

# CLASSIFYING A CHILD AS LABOURER: A SCIENTIFIC STUDY FOR CHILD LABOR CLUSTERING AND DISCRIMINATION

Ahmed F. Siddiqi and Muhammad Khalid Pervaiz

*School of Business & Economics, University of Management & Technology, Lahore, Pakistan, and Department of Statistics, GC University, Lahore, Pakistan*

Received April 2006, accepted August 2006

Communicated by Prof. Dr. M. M. Qurashi

**Abstract:** Child labour is a persistent problem not only in developing countries, as it is usually assumed, but in developed countries as well. The present study analyzed the child labour problem in the Province of Punjab, Pakistan to categorize labouring kids in respect to their household characteristics. Three clusters emerged by splitting the studied household characteristics; “female kids”, “mechanic kids”, and “chotta”. These three types of labourers are different with respect to circumstances, which led them to labour. Such segmentation may help in carving a more objective and pragmatic segment-based policy to tackle this menace. Further, a statistical discriminant analysis was conducted to dig out important factors in child labour dissemination dynamics and to develop a yard-stick to measure the inclination of a household towards child labour. Such measurements may prove helpful in knowing the potential entrants into the household status of having child labourers, which would help in assessing the changing gravity of child labour in a society.

**Keywords:** Clustering analysis, discriminant analysis, labour dissemination dynamics, group centroids, cutting score

## Introduction

While hardly a new phenomenon, the issue of child labour has attracted increasing attention of policy makers, advocates and researchers in the past decade. Child labour is a persistent problem, found throughout much of the developing world (see Assefa and Boyden [1], Rajawat [2]), and to a lesser extent in developed countries Doepke and Zilibotti [3]. While there have been numerous in-depth case studies and attempts to estimate the total number of afflicted children, we still have very little understanding of why there is child labour, or what can be done to successfully eliminate it Ahmed [4]. There is much conjecture, largely based on anecdotes and impressions, about inherent differentials but little rigorous analysis or objective facts. There are also conflicting accounts, reports and theories. Part

of this is no doubt due to the assorted nature of child labour, which fluctuates over time and from place to place. But some of the problems are, surely, due to lack of a consistent, coherent framework for description and analysis, lack of quantitative data, and absence of differential analytics (Anker [5]). Understanding child labour, developing policies and monitoring progress require in-depth positive knowledge and objective analysis of its directly as well as in the indirectly observable differentials (Basu [6]). The availability of detailed and reliable child labour statistics and its analysis on a continuing basis is important. It is particularly important for establishing policy priorities and targets, formulating and implementing interventions, as well as monitoring policies, regulations and programs aimed not only at minimization of the negative consequences of child labour in the short

term, but most importantly for eventual elimination of the practice (Cigno [7]). The objective of this paper was to suggest how data can be used to describe and analyze child labour, and suggest important differential(s) from which to learn about child labour, its patterns, causes and consequences.

There is ample literature available on child labour; ranging from newspaper reports to researched investigations. Experts from different disciplines like Sociology, Economics, Psychology and Statistics have written a lot covering different facets of this menace. There exist hot debates on “Who is a child” or “What is child labour”. ILO/UNESCO sponsored surveys, conducted through NGOs or community based organizations, are available gauging numerical strength of the menace in different locations of the globe. A series of articles/papers are available in research journals pointing out causes and determinants of child labour or dragging out hidden patterns in the spread of child labour. What is missing in all these synergistic efforts is a yardstick based on different characteristics of the household, which may cluster and then discriminate households in respect to their status of having or not having certain characteristics. Such a clustering would segment, on the first hand, child labour in respect to households and their present social, educational and economic settings. On the second hand, the segmentation would also help in carving a more objective and pragmatic segment-based policy to tackle this menace. The same yardstick would also determine and measure propensity of a household to change its status of “child labourer type” to “non child labourer type” or vice-versa. Such measurements may prove helpful in knowing the potential entrants into the household status of having child labourers, which would help in assessing the changing gravity of child labour in a society. It may be helpful in measuring the effects of shop-child-labour policies/steps taken at government echelons. An

attempt has been made here in this paper to develop such a household characteristic-based yardstick.

