Factor analyses for non-normal variables by fitting characteristic functions

Kees van Montfort
Free University Amsterdam
Department of Econometrics
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands
email: kvmontfort@feweb.vu.nl

and

Jan de Leeuw
University of California at Los Angeles
Department of Statistics
Los Angeles
USA
Email: deleeuw@stat.ucla.edu

Abstract:

This paper deals with a factor analysis model. Usually one assumes that the model has normally distributed variables. If these variables are not normally distributed, the well-known estimators using only variances and covariances of the observed variables may lead to biased estimates of the factor loadings. This paper uses characteristic functions to derive an approximately consistent BGLS-estimator of the factor loadings in a model with both normal and non-normal variables.
An example with generated data and an empirical example will evaluate the estimator. Furthermore, the estimator will be compared to well-known estimators using only second order moments, consistent estimators using third order moments and estimators using transformations to get normal variables.
It turns out that our method gives good estimates of the unknown model parameters for sample sizes larger than 50. Our estimation method also avoids the disadvantage of the above mentioned estimation methods.

Key words: factor analysis, non-normal variables, characteristic functions, BGLS-estimates
JEL: C51, C13, C14

Introduction

We consider the factor analysis model

$$\underline{y} = A\underline{f} + \underline{\varepsilon} \tag{1}$$

in which $\underline{y}$ (the observed variables), $\underline{f}$ (the factors) and $\underline{\varepsilon}$ (the residuals) are vectors of centered random variables and A is a matrix of constants (the factor loadings) which have to be estimated. $\underline{y} = (\underline{y}_1, \dots, \underline{y}_n)^T$ and $\underline{\varepsilon} = (\underline{\varepsilon}_1, \dots, \underline{\varepsilon}_n)^T$ are (n x 1)-column vectors, $\underline{f} = (\underline{f}_1, \dots, \underline{f}_p)^T$ is a (p x 1)-column vector and A is a (n x p)-matrix. The residuals $\underline{\varepsilon}_1, \dots, \underline{\varepsilon}_n$ are independent of each other, and independent of the factors $\underline{f}_1, \dots, \underline{f}_p$.
Usually in factor analysis models one assumes that $\underline{y}$, $\underline{f}$ and $\underline{\varepsilon}$ are multivariate normally distributed and have expectation zero. In that case we have for the first and the second order moments of the model variables

$$E\,\underline{y} = 0, \quad E\,\underline{f} = 0, \quad E\,\underline{\varepsilon} = 0 \tag{2a}$$

$$C = A\,W\,A^T + U \tag{2b}$$

in which C, W and U are the covariance matrices of $\underline{y}$, $\underline{f}$ and $\underline{\varepsilon}$. If the factors are independent of each other, then the matrix W is equal to the unity matrix I_p. The matrix U is diagonal and uniquely determined by (2b) (see Anderson and Rubin (1956)). Unless restrictions are imposed on the parameters in A and W, if there is one set of parameters that satisfies equation (2b), there will be an infinite number of such sets (see Lewis-Beck (1994), and Krijnen, Dijkstra and Gill (1998)). So, the matrices A and W are not identifiable.

Usually the matrix A will be estimated by fitting the second order moments of expression (2b). Doing this, one implicitly assumes that $\underline{y}$ is normally distributed. If $\underline{y}$ is not normally distributed the solution space for A is smaller and the estimates may be biased. For non-normal data, Mooijaart (1985) even shows that under some additional conditions the solution space for the matrix A shrinks to a single point.
For estimating A in the correct solution space, among others, we can use Asymptotic Distribution Free (ADF) methods, which fit higher order moments or use higher order moments in the weight matrix (see Bentler (1983), Browne (1984), Mooijaart (1985), and Yuan, Marshall

and Bentler (2002)). The disadvantage of this approach is the need of a very large sample size. Furthermore, the necessary conditions are not always fulfilled.
We also can transform non-normally distributed variables to normally distributed variables (see Muthen (1984)). This approach gives factor loadings for the transformed normally distributed variables instead of factor loadings for the observed non-normally distributed variables, which may render the interpretation of the factor loadings a very difficult task. Moreover, it is not always possible to transform an arbitrary distribution function to a normal distribution function.
Ichikawa and Konishi (1995) investigated the performance of the bootstrap method in normal maximum likelihood factor analysis when the normal distribution assumption is not satisfied. Statistical inference based on resampling can still be valid for nonnormal data. However, as concluded by the authors, because of ill conditions and improper solutions, bootstrap inference should be used with caution.

