HHS Public Access
Author manuscript
Br J Math Stat Psychol. Author manuscript; available in PMC 2018 June 01.

Published in final edited form as:
Br J Math Stat Psychol. 2009 May ; 62(Pt 2): 349–368. doi:10.1348/000711008X304385.

Cross validation issues in multiobjective clustering

Michael J. Brusco[1,*] and **Douglas Steinley**[2]

[1]Department of Marketing, Florida State University, Florida, USA

[2]University of Missouri-Columbia, Columbia, Missouri, USA

Abstract

The implementation of multiobjective programming methods in combinatorial data analysis is an emergent area of study with a variety of pragmatic applications in the behavioural sciences. Most notably, multiobjective programming provides a tool for analysts to model trade offs among competing criteria in clustering, seriation, and unidimensional scaling tasks. Although multiobjective programming has considerable promise, the technique can produce numerically appealing results that lack empirical validity. With this issue in mind, the purpose of this paper is to briefly review viable areas of application for multiobjective programming and, more importantly, to outline the importance of cross-validation when using this method in cluster analysis.

I. Introduction

Combinatorial data analysis (CDA) encompasses a wide class of models and methods that can be applied to a number of quantitative problems in psychology, as well as other disciplines. In their 1992 article in the *Annual Review of Psychology,* Arabie and Hubert (1992) identified important CDA problems related to seriation, cluster analysis, additive trees, and network models. Their review captured the various classes of combinatorial optimization problems in each of these areas, as well as available solution procedures for such problems. Hubert, Arabie, and Meulman (2001) provided an updated review of hierarchical clustering, partitioning, and seriation, as well as a general dynamic programming paradigm that can be used to obtain optimal solutions for problems of modest size. More recently, Hubert, Arabie, and Meulman (2006) presented a suite of MATLAB (MathWorks, 2002) routines for fitting a variety of structures to one- and two-mode proximity data, including linear and circular unidimensional scaling, ultrametric and additive trees, and anti-Robinson (Robinson, 1951) forms.

Recent developments in CDA have also emphasized the need for multiobjective programming methods that enable the quantitative analyst to make trade offs among competing objective criteria (Brusco & Stahl, 2005b, Chapters 6, 11). A key aspect of multiobjective programming is the *non-dominated* (or *Pareto efficient*) set of solutions. To understand this concept, consider a set of H objectives (indexed $1 \leq h \leq H$) and let $\boldsymbol{\tau}$

*Correspondence should be addressed to: Dr Michael J. Brusco, Department of Marketing, College of Business, Florida State University, Tallahassee, FL 32306-1110, USA (mbrusco@cob.fsu.edu).

represent a feasible solution for the relevant optimization problem. Without loss of generality, assume that each of the H objectives should be maximized. In addition, denote $f_h(\boldsymbol{\tau})$ as the objective criterion index for criterion h (indexed $1 \leq h \leq H$). A solution $\boldsymbol{\tau}'$ *dominates* $\boldsymbol{\tau}$ if $f_h(\boldsymbol{\tau}') \geq f_h(\boldsymbol{\tau})$ for all $1 \leq h \leq H$, and $f_h(\boldsymbol{\tau}') \geq f_h(\boldsymbol{\tau})$ for at least one h. If T denotes the set of all feasible solutions for the optimization problem, then the non-dominated set, $T' \subset T$, consists of all solutions, $\boldsymbol{\tau}'$, that are not dominated by any other solution. For small problems, it might be possible to generate the entire set of non-dominated solutions via exhaustive enumeration. Although this option is impractical for most applications, there are a variety of approaches for generating or approximating non-dominated solutions and these are discussed by numerous authors (Ehrgott, 2005; Steuer, 1986; Zeleny, 1982). One of the most popular approaches, which we use in this paper, is to convert the multiobjective program into single scalar functions of the objective criteria. If the resulting scalar optimization problems can be solved exactly, then the corresponding solutions are guaranteed to be non-dominated. When heuristics are applied to the scalar optimization problems, approximations of non-dominated solutions are obtained.

In recent years, there have been a variety of legitimate applications of multiobjective CDA in quantitative psychology and related fields. Most notably, previous research has focused on seriation and scaling (Brusco, 2002; Brusco & Stahl, 2001a, 2005a, 2005b, Chapter 11) and cluster analysis (Brusco & Cradit, 2004, 2005; Brusco & Stahl, 2001b, 2005b, Chapter 6; Steinley & Hubert, 2008). These applications notwithstanding, appropriate cross-validation is necessary when using multiobjective programming to partition a set of objects based on multiple sets of proximity matrices. The problem especially manifests itself when using multiobjective programming to partition a sample of cases on two or more distinct sets of clustering variables with the ultimate goal of using the clustering solution to classify new cases *based on only one of the variable sets.*

To illustrate, consider a common scenario in the marketing literature that focuses on simultaneously accomplishing the following goals: (a) profile customers based on demographic variables and (b) explain variation in a set of response variables such as satisfaction, potential for switching, likelihood of purchasing new products or services, and so on (Brusco, Cradit, & Stahl, 2002; Brusco, Cradit, & Tashchian, 2003; DeSarbo & Grisaffe, 1998; Krieger & Green, 1996). Multiobjective clustering methods can produce solutions with impressive goodness-of-fit statistics for a *training sample* of cases measured on the profiling and response variable sets; however, the results are frequently devoid of any merit from a cross-validation standpoint. The problem stems from the fact that, when new cases are presented or the clustering solution is applied to a *validation sample,* only the profiling information is used for classification purposes. A common assumption is that classification of the validation sample would still yield good explanation of the response variables, but we will show that this often does not materialize because the multiobjective clustering merely capitalizes on chance in the training sample.

It is imperative to unequivocally clarify at this point that we are not suggesting that *all* applications of multiobjective clustering to multiple sets of proximity data are problematic. Indeed, there are situations where multiobjective programming provides an excellent trade off among multiple data sources and the resulting solutions will stand up to cross-validation

(e.g. Handl & Knowles, 2006). However, in the case of our marketing example above, as well as many psychological settings, an analyst could easily be fooled by impressive fit statistics from multiobjective clustering and not realize that the resulting solution cannot be cross-validated.

A brief review of multiobjective CDA is provided in section 2. A biobjective optimization problem corresponding to the clustering of a data set based on a weighted function of explained variation for profiling and response variables is described in section 3. In section 4, we use an example to demonstrate that this procedure produces numerically appealing results for the sample that do not translate well to the entire population. Section 5 reports the results of an application of the biobjective model to a real psychological data set. The paper concludes in section 6 with a brief summary.

2. Literature review – multiobjective CDA

2.1. Multiobjective seriation and scaling

One class of applications of multiobjective programming in CDA pertains to the seriation and scaling of proximity matrices. In some instances, multiobjective programming has been used to obtain a structural fit to a single proximity matrix based on multiple criteria. For example, using a generalization of well-established dynamic programming procedures (Hubert & Golledge, 1981; Lawler, 1964) to generate non-dominated solutions, Brusco and Stahl (2001a) presented a multiobjective procedure for seriation that can be used to identify a permutation of objects that provides a good structural fit to a single asymmetric proximity matrix. The objective function of their model was to optimize a weighted function of two or more structural indices. Brusco and Stahl (2005a) expanded on this problem by developing a biobjective seriation and scaling model for different components of skew-symmetric matrices.