The paper is survey-based, where original data have been gathered from the Province of Punjab, Pakistan. We are not discussing the agriculture-based (rural) child labour. The focus is primarily on the urban Punjab. It has been assumed here that the household demographics, household literacy, and household financial status are the three most instrumental variables in the dynamics of child labour in any community. All other seemingly important variables are either a derivative of these three or are simply not important at all in reality.

## **Materials and Methods**

The identification and classification of the main types of kids engaged in child labour was carried out in 2005 in order to examine appropriate relevant strategies. It was fully recognized that there exist different circumstances and situations influenced by an individual household’s profile that lead a kid to the turmoil of labour. For example, in the same society one might expect different approaches to be pursued by household heads, and child-labour users. Our focus here, however, was on the identification of broad types and the generic approaches suited to each.

The method used was a comprehensive survey which covered both the types and the approaches used. Although the survey was not conducted exclusively for this paper, it was conducted by the same authors and by giving all the relevant information required here to conduct this study (Siddiqi and Pervaiz [8]). The target population in the survey was the Province of Punjab, Pakistan. This is the biggest province with the largest head count, highest literacy rate, and a cosmopolitan populace. Detailed information regarding the social, educational and economic status of the province was already

available in Encarta [9]. Definitions for the different key terms were taken from the Pakistan Population Census. The size of the survey was determined by using a pilot survey and guidance in DeSantis *et al.* [10] and Adcock [11]. Some 13,130 questionnaires were filled out for 1,313 clusters, using a multi-stage probability proportional systematic sampling plan.

The administrative structure divides the Province of the Punjab into 34 districts, which are further split into 108 *tehsils* (subdistricts). All these 34 districts were stratified in this survey into three strata in respect to their geographies. This resulted in 11, 13, and 10 districts in the three strata, respectively. It was planned to select  $S!^{rd}$  districts from each contrived stratum which gave us, respectively, 3, 4, and 3 districts from the developed strata. On average, every district happened to have 3 tehsils, and using the already decided  $S!^{rd}$  sampling rule, a single tehsil was selected from each selected district. All these selections were probability proportional simple random selections, where weights were assigned according to the educational and economic status of the districts and tehsils using Brewer [12] methodology with an extensions proposed by Sampford [13]. A systematic sampling scheme was adopted to select household clusters from each selected tehsil, where each cluster was made up of 10 households by making use the dictum of FBS [14]. This resulted in a total of 13,130 households that were contacted during the survey. Specially trained interviewers were hired to collect information on household demographics, literacy, and economic status using an exclusively designed 45-question questionnaire (see Siddiqui and Pervaiz [15] for the questionnaire). Microrepresentativity analysis was conducted to ensure the sample representativeness. Serious efforts, both at manual and software level, were made to ensure the accuracy and reliability of the data. A few re-do were also there to rectify errors at surveyors' level. SPSS and JMP were primarily used for all sorts of statistical and

graphical analysis of the data.

### ***Clustering and Discriminating Index for Child Labour in Punjab***

The clustering and discrimination are two quite similar yet different techniques focusing on two different facets. The clustering focuses on grouping respondents on the basis of similar characteristics, while the discrimination is a statistical exercise to develop an index for categorization (see Aldenderfer and Blashfield [16] and Smith and Albaum [17] for details). Such an index helps in extracting important factors that may force a household to send its kids for labour. It also assigns a number to each household that helps in knowing its propensity towards child labour.

