T E C H N I C A L

R E P O R T

0693

A COMPARISON OF TWO APPROACHES TO INVESTIGATE THE CATEGORICAL VS. DIMENSIONAL NATURE OF PSYCHOLOGICAL PHENOMENA

HIDEGKUTI, I. and P. DE BOECK

I A P S T A T I S T I C S N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Running head: Comparison of DIMCAT and taxometrics

A comparison of two approaches to investigate the categorical vs. dimensional nature of psychological phenomena

István Hidegkuti and Paul De Boeck

K.U. Leuven

Belgium

Abstract

To find out the categorical or dimensional nature of the latent structure of psychological data, several methods are available, of which taxometrics is the most well-known and most widely used, mainly in the field of psychopathology. Taxometrics stands for a series of methods that are based on Coherent Cut Kinetics. A recent development to address the continuity vs. discontinuity controversy is the Dimension/Category (DIMCAT) framework that is based on item response theory. A simulation study with both rating-scale and continuous data is presented to compare the performance of both methods under various circumstances. Of the taxometric methods MAXCOV, MAMBAC and L-Mode are used. It turns out that DIMCAT and L-Mode perform very well, while MAXCOV and MAMBAC deliver a rather poor performance when using rating-scale data. Furthermore, it is also established that among the several measures suggested in the taxometric literature, the Comparison Curve Fit Index seems to be the best index for decision making about the underlying structure.

Introduction

Taxometric methods are developed to investigate whether a phenomenon under study, such as depression or giftedness, is categorical in comparison with normality. The methodology in question requires a hypothesized category, such as depression or giftedness, a sample of people from the hypothesized category and a substantial proportion of these among the total sample, and finally, a set of indicator variables with validity for the hypothesized category. That the category is hypothesized means that the operational classification of persons into a category does not imply that the category is a genuine category. The alternative to a genuine category is a continuum or dimension with a cutpoint as the only basis for a classification. Taxometric methods are aimed at differentiating between a dimensional structure and a genuine categorical structure, starting from a data set with persons described in terms of indicator variables.

In order for the taxometric methodology to work one needs (a) a hypothetical focus category, called the taxon group, which can be differentiated from a contrast category, called the complement group, commonly the category of normality; (b) a sample of persons belonging to the focus category or contrast category, with a substantial proportion of both; (c) a small set of indicator variables with validity for the differentiation between the two categories (the focus category and the contrast category), with the indicator variables being continuous variables (ideally), ordered-category variables or binary variables (possible in practice). (d) It is not required that the actual membership of the persons in the two categories is known, it suffices that a substantial proportion of both categories is present in the sample. (e) The methodology is not based on a specific model for the data. However, for the answer obtained through an application of the methodology to be categorical, the distance between the categories on the indicator variables needs to be above 1.25 SDs (Meehl, 1995) which is considered to be a large effect size in psychology (Cohen, 1988).

A common type of data sets in psychology stems from the application of a questionnaire measuring a trait or attitude, presented to a sample of groups, a focus group and a control group. The issue may then be whether the trait or attitude shows a categorical or a dimensional structure. For example, using a depression questionnaire, one of the many available, for depressed persons and normals, one may wonder whether depression is a continuum or a category when compared to normality. In fact, each time such data are available, one may investigate the dimensional versus categorical nature of the differentiation made by the questionnaire.

The DIMCAT approach is developed with this kind of application in mind, but with a broader domain of application if wanted. (a) More than two categories can be investigated at the same time. (b) The indicator variables may vary as to their validity for the categories (positive, negative, none), but a substantial proportion of them needs to have at least some validity, no matter in which sense (positive or negative). (c) The indicator variables may be a few or many, and they may be of any kind: binary, rating scale, continuous, etc., although in the original formulation they are binary. (d) The categories under study may be observed (e.g., a diagnosis) or unobserved, and hence, latent.

Because its generality, DIMCAT should be able to perform reasonably well also in situations considered ideal for taxometrics. An important restriction of DIMCAT is that it requires a model to start from. In the original formulation, the two parameter item response model (2PL; Birnbaum, 1968) is used as the basic model. Also models for rating-scale and even for continuous data can be used, of the linear type for the latter, and multidimensional models can be used as well, but always a certain type of model needs to be chosen, and the answer to the issue of categoricality may depend on the model that is chosen. The two approaches will be described more formally and more in detail in separate sections on taxometrics and DIMCAT.

In order to compare both approaches on their performance for differentiating between a categorical and a dimensional structure, two studies are set up. In the first study, data are generated from a rating-scale item response model with 11 indicator variables. This is a typical case when a questionnaire is used for a specific trait or attitude. These kinds of data may be considered DIMCAT data, as they represent the ideal case for DIMCAT. For reasons of comparison both DIMCAT and taxometrics will be applied in this study.

In the second study, data generated by Meehl and Yonce (1994) will be used, with four continuous indicator variables. These data were used by Meehl and Yonce (1996, 1994) to validate the taxometric methods, and these methods worked rather well. These kinds of data may be considered taxometric data, as they represent the kind of data taxometric methods are made for. It is an interesting issue how well DIMCAT will perform when applied to these data.

In sum, the design of the two studies taken together is a two by two design: DIMCAT versus taxometric data analyzed with a DIMCAT versus taxometric methodology.

Taxometrics

Of the numerous taxometric methods MAXCOV, MAMBAC and L-Mode will be applied in this paper and hence presented here briefly.

MAXCOV

Being the first published taxometric method, MAXCOV is the most well-known and most popular technique of the taxometric family up to now. MAXCOV (e.g., Meehl, 1973; Meehl & Yonce, 1996; Waller & Meehl, 1998) is based on the General Covariance Mixture Theorem (Waller & Meehl, 1998) which describes the covariance of two indicator variables (denoted by x and y here) in a population consisting of two categories (or groups, classes etc.) as follows:

$$cov(x,y) = Pcov_t(x,y) + Qcov_c(x,y) + PQ(\bar{x_t} - \bar{x_c})(\bar{y_t} - \bar{y_c}) \tag{1}$$

Where $cov(x,y)$ is the covariance of x and y in the full population, and the subscripts t and c are indices for the taxon group (i.e., the group or class in question) and the complement group, respectively. P is the class size or proportion (also called base rate) of the taxon group and Q is the class size of the complement group ($Q = 1 - P$), whereas $(\bar{x_t} - \bar{x_c})(\bar{y_t} - \bar{y_c})$ is the cross-product of the mean differences between the two categories.

In the ideal case, the indicator variables covary only in a mixed set of cases, and thus between categories, but not within categories, so that the General Covariance Mixture Theorem reduces to:

$$cov(x,y) = PQ(\bar{x_t} - \bar{x_c})(\bar{y_t} - \bar{y_c}) \tag{2}$$

The assumption of zero within-class covariance is a very severe one, but fortunately simulation studies have shown that MAXCOV is robust even to sizeable violations of this assumption (Meehl & Yonce, 1996). Meehl (1995) suggested that within-category correlations as high as .30 are still acceptable. Furthermore, he claimed that even higher within-category correlations might be acceptable if the correlations are similar in the two categories. However, in the taxometrics literature .30 is used as an upper limit for within-category correlations.

In practice, there are three indicator variables at a time involved in MAXCOV: two indicator variables as output variables and the third indicator variable as an input variable. All these three indicators are assumed to be valid for the differentiation of the two categories. In order for the method to be applied, the input variable is divided into subsequent sections and in each section the covariance of the two output indicators is calculated and plotted. In principle, at one extreme of the input variable there are (almost) only taxon group members, whereas at the other extreme there are (almost) only complement group members, with different mixing proportions in the sections in between. According to Equation 2, the covariance of the output variables would approach zero at the extremes, and would increase towards the interval where the proportions are equal. For a truly categorical structure the plot of the covariances shows a peaked curve, the peak appearing in the section where the proportions of the two groups approach equality in the best way. Based on Equation 2, the base rate of a taxon can be calculated from the location of the interval with the maximal covariance (called the Hitmax interval).

When applying MAXCOV, the indicator variables can be used in all possible output, output, input triplets, which yields $I*(I-1)*(I-2)/2$ possible MAXCOV analyses (with I being the number of indicator variables). As all other taxometric methods, MAXCOV is heavily based on the visual inspection of the resulting plots of the analyses. Fortunately, also several consistency tests are developed to corroborate possible decisions from a visual inspection of the plots (see later).

MAMBAC

MAMBAC (Mean Above Minus Below A Cut; Meehl & Yonce, 1994) is the simplest taxometric method in which two indicator variables are used at a time, one being the input indicator and the other the output indicator. The cases are sorted along the input indicator that is assumed to separate the hypothetical groups in a valid way (e.g., depressed and nondepressed). Cuts are made along the input indicator and the mean score on the output variable is calculated for the cases falling below the cut as well as for the cases falling above the cut, and then the difference between the two means is plotted. In the next step the cut is moved upwards along the input indicator, and the same procedure is repeated a number of times, so that a MAMBAC graph of the plotted differences of the means is obtained. In case there is an underlying categorical entity, the MAMBAC curve is supposed to show a clear peak, or hump. In case there is an underlying dimensional structure, the MAMBAC curve is supposed to have a dish shape. The MAMBAC procedure can be repeated with all possible pairs of input and output variables.

