

Assessing the Disprof Test for Significant Clustering: Is It Suitable for Strategic Groups Research?

Charles Carroll

Univeristy of Groningen, Netherlands

ABSTRACT

Research on strategic groups has been hindered by the lack of significance testing in cluster analysis. Consequently, researchers cannot determine if a cluster analysis has invented statistical groupings as an analytic convenience or actually discovered discrete strategic groups that could constitute pockets of oligopolistic competition within that industry. Ideally, a permutation technique would impose the conditions of the null hypothesis (no clustering) while preserving all of the other characteristics of that data. Unfortunately, the closest approximation for cluster analysis is achieved by independently permuting each variable. This destroys all multivariate structure in the data including, but not limited to, multivariate clustering. Hence, a significant result indicates the existence of multivariate structure which might include multivariate clustering. Further, this permutation approach cannot detect univariate clustering. Several programs have recently become available. A Monte Carlo study examines the *DISPROF* function in the *Fathom toolbox* in *Matlab*. Type I and Type II error rates as well as clustering accuracy are reported for a variety of conditions that are relevant for strategic groups research. Under favourable conditions, this program is remarkably powerful, but under less favourable conditions, the results are horribly misleading. A permutation test should not be used alone. A multimethod approach is proposed that exploits the complementarity of a permutation test and a Monte Carlo test; the weaknesses of one correspond to the strengths of the other. Indeed, the Monte Carlo test specifically rules out the most troubling source of Type I errors for the permutation test.

Key words: Cluster Analysis, Permutation Test, Significance Testing, Strategic Groups, Monte Carlo Study

1. STRATEGIC GROUPS AND CLUSTERING

A basic insight from economics is that the nature of competition among rival firms depends on the number of firms involved. Simply stated, in an industry with many homogeneous firms, the interactions would presumably lean towards perfect competition. The notion of strategic groups is intriguing in that the industry is divided into groups such that strategies are homogeneous within each group and heterogeneous across groups (McGee and Thomas, 1986; Porter, 1979). Highly similar firms would be predisposed to targeting similar suppliers to attain the resources needed to fuel their common strategy; they would also be predisposed to targeting similar buyers for their products and services. Thus, firms within a strategic group would be more likely to influence each other's performance than firms belonging to different groups.

Group members are likely to interact with each other using a mixture of cooperative and competitive behaviours in an effort to manage their interdependence (Brandenburger and Nalebuff, 1996), and the intensity of those interactions should be stronger within groups than between groups. Hence, it is possible that rivalry within the industry could resemble pockets of oligopolistic competition rather than industrywide perfect competition. That is, strategic groups could reflect the structure of rivalry within the industry (Tang and Thomas, 1992).

Further, if these groups differ in terms of the intensity of competitive and/or collusive interactions among their members, then the performance of those groups could differ in due solely to the way the firms interact within those groups. These differences in performance are referred to as true *group-effects* in that they result from the dynamics within the groups. We refer to this line of reasoning as the *interdependent view* of strategic groups to reflect the bonds connecting the firms within these coherent groups.

The central challenge within strategic groups research is to demonstrate that firms actually do form distinct groups within an industry. Cluster analysis (more specifically, hierarchical cluster analysis using Ward's method) has been the most popular

method for identifying strategic groups, but it traditionally lacks a test for significant clustering. The algorithm always reports groups that minimize the total within-group variance. Unfortunately, it is never clear if the analysis has discovered existing groups or merely created groupings by arbitrarily partitioning a set of relatively homogeneous firms.

Hatten and Hatten (1987) reject the interdependent view because a traditional cluster analysis simply cannot determine whether discrete groups actually exist. They argue that the identified clusters reflect artificially imposed groupings that are handy as an analytic convenience. Researchers can aggregate firm-level data to larger categories or types of firms within the industry. In this view, grouping firms into strategic groups is analogous to grouping customers into market segments based on similar demographic characteristics. There is no reason to assume that the members of these statistical groupings would interact as a cohesive social group. Hence, we refer to this as the *independent view* of strategic groups to reflect the lack of cohesive bonds between the members within each grouping.

A test for significant clustering is needed to determine if a cluster analysis has actually discovered separate strategic groups (the interdependent view) or simply created groupings from a set of relatively homogeneous firms (the independent view). Barney and Hoskisson (1990) argue that the inability to distinguish between these two conditions is so profound that it undermines the validity of the entire field of research.

Fortunately, researchers in ecology and oceanography have been expanding the use of computationally intensive techniques. In particular, Clarke, Somerfield and Gorley (2008) demonstrate a permutation test for significant clustering that could be quite useful. However, they did allude to a few problems associated with this technique. In spite of these weaknesses, several programs have recently become (commercially) available for significance testing within cluster analysis. In this paper, we explore whether these newly available programs would be appropriate for use in strategic groups research. A Monte Carlo study is performed on one of the programs (the *DISPROF* function in the *Fathom Toolbox* within *Matlab*) to determine how well it performs under a variety of relevant conditions.

2. A PERMUTATION TEST FOR CLUSTERING

After decades of effort, the problem of mathematically deriving a significance test for cluster analysis appears to be intractable (McKelvey, 1982). A pragmatic search for alternative approaches has led to (limited) interest in developing computationally intensive significance tests based on permutation and Monte Carlo techniques. While these discussions date back at least to the late 1960s, McKelvey (1982, p. 427) stated that those early efforts “do not offer much that is either well accepted, well worked out, or specific. The problem of statistical substantiation is largely unsolved at this time.”

One factor inhibiting the wide-spread acceptance of permutation tests is the rather unsatisfying operationalization of the null hypothesis. In general, the goal of a permutation tests is to preserve all the known and unknown characteristics of the data while ensuring that the phenomenon of interest (clustering) is reduced to random noise. The analysis is repeated, say, 999 times to generate the *null distribution* for that clustering statistic when it is known that no clustering exists. Notably, this null distribution also includes all the statistical artefacts that arise from the unique set of quirks in the given data. In principle, this should automatically cancel out the impact of those statistical artefacts on the significance test; the likelihood of obtaining the observed value of the statistic is empirically conditioned on the specific data at hand. If the observed value of the clustering statistic is an outlier (in this case, an extremely low value) when compared to the null distribution, then the null hypothesis that there is no clustering in the data would be rejected and researchers could infer that firms cluster together more than would be expected by chance.

