Tjalling C. Koopmans Research Institute

How to reach the authors

Please direct all correspondence to the first author.

Martijn A. Boermans *^
*HU University of Applied Sciences,
Faculty of Economics and Management
P.O. Box 13102
3507 LC Utrecht
E-mail: M.A.Boermans@hu.nl
Mark A.C. Kattenberg
^Utrecht University
Utrecht School of Economics
Kriekenpitplein 21-22
3584 EC Utrecht
The Netherlands.
E-mail: M.A.C.Kattenberg@uu.nl

Estimating reliability coefficients with heterogeneous item weightings using Stata: A factor based approach

Martijn A. Boermans[a,b]
Mark A.C. Kattenberg[b]

[a]HU University of Applied Sciences,
Faculty of Economics and Management

[b]Utrecht School of Economics
Utrecht University

November 2011

Abstract
We show how to estimate a Cronbach's alpha reliability coefficient
in Stata after running a principal component or factor analysis. Alpha evaluates
to what extent items measure the same underlying content when the items are
combined into a scale or used for latent variable. Stata allows for testing the
reliability coefficient (alpha) of a scale only when all items receive homogenous
weights. We present a user-written program that computes reliability coefficients
when implementation of principal component or factor analysis shows heteroge-
neous item loadings. We use data on management practices from Bloom and Van
Reenen (2010) to explain how to implement and interpret the adjusted internal
consistency measure using afa.

Keywords: afa, factor analysis, Cronbach's alpha, reliability, heterogeneous scale
construction,latent variable

JEL classification: C49, C87, M21

1. Introduction

Survey methods are becoming increasingly popular in economic research. Many questionnaires are designed in such a way that a latent, conceptual variable (often referred to as a scale) is measured using a wide range of hand-picked question items. Scale construction is often based on simple addition of items (homogeneous weighting). To measure the internal consistency of a constructed scale, typically the Cronbach's alpha is calculated, which indicates how reliable the combination of items are as a proxy for the conceptual variable under homogeneous weighting.

Principal component or factor analysis (hereafter: *FA*) are commonly employed as data reduction tools to make scales which proxy latent variables without the assumption that items have homogeneous weights. In other words, *FA* allows some items to be more strongly associated with the core concept than others; items endogenously receive different weights. As Raykov (2001) shows, in many studies Cronbach's alpha is misreported when latent variable analysis reveals heterogeneous item loadings. In addition to the weighting issue Cronbach's alpha is flawed when the combination of items is not unidimensional. In principal *FA* can be used to test whether the underlying structure consists of one or several latent dimensions, yet this makes calculation of the internal consistency of conceptual variables more complex, especially when the conceptual variable is multidimensional (Brown, 2006). Therefore this paper provides a friendly user-written program `afa` to reduce this complexity.

As Boermans, Roelfsema, and Zhang (2011) argue *FA* are both an eclectic and neutral approach to making scales comprising of various items. Researchers cannot control the weighting of variables nor the dimensions of the conceptual variable on an a priori basis. In contrast to simple addition of items (combinatorics) where researchers may bundle certain items together homogenously, *FA* constructs *J* scales equal to the number of loaded items based on the underlying correlation of variables and proper (heterogeneous) weighting of items. Hence, key difference between scale construction where the Cronbach's alpha is calculated and *FA* is that in the former all items require a similar weighting into a single final scale, whereas in *FA* each variable (item) can have different weightings for each underlying factor *J* (scale).

To motivate our paper, as of present Stata allows for testing the reliability coefficient (alpha) only when all items are added under a homogenous weighting scheme using the command `alpha`. It is not possible with Stata to estimate the internal consistency or reliability coefficient (alpha) of a constructed scale with heterogeneous loadings, nor of a factor or of multidimensional concepts and factors. *afa* assigns heterogeneous weights to a possibly multidimensional scale construct.