Whereas the multiobjective seriation and scaling models developed by Brusco and Stahl (2001a, 2005a) focused on the analysis of a single source of proximity data, other multiobjective models have been developed within the context of a set of proximity matrices available for the same set of objects. For example, Brusco (2002) proposed a multiobjective programming model for finding a permutation that provides a good fit for each of a set of matrices, even when antagonistic relationships existed among some of the matrices. Brusco, Stahl, and Cradit (2002) subsequently expanded this model to the case of multiobjective multidimensional scaling of a set of proximity matrices. In each of these multiobjective programming applications, the proximity matrices were obtained from confusion experiments on the same set of objects under different experimental conditions; however, the models are also applicable for proximity data produced from paired comparison, sorting, or other collection methods.

2.2. Multiobjective clustering

Multiobjective programming methods have also received considerable attention in clustering applications. These applications span a number of areas of scientific inquiry, including computational biology (Handl, Kell, & Knowles, 2007; Handl & Knowles, 2006, 2007;

Ripon, Tsang, Kwong, & Ip, 2006), marketing (Brusco *et al.*, 2002, 2003), psychology (Brusco & Cradit, 2004, 2005; Brusco & Stahl, 2001b; Steinley & Hubert, 2008), as well as the more general statistical and pattern recognition literature (Delattre & Hansen, 1980; Ferligoj & Batagelj, 1992; Law, Topchy, & Jain, 2004). In addition to the emergence of multiobjective programming methods in different scientific disciplines, the role of multiobjective programming also differs across applications.

In one of the earliest multiobjective clustering papers, Delattre and Hansen (1980) developed an algorithm that generates the entire non-dominated set for a biobjective clustering problem using *partition split* and *partition diameter* as the objective criteria. Partition split, which is analogous to Johnson's (1967) minimum method in hierarchical clustering, is the minimum distance between any two objects not in the same cluster. Partition diameter, which is analogous to Johnson's (1967) maximum method in hierarchical clustering, is the maximum distance between any two objects that are in the same cluster. Brusco and Cradit (2004) also utilized the partition diameter criterion in their biobjective clustering approach for confusion data. Drawing from the pioneering work of Hubert (1973, 1974a), Brusco and Cradit established compact partitions by minimizing diameter and, subsequently, selected from among alternative optima by minimizing the number of inconsistencies in a digraph corresponding to the confusion matrices. Brusco and Cradit (2005) also exploited the potential for alternative optima for the diameter criterion, evaluating a variety of secondary criteria. Steinley and Hubert (2008) demonstrated that the K-means algorithm (Howard, 1966; MacQueen, 1967), when used in conjunction with an object-order constraint obtained based on an anti-Robinson criterion, was especially effective at recovering the underlying (known) structure in synthetic data sets. Moreover, relative to the use of K-means alone, the order constraint enabled the identification of more interpretable solutions for several empirical data sets.

In computational biology and pattern recognition, there are many instances where the effective clustering of a data set requires trade offs among competing criteria such as compactness, separation, and, connectedness. For example, Law *et al.* (2004) implemented a consensus clustering approach that uses partitions established by a variety of single-objective clustering approaches (e.g. K-means, single-linkage, model-based methods, etc.). Methods of this type are often termed 'ensemble approaches' (Strehl & Ghosh, 2002). In contrast to ensemble methods, Handl and Knowles (2007) adopted a direct approach using a multiobjective genetic algorithm that incorporates minimum spanning trees (see Hubert, 1974b) to produce well-connected solutions and a K-means algorithm (Howard, 1966; MacQueen, 1967) for achieving compactness.

3. A biobjective clustering problem

3. 1. Problem overview

Whereas section 2.2 identifies viable implementations of multiobjective programming in cluster analysis, there are other situations where the approach is of dubious value. To illustrate a the importance of cross-validation in multiobjective CDA, we return to the biobjective clustering problem that has its origin in the marketing literature (Brusco *et al.*, 2002; DeSarbo & Grisaffe, 1998; Krieger & Green, 1996). Although the same perils pertain

to the case of three or more objectives (see e.g. Brusco *et al.*, 2003), we restrict our attention to the biobjective case for the purposes of case of presentation and clarity.

The goal of the biobjective clustering problem is to obtain clusters that are 'reachable' with respect to a set of demographic/psychographic profiling variables, yet at the same time facilitate adequate explanation of one or more response variables. Krieger and Green describe an application corresponding to a company (using the disguised name 'Alpha Company') that markets home equity loans. A telephone survey was used to collect information from a sample of customers. The information gathered included responses to psychographic statements and an evaluation of the likeability of the company. Psychographic variables measure attitudes and opinions that help to establish personality traits, values, and other psychological characteristics of the respondents. The sample was initially classified (using K-means) on the basis of the psychographic data. Next, the company attempted '… to find out if the original segmentation could be "loosened" a bit to better predict respondent liking of the Alpha Company' (Krieger & Green, 1996, p. 356). An algorithm was applied to the K-means solution, reducing explained variation in the psychographic variables by 15%, but increasing explained variation in the response variable (likeability) by more than 100%. Krieger and Green (1996, p. 360) recognized that their algorithm capitalized on chance and suggested that using the results from a training sample to classify a validation sample from the same population would result in a degradation of explained variation; however, no information was presented with respect to the extent of this problem.

DeSarbo and Grisaffe (1998) demonstrated impressive explained variation results for a biobjective application pertaining to a utility company that sought to achieve strong prediction of overall satisfaction for its industrial clients, while maintaining homogeneous clusters with respect to firm-specific profile measurements. Similarly, Brusco *et al.* (2002) classified firms into homogeneous groups based on telecommunication activity while also enabling adequate prediction of long-distance spending. Brusco *et al.* used a split-half procedure to validate their results; however, they used *both* the profiling and response measures in the validation process. As noted previously, in many marketing and psychological applications, only the profiling measures would be available for classifying new cases.

To place the biobjective problem in a psychological context, consider a situation where a university department wishes to use demographic information on an incoming freshman class (e.g. SAT score, high-school GPA) to establish homogeneous clusters of students, while also yielding satisfactory prediction of college-GPA at the end of the first year. A second example is the development of a homogeneous classification of participants based on their histories of depressive disorders that also yields prediction of symptoms displayed during smoking cessation (Burgess *et al.*, 2002). A third illustration is the construction of clusters that facilitate explanation of social acceptance and adjustment, yet are also homogeneous with respect to aggressive and prosocial behaviours (Haselager, Cillessen, Van Lieshout, Riksen-Walraven, & Hartup, 2002).

3.2. Notation

S: a set of n indices corresponding to objects sampled from a population, $\{1,2, \ldots, n\}$;

U: the number of profiling variables, indexed $1 \le u \le U$;

V: the number of response variables, indexed $1 \le v \le V$;

$\mathbf{X}$: a matrix of profiling variable measurements with elements x_{iu} representing the measurement for object i on profiling variable u, for $1 \le i \le n$ and $1 \le u \le U$;

$\mathbf{Y}$: a matrix of response variable measurements with elements y_{iv} representing the measurement for object i on response variable v, for $1 \le i \le n$ and $1 \le v \le V$;

K: the number of clusters;

P_K: a partition of S into K clusters, $P_k = \{C_1, \ldots, C_K\}$, where C_k. is a set of n_k object indices assigned to cluster k for $1 \le k \le K$;

Π_K: the set of all possible partitions of n objects into K clusters. The cardinality of Π_K can be computed using a formula for the Stirling number of the second kind.

