Journal of Classification 16:63-89 (1999)

A Generalized Rand-Index Method for Consensus Clustering of Separate Partitions of the Same Data Base

Abba M. Krieger
University of Pennsylvania

Paul E. Green
University of Pennsylvania

Abstract: One of the recent trends in industry-based cluster analysis, especially in marketing, is the development of different partitions (e.g., needs-based, psychographics, brand choice, etc.) of the same set of individuals. Such individualized clusterings are often designed to serve different objectives. Frequently, however, one would also like to amalgamate the separate clusterings into a single partition — one that parsimoniously captures commonalities among the contributory partitions. In short, the problem entails finding a consensus partition of T clusters, based on J distinct, contributory partitions (or, equivalently, J polytomous attributes). We describe a new model/algorithm for implementing this objective. The method's objective function incorporates a modified Rand measure, both in initial cluster selection and in subsequent refinement of the starting partition. The method is applied to both synthetic and real data. The performance of the proposed model is compared to latent class analysis of the same data set.

Keywords: Consensus clustering; Categorical variable clustering; Latent class analysis; Hubert-Arabie modified Rand index.

1. Introduction

In many applications of cluster analysis, it is not unusual for the analyst to construct alternative partitions using the same data set. These separate

---

Authors' address: Abba M. Krieger and Paul E. Green, Suite 1450, Dietrich Hall, University of Pennsylvania, Philadelphia, PA 19104, USA;
email: abba@stat.wharton.upenn.edu

partitions reflect different viewpoints of how, for example, the classifications are to be used by a firm. To illustrate, in the banking industry, multiple classifications of the same customer base could entail different sets of variables, such as account assortments and balances, psychographic scores, attitudes toward risk taking, and so on.

An amalgamated clustering is a type of consensus method in cluster analysis. As the name suggests, consensus methods extract commonalities across multiple classification trees (and, less frequently, partitions) of the same objects. (By "partition" we mean a set of mutually exclusive and collectively exhaustive classes, such that any object is in one and only one class.) An excellent survey of this research area can be found in Day (1986). Related articles by Margush and McMorris (1981), McMorris and Neumann (1983), Barthélemy, Leclerc, and Monjardet (1985). Neumann and Norton (1986) and Vach (1994) also describe various aspects of this methodology. DeSarbo, Carroll, Clark, and Green (1984) describe a procedure, called Synclus, by which separate data sets for the same individuals can be amalgamated into a single partition. Those authors' model derives both (original) variable weights and the k-means (MacQueen 1967) based segments. Synclus was one of the first models explicitly to incorporate separate batteries of variables for a weighted type of consensus analysis.

1.1 Earlier Research: Multiple Correspondence Analysis Followed by k-Means Clustering

Because separate partitionings of the same data base result in a set of nominal, multi-state attributes (one for each partitioning), an amalgamated segmentation is structurally similar to the clustering of a set of unordered categorical variables irrespective of whether the contributory variables are, themselves, cluster-based partitions. In such cases multiple correspondence analysis (MCA), also known as homogeneity analysis or dual scaling, has been proposed as a technique for finding a spatial representation of objects, e.g., individuals or households. The individuals' spatial coordinates can then be clustered. General background on these methods can be found in Greenacre (1984), Gifi (1990), Nishisato, (1980, 1984, 1993, 1996), Heiser (1981), Meulman (1982), and de Leeuw (1984).

As an early example of MCA, followed by the k-means clustering of respondent coordinates, Lebart, Morineau, and Warwick (1984, pp. 132-143) reported an application where data were available on 1,000 individuals, described by 17 nominal variables. The authors chose six of these variables (e.g., sex, age, education) with a total of 25 categories. Multiple correspondence analysis was used to obtain a joint space of the 1,000 individuals and each of the six active-variable categories in eight dimensions. The 1,000

individuals were then clustered by k-means, based on their 8-dimensional coordinates. Seven segments were obtained and later described using the full set of categorical variables. Other applications of MCA, followed by a clustering of objects in terms of their reduced-spaced coordinates, have been described by Green, Krieger, and Carroll (1987), Green, Schaffer, and Patterson (1988), van Buuren and Heiser (1989), and Nishisato (1984). There is no compelling theoretical rationale regarding the appropriateness of this two-stage procedure to clustering. Rather, MCA is used as a device to convert qualitative variables into continuous ones so that k-means clustering can be easily applied.

1.2 Later Research: Latent Class Modeling

More recently, latent class analysis and related methods have been applied to sets of polytomous variables. Contributions include research by DeSarbo and Cron (1988), DeSarbo, Wedel, and Ramaswamy (1992), Poulsen (1990), van der Pol and de Leeuw (1986), Dillon and Kumar (1994), Ramaswamy, Chatterjee, and Cohen (1996), Dillon, Madden, and Mulani (1983), and Dillon and Mulani (1989). Accordingly, we later compare a latent class analysis approach to our proposed model.

1.3 Proposed Method and Format

The current authors propose a new model (called SEGWAY) for developing a single clustering of a set of separate partitions, obtained from different subsets of variables appearing in a common data base. Alternatively, SEGWAY can be used to provide a single clustering where the input data are a set of unordered categorical variables. The procedure is based on a generalization of the Rand criterion measure (Hubert and Arabie 1985). The original Rand index was prepared as a way to measure agreement between two partitions, rather than as a clustering tool. The Rand measure is used here as a criterion function in both the initial selection of a higher-order segmentation and in its later refinement. Unless noted otherwise, references to the Rand measure assume that the Hubert and Arabie modified (to adjust for chance agreements between two partitions of the same data set) version has been computed.

We briefly describe the SEGWAY model and some of its properties. We next employ synthetic data analyses to illustrate the model's behavior over different characteristics of data sets. The model is then applied to a real data set consisting of 1975 individuals and four contributory partitions. The model is compared empirically with latent class analysis. We conclude the paper with a summary discussion. Appendix A describes the proposed

algorithm more formally. Appendix B provides details of the synthetic data simulations. Appendix C describes a formal relationship between the chi-square statistic and Rand.

2. The Amalgamated Clustering Model

To motivate subsequent discussion we assume that various clustering tools have been applied to various subsets of variables of the same data base. Each individual has been separately assigned to one and only one cluster in each of the separate partitions. (The number of clusters per selected base is allowed to vary across clusterings.) The summary matrix, consisting of I individuals and J partitions is denoted as

$$C_{ij}\,;\, i = 1,2,\ldots,I;\, j = 1,2,\ldots,J$$

where the j-th partition has M_j classes, indexed as $m = 1,2,\ldots,M_j$.

The problem is to create a new partition, called D, with T classes. We choose as our objective function the highly popular Rand measure, denoted as $R_j(D)$, to represent the degree of association between clustering j and a to-be-found, amalgamated clustering D. The objective function is written as:

$$V(D) = \sum_{j=1}^{J} W_j R_j(D) \tag{1}$$

where the W_j are researcher-supplied, nonnegative weights, such that $\sum_{j=1}^{J} W_j = 1.0$, and $R_j(D)$ is the Rand measure computed between the original j-th clustering and the new (amalgamated) clustering.