L-Mode

L-Mode (short for Latent Mode; Waller & Meehl, 1998) is a relatively new taxometric method that is based on factor analysis and more specifically on the notion that the factors can represent not only latent continua, but also latent categories. L-Mode is applied on the whole set of indicator variables which are all treated equally. The distinction between the input variable and output variables does not apply.

L-Mode is based on the notion that a factor can not only represent an underlying continuum but also an underlying dichotomous variable as well. In L-Mode, a common factor analysis is performed and the resulting individual factor scores on the first factor are plotted. In case of a categorical structure, the factor score plot is expected to show bimodality, whereas in case of a dimensional structure the factor score plot is expected to show unimodality or a rather flat distribution without clear peaks. In case of bimodality the estimates of the base rate can be calculated from the location of either of the two modes. Here again, the visual inspection of the plots plays an important role in the decision making, but also consistency tests may be (and should be) employed.

Generating empirical sampling distributions to help interpret results

In the taxometric approach the decision about dimension-likeness vs. category-likeness is primarily based on the visual inspection of the taxometric graph. However, taxometric graphs most of the time are not quite easy to interpret just by visual inspection, as many times the graphs are rather unclear, or ambiguous. To address this problem, John Ruscio and his colleagues (J. Ruscio, Ruscio, & Meron, in press) developed a method to help interpret taxometric results. Their method is an application of the bootstrap methodology. The essence of their method is to use important features of the data (e.g., indicator mean, skew, sample size, correlation structure, etc.) to generate artificial data sets based on a dimensional underlying structure and a categorical underlying structure, while sustaining the important data characteristics. The generated artificial data sets are submitted to the same taxometric analyses than the empirical data, and the resulting taxometric graphs are compared to find out which structure is more suitable to describe the empirical data.

Measures to evaluate the category- versus dimension-likeness of different phenomena in taxometric studies

Although in the taxometric research the emphasis has been on the visual inspection of the graphs resulting from the various techniques, there were several attempts to find some measures that could be used to the decision making about the underlying nature of the investigated phenomena. From the beginning, consistency tests were strongly recommended to reassure the decisions about the underlying nature. Most studies employed consistency tests that compare the results across different taxometric procedures. As an alternative, the consistency of base rate estimates can be used as a criterion. In MAXCOV and MAMBAC the built-in consistency test is based on the variability of the base rate estimates stemming from the analyses of the different triplets, or pairs of indicators for MAXCOV and MAMBAC, respectively. In case of L-Mode, the built-in consistency test is the difference between the base rate estimates stemming from the estimation based on the left mode and the right mode of the L-Mode graph. Quantitative consistency measures such as the standard deviation of the base rates or the difference between the base rates are used as a decision aid. Unstable base rates must be considered as counterevidence for a categorical structure.

Another measure that can be used in taxometric decision-making is the Goodness of Fit Index (GFI) which was described by Jöreskog and Sörbom (1988) for Structural Equation Models. The GFI is a measure to express the consistency of the observed variance-covariance matrix and the variance-covariance matrix predicted from the results of a given analysis. The use of this index in taxometrics was proposed by Waller and Meehl (1998), who suggested .90 as a threshold, stating that a categorical structure is unlikely to produce a GFI value below .90, whereas a dimensional structure is unlikely to result in a GFI value above .90 based on a small simulation study. However, other simulation studies (e.g., Haslam & Cleland, 2002) suggested that the GFI is a rather poor index of the underlying structure.

The graphs of simulated sampling distributions can also provide a measure to help the taxometric decision making. The taxometric graph of the empirical data can be compared with the graphs of the simulated categorical and simulated dimensional data, not only visually but also using a similarity measure. Ruscio et al. (in press) suggested the Comparison Curve Fit Index (CCFI) as a measure to compare the similarity of the graphs.

The Dimension/Category framework

A recent development in the methodology to investigate the categorical vs. dimensional structure of individual differences (or differences in general) is the DIMCAT framework (De Boeck et al., 2005). DIMCAT is based on item response modeling (IRT), and it provides a broad and elaborated framework for the distinction between categories and dimensions. Given a hypothesized pair of categories (e.g., depression and normality), the issue is how much category-like or how much dimension-like the distinction is.

In the DIMCAT framework, various types of categories are described. Distinctions are basically made along two axes. The first axis differentiates *qualitative differences* (i.e., the differences cannot be understood as referring to a degree of something) and *quantitative differences* (i.e., the differences are differences in the degree of something; e.g., category B having more of the same category A has). The hypothesized categories can be homogeneous or heterogeneous, and they can be qualitatively different or only quantitatively different. The heterogeneity versus homogeneity axis (second axis) refers to a within-category feature, whereas the latter refers to a between-category feature. The heterogeneity versus homogeneity axis differentiates between *heterogeneity* (systematic variance) and *homogeneity* (no systematic variance) within categories. The axes refer to the latent structure and not the manifest variables. For example, homogeneity is not homogeneity in terms of manifest variables, but in terms of the latent variable(s). In terms of the manifest variables, it means that they are not correlated within categories, which is the basic assumption of taxometrics and of the common type of latent class analysis.

Furthermore, there are two types of qualitative differences: differences in indicator location, and differences in indicator relevance, to be explained in the following. Differences in *indicator location* refer to a lack of parallelism in (the ordering of) the indicator means from one category to another. For example, the prevalence of symptoms may show a different pattern depending on the category.

Differences in *indicator relevance* mean that depending on the category other indicators are more relevant in determining a degree within the category (if there is such a degree). Differences in relevance imply that the indicator values contribute to the within-category degree in a different way depending on the category and hence that a different quality is represented depending on the category. Differences in relevance imply also a different correlational structure within the categories and therefore refer to what is commonly seen as lack of factorial equivalence. However, since the full factor model also explains the means, factorial equivalence in a broader sense also includes the equivalence of the means (e.g., Meredith, 1993). For this reason De Boeck et al. (2005) deem the equivalence of the factor loadings as factorial equivalence in the limited sense.

When the between-category differences are quantitative, a nested (secondary) axis differentiates between abrupt and smooth differences. Abrupt differences refer to a clear multimodality in the distribution along a common underlying dimension, whereas for smooth differences there is a large overlap, so that the sum of the distributions of the categories appears as unimodal.

The alternative for a categorical structure is a dimensional structure. It is a structure with only quantitative differences among the persons, and with no other basis for categorization than an arbitrary cutpoint.

In the DIMCAT framework, the category versus dimension distinction is seen as multifaceted and gradual. (Meehl also shared this idea stating that "taxonicity itself is not taxonic" (Meehl, 1979, p. 566), but in taxometrics this issue was not explicated any further with the exception of J. Ruscio and Ruscio (2002)). Following DIMCAT, a structure can be categorical in different senses, and to different degrees. Qualitative differences between the categories are more category-like than quantitative differences, and the larger the qualitative differences are, the more category-like the structure is. Homogeneity within the categories is more category-like than heterogeneity, and the smaller the within-category variance is, the more categorical the structure is. Abrupt differences between categories are more category-like than smooth differences, and the more abrupt the differences are, the more categorical the structure is.

To find out the true nature of a given phenomenon, a series of item response models are fitted, and then the decision about what kind of category-likeness the phenomenon in question shows, is based on the comparison of the goodness-of-fit statistics of the various models.

To present the models, we assume the indicator data are of a rating-scale type.

Given M response alternatives indexed by m ($m = 1, \ldots, M$), the Modified Graded Response Model (MGRM; Muraki, 1990) defines the cumulative (conditional) probability of alternative m ($m = 1, \ldots, M$) or higher for indicator i, given the category c person p belongs to as:

$$P(Y_{pi} \geqslant m|\theta_{pc}, c_p) = \frac{\exp(\alpha_{ic}(\theta_{pc} - \beta_{ic} + \omega_m + \gamma_c))}{1 + \exp(\alpha_{ic}(\theta_{pc} - \beta_{ic} + \omega_m + \gamma_c))} \tag{3}$$

and the probability of responding in response category m is:

$$P(Y_{pi} = m|\theta_{pc}, c_p) = P(Y_{pi} \geqslant m|\theta_{pc}, c_p) - P(Y_{pi} \geqslant m + 1|\theta_{pc}, c_p), \tag{4}$$

except that for the lowest rating-scale response alternative $P(Y_{pi} \geqslant m|\theta_{pc}, c_p) = 1$ and for the highest one $P(Y_{pi} \geqslant m + 1|\theta_{pc}, c_p) = 0$, because $M + 1$ is not possible.

In Equations 3 and 4, Y is the response, α is the indicator discrimination parameter (indicating relevance), θ is the latent dimension of category c ($c = 1, \ldots, C$) along which persons p ($p = 1, \ldots, P$) are normally distributed (or in other words, it is the category-specific random effect), β is the location parameter of indicator i ($i = 1, \ldots, I$) in category c (together, the βs define the profile), γ is the category (group) difference parameter, and ω is the response alternative threshold parameter for alternative m (but ω_1 is irrelevant because $P(Y_{pi} \geqslant 1) = 1$). It is assumed that $\theta_{pc} \sim N(0, \sigma^2)$. For reasons of identification one of the $m - 1$ relevant thresholds must be set equal to 0.