In fact, the approaches of the above two paragraphs do not decrease the solution space for the factor loading matrix A and still allow biased estimators. Another possibility to deal with the estimation of nonnormal factor analysis is the fitting of the distribution functions or the characteristic functions. Characteristic functions are more appropriate than distribution functions because they can be estimated by smaller sample sizes (see for instance, De Leeuw and Van Rijckevorsel (1981). Furthermore every random variable has a corresponding characteristic function, even if the higher order moments do not exist. For instance, the Cauchy distribution has no first, second and higher order moments, while the characteristic function exists.
The characteristic function of a variable is a transformation of the corresponding density function. A characteristic function uniquely determines the distribution of a random variable. If we use only means and covariances of variables we only deal with a limited amount of information about the characteristic functions around the point $t = 0$. The mean of a random variable is the first order derivative of the corresponding characteristic function in the point $t = 0$, multiplied by i. The second order moment of a random variable is the second order derivative of the characteristic function in the point $t = 0$, multiplied by -1.

In this paper we will deal with the estimation of a factor analysis model for observed variables which can be either normal or non-normal distributed. Therefore we use characteristic functions. In section one we discuss the derivation of an approximately Best Generalized Least Squares (BGLS) estimator. Furthermore in section 2 we deal with the approximation of characteristic functions by B-splines, which we need for our estimation method. In section 3 we discuss some asymptotic properties of our estimator. Finally in sections 4 and 5 we discuss an example with generated data and an empirical example. Our estimator will be compared to other well-known estimation methods.

1. Estimation algorithm

In this section we derive estimators of the factor loading matrix with the help of the relations for the first and second order moments and additional characteristic function relations. Relation (2) is valid for every distribution of the observed variables. The following additional relation between the logarithms of the characteristic functions of the random model variables also holds:

$$\ln c(t_1, \ldots, t_n) = \ln g\Big(\sum_{i=1}^{n} A_{i1} t_i, \ldots\ldots, \sum_{i=1}^{n} A_{ip} t_i\Big) + \sum_{i=1}^{n} \ln h_i(t_i), \qquad (3)$$

where $c(t_1, \ldots, t_n)$ is the characteristic function of vector $\underline{y}$ in the point $(t_1, \ldots, t_n)$, $g(t_1, \ldots, t_p)$ the characteristic function of vector $\underline{f}$ in the point $(t_1, \ldots, t_p)$, and $h_i(t)$ the characteristic functions of ε_i in the point t. If the factor loading matrix A is identified by the factor model (1), than this matrix is also identified by relation (3).

In orthogonal factor analysis, where the factors are independent of each other, the expression for the logarithm of the characteristic function of vector $\underline{f}$ reduces to the sum of the logarithms of the characteristic functions of the separate factors f_1 up to f_p:

$$\ln g\Big(\sum_{i=1}^{n} A_{i1} t_i, \ldots\ldots, \sum_{i=1}^{n} A_{ip} t_i\Big) = \sum_{j=1}^{p} \ln g_j\Big(\sum_{i=1}^{n} A_{ij} t_i\Big)$$

To obtain simpler expressions we will work with the above decomposition of the logarithm of the characteristic function of $\underline{f}$. Nevertheless, the results can be generalised to the more general case in which the factors are not necessarily independent.

With the help of relation (3) we can define the following Best General Least Squares (BGLS) function Ψ, which fits the characteristic function of the vector $\underline{y}$:

$$\boldsymbol{\Psi = Z^T V^{-1} Z} \qquad (4)$$

, with **Z** a column vector

, $\mathbf{Z_i}$ equal to the logarithm of the empirical characteristic function corresponding to the left hand side of (3) minus the right hand side of (3), for an arbitrary fixed value of $\tau_i = (t_1^{(i)}, \ldots, t_n^{(i)})$

and **V** the asymptotic dispersion matrix of the vector **Z**.

