Nonparametric inference in nonlinear principal components analysis: Exploration and beyond

Linting, M.

Chapter 3

Stability of Nonlinear Principal Components Analysis: An Empirical Study Using the Balanced Bootstrap

Principal components analysis (PCA) is used to explore the structure of data sets containing linearly related numeric variables. Alternatively, nonlinear PCA (NLPCA) can handle possibly nonlinearly related numeric as well as nonnumeric variables. For linear PCA, the stability of its solution can be established under the assumption of multivariate normality. For nonlinear PCA, however, standard options for establishing stability are not provided. In this paper, we use the nonparametric bootstrap procedure to assess the stability of NLPCA results, applied to empirical data. We use confidence intervals for the variable transformations, and confidence ellipses for the eigenvalues, the component loadings and the person scores. We discuss the balanced version of the bootstrap, bias estimation, and Procrustes rotation. To provide a benchmark, the same bootstrap procedure is applied to linear PCA on the same data. Based on the results, we advise to use at least 1000 bootstrap samples, to use Procrustes rotation on the bootstrap results, to examine the bootstrap distributions along with the confidence regions, and to merge categories with small marginal frequencies to reduce the variance of the bootstrap results.

3.1 Introduction

Chapter 2 contained a didactic description of the data reduction method *nonlinear principal components analysis* (nonlinear PCA), illustrated with an elaborate application of this method to an empirical data set. Nonlinear PCA was compared to standard PCA, discussing its advantages and disadvantages, and highlighting its ability to effectively deal with all types of possibly nonlinearly related variables (including nominal and ordinal ones). In Chapter 3, we will shift the focus from establishing the advantages of nonlinear PCA to how the stability of nonlinear as well as linear PCA can be assessed. Contrary to Chapter 2, the focus will be on similarities instead of differences in the (stability of) the two methods.

Standard PCA reduces a large number of variables to a limited number of principal components, which are uncorrelated linear combinations of the original variables that reproduce as much as possible the information in the data. PCA assumes relationships between variables to be linear, and is thus referred to as linear PCA. In addition, to obtain a sensible interpretation of the PCA solution, variables should be of a numeric (interval or ratio) level of measurement. However, research in the social and behavioral sciences often results in data that are nonnumeric, with measurements recorded on scales having an uncertain unit of measurement. Data would typically consist of qualitative or categorical variables that describe the persons in a limited number of ordered or unordered categories. Even when the data are numeric, nonlinear relationships between the variables are common. To deal with data with such characteristics, *nonlinear* PCA has been developed as an alternative to linear PCA (Gifi, 1990) (for further references, see Chapter 2 of this thesis and Meulman, Van der Kooij, and Heiser (2004)).

Nonlinear PCA uses a procedure called optimal quantification, with a two-fold purpose. First, it assigns numeric values to categories of nonnumeric variables. These numeric values are called category quantifications; the category quantifications of a variable together form the so-called *transformation* of this variable. Then, the objective is to maximize the association between the quantified variables, or, in other words, to maximize the variance-accounted-for (VAF) by the principal components. Second, optimal quantification can deal with nonlinear relationships between variables. Here, the objective is to make these relationships linear by allowing for nonlinear transformations of the variables. If the variables are optimally quantified, the fit of the linear PCA model will be maximized.

Nonlinear PCA is also referred to as *categorical* PCA, because it can deal with categorical variables. In the type of nonlinear PCA described in this chapter, *all* variables – with nominal, ordinal, or numeric measurement levels –

are regarded as categorical. For instance, a numeric (interval or ratio) variable may be seen as a categorical variable with as many different categories as there are different scores on this variable. The researcher can specify an analysis level for each of the variables in the data set, and each variable is quantified in accordance with the requirements of its analysis level. Quantification of a numeric variable results in its standardization (i.e., z-scores, as in linear PCA), and therefore, the order of the original categories as well as the equal spacing between the categories will be maintained.[1] Quantification of the categories of an ordinal variable maintains the same order as the categories, and finds an optimal spacing between the categories. For nominal variables, both an optimal order and an optimal spacing will be obtained. Note that the analysis level is specified by the researcher and is not fixed by the measurement level of a variable. For a detailed description of the method of nonlinear PCA, we refer to Meulman, Heiser, and SPSS (2004), and Chapter 2 of this thesis.

Both linear and nonlinear PCA are often used in the context of exploratory research. However, there is no reason that these procedures should be deprived of inferential statistics for establishing, for example, the stability or robustness of a solution (defined in more detail below). For instance, for *linear* PCA on the covariance matrix, asymptotic distributions of the component loadings have been established (Anderson, 1963; Girshick, 1939). For linear PCA on the correlation matrix, Ogasawara (2004) has derived asymptotic standard errors for component loadings. However, such approaches rely on the assumption of multivariate normality, which may not apply in practice. Specifically for nonlinear PCA, a nonparametric approach seems more natural.

This chapter is focused on applying the nonparametric bootstrap procedure (Efron, 1982; Efron & Tibshirani, 1993) to establish confidence regions for several nonlinear PCA results. As linear PCA is the standard method, we use the stability of the linear PCA solution (assessed by exactly the same procedure) as a benchmark to judge the stability of nonlinear PCA. Clearly, nonlinear PCA cannot be expected to be more stable than linear PCA, because in most cases, many more outcome values have to be estimated in nonlinear PCA (each category obtains a quantification). In linear PCA, different versions of the bootstrap (parametric and nonparametric) have been applied to assess stability of the results (for example, see Efron & Tibshirani, 1993; Lambert et al., 1991; Milan & Whittaker, 1995). Timmerman et al. (in press) compared the asymptotic approach to the bootstrap approach, and found that the bootstrap is more flexible and under most conditions more accurate than

[1] In fact, if all variables are treated numerically, nonlinear PCA will give exactly the same results as linear PCA, because in that case no optimal quantification is required. In other words, nonlinear relationships between variables will not be discovered if only numeric analysis levels are used.

the asymptotic approach.

In the first section of this chapter, we start with defining the concept of stability in the context of nonlinear PCA. Then, we discuss the bootstrap procedure and its validity in establishing the stability of nonlinear multivariate analysis methods. We show how the confidence regions derived from the bootstrap results can be graphically displayed. In the second section, we apply nonlinear PCA to an empirical data set from the NICHD Study of Early Child Care (NICHD Early Child Care Research Network, 1996), and thoroughly examine the stability of the eigenvalues, component loadings, person scores (referred to as component scores in linear PCA), and quantified variables obtained by nonlinear PCA. A solution for the apparent instability of some of the results is proposed. Finally, the bootstrap results from nonlinear PCA are compared to bootstrap results from linear PCA on the same data set. In the third section, we state our final conclusions and give some general guidelines.

3.2 Assessing Stability of Nonlinear PCA

In Gifi (1990), stability of an analysis method is defined as the degree of sensitivity of the analysis to variations in the data or model parameters. A solution is said to be stable when “a small and unimportant change in data, model, or technique leads to a small and unimportant change in the results” (p. 36). In the current chapter, we limit our examination of stability to *data selection*. In other words, we define stability as the degree of sensitivity of nonlinear PCA to changes in the data. Small changes in the data should lead to only small changes in the output of the analysis.

Greenacre (1984) made a distinction between external and internal stability (also see Markus, 1994). *External stability* refers to whether the sample results may be generalized to the population. If a sample is representative for the population and given the sample size is large enough, we expect the results for that sample to only differ slightly from the results of another sample of the same size. A possible source for external instability is that a sample does not characterize the population structure, for example because it is too small. *Internal stability* refers to whether the results provide a good characterization of the structure of the sample at hand. Outliers are possible sources of internal instability.

In the context of external stability, a sample value can be seen as an estimate of the population value. This estimate is expected to vary across samples. One way to find out how much the estimate is likely to vary is to construct a confidence region, which will cover the population value with a probability specified by the researcher, for example 90%. In the following paragraphs, we will show how to establish confidence regions for the nonlinear PCA results. These confidence regions can provide information on internal as well as external stability.

3.2.1 The nonparametric bootstrap procedure

The classical method for assessing stability is through statistical inference, thereby relying on distributional assumptions. Because the standard assumptions, like (multivariate) normality, are often unrealistic for data collected in the behavioral sciences, nonlinear PCA, like linear PCA, explicitly avoids making such assumptions. Therefore, we employ the nonparametric bootstrap procedure (Efron, 1982; Efron & Tibshirani, 1993) to assess the stability of the nonlinear as well as the linear PCA solution. The nonparametric bootstrap procedure embodies randomly drawing, with replacement, B bootstrap samples from the original $n \times m$ data set, with n the number of persons and m the number of variables. The original sample is termed the parent sample. Each bootstrap sample contains persons from the parent sample, but some persons may occur several times, whereas others may not occur at all in a particular sample. In this way, a large number of B bootstrap samples consisting of n persons and m variables, is obtained. Subsequently, the analysis is performed on each of the bootstrap samples, which gives B values for each of the outcome values of interest. For each outcome value, these B bootstrap values form a bootstrap distribution from which a confidence region can be computed. In this thesis, we will compute bootstrap percentile intervals (including two-dimensional intervals, displayed by ellipses). Such intervals can be used to estimate the stability or robustness of analysis results, and if bias is small and the bootstrap distribution is approximately normal, they can be used to estimate the population parameter.

3.2.2 Validity of the bootstrap in nonlinear multivariate analysis

The performance of the bootstrap procedure has not yet been evaluated for all outcomes of nonlinear PCA. Markus (1994) performed a meta-study on the validity of the bootstrap method in assessing the stability of multiple correspondence analysis (MCA or HOMALS), which is identical to nonlinear PCA if all variables are treated as multiple nominal; here, the categories of all variables are represented as group points in the principal component space, with each group point indicating the center of all component scores for the persons that scored that particular category (for example, see Chapter 2). Markus also investigated some nonlinear PCA results when variables were treated ordinally. Although the results for the variables with an ordinal analysis level were explored in less detail than those for variables with a multiple nominal analysis level, they generally lead to the same conclusions.