The MGRM is the cumulative logit equivalent of the Rating Scale Model (RSM; Andrich, 1978), which is based on adjacent logits. The MGRM cuts the distribution over a latent continuum into sections corresponding to the values on the rating scale. The MGRM is a model for rating scale data, for the case it is reasonable to assume that the distances between the response alternatives of the rating scale are the same for all indicators, but not necessarily equal along the scale (the same scale with possibly unequal distances). One may think of a continuum with – depending on the indicator – a series of thresholds or cutoffs on the one hand, and a moving normal distribution, one for each person, so that for each pair of a person and an indicator, the thresholds or cutoffs determine the probabilities of the response alternatives.

Other models can be used instead of the one in Equation 3, but it seems a reasonable assumption that the scale does not depend on the indicator when the same response alternatives are used (Embretson & Reise, 2000). Therefore, an alternative to the MGRM is the Rating Scale Model (RSM; Andrich, 1978) which has in common with the MGRM that the category distances do not change from indicator to indicator, but the driving principle is that not cumulative log odds are modeled as in Equation 3, but adjacent log odds, preferring response alternative m on response alternative $m-1$.

Note that in the original paper on DIMCAT (De Boeck et al., 2005) c_p is a manifest and thus observed variable, but this is not a necessary condition. Here, c_p will be considered to be a latent variable, meaning that the categories and category memberships are not observed.

The model in Equation 3 and its restricted variants will be used here to make the distinctions along the two main axes. The most general model (Equation 3) is called the QUAL1&2-HET model, with QUAL referring to qualitative differences, 1 and 2 meaning that both types of qualitative differences (lack of discrimination equivalence and lack of location equivalence, respectively) are present, whereas HET means that there is within-category heterogeneity. As a next step the discriminations are set to be equal across groups resulting in the QUAL2-HET model (qualitative differences are restricted to location nonequivalence). Restricting also the locations to be equal across groups yields the QUAN-HET model, where only quantitative differences are taken into account, but not qualitative differences. The QUAL2-HET and the QUAN-HET models have their counterparts in models with within-group homogeneity, called the QUAL2-HOM and the QUAN-HOM models, respectively. A QUAL1&2-HOM model does not make sense because discriminations imply heterogeneity. A purely dimensional model, called the *uniform MGRM*, is defined by Equation 3 without any category-specific parameters, implying that the same MGRM applies to all persons. In other words stated, there is only one category and all persons belong to this category.

In order to compare the performance of the above-described two approaches, taxometrics and DIMCAT, two simulation studies are planned. In Study 1 rating-scale data are used, which can be considered the most typical data type in questionnaires used for psychological assessment (e.g., Beck Depression Inventory, Beck, Steer, & Brown, 1996; Center for Epidemiologic Studies Depression Scale, Radloff, 1977; NEO PI-R, Costa & McCrae, 1992; Rosenberg Self-Esteem Scale, Rosenberg, 1965). In the second study, simulated data sets with continuous indicator variables were employed that were used earlier to evaluate taxometric procedures. The data type in the first study can be considered as more suited for the DIMCAT approach, although rating-scale and even dichotomous data are not uncommon in the taxometric practice (e.g., Beach & Amir, 2003; Hankin, Fraley, Lahey, & Waldman, 2005; A. M. Ruscio & Ruscio, 2002). Note that recently the trend is to use (quasi-) continuous data or at least rating-scale data with a relatively large number (7-10) of response alternatives in taxometric studies (e.g., Franklin, Strong, & Greene, 2002; Marcus, John, & Edens, 2004; Rothschild, Cleland, Haslam, & Zimmermann, 2003). As far as the continuous data in Study 2 are concerned, these are clearly suited for taxometric analyses, but their suitability for DIMCAT analyses may not be that clear, as earlier it was stated that DIMCAT is based on IRT. For continuous data, it is quite natural to use the linear model, and hence, to replace Equation 3 with

$$Y_{pi}|c_p = \alpha_{ic}\theta_{pc} - \beta_{ic} + \gamma_c + \epsilon_{pi} \tag{5}$$

with β_{ic} now being the mean of indicator i in category c, and ϵ_{pi} being a normally distributed error term which is independent of p, i and c. The other symbols have the same meaning as in Equation 3.

Study 1

Data generation

Common features of the data sets

The data sets used for the analyses were generated based on the model in Equation 3. All data sets consisted of eleven indicator variables ($I = 11$) each of which have four response alternatives on a rating scale ($M = 4$), with a total sample size of 400 ($N = 400$), and 200 in each of both groups. The discrimination parameters are fixed to 1 ($\alpha_{ic} = 1$ for all i and c), the indicator locations (β_{ic}) range from -2 to 2 by steps of 0.4, the person category difference parameter (γ_c) is fixed to zero for the reference category, and is 0 or 2 for the other category (see later), and the response alternative thresholds are -1, 0 and 1 for ω_2, ω_3 and ω_4, respectively. The implication is that the item location coincides with the second threshold of the item.

Design factors

There were three factors in the data generation design, varied in an orthogonal way. The first factor was the value of the category difference parameter for the second group taking the values of either 0 or 2 ($\gamma_2 = 0$ or $\gamma_2 = 2$), so that the second class differs from the first in a quantitative way when $\gamma_2 = 2$ or does not differ from the first when $\gamma_2 = 0$, independent of whether the categories also differ in a qualitative way or not.

The second factor was the relation of the location parameters across the two classes. The same 11 values were used in all three cases, but the location parameters were either identical in the two groups ($r_{\beta_{i1}\beta_{i2}} = 1$), or uncorrelated ($r_{\beta_{i1}\beta_{i2}} = 0$), or they were perfectly negatively correlated ($r_{\beta_{i1}\beta_{i2}} = -1$). The different values of this factor, define an absence of qualitative differences (when $r_{\beta_{i1}\beta_{i2}} = 1$), or define a qualitative difference type (when $r_{\beta_{i1}\beta_{i2}} = 0$ or $r_{\beta_{i1}\beta_{i2}} = -1$). The case of $r_{\beta_{i1}\beta_{i2}} = -1$ is one with extreme qualitative differences with respect to location. Note that the correlation between the location parameters is not based on individual differences and is unrelated to overall individual differences.

When a location correlation of 0 or -1 ($r_{\beta_{i1}\beta_{i2}} = 0$ or $r_{\beta_{i1}\beta_{i2}} = -1$) is combined with a group main difference of 2 ($\gamma_2 = 2$), there are both qualitative <u>and</u> quantitative differences present in the data. If the group difference is 0 ($\gamma_2 = 0$), whereas the location correlation is 0 or -1, there are only qualitative category differences. If the location correlation is one ($r_{\beta_{i1}\beta_{i2}} = 1$) and the group means do not differ ($\gamma_2 = 0$), the model is purely dimensional because homoscedasticity and equal discriminations are assumed.

The third factor defines within-category homogeneity or within-category heterogeneity by using $\theta_{pc} \sim N(0,0)$ or $\theta_{pc} \sim N(0,1)$, respectively, yielding homogeneous categories or heterogeneous categories, respectively.

The combination of these three factors describe twelve data configurations for each of which 20 data sets were generated.

Taxometric results

The taxometric analyses were accomplished using John Ruscio's (2006) free program code. As the data in Study 1 are rating-scale data (DIMCAT data), not the original MAXCOV method (described above) is used, but the so-called Short Scale MAXCOV (e.g., Gangestad & Snyder, 1985; Ruscio, 2000; Schmidt, Kotov, & Joiner, 2004) which is designed to be used with variables that do not form a (quasi) continuous scale. In Short Scale MAXCOV, two output indicators are picked from the indicator set, just as in case of the original version, but instead of using one of the remaining indicators as the input variable, the sum of all remaining indicators is used to form a composite input variable. An important assumption is that the sum score has validity for the category distinction.

Although in the original form of MAXCOV the covariances are plotted along the nonoverlapping ordered sections of the input indicators, the analyses can also be accomplished using overlapping windows instead. The method was originally used for another taxometric procedure called MAXEIG (Waller & Meehl, 1998). The advantage of the procedure is that it provides MAXCOV plots that are easier to interpret without using smoothing techniques.

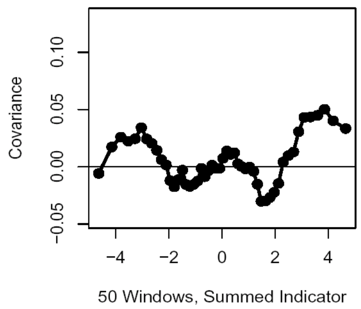
The MAMBAC analyses were performed in a way similar to Short Scale MAXCOV; one indicator was taken as the output indicator, whereas the the remaining indicators were summed to form a composite input indicator. Also for MAMBAC, overlapping windows were used instead of nonoverlapping windows.

For the L-Mode, the analyses are simply common factor analyses and the factor scores on the first factor are plotted.

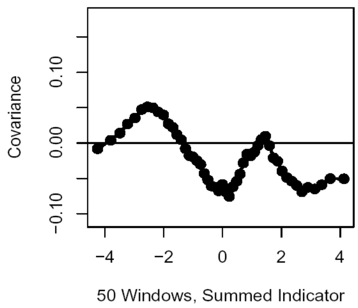
As it was already mentioned, the visual inspection of the graphs plays an important role in taxometrics. However, presenting all the plots resulting from the simulation study would not be feasible. Only representative graphs for each cell in the generation design will be shown. As there are many plots: 55 MAXCOV plots, 11 MAMBAC plots and 1 L-Mode plot for each of the 20 data sets within a cell of the design, only representative plots will be presented. Beyond these graphs, also the summary of the consistency tests will be presented.

