Ranking Police Administration Units on the Basis of Crime Prevention Measures using Data Envelopment Analysis and Clustering

Manish Gupta[1], B. Chandra[2]* and M. P. Gupta[3]

ABSTRACT

In this paper, a novel approach has been proposed to rank police administration units on the basis of their effective enforcement of crime prevention measures using Data Envelopment Analysis (DEA) and Clustering. The proposed approach will offer an effective mechanism not only to rank police administration units but also provide an evaluation tool to monitor the implementation of crime prevention measures at various levels of police administration. The paper discusses two major phases of the proposed approach. In the first phase, clustering is used to identify the crime zones and to form homogeneous groups in crime data. In the second phase, police administration units in a particular crime zone are ranked using DEA. The effectiveness of the proposed approach has been illustrated on Indian crime data. The comparative results of DEA with clustering and DEA without clustering are also given to highlight importance of linking DEA with clustering.

Keywords: *Data Envelopment Analysis, Clustering, Performance Measurement, Police Administration.*

1. Introduction

In the present scenario, crime deterrence has become an upheaval task with an enormous increase in the crime. Several crime prevention measures such as e-governance initiatives, CCTV vigilance, police patrolling, special task force (STF) etc are undertaken by various police administration units i.e. state, district and police station. There is a need to monitor these crime prevention measures on regular basis so that shortcomings of these measures and implementation related issues could be highlighted. In this paper, Data Envelopment Analysis (DEA) (Anderson & Petersen, 1993; Banker et al., 1984; Charnes et al., 1978; Kao & Hung, 2005; Sexton & Silkman, 1986) has been applied in combination with clustering technique to rank police administration units on the basis of their effective enforcement of crime prevention measures. The approach will offer a computer-based environment not only to rank police administration units but also provide an evaluation tool to monitor the implementation of crime prevention measures at various levels of police administration.

DEA has been proposed by Charnes, Cooper & Rhodes (1978) as a technique for measuring efficiencies of various decision-making units (DMUs). DEA is a mathematical model that measures the relative efficiency of DMUs with multiple inputs and outputs without production function. The results of DEA determine

[1] Institute for Systems Studies and Analyses, Metcalfe House, Delhi, India
[2] Department of Industrial Engineering and Management, Indian Institute of Technology, Kanpur, India
[3] Department of Management Studies, Indian Institute of Technology, Delhi, India
* *Corresponding Author:* (E-mail : bchandra104@yahoo.co.in, Telephone: +91-11-26591493)

Pareto Frontier which is attained & marked by specific DMUs on the boundary envelope of input/output variable space if DMU lie on the envelope than efficient otherwise inefficient. Instead of using fixed weights for all DMUs under evaluation, DEA computes a separate set of weights for each DMU. Each DMU will choose weights so as to maximize self-efficiency with the constraints that the efficiency of no DMU exceeds 1 when using the same weights. DEA also provide reasons for inefficiency by comparing inefficient units with the peer group of efficient units. It also sets target levels for the inefficient units to become efficient one.

In the past, DEA has been widely applied for relative performance measurement in socio-economic sector (Banker et al., 2007; Berger & Humphrey, 1997) and public services (Charnes et al., 1994; Cooper et al., 2007). Banker, Chang, Natarajan (2007) has used DEA technical inefficiency measure to estimate allocative inefficiency in the public accounting industry using revenue and personnel data for the top U.S. public accounting firms over 1995–1998. Berger & Humphrey (1997) has reviewed 130 studies on DEA applied to financial institutions in 21 countries. Charnes, Cooper & Rhodes (1978) provides a collections of various papers on applications of DEA in public services such as health, educations etc. Bergendahl & Lindblom (2008) has applied DEA to evaluate the performance of Swedish savings banks based on service efficiency. DEA has also been used for measuring efficiencies of police forces (Butler & Johnson, 1997; Carrington et al., 1997; Drake & Simper, 2000, 2003; Thanassoulis, 1995; Verma & Gavirneni, 2006). But the results of stand-alone usage of DEA may not be satisfactory since crime locations on which DEA is applied are heterogeneous in context of crime density, area and size of the population. The major drawbacks of using DEA in standalone mode i.e. DEA without clustering for measuring efficiencies of police administration units as DMUs is the creation of insignificant peer groups for inefficient unit. A DMU should not have a peer group consisting of DMUs from different crime zones since two DMUs from different crime zones cannot be benchmark for each other in DEA models. To overcome these limitations, clustering is used to identify the homogeneous groups of similar crime density, area and population.

Clustering (Jain et al., 1999) as unsupervised technique is the process of organizing objects into groups such that similarity within the same cluster is maximized and similarities among different clusters are minimized. Clustering techniques have been applied in various areas such as information retrieval, pattern recognition etc. In this paper, clustering technique is applied to Indian crime data to cluster the crime locations and to find crime zones in India, as the density of crime incidents will be continuous over an area, being higher in some parts and lower in others. The paper describes two major phases of the proposed approach. The first phase uses clustering to identify the crime zones and to form homogeneous groups in the area of interest. In the second phase, police administration units as DMUs in a particular crime zone are ranked using DEA based on their efficiency score. The proposed approach has been effectively applied on Indian crime data. Comparative results of constant return to scale (CRS) model (Charnes et al., 1978) and variable return to scale (VRS) model (Banker et al., 1984) for DEA with clustering and DEA without clustering are also shown in the paper.

Section 2 highlights some of the earlier work done in measuring efficiencies of police forces using DEA as well as the limitations of those approaches. The two-phase methodology has been described along with brief details of clustering and DEA technique in section 3. Section 4 shows the results of clustering and DEA on Indian crime data for year 2006 under various crime heads such as murder, rape kidnapping etc. The Comparative evaluation of both the approaches i.e. DEA with clustering and DEA without clustering for CRS model and VRS model are shown in section 5. Concluding remarks is given in the last section of the paper.