The paper commences as follows. In Section 2 we describe how the Cronbach's alpha is commonly employed as measure of scale reliability. We also briefly explain the core concept of *FA* which endogenously derive the weightings of items, and also test for multidimensionality. Section 3 provides the syntax of the `afa` command. An example of the implementation of the program is presented in section 4. In order to make `afa` user friendly, the output is similar to the standard outputs of the `alpha` and `factor` commands. We use survey data from Bloom and Van Reenen (2010) who study management practices based on a set of question items. Section 5 concludes.

2. Theoretical Framework

This part briefly stipulates Cronbach's alpha. Next, an outline is given of principal component and factor analysis, or *FA*. We refer to both principal component and factor analysis with the acronym *FA* since these methods are closely related (eg Velicer and Jackson (1990)) and are implemented in Stata with the same command `factor`. Many different conceptualizations of *FA* have been established for various purposes including the confirmatory factor analysis (`confa`) in Stata (see Kolenikov (2009)).

2.1. Cronbach's alpha

In most empirical research Cronbach's alpha is estimated to obtain a measure of reliability of a set of question items (Henson, 2001).[1] Cronbach (1951) refers to the internal consistency as the proportion of the test variance that can be attributed to a group of items, which measures the reliability coefficient alpha:

$$\alpha = \frac{I}{I-1} \left(1 - \frac{\sum_{i=1}^{I} \sigma_i^2}{\sigma_X^2} \right) \tag{1}$$

where $n = 1, 2 \ldots I$ are the number of items in the scale, σ_i^2 is the variance of item i, and σ_X^2 is the variance of the observed total test scores. A high value of the Cronbach's alpha implies a higher internal consistency of the construct X.[2] Yet it does not imply unidimensionality of X, however, the interpretation of X depends on the dimensionality and the construct validity (Cortina (1993); Schmitt (1996)). Another issue related to Cronbach's alpha is that researchers can freely decide on the number and subsets of items to include in scale construction. One can imagine that this may lead to a bias where certain dimensions of the underlying construct can be over- or underrepresented, because of the selection of question items by subjective criteria (e.g. omitted variable bias or measurement error).

2.2. Principal Component and Factor Analysis

FA is used for data reduction as it concentrates variation of a large number of observed variables into far fewer aggregated dimensions (Jolliffe (2002)). Similarly it can be used for scale construction of latent variables as it endogenously classifies observed variables (items) into fewer latent dimensions (constructs). In matrix notation *FA* expresses the variation in the vector of (demeaned) items as a linear combination of latent dimensions, the so-called factors:

$$x - \mu = \Lambda F + \epsilon \tag{2}$$

---

[1]The widespread usage of the Cronbach's alpha has received some criticism since the Cronbach's alpha provides a lower bound score of reliability. Alternative reliability test score are available such as the older Guttman's $\lambda 2$ and McDonald's ω_h, where Cronbach's alpha is a special case of those two estimates (Zinbarg, Yovel, and Li (2005), Sijtmsa (2009)). Because its sensitivity to multidimensionality, the interpretation of the Cronbach's alpha as internal consistency measure should be done with caution. Nonetheless, Cronbach's alpha remains the most commonly used reliability coefficient and internal consistency measure.

[2]Traditionally, reliability coefficients below 0.5 are seen as unreliable, measures between 0.5 and 0.7 are modest, and levels of the Cronbach's alpha above 0.7 indicate acceptable levels.

where x is a $n \times i$ matrix containing n observations for i items and μ is a $n \times i$ matrix containing the item means ($\mu = n^{-1} \sum_{i=1}^{n} x_i$). F is a $j \times i$ matrix of latent factors [3], Λ is a $n \times j$ matrix of factor loadings and ϵ is a $n \times i$ matrix of random errors. It follows from equation 2 that the observed items are modeled as linear combinations of the factors with additional error terms. The estimated factors are orthogonal.