MAXCOV

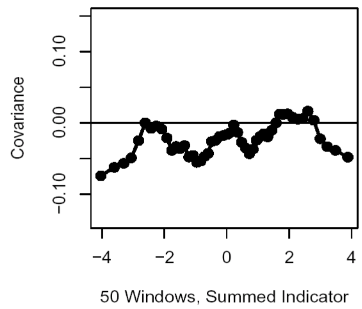
MAXCOV does not seem able to indicate taxonicity (category-likeness) with the data configurations of Study 1. Figure 1 shows a representative averaged MAXCOV graphs for each of the twelve data structures.



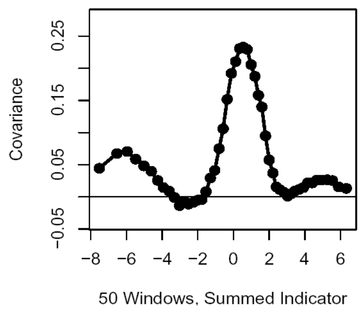
(a) correlation is 1
main effect is 0



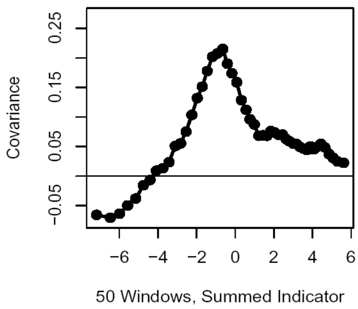
(b) correlation is 0
main effect is 0



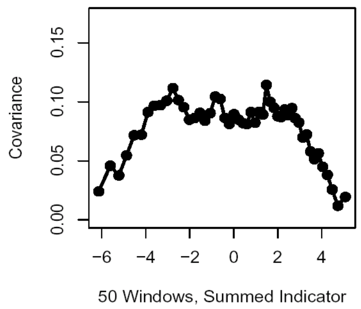
(c) correlation is -1
main effect is 0



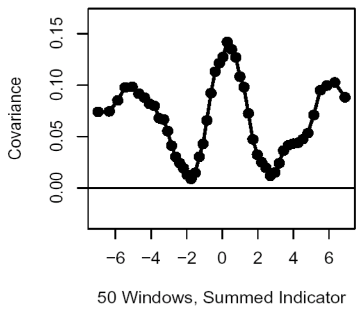
(d) correlation is 1
main effect is 2



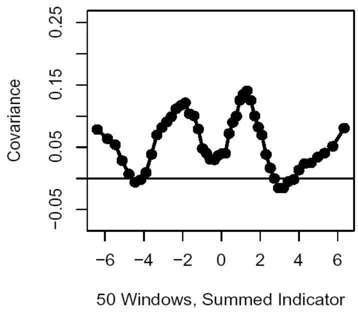
(e) correlation is 0
main effect is 2



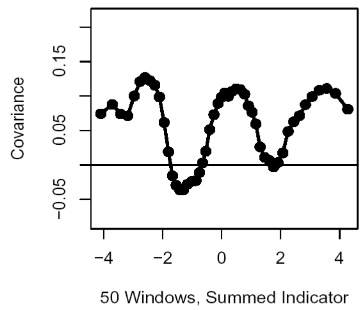
(f) correlation is -1
main effect is 2



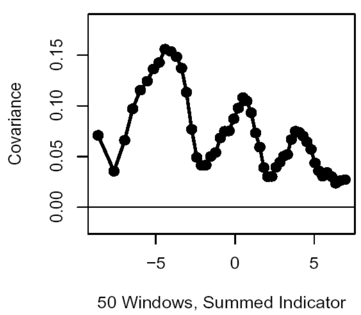
(g) correlation is 1
main effect is 0



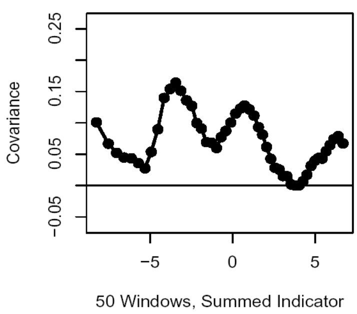
(h) correlation is 0
main effect is 0



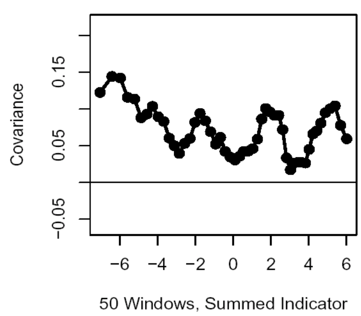
(i) correlation is -1
main effect is 0



(j) correlation is 1
main effect is 2



(k) correlation is 0
main effect is 2



(l) correlation is -1
main effect is 2

Figure 1: MAXCOV graphs for for homogeneous categories (a to f) and heterogeneous categories (g to l) depending on (the degree of) qualitative differences and main effect.

Based on the visual inspection of these graphs, one would conclude in most cases that the underlying structure is dimensional, whereas in most of the cells (10 out of 12) the structure is in fact category-like. The exceptions where category-likeness can be established indeed are the cells where a quantitative difference (main effect of 2) is combined with homogeneity without extreme qualitative differences (panel d and e of Figure 1, respectively).

For MAXCOV, a "built-in" consistency test is the variability of the base rate estimates stemming from the covariance plots of indicator triplets (two output indicators and one input indicator). For each of the data sets in this simulation study, there are 55 estimates of the base rate. Kotov et al. (2004) suggest 0.1 SD as a rule of thumb, that is, when the standard deviation of the base rate estimates is smaller than 0.1, the underlying structure is most likely to be categorical. The underlying structure is most likely to be dimensional when the standard deviation of the base rate estimates is greater than or equal to 0.1. Table 1 contains the percentages of correct MAXCOV decisions based on the SD of the base rate estimates, on the GFI and on the CCFI. It can be clearly seen from these results that with the data configurations of Study 1 the rule of thumb of 0.1 SD of the base rate estimates does not work. Even when the averaged curve suggest a categorical underlying structure, the SD of the base rate estimates is larger than 0.1, suggesting a dimensional underlying structure. This result is true for every data set of every data configuration. Hence, for the true dimensional structures the SD-based decisions are perfect, whereas for the categorical structures the SD-baseddecisions are never correct.

For the GFI the results are more promising for the data configurations with a main effect of 2 ($\gamma_2 = 2$). When the main effect is zero ($\gamma_2 = 0$) the GFI never results in a correct decision, while for a main effect of two the results are perfect when only quantitative differences are present ($r_{\beta_{i1}\beta_{i2}} = 1$), with only slightly worse results when the quantitative differences are combined with qualitative differences. However, the correct decision rate deteriorates to zero when extreme qualitative differences are combined with within-class heterogeneity.

Table 1: Percentages of correct MAXCOV decisions based on the threshold of the standard deviation (SD) of the base rate estimates, on the GFI and on the CCFI

	var theta	main effect	corr of items	SD	GFI	CCFI
a	0	0	1	100	0	68.42
b	0	0	0	0	0	25
c	0	0	-1	0	0	25
d	0	2	1	0	100	84.21
e	0	2	0	0	100	55.56
f	0	2	-1	0	85	5
g	1	0	1	100	0	94.74
h	1	0	0	0	0	5
i	1	0	-1	0	0	25
j	1	2	1	0	100	10
k	1	2	0	0	56.25	6.25
l	1	2	-1	0	0	10

Of the three measures described above, the CCFI seems to provide the highest correct decision rate when considering all data configurations. The CCFI produces correct decision rates larger than zero for all data configurations, with the highest hit rate for a categorical structure if no qualitative differences occur and when the categories are homogeneous (hit rate is 84.25). This is because taxometrics is not specifically aimed at revealing qualitative between-class differences (see Hidegkuti & De Boeck, in press) and because latent homogeneity is assumed as explained earlier (although the method is robust against mild heterogeneity). For the two dimensional structures (main effect of zero and no qualitative differences) also high hit rates are obtained (68.42 and 94.74).

It must be concluded that with the data configurations in Study 1 the MAXCOV procedure provides a rather poor performance in detecting categorical structures. However, there are substantial differences between the different measures that were suggested to help the taxometric decision making. The CCFI seems to be the best measure, although also this measure fails most of the time when qualitative differences play a role. It has to be noted though, that the qualitative differences employed here (mainly the case of perfect negative correlation of the indicator locations across the two classes) represent a rather extreme case of qualitative differences.