2. Related Work

Thanassoulis (1995) has applied DEA to measure relative police force efficiency of English and Welsh police force. It focuses solely on the clear up rates for violent crime, burglary and other crimes, while also including the total numbers of each crime as inputs in the DEA analysis alongside the number of officers. It does not consider the major input measure i.e. total expenditure cost since all the effort and crime prevention measures are directly depend upon the money invested for the purpose. The limitation of the paper is in constructing weight restriction for measuring efficiencies since it is subjective in nature.

Drake & Simper (2000) measured the size efficiency of English and Welsh police forces using DEA and multiple discriminant analysis. In this paper, inferences about the optimal size and structure of the English and Welsh police forces are made using DEA efficiency results and the issue of the statistical significance of the differences in efficiency scores across staff size groups is rectified using analysis of variance (ANOVA) and discriminant analysis techniques. It deals only with the problem of size and structure in English and Welsh policing and not with the overall efficiency of police forces.

Drake & Simper (2003) compare four different distance function models i.e. DEA, free disposal hull (FDH) (Tulkens, 1993), super-efficiency DEA (Anderson & Petersen, 1993) and stochastic frontier analysis (Banker, 1993; Banker et al., 1992) in order to assess police force efficiency of English and Welsh police force. It does not highlight limitations of parametric and non-parametric approaches in case of different crime zones which is present in the data.

Verma & Gavirneni (2006) measured police efficiency in India as an application of data envelopment analysis. This work provides a rationale for identifying good performance practices. It helps in generating targets of performance, the optimum levels of operations, role models that inefficient departments can emulate and the extent to which improvements can be made over a period of time. The paper measures the performances of state police units in India and the results suggest ways in which some State police departments can improve their overall efficiency. But this paper did not consider the aspects of heterogeneous nature of state polices presents in India since some states has larger area and population as compare to other states.

The paper differs from the earlier studies due to the use of DEA in combination of clustering and carries out analysis on the homogeneous groups of police administration units not on all units as earlier done.

3. Methodology

The paper discusses two major phases of the proposed approach. In the first phase, clustering is used to identify the crime zones and to form homogeneous groups in the area of interest. In the second phase, police administration units as DMUs in a particular crime zone are ranked using DEA. The brief descriptions of clustering with some of its popular algorithms and DEA along with its models are given in the subsequent sections.

3.1 Clustering Techniques

In this section, some of the widely known clustering algorithms like K-means clustering, Hierarchical clustering and Self Organizing Map (SOM) have been described in brief.

K-means (McQueen, 1967) is one of the most popular clustering algorithms. K-means is a partitioning method, which creates initial partitioning and then uses iterative relocation technique that attempts to improve the partitioning by moving objects from one group to another. The algorithm is used to classify a given data set into fixed number of clusters (K). K-means uses the concept of centroid, where a centroid represents the center of a cluster. In the process, K centroids, one for each cluster is defined apriori. Then each object of the data set is assigned a group with the closest centroid. The positions of k centroids are recomputed when all objects have been assigned to any of the clusters. The process is repeated until the centroids are no longer move.

Hierarchical clustering (Johnson, 1967) groups the data objects by creating a cluster tree called dendrogram. Groups are then formed by either agglomerative approach or divisive approach. The agglomerative approach is also called the bottom-up approach. It starts with each object forming a separate group. Groups, which are close to each other, are then gradually merged until finally all objects are in a single group. The divisive approach is also called as top-down approach. It begins with a single group containing all data objects. Single group is then split into two groups, which are further split and so on until all data objects are in groups of their own. The drawback of Hierarchical clustering is that once a step of merge or split is done it can never be undone.

SOM (Kohenen, 1990) is a neural network based unsupervised clustering. It maps high dimensional data into a discrete one or two-dimensional space. SOM performs clustering through a competitive learning mechanism. In the process, several units compete for the current object and the unit whose weight vector is closest to the current object becomes the winning or active unit. Only the winning unit and its nearest neighbours participate in the learning process using Mexican Hat function.

3.2 Data Envelopment Analysis

Data envelopment analysis (DEA) was first put forward by Charnes et al. in 1978. DEA is used for evaluating the relative efficiency of decision making units (DMUs) which produce multiple outputs and multiple inputs, via weights attached to input-output measures. DEA uses linear programming problems to evaluate the relative efficiencies and inefficiencies of peer decision-making units. DEA is a nonparametric approach that does not require any assumptions about the functional form of the production function. In the simplest case, where a unit has a single input (X) and output (Y), efficiency is defined as the output to input ratio: Y/X. DEA usually deals with unit k having multiple inputs X_{ik} where $i = 1,2,\ldots,m$ and multiple outputs Y_{rk}, where $r = 1,2,\ldots,s$ which can be incorporated into an efficiency measure. Efficiency measure for DMU k is given by

$$h_k = \underset{u_r, v_i}{Max} \frac{\sum_{r=1}^{s} u_r Y_{rk}}{\sum_{i=1}^{m} v_i X_{ik}} \qquad \ldots (1)$$

where the weights, u_r and v_i, are non-negative. A second set of constraints requires that the same weights, when applied to all DMUs, do not provide any unit with efficiency greater than one. This condition appears in the following set of constraints:

$$\frac{\sum_{r=1}^{s} u_r Y_{rj}}{\sum_{i=1}^{m} v_i X_{ij}} \le 1 \quad for \;\; j = 1,\ldots,n. \qquad \ldots (2)$$

Each unit k is assigned the highest possible efficiency score ($h_k \le 1$) that the constraints allow from the available data, by choosing the optimal weights for the outputs and inputs. If unit k receives the maximal value $h_k = 1$, then it is efficient, but if $h_k < 1$, it is inefficient. Basically, the model divides the units into two groups, efficient ($h_k = 1$) and inefficient ($h_k < 1$) by identifying the efficient frontier of the data. Once the efficient frontier is determined, inefficient DMU's can improve their performance to reach the efficient frontier by either increasing their current output levels or decreasing their current input levels. Cook & Seiford (2009) has reviewed various DEA models developed during last three decades which includes basic DEA models such as constant return to scale (CRS) model (Charnes et al., 1978), variable return to scale (VRS) model (Banker et al., 1984) and specific models like free disposal hull (FDH) model (Tulkens, 1993), cross evaluation (Sexton & Silkman, 1986) and minimum distance models (Frei & Harker, 1999). The two most commonly used models i.e. CRS and VRS are described in brief in the next subsection.

