Multilevel Homogeneity Analysis

Jan de Leeuw and George Michailides
UCLA Statistics

April 10, 1999

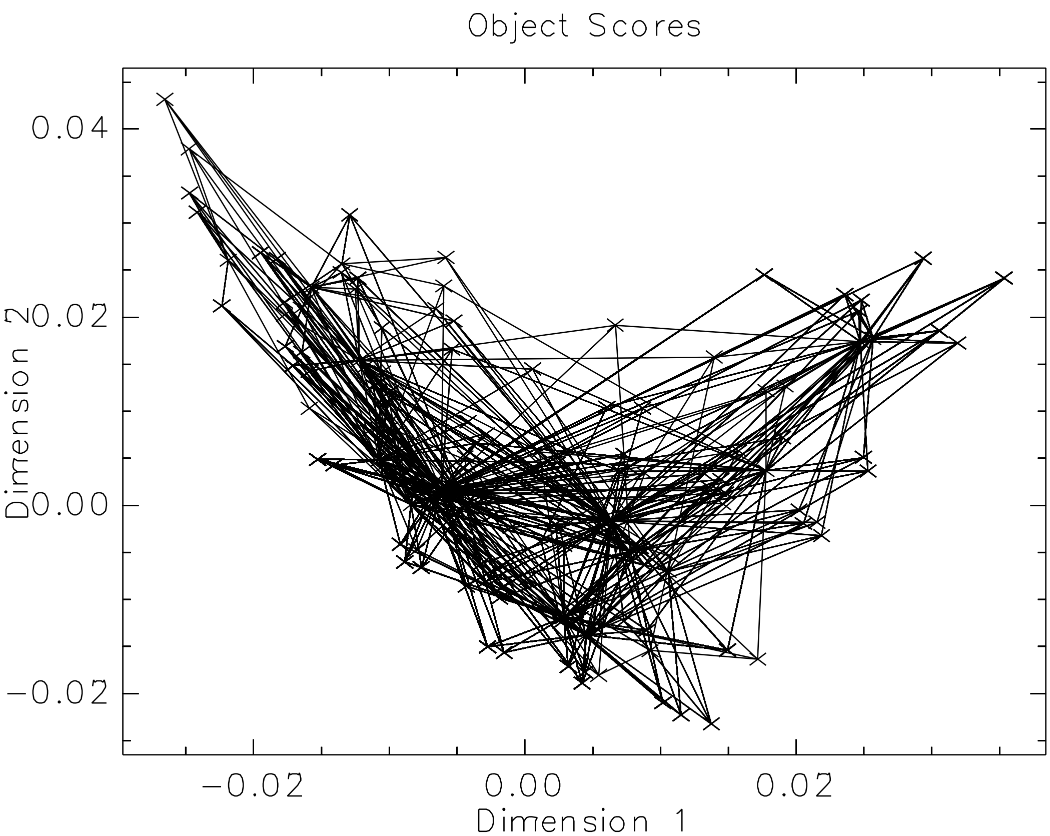
Analysis of large educational datasets, carried out over the last few years at UCLA Statistics, has evolved around two basically different data analysis techniques.

The first of these techniques is homogeneity analysis, also known as MCA, multiple correspondence analysis. It is a technique closely related to principal component analysis and multidimensional scaling. The emphasis is on making visual representations of a low-dimensional projection of the data.

