Exploratory factor analysis revisited:

How robust methods support the detection of hidden multivariate data structures in IS research

H. Treiblmaier and P. Filzmoser

Forschungsbericht SM-2009-5
November 2009

Kontakt: P.Filzmoser@tuwien.ac.at
Exploratory Factor Analysis Revisited:
How Robust Methods Support the Detection of Hidden Multivariate Data Structures in IS Research

Horst Treiblmaier*
Institute for Management Information Systems
Vienna University of Economics and Business
Augasse 2-6, 1090 Vienna, Austria

Peter Filzmoser
Department of Statistics and Probability Theory
Vienna University of Technology
Wiedner Hauptstraße 8-10, A-1040 Vienna, Austria

Abstract

Exploratory factor analysis is commonly used in IS research to detect multivariate data structures. Frequently, the method is blindly applied without checking if the data at hand fulfill the requirements of the method. In this paper, we investigate the influence of sample size, data transformation, factor extraction method, rotation and number of factors on the outcome. We compare classical exploratory factor analysis with a robust counterpart which is less influenced by data outliers and data heterogeneities. Our analyses reveal that robust exploratory factor analysis is more stable than the classical method.

Key words: Exploratory Factor Analysis, Factor Analysis, Robust Statistics

---

1 Augasse 2-6, A-1090 Vienna, Austria
Tel.: +43/1/31336/4480
Fax: +43/1/31336/746
E-Mail: Horst.Treiblmaier@wu.ac.at
Exploratory Factor Analysis Revisited:
How Robust Methods Support the Detection of Hidden Multivariate Data Structures in IS Research

Introduction

The wealth of empirical research papers in top-tier IS journals suggests that the information systems discipline frequently uses quantitative research methods. In order to enhance the informational value of the results, there is an ongoing discussion on how to improve existing methods, avoid potential sources of error, and ultimately introduce new ways of data analysis [6, 9, 27]. Many methodological papers being published in leading IS journals focus on topics which are ‘in fashion’ in the community. Examples include the huge popularity of PLS in IS research [7] or, more recently, the interest in formative indicators [36].

In recent years, a strong focus has been placed on the confirmation of theory-based models, for example by using LISREL or PLS. In order to test their hypotheses, researchers conduct surveys and rely on the results of previously tested measurement instruments. They concentrate on confirmatory aspects, such as achieving a sufficient fit for the model, but often overlook the important role of exploratory statistical techniques, which are frequently used prior to a confirmatory study.

In general, the decision of whether to use either an exploratory or a confirmatory procedure depends on the availability of a theory that explains how the hypothesized construct corresponds to the respective items. In spite of the amount of attention which is given to confirmatory techniques, Conway et al. [11, p. 147]
emphasize that "researchers tend to make better decisions when EFA plays a more consequential role in their research."

Additionally, there has been a long-standing discussion about the eligibility of survey data for various data analysis methods. Especially data generated using Likert-type items – a common technique in IS research – frequently do not meet the requirements of the method applied, such as normal distribution of the individual variables or multivariate normal distribution. In many cases, latent variables represent users’ attitudes, norms, values and intentions. To measure these, respondents are typically asked to express their level of agreement on a scale (e.g.: 5-point Likert scale: "strongly agree", "agree", "neutral", "disagree", "strongly disagree"). Although the underlying latent construct may be seen as a continuum, the items are measured on an ordinal scale. In this paper, we use a 100-point slider in an online survey to create a metric scale and examine the use of robust factor analysis rather than classical factor analysis for detecting hidden data structures. It consists of three major parts.

- First, we discuss the capabilities and problems of EFA, the latter of which are frequently overlooked in current IS research. We especially want to demonstrate that researchers have a multitude of options such as data transformation, factor extraction method, rotation method, and choosing the optimum number of factors when conducting EFA. If they try out various configurations and disregard basic statistical assumptions, they may eventually obtain the results they desire. Previous research has shown that many applications of EFA in the literature are to some extent questionable [38]. Additionally, we introduce the concept of robust factor analysis in IS research, which proposes a viable alternative that eliminates many of the shortcomings of non-robust exploratory factor analysis. By reducing the influence of outliers, robust statistics helps to overcome the strict model assumptions of classical methods, such as the necessity of
normal or multivariate normal distribution, which is rarely given in survey-based research.

- Second, we present several examples of how EFA is currently used in IS research. Although most of the attention is given to confirmatory procedures, EFA is frequently used when new scales are developed or the validity of a measurement model is assessed.

- Third, we use robust factor analysis to examine data which we gathered in an Internet-based survey and compare the results with those obtained by using classical factor analysis and a qualitative study. Finally, we discuss our findings and their implications for future research.

**Exploratory Factor Analysis**

Factor analysis, which was popularized by Charles Spearman [45], has been one of the most widely used statistical techniques in psychological research [14]. It is therefore of paramount importance for all social sciences investigating human behavior. Its major objective is to reduce a number of observed variables to fewer unobserved factors in order to enhance general interpretability and to detect hidden structures in the data. Frequently, these structures are used as constructs in sophisticated models displaying aspects of human behavior. Table 1 illustrates that researchers using exploratory factor analysis have to make decisions on issues such as robust versus non-robust procedures, data transformation, factor extraction method, factor rotation and the number of factors retained. All of these decisions strongly influence the final results. Researchers need to be aware that there are many factor solutions to one correlation matrix and that their final solution represents just one of many possible choices [48, p. 420]. Accordingly, a statement by Chin et al. [8, p. 350] highlights the researcher’s freedom in and responsibility for determining the output: "Though the scree plot suggested a smaller number of factors than the Eigenvalue rule, we opted to err on
the conservative side by including more factors to avoid the possibility of missing relevant ones”. Furthermore, it has to be pointed out that the list of options shown in Table 1 is by no means exhaustive, but only presents a selection of choices commonly made.