Constant Return to Scale (CRS) Model

CRS model (Charnes et al., 1978) is often referred to as the CCR model based on its founder's name i.e. Charnes, Cooper and Rhodes. The model assumes that the production function exhibits constant returns-to-scale i.e. an increase in inputs leads to a proportionate increase in its outputs. The model can be written into a linear program, which can be solved relatively easily and a complete DEA solves *n* linear programs, one for each DMU.

$$
\begin{aligned}
& h_k = Max \sum_{r=1}^{s} u_r Y_{rk} \\
& subject \quad to \\
& \sum_{i=1}^{m} v_i X_{ij} - \sum_{r=1}^{s} u_r Y_{rj} \geq 0 \quad for \quad j = 1, \dots, n. \\
& \sum_{i=1}^{m} v_i X_{ik} = 1 \\
& u_r \geq 0 \quad for \quad r = 1, \dots, s. \\
& v_i \geq 0 \quad for \quad i = 1, \dots, m.
\end{aligned}
\qquad \dots (3)
$$

It should be noted that the results of the CRS input-minimized or output-maximized formulations are the same.

Variable Return to Scale (VRS) Model

VRS model (Banker et al., 1984) is often called on the name of its originator BCC i.e. Banker, Charnes and Cooper. VRS model adds an additional constant variable, c_k, in order to permit variable returns-to-scale i.e. an increase in inputs does not produce a proportional change in its outputs. It should be noted that the results of the VRS input-minimized or output-maximized formulations are different, which is not the case in the CRS model. Thus, in the output-oriented VRS model, the formulation maximizes the outputs given the inputs and vice versa.

$$
\begin{aligned}
& h_k = Max \sum_{r=1}^{s} u_r Y_{rk} + c_k \\
& subject \quad to \\
& \sum_{i=1}^{m} v_i X_{ij} - \sum_{r=1}^{s} u_r Y_{rj} - c_k \geq 0 \quad for \quad j = 1, \dots, n. \\
& \sum_{i=1}^{m} v_i X_{ik} = 1 \\
& u_r \geq 0 \quad for \quad r = 1, \dots, s. \\
& v_i \geq 0 \quad for \quad i = 1, \dots, m.
\end{aligned}
\qquad \dots (4)
$$

DEA does not use common weights, as do multiple criteria decision theory models, which usually rank the elements based on the multiple criteria (inputs and outputs), and usually provide common weights. The weights vary among the units and this variability is the essence of DEA.

4. Results and Discussions

The two-phase methodology has been applied on crime related data of Indian state police forces. It is needed to describe briefly the Indian police system to understand nature of crime data. Indian constitution assigns responsibility for maintaining law and order to the states and union territories (UT), and almost all routine policing, including apprehension of criminals, is carried out by state-level police forces. India is divided into 28 states and 7 union territories (UT). Measuring efficiencies of state police forces has remained a constant area of governmental concern since these states and UT are having diversities in area, population and crime density. So, first phase of the methodology is to identify crime zones of states with similar crime density using clustering techniques.

4.1 Identification of Crime Zones Using Clustering Techniques

Clustering technique has been employed to Indian crime data to cluster the crime locations as the density of crime incidents differs from one state to another. Since the number of crime zones is known apriori, K means clustering has been used for clustering the crime locations. Crime data contains the crime records of all 28 states and 7 union territories of India for year 2006 under 12 crime heads i.e. Murder, Attempt to murder, Culpable Homicide (C.H.) Not Amounting To Murder, Kidnapping & Abduction, Rape, Cruelty by husband, Dowry Deaths, Dacoity, Preparation And Assembly For Dacoity, Robbery, Riots and Arson. The states have been grouped into three crime zones such as High Crime Zone (H), Moderate Crime Zone (M) and Low Crime Zone (L) based on the densities of various crimes. Table-1 shows the crime densities and crime zones of 28 states & 7 UTs of India for year 2006. Clustering results are shown in the last column of Table-1 as crime zones. Bihar, Karnataka, Kerala, Madhya Pradesh, Maharashtra and Uttar Pradesh are grouped into high crime zone since crime density of all crime heads of these states are much higher than any other states whereas, Andhra Pradesh, Gujarat, Rajasthan and West Bengal with lesser crime density than high crime zone states are clustered into moderate crime zones. The rest of the 25 states and UTs come into the low crime zone category. These homogeneous groups i.e. crime zones are used to measure the efficiencies of police administration units i.e. states and UTs by applying DEA.

4.2 Ranking Police Administration Units using DEA

In the second phase, DEA technique has been used to measure the efficiencies of police administration units as DMUs over their respective crime zone as identified in the first phase. Furthermore, police administration units are ranked based on the efficiency scores. The selection of criteria for analysis i.e. input and output measures play a crucial role in DEA results. The next subsection describes the input and output measures considered for measuring efficiencies of police administration units for effective implementation of crime prevention measures.

Selection of Input /Output Measures

We have identified only the relevant input and output measures from crime prevention measures before applying DEA. The selected input measures for analysis are civil and armed police force strengths and total police expenditure. A police administration unit has high efficiency if it is utilizing its resources well both personnel and financial resources to achieve the desired objectives. Total police expenditure consists of all the money invested for implementing all crime prevention measures.

Outputs are selected by considering the fact that DEA model tries to maximize the output measures to make DMUs as efficient. Therefore, apprehension of criminals i.e. number of person arrest is considered as the first output measures for analysis. The second output measure is the output function of crime rate defined by

$$\text{Output Function of Crime Rate} = 1/\text{Crime Rate} \qquad \dots (5)$$

Here, crime rate is obtained by dividing total crime density of the state/UT with total population of that state/UT since the police of a state/UT is called efficient if its crime rate is low i.e. the output function of crime rate is high.