This approach works beautifully for many techniques based on the general linear model because it is relatively easy to separate the sets of independent and dependent variables from each other while preserving the properties of both sets. For example, Carroll (2006) used a permutation test to extend the use of canonical correlation analysis (the most general form of the general linear model) to network analysis applications.

Unfortunately, cluster analysis poses an unusually difficult challenge for permutation tests. The permutation method is based on Legendre and Legendre’s (1998) Model 1, environmental control over individual species. Note that they describe the technique for permuting datasets in which the rows represent variables and the columns represent observations (the convention used in biology). We have switched the references here to conform to the conventions used in business research.

When attempting to wipe out any clustering among firms, the closest approximation is obtained by permuting every variable independently. That is, the values in each column are reshuffled (permuted). This perfectly preserves the characteristics of each variable. However, instead of removing any clusters, this permutation technique removes any *multivariate structure* from

the data. This would include *multivariate clustering*, but it would also include any other multivariate characteristics, such as correlations among variables. Thus, a significant finding indicates that it is unlikely that the observed value of the clustering statistic could be generated from the given data in the absence of multivariate structure. That is, it provides support for the existence of multivariate structure in general, but it is not clear whether that specifically includes multivariate clustering.

A related problem with this permutation method is that it is blind to *univariate clustering*. The permutation technique perfectly preserves the characteristics of each variable in every permutation. Typically, this would be one of the desirable features of a permutation test. Unfortunately, in cluster analysis the impact of univariate clustering on the test statistic would be identical in every permutation used to generate the null distribution. This cancels out the impact of univariate clustering the same way it cancels out other statistical artefacts. That is, the observed value of the clustering statistic would not appear unlikely when compared to the null distribution because the latter was generated with exactly the same effects.

While a permutation test appears to have some flaws, the question is, could it still be useful? Researchers in ecology and oceanography have discussed these details and illustrated the use of permutation tests (Clarke et al, 2008; Legendre and Legendre, 1998; Somerfield and Clarke, 2013). In particular, Clarke et al (2008) used a similarity profile (SIMPROF) analysis to assess the degree of clustering. That paper has inspired the development of a wave of new programs that recently hit the market: *PRIMER* (Clarke and Gorley, 2015), *Clustsig* (Whitaker and Christman, 2014) and *Matlab* (Jones, 2015).

The permutation tests appear to be gaining traction in biology despite the problems noted above. Our goal in this paper is to run a Monte Carlo study to assess the performance of these new tests to determine if they would they be useful for strategic groups research? All of these programs use similarity profiles (SIMPROF) for similarity measures and/or dissimilarity profiles (DISPROF) for distance measures. We arbitrarily selected the *DISPROF* function in the *Fathom Toolbox* within *Matlab* (Jones, 2015) for this assessment. Since it operates on distance measures, the remainder of the paper will focus on dissimilarity profiles for the sake of consistency.

2.1. Dissimilarity Profiles and Clustering

A dissimilarity profile is generated from the upper (or lower) triangle of a symmetric actor-by-actor matrix containing the distance between all possible pairs of firms. The values are stripped off as a vector, sorted by their absolute value, and then plotted. Figure 1 illustrates this for a sample of 105 firms; the distance between the 5,460 possible pairs of firms are sorted in ascending order and plotted to create these elegant curves. The solid grey line reflects the mean values at each point in the rank ordering using permuted data, and the dashed grey lines reflect the 99 percent confidence intervals at each point.

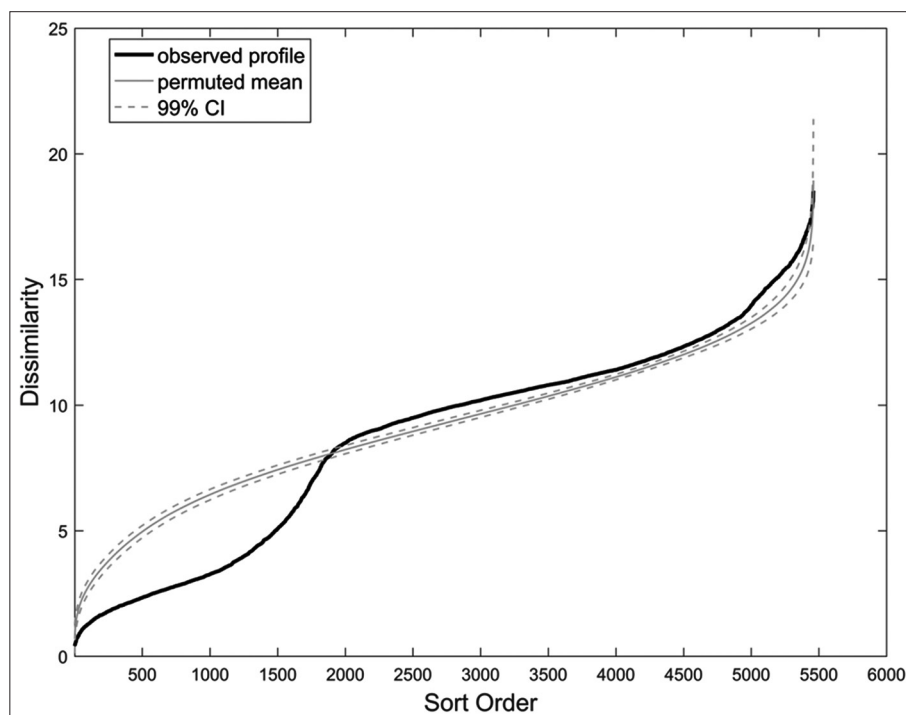


Figure 1: Dissimilarity Profile using four clustered variables (n = 105)

Clustering implies that subsets of firms clump together (creating short distances within groups) and also that those clumps are clearly separated from each other (creating long distances between groups). The solid black line in Figure 1 illustrates the profile generated from the original data when clustering is present.