One such formula is provided by Hand (1981):

$$\frac{1}{K!}\sum_{k=0}^{K}(-1)^k\binom{K}{k}(K-k)^n. \quad (1)$$

$W_{\mathbf{X}}(P_K)$: the within-cluster sum-of-squares for the profiling variables that is produced from partition P_K. This quantity is computed as follows:

$$W_{\mathbf{X}}(P_K) = \sum_{k=1}^{K}\sum_{i \in C_k}\sum_{u=1}^{U}(x_{iu} - \bar{x}_{uk})^2, \quad (2)$$

where

$$\bar{x}_{uk} = \frac{1}{n_k}\sum_{i \in C_k} x_{iu} \quad \text{for } 1 \le u \le U \text{ and } 1 \le k \le K. \quad (3)$$

$W_{\mathbf{Y}}(P_K)$: the within-cluster sum-of-squares for the response variables that is produced from partition P_K. This quantity is computed as follows:

$$W_{\mathbf{Y}}(P_K) = \sum_{k=1}^{K}\sum_{i \in C_k}\sum_{v=1}^{V}(y_{iv} - \bar{y}_{vk})^2, \quad (4)$$

where

$$\bar{y}_{vk} = \frac{1}{n_k} \sum_{i \in C_k} y_{iv} \quad \text{for } 1 \le v \le V \text{ and } 1 \le k \le K. \quad (5)$$

$\mathrm{TSS}_{\mathbf{X}}$: the total sum-of-squares for the profiling variables, which is computed as:

$$\mathrm{TSS}_{\mathbf{X}} = \sum_{i=1}^{n} \sum_{u=1}^{U} (x_{iu} - \bar{x}_u)^2, \quad (6)$$

where

$$\bar{x}_u = \frac{1}{n} \sum_{i=1}^{n} x_{iu} \quad \text{for } 1 \le u \le U. \quad (7)$$

$\mathrm{TSS}_{\mathbf{Y}}$: the total sum-of-squares for the response variables, which is computed as:

$$\mathrm{TSS}_{\mathbf{Y}} = \sum_{i=1}^{n} \sum_{v=1}^{V} (y_{iv} - \bar{y}_v)^2, \quad (8)$$

where

$$\bar{y}_v = \frac{1}{n} \sum_{i=1}^{n} y_{iv} \quad \text{for } 1 \le v \le V. \quad (9)$$

$Z_{\mathbf{X}}(P_K)$: the proportion of variation in the profiling variables that is explained by partition P_K. This quantity is computed as follows:

$$Z_{\mathbf{X}}(P_K) = 1 - \frac{W_{\mathbf{X}}(P_K)}{\mathrm{TSS}_{\mathbf{X}}}. \quad (10)$$

$Z_{\mathbf{Y}}(P_K)$: the proportion of variation in the response variables that is explained by partition P_K. This quantity is computed as follows:

$$Z_{\mathbf{Y}}(P_K) = 1 - \frac{W_{\mathbf{Y}}(P_K)}{\mathrm{TSS}_{\mathbf{Y}}}. \quad (11)$$

α: a weighting parameter (0 ≤ α ≤ 1) that represents the proportion of weight allocated to the explained variation for the profiling variables.

3.3. Model formulation and solution procedure

To obtain non-dominated solutions for the biobjective clustering problem, we formulate the following scalar optimization problem:

$$\text{Maximize:}\,\alpha(Z_{\mathbf{X}}(P_K)) + (1-\alpha)(Z_{\mathbf{Y}}(P_K)), \quad (12)$$

$$\text{Subject to:}\,P_K \in \Pi_K. \quad (13)$$

For computational purposes, we substituted the right-hand-sides of equations (10) and (11) into equation (12) and, after some minor algebra, obtained the following equivalent objective function:

$$\text{Minimize:}\,\frac{\alpha}{\text{TSS}_{\mathbf{X}}}(W_{\mathbf{X}}(P_K)) + \frac{(1-\alpha)}{\text{TSS}_{\mathbf{Y}}}(W_{\mathbf{Y}}(P_K)). \quad (14)$$

This reformulation enabled us to obtain optimal solutions to the scalar optimization problem using a modified version of a branch-and-bound algorithm developed by Brusco (2006) for within-cluster sum-of-squares partitioning (see also Brusco & Stahl, 2005b, Chapter 6).

In applied analyses, there are two key parameter choices for the biobjective clustering model. First, it is necessary to determine the number of clusters, K. In Krieger and Green's (1996) application, K was selected by company management based on interpretability issues. Brusco *et al.* (2002) used the Caliński and Harabasz (1974) pseudo-F statistic for the K-means solution to choose an appropriate value of K (see Dimitriadou, Dolničar, & Weingessel, 2002, for a variety of other indices for selecting K). The determination of α is the second parameter choice. We recommend running the biobjective model multiple times, varying α from 0 to 1 in increments of .1. This process typically facilitates the identification of a range for α that can be investigated more thoroughly. For example, an analyst might initially identify .7 and .8 as possible candidates for α. Subsequent executions of the algorithm can be implemented for α = .725, .75, .775, etc. to narrow in on an effective choice.

Exact solutions to the scalar optimization problem using different values of α produce what are known as 'supported' non-dominated solutions (Erhgott & Gandibleaux, 2000; Ulungu & Teghem, 1994). For multiobjective combinatorial optimization problems, there are frequently a large number of 'unsupported' non-dominated solutions that cannot be identified using the scalar optimization problem with any weighting scheme. This limitation notwithstanding, we have found that the scalar optimization problem typically yields satisfactory trade offs among competing criteria.

4. A numerical demonstration

4.1. Data generation

To provide a numerical illustration of the potential pitfalls associated with multiobjective clustering, we produced synthetic variable measurements using a data generation process comparable to one recently employed by de Craen, Commandeur, Frank, and Heiser (2006). First, we generated a pseudo-population of $N = 3{,}600$ profiling variable measurements using bivariate normal distributions with equal variances. The data were generated to produce $K = 3$ equally sized clusters in the $U = 2$ – dimensional space. The variable means were (30, 70), (50, 50), and (70, 70) for clusters 1, 2, and 3, respectively. The 2 × 2 diagonal covariance matrices were 100 $\mathbf{I}_{2\times 2}$ for each of the three clusters (where $\mathbf{I}_{2\times 2}$ is a 2 × 2 identity matrix).

The measurements for the $V = 2$ response variables were generated in precisely the same manner as the profiling variables. Next, the 3,600 profiling measurements were paired with the 3,600 response measures after random relabeling of the latter. This process produces an extreme situation where the profiling variables are completely unrelated to the response variables. We use these data to demonstrate that, even in this extreme condition, numerically appealing results can be obtained using multiobjective programming methods. We deem this demonstration important in light of the fact that 'antagonistic' relationships between the profiling and response variables have been observed in the literature (Brusco *et al*., 2002, 2003; Krieger & Green, 1996). Section 5 focuses on the analysis of an empirical data set that exhibits moderately strong relationships among the profiling and response variables.