#### *Clustering*

This method involved classification of child labourers into different clusters according to their characteristics studied in the questionnaire. The sampling design, discussed in previous section (section 0), consisted of splitting all available data into 3 strata in order to give us a starting point here for clustering of households. We used the Quick Cluster algorithm available both in SPSS & JMP. McIntyre and Blashfield [18] and Aldenderfer and Blashfield [16] have written excellent commentaries on cluster analysis and its different techniques. Quick Cluster is an alternative to hierarchical clustering. It is intuitively appealing as it enables automatic relocation of cases to regions of population density regardless of input order (Hair and Anderson [19]). In addition, it is less demanding of computer space and time for a large sample such as this. The objective of Quick Cluster is to form a predetermined number of clusters from a large sample in such a way that the clusters display a high degree of internal similarity while being distinct from each other. Quick Cluster has three steps:

- In step 1, initial cluster centres are selected.  $K$  cases (where  $K$  is the target number of clusters) are chosen with well separated, non missing values on each of the cluster variables.
- In step 2, Quick Cluster assigns each case in turn to the nearest cluster centre on the basis of Euclidean distance across variables with non missing values. As each case is assigned, the procedure updates the cluster centre to a mean for all the cases in that cluster. As the process of allocating cases continues, the centres of the clusters migrate to concentrations of observations.
- The third and final step reassigns each case to the nearest of the classification centres. This reassignment yields the final clusters. The final cluster centres are then calculated as the mean on each variable for all non missing cluster members.

### *Discriminant analysis*

An attempt was also be made to conduct a discriminant analysis to classify any household assuaging or dissuading its kids for/from labour. Discrimination can be carried out at child-labour Yes/No basis to develop a cut off point that would help in forecasting the status of households for child labour. Smith and Albaum [17] have discussed statistical discriminant analysis in detail. Again, SPSS was used to develop “Fisher linear discriminant function-opting step-wise method” with the smallest  $F$  ratio as the yardstick, assuming that all groups had identical prior probabilities

## **Results**

### *Clustering*

Because the number of clusters is predetermined for Quick Cluster, there can be a problem in identifying the number of clusters necessary

to give a good solution for any set of data. In our analyses we examined the 3, 4, 5 and 6 cluster solutions. The studied household characteristics were listed in the first column of Table 1. A 3-cluster solution gave a fairly good solution with amply and recognizable groups well separated from each other. The 4-cluster solutions resulted in large clusters with unacceptably diverse memberships; merely 14 (31%) characteristics were significantly different among the different clusters. The 5 and 6 cluster solutions resulted in a small “splinter group” of 15 characteristics. When the three substantial groups from the 6 cluster solution were compared with the three groups from the 3-cluster solution, a high degree of correspondence was observed. The three cluster solution was thus adopted as a good representation of the data. Table 2 explicates the distances between these three clusters. Within these clusters, the different characteristics of the households defined the distinctive personality of the households and thus of the child labourer(s) residing there. Table 1 shows analysis of variance (ANOVA) for this three-cluster solution, and reveals 20 (54%) characteristics that were significantly different among the different clusters. The very last column of the Table 1 gives the associated  $p$ -values for the hypotheses of no mean difference among the three clusters. All other characteristics were not well demarcated by this 3-cluster solution. Such a 3-cluster solution also vindicated the 3 strata stratification plan in the sampling design.

Table 1 differentiates among these three clusters for the different households characteristics. The 1st column lays down the household characteristics, while the other three columns differentiate among the status of these characteristics in the three developed clusters. It is always tricky to name a cluster but here the characteristic numbers 3 and 4 were very helpful in his regard. The following results of the study reveal the cluster characteristics.