Table 1: Crime densities & Crime Zones of 28 states & 7 UTs of India for year 2006

State/UT	Murder	Attempt To Commit Murder	C.H. Not Amounting To Murder	Kidnapping & Abduction	Rape	Cruelty By Husband And Relatives	Dowry Death	Dacoity	Preparation And Assembly For Dacoity	Robbery	Riots	Arson	Crime Zone
Andhra Pradesh	2766	1860	123	2030	1049	9164	519	178	3	679	2916	1012	**M**
Arunachal Pradesh	60	33	3	75	37	14	1	28	0	74	6	20	**L**
Assam	1207	414	32	1825	1244	2548	105	319	12	544	2684	488	**L**
Bihar	3249	3303	326	2619	1232	1689	1188	1001	140	2169	8259	785	**H**
Chhattisgarh	1098	825	15	239	995	717	103	160	24	439	905	262	**L**
Goa	39	17	1	16	21	14	0	7	0	17	63	38	**L**
Gujarat	1165	471	24	1128	354	4977	50	290	12	970	1534	321	**M**
Haryana	873	585	59	635	608	2254	255	104	224	410	1142	156	**L**
H P	111	78	20	130	113	259	3	7	0	29	566	115	**L**
J & K	487	854	40	789	250	135	10	10	0	112	1197	203	**L**
Jharkhand	1492	1078	97	635	799	668	281	536	45	779	2650	178	**L**
Karnataka	1627	1484	68	563	400	2129	244	202	247	1349	6183	268	**H**
Kerala	393	347	87	294	601	3708	25	129	125	691	6365	435	**H**
Madhya Pradesh	2309	2370	138	808	2900	2989	764	151	121	1770	2308	815	**H**
Maharashtra	2656	1680	100	1261	1500	6738	387	663	323	2574	7453	1188	**H**
Manipur	205	265	4	130	40	10	0	2	27	8	60	109	**L**
Meghalaya	157	42	8	57	74	13	6	57	4	65	7	28	**L**
Mizoram	25	19	7	6	72	1	0	7	0	10	0	25	**L**
Nagaland	123	46	8	31	23	3	0	16	0	84	7	16	**L**
Orissa	1159	752	36	704	985	694	457	239	76	1184	1535	371	**L**
Punjab	817	898	165	591	442	801	130	35	90	142	3	68	**L**
Rajasthan	1209	1694	79	1970	1085	7038	394	37	97	631	1767	551	**M**
Sikkim	21	16	2	8	20	6	0	0	0	9	12	1	**L**
Tamil Nadu	1363	1599	20	906	457	1248	187	95	4	450	1838	460	**L**
Tripura	154	52	1	105	189	471	35	18	1	71	154	35	**L**
Uttar Pradesh	5480	4997	1543	3318	1314	5204	1798	218	86	2024	3774	299	**H**
Uttaranchal	274	215	40	212	147	358	80	31	2	157	489	39	**L**
West Bengal	1425	672	401	1355	1731	7414	445	177	1263	426	2385	111	**M**
A & N Islands	4	4	1	5	6	7	0	0	0	5	10	9	**L**
Chandigarh	12	21	8	65	19	102	10	1	3	34	44	5	**L**
D & N Haveli	9	0	1	18	6	5	1	5	0	5	8	5	**L**
Daman & Diu	6	3	0	3	3	2	0	8	0	1	24	5	**L**
Delhi	476	510	75	1442	623	1728	137	14	200	541	87	33	**L**
Lakshadweep	0	2	0	0	0	1	0	0	0	0	12	4	**L**
Pondicherry	30	24	3	18	9	19	3	2	0	3	194	22	**L**

Table2: Data of all Indian states/UTs for year 2006 under selected I/O measures

STATE/UT	Civil Police Strength	Armed Police Strength	Total Police Expenditure (Rs. In Crores)	Arrested Person	Output Function of Crime Rate
Andhra Pradesh	66643	12653	1115.97	227935	464.17
Arunachal Pradesh	2961	2481	96.75	2849	511.77

Assam	28341	23708	621.06	58943	667.74
Bihar	43273	8350	897.8	180446	907.27
Chhattisgarh	18147	11948	336	58502	508.89
Goa	3077	445	66.6	3225	703.72
Gujarat	67761	16217	831.32	159810	455.31
Haryana	29201	5279	645.1	66784	458.59
Himachal Pradesh	8818	3255	195.04	19993	492.32
Jammu & Kashmir	39103	20106	898.39	30778	564.05
Jharkhand	25730	3036	725.64	45674	806.13
Karnataka	48011	4041	996.41	142252	478.71
Kerala	35687	7857	606	142301	319.98
Madhya Pradesh	55298	21067	907.99	310782	345.19
Maharashtra	140089	13539	1858	290546	545.24
Manipur	5204	9047	185.62	934	891.82
Meghalaya	5635	3295	124.21	1699	1283.20
Mizoram	2953	4416	103.36	2215	462.61
Nagaland	5497	6450	284.59	906	1941.07
Orissa	27913	10839	429.58	85592	742.46
Punjab	52196	19731	1195.49	45391	813.27
Rajasthan	54766	10622	875.52	185350	442.29
Sikkim	1984	1682	70.75	737	826.46
Tamil Nadu	69913	14327	1323.27	177582	439.36
Tripura	8166	12136	282.63	5114	872.08
Uttar Pradesh	119893	33040	2761.25	217758	1456.85
Uttaranchal	9518	4009	269.7	13471	1101.05
West Bengal	61393	19178	794.28	110346	1265.65
A & N Islands	2204	541	49.44	797	585.80
Chandigarh	3644	419	77.51	3381	326.62
D & N Haveli	217	0	4.8	596	574.71
Daman & Diu	222	0	3.95	372	625.00
Delhi	46694	9739	1389	54198	279.99
Lakshadweep	310	0	4.44	237	837.50
Pondicherry	1437	669	43.14	6187	223.38

Table 2 shows the data of all Indian states/UTs for year 2006 under selected input/output measures for measuring efficiencies of police forces of all states/UTs as DMUs for DEA model. Constant return to scale (CRS) model and variable return to scale (VRS) model of DEA have been applied on the data given in Table-2. The efficiency scores of CRS model and VRS model of DEA with clustering and DEA without clustering are given in the next section of the paper.