This figure generated by the DISPROF function in Matlab is perhaps a bit misleading in that it suggests that significance is assessed at every point in the sorted order by checking if the observed value (the black line) falls outside of the envelope of values generated under the null hypothesis (the dotted grey lines). However, Clarke et al (2008) point out that the values in an actor-by-actor dissimilarity matrix are highly interdependent. All of the distances in the i th row and the i th column are influenced by the positioning of the i th firm. Since the distances are not independent, a simple Bonferroni correction for the number of tests performed would not be appropriate.

To avoid the problem of multiple tests, Clarke et al (2008) create one global statistic, π , by taking the sum of the distance between the solid black line (the observed values) and the solid grey line (the mean values based on permuted data) at every point in the ranking. To determine if the obtained value of π is significant, a second round of 999 permutations is performed, and π is calculated for each new permutation. This generates a distribution of π . Comparing the observed value of π using the original data to the distribution of π generated from the 999 permutations provides an exact significance test for π .

This significance test for π is applied at each branching point in the dendrogram that was generated from the original data. Notably, an agglomerative clustering algorithm was used to build the dendrogram from the bottom up; it started with many separate firms and merged them together until it ended up with one large cluster containing all of those firms. In contrast, the significance test is divisive; it starts at the top of the completed dendrogram and splits large clusters into smaller ones if π is significant at that branching point. The algorithm moves further down the resulting branches. The method recalculates π at each point on the dendrogram using the subset of rows and columns from the dissimilarity matrix that correspond with the ever shrinking subset of firms in the cluster at that point. The algorithm continues moving down each branch until one of the two stopping criteria is met: the cluster contains only one firm (a singleton) or the test for π is not significant at that branching point. The algorithm abandons the branch at that point and begins testing any remaining branches of the dendrogram. Thus the search moves branchwise through the dendrogram, moving down one branch until the stopping criteria are met, and then moving to the next branch. The program then indicates the maximum number of significant clusters. Clarke et al (2008) note that researchers have some flexibility in the cluster solution that they select. Since the dendrogram is traversed branchwise from the top down, if a given branching point is significant, then all of the branching points higher on that branch are also significant. Thus, solutions with fewer clusters on that branch are also acceptable, but those with more clusters (lower on the given branch) are not.

Note that solutions do not draw a horizontal line through the dendrogram that divide it into significant and nonsignificant sections. The testing algorithm can move very deep on some branches and stop much higher on other branches. Thus, a solution identified from the dendrogram need not correspond with the order in which clusters were merged in the agglomeration schedule.

Since the significance test is applied to (many) branching points in the dendrogram, this process could involve a large number of significance tests resulting in a high family-wise error rate. Researchers can specify the p-value (α) used to test each branching point. Clarke et al (2008) recommend using a stringent value for α (say, .01 or .001) as an informal way to compensate for repeated testing. They argue that a more sophisticated correction for family-wise error rates is perhaps unwarranted. "Too pedantic an approach to significance levels here, however, seems counter-productive, given the likely decline in power of the test" (p. 59) as the algorithm moves further down each branch and tests ever smaller groups of firms. They argue that this makes the process "appropriately and fortuitously self-limiting" (p. 59).

The following Monte Carlo study examines how effective these suggestions are for managing Type I and Type II error rates. It also simulates a variety of conditions that are relevant for strategic groups research. The goal is to determine if the recent wave of available software would satisfy the long-standing demand in strategic groups research for tests of significant clustering.

3. SETUP FOR THE MONTE CARLO STUDY

The most popular method for identifying strategic groups is hierarchical cluster analysis using Ward's method (Ward, 1963) with Euclidean distance measures (Harrigan, 1985; Ketchen and Shook, 1996; Nath and Gruca, 1997; Scheibler and Schneider, 1985). This tends to find relatively compact hyperspheres (as opposed to long stringy clusters) which makes it more plausible

that the firms are direct rivals with many if not all of the other members of that group. Singletons (groups consisting of only one firm) are also relatively rare, which is desirable given that the goal is to capture group phenomena.

A permutation test is computationally intensive because it repeats the cluster analysis on a large number of permutations (say, 999) to generate the null distribution of the desired statistic. All those repetitions are performed on one sample (one trial). This Monte Carlo study evaluates the Type I and Type II error rates of the significance test by running a large number of these trials for each condition, and this study addresses many relevant conditions. Consequently, this Monte Carlo study only runs 100 trials per condition unless otherwise stated. This sacrifices a bit of precision in the results for each condition in favour of allowing a larger number of conditions to be tested.

In their examples, Clarke et al (2008) typically set α to .05 for each branching point in the dendrogram, but they recommend using a more stringent level of α (say, .001) if the test is likely to be repeated across a large number of branching points in the dendrogram. In this Monte Carlo study, we vary the input value for α in order to assess the impact on the overall Type I and Type II error rates.

4. CONDITIONS WITHOUT CLUSTERING

4.1. Uncorrelated Random Variables

To establish a baseline for the Type I error rate, a population dataset is created with no clustering and no other multivariate structure in it beyond that due to chance. The DISPROF algorithm starts at the top of the dendrogram (all firms are in one large “cluster”), and the initial test assesses if there is any clustering (or more accurately, any multivariate structure) at all in the data. The Type I error rate on this baseline condition is simply the number of false positives for that initial test. This should be roughly equal to the input value for α . Note that α does not need to be adjusted for multiple tests since the focus is only on the first test at the top of the dendrogram to determine how often the test incorrectly indicates that significant clustering exists in completely random data. Thus, the input level for α was .05.

Sample size was set to 100 firms. To ensure a bit more precision in assessing the Type I error rate, 1,000 samples (trials) were drawn from the simulated population ($N = 10,000$ firms). The four variables were uncorrelated ($r = 0.00$) and each had a random normal distribution (mean = 0, SD = 1). While 50 significant trials (5%) were expected, 62 trials were incorrectly flagged as having significant clustering. Thus, the baseline Type I error rate was slightly higher than expected.

4.2. Correlated Random Variables

The Type I error rate (reporting clusters when none exist) becomes more pronounced when the variables are correlated. To understand this, imagine that there are two variables that are highly correlated with each other, but otherwise have random uniform distributions. The points (representing firms) are evenly spread across the range of each variable. When plotted in 2-dimensional space, the points would be densely packed along a straight line, and the distance between adjacent points would be very short.