The purpose of Markus (1994) was to investigate whether and under what circumstances the bootstrap method provides correct and useful information on the stability of nonlinear multivariate analysis techniques. To assess the validity of the bootstrap, she simulated a finite population on which she performed a nonlinear multivariate analysis to establish the population values of its outcomes. Then, she performed a bootstrap study on one sample drawn from the same finite population. She determined a bootstrap distribution and $(1-\alpha) \times 100\%$ bootstrap confidence regions for each of the outcome values of interest. This bootstrapping procedure was replicated 100 times on 100 different samples from the population, so that 100 confidence regions were obtained for each outcome. Finally, the proportion of times that the population value lay within the estimated confidence region – the *coverage percentage* – was established.

Markus's results (1994) showed that the coverage percentages of the bootstrap percentile confidence regions in nonlinear multivariate data analysis were satisfactory, meaning that a $(1-\alpha) \times 100\%$ bootstrap confidence region covers the population value of the statistic with a probability of approximately $p = 1 - \alpha$. Coverage percentages were particularly satisfactory if 90% or 95% confidence regions were established. This result is consistent with Efron's recommendation (1988) to use 90% confidence intervals (rather than, for instance, 99% confidence intervals), because bootstrap confidence intervals perform better if not pushed too far toward extreme coefficients.

Markus compared the standard deviations of MCA category quantifications in some bootstrap samples to corresponding asymptotic standard deviations, assessed by the so-called "delta method" (for other applications of the delta method in nonlinear multivariate data analysis, see Gifi, 1990; Meulman, 1984; Van der Burg & De Leeuw, 1988). The standard deviations from both methods were compared to criterion standard deviations, established by drawing 10,000 Monte Carlo samples (sample size = 100) directly from the known finite population. She concluded that the asymptotic and bootstrap results were fairly similar, but that the bootstrap procedure yielded more accurate estimates of the criterion standard deviations than the asymptotic method, and tended to be more conservative in estimating the variability of the quantifications. In line with these results, Efron (1988) asserted that the nonparametric bootstrap and parametric methods provide nearly equivalent inferences.

In her study, Markus (1994) varied the number of bootstrap samples, B, drawn from the population, and the sample size n. She found that for the bootstrap generated confidence regions to be valid in assessing internal as well as external stability, B should be at least 1000. In addition, she provided the guideline that n should be at least 200 to lead to an acceptable range of coverage percentages. Smaller samples led to large confidence regions, as they are expected to, since smaller sample size implies less accuracy. The general conclusion may be that "a solution resulting from a small sample may severely differ from the population solution, but the bootstrap samples will give a correct impression of the variation of the parameter" (p. 168).

A final important result is that in the MCA analyses, categories with small marginal frequencies (i.e., categories that contain relatively few observations) tended to be problematic. Merging categories results in higher and more accurate coverage percentages. Categories with small marginal frequencies also caused problems in the Monte Carlo study and the asymptotic estimations of stability. The criterion Monte Carlo standard deviations were considerably larger for categories with small marginal frequencies than for categories with intermediate or high marginal frequencies. Also, for categories with small marginal frequencies, the asymptotic as well as the bootstrap estimates of the criterion standard deviations were less accurate. The asymptotic estimations were somewhat less sensitive to small marginal frequencies than the bootstrap estimations, but overall less accurate.

Markus mainly studied the case where all variables had a multiple nominal analysis level. As this is the least restrictive analysis level in nonlinear PCA, we argue that her conclusions can be generalized to the more restricted case that we consider in the current chapter. This argument is supported by the fact that, when investigating the stability of some nonlinear PCA outcomes with variables treated ordinally, Markus found results similar to those with variables treated multiple nominally. In line with several other authors who have used the bootstrap to examine the stability of linear PCA (for example, Timmerman et al., in press) as well as the stability of nonlinear multivariate data analysis methods (Gifi, 1990; Greenacre, 1984; Heiser & Meulman, 1983; Meulman, 1982; Saporta & Hatabian, 1986; Van der Burg, 1988; Van der Burg & De Leeuw, 1988), we believe the bootstrap is a valid method for establishing the stability of results obtained by nonlinear PCA.

3.2.3 The bootstrap procedure in nonlinear PCA

For the application of the bootstrap procedure to nonlinear PCA, there are some considerations about the exact procedure to be used. These considerations are discussed below.

The balanced bootstrap

Because some persons may occur several times in one bootstrap sample, whereas others do not occur at all, persons may not be equally represented in the total of bootstrap samples. The balanced bootstrap (Efron & Tibshirani, 1993) can be used to ensure that every person appears a total number of exactly B times in B bootstrap samples. Efron and Tibshirani (1993) found that the average performance of the balanced bootstrap was about the same as that of the simple bootstrap estimate. Markus (1994) did find better results for the *variables* with the balanced bootstrap than with the unbalanced version, but this effect diminished when sample sizes (n) became larger. Because in the current study, we wish to establish the stability for the *person scores* as well, it is important that each person appears an equal number of times in the total of bootstrap samples. Therefore, we balanced the bootstrap in the following way.

Rather than simply sampling with replacement, we put B copies of the array of numbers from 1 to n into a vector $\mathbf{k}$ of length nB. For example, if we have a data set with 594 persons (as in the data set used in the application section of this chapter), a vector of length $594 \times 1000 = 594,000$ is constructed, containing the sequence of the numbers from 1 to 594, repeated 1000 times. These numbers are used to identify each person in the data set (and the entire row representing the scores on the measured variables for that person). Then, we randomly permute the vector $\mathbf{k}$, destroying the order of the identification numbers, and call the permuted vector $\mathbf{k}_p$. The first bootstrap sample then contains the rows of the data matrix (persons) indicated by the first n elements of $\mathbf{k}_p$, the second bootstrap sample contains the rows indicated by the elements $n+1$ to $2n$, and so on. For our example, $\mathbf{k}_p$ is used for 1000 bootstrap samples, from the first sample containing the persons identified by the first 594 numbers from the vector $\mathbf{k}_p$, unto the 1000^{th} sample containing the persons identified by the $593,406^{\text{th}}$ to the $594,000^{\text{th}}$ number. This procedure ensures that each person appears exactly B times in the total of B bootstrap samples, although the number of times a person appears in a single sample varies.

After the persons appearing in each bootstrap sample are selected as described above, nonlinear PCA is performed on each bootstrap sample. Because we want to construct confidence regions for the person scores, we need to know which person is which in each bootstrap sample to be able to construct an ellipse for each person separately. Therefore, during the resampling process, we retain the order of the persons in the original data set by using a weighting system: After we draw a bootstrap sample, the number of times that each person in the original data occurs in that bootstrap sample is counted and stored in the vector $\mathbf{w}$. Then, in the nonlinear PCA analysis on that bootstrap sample, **w** is used as a weight vector for the persons in the original data set. For example, persons with weight 0 do not contribute to the nonlinear PCA solution, whereas persons with weight 2 contribute as if they appeared twice in the data set. Clearly, in the weighting process, the order of the persons remains the same, enabling us to establish a confidence region for each of the persons. After performing PCA on all of the 1000 bootstrap samples, such a confidence region for each person is assessed on the collected object scores from all of the bootstrap samples in which that particular person appeared.

For each of the outcome values of interest, 1000 values are obtained from the bootstrap procedure. Subsequently, 90% confidence regions are established for the eigenvalues, component loadings, person scores, and category quantifications. The construction of a confidence region for a person score is based on a subset of the bootstrap samples, namely those samples in which that particular person appeared.

Bias

Despite the fact that the bootstrap distribution of a statistic approximates the sampling distribution of this statistic, and has the same shape and spread, it is not centered at the population value, but at the statistic from the original sample. Therefore, the bootstrap distribution does not reveal the center of the sampling distribution directly. In an unbiased situation, the mean of the bootstrap distribution would equal the original sample statistic, and may be viewed as an estimate of the population parameter. If there is bias, however, the bootstrap distribution reveals this by showing a difference between the center of the bootstrap distribution and the originally found statistic. When the difference between these two values is large, the bootstrap percentile confidence intervals do not work well (Efron & Tibshirani, 1993), and bias correction would be called for. However, Efron and Tibshirani (1993) note that bias correction could be dangerous in practice, because the bias corrected estimator may have a substantially greater standard error than the sample statistic. According to Markus (1994), in the context of nonlinear multivariate analysis, bias correction should be applied to the eigenvalues when bias is large, but is harmful to the category quantifications and component loadings. In addition, the effect of bias correction diminishes for sample sizes (n) of 500 or larger.

Rotation

In a graphical representation of the component loadings, an orthogonal rotation in both linear and nonlinear PCA solutions does not effect the configuration of points, nor does it effect the variance-accounted-for (VAF) by the components represented in the figure.[2] Thus, for two different samples, the results may appear very different, but can in fact be quite similar. In the extreme case, the rotational indeterminacy makes it possible for all of the variables to load highly positive on a component in one bootstrap solution, and highly negative on the same component in another, only due to a reflection. This effect will lead to artificially large confidence regions. Therefore, allowing for reflection is essential for a proper representation of the stability of the nonlinear PCA solution.