--- Table 1 about here ---

Robust Versus Non-robust

Exploratory factor analysis relies on the estimation of the correlation matrix. Once this correlation matrix is available, the loadings and uniquenesses, and subsequently the factor scores are estimated. The correlation matrix is usually estimated with the sample correlation matrix, which is the empirical sample covariance matrix standardized by the empirical variances. This classical approach is most adequate if the data are multivariate normally distributed. However, if the data distribution is deviating from this ideal distribution, the estimated correlation matrix can be severely biased. Figure 1 shows this effect. The estimated correlations are visualized by the ellipses which, in case of bivariate normal distribution, are supposed to contain 97.5% of the data points. The ellipse with the dashed line is based on the classical estimates, the one with the solid line uses robust estimates. While the latter ellipse covers the homogeneous majority of observations, the ellipse with the classical estimates is blown up by the deviating data points, which leads to an unrealistic estimation of the correlation, which in turn constitutes the basis for the factor analysis.

--- Figure 1 about here ---

Several alternative approaches have been introduced in the literature to reliably estimate the correlation matrix in presence of deviations from multivariate nor-
mality. A prominent way is to use the Minimum Covariance Determinant (MCD) estimator, which looks for a subset of observations (e.g. at least 50% of the observations) with minimal determinant of the empirical covariance matrix. The covariance matrix formed by this subset is multiplied by a consistency factor, and it is highly robust (up to 50% of the observations could deviate from the multivariate normal distribution). The robust correlation matrix is obtained by standardizing the robust covariances with the square-root of the diagonal elements. The MCD estimator [41] is attractive for its high robustness, but also for a fast algorithm being available in several widely used statistical software packages, such as R, S-Plus, SPSS, SAS, and STATA.

Plugging in the robust MCD correlation matrix into factor analysis also leads to a highly robust estimation of the loadings and uniquenesses [37]. Whereas outliers could have an unduly high influence on the estimation of these parameters using the classical correlation matrix, their influence is limited when using the MCD correlation matrix.

**Sample Size**

Clearly, larger samples outperform smaller samples due to the reduction in the probability of errors, more accurate population estimates and a better generalizability of the results. If the overall sample size is too small, errors of inference can occur [33]. Various recommendations pertaining to sample size can be found in the literature. While some authors highlight the importance of absolute sample size [10], most researchers focus on the ratio between subjects and variables. Recommendations frequently include ratios of 5:1 or 10:1 [20]. However, MacCallum et al. (2001) illustrate that this might be an oversimplification. They show that population factors in data can be adequately recovered if communalities are high and that sample size is of minor importance in this case. When researchers
are confronted with low communalities and poorly overdetermined factors, they recommend sample sizes that are much larger than usually suggested (e.g. a 20:1 ratio) [28]. For a robust method the sample size should be even larger because deviating data points will be downweighted.

**Data Transformation**

Essentially, data transformation is used to obtain a particular type of distribution, which may be, for example, normal or symmetric. In addition, it is used to establish a simple systematic relationship between an independent and a dependent variable and to stabilize the variance. Real-world data are frequently characterized by skewness, heteroscedasticity and outliers [42], which should be taken into account when applying statistical procedures, depending on specific requirements. As will be discussed below, the proper application of a factor extraction method is contingent upon the distribution of the data. If robust factor analysis is used, it is sufficient that the majority of the observations fulfills the distributional assumptions.

**Factor Extraction Method**

The overall goal of the research determines the selection of the best factor extraction method. Broadly, most models can be categorized as either a component model or a common factor model [18]. The goal of a principal component analysis, which is by far the most popular type of a component model, is to retain as much of the original measures’ total variance as possible. Common factor models, meanwhile, differentiate between variance attributable to common factors and variance caused by unique factors. Conway and Huffcutt (2003) therefore conclude that “if a researcher’s purpose is to understand the latent structure of a set of variables (which will usually be the case), then the use of a common factor model such as principal axis or maximum likelihood factoring represents a
high-quality decision. (...) Given that most researchers do attach meaning beyond the observed variables, the common factor model will generally be the better choice. [11, p. 150] Accordingly, Preacher et al. (2003) state that "it is strongly recommended that principal component analysis (PCA) be avoided unless the researcher is specifically interested in data reduction [38, p. 39]", and Widaman (1993) argues that "... the results suggest that principal component analysis should not be used if a researcher wishes to obtain parameters reflecting latent constructs or factors [55, p. 263]."

Furthermore, the application of the correct factor extraction method depends upon the distribution of the data, which is frequently neglected in social sciences research. A multivariate normal distribution is required when using maximum likelihood as the factor extraction method, whereas a principal component analysis and a principal factor analysis require elliptical symmetry. In these cases, normal distribution is not a prerequisite, but the results may still be strongly influenced by the occurrence of non-normally distributed data and outliers because of their dependence on the correlation and the covariance matrix [40]. In such cases a robust method is preferable.

**Rotation Method**

A number of factor rotations exist, which differ as to the correlations between the factors. Orthogonal rotations, such as Varimax, Quartimax and Equamax, do not allow for correlations, whereas oblique rotations, such as Oblimin, Quartimin and Promax, consider correlated factors. Previous research has shown that an orthogonal method, e.g. Varimax, produces stable results [40]. However, when factors are actually correlated, an orthogonal rotation produces an unrealistic solution, while an oblique rotation better reproduces reality [11]. Additionally, an oblique rotation does not require factors to be correlated, so that the correlations
between the factors will be close to zero, if the actual data structure is orthogonal [22]. Therefore, oblique rotation is preferred by some researchers [14, 16].