If those variables were then independently permuted, the densely packed points along that straight line would be scattered fairly evenly across that 2-dimensional space. This would reduce the number of firms that were extremely close together. The dissimilarity profile for the permuted data would have a shallower slope than that obtained from the original data, resulting in π being statistically significant.

This illustrates the problem with misspecifying the null hypothesis. While we wish the null hypothesis referred to the absence of clusters, it actually refers to the absence of multivariate structure. Technically speaking, the significance test would be responding correctly to the presence of correlations (multivariate structure), but for the purpose of cluster analysis, this is treated as a Type I error.

To test this, four variables (v1 to v4) are created with random normal distributions, and their correlations are varied to mimic common patterns observed in the field. In one condition, two pairs of variables (v1 with v2; v3 with v4) are moderately correlated ($r = .50$) and all other pairs are weakly correlated ($r = .20$). In the second condition, the respective correlations are stronger ($r = .80$ and $r = .30$). Again the significance level set for the tests at each branching point are varied ($\alpha = .01$ and $\alpha = .001$) to determine the impact that α has on the results.

The results in Table 1 indicate that when the correlations are moderate ($r = .50$ and $r = .20$) and a relatively stringent probability level ($\alpha = .001$) is used, this test reported significant clustering (two or more clusters) in 69% of the trials. When the probability level was relaxed a bit ($\alpha = .01$), the test reported significant clustering in every sample—a 100% error rate.

The Type I error rate becomes even more pronounced when the correlations are stronger ($r = .80$ and $r = .30$). Even when the more stringent significance level was entered ($\alpha = .001$), this test reported finding a minimum of 4 clusters in every sample. With the less stringent level ($\alpha = .01$), as many as 13 clusters were identified when it was known that no clustering existed. As expected, this permutation test is sensitive to types of multivariate structure other than just clustering. Given that the data used in strategic groups research typically include correlated variables, it is virtually guaranteed that this test would mislead researchers. This is a serious concern that must be addressed before these recently available programs can be used in the field.

5. CONDITIONS WITH CLUSTERING

Having established how this method behaves when no clustering exists, we now begin adding clusters to the data. This design systematically varies (a) the number of variables involved in clustering, (b) the sample size and (c) the significance parameter, α , that is used at each branching point. We start with the most difficult conditions and move gradually to the most favourable conditions to illustrate where this method struggles as well as what it is capable of doing under the right conditions. The performance of the permutation test is assessed in terms of Type I and Type II errors. The Jaccard index is also used to examine how often firms are correctly paired together in the same group (Milligan, 1981).

We start by creating clusters on one variable (the impossible condition), then two variables (the minimum required condition for this method to work), and finally on all four variables (the highly favourable condition). This progression (first the bad news, then the good news) is used to illustrate the critical role that multivariate structure plays in the performance of this permutation test.

5.1. Univariate Clusters

Univariate clustering is easy to see simply by looking the histograms of the variables. However, this permutation method cannot detect univariate clusters (Clarke et al, 2008). If univariate clusters exist, then the dissimilarity profile for the original data (see the black line in Figure 1) would be relatively steep. However, this univariate clustering would be perfectly preserved in every permutation. Assuming that there were no other effects in the data, the grey lines generated from the permuted data would show the same effect, thereby mimicking and enveloping the black line. Thus, Type II errors are inevitable; this permutation technique should be completely blind to univariate clustering.

Table 1: Number of clusters incorrectly identified (false positives) given four correlated variables with no clustering

| Correlations | Moderate ($r=0.50$ & 0.20) | | High ($r=0.80$ & 0.30) | |
|-------------------|--------------------------------|------|----------------------------|------|
| | 0.001 | 0.01 | 0.001 | 0.01 |
| Prob.(α) | | | | |
| Clusters | | | | |
| 1 | 31* | 0* | 0* | 0* |
| 2 | 49 | 45 | 0 | 0 |
| 3 | 15 | 25 | 0 | 0 |
| 4 | 5 | 13 | 4 | 0 |
| 5 | 0 | 15 | 19 | 2 |
| 6 | 0 | 1 | 23 | 10 |
| 7 | 0 | 1 | 20 | 15 |
| 8 | 0 | 0 | 19 | 21 |
| 9 | 0 | 0 | 12 | 21 |
| 10 | 0 | 0 | 2 | 17 |
| 11 | 0 | 0 | 1 | 12 |
| 12 | 0 | 0 | 0 | 1 |
| 13 | 0 | 0 | 0 | 1 |

* Frequency of finding the correct number of clusters in each condition

To confirm this, the clusters were made as easy to detect as possible. A much larger sample ($n = 300$ firms) was used to create three clusters (100 firms each). Again, the population ($N = 10,000$ firms) contains four uncorrelated variables ($v1$ to $v4$), each of which has a random normal distribution (mean = 0, SD = 1). Groups were created on one variable ($v1$) by subtracting 10 from the values of one group, adding 10 to the values of another group, and leaving the third group as is. Thus, the three group means on $v1$ were either 10 or 20 standard deviations away from each other. The significance test was based on 999 random permutations, and the Monte Carlo study repeated this significance test on 100 unique samples (trials).

As expected, this permutation method was not able to detect these univariate clusters in spite of the fact that they were made painfully clear. When α was set to .01, none of the 100 trials were flagged for significant clustering. When the analysis was rerun with a less conservative significance level ($\alpha = .05$), 6 of the 100 trials reported significant clusters (see Table 2). Notably, this is the same as the Type I error rate found in the baseline condition (random, uncorrelated variables with no clustering). This confirms the blind spot that Clarke et al (2008) warned about.

5.2. Multivariate clusters

In order to determine if this permutation test would be useful for research on strategic groups, the pattern of clusters (see Figure 2) was designed to reflect some of the relevant issues. Specifically, the number of firms in a group can influence the degree of competition and/or cooperation within that group. Small groups would facilitate oligopolistic competition while large groups would lean more towards perfect competition. Combining groups of different sizes in the same industry creates the possibility (but does not guarantee) that performance differences would exist across groups due to differences in the dynamics within those groups (true group-effects).