The above function (4) contains the unknown factor loadings and the logarithms of the unknown characteristic functions of the common factors and residuals. The form of these characteristic functions is restricted by the first and second order moment equations in (2).

We will approximate the logarithms of the characteristic functions of the common factors and the residuals in expression (4) by B-splines. Spline functions have found great use as approximating functions. Ordinary polynomials are inadequate in many situations. Wold (1974) mentioned that in many situations the behaviour of a function in one region might be totally uncorrelated to the behaviour in another region. Polynomials along with most other mathematical functions have the property, that their behaviour in a small region determines their behaviour everywhere. Splines do not suffer from this handicap since they are defined piecewise. Furthermore De Leeuw and Van Rijckevorsel (1981) note that the use of B-splines takes away the disadvantage of the crudeness or discontinuity of the step functions at the breakpoints, while maintaining the advantage of not approximating strictly linear functions only.

The logarithms of the characteristic functions of the factors and the residuals are multiple-valued functions. The single-valued interpretation of them is true for the real part but for the imaginary part we have equality modulo 2π. Thus for instance we have

$\text{Im} \ln g_i(t) = \text{Im} \ln g_i(t) + 2\pi K$ with K an arbitrary integer. This is a bit inconvenient to work with numerically. Therefore we only work with the real parts of the logarithms of the characteristic functions of the factors and the residuals. For instance,

$$\text{Re} [\ln g_i(t)] = \ln \{ (\text{Re}\, g_i(t))^2 + (\text{Im}\, g_i(t))^2 \}^{1/2} \quad (i = 1,\ldots,p).$$

Furthermore we mention that for all well-known distribution functions the real part of the logarithm of the corresponding characteristic functions is symmetric around the point zero, strictly increasing for $t < 0$ and strictly decreasing for $t > 0$. This paper only deals with the part $t > 0$ of Re [ln $g_i(.)$] ($i = 1,\dots p$), which has a simple monotone decreasing form.

In this paper we will obtain BGLS-estimates of the elements of the factor loading matrix A by minimizing the BGLS-function (4) under the restrictions of relation (2). We will minimize function (4) by minimizing it over the unknown spline parameters and the unknown factor loadings alternatingly. At one time we fix the set of the spline coefficients and minimize over the model parameters, the other time we fix the model parameters and minimize over the spline coefficients with the restrictions of relation (2). We continue to do so until no further improvement is made. This principle has worked very well in many related data analysis contexts (see for instance, Young (1981), Wold (1974, 1982) and Winsberg and Ramsay (1980)).

2. B-spline approximation

We use B-splines of the order 3, subdivide the arbitrary interval [0,1] into (k+1) subintervals and choose $knot_j = j * 0.1$ ($j = -3,\dots,k$). Approximating the logarithms of the characteristic functions of the factors and the residuals by B-splines we get the following expressions:

$$\mathrm{Re}\,[\,\ln g_i(t)\,] \approx \sum_{j=-3}^{k} \alpha_{i,j} B_j(t)\,, \qquad (0<t<1;\ i = 1,\dots,p) \tag{5a}$$

$$\mathrm{Re}\,[\,\ln h_i(t)\,] \approx \sum_{j=-3}^{k} \alpha_{p+i,j} B_j(t)\,, \qquad (0<t<1;\ i = 1,\dots,n) \tag{5b}$$

with $\alpha_{i,j}$ the spline coefficients and $B_{-3}(.),\dots, B_k(.)$ a basis for the interval [0,1] (For instance $B_j(t) = (knot_{j+4} - knot_j) \sum_{u=j}^{j+4} \{ \prod_{v=j,\, v\neq u}^{j+4} (knot_v - knot_u)^{-1}\}(t-knot_u)^3_+$; see Powell (1981, page 229).

If the factors are correlated, we can use multivariate B-splines instead of univariate B-splines.