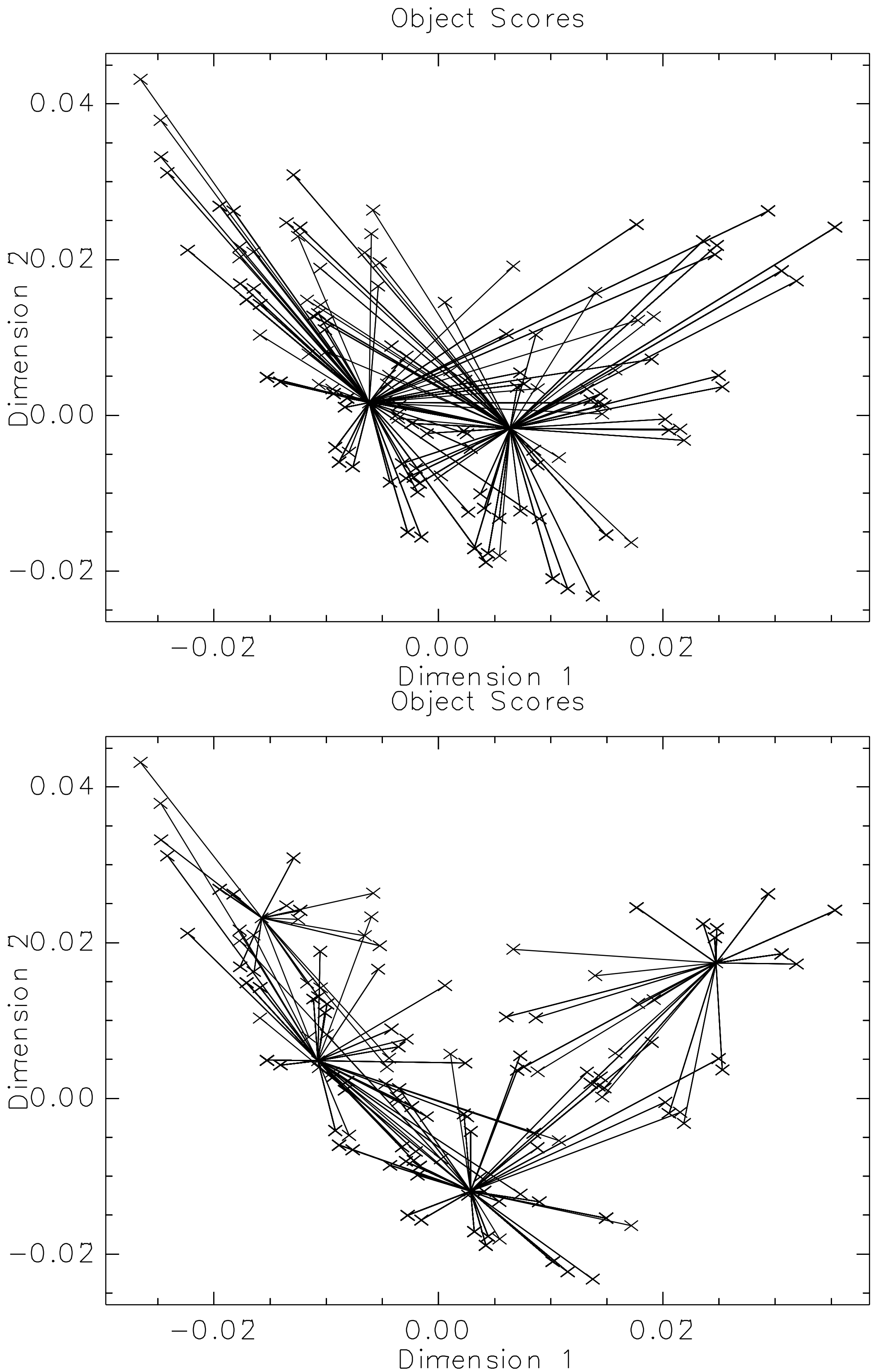
The second technique is multilevel analysis, also known as HLM, hierarchical linear model analysis. This is, in its simplest and most well-known form, a linear regression method which takes the hierarchical nature of the data into account.

This hierarchy comes about, because students are in classes, classes are in schools, schools are in districts, and so on. We may have variables describing units at all levels.

In MCA the key concept is the graph-plot. Here is a graph-plot for the GALO data (1024 kids in 32 schools in 1959).



This is perhaps a bit easier to understand if we realize that this MCA is based on the four variables Gender, IQ, SES, and Teacher's Advice. The graph-plot is an overlay of the four star-plots, of which we only show the ones for Gender and SES.


Object Scores
0.04
0.02
0.00
−0.02
Dimension 2
−0.02
0.00
0.02
Dimension 1
Object Scores
0.04
0.02
0.00
−0.02
Dimension 2
−0.02
0.00
0.02
Dimension 1

There is a version of these plots in a formula.

The objects scores are in an $n \times p$ matrix X, the category quantifications of variable j are in an $k_j \times p$ matrix Y_j. The lines in the star plot have squared length (star-loss)

$$\sigma_j(X, Y_j) = \mathbf{SSQ}(X - G_j Y_j),$$

where G_j is the $n \times k_j$ indicator matrix that codes the correspondence between the objects and the categories they are in.