2.1 The Rand Measure

As Milligan and Cooper (1987) note, the Hubert-Arabie modified Rand measure has become the index of choice in comparing the agreement between two separate partitions of the same data set. This measure adjusts for chance agreement and is not restricted to comparing partitions with the same number of segments. Table 1 shows the logic underlying the basic Rand measure and the explicit formula for both the unadjusted index and the Hubert and Arabie modification. As noted, the Rand index compares agreement between pairs of entities (e.g., individuals) across two different partitions.

Complete independence between the two partitions yields a Rand index of essentially zero. Complete association yields an index of 1.0. We utilize the Hubert and Arabie index in the following ways:

Table 1
The Structure of the Modified Rand Measure

First Partition	Second Partition		
	Pair of Items in Same Cluster	Pair of Items in Different Clusters	Marginals
Pair of Items in Same Clusters	A	B	A + B
Pair of Items in Different Clusters	C	D	C + D
	A + C	B + D	N

Notes: Unadjusted Rand Index = (A + D)/N, where N = A + B + C + D = I(I - 1)/2, where I = number of individuals

Hubert and Arabie's Modified Rand Index:

$$\text{Rand(adj)} = \frac{N(A + D) - [(A + B)(A + C) + (C + D)(B + D)]}{N^2 - [(A + B)(A + C) + (C + D)(B + D)]}$$

1. A Rand measure (in conjunction with a greedy algorithm) can be used to find an initial amalgamated partition of the individual profiles (Stage 1).
2. A Rand measure is then used to reassign individuals across clusters after the starting partition has been found (Stage 2).
3. Alternatively, Stage 2 can be applied to an initial partition found by some other procedure (e.g., random assignment of individuals to clusters).

2.2 The Algorithm

Finding the partition D to maximize the Rand measure is NP hard. Accordingly, a two-stage algorithm has been developed. As noted above, Stage 1 consists of choosing a starting partition via one of several options. Stage 2 then considers the systematic reassignment of individuals to maximize $V(D)$.

Stage 1

Several methods are available for finding a starting partition of the individuals: (a) random assignment; (b) multiple correspondence analysis, followed by k-means clustering; (c) latent class analysis; and (d) a greedy heuristic (utilizing the Rand measure, as noted above).

We consider each procedure in turn:

1. Random Assignment: If we were to assign individuals to clusters randomly (i.e., ignoring the clustering data), we would expect $V(D)$ to be approximately zero. However, because in Stage 2 we iteratively reassign individuals to maximize $V(D)$, the purpose of the initial (random) assignment is to look for possible local optima. Because we can vary the initial random start, we can also get a sense for the prevalence of local optima and the sensitivity of the final solution to starting partitions.
2. Multiple correspondence analysis (MCA): As described earlier, one could apply MCA to the input matrix, consisting of individuals by concatenated initial partitions. The resulting analysis yields a set of coordinates for each individual. These profile data can then be clustered (e.g., via k-means) to obtain a "representative" starting partition that reflects association across the contributory partitions. (One problem with this approach is to determine the appropriate number of MCA dimensions to retain for the "person" clustering step.)
3. Latent class analysis: One can fit a traditional latent class model to the given J segmentations. Doing so implies that, given the latent class structure, within-class local independence applies to the resulting segments. Various methods such as the EM Algorithm (Ramaswamy, Chatterjee, and Cohen 1996), can be used to obtain the latent classes. The resulting latent class solution can also provide a starting partition for Stage 2.
4. Greedy heuristic: We can begin by creating a higher level clustering consisting of as many clusters as there are unique vectors of $C_i = (c_{i1}, c_{i2}, \ldots, c_{iJ})$ and then combine pairs of clusters to reduce the number of higher level clusters by one at each step. The pair chosen at each step is the one that maximizes $V(D)$. We stop when the procedure produces the desired number of clusters T that serve as the starting partition. (Further details of this greedy heuristic are described in Appendix A.)

Stage 2

At this point, we have an initial partition obtained by one of the methods suggested above. The problem now is to reassign individuals so as to maximize $V(D)$. We consider moving each individual from its current cluster to any of the other clusters (i.e., a singleton reassignment approach to combinatorial optimization). All $I(T-1)$ possible moves are considered and the one that maximizes $V(D)$ is chosen. This procedure is repeated until none of the $I(T-1)$ possible moves leads to an increase in $V(D)$.

This approach need not necessarily lead to a global optimum. Various constraints can also be imposed on the final solution. For example, constraints can be introduced to preclude finding a clustering that is identical to one of the contributory partitions. One can also input the minimal number of individuals that must appear in each class. These possibilities are illustrated later in the paper. Appendix A provides a detailed and more formal account of Stage 2 of the SEGWAY algorithm.

3. A Monte Carlo Simulation of SEGWAY Under Different Data Generation Conditions

Before presenting an empirical application of SEGWAY, it is appropriate to describe the model's behavior under various researcher-specified assumptions. We set up a group of initial conditions specifying the character of the contributory partitions that SEGWAY takes as input data.

3.1 Simulation Design

A Monte Carlo study was designed with the following features:

1. Six different partitions were generated, each with four clusterings. The first four partitions were used for calibration purposes and the last two for cross tabulation.
2. Partitions were generated according to a latent class model (Dillon and Kumar 1994).[1] Each of the four latent classes was assumed to have an equal probability of occurring. Each of the six partitions was initially generated independently, conditioned by latent class. The probability distribution of clusters varied, by latent class, according to a parameter ρ.
3. There were four sets of conditions, designed according to the following settings for (which, in turn, is defined mathematically in Appendix B).
   a. All $\rho = 0$ (i.e., independence);
   b. All $\rho = 1$ (weakly dependent);
   c. All $\rho = 2$ (strongly dependent);
   d. $\rho = 0$ for the first three clusterings and $\rho = 2$ for the last three clusterings (mixed).

1. Details of this procedure appear in Appendix B.

Table 2
Results of Monte Carlo Simulation, Comparing the Cross Validation of "True" Latent Class, SEGWAY, and the Data-Based Latent Class Algorithm (Cell Values are Hubert-Arabie Modified Rand Index Measures)

		Calibration Clusterings				Validation Clusterings	
		1	2	3	4	5	6
Independent	Theoretical LC	.273	.278	.181	.280	.274	.278
	Segway	.628	.282	.265	.287	.287	.279
	Data-Based LC	.439	.382	.548	.430	.278	.278
Weakly Dependent	Theoretical LC	.518	.530	.489	.535	.522	.513
	Segway	.647	.648	.731	.654	.507	.543
	Data-Based LC	.604	.602	.616	.658	.500	.520
Strongly Dependent	Theoretical LC	.583	.580	.597	.604	.586	.590
	Segway	.824	.799	.821	.799	.745	.758
	Data-Based LC	.746	.724	.711	.744	.639	.631
Mixed Dependence	Theoretical LC	.273	.278	.281	.604	.586	.590
	Segway	.314	.288	.298	.902	.672	.683
	Data-Based LC	.396	.358	.330	.608	.523	.528

4. Three higher-order clusterings were analyzed. The first higher-order clustering is simply the original latent classes that generated the data. The second is based on SEGWAY, using a random starting partition. The third higher-order clustering is based on empirically finding the probabilities that each individual belongs to each of the latent classes and then assigning each individual to the latent class with the highest posterior probability. We use a standard approach, based on the EM algorithm (Dempster, Laird, and Rubin 1977).