Let $\dot{B}(0)$ be the first order derivative and $\ddot{B}(0)$ the second derivative of the function B(.) in the point $t = 0$. From the properties $g_i(0) = 1$ and $h_i(0) = 1$, and the assumptions E $\underline{f}_i = 0$ and E $\underline{\varepsilon}_i = 0$ it follows that for every $i = 1,\dots,n + p$

$$\sum_{j=-3}^{k} \alpha_{i,j} B_j(0) = 0 \text{ and } \sum_{j=-3}^{k} \alpha_{i,j} \dot{B}_j(0) = 0.$$

Furthermore from the assumption E $\underline{f}_i^2 = 1$ and E $\underline{\varepsilon}_i^2 = U_i$ it follows that

$$\sum_{j=-3}^{k} \alpha_{i,j} \ddot{B}_j(0) = -1 \text{ for every } i = 1,\ldots,p$$

$$\text{and } \sum_{j=-3}^{k} \alpha_{p+i,j} \ddot{B}_j(0) = -U_i \text{ for every } i = 1,\ldots,n.$$

Since $B_j(0) = 0$, $\dot{B}_j(0) = 0$ and $\ddot{B}_j(0) = 0$ for every $j > -1$ we get the results

$$\alpha_{i,-3} B_{-3}(0) + \alpha_{i,-2} B_{-2}(0) + \alpha_{i,-1} B_{-1}(0) = 0 \qquad \text{(with } i=1,\ldots,p+n) \qquad (6)$$

$$\alpha_{i,-3} \dot{B}_{-3}(0) + \alpha_{i,-2} \dot{B}_{-2}(0) + \alpha_{i,-1} \dot{B}_{-1}(0) = 0 \qquad \text{(with } i=1,\ldots,p+n) \qquad (7)$$

$$\alpha_{i,-3} \ddot{B}_{-3}(0) + \alpha_{i,-2} \ddot{B}_{-2}(0) + \alpha_{i,-1} \ddot{B}_{-1}(0) = -1 \qquad \text{(with } i=1,\ldots,p) \qquad (8a)$$

$$\alpha_{i,-3} \ddot{B}_{-3}(0) + \alpha_{i,-2} \ddot{B}_{-2}(0) + \alpha_{i,-1} \ddot{B}_{-1}(0) = -U_i \qquad \text{(with } i=p+1,\ldots,p+n) \qquad (8b)$$

Relations (6), (7) and (8), which follow from relation (2), determine the spline coefficients $\alpha_{i,-3}(0)$, $\alpha_{i,-2}(0)$ and $\alpha_{i,-1}(0)$ for $i = 1,\ldots,p+n$ uniquely.

Furthermore we assume that for every $i = 1,\ldots,n + p$: $\alpha_{i,0} > \alpha_{i,1} > \ldots > \alpha_{i,9}$ (condition (9)). From the results of De Boor (1978; page 109,110,154 and 155) it follows that this is a sufficient condition for the corresponding functions Re [ln $g_j(.)$] and Re [ln $h_i(.)$] ($j = 1,\ldots,p$; $i = 1,\ldots,n$) to become monotone strictly decreasing on the interval [0,1]. For order k<4 of the splines it even is a necessary condition. If we define $\alpha_{i,j} \equiv \sum_{z=-3}^{j} \delta_{i,z}$, then we can reformulate expression (5a) and (5b) in the following way:

$$\text{Re} [\ln g_i(t)] \approx \sum_{j=-3}^{k} [\sum_{z=-3}^{j} \delta_{i,z}] B_j(t) \qquad (i = 1,\ldots,p) \qquad (10a)$$

$$\mathrm{Re}\,[\,\ln h_i(t)\,] \approx \sum_{j=-3}^{k} \;[\, \sum_{z=-3}^{j} \delta_{p+i,z}\,]\, B_j(t) \qquad (i = 1,\ldots,n) \tag{10b}$$

with $\delta_{i,0}, \ldots, \delta_{i,9} < 0$ and $\delta_{i,-3}$, $\delta_{i,-2}$, $\delta_{i,-1}$ fixed from relation (6), (7) and (8).