The squared length of the lines in the graph plot (graph-loss) is simply the sum over j of the m star-losses, and the technique of homogeneity analysis locates the objects in $\mathcal{R}^p$ such that graph-loss is minimized, where X is restricted by $u'X = 0$ and $X'X = nI$.

In this formulation, the fact that we square the lengths, and the particular choice of normalization, are both not essential for the definition of the technique.

There is no school information incorporated, or shown, in the MCA results. But there are at least three easy ways in which we could incorporate school information.

1. Add school as an additional variable. This means we want to make the school-stars small.

2. Add school as a passive variable. This means if we find out after the analysis if the school stars are small.

3. Perform a separate MCA for each school.

In this talk/paper we will study these various options, and extend them by imposing restrictions on the category quantifications.

But first we have to discuss the second class of techniques.

In HLM we have the following situation. An outcome y_k is related to a number of predictors X_k by the rule $y_k \approx X_k\beta_k$. Here index k is used for a group, such as a class or school. We restrict our attention here to the case of two levels, students and schools.

It looks as if the situation simply requires fitting a separate regression model for each school, but this may be impractical (too many schools, not enough students per school) or otherwise unwise (schools are supposed to be related, they are all in the same system). In multilevel analysis this relatedness is conceptualized with a second level model of the form $\beta_{ks} \approx z_k'\gamma_s$, i.e. with a regression model in which the first-level regression coefficients are the outcomes of a number of second level predictors (school-variables). We do no go into the variance component error-structure of the HLM, because it does not seem to be an essential part of the model.

Since

$$y_{ik} \approx \sum_{s=0}^{p} x_{kis} \beta_{ks},$$

we see that

$$y_{ik} \approx \sum_{s=0}^{p} \sum_{t=0}^{q} \{x_{kis} z_{kt}\} \gamma_{st}.$$

Thus we see that HLM is a regression model in which each predictor is an interaction between a student and a school variable. Because of the convention $x_{ki0} \equiv 1$ and $z_{k0} \equiv 1$ we see that the student and school level variables themselves are also in the model.

In the balanced case, in which $x_{kis} = x_{is}$, we can write

$$Y \approx X \Gamma Z'.$$

Also, observe that the restrictions on the regression coefficients can be written as $B = Z\Gamma'$.

It seems that the two techniques are not really related. MCA is a technique for the analysis of interdependence, HLM for the analysis of dependence. Nevertheless, the general HLM idea, of restricting structural parameters in each of the separate regressions, can be extended to MCA. And the general MCA idea, to have low-rank approximations to data and parameter structures, can be extended to HLM.

Let us first extend HLM in a straightforward way. We have seen that the restriction imposed on the regression coefficients in the various groups is $B = Z\Gamma'$, where B is the $K \times p$ matrix of regression coefficients, Z is the $K \times q$ matrix of group-level predictors, and Γ is the $p \times q$ matrix of regression coefficients.

The extension we have in mind is also of the form $B = Z\Gamma'$, but now we leave the possibility open that Z is completely or partly unknown. Thus, if Z is completely unknown, we incorporate a rank q approximation of B into the technique, i.e. we perform a little principal component analysis on the regression coefficients.

This can be extended, in the obvious way, with optimal scaling, i.e. some of the columns of Z can be known completely, known up to order, and so on. Since maximum likelihood for HLM is really a two-step procedure, in which we impose the restrcitions on B in the second step, efficient algorithms can be applied.

Instead of imposing the school-level variables, we discover them, using PCA.

In the rest of the talk we concentrate on adapting MCA in the HLM direction.

For this it is convenient to extend graph-loss to

$$\sigma(X,Y)=\sum_{k=1}^{K}\sum_{j=1}^{m}\mathbf{SSQ}(X_k-G_{jk}Y_{jk}),$$

where we have assumed that the same m variables are observed in each of the K groups.

Minimizing this over X and Y, with normalization $X_k'X-k=n_kI$, is the same as K separate MCA's. If we impose the normalization $X'X=nI$, then it is only one MCA, but one in which the G_{jk} within each group are replaced by their direct sum.