**Table 1.** ANOVA, signifying the difference between clusters.

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Stratum	5719.805	2	.000	20717	.	.
Location	67077.828	2	.733	20717	91567.075	.000
Gender	.104	2	.089	20717	1.165	.312
Mother tongue	2550.856	2	.870	20717	2932.053	.000
Age group	.251	2	.995	20717	.252	.777
Education	1.185	2	.494	20717	2.400	.091
Number of persons living	6028.742	2	2.390	20717	2522.148	.000
Number of families living	.637	2	.532	20717	1.197	.302
Where do you live?	63.768	2	.742	20717	85.981	.000
Number of rooms	113.219	2	.205	20717	551.512	.000
Number of persons working	3646.313	2	2.543	20717	1434.081	.000
Number earning money	3011.777	2	2.096	20717	1437.230	.000
Number of persons literate	762.799	2	.351	20717	2174.051	.000
Male literates	45.926	2	.75 1	20717	61.140	.000
Children between ages (5-15)	5927.919	2	2.081	20717	2849.248	.000
Number of males working	732.551	2	.827	20717	885.678	.000
Number of females working	715.308	2	.810	20717	883.616	.000
Living with parents	.126	2	.254	20717	.497	.609
Source of income	759.007	2	.365	20717	2080.294	.000
Income	3.131	2	.027	20717	115.747	.000
Send children to same employment	.000	2	.000	20717	.	.
If no, why	.314	2	.479	20717	.656	.519
Are you in debt'?	.000	2	.000	20717	.	.
How much debt'?	1.620E+12	2	47202.483	20717	3.4E+07	.000
Debt, from where	.605	2	.744	2U717	.814	.443
Kids go to school	.000	2	.000	20717	.	.
School is	.047	2	.179	20717	.264	.768
Distance from school	2346.123	2	.362	20717	6479.494	.000
Reasons for not sending to school	794.588	2	.739	20717	1075.890	.000
Kids having literate Fathers	.016	2	.036	20717	.440	.644
Kids having literate mothers	.002	2	.001	20717	1.694	.184
Kids having literate parents	.000	2	.001	20717	.505	.603
Number of kids working at shops	.067	2	.230	20717	.292	.747
Number of kids working as home servants	.292	2	.228	20717	1.282	.278
Number of kids working at construction Ind.	.238	2	.211	20717	1.130	.323
Like a primary school open in locality	.063	2	.240	20717	.263	.769
Child labour affects the health of kids	.649	2	.239	20717	2.712	.006
Distance from medical center	1.269	2	.495	20717	2.562	.007
Which vocational training?	.715	2	2.326	20717	.307	.735
Removal of kids affects financial status	.217	2	.145	20717	1.495	.224
Give pocket money to your kids	.000	2	.000	20717	.	.
How much pocket money?	.761	2	.499	20717	1. 52 5	.218
Would your kids steal something	.095	2	.226	20717	.420	.657
Would your kids go criminal'?	.074	2	.046	20717	1.586	.205
Do you have these facilities in your house	38.102	2	I .270	20717	30.008	.000
What recreation you usually have'?	2.399	2	.691	20717	3.471	.031
Offer new clothes to kids on last Eid	.179	2	.310	20717	.577	.562

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

**Table 2.** Distance between cluster centres.

Cluster	1	2	3
1		29999.102	19999.633
2	29999.102		9999.472
3	19999.633	9999.472	

This cluster was represented predominantly by households with absolutely illiterate and young (20-40 year age group) heads of the household. The household was mostly “Rented and Pakka”, having 3 to 5 rooms with less than 3 adult earning hands. The household was not under any sort of debt. Gender wise, mostly the female kids were working in the services sector principally as servants in homes, which gave cluster its name--the Cluster of Females. The household heads were not willing to send their