**Number of factors**

There are multiple ways to determine the number of factors to be extracted from the data. Besides using an *a priori* criterion – which may be useful in replication studies – a researcher may decide to use Eigenvalues (latent root criterion) greater than 1, a visual scree plot, or a specific amount of variance to be explained by the factors [20]. Researchers may be confronted with the problem of specifying too few factors (underfactoring) or too many factors (overfactoring). Previous research suggests that the latter leads to fewer errors when factor loadings are estimated [15]. However, specifying too many factors might lead to the creation of constructs with little theoretical value [14]. Outliers or deviating data points can lead to an unrealistic estimation of the number of factors because they can artificially increase the Eigenvalues of the correlation matrix.

**Further Issues**

In addition to the issues discussed above, exploratory factor analysis poses several problems which are frequently overlooked. One major error is the assumption that the measurement errors of the items are uncorrelated [39]. Another important requirement often neglected is the metric measurement of the variables. A further problem – in addition to the proper application of factor analysis – is the interpretation of the results, which is controversially discussed in the literature. One major issue is the minimum threshold for validity. Tinsley and Tinsley (1987, p. 422) state that factor loadings of .30 or higher, which explain approximately 10% of a variable’s variance, should be considered when interpreting a factor [48]. Similarly, Tabachnick and Fidell (2001) go as low as .32 as the minimum acceptable factor loading for an item [47]. Interestingly, Peterson (2000) reports
the same number as the average factor loading in a number of studies using exploratory factor analysis [35]. In social sciences research, items tend to load on a number of factors, so that the threshold values are usually more stringent. Some authors suggest that a factor loading higher than .60 on the "parent factor" and a loading of less than .40 on a "foreign factor" indicate convergent and discriminant validity of the constructs [8, 51].

**Exploratory Factor Analysis in Information Systems Research**

In order to assess the relevance of factor analysis for information systems research, we performed a full-text search (i.e. including both citation and document text) in the databases EBSCOhost and the ProQuest using three different search terms ("factor analysis", "exploratory factor analysis" and "principal component analysis"). We decided to concentrate on four top IS journals, which frequently publish empirical research and demand a high level of methodological rigor: *Information Systems Research (ISR)*, *MIS Quarterly*, *Journal of Management Information Systems (JMIS)*, *Information & Management (I&M)*. The results can be found in Table 2, with the figures in brackets indicating the number of papers published since the year 2000. Even when taking into account that the results partially overlap and not all papers actually carry out an EFA, the findings clearly indicate that techniques of factor analysis are frequently used and discussed in empirical IS research.

--- Table 2 about here ---

Exploratory factor analysis is frequently applied in order to discover patterns of multidimensional constructs, which are subsequently used for the development
of measurement scales. Especially when new frameworks or scales are developed, EFA plays a major role in detecting hidden data structures. Even though the major focus of many current papers published in IS top journals lies on model confirmation rather than on scale development, researchers often use pre-studies to develop or refine their measurement instruments [29]. Most frequently, however, EFA is applied as a kind of pre-study to confirm the validity of the scales used. Examples will be given below to illustrate the practical application of EFA.

**Development of New Frameworks and Scales**

In order to uncover the underlying data structure of 118 items pertaining to consumers’ perception of data quality, Wang and Strong (1996) use a series of factor analyses. As a result of their exploratory research, they identify the most important dimensions of data quality, viz. intrinsic, contextual, representational, and accessibility data quality. In order to test for the stability of their results, they vary the number of factors and eliminate those items with insignificant loadings [53]. Torkzadeh and Dhillon (2002) use a sophisticated approach to identify factors influencing the success of Internet commerce. They conduct a survey with two phases. The first one is intended to reduce the overall number of items, while the second one served to fine-tune the instrument. Most notably, they use different methods of rotation (i.e. orthogonal and oblique) and different procedures to determine the number of factors (i.e. Eigenvalues and scree plot) to make their results more stable [49].

**Validation of the Measurement Model**

Exploratory factor analysis is frequently used to validate a measurement model. Zhu and Kraemer (2005), for example, compare the results from a principal component analysis (Equamax rotation) with the results from a confirmatory fac-
tor analysis. They use a structural equation modeling approach (PLS) to determine the robustness of the measurement model [56]. Koh et al. (2004) report that the results of an exploratory factor analysis (principal component analysis with orthogonal and oblique rotation) produce their hypothesized factor solution [25]. Similarly, Ba and Pavlou (2002) use EFA with Varimax rotation in order to account for convergent validity [2] (see also [3]). This procedure is proposed by Segars and Grover (1993), who point out that "... such results provide evidence of convergent and discriminant validity of scale items". However, they also note that "exploratory factor models provide no explicit test statistics for ascertaining whether convergent and discriminant validity are achieved" [43, p. 519]. Accordingly, Karimi et al. (2004) stress that "the more commonly used exploratory factor analysis does not validate the convergent and discriminant validities of latent variable indicators" [24, p. 185]. Table 3 shows examples of EFA applied in IS research, as discussed above, and states the details reported by the authors.

--- Table 3 about here ---

**Methodology**

**Survey**

In order to compare various methods of factor analysis, we decided to gather primary data rather than rely on an existing dataset. We developed a research question that is of interest to IS researchers from various sub-disciplines. Current customer relationship management (CRM) literature describes how companies strive to build (online) relationships with their (prospective) customers. However, when scrutinizing the general concept of "relationship", it turns out that many definitions remain vague and concentrate on describing activities rather than the core terms [21, 50]. Existing literature indicates that many factors such as usa-
bility, ease of use [1], enjoyment [51], service [31], interactivity [34] and individualization [26] might be responsible for perceived online relationships between customers and vendors. With "perceived online relationship" being a supposedly multidimensional construct, it represents an ideal starting point for our research.