There are three large groups (A, B, and C) that represent mainstream strategic positions in an industry. The wide gaps between these large groups make them very easy to distinguish. Any analysis that fails to identify these three groups would be fairly useless. To systematically increase the level of difficulty, each mainstream position has a small, specialized niche position nearby (D, E, and F). The distance between the large group and the corresponding small group is varied (see the bold numbers in Table 3) so the difficulty of identifying the separate groups ranges from moderately difficult (A and D) to extremely difficult (C and F).

Since this permutation method is blind to univariate clustering, the centroids of the small clusters were positioned at roughly 45 degree angles from the centroids of the associated large clusters. This multivariate (rather than univariate) displacement should make the niche clusters easier to distinguish.

The number of firms within each group also influences the difficulty of identifying distinct groups. The large mainstream groups contain 10 firms in the harder condition and 30 firms in the easier condition. The small niche groups always contain 5

Table 2: Effects of the number of variables involved in clustering, the sample size and α on both the number of clusters found and the mean Jaccard index indicating assignment accuracy

| Clusters ^a | 1 variable | 2 variables | | | | All 4 variables | | | |
|-----------------------|------------|-------------|------|-------|------|-----------------|------|-------|------|
| | Firms (n) | 300 | 45 | | 105 | | 45 | | 105 |
| Prob. (α) | 0.05 | 0.001 | 0.01 | 0.001 | 0.01 | 0.001 | 0.01 | 0.001 | 0.01 |
| Clusters | | | | | | | | | |
| 1 | 94 | 50 | 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0* | 43 | 40 | 80 | 75 | 0 | 0 | 0 | 0 |
| 4 | 1 | 7 | 45 | 17 | 22 | 0 | 0 | 0 | 0 |
| 5 | 1 | 0 | 9 | 3 | 3 | 36 | 11 | 16 | 11 |
| 6 | 0 | 0* | 0* | 0* | 0* | 63* | 85* | 80* | 77* |
| 7 | 1 | 0 | 0 | 0 | 0 | 1 | 4 | 4 | 12 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Jaccard | 0.35 | 0.35 | 0.55 | 0.75 | 0.75 | 0.90 | 0.95 | 0.97 | 0.97 |

* Frequency of finding the correct number of clusters in each condition.

^aThe number of variables (out of 4) that contribute to clustering; remaining variables are random.

firms so they should be relatively difficult to detect. Thus, the total sample size is 45 firms (3x10 + 3x5 firms) in the difficult condition versus 105 firms (3x30 + 3x5 firms) in the easier condition. Thus, the test pattern shown in Figure 2 is implemented by positioning groups along the coordinates listed in Table 3 and the scatter plots in Figure 3 illustrate the appearance of those clusters given large and small sample sizes.

In this Monte Carlo study, firms are drawn from a simulated population (N = 10,000 firms) with four uncorrelated variables (v1 to v4), each of which has a random normal distribution (mean = 0, SD = 1). Initially, all groups have their centroids roughly at the origin in this four-dimensional space. The smaller groups are created by dividing the random normal variables by 2 (mean = 0, SD = 0.5), while the larger groups are left as is (mean = 0, SD = 1). This might make it slightly easier for an agglomerative cluster analysis to identify the niche groups in spite of the small number of firms, because these firms have been squeezed closer together in the process of reducing the standard deviation for the niche groups. Still, the scatter plots in Figure 3 look like plausible representations of the issues at hand. In this context, Type I errors refer to reporting more clusters than actually exist in the data. Type II errors occur when fewer clusters are reported than actually exist.

5.2.1. Two clustered variables and two noise variables

Clarke et al (2008) note that this permutation test requires at least two variables to be involved in the clustering, and that the power increases with the number of variables. In this condition, only two variables contribute to the clustering. This illustrates how the test performs given the minimum required conditions. The x-coordinates (see Table 3) are added to one variable (v1) and the y-coordinates are added to another variable (v2) in order to create the cluster pattern illustrated in Figure 2. Variables v3 and v4 are left as random normal variables (mean = 0, SD = 1) with no clustering.

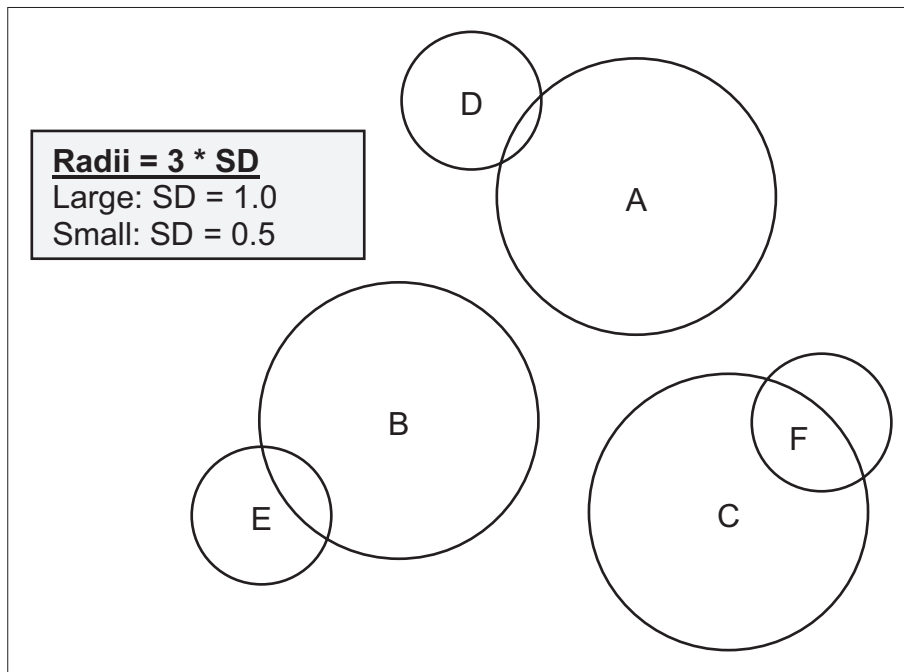


Figure 2: Positions of the clusters

Table 3: Coordinates of centroids and the distance between them

| Group | Radius ^a | x | y | Distance between centroids | | | | | | |
|-------|---------------------|------|----|----------------------------|-------|------|-------|-------|-------|---|
| A | 3.0 | 1 | 3 | A | | | | | | |
| B | 3.0 | -4 | -2 | B | 7.07 | | | | | |
| C | 3.0 | 3 | -4 | C | 7.28 | 7.28 | | | | |
| D | 1.5 | -2.5 | 5 | D | 4.03 | 7.16 | 10.55 | | | |
| E | 1.5 | -7 | -4 | E | 10.63 | 3.61 | 10.00 | 10.06 | | |
| F | 1.5 | 5 | -2 | F | 6.40 | 9.00 | 2.83 | 10.26 | 12.17 | |
| | | | | Group | A | B | C | D | E | F |