MAMBAC

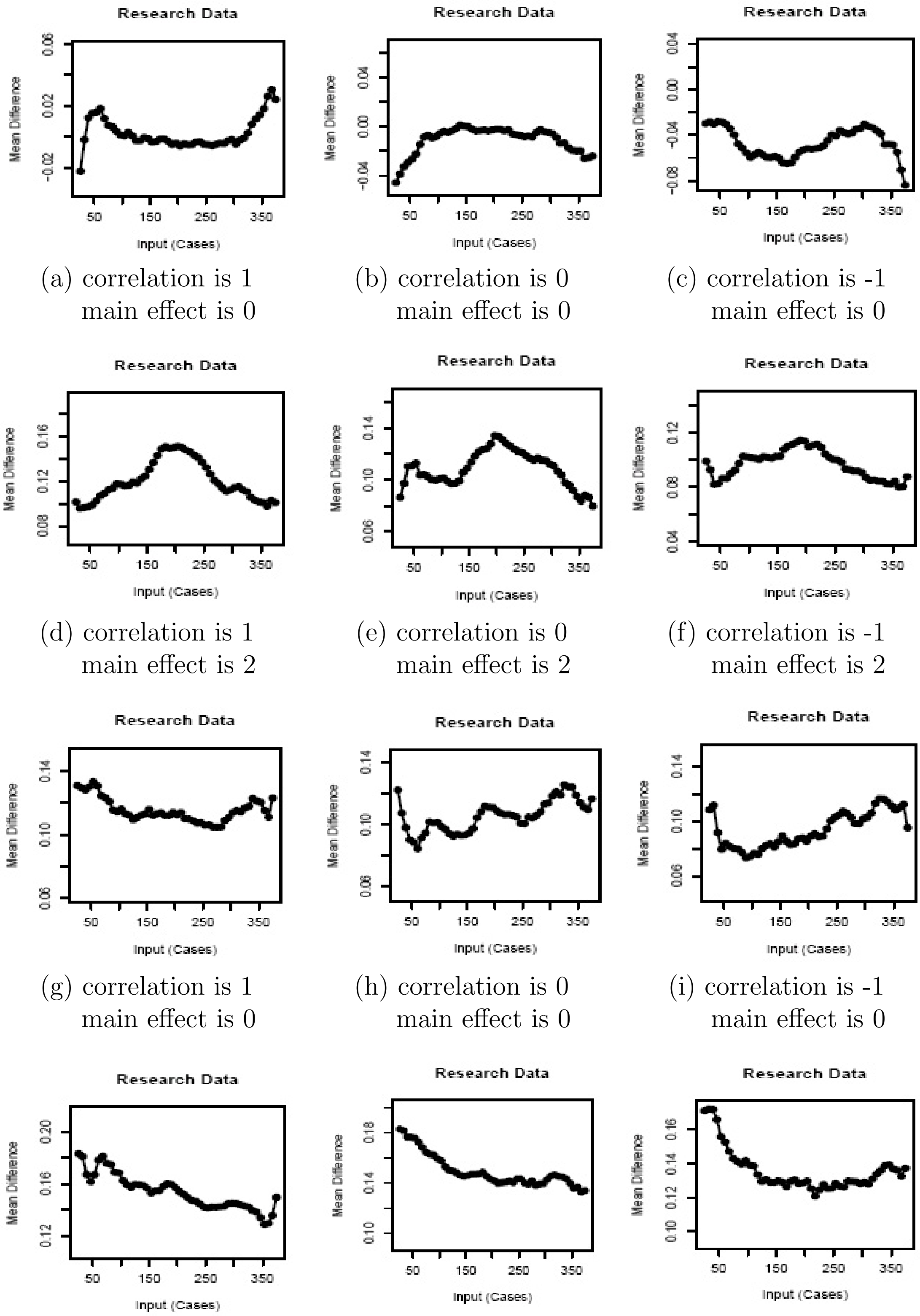
The MAMBAC results are rather similar to those of the MAXCOV analyses. The representative averaged MAMBAC graphs appear in Figure 2, while the percentages of correct decisions based on the threshold on the standard deviation of the base rate estimates, the GFI and the CCFI are shown in Table 2.

The visual inspection of the representative MAMBAC graphs suggests the same conclusions that were obtained with the visual inspection of the MAXCOV graphs, namely that a category-like structure would be concluded only in the case of data configurations with a main effect of two and within-class homogeneity.

As far as the measures are concerned, the SD of the base rate estimates and the GFI yield the exact same hit rates as in case of MAXCOV, hence they are not discussed any further here. Also the CCFI results are similar to those obtained with MAXCOV, but more polarized.

L-Mode

Figure 3 shows the representative L-Mode graphs. As can be seen, L-Mode performs very well in most cases, that is, the plot of the factor scores shows unimodality when only one class is present or in other words stated, when the structure is dimensional (panels a and g of Figure 3), and it shows clear bimodality for all but two categorical structures. For a mild form of qualitative differences (correlation of the indicator locations across classes



(a) correlation is 1
main effect is 0

(b) correlation is 0
main effect is 0

(c) correlation is -1
main effect is 0

(d) correlation is 1
main effect is 2

(e) correlation is 0
main effect is 2

(f) correlation is -1
main effect is 2

(g) correlation is 1
main effect is 0

(h) correlation is 0
main effect is 0

(i) correlation is -1
main effect is 0

(j) correlation is 1
main effect is 2

(k) correlation is 0
main effect is 2

(l) correlation is -1
main effect is 2

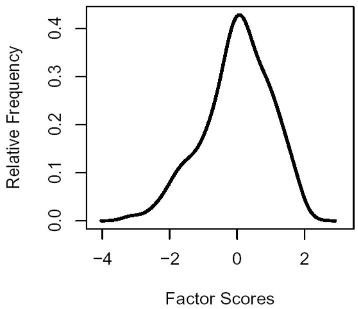
Figure 2: MAMBAC graphs for homogeneous categories (a to f) and heterogeneous categories (g to l) depending on (the degree of) qualitative differences and main effect.

Table 2: Percentages of correct MAMBAC decisions based on the threshold of the standard deviation (SD) of the base rate estimates, on the GFI and on the CCFI

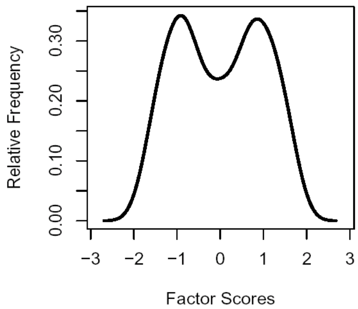
	var theta	main effect	corr of items	SD	GFI	CCFI
a	0	0	1	100	0	47.37
b	0	0	0	0	0	0
c	0	0	-1	0	0	0
d	0	2	1	0	100	100
e	0	2	0	0	100	72.22
f	0	2	-1	0	85	50
g	1	0	1	100	0	100
h	1	0	0	0	0	0
i	1	0	-1	0	0	0
j	1	2	1	0	100	0
k	1	2	0	0	56.25	6.25
l	1	2	-1	0	0	5

is 0) combined with heterogeneity and a main effect of zero (panel h in Figure 3) the bimodality is only minor. Also when the categories are heterogeneous and the difference between the categories is quantitative, one would not recognize the two underlying classes based on the plot of the factor scores (panel j of Figure 3). When the difference is qualitative and the two categories differ as to their level as well (main effect of two), the bimodality is always very clear.

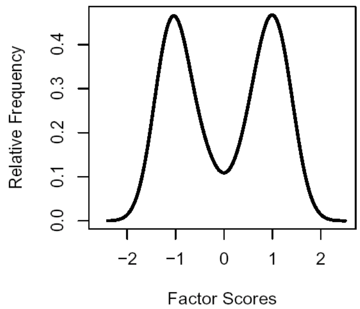
As it was mentioned before, the visual inspection of the graphs can be supported by consistency tests. For L-Mode, a "built-in" consistency test is to compare the base rate estimates calculated from the two modes. When there is a categorical structure in the background of the data, the base rate estimates are supposed to be fairly similar, whereas in case of a dimensional underlying structure, the base rate estimates are not expected to be similar. Therefore, the difference between the base rate estimates calculated from the two modes can be used as an indication of the underlying structure. Unfortunately, for this consistency test there is no recommended threshold value. Therefore, a threshold of 0.1 was used as in case of the SD of the base rate estimates for the MAXCOV and