3. The afa command

`afa` makes use of the build-in Stata commands `alpha` and `factor` to test whether a set of items forms a multidimensional construct. When the construct is multidimensional, `afa` computes Cronbach's alpha for each dimension of the construct using the factor loadings as (heterogeneous) weights. When the construct is unidimensional, `afa` may differ from Cronbach's alpha as it weights the items heterogeneously based on their factor loadings.

`afa` *varlist* performs a factor analysis (or principal component analysis) on the items specified in *varlist* to heterogeneously weight the items into a latent scale.[4]

When J factors are detected, $i \times j$ rescaled items $\tilde{i}_j$ are computed by multiplying item i with its factor loading λ_{ij}.

$$\tilde{i}_j = i\lambda_{ij}. \tag{3}$$

Thus `afa` heterogeneously weights the items in the scale according to equation 3. Next, Cronbach's alpha α_j is computed for each factor j using the rescaled items $\tilde{i}_j$ instead of i in equation 1.

3.1. Syntax

```
afa varlist [if] [in] [, method(mname) retainev(number)
foptions(foptions) varimax promax(integer) post(post) blank(number) i]
```

method(*mname*)	Control the FA method.
	mname is:
	pf for principal factors (default).
	pcf for principal-component factors.
	ipf for iterated principal factors.
	ml for maximum likelihood factors.
	pca for principal component analysis.
retainev(#)	Control the maximum number of factors / components to be shown in the output window.
foptions(*fname*)	Set FA options.
	fname is:

---

[3] There are several information criteria available to determine the number of endogeneously derived factors, or the dimensionality of the underlying structure, (see e.g. Alessi, Barigozzi, and Capasso (2010)) where a standard rule is to include only factors with an eigen value above unity.

[4] There are several criteria to select the number of derived factors, including the Kaiser little jiffy based on eigenvalues, and the Cattell scree-plot.

	factors(#) sets the number of factors or components to be estimated.
	mineigen(#) defines the minimum Eigenvalue required for an estimated factor/component to be shown in the output window.
`rotate(rname)`	Control the rotation method after estimating the factors/ components. The default value is not to rotate. When rotation is required, *rname* is: *varimax* uses varimax roatation. *promax* uses oblique promax rotation.
`blank(real)`	Factor loadings with a value smaller than real are estimated, but not reported in the factor loadings matrix. *real* should be a real number.
`post(pname)`	Allows to run factor postestimation commands from within `afa`. Alternatively, they can be called after the `afa` command has been executed using saved results of the `factor` command. *pname* can be: *estat anti* provides the anti-image correlation and covariance matrices. *estat factors* provides AIC and BIC model-selection criteria for different numbers of factors. *estat kmo* provides Kaiser-Meyer-Olkin measure of sampling adequacy. *estat residuals* provides matrix of correlation residuals. *estat smc* provides squared multiple correlations between each variable and the other variables. *estat structure* provides correlations between variables and common factors *estat summarize* provides estimation sample summary *loadingplot* provides a plot of the factor loadings *scoreplot* provides a plot of score variables *screeplot* provides a plot of the eigenvalues multiple postestimation commands can be specified, they should be separated with a "*".
`i`	Display item-test and item-rest correlations for each dimension.

3.2. Description

The `afa` performs *FA* and estimates the Cronbach's alphas α_j using rescaled items. As such, `afa` is an extension of the `factor` and `alpha` commands. When `afa` is called an unrotated factor analysis is performed, using the *if* and *in* conditions. The researcher controls the factor method by specifying the option `method( )`.[5]

   After *FA* `afa` lists the results by showing the matrix of factor loadings and the uniqueness of the items. Presentation of the matrix of factor loadings can be controlled using the `blank()`

---

[5]By default the principal factor method is used. *pf* indicates factor analysis using the principal factor method, the options *pcf*, *ipf* and *ml* specify the factor methods principal-component factors, iterated principal factors and maximum-likelihood factors. To perform a principal component analysis, the factor method *pca* should be specified.

option. When `blanks(x)` is specified all factor loading smaller than *x* will be showed as blanks. However, they will still be used in the further analysis. Together with Cronbach's alpha α_j (Scale Reliability Coefficient), the average interitem correlation, the number of items in the scale and the eigenvalue of factor *j* are listed. When a lists of items is multidimensional, items will load strongly on different factors. By rescaling the items with their factor loadings, `afa` controls for multidimensionality and computes Cronbach's alphas α_j for each latent dimension.