3.2 Simulation Results

Results of the simulation are shown in Table 2. All table entries represent Rand values. Each row of three calibration conditions consists of: (a) theoretical latent classes used to generate the data; (b) the SEGWAY solution based on the actual generated data; and (c) a latent class solution based on the actual generated data. Not surprisingly, the theoretical latent class condition performs worst, because it does not capitalize on the "observed" data.

Our primary interest is in the comparative performance of the three calibration conditions when cross validated with the simulation-generated holdout classes, designated as columns 5 and 6 in Table 2. Results differ by condition: (a) when the clusterings are generated independently ($\rho = 0$), all three calibration sets, of course, cross validate with columns 5 and 6 about the same; (b) however, as the dependence across latent classes increases, the

extent to which SEGWAY predicts the holdout clusterings (in terms of the hit ratio) improves, relative to the empirically-based latent class method;[2] (c) when $\rho = 2$ (high dependence), SEGWAY does particularly well, compared to latent class. These results are encouraging, because the data were *initially generated according to a latent class structure.*

Why does SEGWAY do so well? Under a special set of conditions (i.e., equal row marginals and equal column marginals), maximizing Rand is linearly related to maximizing chi square which, in turn, is more in keeping with the common prediction problem of maximizing the hit ratio in a holdout sample. Appendix C details the relationship between Rand and the chi-square statistic under the foregoing conditions.

Furthermore, it is not clear how the process of finding latent classes (via a maximum likelihood approach) relates to the objective of maximizing a hit ratio between calibration and holdout sample. A second reason is that latent class solutions obtain posterior probabilities that each object belongs to each latent class. Researchers typically assign objects to the *modal* class. Doing so is tantamount to choosing probability 1.0 for the modal class and 0.0 for all other classes. These reasons may, in part, explain the somewhat poorer performance of the empirically-based latent class alternative when the latent classes themselves exhibit higher dependence, as demonstrated in the present Monte Carlo simulation.

4. An Empirical Example: Clustering Options for Sport Utility Vehicles

We now turn to a (disguised) empirical application of the SEGWAY algorithm. Since the early 1990's, with the introduction of the Ford Motor Company's *Explorer*, sport utility vehicles have captured the public's fancy. Five of the largest sellers in 1996 were the Ford *Explorer*, Chevrolet *Blazer*, Jeep *Grand Cherokee*, Toyota *4Runner*, and Nissan *Pathfinder*.

Our disguised firm, Alpha, is planning on introducing a new sport utility vehicle. One of the issues confronting Alpha's management is the promotion and sale of four optional (at extra cost) features: electronic navigational system; CD player; rear TV monitor; and security system.

---

2. The "hit ratio" is defined on a square table summarizing the association between separate partitions of the same individuals. Given a specific ordering of the row set of clusters, the column set of cluster labels are permuted so as to maximize the trace (sum of the diagonal elements of the square matrix). The hit ratio is defined as the ratio of the trace over the total number of individuals.

4.1 Market Survey

In the Spring of 1996, Alpha initiated a survey of prospective buyers for the new sport utility vehicle. The survey was conducted at various dealerships throughout the U.S. Respondents were shown color photographs and descriptions of each extra-cost option in randomized order. For each such option the respondent was asked to indicate if he/she would purchase the option if it were offered at the stated price. In addition, respondents were asked a standard series of background questions. Information was obtained on six categorical background variables. A total of 1975 respondents supplied both option evaluations and background data. Table 3 shows a description of the six background attributes: home ownership; living area; working status; commuting-to-work method; marital status; and occupation.

Each response to the willingness-to-buy question yields an *à priori* market partitioning, based on a yes or no answer; hence, there are four *à priori* partitions, in total. Rand measures were first computed between all six pairs of willingness-to-buy responses. These are shown in Table 4. As noted from Table 4, the highest pairwise associations between partitions are Electronic navigation with CD player (.312) and Rear TV monitor with Security system (.235).

We next consider two questions related to applying the SEGWAY model. The proposed two-stage algorithm consists of, first, finding a starting partition, followed by reassignment of individuals to maximize $R(D)$. The SEGWAY algorithm provides only a local optimum, similar to other approaches, such as k-means and latent class analysis. Hence, we first consider the question of rational versus random starting partitions. We then examine, via Monte Carlo methods, the robustness of SEGWAY solutions to changes in random starting partitions.

4.2 Rational Versus Random Starts

We first apply the SEGWAY model, using the greedy start, to the 1975×4 matrix, consisting of the four separate input partitions, as described above. Illustratively, we seek three clusters. In this application, the number of possible option profiles is 2^4, or 16. Not surprisingly, all 16 profiles are present. The first row of Table 5 indicates that the Stage 1 greedy-heuristic solution is a local maximum; the Rand index could not be increased by moving respondents across clusters (in Stage 2). The cluster sizes of the SEGWAY partition are 1086, 767, and 122. As noted, the overall (amalgamated) Rand is .440. The highest Rand indexes between it and the four contributory partitions are .511 (Electronic navigation) and .632 (CD player).

Table 3
Description of the Six Background Attributes

Attribute	Levels
Home ownership	Own; rent
Living area	Urban; suburban; town; rural
Working status	Full-time; part-time; self-employed; retired; student; housewife; currently unemployed
Commuting method	Car; public transport; walking; work at home; car pool
Marital status	Married; single; single with partner; widowed; divorced/separated
Education	Some high school; high school grad; vocational; some college; college grad; post-graduate work

Table 4
Pairwise Hubert-Arabie Modified Rand Measures Based on Willingness-to-Buy Partitions

	Second Partition		
First Partition	CD Player	Rear TV monitor	Security system
Electronic navigation	.312	.126	.092
CD player		.153	.112
Rear TV monitor			.235

Table 5
Summary Hubert-Arabie Modified Rand Measures with Original and Amalgamated Partition for Each of Four Starting Configurations

	Modified Rand Between Original Partition and Amalgamated Partition					Amalgamated Rand
Starting Partition	1	2	3	4	No. of Iterations	
SEGWAY (greedy start)	.511	.632	.373	.244	0	.440
Multiple Correspondence Analysis	.402	.443	.549	.306	92	.425
Random start 1	.488	.655	.369	.249	1295	.440
Random start 2	.488	.655	.369	.249	1277	.440

Table 6
Summary of Percentage of Hits Between Each Pair of Solutions, Based on Four Different Starting Partitions

	Second Partition		
First Partition	MCA-based	Random 1 start	Random 2 start
SEGWAY (greedy start)	77.8	98.5	98.5
MCA-based start		76.3	76.3
Random 1 start			100.0

We next considered a "rational" starting configuration for stage 1. Multiple correspondence analysis was applied to the original 1975×4 matrix of the four nominal (yes/no) responses. A 3-dimensional solution for the respondent score matrix was obtained.[3] As noted in Table 5, 92 iterations were required in Stage 2 to reach a local optimum. The amalgamated Rand index, as shown in Table 5, is .425, which is slightly below that for the greedy heuristic start. The final cluster sizes from the MCA-based start are 1039, 560, and 376.