One of the major goals in the beginning of many research projects is to ensure a sufficient level of content validity, which can be defined as the degree to which items in a measuring instrument represent the content universe to which the instrument will be generalized [12, 5]. We therefore created a pool of items by conducting a literature research in IS and marketing papers dealing with (online) relationships, (e)CRM and related topics. In this first phase, we tried to generate as many items as possible, even if this led to redundancy, which was dealt with in a subsequent phase [49]. We developed a total of 24 items, which were used in the literature as antecedents of online relationships between consumers and vendors. In a next step, we conducted an online survey in cooperation with a newspaper company, which included a link to our questionnaire in two weekly newsletters sent out to a total of 85,500 registered recipients. No incentive was given for filling out the questionnaire. The survey was carried out between 08/25/04 and 09/16/04. We used slider bars ranging from 1 ("strongly disagree") to 100 ("strongly agree") and gathered a total of 396 responses. After removing incomplete records, 389 usable answers remained. The overall goal of the survey was to measure constituents of online relationships between customers and companies. Although the original questionnaire was in German, no translation and back-translation process was necessary, since we did not use items from existing literature. Instead, the results were translated by the authors and double-checked by a native speaker. The low response rate of 0.5% can be attributed to the fact that we did not use an incentive and only a part of the recipients actually read the whole newsletter. In our case, this is no major problem since the objec-
tive of this study is not the generalization of the results (external validity), but rather the analysis of a factor structure.

**Scales**

The terms nominal, ordinal, interval and ratio scales were coined in the 1940s by S.S. Stevens, an American psychologist. Although these terms have faced strong criticism [52], this basic classification still prevails in the scholarly literature. An important issue is the treatment of survey data, which are frequently collected in Likert-type scales. Integer values are then assigned to each category, and the analysis is carried out as though the data had been measured on an interval scale, which may lead to biased results. This is especially the case when the product moment correlation coefficient is used [32, p. 391].

In order to be able to use robust factor analysis, we had to develop a measurement instrument that better represents a ratio scale than the frequently used Likert-type scales do. Typically, a loss of information occurs, when categorizing an unobserved continuous variable into an ordered categorical variable [46]. We therefore decided to use a magnitude scale, which represents a valid and reliable alternative to category scales. This type of measurement was popularized by S.S. Stevens and is based on the assumption that there is no fundamental difference between physical measurement and psychophysical measurement [17].

We created a visual analogue scale (VAS), which is simply a line with well-defined end-points on which the respondents indicate their level of agreement [19].

--- Figure 2 about here ---

Figure 2 shows a screenshot of the questionnaire we used. Based on a number of pretests, we decided to color the sliders red at both ends and blue in the mid-
dle. This makes it easier for respondents to differentiate among different levels of agreement, and it also prevents the accumulation of extreme values. In order to avoid default values, the users actually had to click on the survey for the slider to appear. This enabled us to count missing values as well.

**Results**

Prior to our comparative analyses, we tested the eligibility of the data for factor analysis by using the Kaiser-Meyer-Olkin measure of sampling adequacy (MSA) (Kaiser et al. 1974). An MSA value of .86 indicated a good (“meritorious”) factorability of the correlation matrix\(^2\).

**Comparison between Classical and Robust Factor Analyses**

In a first step we used a Shapiro-Wilk test [44] in order to test whether our sample has a normal distribution. We found that none of our variables was normally distributed. Since factor analysis is sensitive to non-normally distributed data, we used a logit transformation to create a better symmetry and to avoid the default range of 1 to 100. Nevertheless, normal distribution could not be achieved for any of the variables, since many of the original data values were equal to 1 or 100, thus corresponding to the extreme positions of the slider in the VAS. Therefore, a power-transformation or a Box-Cox transformation does not yield normally distributed data. Since the maximum likelihood method, which requires multivariate normally distributed data, is therefore not appropriate, Principal Factor Analysis (PFA) was our choice for the factor extraction method. PFA is based on a decomposition of the (reduced) correlation matrix, which is sensitive to severe deviations from an elliptically symmetric form of the data distribution and to out-

\(^2\) We also performed a Bartlett sphericity test, which was statistically significant \((p < .001)\), thus indicating the eligibility of the data. This test requires normally distributed data and is sensitive to deviations from this assumption. Since our data are non-normally distributed, this result has to be interpreted with caution.
liers in the data. Hence, PFA based on a robust correlation matrix, which results in a robust factor analysis (FA), will be more appropriate for the data analyzed in this paper [37]. To compute the robust correlation matrix, the MCD (Minimum Covariance Determinant) estimator was used, because it is highly robust, fast to compute, and widely available in statistical software packages [41]. Both the robust and the non-robust (classical) FA suggested 5 to 6 factors according to the scree plot and the Eigenvalue criterion. We opted for the larger number of factors in order to avoid underfactoring. An orthogonal rotation such as Varimax resulted in a very dominant first factor and in a rather weak discrimination of the factors according to their loadings. Therefore, an oblique rotation seemed to be the better choice, since it is reasonable to expect that the several dimensions of online relationships correlate with each other. We used Oblimin rotation, but observed very similar results for a variety of other oblique rotation methods (Quartimin, Covarimin, McCammon, Promax, etc.). The resulting biplots of the classical and the robust FA for the first pair of factors are shown in Figure 3.

--- Figure 3 about here ---

The biplots in Figure 3 show striking differences in both the configuration of the variables (arrows) and the configuration of the observations. The latter were plotted using two different symbols: "+" if the observation was identified as an outlier by the robust method, and "." for all other observations. Obviously, the factors in the classical analysis were "attracted" by the outliers, because the factors point in their directions (see Figure 3, left). This is different for the robust analysis, which focuses on the core of the data.