In order to reduce the number of parameters (or smooth) we can first think of using equality restrictions. We can require some of the Y_{jk} to be equal, or some of the X_k to be equal. This easpecially makes sense in balanced situations, such as panel or repeated measures data. There the groups are measurements on the same individuals and the same variables, at different time points. Thus we could write X_t and Y_{jt}, and because individuals are the same we could require $X_t = X$, for instance.

This has been explored before in a series of papers on extending MCA to time-varying measurements (Deville, Saporta, Van der Heijden, De Leeuw, Van Buuren).

These equality restrictions stay very close to classical MCA/CA, and we will not discuss them any further. See Michailides (UCLA Thesis, 1996).

Instead, we shall illustrate the restrictions $Y_{kj} = Q_j B_k$, with B_j diagonal. This means, in terms of the graph plot, that the Y_{kj} are no longer centroids. There is only one set of category quantifications Q_j for variable j, and the dimensions of these quantifications are stretched/shrunk differentially in the various groups.

There is a clear connection with three-way methods here. For each fixed j, the Y_{kj} satisfy a restricted PARAFAC/CANDECOMP model of the form $Y_{kj} = Q_j B_k I_j$, which could also be extended to $Y_{kj} = Q_j B_k P_j$. As in the three-way case, rotational indeterminacy of the solution disappears.

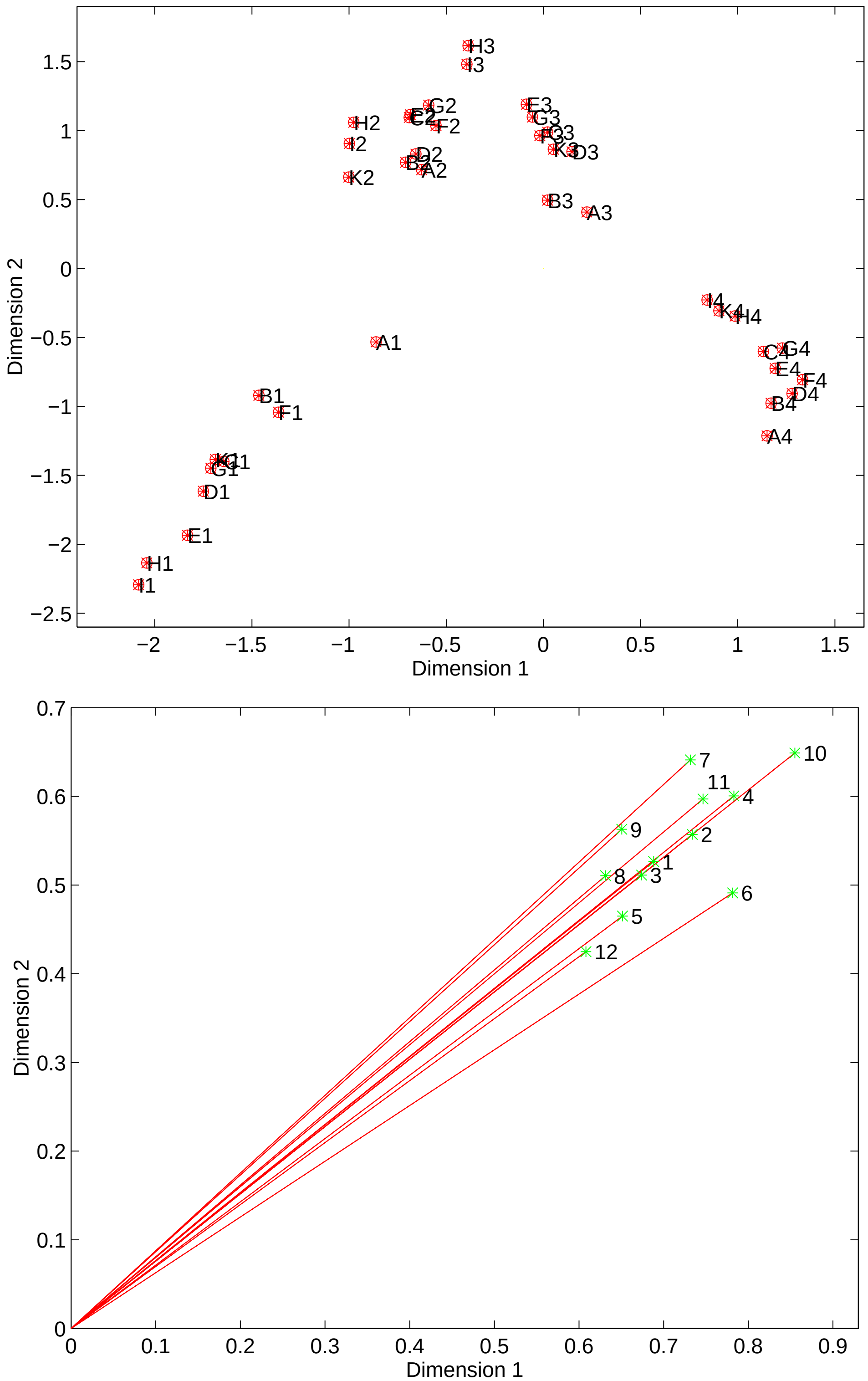
We apply this technique, which uses the classical alternating least sqaures methods based on the least squares partitionings of the loss function, to the NELS-88 example. In this example 11 variables describing problem behavior, and 498 students in 12 schools were selected.