Next, two different randomly selected starting partitions were constructed with an essentially equal number of cases in each of the three clusters of the initial amalgamated partition. As Table 5 shows, the number of iterations required in Phase 2 of SEGWAY is 1295 for the first random start and 1277 for the second random start. Interestingly, the two random configurations converge to the same amalgamated solution. Cluster sizes are 1086, 797, and 92, for each of the two random-start solutions. A permutation of labels for the second partition indicates complete agreement between the two partitions, despite differences in starting configurations and number of iterations.

Hence, in this empirical data set all four solutions result in fairly similar amalgamated partitions. This conclusion was verified by constructing two-way cross tables for all pairs of final clusterings. First, the ordering for rows was fixed. Next, we permuted the columns to maximize the trace of the cross table; we then found the number of "hits" between each pair of clusterings. Table 6 shows summary results according to hit ratios.

As evinced in Table 6, the MCA-based start averages a hit ratio of 76.8 percent across the three remaining partitions. The average hit ratio for all paired clusterings of the remaining three starting configurations is 99.0 percent.

4.3 Sensitivity of SEGWAY Solution to Random Starting Partitions

Given the fact that SEGWAY obtains only a local optimum, one wonders if the reassignment part (i.e., Stage 2) of the computer program is "robust" to changes in starting partitions. To examine this question, a computer program was written to generate 100 random starting partitions of three

3. Although not shown, MCA solutions were also obtained for 1, 2, and 4 dimensions. The 3-dimensional solution provided the best cross-validation with the six background attributes of Table 2. This persons-by-scores matrix was then clustered (by k-means) to obtain a starting partition of individuals for refinement by Stage 2 of SEGWAY.

segments each. SEGWAY was then applied to each such random start, and amalgamated Rand measures were computed, similar to those shown in the last column of Table 5. (As noted in Table 5, the two explicit random starts each resulted in an amalgamated Rand of .440.) Apparently Stage 2 of SEGWAY is quite robust. The average higher-order Rand of the 100 simulated, random starting partitions was .440. Better still, the minimum amalgamated Rand was .4392, and the maximum was .4409. Clearly, for this data set, at least, Stage 2 of SEGWAY is quite insensitive to randomly determined starting partitions.

5. Comparing SEGWAY to Latent Class Analysis

As described earlier, the latent class model represents the principal competitor to SEGWAY in developing either higher-order clusterings or in partitioning multistate, categorical data. Accordingly, we now apply the traditional latent class model to the same empirical data set, consisting of the four dichotomous, contributory partitions. A latent class analysis program was written to implement this task. A random procedure was used to obtain an initial partition of three clusters. These clusters were then modified by a latent class, EM algorithm. A local optimum was reached after 38 iterations. The resulting three cluster sizes were 861, 796, and 318, respectively (compared to SEGWAY's cluster sizes of 1086, 767, and 122).

To provide a description of the closeness of SEGWAY (greedy start) and the latent class analysis solution, we found (after column permutation) that 74.9 percent of the individuals were clustered similarly. They differed in the sense that latent class analysis assigned about one-fifth of SEGWAY's largest cluster to SEGWAY's cluster 3. Other assignments were relatively close between the two methods.

5.1 External Comparisons

A more compelling empirical exercise is to compare SEGWAY and latent class analysis on exogenous variables not used in the internal analysis. As noted earlier, Table 3 lists six qualitative background variables for the same individuals in the survey. In industry studies it is not unusual to collect such background data to examine inter-profile differences in an effort to reach types of respondents who could behave differently across clusters. The original designers of this study selected background variables on the basis of their judged relationships to controllable variables (i.e., vehicle options). This external comparison provides a more level playing field for comparing SEGWAY with latent class analysis.

The first question is: which of the two clusterings is more highly related to the six ''holdout'' attributes (that played no role in developing the partitions)? To examine this question, we prepared cross-tabulations of SEGWAY (greedy start) and latent class, respectively, with each of the six background variables, in turn. We then computed the p-value associated with the χ^2 statistic of each cross table. Smaller p-values, of course, indicate higher association between clustering and background variable.

As we note in Table 7, the partition obtained from SEGWAY shows, on average, a better cross validation with the six background variables than that associated with latent class, namely, an average p-value of .058 for SEGWAY versus .118 for latent class. However, on a variable-by-variable basis, the results are generally very close. The large difference (most affecting the average) results primarily from SEGWAY's much better performance on the ''Commuting Method'' variable, which itself is non-significant.

5.2 Product Profiling

The second question of interest is: do the SEGWAY (greedy start) and latent class partitions also show different product profiles? To examine this question, we computed cluster profiles for SEGWAY (greedy start) and latent class analysis for each of the four willingness-to-buy questions. The profile results appear in Tables 8 and 9, respectively. Table 9 for the latent class solution shows somewhat different product profiles when compared to those of Table 8 for the SEGWAY (greedy start) solution, after permuting columns to maximize the trace of the original cross tabulation between the two cluster memberships. Clusters 1 and 2 in Table 9 seem to correspond roughly to their counterparts in Table 8. The largest difference between the two appears in cluster 3 (as noted earlier).

5.3 Further Description of the SEGWAY Segmentation

At this point, we focus on the SEGWAY solution (first row of Table 5), whose overall Rand index is .440. As shown earlier, the SEGWAY solution consisted of three clusters with respective sizes of 1,086, 767, and 122 respondents. Table 8 summarizes the profiles of ''yeses'' to the willingness-to-buy question for each cluster. As earlier observed from Table 8, cluster 1 evinces high interest (over 80 percent) in all of the four extra-cost options. Cluster 2 shows extremely low interest in the navigational system and CD player and only moderate interest in the two remaining options. Cluster 3 shows reasonably high interest in the first two options and relatively low interest in options 3 and 4.

Table 7
Cross Validation p-Values Obtained from Each Amalgamated Partition and the Six Background Variables

Partition	Background Variable						Average
	Home Ownership	Living area	Working status	Commuting method	Marital status	Occupation	
SEGWAY (greedy start)	.000	.002	.001	.344	.002	.000	.058
Latent class	.000	.001	.002	.691	.012	.000	.118

Table 8
SEGWAY (Greedy Start) Cluster Profiles for Vehicle Option Preferences

Clusters	Size	Percentage of Respondents Saying "Yes" to Willingness-to-Buy Question			
		Navigation	CD player	TV monitor	Security
1	1086	81%	91%	95%	84%
2	767	2	5	39	37
3	122	76	63	25	36
	1975				

Table 9
Latent Class Based Profiles for Vehicle Option Preferences

Latent Classes*	Size	Percentage of Yeses to Willingness-to-Buy Question			
		Navigation	CD player	TV monitor	Security
1	861	82%	88%	100%	100%
2	796	2	8	38	39
3	318	85	86	64	21
	1975				

* Permuted to maximize the trace in the cross table with the SEGWAY (greedy) solution

Table 10
SEGWAY (Greedy) Cluster Profiles for Selected Levels of Five Demographic Variables
(Table entries are the percentage of individuals in each cluster displaying column caption)

Clusters	Size	Percentages Involving: *				
		Home owner	Urban/ suburban dweller	Full-time employed	Married	College/ post-grad
1	1086	82%	75%	59	35%	(12)
2	767	(54)	(70)	63%	35%	15
3	122	78	72	(54)	(32%)	18
	1975					

* highest (lowest)

How do the SEGWAY-based clusters differ in terms of the background attributes? Table 10 compares the three SEGWAY clusters using five of the six demographic features. (The ''non-significant'' commuting method variable was excluded.) As noted from Table 10, SEGWAY cluster 1 is highest with respect to home ownership and urban/suburban dweller, but lowest with regard to college/post-graduate. Cluster 2 is highest with respect to full-time employment and lowest with respect to home owner and urban/suburban dweller. Cluster 3 is highest on college/post-grad and lowest on full-time employment and married. Hence, some differentiation with respect to background variables is also noted across the three clusters.