^aThe radius of each circle is three times the standard deviation of that group

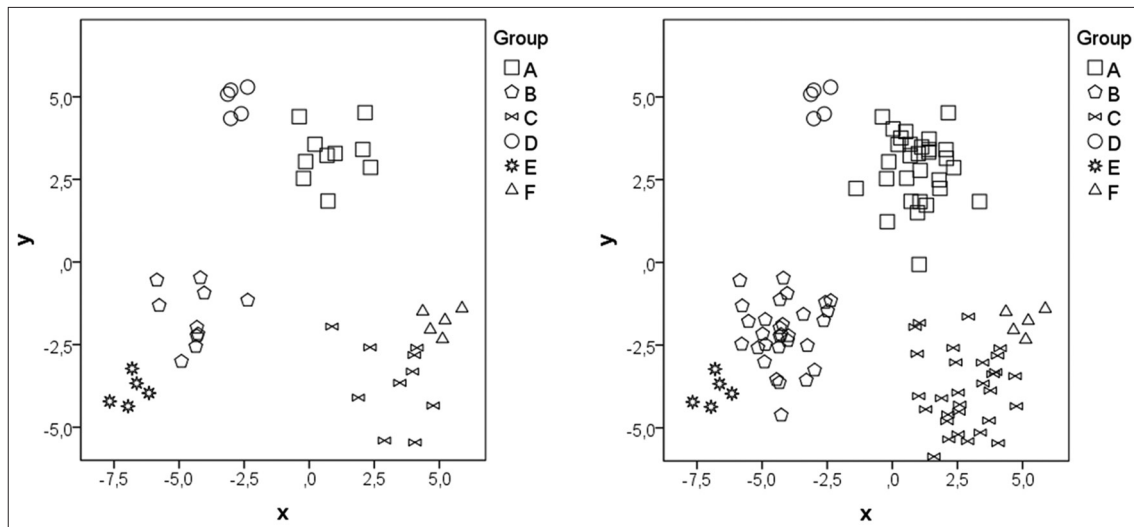


Figure 3: Scatter plot of clusters: n = 45 (left), n = 105 (right)

The most challenging variation in this condition had a small sample size (n = 45) and a stringent significance level ($\alpha = .001$) for testing each branching point in the dendrogram. Under these difficult conditions, half of the trials failed to reject the null hypothesis (see Table 2). Even the most obvious mainstream groups were not significant in half of the trials. In the other half of the trials, the test identified the three large groups and occasionally the most distinct niche group (D).

Clarke et al (2008) suggest relaxing the significance level a bit when dealing with small sample sizes or when more sensitivity is needed to detect small groups. Setting α to .01 did indeed improve the performance. The program identified the three largest groups almost every time and occasionally found one or two of the smaller niche groups as well. Relaxing α made it possible to identify more of the existing clusters in the dendrogram, thereby allowing the accuracy on the Jaccard index to jump from 35% to 55%. Although it was not tested, it seems plausible that relaxing α even further to .05, as Clarke et al (2008) did in their examples, would be helpful under the minimalistic conditions involving a small sample and only two variables involved in clustering.

Increasing the sample size (n = 105) also led to marked improvements in performance. The three large groups were identified in all of the trials and occasionally one or two of the smaller groups were also found. The accuracy on the Jaccard index jumped to 75%.

Admittedly, this DISPROF permutation test in Matlab struggles under these challenging conditions. The clusters only involve two variables (the weakest possible condition). Looking at the scatter plots in Figure 3, it is not surprising that the small niche groups positioned very close to the larger groups are not detected. Of course, strategic groups in practice might not be positioned so close together. Indeed, groups that are too close together might not be sufficiently isolated from neighbouring groups to avoid the spread of cut-throat competition. Still, the results under this condition are disconcerting for strategic groups applications, because it is the juxtaposition of small and large groups (reflecting oligopolistic versus perfect competition, respectively) that gives rise to the aforementioned true group-effects in performance.

5.2.2. Four clustered variables

The preceding (unfavourable) conditions were intentionally designed to demonstrate the weaknesses of this technique—the *lower bounds*, if you will. Having exposed the major flaws, it is time to see what the program can do under highly favourable conditions—loosely speaking, the *upper bounds* of performance for this test. In this condition, all four variables contribute to the clustering. For simplicity, two variables (v1 and v2) are shifted along the x-axis, while the remaining two variables (v3 and v4) are shifted along the y-axis (see Table 3) to displace the centroids of the groups as described earlier.

While the four variables were initially uncorrelated, the displacement of each group mean is large compared to its standard deviation. This displacement creates strong correlations ($r > .90$) between v1 and v2 as well as between v3 and v4, while all other correlations remain close to zero. Essentially, this creates two orthogonal factors (factor-x and factor-y), and this multivariate structure corresponds perfectly with the formation of the six clusters.

This condition reveals that multivariate structure is a double-edged sword. Earlier conditions demonstrated that the presence of correlations without clustering caused the Type I error rate to explode (see Table 1). In sharp contrast, the correlations in this condition correspond directly with the clustering. This resulted in remarkably good performance (see Table 2). In three of the four variations on this condition, the correct number of clusters (six) was identified in the vast majority of the trials, and the accuracy in pairing firms together in the correct groups was nearly perfect. The relatively difficult variation ($n = 45$, $\alpha = .001$) had the poorest performance. Still, the most challenging cluster (F) was detected in nearly two thirds of the trials and the accuracy on the Jaccard index only dropped to 90%.