Reflection is a restricted form of rotation. A more sophisticated option is to use the rotational freedom of the nonlinear PCA solution by optimally rotating the component loadings from each of the B bootstrap samples to be as close as possible to the component loadings from the nonlinear PCA solution in the parent sample. This is a useful option if one is not interested in the relative dominance of the first component over the second, and wishes to remove uninteresting variation. In that case, the Procrustes procedure can be used to perform an optimal rotation of the bootstrap solutions (which also takes care of reflection). Procrustes rotation has been proposed by Cliff (1966) and Schönemann (1966), and is a rotation towards a target, here the component loadings found in the nonlinear PCA solution in the parent sample. We apply the orthogonal variant of Procrustes rotation, leaving the angles between the variable vectors unchanged. The person scores are rotated along with the component loadings. The need for rotation in *linear* PCA is discussed more extensively by Timmerman et al. (in press).

The interpretation of the bootstrap results

The bootstrap procedure implies repeating a specific analysis on different samples, all drawn from the same parent sample, and the bootstrap distribution can be viewed as an approximation of the sampling distribution. The objective of the bootstrap is to show how a statistic would vary due to random sampling. In the context of nonlinear PCA, the category quantifications differ with every bootstrap sample. Otherwise stated, although the class of quantification (e.g., ordinal) is specified by the analyst, the specific implementation of this quantification will vary across bootstrap samples, as each bootstrap sample contains a different group of persons. We do not interpret this fact as if different variables are analyzed in each analysis. The quantification of the original variables gives the optimal transformations; the bootstrap quantifications indicate how stable these transformations are. Stability of the

[2] Rotation does change the *relative* VAF by each of the components, and after rotation the first component is no longer the "principal component" in the sense of maximizing the VAF.

component loadings must also be viewed in the light of the definition given previously: Component loadings are stable if slight changes in the data lead to only slight changes in the results. In other words, the analysis should be robust to changes under data selection. Since the quantification process is considered an integral part of the analysis, it should be performed for each bootstrap sample. If the quantifications come out substantially different in the analysis of the bootstrap samples, the component loadings will be unstable as well. If, on the other hand, the component loadings have small confidence regions, the quantified variables can not be very different in the analysis of the bootstrap samples. In the latter case, we may refer to the solution as stable. A similar reasoning holds for the bootstrap results for the person scores.

3.2.4 Confidence ellipses

After performing nonlinear PCA in each bootstrap sample, confidence intervals for each outcome value can be established per component and displayed in a table. However, by looking at confidence intervals per component, the joint information concerning both components simultaneously is disregarded. The latter information can be revealed in a graphical representation. Because eigenvalues, component loadings and person scores from a two-dimensional nonlinear PCA can be represented as points in two dimensions (see Chapter 2), it is possible to use ellipses to indicate the confidence regions of these outcome values. The procedure of constructing confidence ellipses is described in an unpublished manuscript by Meulman and Heiser (1983), and used, among others, in Heiser and Meulman (1983), and Groenen, Commandeur, and Meulman (1998). The method has some attractive features: It is nonparametric (like PCA and the bootstrap), easy to implement, computationally fast, and visually insightful.

Confidence ellipses are constructed as follows, with the eigenvalues used as an example. In each bootstrap sample, an eigenvalue is found for each principal component. A combination of two eigenvalues, one for the first and one for the second component, can be represented as a point in two dimensions, with the first component on the x-axis and the second component on the y-axis. For each bootstrap sample, such a point for a combination of eigenvalues, called a bootstrap point, can be displayed. The B bootstrap points obtained for B bootstrap samples, form a cloud of points, referred to as the bootstrap cloud. The centroid of this cloud is the point that has as x- and y-coordinates the means of the bootstrap eigenvalues on the first and second component. A 90% confidence ellipse is the best fitting ellipse around 90% of the bootstrap points closest to the centroid of the bootstrap cloud, while retaining the orientation of the cloud in two dimensions. Thus, 90% confidence ellipses are *not* established for each component separately. A technical explanation how to construct such confidence ellipses can be found in Appendix C.

In general, the procedure of establishing confidence regions may be applied to the p eigenvalues of a p-dimensional nonlinear PCA. For a three-dimensional PCA, a three-dimensional variant of an ellipse (an ellipsoid) should be used, containing 90% of the points of a three-dimensional bootstrap cloud with as axes the three principal components. For higher-dimensional solutions a more elaborated procedure is called for, for instance looking at confidence ellipses for pairwise combinations of principal components (Component 2 versus 1, Component 3 versus 1, Component 3 versus 2, and so on).

3.3 Application

In this section, we apply the nonparametric bootstrap procedure to an empirical data set collected by the National Institute of Child Health and Human Development (NICHD) for their Study of Early Child Care (NICHD Early Child Care Research Network, 1996). First, we briefly describe the data set and discuss the nonlinear PCA solution for these data. Second, the bootstrap procedure is applied for assessing the stability of the eigenvalues, component loadings, person scores, and quantified variables. Third, we propose a solution for the instability in some of the outcomes. Finally, the bootstrap results from the nonlinear PCA solution are compared to the bootstrap results from the linear PCA solution for these data.

Several statistical packages contain programs that perform nonlinear PCA. In accordance with Chapter 2, we use the CATPCA program from the SPSS Categories Package (Meulman, Heiser, & SPSS, 2004) to apply nonlinear PCA in the current chapter. The SAS procedure PRINQUAL (SAS, 1992), which optimizes a similar criterion as CATPCA, can be considered as an alternative. For applying a stability study on the nonlinear PCA results, it is possible to combine PRINQUAL with the JACKBOOT procedure. However, as JACKBOOT is a general purpose jackknife and bootstrap macro, there is no standard provision for specialized output for nonlinear PCA. For example, JACKBOOT will provide univariate bootstrap confidence intervals, but not the confidence ellipses that for representing the stability for a two-dimensional solution. In the current study, we focus on the graphical representation of the bootstrap results for such a two-dimensional solution. SPSS macro files that can be used to perform the bootstrap for nonlinear PCA and a corresponding user guide are available on request. For *linear* PCA, SYSTAT provides bootstrap results, as well as the asymptotics when appropriate.

3.3.1 The Observational Record of the Caregiving Environment (ORCE)

The example data set we used contains a selection of 594 6-month olds who were observed in nonmaternal child care (father care, grandparent care, in-home sitter, child care home, or child care center) using the Observational Record of the Caregiving Environment (ORCE). We selected 21 ORCE variables aimed at measuring caregiver child interactions. At the end of an observation cycle, the observer had to rate certain described behavior from "not at all characteristic" (1) up to "very characteristic" (4) for the caregiver-child interactions during that cycle. The first eight ORCE variables were averages of these ratings over (in most cases) four 30-minute observation cycles. The final thirteen of the ORCE variables were so-called "behavior scales", indicating the average number of times a described type of behavior occurred during an observation cycle.

On the basis of its data theoretical philosophy and for computational efficiency, CATPCA is designed to handle discrete variables. We used the simplest form of discretization: We rounded the scores on all of the variables, before entering them into the analysis. Means, standard deviations and measures for skewness and kurtosis for the ORCE variables included in the analysis are given in Table 3.1. From this table we can conclude that all of the ORCE variables are skewed, specifically "Intrusiveness" (3), "Negative regard" (7), "Reads aloud" (12), "Stimulates cognitive development" (15), "Stimulates social development" (16), "Restricts activity" (18), "Negative speech" (20), and "Negative physical action" (21), which are also highly peaked.

The ORCE data have some features that implicate nonlinear PCA as an appropriate analysis method to handle these data. These features include: (a) the fact that the data contain ratings, of which the numeric characteristics might be questioned, (b) the possibility of nonlinear relationships between variables, and (c) the presence of missing values in the data (see Chapter 2). In the context of the didactic aspects of bootstrap study, the ORCE data – because of their skewed distributions – are suitable to investigate the stability of categories with small as well as large marginal frequencies.

Nonlinear PCA solution for the ORCE variables

In the bootstrap study, nonlinear PCA was performed on the rounded ORCE data, using the program CATPCA. In Chapter 2, two nominal variables concerning the type of nonmaternal care and caregiver education were included in the analysis, showing the advantage of nonlinear over linear PCA. In the bootstrap study, we wished to compare the stability of nonlinear PCA to the

Table 3.1: *Descriptives for the rounded ORCE variables. Skew. = skewness; Kurt. = kurtosis. Standard deviations of skewness and kurtosis are approximately 0.10 and 0.20 for this example ($n = 594$)*

	Variable	Range	Mean	SD	Skew.	Kurt.
1	(Responds to) distress	1–4	3.24	0.80	-0.65	-0.58
2	(Responds to) nondistress	1–4	2.99	0.74	-0.30	-0.34
3	Intrusiveness	1–4	1.20	0.48	2.69	8.23
4	Detachment	1–4	1.69	0.82	0.95	0.05
5	Stimulation	1–4	2.05	0.76	0.27	-0.41
6	Positive regard	1–4	3.19	0.77	-0.60	-0.28
7	Negative regard	1–3	1.03	0.17	6.81	50.57
8	Flatness	1–4	1.42	0.67	1.54	1.74
9	Positive affect	0–32	4.86	4.33	1.95	6.06
10	Positive physical contact	0–55	19.98	10.39	0.36	-0.14
11	(Responds to) vocalization	0–26	4.82	4.56	1.64	3.35
12	Reads aloud	0–11	0.41	1.23	4.36	22.91
13	Asks question	0–46	12.41	8.07	0.89	0.73
14	Other talk	1–57	24.23	12.08	0.40	0.52
15	Stim. cognitive development	0–34	3.12	3.86	2.70	12.75
16	Stim. social development	0–9	0.83	1.33	2.52	8.26
17	Facilitates behavior	0–55	18.78	9.67	0.60	-0.06
18	Restricts activity	0–21	1.39	1.99	3.22	18.94
19	Restricts physically	0–76	20.90	14.24	0.56	-0.20
20	Negative speech	0–3	0.07	0.30	4.80	27.49
21	Negative physical	0–2	0.02	0.14	9.90	109.19

stability of linear PCA under conditions in which these stabilities could be expected to show similar results. Under these conditions, strong differences in the results would suggest that the nonlinear transformations in nonlinear PCA introduced appreciable uncertainty into the solution. As it is not possible to include nominal variables in standard linear PCA and thus a stability comparison with nominal variables would not be possible, we included only the behavior scales and ratings in the current analysis.[3] In accordance with Chapter 2, we performed a two-dimensional nonlinear PCA, with all missing values treated as "passive," meaning that persons do not contribute to the solution for a variable they have a missing value on, but do contribute to the solution for the other variables.