The loading plots shown in Figure 4 illustrate the differences in factor loadings for all factors of the classical and the robust analysis. The horizontal axis is scaled according to the relative amount of variability explained by each single
factor in the FA model, excluding the unexplained part of the variability (uniqueness) of each variable. Additionally, the percentage values at the top display the cumulative explained variance for the total data variability. It is thus possible to see at one glance how much of the total variance is explained and how important the single factors are for this explanation. The vertical axis is scaled from +1 to –1 and shows the factor loadings. Dashed lines at values of +0.5 and –0.5 help to distinguish the important (> +0.5, < –0.5) from the less important variables. Names of variables with absolute loadings of < 0.3 are not plotted because their contribution to the factors is negligible [40].

--- Figure 4 about here ---

Figure 4 illustrates the difference of the biplots shown in Figure 3 with some factors changing their order due to their relative importance. Especially the classical method is sensitive with respect to changes in the order of the factors, because outliers artificially increase the explained variance of some factors, giving them unduly more importance. There are also major differences in the composition of the loadings for both factor analyses, which will also lead to different interpretations.

**The Influence of Data Transformation**

In order to assess the effect of data transformation, we also analyzed the original untransformed data. In a further step we simulated a 5-point Likert scale by creating five categories with intervals from 1-20, 21-40, 41-60, 61-80, and 81-100. These data resemble the results that would be obtained from a standard questionnaire using a 5-point Likert scale. To make them comparable to previous results (see Figure 4), we also applied PFA (robust and classical) with 6 factors and Oblimin rotation. However, the robust method does not work for the categor-
ical data, because the algorithm for the MCD estimator leads to singularities. Figure 5 shows the resulting loading plots.

--- Figure 5 about here ---

Figure 5 shows that – apart from changes in the order of the factors – the classical analyses are quite similar, although we note several differences to the classical FA for the logit-transformed data (see Figure 4, top). This is not surprising, given that the raw data and the categorized data are very skewed, which leads to biased correlations. But even the robust method gives rise to substantial differences caused by severe deviations from elliptical symmetry.

Table 4 summarizes the results in a concise format. The original wording of the variables '01' to '24' is shown in the rows of the table. The five columns on the right contain the various factors (labeled A to F) of the different analyses shown in Figures 4 and 5. In order to improve readability, we do not show the loadings in numbers, but only the respective label A – F of the factor on which the items load. Since all analyses resulted in a six factor solution, we use all labels in each of the columns. Capital letters in these columns refer to absolute loadings higher than 0.5, and lower-case letters indicate absolute loadings between 0.3 and 0.5.

To facilitate comparison between the five analyses, we changed the labeling of the factors, rather than simply adopting the labels F1 to F6 from Figures 4 and 5, which are determined by the amount of variance explained (i.e. the relative importance of the factors). The labels of the classical factor analysis with logit transformation (CL) are directly taken from Figure 4 (i.e. A = F1, B = F2 …). When applying a robust factor analysis with logit transformation (RL), the relative importance of the factors changes. As can be seen from Figure 4, the factor F2 of the robust analysis contains exactly the same items as F3 in the classical analysis. We therefore decided to relabel F2 from the robust solution into 'C' in-
instead of 'B', to facilitate the comparison. In some cells in Table 4 there are two letters, which shows that the variable loads on two different factors.

--- Table 4 about here ---

Table 4 indicates that some factors are quite stable across different methods of analysis, e.g. factor 'C', while other factors such as 'B' and 'D' slightly change across the methods, and factors 'A', 'E', and 'F' are somewhat unstable. In other words, the latter ones are those factors which are influenced by the method chosen by the researcher. Moreover, variables such as '12' or '21' are poorly represented by all factor models, which highlights an additional advantage of factor analysis over principal component analysis, because the interpretation of the resulting factors will be improved by suppressing "unsuitable" variables.

The summary of the results presented in Table 4 shows the general sensitivity of factor analysis. By ignoring the required distributional assumptions, the resulting factors will have a slightly different composition as opposed to an analysis where the data have been transformed appropriately. But even after transformation, data inhomogeneities or outliers can still be present, leading to a biased classical analysis. As discussed above, this can be avoided when using robust methods.

**The Impact of Noise**

Robust methods also help to achieve better stability in the results with respect to noise in the data, which can be caused e.g. by inaccurate or wrong answers from survey respondents. In our case, the slider bars for generating the data could have been moved to an arbitrary position between 1 and 100. This will result in a 3

---

3 Since the labeling of the factors is arbitrary, this relabeling does not alter the results (i.e. the overall factor structure). However, it has to be noted, that in Table 4 the letters do not show the relative importance of the factors anymore, i.e. 'A' is simply a name for a factor and it does not indicate any ranking.
less pronounced structure of the data, and consequently leads to changes in the factors. The impact of noise in the data to factor analysis can be simulated by replacing data records with uniformly distributed data in the range 1 to 100, and subsequently comparing the resulting loadings with the loading matrix obtained for the original data. Since not only the loadings, but also the factors can change their position, a useful comparison is only possible by using a target rotation [4]. The goal of this rotation is to bring the loading pattern resulting from the modified data as much as possible in agreement with the original loadings, using an orthogonal or oblique rotation. The simulation was done for the untransformed as well as for the logit-transformed data. We used classical and robust factor analysis with Oblimin rotation, and replaced 10, 20, and 30% of the observations with uniform random noise, respectively. For each percentage, 100 simulations were done. The results can be found in Figure 6, with the vertical axis showing the absolute frequencies and the horizontal axis depicting the loading differences.