Substituting expression (10) into equation (3) we get:

$$\mathrm{Re}\,[\,\ln c(t_1,\ldots,t_n)\,] \approx \sum_{i=1}^{p} \;\{\; \sum_{j=-3}^{k} \;[\, \sum_{z=-3}^{j} \delta_{i,z}\,]\, B_j(t_1 A_{1,i} + \ldots + t_n A_{n,i})\,\}$$
$$+ \sum_{i=1}^{n} \;\{ \sum_{j=-3}^{k} \;[\, \sum_{z=-3}^{j} \delta_{p+i,z}\,]\, B_j(t_i)\,\} \tag{11}$$

with $t_1, \ldots, t_n \in [0,1]$.

The left- and right-hand sides of (11) are functions of $(t_1, \ldots, t_n)$. For every different value of combination $(t_1, \ldots, t_n)$ relation (11) generates a different equation. The unknown spline coefficients are determinated uniquely with only a finite number of these approximate equations from expression (11) (see Winsberg and Ransay (1980)). We only use a finite number of values for the combination $(t_1, \ldots, t_n)$ to determine the unknowns factor loadings and spline coefficients in (11). Naturally the number of equations has to be greater than the number of knots and the number of unknown model parameters.

Now we can summarize the estimation procedure as follows.

Step 1: Initial step.

Using the sample covariance matrix of $\underline{y}$ we can estimate the covariance matrix U of the unique residuals $\underline{\varepsilon}$ and the first three spline coefficients for the logarithms of the characteristic functions of the separate factors and residuals. This estimation can be done by a classical factor analysis estimation procedure, which uses first and second order moments only.

Step2: Minimizing over unknown spline coefficients.

Minimizing the GLS-function (4), into which expression (11) is substituted, over the unknown spline coefficients for fixed modelparameters is a quadratic optimization problem with the linear inequality constraints. Because of the monotonicity constraints finding the optimal unknown spline coefficients is a convex programming problem, which can be solved by using the rapid manifold suboptimization method by Zangwill (1969; section 8.3).

Step 3: Minimizing over unknown model parameters.

On the other hand minimizing GLS-function (4) over the model parameters is a nonlinear optimization problem. We use the Newton-Raphson method (see also Zangwill (1969)).

Step 4: Steps 2 and 3 can be repeated until converge is guaranteed.

3. Asymptotic properties

Until now we dealt with a fixed arbitrary number of knots on the interval [0,1]. In this section the number of knots of our approximation splines is not fixed. We derive the number of knots k_{sl} as a strictly increasing function of the sample-size sl.

Let the functions $\ln g_1(.),\ldots, \ln g_p(.)$, $\ln h_1(.),\ldots, \ln h_n(.)$ be 2 times continuous differentiable on the interval [0,1] and be approximated by spline functions of order three. Under these conditions, we can use the results of Powell (1981, chapter 20) to determine the approximation error for the B-spline-approximations of the logarithms of the characteristic functions of the factors and the residuals. It is not useful to make the approximation errors smaller than the variances of the logarithms of the empirical characteristic function of $\underline{y}$ in different points $\tau = (t_1 ,\ldots, t_n)$ $(t_1 ,\ldots, t_n \in[0,1])$. Now for a fixed sample-size sl we define the number of knots k_{sl} in the following way:

k_{sl} is the largest integer for which the following constraints are satisfied

$$1/8\ [\max_{t \in [0,1]} (\partial^2 \ln g_j(t) / \partial t^2)]\ d^2 \geq [\max_{t1 ,\ldots, tn \in[0,1]} (\mathrm{var} \ln \hat{c}\,(t_1 ,\ldots, t_n))] \qquad (j = 1,\ldots,p) \qquad (12a)$$

and

$$1/8\ [\max_{t \in [0,1]} (\partial^2 \ln h_j(t) / \partial t^2)]\ d^2 \geq [\max_{t1 ,\ldots, tn \in[0,1]} (\mathrm{var} \ln \hat{c}\,(t_1 ,\ldots, t_n))] \qquad (j = 1,\ldots,n) \qquad (12b)$$

with d the maximal distance between two knots.

The lef- hand sides of (12a) and (12b) represent the spline approximation-upper-bounds for

$\ln g_1(.),\ldots, \ln g_p(.)$, $\ln h_1(.),\ldots, \ln h_n(.)$. The right-hand sides of the above constraints represent the maximum of the variances of the logarithms of the empirical characteristic function of $\underline{y}$ in several points $\tau = (t_1 ,\ldots, t_n)$.