In Chapter 2, we extensively described and interpreted many aspects of

[3]At the moment, the macro files used to perform the bootstrap for nonlinear PCA only incorporate single analysis levels. This issue is mentioned more extensively in the discussion.

the nonlinear PCA solution for the ORCE data, and showed how to use several analysis options. In the current chapter, we will focus on the eigenvalues, component loadings, and person scores from a relatively uncomplicated analysis. As the global results of the nonlinear PCA analysis were very similar to those found in Chapter 2, we will only briefly discuss them here.

The eigenvalue of the first component of the nonlinear PCA solution with all variables treated ordinally was 7.12, indicating that the first component accounted for approximately 33.9% (7.12/21) of the variance in the 21 optimally quantified variables. The second eigenvalue was 2.08, indicating a VAF of approximately 9.9%, which makes the total VAF by the two components almost 44%.

In Figure 3.1, the variables are represented as vectors. The direction of these vectors is determined by the component loadings; the squared length indicates the VAF. When vectors are long, the cosines of the angles between the vectors approximate the correlations between the quantified variables. The component loadings indicate correlations between the principal components and the quantified variables. From Figure 3.1, we can conclude that the first component depicts positive caregiving behaviors on the left-hand side versus disengaging behaviors on the right-hand side of the figure, and the second component is comprised of overtly negative behaviors.[4] As the solution for the ORCE data showed quite a simple structure, we did not apply any form of rotation.

The points for the children (given by the object/person scores) include two distinct outliers (also see Chapter 2). However, these outliers do not dominate the solution (they are not influential), which is indicated by the fair amount of spread in the points for the other persons.

The nonlinear PCA solution described above seems insightful, but the question is whether it can be generalized to the population of children experiencing nonmaternal care. In fact, the solution could be dependent on the sample, which is sometimes called *capitalization on chance*. In the next section, we will assess the stability of the nonlinear PCA solution by analyzing 1000 bootstrap samples, to find out to what extent differences in the data lead to differences in the results. If the results for the bootstrap samples are similar, we would infer that they are also similar to the results in the population.

[4]Due to rotational freedom of the (nonlinear) PCA solution, the component loadings on the first component are reflections of the component loadings in Chapter 2.

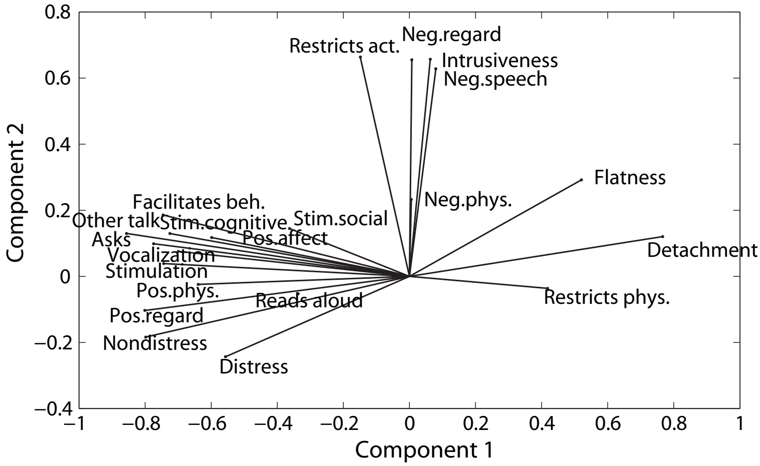

Figure 3.1: *Component loadings from nonlinear PCA on 21 ORCE variables, indicating behavior scales and ratings of caregiver-child interaction in nonmaternal child care.*

3.3.2 Balanced bootstrap results for the nonlinear PCA solution

We performed the nonparametric balanced bootstrap procedure, with B, the number of bootstrap samples, equal to 1000. Because the CATPCA program emphasizes graphical representation of the analysis results, we will show in the next subsections how the bootstrap results can be displayed in a way that reflects this emphasis.

Confidence regions for the eigenvalues

Figure 3.2 displays the 90% confidence ellipse for the eigenvalues from the two-dimensional nonlinear PCA for the example data set. In Figure 3.2, the bootstrap estimate of bias is shown as the white bar between the centroid of the bootstrap cloud (indicated by a circle) and the point representing the eigenvalues from the original parent sample (represented by a square). There is some – but not much – bias, especially in the second component. Markus (1994) indicated that, in spite of bias in the bootstrap estimate of the population value, bootstrap confidence regions still give a correct representation of the variation in this value (also see the section on the validity of the bootstrap). Because the sample is quite large ($n = 594$), and bias for the eigenvalues is limited, in accordance with Markus (1994), we decided not to correct for bias in the current study.

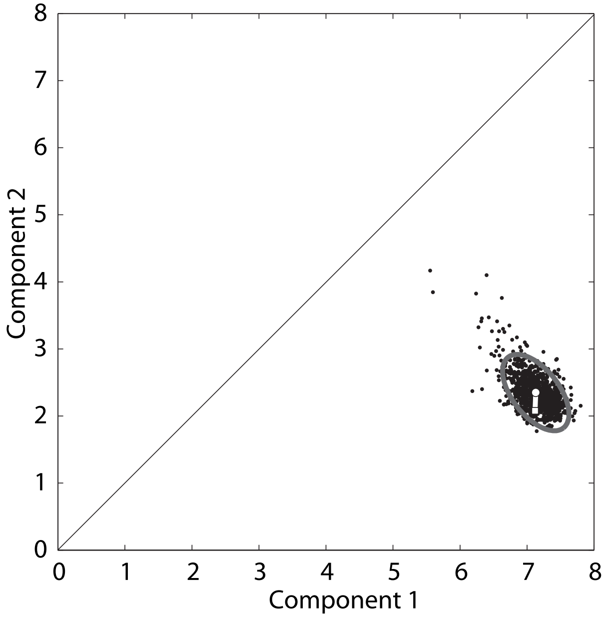

Figure 3.2: *90% confidence ellipse for the bootstrapped eigenvalues. The white circle indicates the centroid of the bootstrap cloud, the white square indicates the eigenvalues in the original parent sample, and the white bar indicates bias.*

The confidence ellipse in Figure 3.2 gives an insightful representation of the stability of the eigenvalues from nonlinear PCA. The confidence ellipse is fairly small, indicating a quite stable solution. It is clear that the first eigenvalue is systematically larger than the second (the first fluctuates around 7 and the second around 2.5). Note that, if we project the bootstrap points within the ellipse onto the axes representing the two components, the dispersion on the first and second axis is about the same. Therefore, we conclude that the eigenvalues are approximately equally stable in both components. Nonlinear PCA optimizes the sum of the first p eigenvalues simultaneously. The degree of negative tilt of the bootstrap cloud indicates that the sum of the two eigenvalues remains fairly constant in the bootstrap analyses. (In fact, the major axis of the ellipse is nearly parallel to the line indicating constant sums, which runs through the points (0,8) and (8,0).) If the sum of the eigenvalues stays approximately the same in the bootstrap samples, and if the second eigenvalue turns out to be larger than the eigenvalue from the parent sample, the first eigenvalue has to become smaller (and vice versa) in that bootstrap sample. In that case, the first and second eigenvalue will be negatively related, which is reflected by the direction of the bootstrap cloud.

Confidence regions for the component loadings

We also constructed confidence ellipses for the component loadings from the nonlinear PCA solution. We used orthogonal Procrustes rotation towards the nonlinear PCA solution for the observed sample to reduce irrelevant instability (see the previous section on rotation). In Figure 3.3, we show the effect of Procrustes rotation by comparing the results with rotation (indicated by the bold lines) to the results when only reflection of the axis has been applied when necessary (indicated by the thin lines). The lines connecting the black dots (centroids of the bootstrap clouds) and the white squares (original component loadings) indicate bias. (Note that these lines are “double” if the centroids with and without Procrustes rotation are not equal, which occurs only in a few occasions.)

This figure shows that for these data an orthogonal Procustes rotation hardly reduces the variance. This is another indication of the fact that the first two eigenvalues are well separated and thus that the principal axis orientation is stable. If the first and second eigenvalues would have been much closer, the need for a Procrustes rotation would be much more apparent. For equal eigenvalues, the principal axis orientation is completely undetermined. In that case, the orientation of the axes for the bootstrap samples is more or less random, leading to large, circularly shaped bootstrap clouds. For almost equal eigenvalues, Markus (1994) reported that crescent shaped bootstrap clouds might appear, which will result in confidence regions becoming disproportionately large. When rotation is applied, these crescent shapes disappear from the bootstrap clouds, resulting in much smaller confidence regions. For the present data, reflection of the axes would have been sufficient. In other cases, however, especially when the relationship between variables is unstable, it might not be clear by observation whether a reflection should be applied. Because an orthogonal Procrustes rotation takes care of an optimal reflection and is never harmful, we have incorporated it in our bootstrap procedure.