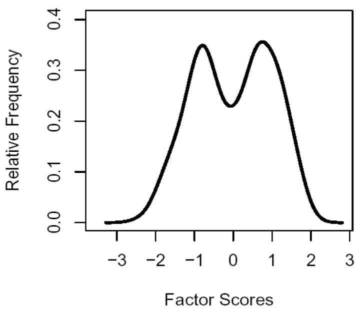
(a) correlation is 1
main effect is 0



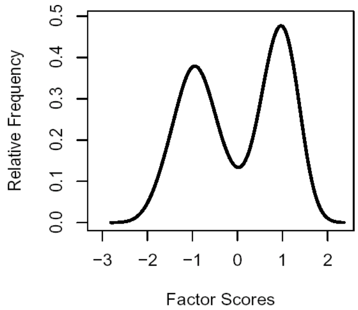
(b) correlation is 0
main effect is 0



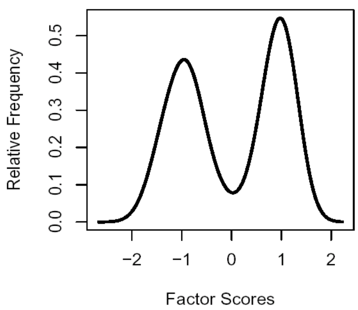
(c) correlation is -1
main effect is 0



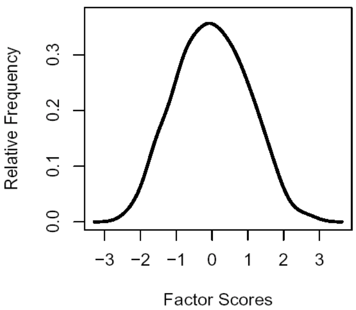
(d) correlation is 1
main effect is 2



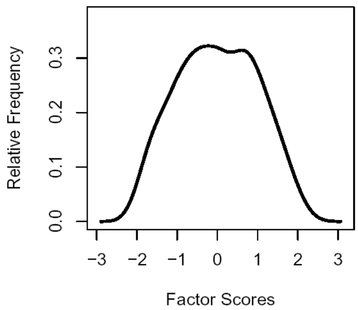
(e) correlation is 0
main effect is 2



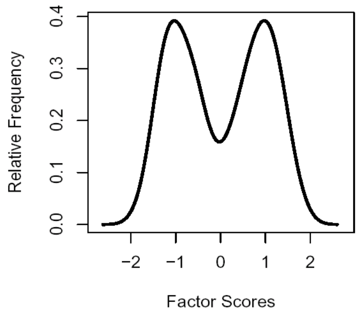
(f) correlation is -1
main effect is 2



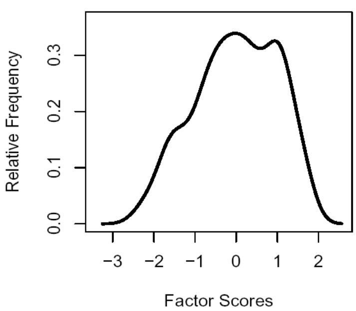
(g) correlation is 1
main effect is 0



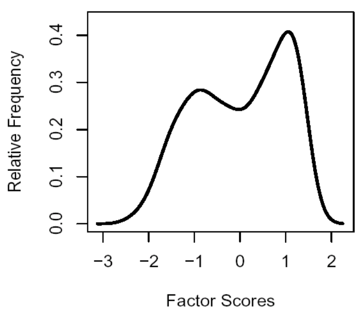
(h) correlation is 0
main effect is 0



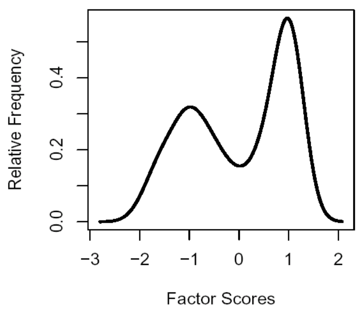
(i) correlation is -1
main effect is 0



(j) correlation is 1
main effect is 2



(k) correlation is 0
main effect is 2



(l) correlation is -1
main effect is 2

Figure 3: L-Mode graphs for homogeneous categories (a to f) and heterogeneous categories (g to l) depending on (the degree of) qualitative differences and main effect.

Table 3: Percentages of correct L-Mode decisions based on the threshold of the difference of the base rate estimates and on the GFI

	var theta	main effect	corr of items	Difference	GFI
a	0	0	1	100	0
b	0	0	0	75	0
c	0	0	-1	100	0
d	0	2	1	26.32	100
e	0	2	0	100	100
f	0	2	-1	100	85
g	1	0	1	100	0
h	1	0	0	0	0
i	1	0	-1	100	0
j	1	2	1	0	100
k	1	2	0	37.5	56.25
l	1	2	-1	90	0

MAMBAC procedures.

Table 3 shows the percentage of correct decisions based on the difference between the base rate estimates based on the two modes, and the percentage of correct decisions based on the GFI. The hit rates yielded by the GFI are identical with those stemming from the MAXCOV and MAMBAC analyses. CCFI seemed not to work for L-Mode (John Ruscio, personal communication, October 18, 2006) and hence, it was not considered. The hit rates based on the difference of the base rate estimates corroborate the conclusions that were drawn based on the visual inspection of the graphs. The results are very good in general, with two clear exceptions for the cells that were identified earlier as problematic from a visual inspection (cells h and j). For two other cells (d and k) the results are rather poor without a clear explanation, although in both cases the bimodality is relatively less pronounced. The GFI results are clearly inferior to those of the difference consistency norm.

As one can see, the results based on the base rate difference corroborate the conclusions drawn from the representative plots, but the visual inspection suggests a better differentiation.

DIMCAT results

Of the various possible models, six were used to reveal the underlying structure of the generated data sets. These models were: (1) a one-class model (Model 1, the polytomous version of the 1PL model) which is actually a purely dimensional model, (2) a two-class latent class model allowing only quantitative differences (Model 2, category main effect), (3) a two-class latent class model allowing both quantitative and qualitative differences (Model 3), and (Models 4, 5, 6) the homogeneous counterparts of the previous three models. The DIMCAT analyses were carried out with the M*plus* software (Muthén & Muthén, 2006). The analyses were performed using nonadaptive Gaussion quadrature with 15 quadrature points.

Since it is not uncommon that a latent class analysis results in a local optimum (e.g., Lubke & Muthén, 2005), the DIMCAT analyses involving latent classes were performed using 15 starting value sets for each analysis. After 10 iterations the loglikelihood values were evaluated and the 5 starting value sets resulting in the best loglikelihood values were used to run the analysis until convergence was reached.

The comparison of the six models and the decision was made based on two information criteria, the AIC and the BIC. The decision about the underlying structure is considered to be correct if the true model (i.e., the one used to generate the data) is the best fitting one (i.e., has the smallest AIC and BIC values). The percentage of correct decisions for each of the twelve data configurations appear in Table 4, for both, the AIC and the BIC. For the combinations on line a and line g, the correct decision is dimension-like, whereas on all other lines the correct decision is category-like.

In general, DIMCAT performs very well in revealing the latent structure of these

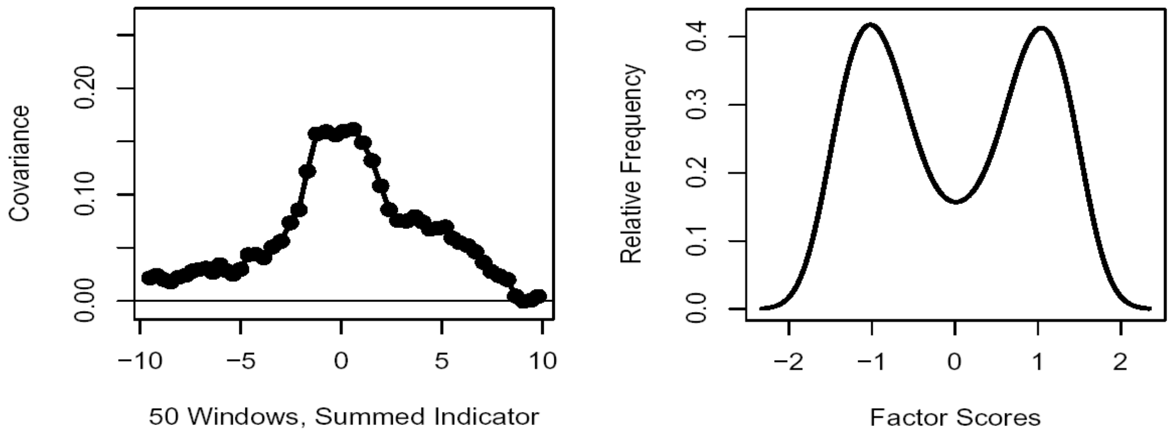
Table 4: Percentage of correct DIMCAT decisions (Study 1)

	var theta	main effect	corr of items	AIC	BIC
a	0	0	1	0	26.7
b	0	0	0	100	100
c	0	0	-1	100	100
d	0	2	1	100	100
e	0	2	0	95	100
f	0	2	-1	95	100
g	1	0	1	100	100
h	1	0	0	100	100
i	1	0	-1	100	100
j	1	2	1	5	0
k	1	2	0	100	100
l	1	2	-1	100	100

data sets, with two exceptions. The first exception is the cell where homogeneity is combined with identical indicator locations in both groups and a person category main effect of zero (neither quantitative nor qualitative differences). For this combination, the DIMCAT analyses tend to prefer the model with heterogeneity, almost never the one with homogeneity. The second exception is the cell where heterogeneity is combined with quantitative differences, but not with qualitative differences (line j in Table 4). This is a category-like combination which was also problematic for L-Mode to detect. The source of this problem is most likely that the mixture of two normal distributions with a variance of 1 and a mean difference of 2, yields a joint distribution which is unimodal, making it almost impossible to separate the two classes.

To further investigate this issue, the analyses for the problematic data configuration was repeated but now using a category main difference of 4 (instead of 2) to generate the data. A representative MAXCOV graph as well as a representative L-Mode graph appear in Figure 4.

It is clear from the graphs, that with a larger category main difference, both



MAXCOV L-Mode

Figure 4: Represenative MAXCOV and L-Mode graphs for the cell with heterogeneity, identical indicator locations and a category main effect of 4

MAXCOV and L-Mode work fine. This finding is also corroborated for L-Mode based on the difference between the base rate estimates calculated from the two modes, which was 0.04 on average, with a standard deviation of .02, whereas in the original study it was 0.72 with a standard deviation of 0.2 (line j of Table 3). In case of the MAXCOV analyses, the mean standard deviation of the base rate estimates is 0.13 with a standard deviation of 0.02, which is a substantial improvement in comparison with the original study (mean standard deviation of 0.22 with a standard deviation of 0.02). However, the standard deviation of the base rate estimates was smaller than the recommended threshold in only 1 out of 20 data sets. Also the DIMCAT results improved, so that the correct data structure was found in all twenty data sets, based on either the AIC or the BIC, whereas in the original study the correct data structure was found only in 5 and 0 percent of the cases, based on the AIC and the BIC, respectively.