4.2. Obtaining biobjective solutions for a sample

We took a random sample of $n = \sqrt{N} = 60$ observations from the pseudo-population and solved the biobjective clustering problem presented in section 3 for a variety of values of α. For this sample size, we were able to solve each scalar optimization problem exactly using the branch-and-bound method. Accordingly, all of the obtained solutions are guaranteed to be in the non-dominated set. Table 1 provides the values of $Z_{\mathbf{X}}(P_K)$ and $Z_{\mathbf{Y}}(P_K)$ for each of 15 non-dominated solutions realized for different values of α. A graphical representation of non-dominated points is provided in Figure 1.

Figure 1 provides a visual indication of the marked improvements in one criterion that can be achieved for a small sacrifice in the maximum value of the other criterion. As shown in Table 1, the maximum value $Z_{\mathbf{X}}(P_K)$ is .72988. Because of the antagonistic relationship between the profiling variables and the response variables, this partition yields a poor value of explained variation for the response variables ($Z_{\mathbf{Y}}(P_K) = .06050$). However, for only a 1.87% reduction in $Z_{\mathbf{X}}(P_K)$ to .71622, the explained variation in the response variable more than doubles to $Z_{\mathbf{Y}}(P_K) = .14769$. Similarly, a modest 4.21% reduction in $Z_{\mathbf{X}}(P_K)$ to .69916 results in $Z_{\mathbf{Y}}(P_K) = .18380$, which is more than triple the value realized when $\alpha = 1$. It is precisely these large gains in explained variation for response measures realized for small sacrifices in explained variation for profiling variables that motivated Krieger and Green's (1996) affinity for this biobjective perspective. Such information is numerically appealing to marketing managers and behavioural science analysts; however, we pose the question: 'What do these results really mean?'

4.3. Implications of biobjective results for the population

We selected the non-dominated solution corresponding to $\alpha = .6675$ for further analysis because the percentage reduction in the maximum possible value for $Z_{\mathbf{X}}(P_K)$ was 12.97% (i.e. from .72988 to .63522), which is in the range of 10–15% considered by Krieger and Green (1996) and Brusco *et al.* (2002). Figure 2 shows the change in the cluster structure for the profiling variables when reducing α from 1 to .6675. The top panel of Figure 2 shows three distinct clusters labelled by ‘ + ‘, ‘○’, and ‘□’, which are realized when $\alpha = 1$. The bottom panel shows that a reduction of α degrades this structure. The ‘ + ’ cluster remains mostly intact, with only minor reassignments. However, the ‘○’ and ‘□’ clusters are mixed together, which is indicative of the reduction in explained variation.

The 12.97% reduction in $Z_{\mathbf{X}}(P_K)$ enables a more than 400% improvement in $Z_{\mathbf{Y}}(P_K)$ from .06050 to .31387. Figure 3 provides a visual reflection of the improvement in the cluster structure for the response variables when reducing α from 1 to .6675. The top panel of Figure 3 reveals a somewhat random smattering of ‘ + ’, ‘○’, and ‘□’ labels that is yielded when $\alpha = 1$. The bottom panel shows that reduction of α to .6675 provides enhanced clustering of the response measurements. Although far from a perfectly clean result, most of the ‘○’ labels have been pushed to the left on the v_1 axis, whereas most of the ‘□’ labels have been pushed to the right. The v_2 axis, however, does little to distinguish the response variable clusters.

The top panel of Table 2 provides a numerical summary of the non-dominated clustering solution obtained for $\alpha = .6675$. For the sample data, the profiling variable means indicate a distinct separation of the clusters on both variables. In contrast, the means for the response variables show some separation on variable v_1, but little on v_2, which is consistent with the bottom panel of Figure 3.

Now, for the all important question: given the biobjective solution for the sample, can we classify new cases based solely on their profiling measurements and achieve explained variation results for profiling and response measures comparable to $Z_{\mathbf{X}}(P_K) = .63522$ and $Z_{\mathbf{Y}}(P_K) = .31387$? To answer this question, we assigned each object in the pseudo-population to the same cluster as its nearest neighbor in the training sample. The results are summarized in the bottom panel of Table 2. This process produced three clusters of roughly equal size that are clearly well-separated with respect to both profiling measures. The proportion of explained variation for the profiling measures in the classified pseudo-population is an impressive .55478. Unfortunately, the results for the response variables are miserable. The cluster means on the response measures show no separation whatsoever, and the proportion of explained variation is an abysmal .00056.

The results of the illustration are clear. It is inappropriate to conclude that the measured explained variation for response variables in the sample will translate into a comparable level of explained variation for subsequent subsamples of the population. The results of the illustration are not sample specific. We repeated this analysis for many samples and always obtained comparable results. The underlying problem with the biobjective approach in this context is that it is merely exploiting the chance variation in the sample to induce numerically alluring results that have no pragmatic explanatory or predictive value.

5. An empirical application

5.1. The empirical data set

The numerical illustration in section 4 provided a demonstration of the shortcomings of the biobjective clustering model using a synthetic data set. In this section, we present additional evidence via an application to an empirical data set pertaining to the study of reflective judgment (Brusco, Cradit, Steinley, & Fox, 2008; Kitchener, Jensen, & Wood, 1999; Wood, Kitchener, & Jensen, 2005), The data are measurements for a sample of 3,618 participants on a single response variable, (v_1) reflective judgment score, and four profiling variables: (u_1) = ACT score; (u_2)= endorsement of a written essay; (u_3)= high-school GPA; and (u_4)= college GPA.

A correlation matrix for the response variable and four profiling variables is provided in Table 3. The multiple correlation coefficient for v_1 with respect to the four profiling variables is .526. Therefore, unlike the synthetic data set analyzed in section 4, there is a modest relationship between the response variable and profile variables in these data. Our goal is to identify a partition of participants that ensures homogeneity with respect to the profiling measurements, yet also provides adequate explanation of reflective judgment. Accordingly, the results of the biobjective clustering model could be used to classify additional participants based solely on their available profiling measurements, yielding a prediction for their reflective judgment score.

5.2. Data preparation

We randomly split the full data set into two equally sized samples each consisting of n = 1,809 participants. The training sample, S1, was analyzed using biobjective clustering for model selection purposes. The validation sample, S2, was used as a *holdout* sample to assess the explanatory power of the model. Variable u_1 (ACT score) exhibited a substantially greater range and variance relative to the other profiling variables and the response variables. For this reason, we applied the variance-to-range data transformation developed by Steinley and Brusco (2008) to the data prior to analysis. Steinley and Brusco demonstrated that this procedure facilitated appreciably better cluster recovery than three alternative methods: (a) no transformation; (b) z score transformation; and (c) range transformation.

The exact algorithm used in section 4 was not computationally feasible for the reflective judgment sample of n = 1,809 participants. For this reason, we used a simulated annealing heuristic developed by Brusco *et al* (2002) to approximate the non-dominated set for the biobjective clustering problem. We applied the heuristic to S1 using 11 different values of α (α = 1, .9, .8, .7, .6, .5, .4, .3, .2, .1,0). Based on prior research that suggested the appropriateness of a four-cluster solution (Brusco *et al.,* 2008), all results were obtained for K = 4 clusters. For each value of α, we restarted the heuristic 10 times, with each restart using a different random assignment of participants to clusters. The solution corresponding to the restart that yielded the largest value for equation (12) was stored as the best-found solution.