4. Example

4.1. Data

In this section we use data on management practices from 694 American manufacturing firms to illustrate how to derive the proper Cronbach's alpha(s) after *FA*. In total we have complete data for 628 firms. The information is obtained from Bloom and Van Reenen (2007, 2010), who define and score 18 basic management question items on a Likert-scale from 1 (worst practice) to 5 (best practice). They argue that a combination of all items reflects 'management practices' by taking the average of the 18 items (unidimensional, homogeneous weighting).

The motivation to analyze this dataset is threefold. First, Bloom and Van Reenen themselves are ambiguous about the usage of a single latent variable for management practices in their theoretical framework. Thus we ask: can management practices be measured as a single construct, and is it internally consistent? In related work on firms from Germany, France, UK, and the US, Bloom and Van Reenen (2007) argue that management practices can be grouped by four areas: *operations*, *targets*, *monitoring* and *incentives*. In contrast, in the 2010 study based on 12 countries they disentangle three dimensions of management practices: *lean manufacturing*, *performance* and *talent*. Moreover, Bloom and Van Reenen (2010) also highlight a three dimensional classification of management practices, namely *monitoring*, *targets* and *incentives*. In each study they diverge from such classifications and empirically estimate a single scale (again: unidimensional, homogeneous weighting). To this end we are interested in showing which classification of the items fits the data best. We ask: what dimensional structure of the management practices items is endogenously derived? The application of `afa` can provide more insights into both the multidimensionality of the management items without any theoretical priors, and the reliability (alpha) of the management practice construct (or factors). Second, the set-up of the survey is a 'best practice'-case of survey research.[6] It is completed in a 'double blind' fashion since the managers are unaware they are being scored and interviewers do not know the firms they are scoring. Given the high efforts undertaken to complete the survey (also across countries) it is of interest to supply reliability tests of the derived management practices factors. Finally, our focus lies on the US because many firm-level studies on management and organization have been undertaken in the US. US management practices are also top-ranked by Bloom and Van Reenen (2010).

---

[6]Bloom and Van Reenen (2007) argue that the application of such an evaluation tool provides some quantification of the firm's management practices. Each interview took on average fifty minutes and was held in 2004 from the Centre for Economic Performance at the London School of Economics. The response rate was 54 percent. In order to limit response biases and obtain accurate data they installed several procedures. The managers that were interviewed by telephone did not know they were scored. They used open question rather than closed question format (e.g. 'can you tell me how you promote your employees?'). The interviewers did not know anything about the company's financial status. In a later stage they show that performance and management practices are positively related, which validates the management practice measures.

4.2. Implementing Cronbach's alpha using principal component and factor analysis

In this part it is explained how to estimate a reliability coefficient (alpha) for a constructed scale (with homogeneous weights) based on the command `alpha`. Next, *FA* is applied using `afa` to inform about the dimensionality of the data and to obtain our factors. It is shown how to determine the corrected Cronbach's alpha after *FA* based on the single command `afa`.