We can conclude from Figure 3.3 that most of the component loadings are quite stable (their confidence ellipses are fairly small). However, some of the variables – “Intrusiveness” (3), “Negative regard” (7), “Flatness” (8), “Reads aloud” (12), “Restricts physically” (19), and “Negative speech” (20) – have produced relatively large confidence ellipses and show some bias, and the instability occurs mainly in the second component (indicated by the fact that the ellipses are longer vertically than horizontally). A possible reason for the instability of variables can be found in the shape of their distributions: Table 3.1 shows that the variables with large confidence ellipses are highly skewed, with some categories having small marginal frequencies. For example, the ratings indicating behavior like “Intrusiveness” (3) and “Flat-

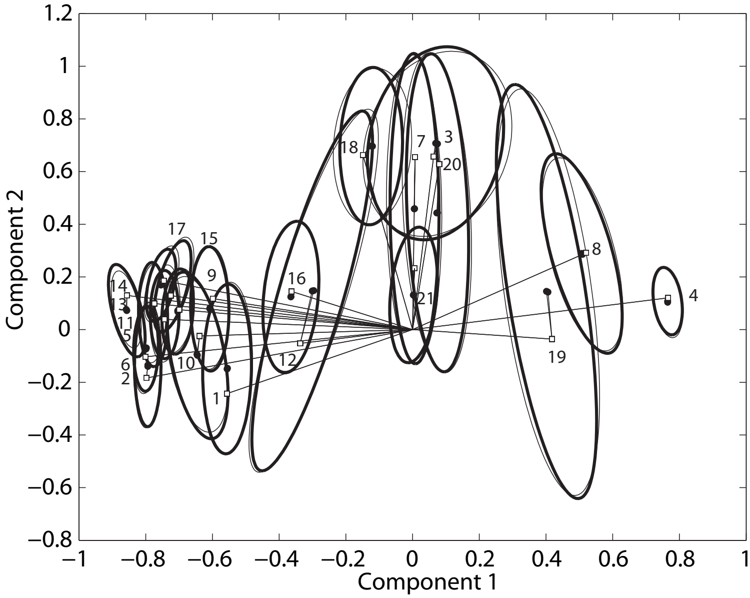

Figure 3.3: *90% confidence ellipses for the rotated component loadings. The bold lines indicate the confidence ellipses for the rotated component loadings. The thin lines indicate the confidence ellipses when only reflection is applied. Black circles indicate the centroids of the bootstrap clouds. White squares represent the component loadings in the original parent sample. The lines connecting the circles and squares indicate bias. 1=Distress, 2=Nondistress, 3=Intrusiveness, 4=Detachment, 5=Stimulation, 6=Positive regard, 7=Negative regard, 8=Flatness, 9=Positive affect, 10=Positive physical contact, 11=Vocalizations, 12=Reads aloud, 13=Asks question, 14=Other talk, 15=Stimulates cognitive development, 16=Stimulates social development, 17=Facilitates behavior, 18=Restricts activity, 19=Restricts in physical container, 20=Negative speech, 21=Negative physical action.*

ness of affect" (8), have small marginal frequencies in the categories indicating that the negative behavior is highly characteristic of the caregiver behavior toward the observed child. In some of the bootstrap samples, children experiencing these rare caregiver behaviors may be absent, whereas in others, they may appear several times. Therefore, a category based on those children who have experienced a rare behavior will produce at best an unstable quantification, and even no quantification at all in some of the bootstrap samples. The more extreme categories of the variables determining the first component show higher frequencies than those of the variables determining the second component, hence there is a higher degree of stability in the first component.

The problem of categories with small frequencies may be illustrated by the two largest confidence ellipses in Figure 3.3: "Reads aloud" (12) and "Restricts in a physical container" (19). When examining the distribution of the bootstrap points within those ellipses, we found that for both these variables, the distribution of bootstrap points is a mixture of two distributions: One distribution (consisting of about 800 bootstrap points) that is centered around the point representing the sample component loadings, and a distribution (consisting of about 200 bootstrap points) that has much higher loadings on the second component, depending on whether some children associated with very rare caregiver behavior were included in the bootstrap sample or not. Obviously, bootstrap results that reflect two different types of solutions point out a problem. This is why categories with small marginal frequencies should be avoided in the nonlinear PCA solution (also see Gifi, 1990; Markus, 1994). Note that in the present case it is the small marginal frequencies and not the skewness per se that creates apparent instability of some of the component loadings. With a much larger sample, these component loadings would be expected to be more stable.

3.3.3 A solution to the instability problem: Merging categories with small marginal frequencies

Because categories with relatively small marginal frequencies lead to unstable bootstrap results, we merged such categories, and repeated our bootstrap study on the recoded data. This decision is in accordance with Markus (1994) who stated that merging categories is beneficial for coverage percentages of bootstrap confidence regions and increases stability. Merging rare categories usually does not have a large effect on the eigenvalues (Gifi, 1990). For the ORCE variables, two or more adjacent categories of a variable were merged when:

1. the marginal frequency of one of the categories was less than 15, or

2. a variable had many (in the case of the ORCE data, more than 15) categories and the quantifications were close together or tied, and

3. the resulting number of categories was at least two. (This rule was used to avoid variables losing all variation.)

We used 15 as the minimal marginal frequency. The categories with small marginal frequencies were mostly the extreme categories and we wanted to ensure that those extreme categories (the tails of the distribution) contained at least 2.5% of the total number of scores. The 2.5% threshold seems a reasonable choice, because in a normal distribution, the two outer tails, containing 2.5% of the scores, are often considered extreme. In addition, our threshold is about twice as high as the eight observations per category recommended by Markus (1994), so that we believe to be "on the safe side". For "Negative physical actions", the most extreme category after merging still contained only eight observations, but further merging was not possible, because the minimum of two categories was reached. When variables had many categories, categories were merged only if their category quantifications were close together or tied. If categories with similar category quantifications within a variable are merged, the loss of information is minimal (Gifi, 1990, p. 397).

As an illustration of the merging procedure, the categories of the variable "Stimulates cognitive development" before and after merging are shown in Table 3.2. For this variable, the first five original categories did not need to be merged, because they contained more than 15 observations and had different quantifications. The eighth original category contained 14 observations and thus needed to be merged. As this category was tied with categories 6 and 7, these three were merged into a new category. All of the other categories contained fewer than 15 observations, and were merged with other categories with the most similar quantifications. The resulting recoded variable contained eight categories, each with more than 15 observations. In the nonlinear PCA solution after recoding, only the category quantifications of the two highest categories were notably different from the quantifications before recoding.

Confidence ellipses for the component loadings after merging

After merging categories as described above, the confidence ellipses for the corresponding component loadings have become much smaller, as is shown in Figure 3.4. In this figure, dashed lines indicate the confidence ellipses before recoding, and solid lines the confidence ellipses after recoding. Especially the ellipses for the variables "Intrusiveness" (3), "Reads aloud" (12), "Restricts in

Table 3.2: *Original category numbers, marginal frequencies, and category quantifications for "Stimulates cognitive development" before and after merging.*

Original			After merging		
Cat.	Freq.	Quant.	Cat.	Freq.	Quant.
1	147	-1.37	1	147	-1.36
2	121	-0.46	2	121	-0.48
3	72	0.12	3	72	0.11
4	63	0.40	4	63	0.41
5	52	0.87	5	52	0.87
6	32	0.94	6	71	0.96
7	25	0.94	6		
8	14	0.94	6		
9	13	1.33	7	44	1.39
10	13	1.33	7		
11	7	1.33	7		
12	11	1.38	7		
13	7	1.66	8	24	1.97
14	6	1.82	8		
15	5	1.94	8		
17	1	2.35	8		
18	2	2.35	8		
24	1	2.35	8		
32	1	2.56	8		
35	1	3.57	8		

a physical container" (19), and "Negative speech" (20) show a clear decrease in size (indicating that the bootstrap points show much less variation than before the recoding). The bias, indicated by the distance between the centroid of the bootstrap cloud (the black circle) and the sample component loading (the white square), has been strongly reduced as well.

A loading on a component may be positive in the original sample, and negative in some of the bootstrap samples, even after a reflection. In such a case, the confidence ellipse could contain the value zero on that component. If such a situation occurs for one component, and not for the other(s), and if the loading on that component is small, this may indicate that the corresponding variable does not make an important contribution to that particular component. Such a result allows for a simple interpretation of the solution (comparable to the simple structure pursued by a VARIMAX rotation). If, however, a confidence ellipse contains the origin, and the loadings are small,

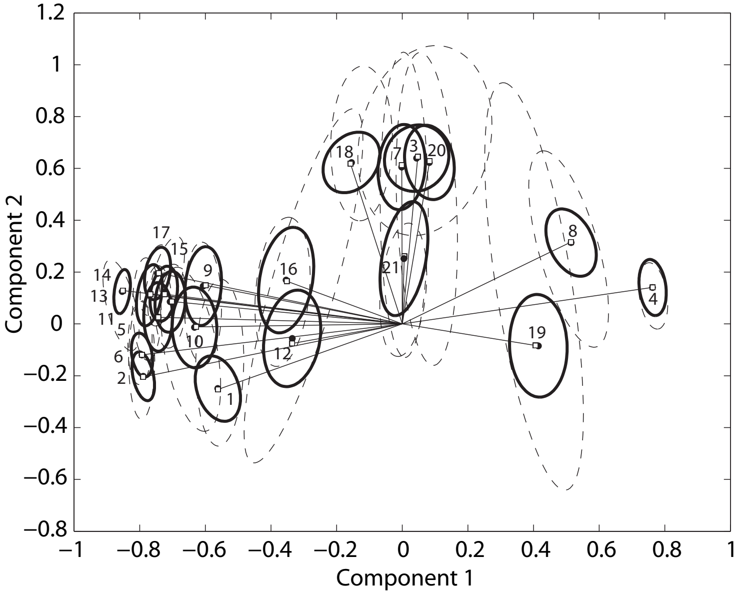

Figure 3.4: *90% confidence ellipses for the component loadings of the recoded ORCE variables. Black circles indicate the centroids of the bootstrap clouds after recoding. White squares represent the sample component loadings after recoding. The dashed ellipses indicate the bootstrap confidence regions for the component loadings from the original data (before merging categories). 1=Distress, 2=Nondistress, 3=Intrusiveness, 4=Detachment, 5=Stimulation, 6=Positive regard, 7=Negative regard, 8=Flatness, 9=Positive affect, 10=Positive physical contact, 11=Vocalizations, 12=Reads aloud, 13=Asks question, 14=Other talk, 15=Stimulates cognitive development, 16=Stimulates social development, 17=Facilitates behavior, 18=Restricts activity, 19=Restricts in physical container, 20=Negative speech, 21=negative physical action.*

the loadings of the corresponding variable are negligible on all components. In such a case, that particular variable can be left out of the analysis, without changing the results and the interpretation of the solution. All confidence ellipses in Figure 3.4 contain the value zero on only one component. Contrary to the results before merging (see Figure 3.3), the ellipses for "Negative speech" (20) and "Negative regard" (7) do not contain the value zero on the second component, which indicates that the loadings on the second component are substantially different from zero. The lower end of the confidence ellipse for "Negative physical action" (21) is close to zero, indicating that the contribution of this variable to the second component is questionable.