Study 2

In the second study, data simulated by Meehl and Yonce (1994) were analyzed. In their study four normally distributed indicator variables were used and the following four factors were varied (but not in an orthogonal way): sample size, base rate, factor loadings, indicator separation for categorical structures. The base rate is the proportion of the taxon group (focus category, e.g., category of depressed people), whereas the indicator separation is the the difference between the means of the indicator in the two categories. The indicator separation is conventionally expressed in terms of *Cohen's d* statistic. For dimensional structures, two factors were varied: sample size and factor loadings. The data sets with a dimensional underlying structure were generated in such a way that the indicator correlations were the same as in the categorical data sets. The different data configurations can be found in Table 5 (together with the results of the MAXCOV analyses). For a more detailed description of the data generation procedure, the interested reader is referred to Meehl and Yonce (1994, pp. 1067-1068).

The data sets were used to test the performance of various taxometric procedures such as MAMBAC (Meehl & Yonce, 1994) MAXCOV (Meehl & Yonce, 1996), and L-Mode (Waller & Meehl, 1998).

Taxometric results

In Study 2, the MAXCOV, the MAMBAC and the L-Mode analyses were performed in the original fashion (not employing summed indicators as for the DIMCAT data), except that also here, overlapping windows were used instead of nonoverlapping windows for MAXCOV and MAMBAC.

MAXCOV

The percentages of correct decisions based on the standard deviation of the base rate estimates, based on the GFI and on the CCFI appear in Table 5. The results are as follows:

(1) The decisions based on the SD of the base rate estimates supports almost exclusively a dimensional underlying structure, independent of the true structure underlying the data. (2) The GFI shows the opposite trend suggesting a category-like latent structure in most cases. (3) The CCFI results are very good, independent of the underlying structure; the CCFI provides a high hit rate for both dimensional and category-like structures. To a certain extent, these findings are expected, as the CCFI has been shown to work well (J. Ruscio, Ruscio, & Meron, in press), while the GFI seemed not to work well in previous unpublished simulations. However, the rather poor performance of the SD and the GFI is still noteworthy. Of the three measures, the CCFI seems to be a good choice to be used in larger scale simulation studies, where the visual inspection of the resulting graphs is not feasible. A visual inspection of the plots is still a decisive factor, but it can effectively be aided by generating dimension-like and category-like comparison data and comparisons of the corresponding results using the CCFI.

MAMBAC

The percentages of correct MAMBAC decisions based on the three criteria are shown in Table 6. The results are similar to those obtained with MAXCOV, although the results are somewhat more extreme. The CCFI results are again the best, being only slightly worse for true dimensional structures than those obtained in the MAXCOV analyses. The results based on the CCFI are very good not only in a relative sense (compared to the other two measures) but also in an absolute sense.

L-Mode

The results of the L-Mode analysis are shown in Table 7. They do not bring any new findings, but they corroborate instead the conclusions drawn in Study 1 and the conclusions based on the MAXCOV and MAMBAC analyses of Study 2. As the CCFI is not used with L-Mode there is no index available here to support the interpretation of the L-Mode graphs. The other two indexes show a pattern similar to the MAXCOV and the

Table 5: Percentage of correct MAXCOV decisions based on the threshold of the SD of the base rate estimates, on the GFI and on the CCFI (Study 2)

Sample	Taxonic data configurations						Nontaxonic data configurations			
size	Indicator separations	Base rate	Factor loadings	SD %	GFI %	CCFI %	Factor loadings	SD %	GFI %	CCFI %
100	2	0.5	0	0	100	100	0.707	100	56	72
300	2	0.5	0	0	100	100	0.707	100	42.86	90.48
600	2	0.5	0	0	100	100	0.707	100	39.13	100
300	2	0.25	0	0	100	100	0.66	100	24	96
600	2	0.25	0	0	100	100	0.66	100	0	100
300	2	0.1	0	84	100	100	0.51	100	0	91.67
600	2	0.1	0	96	100	100	0.51	100	0	91.67
300	1.5	0.5	0	0	100	96	0.6	100	0	88
600	1.5	0.5	0	0	100	100	0.6	100	8	96
300	2	0.5	ind1: .7	0	100	100	ind1: .85	100	96	92
			ind2: .5				ind2: .79			
			ind3: .4				ind3: .77			
			ind4: .2				ind4: .69			
600	2	0.5	ind1: .7	0	100	100	ind1: .85	100	100	95.83
			ind2: .5				ind2: .79			
			ind3: .4				ind3: .77			
			ind4: .2				ind4: .69			
600	ind1: 2	0.5	ind1: .7	0	100	84	ind1: .85	100	63.64	100
	ind2: 1.75		ind2: .5				ind2: .76			
	ind3: 1.5		ind3: .4				ind3: .68			
	ind4: 1.25		ind4: .2				ind4: .54			

When only one factor loading or indicator separation value is indicated that means that the values are the same for all four indicators. In the other cases the four indicator separations are 2, 1.75, 1.5 and 1.25, and the four loadings are .7, .5, .4 and .2.

Table 6: Percentage of correct MAMBAC decisions based on the threshold of the SD of the base rate estimates, on the GFI and on the CCFI (Study 2)

Sample	Taxonic data configurations						Nontaxonic data configurations			
size	Indicator separations	Base rate	Factor loadings	SD %	GFI %	CCFI %	Factor loadings	SD %	GFI %	CCFI %
100	2	0.5	0	0	100	96	0.707	100	12	68
300	2	0.5	0	0	100	100	0.707	100	0	95.24
600	2	0.5	0	0	100	100	0.707	100	0	100
300	2	0.25	0	0	100	100	0.66	100	0	92
600	2	0.25	0	0	100	100	0.66	100	0	95.83
300	2	0.1	0	12	100	100	0.51	100	0	66.67
600	2	0.1	0	12	100	100	0.51	100	0	79.17
300	1.5	0.5	0	0	100	100	0.6	100	0	68
600	1.5	0.5	0	0	100	100	0.6	100	0	96
300	2	0.5	ind1: .7	0	100	100	ind1: .85	100	44	96
			ind2: .5				ind2: .79			
			ind3: .4				ind3: .77			
			ind4: .2				ind4: .69			
600	2	0.5	ind1: .7	0	100	100	ind1: .85	100	41.67	75
			ind2: .5				ind2: .79			
			ind3: .4				ind3: .77			
			ind4: .2				ind4: .69			
600	ind1: 2	0.5	ind1: .7	0	100	100	ind1: .85	100	0	95.45
	ind2: 1.75		ind2: .5				ind2: .76			
	ind3: 1.5		ind3: .4				ind3: .68			
	ind4: 1.25		ind4: .2				ind4: .54			

When only one factor loading or indicator separation value is indicated that means that the values are the same for all four indicators. In the other cases the four indicator separations are 2, 1.75, 1.5 and 1.25, and the four loadings are .7, .5, .4 and .2.

MAMBAC analyses. The difference of the base rate estimates for L-Mode proves to be a somewhat better criterion than the standard deviation of the base rate estimates for the MAXCOV and MAMBAC procedures. However, it is still not nearly as good as the CCFI in case of the other two procedures.

Generally, it can be concluded that two of the three taxometric procedures applied in this study, namely MAXCOV and MAMBAC work very well when using the CCFI as a criterion, whereas the other two criteria, the standard deviation (or difference in case of L-Mode) of the base rate estimates, and the GFI have a rather poor performance. The difference of the base rate estimates in case of L-Mode proves to be a better working index then the SD of the base rate estimates (in case of MAXCOV and MAMBAC) but not as good as the CCFI for the other two taxometric methods considered in this paper. Although two of the three indices do not seem to work well, it must be concluded that the taxometric methods can indeed differentiate between categorical and dimensional structures quite well.

DIMCAT results

To perform the DIMCAT analyses, four models were used: (1) a one-class single factor analysis model, (2) a two-class single factor analysis model, and (3, 4) the homogeneous counterparts (the variance of the underlying factor is 0) of these two models. The αs (see Equation 5) were fixed to one just as in the first study the discrimination parameters were fixed.

The percentage of correct decisions (based on the AIC and BIC) about the underlying structure appear in Table 8. As can be seen, the results are very good in general. The AIC produces somewhat less correct decisions, but in general it provides a fairly good hit rate. The BIC, on the other hand, proves to be a perfect or nearly perfect criterion for all the taxometric data configurations.

Table 7: Percentage of correct L-Mode decisions based on the threshold of the differences of the base rate estimates and on the GFI (Study 2)

Sample size	Taxonic data configurations: Indicator separations	Base rate	Factor loadings	Difference %	GFI %	Nontaxonic data configurations: Factor loadings	Difference %	GFI %
100	2	0.5	0	20	100	0.707	100	12
300	2	0.5	0	40	100	0.707	100	0
600	2	0.5	0	20	100	0.707	100	0
300	2	0.25	0	40	100	0.66	100	0
600	2	0.25	0	40	100	0.66	100	0
300	2	0.1	0	48	100	0.51	100	0
600	2	0.1	0	48	100	0.51	100	0
300	1.5	0.5	0	0	100	0.6	100	0
600	1.5	0.5	0	0	100	0.6	100	0
300	2	0.5	ind1: .7	8	100	ind1: .84	100	40
			ind2: .5			ind2: .79		
			ind3: .4			ind3: .77		
			ind4: .2			ind4: .69		
600	2	0.5	ind1: .7	0	100	ind1: .84	100	41.67
			ind2: .5			ind2: .79		
			ind3: .4			ind3: .77		
			ind4: .2			ind4: .69		
600	ind1: 2	0.5	ind1: .7	0	100	ind1: .85	100	0
	ind2: 1.75		ind2: .5			ind2: .76		
	ind3: 1.5		ind3: .4			ind3: .68		
	ind4: 1.25		ind4: .2			ind4: .54		

When only one factor loading or indicator separation value is indicated that means that the values are the same for all four indicators. In the other cases the four indicator separations are 2, 1.75, 1.5 and 1.25, and the four loadings are .7, .5, .4 and .2.