Using (12a,12b) and an uniform knot placement we can easily see that k_{sl} is an increasing function of sl. If $sl \to \infty$ then $k_{sl} \to \infty$. For $sl \to \infty$ the logarithms of the characteristic functions of the factors and the residuals will be approximated exactly. In that case Feuerverger and McDunnough (1981) prove the BGLS-estimates from (4) converge almost surely to the true values of the unknown parameters. Furthermore the estimators are asymptotically normally distributed with asymptotic variance $\dot{Z}\ \mathbf{V}^{-1}\dot{Z}$ ($\dot{Z}$ is the derivative from $\mathbf{Z}$ with respect to the unknown parameters). Feuerverger and McDunnough (1981) also show that by using a sufficiently extensive grid $\{ \tau_i = (t_1^{(i)} ,\ldots, t_n^{(i)}),\ i = 1,\ldots, \text{dimension } \mathbf{Z}\}$ the asymptotic variances of our BGLS- estimators can be made arbitrarily close to the Cramer-Rao bound. So, our procedure can attain arbitrarily high asymptotic efficiency.

If the model is correct the optimization function Ψ is asymptotically chi-square distributed with degrees of freedom equal to the number of components of the vector $\mathbf{Z}$ minus the number of independent parameters. The number of components of $\mathbf{Z}$ corresponds to the number of different values for the combination $(t_1, \ldots, t_n)$. The number of independent parameters is equal to the sum of the unknown factor loadings and the unknown spline coefficients.

Finally we remark that the dimension of $\mathbf{Z}$ in expression (4), which is equal to the number of different values for the combinations $(t_1 ,\ldots, t_n)$, depends on the number of spline-knots. The dimension of $\mathbf{Z}$ has to be larger than the number of knots k_{sl} on the interval [0,1]. If $sl \to \infty$ and $k_{sl} \to \infty$, then dimension $\mathbf{Z} \to \infty$.

4. Monte Carlo simulation

In this example we discuss, by means of some generated data, our estimation method. Besides our method we also apply some well-known rotation methods for factor loadings estimates like Quartimax and Varimax, which only use first and second order moments. These estimation methods will be run with the help of standard SPSS-software (version 10.0). We shall generate some models in which the factor loadings are identified and we will check whether the correct factor loadings will be estimated. Because sample size may play an important role with respect to the accuracy of the estimators, both sizes of 50 and 200 are used.

Scores on f_1, f_2, e_1, e_2, e_3, e_4 and e_5 are drawn independently from a chi-squared distribution with one and two degrees of freedom for the common factors f_1 and f_2 and from a standard normal distribution for the unique residuals e_1, e_2, e_3, e_4 and e_5. Next the common factors are normalised (expectation 0 and standard deviation 1) and the residuals are multiplied by arbitrary random numbers between 0 and 1. The factor loadings are also chosen randomly in the interval [0,1]. Using the scores for the common factors, the scores for the unique residuals and the values for the factor loadings we calculate the scores for the observed variables y_1, y_2, y_3, y_4 and y_5. This simulation will be repeated 50 and 200 times with the same factor loading matrix, and will lead to 50 and 200 observations for y_1 up to y_6. From these y_1, y_2, y_3, y_4 and y_5 scores the estimates of the initial factor loadings are obtained. Mooijaart (1985) proves that this model is identified with higher order moments. Now we apply the Quartimax and Varimax method, and our characteristic function method to obtain estimates for the factor loadings. Therefore, the empirical characteristic function of $\underline{y}$ is computed in arbitrary values for the combination (t_1,t_2,t_3,t_4,t_5), with each value for t between 0 and 1/5. The number of replications in this study is equal to 1000.

Our estimation procedure is programmed with the programming language MATLAB. The number of internal knots on the interval [0,1] for the B-splines approximation will be two (uniform knot placement). The starting values of the factor loadings for our iterative estimation procedure are obtained by the Varimax estimates. In this example we choose the approximation interval [0,1]. If the arguments of the logarithms of the characteristic functions are going beyond the approximation interval [0,1], our method has some complications. In that situation we drop out the t's of which the corresponding arguments are outside the approximation interval. However, using this correction method we are changing the optimisation function.

