_k \Leftarrow q_k/\lambda_l; \ \ k=1,M$

   If($|q_k-p_k|$) $\le 0.000001\ \ Go\ \ to\ \ 3$

   Else

   $p_k \Leftarrow q_k; \ \ k=1,M$

   Go to 2

3. $T_{kj}=q_kq_j; \quad j,k=1,M$

   $s=\sum_{k=1}^{M}q_k^2$

   $g=\sqrt{s/\lambda_l}$

   $w_k=q_k/g \ \ k=1,M$

   $V_{li}=\sum_{j=1}^{M}x_{ij}w_j; \ \ i=1,N$

   $R_{jk} \Leftarrow R_{jk}-T_{jk}(\lambda_l/s); \ \ k=1,M$

   If($l<M$) Go to 1

   Else

   $\lambda_l \Leftarrow \lambda_l/M; \ \ l=1,M$

Comment: By now we have computed M principal components (V_{li}) and weights (λ_l) and now proceed for concordance analysis

$$V_{ji} \Leftarrow V_{ji}\Big/\sqrt{\sum_{i=1}^{N}V_{ji}^2}; \quad j=1,M; i=1,N$$

$\underset{i\neq j}{C_{ij}}=\sum_{l=1}^{M}\lambda_l\beta_l; \ \ i,j=1,M$ where $\beta_l=0$ if($V_{li}-V_{lj}$) < 0 else $\beta_l=1; l=1,M$

$$U_{ji}=V_{ji}\lambda_j; i,j=1,M$$

$$\left.\begin{array}{l} D_{ij}=\max(|U_{li}-U_{lj}|\,\beta_l); i\neq j \\ where\ \ \beta_l=0\ \ if\,(V_{li}-V_{lj})\ge 0\ \ else\ \ \beta_l=1 \end{array}\right\} \begin{array}{l} l=1,M \\ i,j=1,N \end{array}$$

$$\left.\begin{array}{l} \Delta_{ij}=\max\limits_{l}(|U_{lj}-U_{li}|); i\neq j \\ D_{ij} \Leftarrow D_{ij}/\Delta_{ij}; \ \ i,j=1,N \end{array}\right\}; \ \ l=1,M; \ \ i,j=1,N$$

Comment: Computation of Concordance and Discordant matrixes is over.
Given arbitrary values of T_c =(0, 1) and T_d=(0, 1)
$E_{ij} = \beta;\ i, j = 1, N$ where $\beta = 0\ if(C_{ij} > T_c\ and\ D_{ij} > T_d)$ else $\beta = 1$
Comment: Decision may be taken now.
Sub-region i is preferred to sub-region j if $E_{ij} = 1$
End

Note: (a) For computation of E_{ij} based on min operator and gamma operator, vide Singh, D. (14, 1983).
(b) The symbol $\Leftarrow$ means 'is replaced by'.

References:

1. Chattopadhyay, R.N. and S.K. Mishra (1979) "Towards Balanced Development of Bihar: A Case Study in Inter-regional Allocation of Resources", *Geographical Review of India*, Vol. 41:3.
2. Delft, A. and P. Nijkamp (1976) "A Multi-Objective Decision Model for Regional Development, Environmental Quality Control and Industrial Land Use", *Papers of the Regional Science Association*, Vol. 36.
3. Hagood, M.J. et al. (1941) "An Examination of the Use of Factor Analysis in the Problem of Sub-Regional Delineation", *Rural Sociology*, Vol. 6.
4. Hotalling, H. (1933) "Analysis of a Complex of Statistical Variables into Principal Components", *Journal of Educational Psychology*, Vol. 26.
5. Mishra, S.K. and A. Chopra (1978) "Dimensions of Inter-District Disparities in Levels of Development in MP – 1971", *The Indian Geographical Journal*, Vol. 53:1.
6. Mishra, S.K. and S.B. Gaikwad (1979) "Impacts of Economic Development on Welfare and Living Conditions of People of MP", *Indian Journal of Regional Science*, Vol. 11:1.
7. Mishra, S.K. and A.P. Subudhi (1982) "Concordance Analysis for Multi-Criteria Evaluation of Competing Projects; A Computer-Oriented Approach", *National Geographical Journal of India*, Vol. 28: 3-4.
8. Mishra, S.K. (1984) *Evaluation of Public Policies for Agricultural Development in Less Developed Regions*, Deptt. Of Architecture and Regional Planning, IIT, Kharagpur.
9. Nijkamp, P. (1975) "A Multi-Criteria Analysis of Project Evaluation", *Papers of the Regional Science Association*, Vol. 35.
10. Pal, M.N. (1974) "A Method of Regional Analysis of Economic Development with Special Reference to South India", *Indian Journal of Regional Science*, Vol. 5.
11. Pal, M.N. (1975) "Regional Disparities in Levels of Development in India", *Indian Journal of Regional Science*, Vol. 7.
12. Roy, B. (1973) "How Outranking Relation Helps Multiple Criteria Decision Making", in J.L. Cochrane and M. Zeleny (eds.) *Multiple Criteria Decision-Making*. Univ. of South Carolina Press, Columbia.
13. Roy, B. and Ph. Vinke (1981) "Multi-Criteria Analysis: Survey and New Directions", *European Journal of Operations Research*, Vol. 8.
14. Singh, D. (1983) *Multiple Criteria Decision Rating: Preference Modelling using Fuzzy Sets* (unpubs) Ph.D. Thesis, Deptt. Of Mathematics, I.I.T., Kharagpur.
15. Subudhi, A.P. and S.K. Mishra (1982) "Dimensions of Regional Disparities in Levels of Development: A Case Study of Uttar Pradesh", in L.R. Singh (ed.) *Regional Planning and Rural Development*, Thinkers' Library.

Published in *Hill Geographer*, a Journal of the Geographical Society of the North-Eastern Hill Region (India)
Shillong; 1984: vol III(1): pp. 20-28.