Confidence intervals for the variable transformations after merging

As the variable transformations are a key feature of the nonlinear PCA solution, inspection of their stability is called for as well. As was shown in the previous subsections, component loadings may be represented as points in a plot with the axes defined by the two components, and their confidence regions can be represented by ellipses. Category quantifications for ordinal and numeric variables, however, are single values that do not differ per component (see Chapter 2), and their 90% confidence regions have to be represented in a different way. We propose a method based on the transformation plot of the optimally quantified variable versus the original variable, consisting of the following steps:

1. For each variable, establish the bootstrap quantifications for each category.
2. Determine the 5^{th} and the 95^{th} percentile of those bootstrap quantifications.
3. Draw the borders of the 90% confidence region by plotting lines between the points that indicate the 5^{th} percentile and between those indicating the 95^{th} percentile.

Figure 3.5 displays transformation plots for the seven variables that showed the largest confidence intervals before merging, and for one other example variable ("Stimulates cognitive development"). The plots labeled 'A' show the confidence intervals *after* merging of the categories, and the plots labeled 'B' show the confidence intervals *before* merging. In each of the plots, the inner (bold) line indicates the transformation for the variable in the original sample, and the outer lines indicate the 90% confidence interval. It is important to note that the ordinality restriction holds within each bootstrap sample, but does not necessarily hold for the boundaries of the confidence intervals. Thus, for some of the variables (for example, the behavior scale "Stimulates cognitive development" (15) in Figure 3.5), the order of the category numbers is not maintained in the boundaries of the confidence interval.[5] The explanation for this phenomenon is that, if a category's quantification in the original sample is high compared to its bootstrap quantifications, and the preceding category's quantification is low compared to its bootstrap quantifications, the 95^{th} percentile of the bootstrap points for that particular category could be lower than the 95^{th} percentile of the bootstrap points for the preceding category. As a result, the upper boundary of the confidence interval would be decreasing.

For all of the variables in the plots labeled 'B' in Figure 3.5, the categories with the smallest marginal frequencies (i.e., the highest categories) are the least stable, which coincides with the results found for the component loadings. The variables with the categories showing the largest confidence intervals are also the ones with the largest confidence ellipses for the component loadings. These conclusions were to be expected, considering the fact that the correlations between a quantified variable and the principal components (indicated by component loadings) will differ when the quantification of a variable varies considerably between bootstrap solutions.

A comparison of the plots in Figure 3.5 after merging (labeled 'A') to the plots before merging (labeled 'B') shows that merging categories with small marginal frequencies leads to a considerable decrease in the size of the confidence intervals for the transformations. The extremely large confidence intervals for the negative behaviors "Restricts activity" (18), "Restricts in a physical container" (19), and "Negative speech" (20), and also for the positive behavior "Reads aloud" (12), have been diminished. The variable "Negative physical action" (21) still has quite a large confidence interval for its highest category, because it contained only eight observations (even after recoding). In general, the sizes of the confidence intervals have decreased notably, indicating enhanced stability of the category quantifications.

Confidence regions for the person scores after merging

For the construction of confidence regions for the person scores, we followed the same procedure as for the component loadings. In Figure 3.6 the confidence regions for the person scores after merging are displayed. The 2.5% largest ellipses, plus a random selection of 15 other ellipses have been displayed to enhance legibility of the plot. From this figure, we can conclude

[5] Note that the scale of the y-axis for this variable is larger than for the other variables to clearly display this effect.

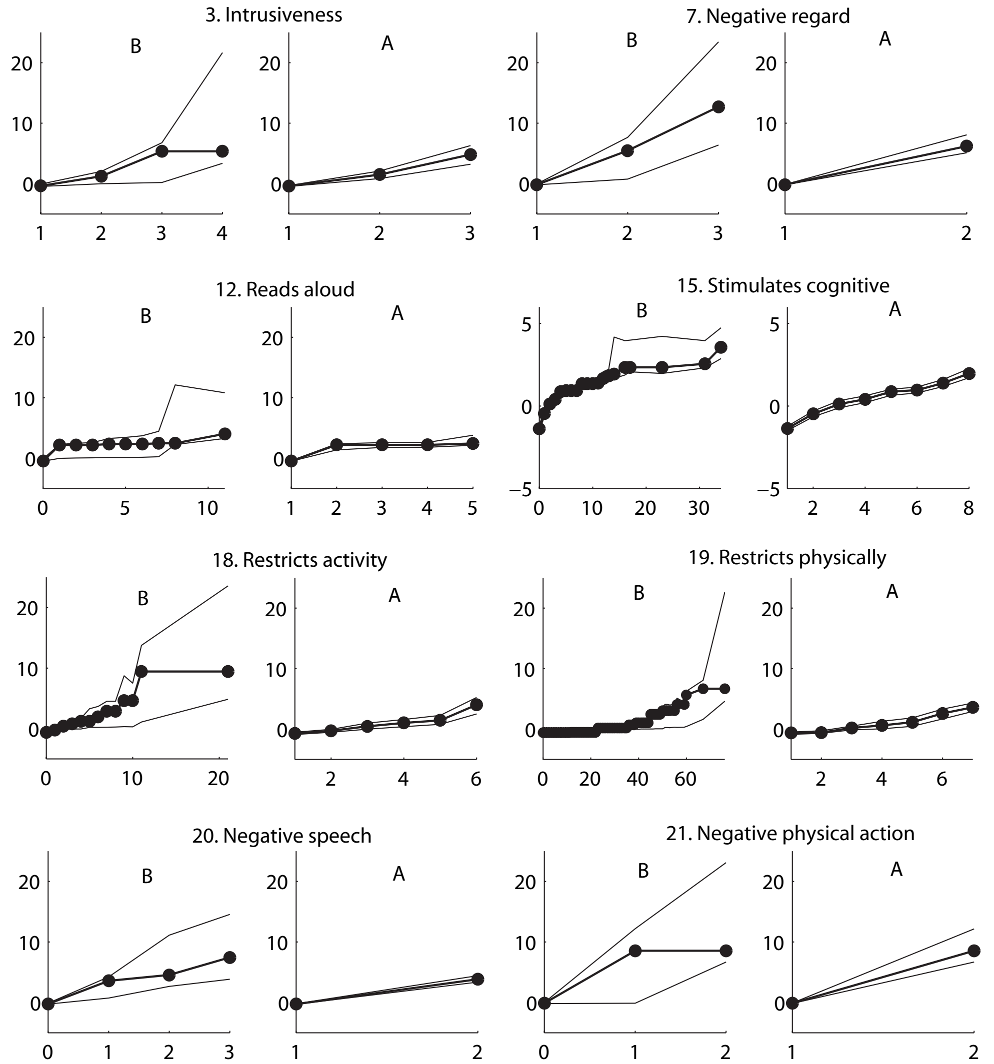

Figure 3.5: *Confidence intervals for variable transformations. The bold (inner) lines indicate the original variable transformation from the observed data. The outer lines indicate the 90% confidence intervals. The x-axis depicts the category numbers and the y-axis depicts the category quantifications. The plots labeled 'B' represent transformations before merging, and the plots labeled 'A' represent transformations after merging of categories with small marginal frequencies. Note that "Stimulates cognitive" is scaled differently from the other variables.*

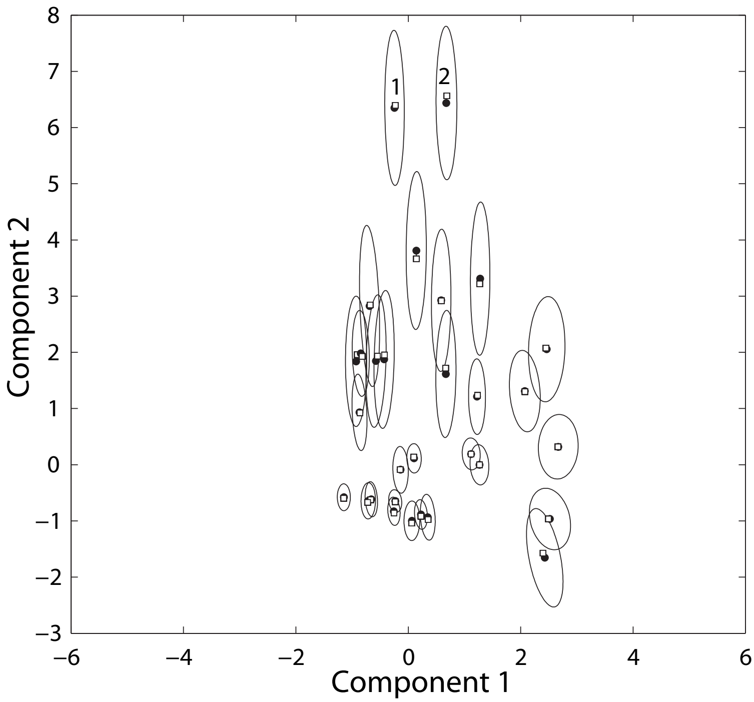

Figure 3.6: *90% confidence ellipses for the person scores from the recoded ORCE data. Black circles indicate the centroids of the bootstrap clouds. White squares represent the original person scores. The 2.5% (i.e., 15) largest confidence ellipses have been selected for display. In addition, a random selection of 15 other ellipses is displayed.*

that the confidence ellipses are quite small, even for the children who were outliers on the second component (indicated by the numbers 1 and 2).