For the untransformed data (Figure 6, left), robust factor analysis shows a much higher stability than classical factor analysis if 10% or 20% of the data are noise. If 30% of the data are arbitrarily chosen, the amount of instability for robust factor analysis is comparable to that of classical factor analysis. For the logit-transformed data (Figure 6, right) the difference in the behavior of classical and robust factor analysis becomes marginal and the stability is in between both cases (classical and robust FA) for the original data. This can be explained by the fact that the logit-transformation reduces the influence of (multivariate) data outliers.

--- Figure 6 about here ---
Data Interpretation

So far, we have concentrated on analyzing the data without paying attention to the interpretation of the results. As was stated above, we decided to choose a Principal Factor Analysis, since we were interested in discovering the latent structure behind our set of variables. Usually the task of finding adequate terms for the factors is an ex-post process done by the researchers conducting the data analysis. Since our main focus lies on the comparison of methods and since we obtained significantly varying results, we decided to use qualitative studies instead to find adequate factor labels. Although triangulation of different methods has been strongly recommended for many years [23, 54], few research projects have actually made use of this approach. In our paper we adhere to the recommendations by Denzin (1978) and apply different methods to the same problem [13]. This enables us to cross-validate our findings and subsequently improve their accuracy [23].

Since the results of the quantitative factor analyses were open to multiple interpretations, we decided to pursue a qualitative approach in order to discover which factor solution resembles a human decision making process. A major advantage of the qualitative study was that the research directors were able to ensure that all of the items were fully understood by the participants. We had all of our items sorted by human subjects, i.e. it was their task to find a suitable ‘factor structure’ by using group discussions and general agreement. In order to account for a potential bias, which may be caused by their previous knowledge, we decided to use two different groups for each stage of the survey. The first group consisted of five experts with a sound knowledge of the Internet working at a large university, and the second group was made up of five graduate business students.
We performed two rounds of a sorting study. The results are shown in Figure 7. In round A, each group was asked to freely sort the items and to find a common denominator. The experts came up with seven groups ("Benefits and Incentives", "Clarity and Transparency", "Individualization", "Responsiveness", "Interactivity", "Entertainment", "No Interpretation"), whereas the students agreed on six groups ("Delivery", "Trust", "Individualization", "Contact", "Service and Entertainment", "No Interpretation"). Subsequently, all the subjects decided which categories were most suitable, i.e. the two groups had to decide on how their categories could be merged. They agreed on the six final groups "Benefits, Incentives", "Transparency, Trust", "Individualization", "Responsiveness, Contact, and Interactivity", "Service, Entertainment" and "No Interpretation". In round B, we formed two new groups, again made up of five experts working at a university and five business students, none of whom had participated in the first round. Their task was to assign the items to the categories defined in round A. Although the objective in the second round was much more clear-cut than in the first one, there was still considerable disagreement between the two groups in respect to five items (see Figure 7). Since this was a qualitative study with small sample sizes, we did not test for significant differences according to demographic and socioeconomic attributes between the groups.

--- Figure 7 about here ---

Interestingly, the agreements and disagreements of the human evaluation presented in Figure 7 strongly reflect the pattern of the loadings in the different factor analyses shown in Table 4. Those items which were assigned to the same factor in round B are printed in boldface. Subsequently, we compared the final results of round B of the qualitative sorting study with the results of the various forms of quantitative factor analysis shown in Table 4.
The first group "Benefits and Incentives" (10, 11, 24) is represented by two factors (C, D) in the quantitative analysis, with all factor analyses unambiguously yielding the same result. The problems of allocating items pertaining to "Transparency and Trust" are clearly reflected by the quantitative analyses. While all analyses assign the same factor 'A' to items 13 and 18 (although some of them are below the threshold of .5), items 17 and 19 cannot be clearly assigned. "Individualization", which is rather straightforward in the qualitative analysis, clearly shows the differences between the various factor analyses performed. The RFA with logit transformation differentiates between two pairs of items (01, 02 vs. 03, 09), while all other analyses identify a single factor. Except for the robust analysis with the original data, no quantitative analysis shows a loading of > .3 for item 12. The group labeled "Responsiveness, Contact, and Interactivity" consists of two different factors according to the quantitative analyses (04, 05 vs. 06, 07). The different views of the students and the experts whose items belong to the group "Service and Entertainment" are well reflected by the robust factor analysis with logit transformation. Not only those items that were undisputed among the subjects (15, 16, 20, 22) but also the other items (08, 17, 23) load on either factor 'A' or 'D'. Similar results are produced by the robust analysis of the original data and the classical analysis with logit transformation. These results show that the uncertainty of experts and students is reflected by inhomogeneities in the survey data leading to unstable factors.