The first pair of plots shows the Q_j and the B_k (two-dimensional solution)

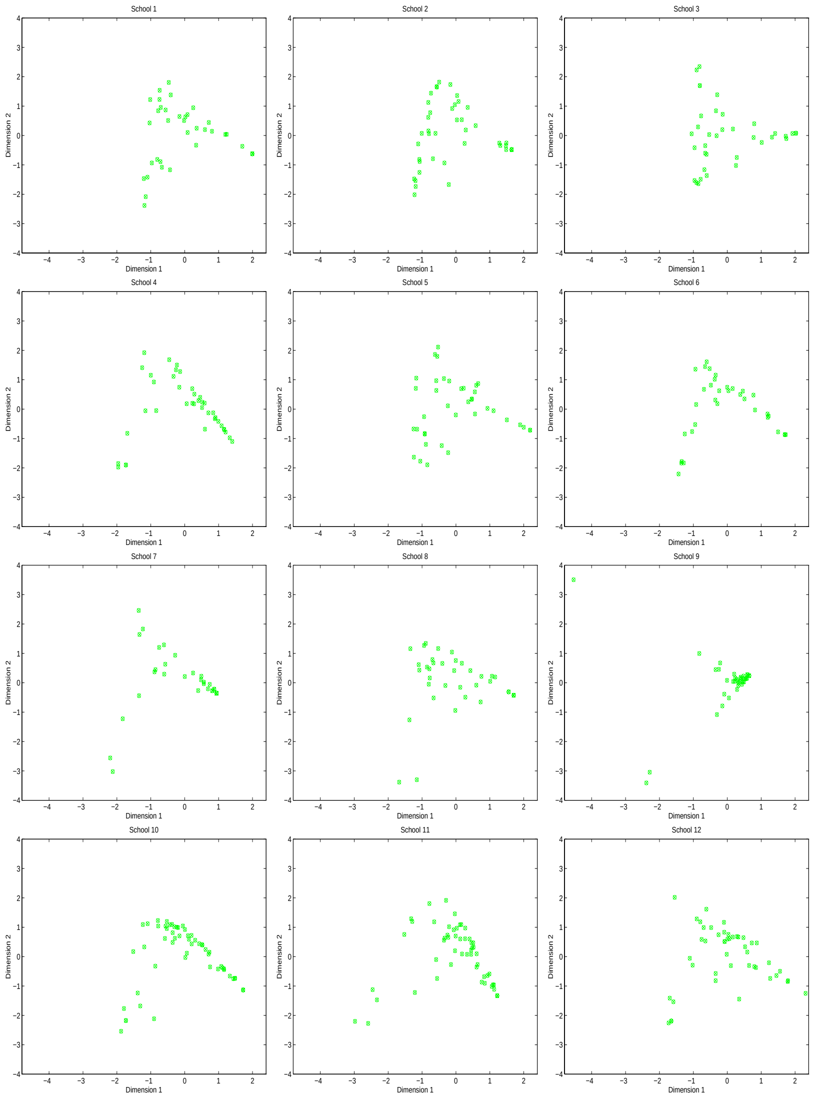
The next two plots show object scores for unrestricted and restricted solutions. In these last plots school 1,2,3 are public urban; 4,5,6 are public suburban; 7,8,9 are public rural; 10,11,12 are private.