**Table 1.** Types of child labourers in Punjab

	Cluster # 1	Cluster # 2	Cluster # 3
Household head: education	<ul style="list-style-type: none"> <li>o Illiterate</li> <li>o Aged (20-40 years')</li> </ul>	<ul style="list-style-type: none"> <li>o Illiterate</li> <li>o Aged (50-60 years')</li> </ul>	<ul style="list-style-type: none"> <li>o Literate (al least primary)</li> <li>o Aged (50-60 years)</li> </ul>
Household type	<ul style="list-style-type: none"> <li>o Rented</li> <li>o Pakka</li> <li>o 3 to 5 rooms</li> <li>o Less than 3 adult earning hands</li> <li>o Abode 1 family only</li> <li>o Under no debt from any source.</li> </ul>	<ul style="list-style-type: none"> <li>o Rented</li> <li>o Pakka</li> <li>o 1 to 3 rooms</li> <li>o Less than 3 adult earning hands</li> <li>o Abode 1 or more families</li> <li>o Under debt often from some private source.</li> </ul>	<ul style="list-style-type: none"> <li>o Rented</li> <li>o Kacha</li> <li>o 1 to 2 rooms</li> <li>o More than 3 adult earning hands</li> <li>o Abode 2 or more families</li> <li>o Under debt often from some bank.</li> </ul>
predominant gender	<ul style="list-style-type: none"> <li>o Female</li> </ul>	<ul style="list-style-type: none"> <li>o Male</li> </ul>	<ul style="list-style-type: none"> <li>o Male</li> </ul>
Work place	<ul style="list-style-type: none"> <li>o Servants in homes mostly for cleanliness</li> <li>o Manufacturing Units</li> </ul>	<ul style="list-style-type: none"> <li>o Auto-workshop</li> <li>o Manufacturing Units</li> <li>o Construction Units</li> </ul>	<ul style="list-style-type: none"> <li>o Tandoors</li> <li>o Small Hotels</li> <li>o Small road-side Shop</li> </ul>
Schooling status	<ul style="list-style-type: none"> <li>o No willingness from household heads</li> <li>o Too expensive to afford</li> <li>o Between 1 to 3km away from home</li> </ul>	<ul style="list-style-type: none"> <li>o No willingness from household heads</li> <li>o Too expensive to afford</li> <li>o Between 1 to 3 km away from home</li> </ul>	<ul style="list-style-type: none"> <li>o Some willingness from household heads</li> <li>o Too expensive to afford</li> <li>o Between 1 to 3 km away from home.</li> </ul>
Vocational Training	<ul style="list-style-type: none"> <li>o No willingness from household heads</li> <li>o Kids want to learn Tailoring, computer and Embroidery</li> </ul>	<ul style="list-style-type: none"> <li>o Some willingness from household heads</li> <li>o Kids want to be Electrician or Auto-mechanic (for cars/trucks)</li> </ul>	<ul style="list-style-type: none"> <li>o Some willingness from household heads</li> <li>o Kids want to learn carpentry, Tailoring</li> </ul>
Health	<ul style="list-style-type: none"> <li>o Having Serious Illness</li> <li>o Backache, Joint Pain, Fever, Respiratory</li> <li>o Medical centres are 1 to 3km awa</li> </ul>	<ul style="list-style-type: none"> <li>o Having Serious Illness</li> <li>o Cough, Fever, Joint Pain</li> <li>o Medical centres are 1 to 3km away</li> </ul>	<ul style="list-style-type: none"> <li>o Having Serious Illness</li> <li>o Fever, Cuts, Stomach</li> <li>o Medical centres are more than 3km away</li> </ul>
Recreation	<ul style="list-style-type: none"> <li>o TV (B/W) at home</li> <li>o Cassette Player</li> <li>o No new clothes for a year</li> </ul>	<ul style="list-style-type: none"> <li>o TV (B/W) at home</li> <li>o Usual visitor of cinema</li> <li>o No new clothes for a year</li> </ul>	<ul style="list-style-type: none"> <li>o No TV at home</li> <li>o Usual visitor of cinemas</li> <li>o New clothes after 3 months</li> </ul>

kids to school. Also, the schools were very expensive and far from their homes. In most cases, the kids wanted to learn some skill but their parents were not willing. Furthermore, most of these households had a television set and cassette player but the kids were not provided new clothes for an entire year. Labourer kids usually suffered from backache, joint ache, and respiratory problems and the nearest medical centre was 3 kilometres away.

### *Cluster 2: Cluster of Mechanics*

This cluster was represented predominantly by households with illiterate, senior (50-60 year age group) household heads. The household was "Rented & Paka", having 1 to 3 rooms with less than 3 adult earning hands and housed one or more families. The household was under severe debt (privately acquired). Most of the kids were male working at "Auto-workshops" and in the manufacturing sector. This cluster was thus designated Cluster of Mechanics. The household heads were not willing to send their kids to school and they could also not afford this luxury. Most of these households had a television set and cassette player but the kids were not provided new clothes for an entire year. The Labourer kids usually suffered from cough, fever, and joint problems and the nearest medical centre was again 3 kilometres away.