Efficiency Scores

The efficiency scores provide the main summary of comparative efficiency. Table-3 shows the efficiency scores of all states/UTs using CRS model and VRS model of DEA with clustering and DEA without clustering.

Table 3: Efficiency Scores

STATE/UT	Efficiency Score with Clustering		Efficiency Score without Clustering	
	CRS	VRS	CRS	VRS
Andhra Pradesh	1	1	0.885	0.963
Arunachal Pradesh	0.243	0.498	0.204	0.455
Assam	0.575	0.732	0.374	0.492
Bihar	1	1	0.992	1
Chhattisgarh	0.968	1	0.578	0.604
Goa	0.348	0.812	0.273	0.795
Gujarat	0.932	1	0.625	0.665
Haryana	0.732	0.98	0.556	0.567
Himachal Pradesh	0.644	0.748	0.414	0.511
Jammu & Kashmir	0.212	0.424	0.143	0.347
Jharkhand	0.565	1	0.488	0.841
Karnataka	1	1	0.88	1
Kerala	0.986	1	0.954	0.985
Madhya Pradesh	1	1	1	1
Maharashtra	0.875	1	0.812	1
Manipur	0.065	0.582	0.065	0.58
Meghalaya	0.104	0.999	0.098	0.992
Mizoram	0.193	0.433	0.164	0.399
Nagaland	0.125	1	0.125	1
Orissa	1	1	0.589	0.738
Punjab	0.242	0.615	0.158	0.482
Rajasthan	1	1	0.889	0.921
Sikkim	0.154	0.761	0.154	0.757
Tamil Nadu	0.863	1	0.594	0.702
Tripura	0.157	0.502	0.132	0.483
Uttar Pradesh	0.58	1	0.376	1
Uttaranchal	0.346	0.852	0.274	0.786
West Bengal	1	1	0.412	0.904
A & N Islands	0.123	0.638	0.116	0.635
Chandigarh	0.317	0.534	0.257	0.386
D & N Haveli	1	1	1	1
Daman & Diu	1	1	1	1
Delhi	0.337	0.459	0.269	0.298
Lakshadweep	1	1	1	1
Pondicherry	1	1	0.795	0.829

All the analyzed states/UTs are given an efficiency score. This score is between 0 and 1. A state/UT with a score of 1 is relatively efficient. Any state/UT with a score of less than 1 is relatively inefficient. States/UTs can also be ranked based on the efficiency score. The efficiency scores of CRS model is less than the efficiency scores of VRS model for both DEA with clustering and DEA without clustering. The results of DEA with clustering approach are better than the DEA without clustering. The comparative evaluation of both the approaches is given in the next section to demonstrate this fact.

5. Comparative Evaluation

The comparative evaluation of both the approaches i.e. DEA with clustering and DEA without clustering is given to judge, which one is the better approach. Comparative efficiency scores and peer groups of both the approaches are shown in this section.