5.3. Computational results

Table 4 provides an approximation of the non-dominated set for the reflective judgment training sample (S1). A graphical representation of non-dominated points is provided in Figure 4. For each value of α. Table 4 reports the values of $Z_{\mathbf{X}}(P_K)$ and $Z_{\mathbf{Y}}(P_K)$ for the training sample, as well as the percentage reduction in $Z_{\mathbf{X}}(P_K)$ associated with the reduction of α by .1. The values of $Z_{\mathbf{X}}(P_K)$ and $Z_{\mathbf{Y}}(P_K)$ for the holdout sample (S2), which are obtained after assigning the S2 cases to the same cluster as their nearest neighbor in the multiobjective solution for S1, are reported in the last two columns of Table 4.

Once again, numerically compelling results for the training sample were evident. For example, reducing α from 1 to .8 provides a modest 8.5% reduction in $Z_{\mathbf{X}}(P_K)$ from .62425 to .57166; however, this modification yields an astounding 637% Improvement in $Z_{\mathbf{Y}}(P_K)$ from .06919 to .51038. Although other estimated non-dominated points in Table 4 are attractive (e.g. the solutions for $.7 \leq \alpha \leq .5$), we selected the solution for α = .8 for further analysis.

Table 5 provides detailed information on the cluster sizes and centroids for the α = .8 solution. The top panel of Table 5 contains the results for S1. The reflective judgment means for S1 reveal considerable cluster separation of the sample with respect to this response variable. The n_1 =435 participants assigned to cluster 1 tend to possess the highest reflective judgment scores, whereas the n_4 = 450 participants in cluster 4 exhibit the lowest. Clusters 2 and 3 consist of participants with slightly less extreme reflective judgment scores relative to clusters 1 and 4, respectively.

Strong cluster separation for S1 is also apparent in the profiling variables, particularly with respect to the high-school and college GPA measurements. Unfortunately, the profiling variable centroids do not have a readily interpretable linkage to the response variable means. For example, the largest high-school (u_3) and college (u_4) GPA averages are achieved in cluster 2, which corresponds to the second largest within-cluster average for reflective judgment. However, for cluster 1, which has the largest within-cluster average for reflective judgment, the average high-school and college GPA are relatively weak (third largest). Similarly, it is noted that cluster 4, which has the smallest within-cluster average for reflective judgment, has the second largest averages for high-school and college GPA scores. The only profiling variable that exhibits strong concordance with the reflective judgment means is endorsement scores for the written essay (u_2).

The bottom panel of Table 5 provides the cluster centroids for the holdout sample, which are computed after assigning each case in S2 to the same cluster as its nearest neighbor in S1. The relative rankings of cluster means for each variable are comparable for S1 and S2. The proportion of explained variation for the profiling variables in S2, $Z_{\mathbf{X}}(P_K)$ = .47998, represents only a modest decrease from the value obtained for S1, $Z_{\mathbf{X}}(P_K)$ = .57166. Unfortunately, the same cannot be said for reflective judgment. Whereas the explained variation for reflective judgment for S1 is $Z_{\mathbf{Y}}(P_K)$ = .51038, the corresponding value for S2 is only $Z_{\mathbf{Y}}(P_K)$ = .09237.

The reflective judgment averages for S2 are markedly different from those in S1 in two respects. First, there is the obvious observation that the cluster averages for S2 are noticeably smaller than those for S1, which reflects the substantial reduction in explained variation. Second, the relative ranking of the averages is different. For S1, the ranking of cluster averages for reflective judgment (from highest to lowest) is 1-2-3-4, whereas the corresponding ranking for S2 is 2-1-4-3. For S2, we observe that the reflective judgment scores are somewhat more commensurate with high-school (u_3) and college (u_4) GPA scores. The largest within-cluster averages for high-school GPA, college GPA, and reflective judgment are realized for cluster 2, whereas the smallest averages occur in cluster 3.

5.4. Analysis of other solutions on the estimated non-dominated set

The problems outlined in the previous subsection are not unique to the solution corresponding to $\alpha = .8$. Table 4 reveals that the $Z_{\mathbf{X}}(P_K)$ values for the holdout sample generally lag slightly below the $Z_{\mathbf{X}}(P_K)$ values for the training sample for $.3 \le \alpha \le 1.0$, whereas these measures are comparable for small values of α. In contrast, the $Z_{\mathbf{Y}}(P_K)$ values for the holdout sample are always quite poor (<.11 in all instances). Most importantly, for $.0 \le \alpha \le .9$, the $Z_{\mathbf{Y}}(P_K)$ values for the holdout sample are drastically less than those for the training sample, illustrating the inability to cross-validate the multiobjective programming solutions in this context. The key point here is that cross-validation problems prevail across nearly the full range of the non-dominated set. Consider, for example, the solution corresponding for $\alpha = .3$, which provides $Z_{\mathbf{Y}}(P_K) = .82785$ and $Z_{\mathbf{X}}(P_K) = .32700$ in the training sample (S1). This particular solution provides stronger explanation of the response variable at the expense of explained variation in the profiling variables. When classifying the cases in S2 by assigning them to the same cluster as their nearest neighbor in S1, the explained variation for the profiling variables is $Z_{\mathbf{X}}(P_K) = .27076$, but explained variation for the response variable is a dismal $Z_{\mathbf{Y}}(P_K) = .06087$.

6. Conclusion

There is no question that the application of multiobjective programming methods to problems of CDA can produce results that are numerically impressive with respect to objective criterion indices. In this paper, we have focused on a fundamental question: Do the impressive indices of fit realized for multiobjective problems necessarily translate into salient findings for quantitative analysts? For some multiobjective problems, we believe the answer to this question is ‘yes’, Multiobjective clustering procedures, for example, can yield better recovery of cluster structure (Steinley & Hubert, 2008), or facilitate the identification of complex structures that cannot be accommodated by any single criterion (Handl & Knowles, 2007; Handl *et al.*, 2007), Multiobjective programming also has pragmatic value for tie-breaking (Brusco & Cradit, 2004, 2005; Delattre & Hansen, 1980) in clustering applications.

In some instances, multiobjective programming can also be gainfully used when multiple sets of proximity matrices are available for the same set of objects. For example. Brusco (2002) used multiobjective procedures to provide a permutation of the digits (0, 1, 2, …, 9) that provided a good structural fit to each of four proximity matrices. We believe that

extreme caution is advised, however, when using multiobjective programming to obtain clustering solutions based on criteria for multiple data sets. This caution is especially important when clustering a sample of objects with the goal of *predicting* one or more response variable measurements upon classification of new objects based only on the profiling variable measurements. We have presented examples using synthetic and real data sets that demonstrate that explained variation for the response measures can be impressively raised in the training sample using multiobjective methods. Unfortunately, these results are obtained through an overfitting of the training sample, and the explained variation performance does not translate to objects in the holdout sample.

Our numerical examples in this paper have been presented within the context of biobjective clustering where the two relevant objective criteria are the percentage of explained variation for the profiling and response variables. These biobjective issues generalize to the multiobjective case where there are multiple sets of response variables, each with their own objective criterion component. Moreover, the same problems pervade the situation where within-cluster regression models are established to represent objective criteria for the response variables (Brusco *et al.*, 2003), The problems we have outlined for multiobjective clustering resemble those associated with clusterwise regression, as recently identified by Brusco *et al.* (2008). Specifically, overfitting the data and an inability to classify new cases to facilitate satisfactory explanation of response measures frequently plague both methods.