6. Introducing *à priori* Weights and Minimal Rand-Value Constraints

The SEGWAY model can incorporate user-supplied *à priori* weights and minimal Rand values (between the higher-order partition and the contributory partitions), at the user's discretion. One can also fix the minimal size of a higher-order cluster. We illustrate the application of constraints in the context of the empirical example. We consider three hypothetical conditions: (a) *à priori* weights of .5, .3, .1, and .1, respectively, for the four contributory partitions: Electronic navigational system, CD player, Rear TV monitor, and Security system; (b) a minimal Rand value of .3 between the higher-order partition and each of the four contributory partitions; and (c) a minimal cluster size of 300 respondents.

As a starting configuration, we first set up a randomly determined initial partition of the 1975 respondents with an essentially equal number of individuals in each of three clusters.

À Priori Weights

We first arbitrarily consider the *à priori* weights condition of .5, .3, .1, and .1, respectively, for the four contributory partitions. We expect, of course, that the results will show a relatively high Rand index between the amalgamated partition and the first contributory partition. Table 11 shows that the algorithm finds only two clusters with an almost equal number of consumers in each. As expected, the first partition (Electronic navigation) exhibits the highest Rand index between it and the amalgamated partition. Both Rear TV monitor and Security system display much lower Rand values. The higher-order Rand index is .615. This value exceeds the counterpart higher-order Rand (.440), noted for the control (i.e., equally-weighted) case, shown at the bottom of Table 11. However, these two Rands are not strictly comparable, given the differences in the contributory partition weights. Indeed, we caution the reader that the *à priori* weights option should be used

Table 11
Results of Applying Different Constraints to Hubert-ArabieModified Rand Coefficients
(Random start partition for each case)

Type of Modification	Amalgamated Rand Index	Individual Rand Indexes			
		Electronic navigation	CD player	Rear TV monitor	Security system
• *À priori* weights of .5, .3, .1, and .1 for contributory partitions 1, 2, 3, and 4 No. of individuals: C-1 982 C-2 993 Iterations = 1302	.615	1.000	.312	.126	.092
• Minimal Rand of .3 for each contributory partition No. of individuals: C-1 1059 C-2 824 C-3 92 Iterations = 445	.436	.527	.577	.339	.300
• Minimal cluster size of 300 respondents No. of individuals: C-1 975 C-2 700 C-3 300 Iterations = 1087	.384	.413	.542	.297	.283
• Control case: no constraints No. of individuals: C-1 1086 C-2 797 C-3 92 Iterations = 1295	.440	.488	.655	.369	.249

with discretion. Solutions can be highly sensitive to disparity in *à priori* weights across partitions.

6.2 Minimal Rand Constraint and Sample Size

Table 11 also shows the results of requiring each contributory Rand index to have a value of at least .3 between the contributory and the amalgamated clustering. Compared to the control case, the amalgamated Rand index drops slightly as a consequence of this constraint. In both cases, however, three clusters are obtained. Table 11 also shows the case where we constrain the clusters of the amalgamated partition so that each has a minimum of 300 respondents. The resulting Rand index drops to .384. We also note that cluster C-3 has 300 respondents. It should be mentioned that user-supplied constraints on Rand index values (between the amalgamated and contributory partitions) and minimal cluster size may result in not finding a solution that satisfies all of these constraints. If so, the algorithm informs the user of this condition. In sum, the use of side conditions increases the flexibility of SEGWAY and permits the user to conduct rudimentary kinds of sensitivity analyses of the data set. However, the user should exercise the usual cautions in applying constraints, because solutions can be sensitive to these factors.

7. Summary

This paper has introduced and illustrated a procedure for finding an amalgamated clustering of a set of contributory partitions, each based on separate sets of variables for the same individuals. Because the contributory partitions consist of nominal variables, the proposed model can also be utilized as a way to cluster categorical data sets. The SEGWAY model is based on a generalization of Rand index maximization, as defined in Table 1.

For comparison purposes, we also considered the highly popular latent class method. We found (Table 7) that the generalized Rand approach cross-validated slightly better with the exogenous (demographic) variables than did the latent class model. We also demonstrated the potential value of various weighting constraints that can be imposed on SEGWAY solutions. These include the assignment of *à priori* weights to contributory partitions, minimal Rand values for each starting position, and minimal sample sizes.

As we have tried to point out, the SEGWAY model represents a competitor to traditional latent class analysis as applied to categorical variables in selecting either the initial partition of a set of categorical variables or an amalgamated clustering of contributory partitions of the same data set.

7.1 Caveats

The generalized Rand index procedure, like similar approaches (e.g., latent class analysis), is subject to the fundamental problems of: (a) local optima and their dependence on starting configuration conditions and (b) determining the "best" number of clusters. Insofar as the Rand approach is concerned, our earlier findings indicated that stage 2 of the algorithm is reasonably insensitive to randomly determined starting partitions, at least for the empirical data set used here.

Following up on these preliminary findings (reported earlier), we examined the comparative behavior of SEGWAY versus latent class analysis in recovering similar partitions under different starts. First, ten sets of randomly determined starting partitions (three clusters each) were developed. We then separately applied Rand and latent class to each starting partition to obtain final partitions for the data set of this paper and computed the number of hits (after column permutation) for each of the 45 distinct pairs of partitions. We found that the average number of hits was 1381(69.9 percent) for latent class and 1923 (97.4 percent) for SEGWAY. Of course, this exercise entails both a small sample and only one empirical data set. Clearly, more investigation of the topic of robustness for each method is needed. Still, the problem of solution robustness is an important one, given that this class of techniques finds only a local optimum.

Finding the ''best'' number of clusters or classes is a still-unsolved problem. The principal difficulty is the lack of a compelling external criterion of ''goodness.'' Such internal criteria as AIC, BIC, and CAIC (Wedel and DeSarbo 1994) need not lead to the ''correct'' number of clusters, as related to an external criterion.

7.2 Future Outlook

So far, the performance of the proposed Rand-based algorithm looks promising. Clearly, more studies are needed of its comparative performance with traditional latent class modeling. In one sense SEGWAY is ''less theoretical'' than latent class analysis. On the other hand, SEGWAY's design is directly related to the objective of finding a ''best amalgamated'' partition that is free of local independence assumptions about the resulting classes. In addition, SEGWAY may be more robust to initial configurations than is latent class analysis. It seems to us that both methods have their respective places, depending upon the assumptions that the researcher wishes to make about the data set's structure.