Interpretation of the nonlinear PCA solution after merging

The recoding of the data has not changed the interpretation of the nonlinear PCA solution: If we compare Figure 3.4 to Figure 3.1, we see that the observed component loadings are in approximately the same locations, indicating that the loadings have not changed much after the recoding of the data. The percentages of VAF by the two principal components (equal to the sums of squared component loadings or the eigenvalues, divided by the number of variables) were only slightly smaller than those of the solution for the original data: 33.5% versus 33.9% in the first component, and 9.7% versus 9.9% in the second. In conclusion, merging categories with small marginal frequencies is quite effective in improving the stability of the nonlinear PCA solution. When, in practice, a researcher encounters data with relatively small marginal frequencies, merging of categories seems sound advice.

3.3.4 Comparing nonlinear PCA to linear PCA

To be able to conclude whether the nonlinear PCA solution, after merging the categories with small marginal frequencies, is sufficiently stable, we need a benchmark. Therefore we performed the same bootstrap study using linear PCA. In this way, the stability of linear PCA acts as the standard to judge the stability of nonlinear PCA. In previous research, linear PCA was found to be as stable as (or in some cases even more stable than) related methods, such as maximum likelihood factor analysis (MLFA) and independent component analysis (ICA) (for example, see Velicer, 1974; Velicer & Fava, 1998). Confidence intervals for linear PCA can be created using the bootstrap procedure (Timmerman et al., in press) or by the asymptotic approach of Ogasawara (2004). Timmerman et al. (in press) and (Markus, 1994) found, respectively for linear PCA and for MCA, that the bootstrap approach is more flexible and under most conditions more accurate than the asymptotic approach.

Because we applied the missing option "passive" to the missing values in the nonlinear PCA in Chapter 2, we wished to apply that same option in the linear PCA to prevent possible differences being due to differential treatment of missing data. However, as this particular treatment of missing data is not available in linear PCA programs[6], we performed linear PCA by using the CATPCA program with all variables treated numerically. We used the original data, without recoding, because that would have been the natural choice for a standard linear PCA analysis.

Comparing stability of nonlinear to linear PCA: The eigenvalues

Table 3.3 shows the eigenvalues and their 90% confidence intervals for both linear and nonlinear PCA. The eigenvalues for nonlinear PCA are only slightly less stable: the size of the 90% confidence interval for the VAF in the first two components is 0.63 and 0.49, respectively (0.57 and 0.44 for linear PCA), and for the total VAF 0.73 (0.65 for linear PCA). When the stability of the two eigenvalues jointly is captured by a 90% confidence ellipse, the area is 0.41 for nonlinear PCA and 0.32 for linear PCA. The cloud of bootstrap points for linear PCA (not shown) is considerably less tilted, displaying more variance in the first component than in the second component, whereas the bivariate

[6] Alternatively, missing in PCA may be handled by multiple imputation.

Table 3.3: *Eigenvalues and their 90% Confidence Intervals (C.i.) from the linear and nonlinear PCA solutions for the same data.*

	Linear PCA		Nonlinear PCA	
Component	Eig	C.i.	Eig	C.i.
1	6.60	6.33–6.90	7.03	6.78–7.41
2	2.01	1.83–2.27	2.03	1.87–2.36
Total	8.61	8.34–8.99	9.06	8.82–9.55

distribution for the eigenvalues in nonlinear PCA (shown in Figure 3.2) has almost equal dispersion in both components.

Comparing stability of nonlinear to linear PCA: The component loadings

To obtain a complete picture of the stability of the component loadings, we examine the areas of the confidence ellipses in two dimensions, which are displayed in Figure 3.7. Here we show ellipses for each variable separately, with regular lines indicating the ellipses for linear PCA and bold lines the ellipses for nonlinear PCA. The open circles indicate the centroids of the bootstrap clouds for linear PCA, and the solid circles the centroids of the bootstrap clouds for nonlinear PCA. By looking at the circles, differences in VAF between the two solutions can be inferred: The further away the center of an ellipse is positioned from the origin (the point with coordinates 0.0, 0.0), the more variance is accounted for.[7] By looking at the size of the ellipses, we can draw conclusions about the stability of the component loadings. Note that the scales of the plots differ across variables, such that the sizes of the ellipses cannot be compared at first sight. To facilitate comparing the two solutions, the ratio of the two areas is given in each plot as well, with the area for nonlinear PCA in the numerator, and the area for linear PCA in the denominator. If this ratio is 1.00 or close to 1.00, the areas are approximately equal. If this ratio is larger than 1.00, the area for nonlinear PCA is the largest, and if the ratio is smaller than 1.00, the area for nonlinear PCA is the smallest.

Eleven of the variables are approximately equally stable in both solutions; eight of those – 2, 4, 5, 6, 11, 13, 14, and 17 – are very stable, and three of those – 8, 9, and 10 – are somewhat less stable. For six of the variables –

[7] Note that some variables have positive and others have negative loadings on the first component.

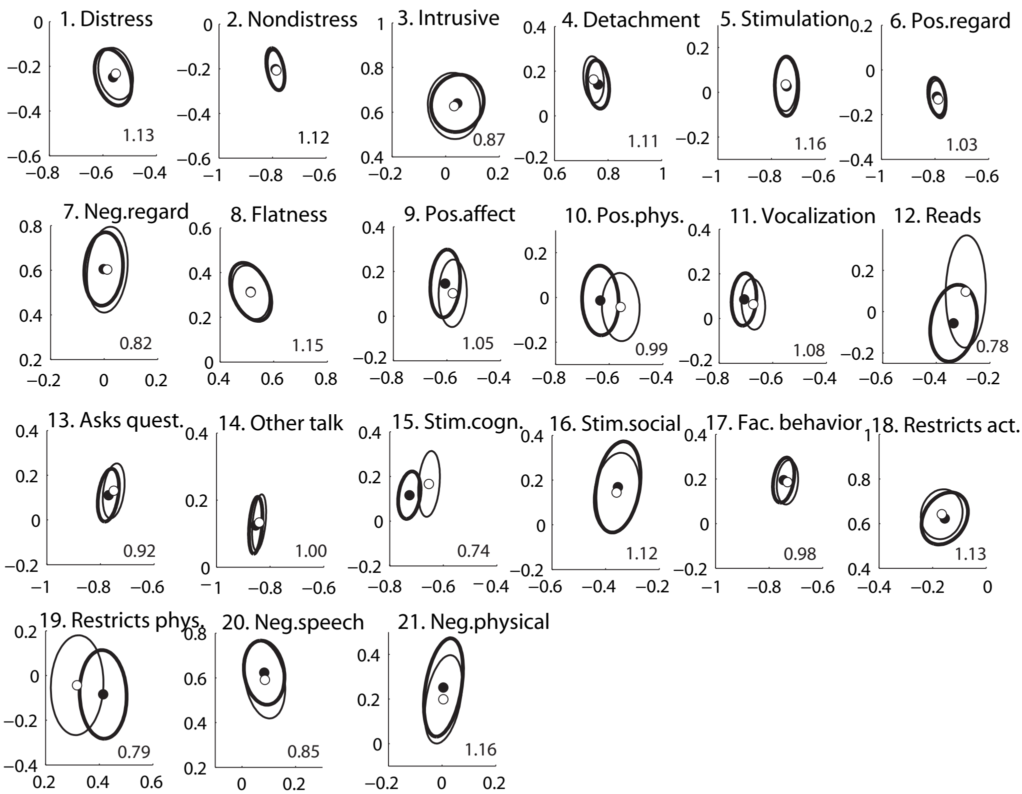

Figure 3.7: *90% confidence ellipses for the component loadings from the numeric solution (indicated by the regular line) and from the ordinal solution (indicated by the bold line). Open circles represent the centroids of the bootstrap clouds from the numeric solution, and solid circles the centroids of the bootstrap clouds from the ordinal solution. In each plot the ratio of the ordinal to the numeric solution is indicated.*

3, 7, 12, 15, 19, and 20 – the areas of the ellipses are considerably smaller in the nonlinear solution (with ratios of 0.87, 0.82, 0.78, 0.74, 0.79, and 0.85, respectively). The four remaining variables – 1, 16, 18, and 21 – show less stability in the nonlinear PCA solution (with ratios of 1.13, 1.12, 1.13, and 1.16). The sum of the areas of the confidence ellipses over all variables is 0.596 in the numeric solution and 0.572 in the nonlinear solution, and the mean ratio of the ellipses is 1.00, indicating no notable difference in the overall stability of the component loadings. The similarity of the stability in both solutions is remarkable, given the fact that the nonlinear PCA results are based on 1000 different transformations for each variable. We need to realize, however, that for didactical purposes the linear PCA is based on the data *before* recoding. This explains the fact that some of the ellipses that were very large for nonlinear PCA before recoding have become smaller after recoding than the corresponding ellipses for linear PCA. This applies to the variables 3, 12, 19, and 20 (see Figure 4), and this phenomenon suggests that the stability of linear PCA is also influenced by categories with small marginal frequencies.