Our goal in writing this paper was not to stifle advancement of multiobjective programming methods in CDA. We believe there are many promising avenues of future research, including: (a) the development of improved methods for estimating non-dominated solutions and (b) identification of other applications in scaling and clustering, as well as other areas of CDA. At the same time, we believe it is imperative to understand the importance of cross-validation when using multiobjective programming for applied data analyses.

References

Arabie P, Hubert LJ. Combinatorial data analysis. Annual Review of Psychology. 1992; 43:169–203.

Brusco MJ. Identifying a reordering of the rows and columns of multiple proximity matrices using multiobjective programming. Journal of Mathematical Psychology. 2002; 46:731–745.

Brusco MJ. A repetitive branch-and-bound algorithm for minimum within-cluster sums of squares partitioning. Psychometrika. 2006; 71:347–363. [PubMed: 28197962]

Brusco MJ, Cradit JD. Graph coloring, minimum-diameter partitioning, and the analysis of confusion matrices. Journal of Mathematical Psychology. 2004; 48:301–309.

Brusco MJ, Cradit JD. Bicriterion methods for partitioning dissimilarity matrices. British Journal of Mathematical and Statistical Psychology. 2005; 58:319–332. [PubMed: 16293203]

Brusco MJ, Cradit JD, Stahl S. A simulated annealing heuristic for a bicriterion problem in market segmentation. Journal of Marketing Research. 2002; 39:99–109.

Brusco MJ, Cradit JD, Steinley D, Fox GL. Cautionary remarks on the use of clusterwise linear regression. Multivariate Behavioral Research. 2008; 43:29–49. [PubMed: 26788971]

Brusco MJ, Cradit JD, Tashchian A. Multicriterion clusterwise regression for joint segmentation settings: An application to customer value. Journal of Marketing Research. 2003; 40:225–234.

Brusco MJ, Stahl S. An interactive approach to multiobjective combinatorial data analysis. Psychometrika. 2001a; 66:5–24.

Brusco MJ, Stahl S. Compact integer-programming models for extracting subsets of stimuli from confusion matrices. Psychometrika. 2001b; 66:405–420.

Brusco MJ, Stahl S. Bicriterion seriation methods for skew-symmetric matrices. British Journal of Mathematical and Statistical Psychology. 2005a; 58:333–343. [PubMed: 16293204]

Brusco, MJ., Stahl, S. Branch-and-bound applications in combinatorial data analysis. New York: Springer; 2005b.

Brusco, MJ., Stahl, S., Cradit, JD. Unpublished working paper. Florida State University; 2002. Multiobjective multidimensional scaling in the city-block metric.

Burgess ES, Brown RA, Kahler CW, Niaura R, Abrams DB, Goldstein MG, et al. Patterns of change in depressive symptoms during smoking cessation: Who's at risk for relapse? Journal of Consulting and Clinical Psychology. 2002; 70:356–361. [PubMed: 11952193]

Caliński T, Harabasz J. A dendrite method for cluster analysis. Communications in Statistics. 1974; 3:1–27.

de Craen S, Commandeur JJF, Frank LE, Heiser WJ. Effects of group size and lack of sphericity on the recovery of clusters in *K*-means cluster analysis. Multivariate Behavioral Research. 2006; 41:127–145. [PubMed: 26782907]

Delattre M, Hansen P. Bicriterion cluster analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1980; 2:277–291. [PubMed: 21868904]

DeSarbo WS, Grisaffe D. Combinatorial optimization approaches to constrained market segmentation: An application to industrial market segmentation. Marketing Letters. 1998; 9:115–134.

Dimitriadou E, Dolničar S, Weingessel A. An examination of indices for determining the number of clusters in binary data sets. Psychometrika. 2002; 67:137–160.

Ehrgott, M. Multicriteria optimization. 2. Berlin: Springer; 2005.

Erhgott M, Gandibleaux X. A survey and annotated bibliography on multiobjective combinatorial optimization. OR Spektrum. 2000; 22:425–460.

Ferligoj A, Batagelj V. Direct multicriteria clustering algorithms. Journal of Classification. 1992; 9:43–61.

Hand, DJ. Discrimination and classification. New York: Wiley; 1981.

Handl J, Kell DB, Knowles J. Multiobjective optimization in bioinformatics and computational biology. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2007; 4:279–292. [PubMed: 17473320]

Handl, J., Knowles, J. On semi-supervised clustering via multiobjective optimization. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006); New York: ACM Press; 2006. p. 1465-1472.

Handl J, Knowles J. An evolutionary approach to multiobjective clustering. IEEE Transactions on Evolutionary Computation. 2007; 11:56–76.

Haselager GJT, Cillessen AHN, Van Lieshout CFM, Riksen-Walraven JMA, Hartup WW. Heterogeneity among peer-rejected boys across middle childhood: Developmental pathways of social behavior. Developmental Psychology. 2002; 38:446–456. [PubMed: 12005387]

Howard, RN. Classifying a population into homogeneous groups. In: Lawrence, JR., editor. Operational research and the social sciences. London: Tavistock; 1966. p. 585-594.

Hubert L. Min and max hierarchical clustering using asymmetric proximity measures. Psychometrika. 1973; 38:63–72.

Hubert LJ. Some applications of graph theory to clustering. Psychometrika. 1974a; 39:283–309.

Hubert LJ. Spanning trees and aspects of clustering. British Journal of Mathematical and Statistical Psychology. 1974b; 27:14–28.

Hubert, L., Arabie, P., Meulman, J. Combinatorial data analysis: Optimization by dynamic programming. Philadelphia, PA: Society for Industrial and Applied Mathematics; 2001.

Hubert, L., Arabie, P., Meulman, J. The structural representation of proximity matrices with MATIAB. Philadelphia, PA: Society for Industrial and Applied Mathematics; 2006.

Hubert LJ, Golledge RG. Matrix reorganization and dynamic programming: Applications to paired comparisons and unidimensional seriation. Psychometrika. 1981; 46:429–441.

Johnson SC. Hierarchical clustering schemes. Psychometrika. 1967; 32:241–254. [PubMed: 5234703]

Kitchener, KS., Jensen, LL., Wood, PK. OERI Award #R309F700860. Denver, CO: University of Denver; 1999. The development and assessment of complex problem solving in college age students: Final Report.

Krieger AM, Green PE. Modifying cluster-based segments to enhance agreement with an exogenous response variable. Journal of Marketing Research. 1996; 33:351–363.

Law, MHC., Topchy, AP., Jain, AK. Proceedings of the IEEE Computer Society on Computer Vision and Pattern Recognition. Vol. 2. Washington, DC: IEEE Computer Society; 2004. Multiobjective data clustering; p. 424-430.

Lawler EL. A comment on minimum feedback arc sets. IEEE Transactions on Circuit Theory. 1964; 11:296–297.

MacQueen, JB. Some methods for classification and analysis of multivariate observations. In: Le Cam, LM., Neyman, J., editors. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability; Berkeley, CA: University of California Press; 1967. p. 281-297.

MathWorks, Inc. Using MATLAB (Version 6). Natick, MA: The MathWorks, Inc; 2002.