**Discussion and Conclusion**

Using factor analysis for exploratory data analysis leaves the researcher with a multitude of options, such as various types of data transformation, the choice of the factor extraction method, factor rotation, and the number of factors to choose. Frequently, these choices are not theoretically justified and basic statistical assumptions are violated. In this paper we illustrate how the choice of the
method influences the results. Additionally, we present the concept of robust factor analysis, which makes less restrictive assumptions about data distribution than classical factor analysis and reduces the influence of outliers. We observed striking differences when comparing the biplots of the classical and the robust factor analyses and using different methods of data transformation. To triangulate our methods, we also conducted a qualitative study with a view to identifying names for individual factors. The results reveal that some groups of variables are clearly assigned to the same factors identified by all factor analysis methods presented in this paper. However, across the various factor analyses, some of the items change their contribution to the factors. This underlines the researcher’s responsibility for a solid statistical analysis. For our sample dataset, a robust factor analysis on the logit transformed data was necessary to deal with skewed and noisy raw data. The robust solution turned out to be more stable than its classical counterpart. In general, robust factor analysis is less sensitive to deviations from model assumptions, such as normality or inhomogeneities of the observations, and might therefore be a preferable solution for researchers who have to deal with data containing noise. Our results also show the limitations of exploratory factor analysis. It is impossible to easily find a single best solution and to justify it. In our paper we therefore use three different criteria to evaluate the quality of the chosen approach. Initially, the method has to be ‘correct’ in terms of basic statistical assumptions. If this criterion is fulfilled, the researcher has to make a choice between several options, which might lead to different solutions. In this paper, we show that the interpretability of the results may vary significantly between the different solutions, and that it is up the researcher to carefully select among the correct solutions. The robust solution may not necessarily be easier to interpret. When it comes to the third criterion, stability, the robust factor analysis provides superior results when the amount of noise is moderate. A final limitation pertains to the scale being used. The visual analogue
scale which we use in this paper is based on a completely different measurement approach than Likert scales [30]. Further research is needed to assess to what extent visual scales may improve existing measurement techniques.
References


Glossary

Communality: Denotes the proportion of a variable’s variance which is explained by a factor structure.

Confirmatory Factor Analysis (CFA): Used to verify the hypothesized factor structure of a set of observed variables. The researcher specifies the relationships between the variables a priori and uses CFA to test the hypotheses.

Eigenvalue: Represents the sum of squared loadings for a factor. It is frequently used to determine the number of factors.

Exploratory Factor Analysis (EFA): Used to find an underlying structure between a set of observed variables without specifying a priori relationships.

Factor Loading (see: Loading)

Heteroscedasticity: Refers to a situation when the variance of the variable differs. Various statistical tests require the equality of the variances (homoscedasticity).

Kaiser-Meyer-Olkin (see: Measure of Sampling Adequacy (MSA))

Loading: Denotes (for standardized data) the correlation between a variable and a factor.

Measure of Sampling Adequacy (MSA): A statistic which is calculated both for the entire correlation matrix and each individual variable. It is used to measure the appropriateness of the raw data to apply a factor analysis.

Minimum Covariance Determinant (MCD) estimator: The minimum covariance determinant (MCD) estimator is a highly robust estimator of multivariate location and scatter. The objective is to find those h observations whose sample covariance matrix has the lowest determinant. The value h determines the robustness, and it can be chosen between about half the observations and all observations.

Rotation: A transformation of the factor loadings in order to approximate a simpler and more meaningful structure.

Orthogonal: Group of rotation methods which imply that the extracted factors are uncorrelated.

Varimax: The variances of the squared factor loadings for each factor are considered, and the sum of these variances is maximized.

Quartimax: The sum of all factor loadings to the power of four is maximized.

Equamax: Maximizes a weighted sum of the Varimax and Quartimax criteria.

Oblique: Group of rotation methods which allow the extracted factors to be correlated.

Quartimin: Minimizes the sum of the cross products of the squared variable loadings.
**Covarimn**: Similar to Quartimin, but adjusts for the overall size of the squared loadings of each factor.

**Oblimin**: Generalizes and combines Quartimin and Covarimn rotation.

**Promax**: Tries to fit a target matrix which has a simple structure.

**McCammon**: Minimizes an entropy ratio.

**Principal Component Analysis**: Reduces the number of observed variables to a smaller number of principal components, which account for the essential amount of variance in the data.

**Principal Factor Analysis**: Similar to principal component analysis, but accounts only for the essential variance that is common to the variables.

**Scree Plot**: Shows the Eigenvalues in descending order of magnitude as a function of the Eigenvalue index. It is used to determine the number of principal components or factors.

**Skewness**: Measures the asymmetry of a distribution.

**Transformation**: Changes the distribution of the data set with the usual objective of approximating a normal distribution.

- **Logit**: Mainly used if the observed data are ratios or proportions.
- **Power**: The data are transformed using power functions.
- **Box-Cox**: One particular way of parametrizing the power transformation.
Table 1

Examples of Influencing Determinants in Factor Analysis.

<table>
<thead>
<tr>
<th>Robustness</th>
<th>Robust Factor Analysis</th>
<th>Non-robust Factor Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Transformation</td>
<td>No Transformation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Logit</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Power</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Box-Cox</td>
<td></td>
</tr>
<tr>
<td>Factor Extraction Method</td>
<td>Classical Extraction Methods</td>
<td>Robust Extraction Methods</td>
</tr>
<tr>
<td></td>
<td>Principal Component Analysis</td>
<td>Robust Principal Component Analysis</td>
</tr>
<tr>
<td></td>
<td>Principal Factor Analysis</td>
<td>Robust Principal Factor Analysis</td>
</tr>
<tr>
<td></td>
<td>Maximum Likelihood</td>
<td>Robust Maximum Likelihood</td>
</tr>
<tr>
<td>Rotation</td>
<td>Orthogonal</td>
<td>Oblique</td>
</tr>
<tr>
<td></td>
<td>Varimax</td>
<td>Oblimin</td>
</tr>
<tr>
<td></td>
<td>Quartimax</td>
<td>Quartimin</td>
</tr>
<tr>
<td></td>
<td>Equamax</td>
<td>Promax</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Covarimin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>McCammon</td>
</tr>
<tr>
<td>Number of Factors</td>
<td>A Priori Criterion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eigenvalue</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scree Plot</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Percentage of Variance</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 1. Estimated Correlation of Simulated Data with Outliers.
Table 2
Factor Analysis Mentioned in Information Systems Research*/*.  