### *Cluster 3: Cluster of Chotas*

This cluster was represented predominantly by households with literate (at least primary educated) senior (50-60 year age group) household heads. The household was "Rented & Kacha", having 3 to 5 rooms with more than 3 adult earning hands and housed two or more families. The household was under severe debt (taken, on average, from banks). Most of the kids were male working at "Tandoors", "Small Hotels" and in the manufacturing sector. This named as Cluster of Chotas--a Punjabi translation for call boys working at shops, small road-side hotels, and the like. The household heads were somewhat willing to send their kids to school even if they could not afford and in spite of the long distance of these from their homes. The households usually did not

have television sets, and cinemas served as a common recreation source. The kids were provided new clothes usually after every 3 months. The kids suffered from stomach problems and other serious illnesses while the nearest medical centre was farther than 3 kilometres from home.

The above segmentation revealed the startling instrumental nature of literacy in the dynamics of child labour. The first two clusters were segmented as economically better-off households (living in pakka houses) yet their heads were not willing to send the kids to schools on the pretext of expense. This was correlated to the illiteracy of the household heads. Correlation calculation using raw data showed a coefficient of 0.643 between illiteracy of the household head and the incidence of child labour. This coefficient was statistically highly significant, stamping the veracity of such a relationship. While the third cluster, not being economically better-off households but being literate, showed willingness to send their kids to school. Another, important revelation was the inclination of the kids towards vocational training. However, such training was not a choice for the female labourer kids despite the latter's willingness to gain it. The kids knew well the type of skill they want to have. One of the disturbing information the segmentation yielded was that the somewhat literate household heads (in cluster 3) were unwilling to send their kids to skill-oriented labour such as auto-workshops and manufacturing units. Also, as more than one family were living together, in cluster 3, such a trend was likely to spread speedily. The labourer kids got sick more often than not but the nature of illness varied among the different kids. Table 1 shows these different types of illnesses. Amazingly, the cause-and-effect relationship between the nature of illness and the nature of jobs was explicit in all three clusters. This would explain the reason for frequent illnesses among the labourer kids.

### *Discriminant Score*

Tables 2 to 8 give the detailed results of such a discriminant analysis.

**Table 2.** Discriminant analysis.

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.956 <sup>a</sup>	100.0	100.0	.699

<sup>a</sup>. First 1 canonical discriminant functions were used in the analysis.

**Table 3.** Standardized canonical discriminant function coefficients.

	Function
	1
Stratum	-.140
Number of Persons Living	-.062
Numbers of Persons Literate	.034
Children Between Ages (5-15)	.217
Number of Males Working	.843
Number of Females Working	.843
How Much Debt?	.168
Distance from School	.025

**Table 4.** Discriminant functions at Group Centroids.

Status of Child Labour	Function
	1
No Child Labourer	-1.682
At Least 1 Child Labourer	.568

Unstandardized canonical discriminant functions evaluated at group means

**Table 5.** Classification function coefficients.

	Status of Child Labour	
	No Child Labourer	At Least 1 Child Labourer
Stratum	-22.070	-22.513
Number of Persons Living	1.115	1.021
Numbers of Persons Literate	-7.753	-7.634
Children Between Ages (5-15)	7.745	8.145
Number of Males Working	-9.478	-7.243
Number of Females Working	-9.509	-7.252
How Much Debt?	.002	.002
Distance from School	19.061	19.135
(Constant)	-139.435	-145.022

Fisher's linear discriminant functions

**Table 6.** Classification results.

Status of Child Labour	Predicted Group Membership		Total	
	No Child Labourer	At Least 1 Child Labourer		
Original Count	No Child Labourer At Least 1 Child Labourer	11654 4833	0 29665	11654 34498
%	No Child Labourer At Least 1 Child Labourer	100.0 14.0	.0 86.0	100.0 100.0