Appendix A

In this appendix we describe the algorithm that finds D, the higher order clustering as discussed in the body of the paper. We let $R_j(D)$ be the adjusted Rand index between partition j and the target, as defined by D. The problem is then to

$$\max_{D} V(D) = \sum_{j=1}^{J} W_j R_j(D), \tag{A1}$$

subject to

$$R_j(D) \geq R_j, \tag{A2a}$$

and

$$\sum_{i=1}^{I} \psi_j(D_i) \geq h, \text{ for each } j, \tag{A2b}$$

where W_j is a set of nonnegative weights indicating the importance of each partition, with $\sum_{j=1}^{J} W_j = 1$; R_j is a user-specified lower bound on the adjusted Rand measure between the target and partition j, and the second constraint (A2b) ensures that each of the target classes has at least h (a user-supplied number of individuals) in it. Note that $\psi_j(D_i)$ is an indicator variable taking

on a value of 1 if $D_i = j$ and 0 otherwise.

Stage 1

We assume that we have an initial solution $D^{(0)}$ that satisfies (A2a) and (A2b). This solution can be obtained by multiple correspondence analysis, randomly generating a solution, or creating a solution with many more levels than desired and then combining levels via a greedy heuristic to reduce the number of levels for the target clustering. If none of these methods produces a feasible solution, then individual cluster assignments are changed to maximize the objective function and move the solution closer to the constraints. Note that if no feasible solution is obtained, then the constraints on the right-hand side of (A2a) and (A2b) are altered.

Because the greedy heuristic may be less familiar to most readers, we now describe its use in obtaining an *initial* partition. The objective is to make an initial assignment of I individuals to T user-specified classes. Ideally, the Stage 1 solution should, in itself, result in a high value of $V(D)$ in Equation (1), even before Stage 2 of SEGWAY is applied. The I individuals can be characterized by $H \leq I$ distinct patterns of the contributory categorical variables. Also, in general, H will be much larger than T, the number of user-desired clusters. We begin with a clustering based on $H > T$ clusters. This initial partition consists of all possible patterns, $C_{i1}, C_{i2}, \ldots, C_{iJ}$ across the I individuals. If all I people have different patterns, then $H = I$. Two (or more) individuals who have the same levels across all the J original categorical variables are assigned to the same cluster.

The number of clusters, at this point, is the number of unique patterns. If H is no greater than T (an unlikely event) the problem is solved. However, in practice $H \gg T$. The greedy heuristic now considers all $\binom{H}{2}$ distinct pairs of clusters for potential merging. The heuristic chooses the pair with the highest $V(D)$. The process is replicated until the number of clusters is reduced from H, one at a time, to the user-specified number of T classes. The user is then free to choose still other values of T, the desired number of starting clusters for later refinement via Stage 2 of the SEGWAY model.

Stage 2

Stage 2 represents the main part of the model. One starts with an initial partition (obtained by any of the methods described in the body of the paper) of I individuals assigned to T clusters. In Stage 2 the algorithm allows a change in any individual cluster assignment pair to maximize $V(D)$. Specifically,

Step 1: Initialization

$$p \leftarrow 0$$
$$NCT(j) \leftarrow \sum_i \psi_j(D_i^{(0)}) :$$
$$NR(n,j) \leftarrow \sum_i \psi_j(C_{in}) ,$$

where $\psi_j(C_{in}) = 1$ if $C_{in} = j$ and 0 otherwise.

$$NCRT(n,j,k) \leftarrow \sum_i \psi_j\,(C_{in})\,\psi_k\,(D_i^{(0)}) ;$$

$$X4 \leftarrow I\,(I-1)/2 ;$$

$$X2 \leftarrow 1/2 \sum_{j=1}^{T} NCT(j)(NCT(j)-1) ;$$

$$X3(n) \leftarrow 1/2 \sum_{j=1}^{M_n} NR(n,j)(NR(n,j)-1) ;$$

$$X1(n) \leftarrow 1/2 \sum_{j=1}^{M_n} \sum_{k=1}^{T} NCRT(n,j,k)(NCRT(n,j,k)-1) ;$$

$$V^{(p)} \leftarrow 2 \sum_{n=1}^{J} W_n \frac{X1(n)X4 - X3(n)X2}{X2(X4 - X3(n)) + X3(n)(X4 - X2)} .$$

Step 2:

For $i = 1, \ldots, I$, and $k = 1, \ldots, T$,

$X2T \leftarrow X2$; $X1T(n) \leftarrow X1(n)$;

1. $X2T \leftarrow XT2 + NCT(k) + 1 - NCT(D_i^{(p)})$;
2. $X1T(n) \leftarrow X1T(n) + NCRT(n,C_{in},k) + 1 - NCRT(n,C_{in},D_i^{(p)})$;
3. $RT(n) \leftarrow 2\dfrac{X1T(n)X4 - X3(n)X2T}{X2T(X4 - X3(n)) + X3(n)(X4 - X2T)}$.

   $V^{(p+1)} = \max_{i,k} \sum_n W_n RT(n)$ subject to the constraints.

   If $V^{(p+1)} \leq V^{(p)}$, stop; otherwise let (i^*,k^*) be the arguments corresponding to the maximum.

Step 3: Update

$$p \leftarrow p + 1 ,$$
$$k \leftarrow D_i^{(p-1)} ,$$

$$D_i^{(p)} \leftarrow \begin{cases} D_i^{(p-1)} & \text{if } i \neq i^* ; \\ k^* & \text{if } i = i^* ; \end{cases}$$

$NCT(k) \leftarrow NCT(k) - 1$;

$NCT(k^*) \leftarrow NCT(k^*) + 1$;

$NCRT(n,C_{i^*n},k) \leftarrow NCRT(n,C_{i^*n},k) - 1$ for all n ;

$NCRT(n,C_{i^*n},k^*) \leftarrow NCRT(n,C_{i^*n},k^*) + 1$ for all n .

$X1(n) \leftarrow X1(n) + NCRT(n,C_{i^*n},k^*) + 1 - NCRT(n,C_{i^*n},k)$;

$X2 \leftarrow X2 + NCT(k^*) + 1 - NCT(k)$.

Go to Step 2.

Appendix B

In this appendix we describe the model underlying the simulation described in the body of the paper. We also describe the design parameter ρ, used in the simulation.

> Assume there a J partitions, where partition j has M_j clusters.
> Let $P(l_1, \ldots, l_J)$ = Prob (any individual belongs to cluster l_j of partition j; $j = 1, \ldots, J$) (B1)

We also assume that J clusters are independently and identically drawn from $P(l_1, \ldots, l_J)$ across the I individuals.

In a latent class model, the probability that an individual is in class $l_1, \ldots, l_J$ of partitions 1 through J, respectively, is independent, conditional on the individual's latent class membership. Hence,

$$P(l_1, \ldots, l_J \mid t) = \prod_{j=1}^{J} P_j(l_j \mid t) \tag{B2a}$$

and

$$P(l_1, \ldots, l_J) = \sum_{t=1}^{T} [\prod_{j=1}^{T} P_j (l_j \mid t)]\, Q(t) \tag{B2b}$$

where T is the number of latent classes and $Q(t)$ is the probability that an individual belongs to latent class t.