Ripon, KSN., Tsang, C-H., Kwong, S., Ip, M-K. Multi-objective evolutionary clustering using variable-length real jumping genes genetic algorithm. Proceedings of the 18th International Conference on Pattern Recognition; Washington, DC: IEEE Computer Society; 2006. p. 1200-1203.

Robinson WS. A method for chronologically ordering archeological deposits. American Antiquity. 1951; 16:293–301.

Steinley D, Brusco MJ. A new variable weighting and selection procedure for *K*-means clustering. Muitivariate Behavioral Research. 2008; 43:77–108.

Steinley D, Hubert LJ. Order constrained solutions in *K*-means clustering: Even better than being globally optimal. Psychometrika. 2008; 73:647–664.

Steuer, RE. Multiple criteria optimization: Theory, computation and application. New York: Wiley; 1986.

Strehl A, Ghosh J. Cluster ensembles - a knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research. 2002; 3:583–617.

Ulungu E, Teghem J. The two-phase method: An efficient procedure to solve biobjective combinatorial optimization problems. Foundations of Computing and Decision Sciences. 1994; 20:149–165.

Wood PK, Kitchener KS, Jensen L. College students' concepts of belief justification regarding open-ended controversies. 2005 Manuscript submitted for publication.

Zeleny, M. Multicriteria decision making. New York: McGraw-Hill; 1982.

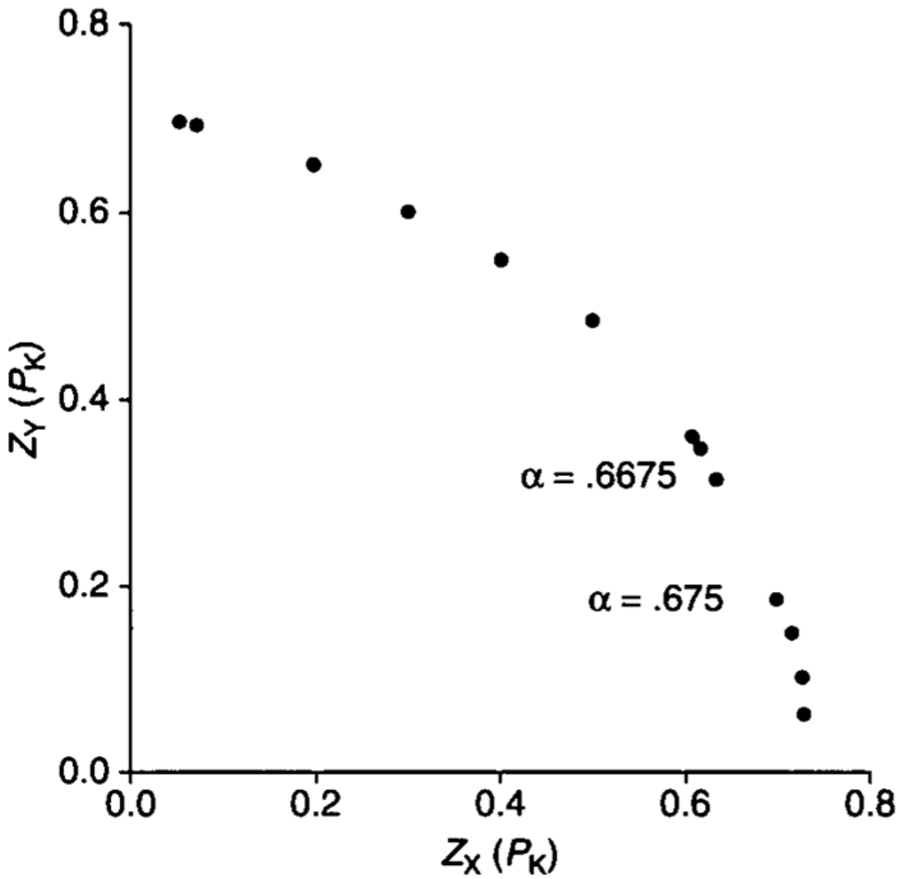
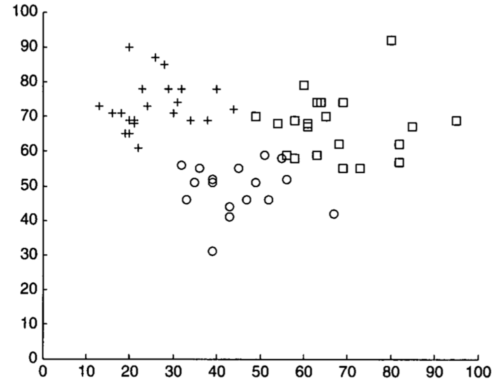


Figure 1.
Points on the efficient frontier obtained using the weighting schemes in Table 1.



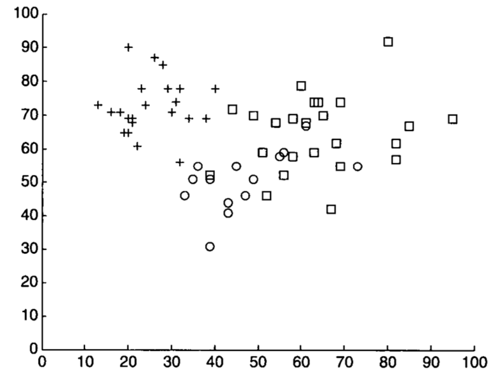
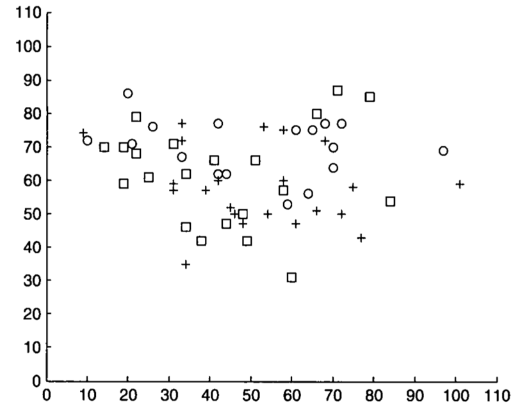


Figure 2.

Scatterplots for the profiling variables for three-cluster bicriterion partitions obtained using α = I (top panel) and α = .6675. (bottom panel). The horizontal axis is v_1, and the vertical axis is v_2.



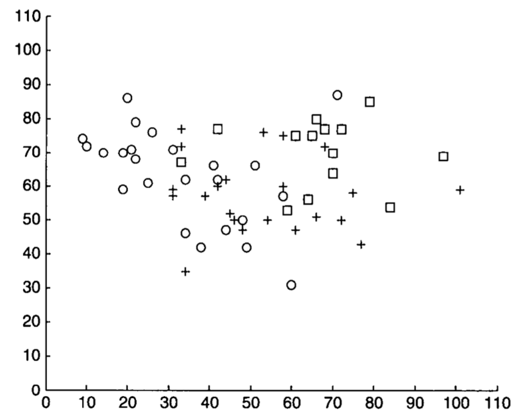


Figure 3.
Scatterplots for the response variables for three-cluster bicriterion partitions obtained using $\alpha = 1$ (top panel) and $\alpha = .6675$, (bottom panel).