3.4 Conclusions and Discussion

In this chapter, we used the nonparametric balanced bootstrap to investigate the absolute stability of nonlinear PCA, and presented a procedure for graphically representing 90% confidence regions for the eigenvalues, the component loadings, the quantified variables, and the person scores. We used the stability of linear PCA as a benchmark to evaluate the relative stability.

To start with, substantively, the results from nonlinear PCA for the data analyzed in this study are quite similar to the linear PCA results, which may leave the reader wondering whether it is worthwhile to use nonlinear PCA instead of linear PCA. Chapter 2 presents a clear illustration that includes nominal variables and shows the advantages of nonlinear PCA. In the current paper, the substantive similarities between linear and nonlinear PCA are a direct result of the focus on the comparison of the stability of the nonlinear and linear PCA solutions under comparable conditions (and linear PCA does not incorporate nominal variables). In this case, strong differences in nonlinear and linear results would suggest that the category quantifications in nonlinear PCA introduced appreciable uncertainty in the solution. Substantively, the situation will be dramatically different if the data set at hand contains variables among which strong nonlinear relationships exist. With such data sets, linear PCA will fail to deal with the nonlinear relational structures, although the outcome of the analysis may turn out to be very stable.

For the application data set, we may conclude that the bootstrap shows that the eigenvalues are quite stable. As Markus (1994) and Timmerman et al. (in press) found that the bootstrap procedure is somewhat conservative compared to asymptotic estimations of stability, we can be quite confident about the stability of nonlinear PCA for these data. We demonstrated, however, that categories with small marginal frequencies cause instability in the transformed variables and in the corresponding component loadings. The explanation for these results is that persons scoring in rare categories sometimes do not appear in a bootstrap sample, and this has a large effect on the category quantifications for these rare categories. The effect of small marginal frequencies of a category is not restricted to that particular category; categories (of other variables) that are related to that category will also become more unstable. Therefore, the degree of instability of a variable and its categories is not merely dependent upon the univariate distribution of that variable, but also upon its relationship with the other variables. The problem with small marginal frequencies is not exclusive to the bootstrap; the asymptotic results in Markus's study also showed more variance for categories with small marginal frequencies, although to a lesser extent.

The instability due to small marginal frequencies can be decreased by leaving variables measuring rarely occurring behaviors out of the analysis. If the variables in question have small component loadings, and the eigenvalues are well-separated, they may be left out without much influence on the results (Gifi, 1990). In the data set under study, however, the unstable variables had reasonably high component loadings, and leaving them out would be insensitive to the critical substantive meaning of rare categories: In this data set, categories with small marginal frequencies define the second dimension which surfaces negative behaviors that may be harmful to the child. A much better strategy in that case, is to merge the rare categories. We have shown that merging rare categories increases the stability of the quantified variables and the component loadings, while the eigenvalues (and thus the variance-accounted-for) remained the same. A theoretical proof has been given in Gifi (1990, p. 397).

It is worthwhile to note why eigenvalues can be stable while the category quantifications and the component loadings are not. If the instability in the component loadings is caused by an indeterminacy of the sign of the loading (apart, of course, from a general reflection), the eigenvalues, which are the sum of squares of the component loadings, will not be influenced by the instability of the component loadings, and thus can still be stable.

In the comparison of nonlinear and linear PCA for the data in this study, the eigenvalues and component loadings from both methods were approximately equally stable. Keeping in mind that nonlinear PCA estimates a large number of extra outcome values (the category quantifications), the decrease in stability from numeric to ordinal treatment of the variables is small. Of course, the stability results for linear PCA may be improved by merging categories as well; this, however, has not been the purpose of the analyses. We wished to demonstrate the effectivity of merging categories with small marginal frequencies by showing that this operation resulted in component loadings for nonlinear PCA that were as stable as or even more stable than the original loadings in linear PCA.

The results of this study, although in line with results of comparable studies in nonlinear multivariate analysis (for example, Markus, 1994), are based on one real-life data set. Simulation studies would be called for to further confirm the bootstrap results from the current study. Combining previous findings from the literature and theoretical knowledge about nonlinear PCA, we can nevertheless formulate some general guidelines for researchers who wish to perform a similar bootstrap study.

- If you have a small sample, do not expect the results to be very stable. The example data set in this chapter contained almost 600 children, and we found reasonably stable results. Markus (1994) found that the size of the confidence regions decreased with increasing sample size. Also, she found that as long as $n \geq 200$, the researcher can expect coverage percentages within an acceptable range. Although smaller samples (just as in other analysis methods) imply less accuracy, the bootstrap gives a correct impression of the stability.

- Use enough bootstrap samples. According to Markus (1994), using 1000 samples will typically be adequate, as we showed in the current study. Given the low cost of computer time, we encourage researchers to use more bootstrap samples.

- Use an orthogonal Procrustes procedure to rotate the bootstrap component loadings towards the original component loadings. This is particularly crucial if the eigenvalues of the principal components are close together. Orthogonal Procrustes rotation automatically takes care of reflection, which is always necessary, since the sign of each component loading is undetermined. Thus, also if the first component is sufficiently dominant over the second (as was the case in the current study), Procrustes rotation will be useful. A Procrustes rotation will never have a harmful effect on the stability results. If a VARIMAX rotation has been applied to the original PCA solution, the Procrustes rotation should be targeted to the VARIMAX solution. If however, the stability of the VARIMAX orientation itself is of interest, each bootstrap analysis outcome should be subjected to a VARIMAX rotation (see Timmerman et al., in press).

- Merge categories with small marginal frequencies. Such categories lead to unstable results for the category quantifications, resulting in unstable quantified variables, component loadings and person scores. In the current study, we drew the line at 2.5% of the cases, which is in accordance with the theory behind normal distributions, where the 2.5% observations in the tails of the distribution are usually considered extreme. We recommend that researchers first perform a nonlinear PCA of the original data, and merge categories that have produced equal or almost equal category quantifications (since this operation will not influence the outcome in any substantial way). Then a bootstrap study can be performed on the recoded data, and the results checked for unstable categories. If categories with relatively small marginal frequencies cause unstable results, these categories should be merged, and the bootstrap study should be repeated to check whether the results are acceptable after recoding.

- An ellipse covering 90% of the bootstrap points is both a simple and insightful device to display a confidence region. To find out whether a confidence ellipse gives a reasonable representation of the spread in the bootstrap cloud, the researcher should check the distributions of the bootstrap points.

In the current study, we excluded the multiple nominal variables analyzed in Chapter 2, because the incorporation of multiple nominal variables would have made it impossible to compare the stability of nonlinear and linear PCA. However, the stability of variables with a multiple nominal analysis level is a topic of substantial interest. In future research, we will extend the bootstrap procedure in the SPSS macro files to incorporate multiple nominal analysis levels. This extension will take some adjustment of the macros, as each category of a multiple nominal variable obtains a separate quantification in each principal component instead of a single quantification across components. However, we do not expect the stability results for multiple nominal variables to differ much from those found by Markus (1994). The problems she stated for multiple nominal variables (such as categories with small marginal frequencies causing instability) also appear to apply to ordinal variables.

Besides for establishing the stability of analysis results, the bootstrap may also be used to estimate the population value. For instance, for a particular category, the average of the quantifications obtained from all bootstrap samples may be viewed as a better and more stable approximation to the population category quantification than the quantification in the observed sample. Bootstrapping the objects and averaging the results is also referred to as "bagging" (Breiman, 1996).

Because of the possibly somewhat conservative nature of confidence ellipses, it would be interesting to try out alternative methods for representing confidence regions. Potentially promising methods include convex hulls (in which irregular shapes instead of ellipses are used to represent confidence regions), minimum volume ellipses (Rousseeuw, 1984) (which involves selection of the confidence ellipse with the smallest possible volume), or two-dimensional versions of boxplots, called bagplots (Gardner & le Roux, 2003; Rousseeuw, Ruts, & Tukey, 1999).

Several nonparametric alternatives to the specific bootstrap approach used in this paper deserve exploration. An alternative to the bootstrap is the Jackknife. Here, a random group of objects is left out from the data set, and this is repeated to obtain different Jackknife samples. A special case of the Jackknife is the 'leave-one-out' method, where only one object is left out. Bootstrap distributions show more variability in the estimated parameters than Jackknife distributions, when the number of objects left out in each sample is small (for instance, in the leave-one-out method, the Jackknife samples only differ by one object). Efron and Tibshirani (1993) showed how estimates of bias and standard errors should be adjusted to that difference in variability. Also, the Jackknife fails if the estimated statistic is not smooth. We expect the bootstrap to perform better in nonlinear multivariate analysis. A simulation study would be needed to show whether this is true. Variants of bootstrap confidence intervals with improved coverage percentages when the bootstrap results show bias and nonnormality, such as the bias corrected and accelerated (BC_a) confidence intervals, have proved to be effective in other research (Efron & Tibshirani, 1993; Timmerman et al., in press) and may be used in nonlinear PCA as well.

A question of potential interest would be whether the contribution of a variable to the nonlinear PCA solution is statistically significant. The bootstrap procedure, however, does not produce exact $p-$values. For that purpose, permutation tests could be used (for example, see Buja & Eyuboglu, 1992; Fisher, 1935; Good, 2000). Permutation tests are a nonparametric approach to inference especially suited for nonlinear or categorical data analysis. In a permutation test, the correlational structure of the data set is deliberately destroyed in each new sample. The null distribution of a result can be empirically generated. The original outcome value should differ sufficiently from its null distribution to be significant. Another nonparametric tool for establishing the validity of a nonlinear PCA solution is cross-validation, in which the generalizability of a solution is investigated by applying nonlinear PCA to an initial data set and applying the estimated values for the results to a cross-validation data set. All these nonparametric methods can be subsumed under the label resampling. These methods form an extremely important toolbox to complete a nonlinear multivariate analysis.