First, in figure 1 the Cronbach's alpha of the overall management practices scale is 0.898 when the scores of all 18-items are combined in a scale under homogeneous weighting. This suggests the internal reliability of the scale is very high, yet it does not necessarily imply unidimenisonality of the management practices construct.[7]

Figure 1: Stata output of `alpha` for 18 management items

```
. alpha lean* perf* talent*

Test scale = mean(unstandardized items)

Average interitem covariance:      .3489744
Number of items in the scale:            18
Scale reliability coefficient:       0.8978
```

Next, using `afa` it is shown that a multidimensional structure under heterogeneous weightings of the items fits the data better. After the implementation of `afa` two tables are obtained. First, following the `factor` command's output, unrotated factor loadings of each item per factor are given, together with the uniqueness of each item in the general structure. Following the `alpha` command's output, the second table below the factor loadings presents the average interitem covariance, the number of items used and the eigenvalues for each respective factor. At the bottom the scale reliability coefficients are presented, that refer to the corrected Cronbach's alpha for each factor. As such, using `afa` one can explore the item loadings per factor and look at the eigenvalue of each factor to determine the number of dimensions. Simultaneously, the researcher takes each factor's construct validity into account based on the reliability scores. Hence, `afa` provides endogenously derived factors that also receive scale reliability coefficient based on the corrected Cronbach's alpha.[8]

Using `afa` we perform a *FA* based on principal component factors (*method(pcf)*). We focus on the factors with eigenvalues above unity (using *retainev(1)*). The factors are rotated using Varimax-rotation to detect the underlying pattern in each factor. To get a better grip of each factor's content, only item loadings greater than 0.5 are shown using the *blank (0.5)* option.

The results based on `afa` in figure 2 reveal a three-dimensional structure.[9] Eight of the ten performance items result in Factor 1. All talent items group together in Factor 2. Two lean items

---

[7]To this end we provide the Cronbach's alphas for the dimensions outlined by Bloom and Van Reenen (2007), p.1393 and Bloom and Van Reenen (2010), p.210 under homogenous weightings. The reliability coefficients for the *operations, targets, monitoring, and incentives* dimensions are 0.816, 0.816, 0.788 and 0.743 respectively. The reliability coefficients for *lean manufacturing, performance and talent* dimensions are 0.853, 0.794 and 0.802 respectively. In the dataset the reliability coefficients for the *monitoring, targets and incentives* dimensions are 0.848, 0.886 and 0.791 respectively. Overall, each distinguished latent class has a high internal consistency as measured by the Cronbach's alpha reliabiliy coefficient.

[8]Notice that using `afa` the corrected alpha after *FA* will depend on the rotation method applied.

[9]Using `post(estat kmo)`, the Kaiser-Meyer-Olkin measure of sample adequacy is an index between zero and unity that can be interpreted as a test for the appropriateness of *FA*. It gives the proportion of variance among the items that can be attributed to the common variance (i.e. that is indicative of underlying latent factor). The overall value of 0.915 suggest a 'marvelous' adequacy (where values below .6 suggest a poor adequacy).

combined with one performance item about process documentation represent Factor 3. These findings suggests that 'management practices' is not unidimensional, but has a three-dimensional structure reflecting *performance*, *talent* and *lean production*. Under heterogeneous weighting the respective reliability coefficients for each factor are 0.882, 0.819 and 0.797. These scores are very high, especially when one considers that the factors are orthogonal (a convenient proporty that a priori constructed scales do not posses). The correlations between the constructed three dimensions are all below 0.2.

Figure 2: Stata output of `afa` command for 18 management items

```
. afa lean* perf* talent*, method(pcf) retainev(1) blank(.5) rotate(varimax)
Factor Analysis using principal-component factors with 3 factors


varimax rotated factor loadings (pattern matrix) and unique variances
```

	Factor1	Factor2	Factor3	Uniqueness
lean1			0.887	0.173
lean2			0.887	0.179
perf1			0.585	0.420
perf2	0.666			0.424
perf3	0.713			0.409
perf4	0.719			0.383
perf5	0.586			0.523
perf6	0.735			0.426
perf7	0.724			0.378
perf8	0.586			0.561
perf9	0.596			0.579
perf10				0.700
talent1		0.676		0.430
talent2		0.658		0.512
talent3		0.570		0.659
talent4		0.764		0.389
talent5		0.674		0.451
talent6		0.647		0.548

```
Cronbach's alpha under factor analysis
Performing Cronbach's alpha on 3 factors. Please wait...
```