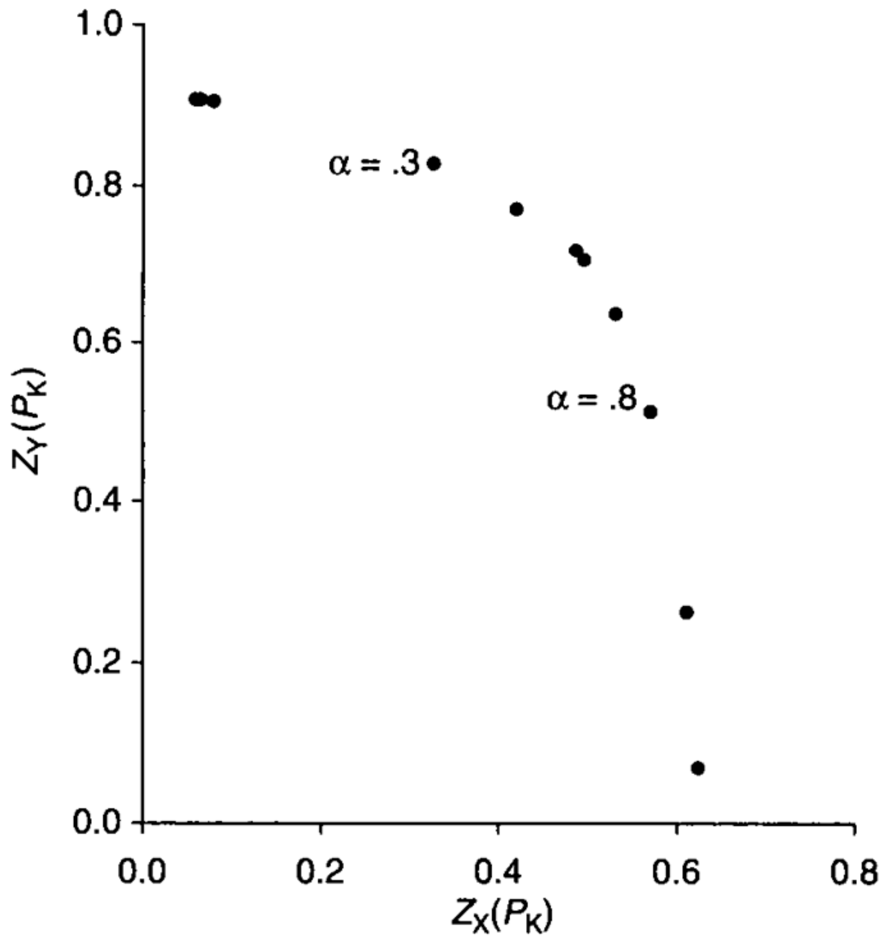


Figure 4.
Points on the estimated efficient frontier for the reflective judgment data, which were obtained using the weighting schemes in Table 3.

Table 1

Explained variation for non-dominated solutions for the sample synthetic data using various values of α

α	$Z_X(P_K)$	$Z_Y(P_K)$	Percentage decrease in $Z_X(P_K)$[a]
1.0000	.72988	.06050	.00
.9000	.72789	.09942	.27
.8000	.71622	.14769	1.87
.7000	.71622	.14769	1.87
.6750	.69916	.18380	4.21
.6675	.63522	.31387	12.97
.6500	.61780	.34644	15.36
.6000	.60905	.35911	16.55
.5000	.49954	.48266	31.56
.4000	.40264	.54740	44.83
.3250	.30214	.59983	58.60
.3000	.19447	.64952	73.36
.2000	.07303	.69122	89.99
.1000	.05474	.69429	92.50
.0000	.05474	.69429	92.50

[a]The percentage decrease in explained variation for the profiling variables relative to the optimal value of $Z_{\mathbf{X}}(P_K)$ obtained when $\alpha = 1$.

Table 2

The top panel contains the cluster profiles for the bicriterion clustering of the *sample* obtained using α = .6675. The bottom panel contains the results for the *population* when each object is assigned to the same cluster as its nearest neighbor in the sample

		Profiling variables $Z_X(P_K) = .63522$		**Response variables $Z_Y(P_K) = .31387$**	
Cluster (k)	**n_k**	**$u = 1$**	**$u = 2$**	**$v = 1$**	**$v = 2$**
1	22	25.32	73.64	53.14	57.68
2	24	63.92	64.58	33.67	63.13
3	14	46.71	50.71	66.43	63.98
		Profiling variables $Z_X(P_K) = .55478$		Response variables $Z_Y(P_K) = .00056$	
Cluster (k)	n_k	$u = 1$	$u = 2$	$v = 1$	$v = 2$
1	1,089	28.07	71.39	50.52	62.88
2	1,749	64.99	64.45	49.71	63.11
3	762	46.21	49.67	51.06	62.91

Table 3

Correlation matrix for the reflective judgment data

	v_1	u_1	u_2	u_3	u_4
v_1	1.000				
u_1	.2763	1.000			
u_2	.4872	.1707	1.000		
u_3	.1603	.4296	.1325	1.000	
u_4	.1982	.4462	.1700	.5451	1.000

Table 4

Explained variation for estimated non-dominated solutions for the reflective judgment training sample (S1) using various values of α, and the corresponding explained variation when used to classify the holdout sample (S2)

	Training sample (S1)			Holdout sample (S2)	
α	**$Z_X(P_K)$**	**$Z_Y(P_K)$**	**Percentage decrease in $Z_X(P_K)$[a]**	**$Z_X(P_K)$**	**$Z_Y(P_K)$**
1.0000	.62425	.06919	.00	.55054	.08252
.9000	.61274	.26243	1.84	.53669	.10780
.8000	.57166	.51038	8.42	.47998	.09237
.7000	.53240	.63554	14.71	.46083	.08694
.6000	.49762	.70486	20.28	.42920	.07217
.5000	.48763	.71747	21.89	.41757	.07073
.4000	.42122	.77097	32.52	.35697	.06885
.3000	.32700	.82785	47.62	.27076	.06087
.2000	.07977	.90420	87.22	.08660	.05313
.1000	.06596	.90673	89.43	.07771	.05165
.0000	.06004	.90709	90.38	.07082	.05064

[a]The percentage decrease in explained variation for the profiling variables relative to the optimal value of $Z_{\mathbf{X}}(P_K)$ obtained when α = 1.

Table 5

The top panel contains the cluster profiles for the bicriterion clustering of reflective judgment dataset S1 obtained using α = .8. The bottom panel contains the cluster profiles for the holdout sample S2 when each object is assigned to the same cluster as its nearest neighbor from the training sample, S1

		Profiling variables $Z_X(P_K)$ = .57166				**Reflective judgment $Z_Y(P_K)$ = .51038**
Cluster (*k*)	n_k	u_1	u_2	u_3	u_4	v_1
1	435	−.155	.503	−.884	−.248	1.285
2	527	.999	.412	2.275	1.046	.908
3	397	−.803	−.443	− 3.035	− 1.092	−1.015
4	450	−.312	−.578	.868	−.022	− 1.410
		Profiling variables $Z_X(P_K)$ =.47998				Reflective judgment $Z_Y(P_K)$ = .09237
Cluster (*k*)	n_k	u_1	u_2	u_3	u_4	v_1
1	487	−.344	.458	−.776	−.345	.205
2	542	1.385	.394	1.799	1.208	.497
3	344	−1.084	−.473	− 2.560	−1.385	−.514
4	436	−.339	−.559	.728	.065	−.375