5.1 Comparative Efficiency Score

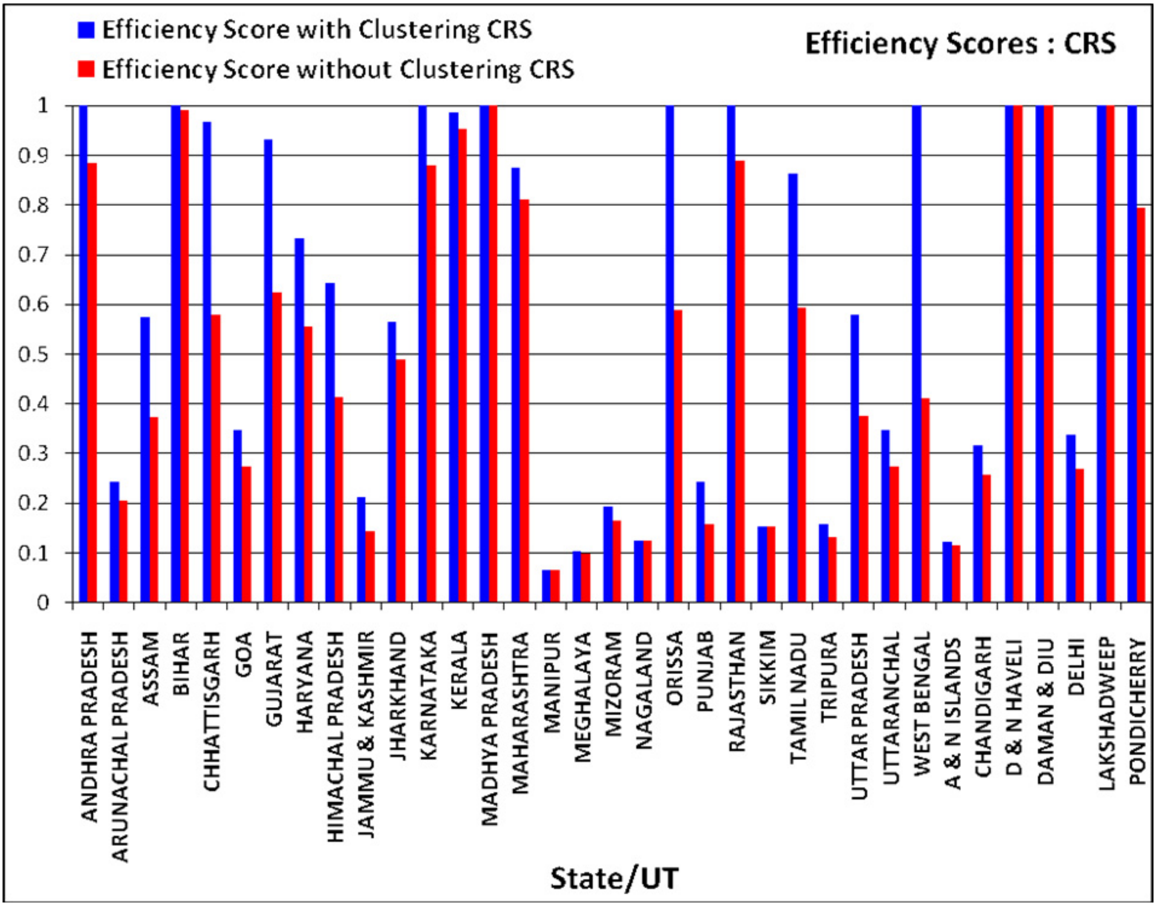

Figure 1: Comparative Efficiency Scores for CRS Model

The mean efficiency score of DEA-CRS model with clustering and without clustering are 0.620 and 0.504 respectively. Furthermore, mean efficiency score of DEA-VRS model with clustering and without clustering are 0.845 and 0.746 respectively. We have applied a paired samples t-test for 5% level of significance level to check that the efficiency scores obtained from both the approaches are statistically same or distinct. According to estimate, t-statistic for DEA-CRS model and DEA-VRS model are 4.904 and 4.712 respectively. Therefore, we conclude that both the approaches are distinct with 100% confidence level. Figure-1 and Figure-2 shows that the efficiency score of DEA without clustering is always less than the efficiency score of DEA with clustering for both DEA-CRS model and DEA-VRS model. It means, DEA without clustering techniques underestimates the states/UTs in measuring their efficiencies and does not produce satisfactory results. Furthermore, DEA with clustering techniques produces the satisfactory results by estimating the DMUs correctly based on their respective crime zones.

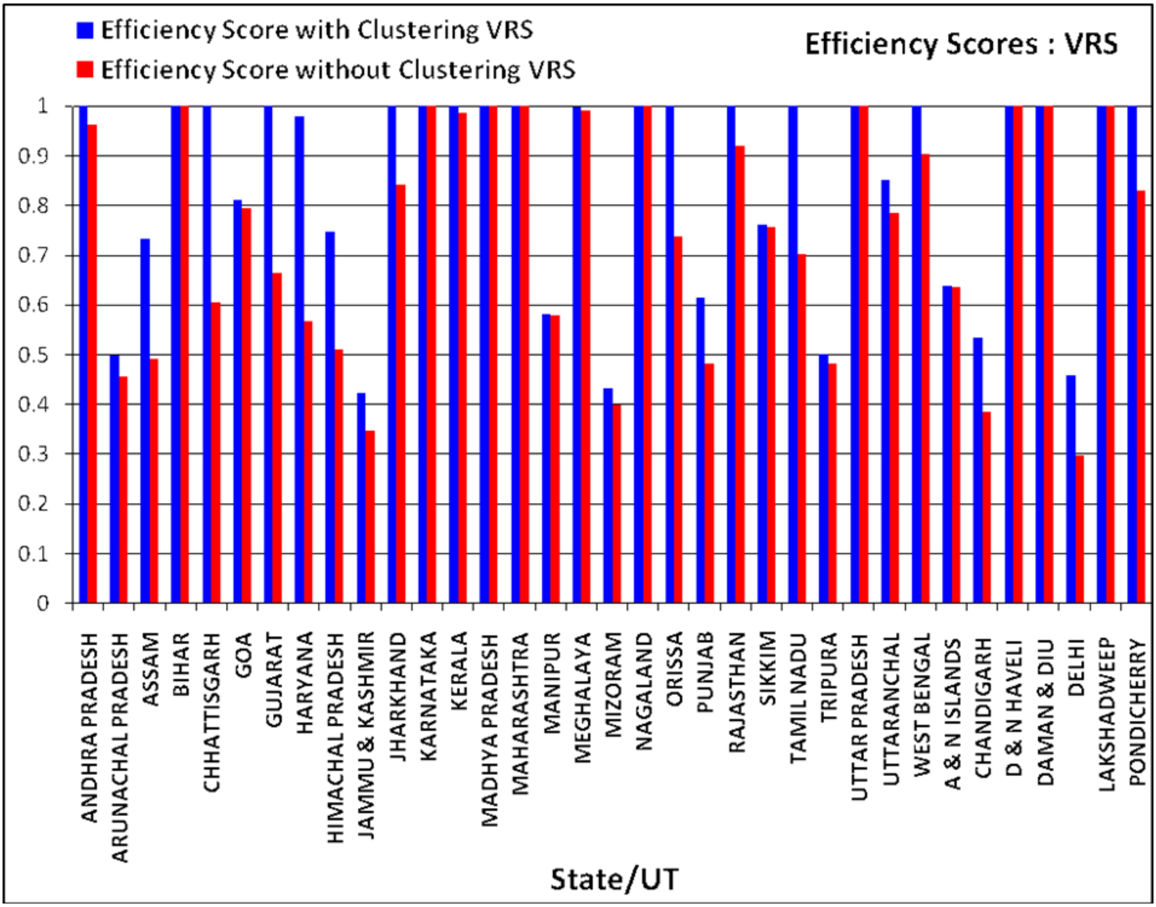

Figure 2: Comparative Efficiency Scores for VRS Model

5.2 Comparative Peer Groups

Table-4 shows the peer groups of DEA-VRS model with clustering and DEA-VRS model without clustering. In Table 4, it is seen that for states like Andhra Pradesh, Karnataka, Kerala, Karnataka etc, peer groups with DEA and clustering have the same states as shown in column 3 of Table 4 since these states are themselves efficient units and there is no need of benchmarking. However for states like Arunachal Pradesh, Assam which are inefficient units, the benchmarking is done with the states (Nagaland, Orissa, Lakshadweep) and (Orissa, Tamil Nadu, Nagaland) which form the peer groups obtained with DEA and clustering.