The eigenvalue in Table 2 shows the 95.6% veracity of the used Fisher discriminant function; confirming 95.6% observations. The first column of Table 3 shows the household characteristics rendered important by the discriminant analysis. The coefficient for the importance is in the corresponding 2<sup>nd</sup> column. Although the characteristics are not sorted in the order of importance even then they gave a fairly good idea of their respective importance. The financial status of a household, represented by characteristics such as “Number of Males Working” and “Number of Females Working”, showed the highest values for these important coefficients. The household demographics, represented by “Children between ages 5-15” came the next in order of importance. The educational status of the household is presented in Table 4 by “Number of Persons Literate” and “Distance from School”. While its importance for the discriminant function was not great, it was still statistically significant. The presence of “Stratum” in this Table was quite meaningful in several senses. It indicates that the child labour differentials differed stratum/region wise.

Table 4 shows the group centroids in the Euclidean space, clearly demarcating the households with or without child labourers. The difference of the sign was quite meaningful in the sense that the discriminant analysis rendered these two categories as opposites of each other. Table 5 shows the respective importance of the different household characteristics for households with and without child labourers.



Table 6 shows the veracity of the developed Fisher linear discriminant function. Evidently, the function correctly classified 89.5% of the cases. Such a high proportion signified the power of the function.

It is interesting to observe in the discriminant function coefficients (Table 3) that the more economically active household contributed more positively to the incidence of child labour, while the more literate household contributed less (rather much less) positively. The population of the household contributed negatively; the more the people in a household the lesser the chance of child labour. The population of kids in a household, however, contributed significantly and positively. The important variables emerging in Table 3 were enough to create a cutting score for the household classification.

*Cutting Score*

A cutting score was calculated as weighted average of the centroids of the two groups. The following depicts development of such a score.

$Z_{CUT} = \frac{n_A Z_B + n_B Z_A}{n_A + n_B}$	Where $Z_{CUT}$ = Critical Cutting Score $Z_A$ = Centroid for Group A $Z_B$ = Centroid for Group B $N_A$ = Observations in Group A $N_B$ = Observations in Group B
---	--

By substituting appropriate values in the formula, we obtained the critical cutting score (assuming equal costs of misclassification):

$$\{1\} Z_{CUT} = \frac{16676(0.356) + 48974(-1.044)}{16676 + 48974} = -0.6884$$

- Classified a household as having child labourers if its discriminant score was greater than -0.6884.
- Classified a household as having no child labourers if its discriminant score was less than -0.6884.

Using the characteristics defined in the first column of Table 3 and their respective coefficients in the second column, a score could be calculated for each household. This score, in comparison with the cutting score calculated in Eq. {1}, may be used to give a verdict about the status of the household as assuaging or dissuading child labour.

**Discussion**

Child labour is a persistent problem not only in developing countries but also in the developed countries. In the present study an attempt was made to classify the existing child labour in respect to household characteristics and development of a child labour index that may help in knowing the status and propensity of a household for child labour. The study shows that the existing child labour in the Province of Punjab (Pakistan) can be split into three major clusters; female kids, mechanic kids, and chotta. These clusters turned out to be well demarcated and statistically significant. In the first instance, such a clustering segments child labour in respect to households and their present social, educational and economic settings, while in the second instance it also helps in carving out a more objective and pragmatic segment-based policy to tackle this menace. The same yardstick tends to also determine and measure the propensity of a household to change its status either from having child labourer to not having them or the other way around. Such measurements may prove helpful in knowing the potential entrants into the household status of having child labourers, which would help in assessing the changing gravity of child labour in a society. The segmentation establishes the surprising differential nature of literacy in the incidence of child labour. The first two clusters segment economically better-off households (living in Pakka houses) while their heads were unwilling to send their kids to schools on the pretext of expense. The third cluster, being literate and not containing economically better-

off households, showed willingness to send their kids to school. The segmentation also reveals that somewhat literate household heads (in cluster 3) were unwilling to send their kids to skill-oriented labour. Further, as more families live together in cluster 3, such a trend is likely to spread more speedily. The discriminant function, developed in Table 3, also reveals that a more economically active household contribute more positively in the incidence of child labour, while a more literate household contributes less positively. The population of the household contributes negatively; the more the people in a household the lesser the chance of child labour. On the other hand, the population of kids in a household contributes significantly and positively. Table 1 and 5 may prove very helpful in carving out some policy for these crippled kids. Both of these Tables give a segment-based picture of the labourer kids. Since the survey was conducted in the Province of Punjab (Pakistan), the results presented should be applicable strictly to this region only. However, the design of the study is powerful enough for application in other parts of the world, especially those having similar socio- economic and educational conditions.