In fact, the power and precision of this technique was so good that the initial cluster patterns proved to be too easy. Initially, groups A and D were the closest pair of groups. The distance between the closest pair of large and small clusters had to be reduced twice (first to that between B and E, and finally to that between C and F) before the cluster pattern became challenging. Frankly, when examining the distance between groups C and F in Figure 3, if they were any closer together, it would be hard to justify calling them two separate groups for the purposes of strategic groups research. Pockets of oligopolistic competition can only exist if the groups are separated (isolated, quarantined for their own protection) from the rest of the industry. Thus, given the right conditions, this test has more than enough power and precision to make fine distinctions. It is capable of identifying small, niche strategic groups even if they are positioned quite close to larger, mainstream strategic groups.

6. DISCUSSION

Given the critical need in strategic groups research for a test of significant clustering, this permutation test merits careful consideration. It is encouraging that this permutation test is extremely powerful and precise under highly favourable conditions, but the test is also terribly misleading under less favourable conditions. The question is whether it would do more harm than good.

Ideally, a permutation test would empirically determine the distribution of the test statistic when it is known that clustering does not exist, and this null distribution would incorporate all the statistical artefacts due to any remaining idiosyncrasies in the data. If the observed value of the statistic were extreme compared to the null distribution, then presumably the difference would be due to clustering.

Unfortunately, the permutation technique cannot isolate the effects of clustering. This technique perfectly preserves the characteristics in each variable when generating the null distribution. This includes univariate clustering. Consequently, the observed value of the clustering statistic should not appear extreme when compared to a null distribution that was generated from permutations containing exactly the same effects. Consequently, this permutation technique cannot detect *univariate clustering*.

Researchers could compensate for this blind spot by applying a separate test for multimodal distributions. Xu, Bedrick, Hanson, & Restrepo (2014) compare four methods and recommend a Bayesian mixture model in general. However, they noted that under certain conditions, other techniques perform better. While the availability of such tests is reassuring, it should be pointed out that the multimodal distributions in their study resembled overlapping clusters rather than the discrete, clearly separated clusters that are associated with the interdependent view and true group-effects. Hence, the tests of multimodal distributions are still less than ideal for assessing univariate clustering.

Multivariate clustering also poses some difficult challenges. Instead of specifically removing multivariate clustering from the permuted data, this technique has a much broader impact in that it removes any *multivariate structure*. Hence, if the observed value of a statistic is extreme compared to the null distribution, then the findings support the existence of multivariate structure, which *might* involve multivariate clustering. Consequently, when no clusters exist, the presence of correlated variables can cause the Type I error rate to explode (see Table 1). Most notably, moderately correlated variables incorrectly led to significant results in every trial even though no clusters were present. However, if those correlations were created by the positioning of distinct clusters, then both Type I and Type II error rates drop dramatically and the results are remarkably precise (see the right side of Table 2).

Clarke et al (2008) discuss a duality between *clustered observations* and *correlated variables*. They argue that clustering the observations creates correlations among the variables, but correlations among variables do not necessarily create clustering of observations. While it is very difficult to isolate the impact of clustering, it is relatively easy to isolate the impact of correlations. For instance, an alternative null distribution could be generated in a Monte Carlo approach using *copulas* to replicate the correlation matrix from the original data (Mair, Satorra, and Bentler, 2012). This would generate the null distribution for the

statistic given the observed correlation matrix. A significant result would indicate that it is unlikely that the observed value of the clustering statistic would have been generated by that pattern of correlations without clustering. This would rule out the major threat to the validity of the permutation test that is so clearly identified in Table 1. Thus, the permutation test is promising, but it should not be used alone. Our ongoing research attempts to extend these secondary analyses to create a multimethod approach.

6.1. A multimethod Approach

Given the historical difficulty in developing tests for significant clustering, it is perhaps overly optimistic to expect to develop a single technique that performs well under all conditions. It might be more fruitful to apply several different tests of significant clustering in a multimethod approach. Every test has weaknesses; a multimethod approach tries to combine techniques that do not share the same weaknesses (Brewer and Hunter, 2006).

To that end, we are currently testing the performance of a Monte Carlo test for significant clustering that expands on the analyses noted above. This makes use of the autofit function in the *SIMPLAN* procedure in *SPSS* (IBM Corporation, 2016). Rather than requiring a separate test for univariate clustering, the proposed Monte Carlo test removes the effects of univariate clustering from the null distribution—something that the permutation test could not do. This is achieved using the Anderson-Darling test for goodness of fit to determine which unimodal probability distributions best describes each input variable. If univariate clustering (a multimodal distribution) is found, the function tweaks the parameters of several unimodal distributions to obtain the best fit to the original histogram with the exception that the peaks and valleys associated with univariate clustering are levelled out.

Further, the autofit function in *SIMPLAN* replicates the original correlation matrix in the simulated data (similar to the use of copulas noted above) while ignoring any multivariate clustering. This approach attempts to tease apart the troubling duality between clustered observations and correlated variables.

So on one hand, the Monte Carlo test creates artificial data that matches the original data as closely as possible except that there is *no univariate or multivariate clustering* in the data. This is a much better operationalization of the null hypothesis than the one imposed by the permutation test (that there is *no multivariate structure* in the data). On the other hand, the Achilles' heel of a Monte Carlo test is the risk that some idiosyncrasies of the observed data are overlooked when generating the simulated data, and those idiosyncrasies could have a dramatic impact on the clustering statistic. This weakness corresponds with the strength of the permutation test which avoids this risk by working with the original data and carefully preserving the characteristics of each variable. So the two methods seem to complement each other nicely. If both techniques yield similar results, that would support for the convergent validity of the two significance tests and the results of the cluster analysis could be accepted with more confidence (Brewer and Hunter, 2006).

7. CONCLUSION

Research on strategic groups has endured decades of harsh criticism due in large part to the lack of significance testing for cluster analysis. A test is needed to determine if firms in a given industry are clumped together (potentially forming pockets of oligopolistic competition) or are more evenly scattered (resulting in industrywide perfect competition).