It is also seen from the Table-4 that both approaches have different peer groups for inefficient units. It is important to note that a state/UT should not have a peer group of states/UTs from different crime zones since a state/UT cannot be benchmark with states/UTs from other crime zones. It is seen from Table-4 that many states/UTs are having peer members from different crime zones with the case of applying DEA without using clustering techniques. For example, Assam is in the low crime zones but having peer members of high crime zone state i.e. Madhya Pradesh and Uttar Pradesh. Similarly, Gujarat is in the moderate crime zone but having peer members from low crime zone state i.e. Nagaland and high crime zone states i.e. Maharashtra, Bihar, Madhya Pradesh. Thus, DEA without clustering fails to recognize suitable peer groups and DEA with clustering approach provides the suitable peer groups in benchmarking the inefficient units to efficient units.

Table 4: Peer Groups

State/UT	Crime Zone	Peer Group of DEA-VRS With Clustering	Peer Group Of DEA-VRS Without Clustering
Andhra Pradesh	M	Andhra Pradesh	Maharashtra , M P, Bihar, Karnataka
Arunachal Pradesh	L	Nagaland, Orissa, Lakshadweep	Nagaland , Lakshadweep, M P
Assam	L	Orissa, Tamil Nadu, Nagaland	Nagaland , M P, U P
Bihar	H	Bihar	Bihar
Chhattisgarh	L	Chhattisgarh	Lakshadweep, M P, Nagaland
Goa	L	Tamil Nadu, Lakshadweep, Nagaland, Jharkhand	Lakshadweep, Nagaland , Karnataka
Gujarat	M	Gujarat	Maharashtra, Bihar, Nagaland , M P
Haryana	L	Jharkhand, Tamil Nadu, D & N Haveli	Lakshadweep, Karnataka, Bihar, D & N Haveli
Himachal Pradesh	L	Pondicherry, Lakshadweep, Tamil Nadu, Orissa	M P, Bihar, Nagaland, Lakshadweep
Jammu & Kashmir	L	Nagaland, Tamil Nadu	M P, U P, Nagaland
Jharkhand	L	Jharkhand	Lakshadweep, Nagaland, Karnataka
Karnataka	H	Karnataka	Karnataka
Kerala	H	Kerala	D & N Haveli, Bihar, M P
Madhya Pradesh	H	M P	M P
Maharashtra	H	Maharashtra	Maharashtra
Manipur	L	Orissa, Nagaland, Lakshadweep	Nagaland , Lakshadweep, M P
Meghalaya	L	Orissa, Nagaland, Lakshadweep	Lakshadweep, Nagaland , M P
Mizoram	L	Lakshadweep, Orissa, Nagaland	Lakshadweep, M P, Nagaland
Nagaland	L	Nagaland	Nagaland
Orissa	L	Orissa	Nagaland , M P, Lakshadweep
Punjab	L	Nagaland, Tamil Nadu	M P, U P, Nagaland
Rajasthan	M	Rajasthan	Bihar, M P, D & N Haveli
Sikkim	L	Orissa, Lakshadweep, Nagaland	Lakshadweep, M P, Nagaland
Tamil Nadu	L	Tamil Nadu	U P, Maharashtra , M P, Bihar
Tripura	L	Nagaland, Lakshadweep, Orissa	Nagaland , Lakshadweep, M P
Uttar Pradesh	H	U P	U P
Uttaranchal	L	Tamil Nadu, Lakshadweep, Nagaland, Jharkhand	Lakshadweep, Nagaland , Karnataka, Bihar
West Bengal	M	West Bengal	M P, U P, Nagaland
A & N Islands	L	Jharkand, Lakshadwep, Nagaland	Karnataka, Lakshadweep, Nagaland
Chandigarh	L	Lakshadweep, Tamil Nadu, Jharkhand, D & N Haveli	Karnataka, Nagaland, Lakshadweep
D & N Haveli	L	D & N Haveli	D & N Haveli
Daman & Diu	L	Daman & Diu	Daman & Diu
Delhi	L	Orissa, Lakshadweep, Nagaland, Tamil Nadu	Nagaland , U P, M P, Bihar
Lakshadweep	L	Lakshadweep	Lakshadweep
Pondicherry	L	Pondicherry	D & N Haveli, M P

Therefore, it can be concluded from this comparative evaluation that DEA in combination of clustering techniques provides better results than DEA in stand-alone mode in measuring efficiencies of police forces.

6. Concluding Remarks

In this paper, the techniques of clustering and Data Envelopment Analysis (DEA) have been applied to measure efficiencies and subsequently to rank police administration units on the basis of their performance in crime prevention measures. These police administration units might be at any level of police administration system i.e. states, district and police station. This paper also illustrates how the concept of clustering is used for effective application of DEA methodology in measuring efficiencies of police forces. The effectiveness of the approach has also been demonstrated for data of India police forces. It can be concluded from the results and discussions that DEA in combination with clustering produces better results than DEA without clustering approach. The proposed approach of measuring efficiencies of police forces is potentially useful to monitor the implementation of crime prevention measures at various levels of police administration on regular basis.

Acknowledgement: We are highly indebted to Shri Sudhir Awasthi, Director, National Crime Records Bureau (NCRB) for funding a project on crime data mining.