	Factor1	Factor2	Factor3
Average interitem covariance	0.073	0.035	0.032
Number of items in the scale	18.000	18.000	18.000
Eigenvalue factor	6.667	1.907	1.281
Scale reliability coefficient	0.882	0.819	0.797

5. Conclusions

Cronbach's alpha is a commonly employed reliability test measure that is used to evaluate constructed scale with homogeneous item loadings. Previous studies that have applied principal component or factor analysis have often misreported the internal consistency of the derived factors, because they do not take into account the heterogeneous weighting. The code `afa` can be implemented to obtain the correct Cronbach's alpha when factors are estimated. Hence, this study shows how researcher can now estimate a Cronbach's alpha after running a principal component

or factor analysis. Using data on management practices we supply such reliability coefficient for management practices as a single construct as well as for endogeneously derived latent factors using `afa`. Overall we obtain high internal consistency scores for each suggested management practices scale. Results by Bloom and Van Reenen (2010) are reconfirmed as three orthorgonal dimensions or factors of management practices related to *performance*, *talent* and *lean production*.

6. Literature

ALESSI, BARIGOZZI, AND CAPASSO (2010): "Improved Penalization for Determining the Number of Factors in Approximate Factor Models," *Statistics and Probability Letters*, 80(23-24), 1806–1813.

BLOOM, N., AND J. VAN REENEN (2007): "Measuring and Explaining Management Practices Across Firms and Countries," *Quarterly Journal of Economics*, 122(4), 1351–1408.

BLOOM, N., AND J. VAN REENEN (2010): "Why Do Management Practices Differ across Firms and Countries?," *Journal of Economic Perspectives*, 24(1), 203–224.

BOERMANS, M., H. ROELFSEMA, AND Y. ZHANG (2011): "Regional Determinants of FDI in China: A Factor-Based Approach," *Journal of Chinese Economic and Business Studies*, 9(1), 23–42.

BROWN, T. (2006): *Confirmatory Factor Analysis for Applied Research*. The Guilford Press, New York.

CORTINA, J. M. (1993): "What is Coefficient Alpha? An Examination of Theory and Applications," *Journal of Applied Psychology*, 78(1), 98–104.

CRONBACH, L. J. (1951): "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, 16, 297–334.

HENSON, R. (2001): "Understanding Internal Consistency Reliability Estimates: A Conceptual Primer on Coefficient Alpha," *Measurement and Evaluation in Counseling and Development*, 34(1), 177–189.

JOLLIFFE, I. (2002): *Principal Component Analysis* 2nd ed. New York, NY: Springer.

KOLENIKOV, S. (2009): "Confirmatory Factor Analysis Using Confa," *The Stata Journal*, 9(3), 329–373.

RAYKOV, T. (2001): "Estimation of Congeneric Scale Reliability Using Covariance Structure Analysis with Nonlinear Constraints," *British Journal of Mathematical and Statistical Psychology*, 54, 315–323.

SCHMITT, N. (1996): "Uses and Abuses of Coefficient Alpha," *Psychological Assessment*, 8(4), 350–353.

SIJTMSA, K. (2009): "On the Use, the Misuse, and the very limited usefulness of Cronbach's alpha," *Psychometrika*, 74(1), 107–120.

VELICER, W. F., AND D. JACKSON (1990): "Component Analysis Versus Common Factor Analysis: Some Issues in Selecting an Appropriate Procedure," *Multivariate Behavioral Research*, 25, 1–28.

ZINBARG, R., R. W., I. YOVEL, AND W. LI (2005): "Cronbach's Alpha, Revelle's Beta, and McDonald's Omega: Their Relations with Each Other and Two Alternative Conceptualizations of Reliability," *Psychometrika*, 70, 123–133.