Detailed discussion of the results is in the paper. For the time being we concluded that some of the irregularties are ironed out, and the interpretatbility is improved.

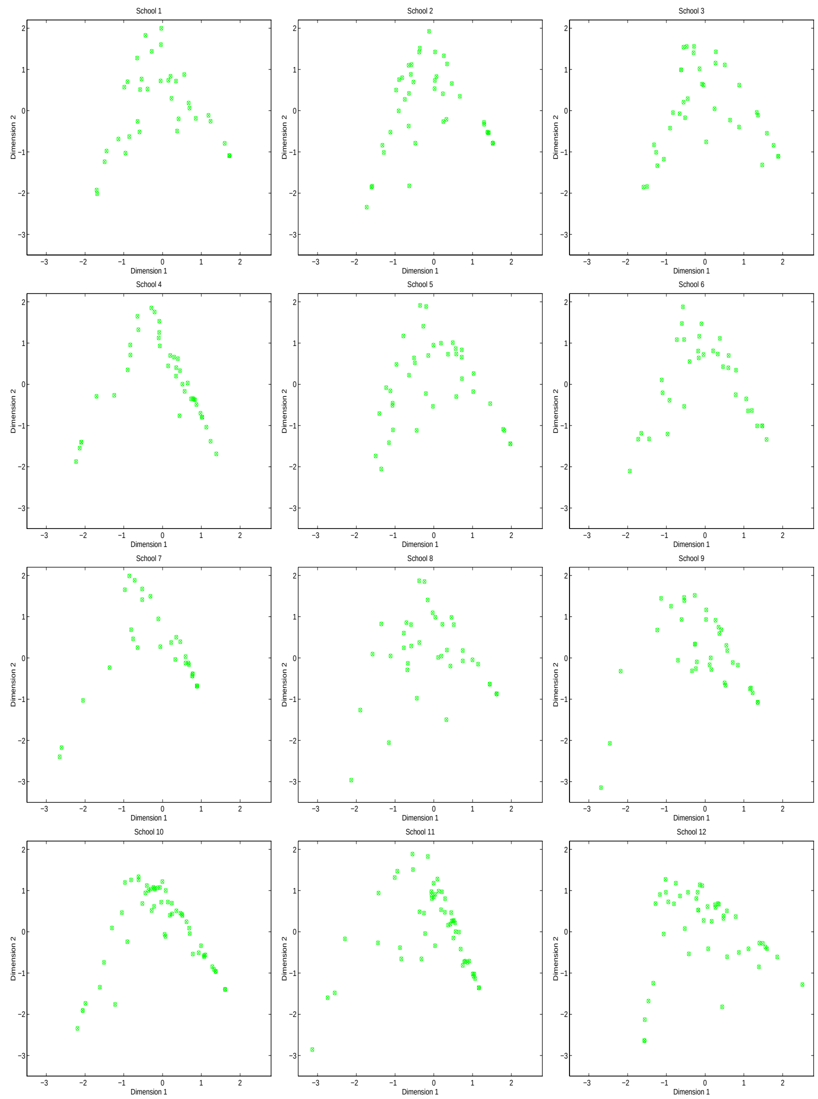
More generally, we conclude that there is much unexplored and possible fruitful terrain between MCA and HLM.



Dimension 1
Dimension 2
H3
I3
G2
E3
G3
H2
F2
I2
D3
D2
A2
K2
B3
A3
A1
H4
G4
E4
F4
D4
B4
A4
B1
F1
G1
D1
E1
H1
I1
Dimension 1
Dimension 2
7
10
11
4
9
2
1
8
3
6
5
12


School 1
School 2
School 3
School 4
School 5
School 6
School 7
School 8
School 9
School 10
School 11
School 12
Dimension 1
Dimension 2

1


School 1
School 2
School 3
School 4
School 5
School 6
School 7
School 8
School 9
School 10
School 11
School 12
Dimension 1
Dimension 2