<table>
<thead>
<tr>
<th></th>
<th>Factor Analysis</th>
<th>Exploratory Factor Analysis</th>
<th>Principal Component Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIS Quarterly</td>
<td>173 (84)</td>
<td>44 (25)</td>
<td>32 (17)</td>
</tr>
<tr>
<td>ISR</td>
<td>54 (37)</td>
<td>16 (10)</td>
<td>15 (10)</td>
</tr>
<tr>
<td>JMIS</td>
<td>159 (74)</td>
<td>32 (25)</td>
<td>36 (13)</td>
</tr>
<tr>
<td>I&amp;M</td>
<td>23 (12)</td>
<td>3 (0)</td>
<td>1 (0)</td>
</tr>
</tbody>
</table>

* Date of Analysis: April 2008

**The number of papers published since the year 2000 is shown in brackets
<table>
<thead>
<tr>
<th>Paper</th>
<th>Goal</th>
<th>Factor Extraction Model</th>
<th>Number of Factors</th>
<th>Rotation</th>
<th>Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang and Strong (1996)</td>
<td>Item reduction</td>
<td>n.s.</td>
<td>20</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Torkzadeh and Dillon (2002)</td>
<td>Instrument development (validity)</td>
<td>PCA</td>
<td>Phase 1: 4 (Eigenvalue &gt; 1) Phase 2: 5 (Eigenvalue &gt; 1, scree plot)</td>
<td>Varimax, oblique rotation</td>
<td>Phase 1: 72.9%, 68.1%  Phase 2: 77.3%, 69.2%</td>
</tr>
<tr>
<td>Koh et al. (2004)</td>
<td>Assessing construct validity (convergent and discriminant validity)</td>
<td>PCA</td>
<td>7 (Eigenvalue &gt; 1)</td>
<td>Varimax rotation, oblique rotation</td>
<td>77.6%, 76.9%</td>
</tr>
</tbody>
</table>

n.s.: not specified
### In an online relationship with a company it is important to me that

<table>
<thead>
<tr>
<th></th>
<th>Absolutely disagree</th>
<th>neutral</th>
<th>Absolutely agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am personally welcomed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I receive congratulations on important dates (e.g., birthday)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I regularly receive individualized newspapers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can express my opinions in forums</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I have the opportunity to give feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can find a contact person at any time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I get answers for my requests quickly</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can check my delivery status at any time</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 2. Screenshot of the questionnaire (translated)**
Fig. 3. Biplots of Classical and Robust Factor Analysis for the First Two Factors.
**Fig. 4.** Classical Factor Analysis versus Robust Factor Analysis: Factor Loadings.
Fig. 5. Classical and Robust Factor Analyses on Untransformed Data.
### Table 4

Comparison of the Factor Analyses.

<table>
<thead>
<tr>
<th>Item</th>
<th>CL</th>
<th>RL</th>
<th>CO</th>
<th>C1-5</th>
<th>RO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In an online relationship with a company it is important to me that</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>01 ... I am personally welcomed</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>02 ... I receive congratulations on important dates (e.g. birthday)</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>03 ... I regularly receive individualized newsletters</td>
<td>B</td>
<td>E</td>
<td>B</td>
<td>B</td>
<td>Be</td>
</tr>
<tr>
<td>04 ... I can express my opinions in forums</td>
<td>E</td>
<td>e</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>05 ... I have the opportunity to give feedback</td>
<td>E</td>
<td>Ef</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>06 ... I can find a contact person at any time</td>
<td>ef</td>
<td>F</td>
<td>ef</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>07 ... I get answers for my requests quickly</td>
<td>af</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>08 ... I can check my delivery status at any time</td>
<td>A</td>
<td>A</td>
<td>f</td>
<td>f</td>
<td>a</td>
</tr>
<tr>
<td>09 ... I receive individualized offers</td>
<td>B</td>
<td>E</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>10 ... I get presents or discounts</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>11 ... I get aggregated rebates</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>12 ... I can customize the website according to my needs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td>13 ... the general terms and conditions are clearly defined</td>
<td>A</td>
<td>A</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>14 ... the website is well structured</td>
<td>A</td>
<td>A</td>
<td>F</td>
<td>F</td>
<td>Af</td>
</tr>
<tr>
<td>15 ... I find the website entertaining</td>
<td>bd</td>
<td>ad</td>
<td>d</td>
<td>d</td>
<td>bd</td>
</tr>
<tr>
<td>16 ... the website offers online games</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>17 ... I can find information about a company’s business policy</td>
<td>Af*</td>
<td>ae</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>18 ... data can be transmitted over an encrypted connection</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>19 ... I can view my personal data at any time</td>
<td>A</td>
<td>A</td>
<td>af</td>
<td>Af</td>
<td>A</td>
</tr>
<tr>
<td>20 ... I receive the ordered products and services on time</td>
<td>A</td>
<td>af</td>
<td>F</td>
<td>F</td>
<td>af</td>
</tr>
<tr>
<td>21 ... I like the website</td>
<td>a</td>
<td>a</td>
<td>d</td>
<td>d</td>
<td>ad</td>
</tr>
<tr>
<td>22 ... I can download software</td>
<td>D</td>
<td>aD</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>23 ... I can participate in sweepstakes</td>
<td>D</td>
<td>D</td>
<td>d</td>
<td>d</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>24</td>
<td>... I can send SMS free of charge</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>RL: Robust Factor Analysis with Logit Transformation</td>
<td>CO: Classical Factor Analysis with Original Data</td>
<td>Lower-Case Letters: Factor Loadings between .3 and .5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1-5: Classical Factor Analysis with Five Categories</td>
<td>RO: Robust Factor Analysis with Original Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Negative loading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 6. Stability of Classical and Robust Factor Analysis for Untransformed and Logit-Transformed Data.
Fig. 7. Results of Qualitative Sorting Study.