In the simulations we assume that $Q(t) = 1/T$. Therefore, we only need $P_j(l_j \mid t)$ for $l_j = 1, \ldots, M_j$ and $t = 1, \ldots, T$. However, there are clearly too many parameters to make sense of the results. For this reason, we parametrize $P_j(l_j \mid t)$.

Table B.1
Illustrative Probabilities, Relating Partitions to Cluster and Latent Class

	Latent Class 1		Latent Class 2	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Partition 1	.8808	.1192	.1192	.8808
Partition 2	.7311	.2689	.2689	.7311

Defining Rho

Specifically, associated with each of the J partitions is a parameter ρ_j; $j = 1, \ldots, J$. We then let

$$P_j(l_j \mid t) = \frac{e^{\rho_j(l_j-\bar{j})(t-\bar{T})}}{\sum_{m=1}^{M_j} e^{\rho_j(m-\bar{j})(t-\bar{T})}} \tag{B3}$$

where $\bar{j} = \frac{M_j + 1}{2}$ and $\bar{T} = \frac{T + 1}{2}$ for centering.

For example, assume there are two latent classes ($T = 2$ and $\bar{T} = \frac{2+1}{2} = 1.5$). Assume there are two partitions, each with two classes ($J = 2, M_1 = M_2 = 2$).

Then $P_1(1 \mid 1)$ = Prob(segment 1 in partition 1 | latent class 1)

$$= \frac{e^{\rho_1(1-1\frac{1}{2})(1-1\frac{1}{2})}}{e^{\rho_1(1-1\frac{1}{2})(1-1\frac{1}{2})} + e^{\rho_1(2-1\frac{1}{2})(1-1\frac{1}{2})}}$$

$$= \frac{e^{\frac{1}{4}\rho_1}}{e^{\frac{1}{4}\rho_1} + e^{-\frac{1}{4}\rho_1}}$$

Similarly,

$$P_2(1 \mid 1) = \frac{e^{\frac{1}{4}\rho_2}}{e^{\frac{1}{4}\rho_2} + e^{-\frac{1}{4}\rho_2}}, \quad P_1(1 \mid 2) = \frac{e^{-\frac{1}{4}\rho_1}}{e^{-\frac{1}{4}\rho_1} + e^{\frac{1}{4}\rho_1}}$$

and

$$P_2(1 \mid 2) = \frac{e^{-\frac{1}{4}\rho_2}}{e^{-\frac{1}{4}\rho_2} + e^{\frac{1}{4}\rho_2}}; \quad (P_j(2 \mid t) = 1 - P_j(1 \mid t)).$$

If $\rho_1 = 4$ and $\rho_2 = 2$, then we have Table B1.

Consider a data set with I rows and two columns, where the entry in row i and column j is either a 1 or 2 depending on whether individual i

belongs to cluster 1 or cluster 2 in partition j. There will be many more rows that are (1,1), predominantly individuals that belong to latent class 1, and (2,2), predominantly individuals that belong to latent class 2 than (1,2) or (2,1). Hence, the associated 2 by 2 contingency table has a large χ^2 value.

The extent to which the probabilities of being in a given class for a given partition differ across the latent classes affects the size of the χ^2 value. Specifically, in the structure of our example,

$$\chi^2 = 16\, I\, [P_1(1 \mid 1) - \tfrac{1}{2}]^2\, [P_2(1 \mid 1) - \tfrac{1}{2}]^2 . \qquad \text{(B4)}$$

Since $P_1(1 \mid 1)$ increases with ρ_1 and $P_2(1 \mid 1)$ increases with ρ_2, then the χ^2 values increase from zero, when $\rho_1 = \rho_2 = 0$, to infinity, as ρ_1 and ρ_2 go to infinity (i.e., $P_1(1 \mid 1) = P_2(1 \mid 1) = 1$). For our example, $\chi^2 = 16I(.8808 - .5)^2(.7311 - .5)^2 = .1237I$.

Appendix C

The purpose of this Appendix is to compare two different measures for the agreement between two clusterings of n objects. The two clusterings can be visualized as an $R \times C$ contingency table with entries n_{rc}. R and C denote the respective number of clusters for the first and second clustering, n_{rc} is the number of data points in cluster r in Clustering 1 and cluster c in Clustering 2. The number of data points in cluster r in the first clustering is $n_{r\cdot} = \sum_{c=1}^{C} n_{rc}$ and the number of data points in cluster c in the second clustering is $n_{\cdot c} = \sum_{r=1}^{R} n_{rc}$.

The two measures that we consider are chi-squared (χ^2) and adjusted Rand ($\Re$). Each of these measures is a function of n_{rc}, $n_{r\cdot}$, and $n_{\cdot c}$ as follows:

$$\chi^2 = \sum_{r=1}^{R} \sum_{c=1}^{C} (n_{rc} - E_{rc})^2 / E_{rc} , \qquad \text{(C1)}$$

where $E_{rc} = n_{r\cdot}\, n_{\cdot c}/n$;

$$\Re = \left[\sum_{r=1}^{R} \sum_{c=1}^{C} \binom{n_{rc}}{2} - A_1 A_2 / T \right] / \left[(A_1 + A_2)/2 - A_1 A_2 / T \right] , \qquad \text{(C2)}$$

and where $A_1 = \sum_{r=1}^{R} \binom{n_{r\cdot}}{2}$; $A_2 = \sum_{c=1}^{C} \binom{n_{\cdot c}}{2}$; and $T = \binom{n}{2}$.

It is assumed that the sizes of the clusters in each clustering ($n_{r\cdot}$ and $n_{\cdot c}$) are fixed. Hence, E_{rc} in (1) and A_1, A_2 and T in (2) do not vary. As a result,

$$\chi^2 = \sum_{r=1}^{R} \sum_{c=1}^{C} n_{rc}^2 / E_{rc} - n \qquad \text{(C1a)}$$

and

$$\Re = \left[\tfrac{1}{2} \sum_{r=1}^{R} \sum_{c=1}^{C} n_{rc}^2 - n/2 - A_1 A_2 / T \right] / \left[(A_1 + A_2)/2 - A_1 A_2 / T \right] \qquad \text{(C2a)}$$

In general, it is difficult to relate χ^2 to $\Re$ because each depends on $n_{r.}$ and $n_{.c}$ in complicated ways. If the sizes of each cluster within each clustering are the *same* however, then χ^2 and $\Re$ are *linearly related* to each other. For example, if there are 100 individuals, one clustering could have 25, 25, 25, 25 individuals and a second clustering could have 20, 20, 20, 20, and 20. Toward this end,

$$n_{r.} = n/R \text{ for } r = 1, \ldots, R\,, \qquad \text{(C3a)}$$

and

$$n_{.c} = n/C \text{ for } c = 1, \ldots, C\,. \qquad \text{(C3b)}$$

Equations (C3a) and (C3b) imply:

$$E_{rc} = n/RC\,, \qquad \text{(C4a)}$$

$$A_1 = n(n-R)/2R\,, \qquad \text{(C4b)}$$

$$A_2 = n(n-C)/2C\,. \qquad \text{(C4c)}$$

Substituting (C4a) into (C1a) yields

$$\chi^2 = a_o + b_o\, x$$

where

$$a_o = -n, \quad b_o = RC/n \text{ and } x = \sum_{r=1}^{R} \sum_{c=1}^{C} n_{rc}^2\,. \qquad \text{(C5a)}$$

Substituting (C4b) and (C4c) into (C2a) yields

$$\Re = a_1 + b_1\, x\,, \qquad \text{(C5b)}$$

where $a_1 = -[n/2 + n(n-R)(n-C)/(2(n-1)\,RC)]/d$,
and $b_1 = 1/(2d)$,

with $d = -n^2[n(R + C - 2) + R + C - 2RC]/[4(n-1)RC]$.