The results presented in this paper are unique in the sense that a peculiar segmentation is attempted on the profile of a child labourer. When the others are exploring means to count the total number of child labourers, or discussing important differentials and determinants of child labour in any society, this segmentation is going one step ahead in addressing this menace. It is showing how the supply of child labourers should be circumscribed more succinctly by addressing child labourers according to their profiles.

### Acknowledgements

The authors are thankful to the Editor for his valuable comments which helped greatly in shaping this article.

### References

1. **Assefa, B. and Boyden, J.** 1988. *Combating child labour*. Geneva: International Labour Office, xiii, 226 pp.
2. **Rajawat, M.** 2004. *Child labour: The Indian perspective*. New Delhi, Kilaso Books, viii, 265 pp.
3. **Doepke, M. and Zilibotti, F.** 2003. *Voting with your children: A positive analysis of child labour laws.*: Centre for Economic Policy Research, London, 53 pp.
4. **Ahmed, I.** 1999. Getting rid of child labour. *Economic and Political Weekly* 34:1815-1822.
5. **Anker, R.** 2000. The Economics of Child Labour: A framework for measurement. *Internat. Labour Rev.* 139:257-280.
6. **Basu, K.** 1999. Child labor: cause, consequence, and cure, with remarks on international labor standards. *J. Econom. Literature* 37:1083-1119.
7. **Cigno, A.** 2003. Globalisation can help reduce child labour. *CESifo Econom. Stud.* 49:515-526.
8. **Siddiqi, A.F. and Pervaiz, M.K.** 2005. Child labour and literacy; regression analytics investigating the role of literacy in alleviating child labour. *European J. Scientific Res.* 4:5-24.
9. **Encarta.** 2006. *Child labor*. In: *Microsoft Encarta 2006, Microsoft Corporation, 2005*. Redmond, WA.
10. **DeSantis, F., Pacijico, M.P. and Sambucini, V.** 2004. Optimal predictive sample size for case control studies. *J. Roy. Statist. Soc., Ser. C* 53:427-441.
11. **Adcock, C.J.** 1997. Sample size determination; a Review. *J. Roy. Statist. Soc. Ser. D* 46:261-283.
12. **Brewer, K.W.R.** 1963. A model of systematic sampling with unequal probabilities. *J. Roy. Statist. Soc. Ser. A* 5:5-13.
13. **Sampford, M.R.** 1967. On sampling without replacement with unequal probabilities of selection. *Biometrika* 54:499-513.
14. **FBS.** 2001. *Household integrated economic survey of Pakistan 1998-99*. Federal Bureau of Statistics, Pakistan, Islamabad.
15. **Siddiqi, A.F. and Pervaiz, M.K.** 2005. *Child labour; a statistical study using multistage probability proportional stratified systematic sampling*. GC University, Lahore, Pakistan.
16. **Aldenderfer, M.S. and Blashfield, R.K.** 1984. *Cluster analysis*. California: Sage Publications.
17. **Smith, S.M. and Albaum, G.S.** 2005. *Fundamentals of marketing research*. Thousand Oaks, California, xii, 881 pp.
18. **McIntyre, R.M. and Blashfield, R.K.** 1980. A Nearest-Centroid technique for evaluating the minimum variance clustering procedure. *Multi-variate Behav. Res.* 15:225-238.
19. **Hair, J.J.F. and Anderson, R.E. Tatham, R.L. and Black, W.C.** 1998. *Multivariate data analysis*. 5th Ed. Singapore, Pearson Education.