Clarke et al (2008) introduced a permutation test for cluster analysis based on the dissimilarity profile describing the distances between all possible pairs of firms. They alluded to problems related to the inability to (a) detect univariate clustering and (b) distinguishing between multivariate clustering versus other forms of multivariate structure. This Monte Carlo study attempts to quantify the severity of these problems. The results reveal that the test is wildly misleading under unfavourable conditions, but it is also remarkably powerful and precise under favourable conditions. Thus, the test might be useful, but it should not be used alone.

A multimethod approach could provide a pragmatic way forward. A Monte Carlo test for significant clustering would complement a permutation test nicely, since the weaknesses of one correspond with the strengths of the other. In fact, the proposed Monte Carlo test provides a validity check in that a significant result rules out the most troubling source of Type I errors for the permutation test—correlated variables without clustering (see Table 1). It is hoped that a multimethod approach will provide a pragmatic way forward for research on strategic groups.

While adding significance tests to cluster analysis should make a substantial contribution to this field, it is important to keep in mind that significance tests say nothing about the effect size (Kline, 2004). For instance, it is not clear how much distance

would be required between groups to prevent cut-throat competition from spreading between them. Similarly, these tests do not guarantee the substantive importance of the findings. Obtaining a significant result does not mean that the observed clusters must possess all of the characteristics associated with the interdependent view. These tests of significant clustering only indicate that firms cluster together more than would be expected by chance. Additional analyses would be required to determine if the groups possess specific hypothesized properties. Even so, the addition of such significance tests would still meet the pent-up demand that has existed among researchers for decades. In particular, it would remove the long-standing criticism from Hatten and Hatten (1987) and Barney and Hoskisson (1990) by finally making it possible to determine if discrete strategic groups exist in the given industry.

REFERENCES

- Barney, J. B., & Hoskisson, R. E. (1990). Strategic groups: Untested assertions and research proposals. *Managerial and Decision Economics*, 11, 187-198.
- Brewer, J., & Hunter, A. (2006). *Foundations of Multimethod Research: Synthesizing Styles*. Thousand Oaks, CA: SAGE Publications.
- Brandenburger, A. M., & Nalebuff, B. J. (1996). *Co-opetition*. New York: Currency Doubleday.
- Carroll, C. (2006). Canonical correlation analysis: Assessing links between multiplex networks. *Social Networks*, 28, 310-330.
- Clarke, K. R., & Gorley, R. N. (2015). PRIMER v7: User Manual/Tutorial [Computer software and manual]. Plymouth, UK: PRIMER-E. Retrieved August 15, 2015, from <http://www.primer-e.com/downloads.htm>
- Clarke, K. R., Somerfield, P. J., & Gorley, R. N. (2008). Testing null hypotheses in exploratory community analyses: Similarity profiles and biota-environmental linkage. *Journal of Experimental Marine Biology and Ecology*, 366, 56-69.
- Harrigan, K. R. (1985). An application of clustering for strategic group analysis. *Strategic Management Journal*, 6, 55-73.
- Hatten, K. J., & Hatten, M. L. (1987). Strategic groups, asymmetrical mobility barriers and contestability. *Strategic Management Journal*, 8(4), 329-342.
- IBM Corporation (2016). IBM SPSS Statistics 24 Command Syntax Reference [Manual]. Retrieved January 17, 2017 from ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/24.0/en/server/Manuals/IBM_SPSS_Statistics_Command_Syntax_Reference.pdf
- Jones, D. L. (2015). Fathom Toolbox for Matlab: software for multivariate ecological and oceanographic data analysis [Computer software]. St. Petersburg, FL: College of Marine Science, University of South Florida. Retrieved August 15, 2015, from <http://www.marine.usf.edu/user/djones/>
- Ketchen, D. J., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, 17(6), 441-458.
- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, D.C.: APA Books.
- Legendre, P., & Legendre, L. (1998). *Numerical ecology* (2nd English ed.). Amsterdam: Elsevier Science.
- Mair, P., Satorra, A., & Bentler, P. M. (2012). Generating nonnormal multivariate data using copulas: Applications to SEM. *Multivariate Behavioral Research*, 47(4), 547-565. <https://doi.org/10.1080/00273171.2012.692629>
- McGee, J., & Thomas, H. (1986). Strategic groups: Theory, research and taxonomy. *Strategic Management Journal*, 7, 141-160.
- McKelvey, B. (1982). *Organizational systematics: Taxonomy, evolution, classification*. Berkeley: University of California Press.
- Milligan, G. W. (1981). A Monte Carlo study of thirty internal criterion measures for cluster analysis. *Psychometrika*, 46(2), 187-199.
- Nath, D., & Gruca, T. S. (1997). Convergence across alternative methods for forming strategic groups. *Strategic Management Journal*, 18(9), 745-760.
- Porter, M.E. (1979). The structure within industries and companies' performance. *Review of Economics and Statistics*, 61(2), 214-227.
- Scheibler, D., & Schneider, W. (1985). Monte Carlo tests of the accuracy of cluster analysis algorithms: A comparison of hierarchical and nonhierarchical methods. *Multivariate Behavioural Research*, 20, 283-304.
- Somerfield, P. J., & Clarke, K. R. (2013). Inverse analysis in non-parametric multivariate analyses: Distinguishing groups of associated species which covary coherently across samples. *Journal of Experimental Marine Biology and Ecology*, 449, 261-273.
- Tang, M.-J., & Thomas, H. (1992). The concept of strategic groups: Theoretical construct or analytical convenience. *Managerial and Decision Economics*, 13(4), 323-329.
- Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236-244.
- Whitaker, D., & Christman, M. (2014). Clustsig: Significant cluster analysis [Computer software and manual]. Retrieved August 15, 2015, from <https://cran.r-project.org/web/packages/clustsig/index.html>
- Xu, L., Bedrick, E. J., Hanson, T., & Restrepo, C. (2014). A comparison of statistical tools for identifying modality in body mass distributions. *Journal of Data Science*, 12, 175-196.