Finally,

$$\Re = [a_1 - a_0 b_1 / b_0] + [b_1 / b_0]\, \chi^2\,. \qquad \text{(C6)}$$

Remark. If the two clusterings are independent, then $\chi^2 = 0$, and $\Re$ is close to, but not necessarily, zero. The former comment is obvious because independence implies $n_{rc} = E_{rc}$. To see that $\Re$ need not be zero, consider (C5b). For example, with $n = 20$, $R = C = 2$ then $\Re = -1/18$.

References

BARTHÉLEMY, J. P., LECLERC, B., and MONJARDET, B. (1985), "On the Use of Ordered Sets in Problems of Comparison and Consensus of Classifications," *Journal of Classification, 3*, 187-224.

DAY, W. H. E. (1986), "Foreword: Comparison and Consensus of Classifications," *Journal of Classification, 3*, 177-182.

DE LEEUW, J. (1984), "The Gifi System of Nonlinear Multivariate Analysis," in *Data Analysis and Informatics*, Eds., E. Diday, M. Jambu, L. Lebart, J. Pages, and R. Tomassone, III, Amsterdam, The Netherlands: North-Holland, 415-424.

DEMPSTER, A. P., LAIRD, N. M., and RUBIN, D. B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, B, 39*, 1-38.

DESARBO, W. S., CARROLL, J. D., CLARK, L. A., and GREEN, P.E. (1984), "Synthesized Clustering: A Method for Amalgamating Clustering Bases with Differential Weighting of Variables," *Psychometrika, 49*, 57-78.

DESARBO, W. S., and CRON, W. L. (1988), "A Maximum Likelihood Methodology for Clusterwise Linear Regression," *Journal of Classification, 5*, 249-282.

DESARBO, W. S., WEDEL, M., VRIENS, M., and RAMASWAMY, V. (1992), "Latent Class Metric Conjoint Analysis," *Marketing Letters, 3*, 273-288.

DILLON, W. R., and KUMAR, A. (1994), "Latent Structure and Other Mixture Models in Marketing: An Integrative Survey and Overview," in *Advanced Methods of Marketing Research*, Ed., R. P. Bagozzi, Cambridge, MA: Blackwell, 295-351.

DILLON, W. R., MADDEN, T. J., and MULANI, N. (1983), "Scaling Models for Categorical Variables: An Application to Latent Structure Models," *Journal of Consumer Research, 10*, 209-223.

DILLON, W. R., and MULANI, N. (1989), "LADI: A Latent Discriminant Model for Analyzing Marketing Research Data," *Journal of Marketing Research, 26*, 15-29.

GIFI, A. (1990), *Nonlinear Multivariate Analysis*, New York: Wiley.

GREEN, P. E., KRIEGER, A. M., and CARROLL, J. D. (1987), "Conjoint Analysis and Multidimensional Scaling: A Complementary Approach," *Journal of Advertising Research, 27*, 21-27.

GREEN, P. E., SCHAFFER, C. M., and PATTERSON, K. M. (1988), "A Reduced-Space Approach to the Clustering of Categorical Data in Market Segmentation," *Journal of the Market Research Society, 30*, 267-288.

GREENACRE, M. J. (1984), *Theory and Application of Correspondence Analysis*, London: Academic Press.

HEISER, W. J. (1981), *Unfolding Analysis of Proximity Data*, Leiden: Department of Data Theory, University of Leiden.

HUBERT, L., and ARABIE, P. (1985), "Comparing Partitions," *Journal of Classification, 2*, 193-198.

LEBART, L., MORINEAU, A., and WARWICK, K. (1984), *Multivariate Descriptive Statistical Analysis*, New York: Wiley.

MACQUEEN, J. (1967), "Some Methods for Classification and Analysis of Multivariate Observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol 1*, Eds., L.M. LeCam and J. Neyman, Berkeley: University of California Press, 281-297.

MARGUSH, T., and MCMORRIS, F. R. (1981), "Consensus n-Trees," *Bulletin of Mathematical Biology, 43*, 239-244.

MCMORRIS, F. R., and NEUMANN, D. A. (1983), "Consensus Functions on Trees," *Mathematical Social Sciences, 4*, 131-136.

MEULMAN, J. (1982), *Homogeneity Analysis of Incomplete Data*, Leiden: DSWO Press.

MILLIGAN, G. W., and COOPER, M. C. (1987), "Methodology Review: Clustering Methods," *Applied Psychological Measurement, 11*, 329-354.

NEUMANN, D. A., and NORTON, V. T., Jr. (1986), "Clustering and Isolation in the Consensus Problem for Partitions," *Journal of Classification, 3*, 281-298.

NISHISATO, S. (1980), *Analysis of Categorical Data: Dual Scaling and Its Applications*, Toronto: University of Toronto Press.

NISHISATO, S. (1984), "Forced Classification: A Simple Application of a Quantification Method," *Psychometrica, 49*, 25-36.

NISHISATO, S. (1993), *Elements of Dual Scaling: An Introduction to Practice Data Analysis*, Mahwah, NJ: Erlbaum.

NISHISATO, S. (1996), "Gleaning in the Field of Dual Scaling," *Psychometrika, 61*, 559-599.

POULSEN, C. S. (1990), "Mixed Markov and Latent Markov Modeling Applied to Brand Choice Behavior," *Internatinal Journal of Research in Marketing, 7*, 5-19.

RAMASWAMY, V., CHATTERJEE, R., and COHEN, S. H. (1996), "Joint Segmentation on Distinct Interdependent Bases with Categorical Data," *Journal of Marketing Research, 33*, 335-350.

VACH, W. (1994), "Presenting Consensus Hierarchies," *Journal of Classification, 11*, 59-78.

VAN BUUREN, J., and HEISER, W. J. (1989), "Clustering N Objects into K Groups Under Optimal Scaling of Variables," *Psychometrika, 54*, 699-706.

VAN DER POL, E., and DE LEEUW, J. (1986), "A Latent Markov Model to Correct Measurement Error," *Sociological Methods and Research, 15*, 118-141.

WEDEL, M., and DESARBO, W (1994), "A Review of Recent Developments in Latent Class Regression Models," in *Advanced Methods of Marketing Research*, Ed., R.F. Bagozzi, Oxford: Blackwell, 352-383.