References

1. Anderson, P., Petersen, N.C.(1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science,* 39 (10), 1261–1264.
2. Banker, R.D. (1993). Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Management Science,* 39 (10), 1265–1273.
3. Banker R. D., Chang H., Natarajan R. (2007). Estimating DEA technical and allocative inefficiency using aggregate cost or revenue data. *Journal of Productivity Analysis,* 27(2), 115-121.
4. Banker, R.D., Charnes, A., Cooper, W.W. (1984). Some models for estimating technical and scale efficiencies in data envelopment analysis. *Management Science,* 30, 1078–1092.
5. Banker, R.D., Maindiratta, A. (1992). Maximum likelihood estimation of monotone and concave production frontiers. *Journal of Productivity Analysis* 3, 401–415.
6. Bergendahl G., Lindblom T. (2008). Evaluating the performance of Swedish savings banks according to service efficiency. *European Journal of Operational Research*, 185, 1663-1673.
7. Berger, A. N., Humphrey, D.B. (1997). Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operations Research,* 98, 175–212.
8. Butler T. W., Johnson W. W. (1997). Efficiency Evaluation of Michigan Prisons Using Data Envelopment Analysis *Criminal Justice Review*, 22(1) 1-15.
9. Carrington, R., Puthucheary, N., Rose, D., Yaisawarng, S. (1997). Performance measurement in government service provision: The case of police services in New South Wales. *Journal of Productivity Analysis*, 8, 415–430.
10. Charnes A., Cooper W. W., Lewin A., Seiford L. M. (1994) Data Envelopment Analysis: Theory, *Methodology and Applications. Boston,* Kluwer Academic Publishers.
11. Charnes, A., Cooper, W.W., Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research,* 2, 429–444.
12. Cook W. D., Seiford L. M. (2009). Data envelopment analysis (DEA) Thirty years on. *European Journal of Operational Research,* 192, 1-17.
13. Cooper W. W., Seiford L. M., Tone K., Zhu J. (2007). Some models and measures for evaluating performances with DEA: past accomplishments and future prospects. *Journal of Productivity Analysis,* 28(3), 151-163.
14. Drake, L., Simper, R. (2000). Productivity estimation and the size-efficiency relationship in English and Welsh police forces: An application of DEA and multiple discriminant analysis. *International Review of Law and Economics,* 20, 53–73.
15. Drake, L., Simper, R. (2003). The measurement of English and Welsh police force efficiency: A comparison of distance function models. *European Journal of Operational Research,* 147, 165–186.
16. Frei, F., Harker, P. (1999). Projections onto efficient frontiers: Theoretical and computational extensions to DEA. *Journal of Productivity Analysis* 11, 275–300.
17. Jain A. K., Murty M. N., Flynn P. J. (1999). Data clustering: a review, ACM Computing Surveys, 31(3), 264–323.

18. Johnson S.C. (1967). Hierarchical clustering schemes, Psychometrika, 32(3), 241-254.
19. Kao C., Hung H.T. (2005). Data envelopment analysis with common weights: the compromise solution approach. *Journal of Operational Research Society,* 56, 1196–1203.
20. Kohenen T. (1990). The Self Organizing Map. *Proc. IEEE,* 78, 1464-1480.
21. McQueen J. (1967). Some methods for classification and analysis of multivariate observations, *Proc. Symp. Math. Statist. And Probability,* 5th, Berkeley, 1, 281–298.
22. Sexton, T. R., Silkman R. H. (1986). Data envelopment analysis: critique and extensions. Measuring Efficiency: An Assessment of Data Envlelopment Analysis. San Francisco, American Evaluation Association, Jossey Bass, Inc. 32.
23. Thanassoulis, E. (1995). Assessing police forces in England and Wales using data envelopment analysis. *European Journal of Operational Research,* 87, 641– 657.
24. Tulkens, H. (1993). On FDH efficiency analysis: Some methodological issues and applications to retail banking, courts and urban transit. *Journal of Productivity Analysis* 4, 183–210.
25. Verma, A., Gavirneni, S. (2006). Measuring police efficiency in India: an application of data envelopment analysis Policing: *An International Journal of Police Strategies and Management,* 29(1) 125-145.

About the Authors

Manish Gupta is a Scientist in Institute for Systems Studies and Analyses and a doctoral student in the Department of Mathematics, Indian Institute of Technology, Delhi (IIT Delhi). He has previously worked as a Junior Research Fellow (JRF) at Department of Management Studies, IIT Delhi. He has been awarded Junior Research Fellowship under Council of Scientific and Industrial Research (CSIR) schemes in 2001. His research interests include Intelligent Decision Support System, Data Mining, Clustering and Artificial Neural Network.

B. Chandra is a Professor in the Department of Mathematics, Indian Institute of Technology, Delhi. She belongs to the computing group of the Department. She has published a number of research papers in reputed International journals in the area of Neural Networks, Classification, Clustering and Association rule mining. .She has been a visiting Professor for a year at the Graduate School of Business, University of Pittsburgh, USA and worked at World Bank at Washington D.C. During 1998-1999 she has been a visiting Professor Penn State University, USA. She has also been a visiting Scientist at INRIA, France. She has been Chairman in the session on Neural Networks at various International conferences held in USA, UK, Canada, France and Singapore. She has given invited lectures at various universities in USA: University of Maryland, University of Hawaii, University of Pittsburgh, Pennsylvania State University Kaist ,Korea, University of Uttercht, Netherlands, Ecole de Mines Paris and university of Bangor U.K. She has been a principal investigator of many research and consultancy research projects in the of Neural Networks and Machine learning. Among the recent ones are Identification of Crime Hotspots using clustering funded by NCRB and Clustering for developing efficient trading strategies for stock market data She is also the author of three books.

M. P. Gupta is Chair-Information Systems Group & Coordinator-Center for Excellence in E-gov at the Department of Management Studies, Indian Institute of Technology (IIT Delhi). His research interests lies in the areas of IS/ IT planning and E-government. Prof. Gupta has authored acclaimed book “Government Online” and edited two others entitled “Towards E-Government” and “Promise of E-Government”, published by McGraw Hill, 2005. His research papers have appeared in National and International Journals/Conference Proceedings. He was the recipient of the prestigious Humanities & Social Sciences (HSS) fellowship of Shastri Indo Canadian Institute, Calgary (Canada) and a Visiting Fellow at the University of Manitoba. He supervised e-government portal “Gram Prabhat” which won the IBM Great Mind Challenge Award for the year 2003. He has steered several seminars and also founded the International Conference on E-governance (ICEG) in 2003 which running into fifth year. He is on the jury of Computer Society of India (CSI) E-gov Awards and also a member of Program Committee of several International Conferences. He is life member of Global Institute of Flexible Systems Management (GIFT) and Systems Society of India.