Table 8: Percentage of correct DIMCAT decisions

Sample size	Taxonic data configurations: Indicator separations	Base rate	Factor loadings	AIC %	BIC %	Nontaxonic data configurations: Factor loadings	AIC %	BIC %
100	2	0.5	0	84	96	0.707	80	100
300	2	0.5	0	88	100	0.707	84	100
600	2	0.5	0	80	100	0.707	80	100
300	2	0.25	0	80	96	0.66	76	100
600	2	0.25	0	80	96	0.66	84	100
300	2	0.1	0	72	96	0.51	92	100
600	2	0.1	0	84	100	0.51	96	100
300	1.5	0.5	0	84	100	0.6	76	96
600	1.5	0.5	0	80	100	0.6	84	100
300	2	0.5	ind1: .7	100	100	ind1: .84	68	100
			ind2: .5			ind2: .79		
			ind3: .4			ind3: .77		
			ind4: .2			ind4: .69		
600	2	0.5	ind1: .7	100	100	ind1: .84	80	100
			ind2: .5			ind2: .79		
			ind3: .4			ind3: .77		
			ind4: .2			ind4: .69		
600	ind1: 2	0.5	ind1: .7	100	100	ind1: .85	88	100
	ind2: 1.75		ind2: .5			ind2: .76		
	ind3: 1.5		ind3: .4			ind3: .68		
	ind4: 1.25		ind4: .2			ind4: .54		

When only one factor loading or indicator separation value is indicated that means that the values are the same for all four indicators. In the other cases the four indicator separations are 2, 1.75, 1.5 and 1.25, and the four loadings are .7, .5, .4 and .2.

Discussion

Although aiming at the same target of revealing the underlying nature of different psychological phenomena, the two approaches have a slightly different emphasis. Taxometrics is best suited for data stemming from different scales which measure the same phenomenon. These scales can be for example parts of a test battery measuring, for example, depression and the total scores on each part may be used as (quasi continuous) indicator variables in the taxometric analyses. This strategy, however, requires data from several scales or questionnaires which may not be readily available in many studies. A possible way to overcome this drawback is to use a single questionnaire consisting of items evaluated on a Likert-type scale, and then combining the items measuring the same (or similar) facet of the phenomenon in question into indicator variables which form a scale, as illustrated in Study 1.

On the other hand, DIMCAT is well-suited for questionnaire-type data, that is, rating-scale data with a limited set of response alternatives on a rating scale. Therefore, DIMCAT appears to be a natural choice when one intends to study the underlying nature of psychological phenomena using data from a single questionnaire, like, for example, the Beck Depression Inventory (Beck et al., 1996).

Another important issue concerns the focus of the two different approaches. Both methods are meant to find out whether there are two (or possibly more in case of DIMCAT) clearly separable categories, but for taxometrics the possible separation necessarily appears on a common underlying dimension. As Waller and Meehl (1998, p. 9) stated 'Thus, the convenient dichotomy taxonic-vs.-dimensional should, strictly speaking, read "taxonic-dimensional vs. *dimensional only.*"' That is, even when taxonicity is found, the taxons (categories) are assumed to be located on the same common underlying dimension. Hence, the categorical vs. dimensional distinction in taxometrics seems to correspond to the distinction between smooth and abrupt quantitative differences in DIMCAT. Thus, in taxometrics no qualitative differences (in the sense used in the DIMCAT framework) are considered. Despite this fact, L-Mode can also detect categories if they are based on qualitative differences (at least the location nonequivalence type) but it cannot determine what the basis of the categorical nature is: quantitative or qualitative differences, or both (Hidegkuti & De Boeck, in press). The method does not differentiate among different types of categorical structures.

As could be seen, the results of the DIMCAT analyses are rather convincing, DIMCAT performs very well in general, for both DIMCAT data and taxometric data. As far as taxometrics is concerned, MAXCOV and MAMBAC works very well when analyzing taxometric data and using the CCFI as a criterion, but their performance becomes less convincing when used with the DIMCAT data configurations of the present study, even when the CCFI is employed as a criterion. L-Mode performs rather well in defining whether the underlying structure is categorical or dimensional, showing a comparable performance to that of DIMCAT in both studies, following a visual inspection or using the difference of the base rate estimates as a criterion. It is noteworthy, that the criterion that worked best for MAXCOV and MAMBAC is not applicable to L-Mode.

The poorer performance of MAXCOV and MAMBAC may stem from the fact that with indicators having only a few response alternatives (in this case four), composite input indicators need to be employed. In many cases these composite indicators do not differentiate between the underlying latent classes, such as when there are only qualitative differences.

A limitation of the second study (using taxometrics data) is that the data generated by Meehl and Yonce (1994) represent fairly idealistic conditions, hence, it would be desirable to repeat this second study with more realistic data including for example skewed indicators, and larger within-category correlations (mainly combined with smaller indicator separations), for example. The reason to use the data sets generated by Meehl and Yonce (1994) is that thus far these were the standard data sets used to evaluate the taxometric methods. In further simulation studies comparing taxometrics and DIMCAT, one should also vary the base rates and the indicator skew.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood theory. In B.N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–381). Budapest: Akadémiai Kiadó.

Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika, 43*, 567–573.

Beach, S. R. H., & Amir, N. (2003). Is depression taxonic, dimensional, or both? *Journal of Abnormal Psychology, 112*, 228–236.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory Manual* (2nd ed.). San Antonio, TX: Psychological Corporation.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical Theories of Mental test Scores.* Reading MA: Addison-Wesley.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory and Five-Factor Inventory Professional Manual.* Odessa, Fl: Psychological Assessment Resources.

De Boeck, P., Wilson, M., & Acton, G. S. (2005). A conceptual and psychometric framework for distinguishing categories and dimensions. *Psychological Review, 112*, 129–158.

Franklin, C. L., Strong, D. R., & Greene, R. L. (2002). A taxometric analysis of the MMPI-2 depression scales. *Journal of Personality Assessment, 79*, 110–121.

Gangestad, S. W., & Snyder, M. (1985). "To carve nature at its joints": On the existence of discrete classes in personality. *Psychological Review, 92*, 317–349.

Haslam, N., & Cleland, C. (2002). Taxometric analysis of fuzzy categories: A Monte Carlo study. *Psychological Reports, 90*, 401–404.

Hankin, B. L., Fraley, R. C, Lahey, B. B., & Waldman, I. D. (2005). Is depression best viewed as a continuum or discrete category? A taxometric investigation of childhood and adolescent depression in a population-based sample. *Journal of Abnormal Psychology, 114*, 96–110.

Hidegkuti, I., & De Boeck, P. (in press). Between-group differences and taxometrics. *Psychological Reports.*

Holland, P. W., & Wainer, H. (1993). *Differential Item Functioning.* Hillsdale, NJ: Lawrance Erlbaum.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications.* Chicago: SPSS.

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10*, 21–39.

Marcus, D. K., John, S. L., & Edens, J. F. (2004). A taxometric analysis of psychopathic personality. *Jornal of Abnormal Psychology, 113*, 626–635.

Meehl, P. E. (1973). MAXCOV-HITMAX: A taxometric search method for loose genetic syndromes. In P. E. Meehl (Ed.), *Psychodiagnosis: Selected papers.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1995). Bootstraps taxometrics. Solving the classification problem in psychopathology. *American Psychologist, 50*, 266–275.

Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports, 74*, 1059–1274.

Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports, 78*, 1091–1227.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543.

Muraki, E. (1990). Fitting a polytomous item response model for Likert-type data. *Applied Psychological Measurement, 14*, 59–71.

Muthén, L. K., & Muthén, B. O. (2006). M*plus* 4.0 [Computer software]. Los Angeles, CA: Muthén & Muthén.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

Rothschild, L., Cleland, C., Haslam, N., & Zimmermann, M. (2003). A taxometric study of Borderline personality disorder. *Journal of Abnormal Psychology, 112*, 657–666.

Ruscio, J. (2000). Taxometric analysis with dichotomous indicators: The modified MAXCOV procedure and a case removal consistency test. *Psychological Reports, 87*, 929–939.

Ruscio, J. (2006). *Taxometric Programs in R.* [Computer software and documentation]. Retrieved June 27, 2006, from http://www.taxometricmethod.com.

Ruscio, A. M., & Ruscio, J. (2002). The latent structure of analogue depression: Should the Beck Depression Inventory be used to classify groups? *Psychological Assessment, 14*, 135–145.

Ruscio, J. & Ruscio A. M. (2002). A structure-based approach to psychological assessment: Matching measurement models to latent structure. *Assessment, 9*, 4–16.

Ruscio, J., Ruscio, A. M., & Meron, M. (in press). Applying the bootstrap to taxometric analysis: Generating empirical sampling distributions to help interpret results. *Multivariate Behavioral Research*.

Schmidt, N. B., Kotov, R., & Joiner, T. E., Jr. (2004). *Taxometrics: Toward a new diagnostic Scheme for psychopathology.* Washington, DC: American Psychological Association.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.

Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua.* London: Sage.