For sample size 50 and 200 the estimation results of our method are acceptable (see Table 1). The estimates of the factor loadings are in the neighbourhood of the true factor loading values. The average differences between the true parameter values and the estimates are within the ranges [0.137;0.149] and [0.091;0.114]. All the standard deviations of the estimates are smaller than 0.094. We repeated the Monte Carlo study with the number of knots on the interval [0;1] set equal to four instead of two. This did not change the estimation results appreciably.

Without knowing the population factor loading matrix we are not able to construct a rotation method, which leads to estimates close to the characteristic functions estimates. The estimation results of the Quartimax and Varimax method do not differ that much from each other (see Table 1). Their bias is about 0.20 larger than the bias of our estimation method. On an average the factor loadings can differ 0.30 from the true values. For sample size 50 and 200 the differences are nearly the same. These results indicate that the classical estimation methods give estimates that are not allowed. This is, because they do not use more information than only first and second order moments.

5. Empirical example

Using the factor analysis method introduced in this paper we estimate the factor loadings in an empirical example. The starting values for the factor loading estimates are equal to the Varimax estimates. Once again, we use the approximation interval [0,1] and two internal knots on this interval for the approximation splines.

The data analysed here were collected by The Economic Social Institute (ESI) of the Free University in Amsterdam measuring five characteristics of 167 entrepreneurs in the field of small and medium-sized enterprises sector (see Table 2). The scores on the characteristics are from 1 (= not important) to 10 (=very important). A Kolmogorov test shows that all the five variables cannot be assumed to come from a normal distribution (significance level 0.05). All the variables were negatively skewed.

We deal with a factor model, which contains two orthogonal common factors. The factor model is fitted only, to show that this can be done. The factor loadings, estimated by the method of this paper, are presented in Table 3. The corresponding standard deviations of the factor loading estimators, computed with the bootstrap method, are drawn between brackets.

If the model is correct, the minimizing function in (4) is asymptotically chi-squared distributed with degrees of freedom equal to the number of elements in the vector **Z** minus the number of independent unknown parameters. The number of elements of **Z** corresponds with the number of different values for the combination (t_1,t_2,t_3,t_4,t_5). The number of independent unknown parameters is equal to the sum of the unknown factor loadings (=10), the unknown spline coefficients corresponding with the common factors and the residuals (=21). So, in this case the optimization function is chi-squared distributed with 40 – 10 – 21 = 9 degrees of freedom. The value for the chi-squared minimizing function is 8.0. The probability level is about 0.40 and we cannot reject the model on the basis of the data.

The estimated factor loading matrix is also used to obtain factor scores for each of the 167 respondents. From these factor scores it follows that both common factors are negative skewed. The Kolmogorov test indicates that the common factors do not come from a normal distribution (significance level 0.05).

Table 3 shows, that the first common factor corresponds with the observed variables representing some internal oriented characteristics of the entrepreneurs. The second common factor is only strongly correlated with the externally oriented variable, i.e., the attitude towards clients.

Table 1, *Monte Carlo results for characteristic function method, Quartimax method and Varimax method.*

	Estimates characteristic function method		Estimates Quartimax Method		Estimates Varimax Method	
	N=50	N=200	N=50	N=200	N=50	N=200
Average diff. A_{11} Stand.dev. diff. A_{11}	0.148 (0.094)	0.106 (0.063)	0.235 (0.179)	0.198 (0.150)	0.253 (0.188)	0.219 (0.174)
Average diff. A_{21} Stand.dev. diff. A_{21}	0.140 (0.068)	0.114 (0.070)	0.229 (0.176)	0.202 (0.157)	0.244 (0.185)	0.224 (0.172)
Average diff. A_{31} Stand.dev. diff. A_{31}	0.137 (0.086)	0.102 (0.062)	0.229 (0.182)	0.196 (0.153)	0.246 (0.193)	0.216 (0.169)
Average diff. A_{41} Stand.dev. diff. A_{41}	0.142 (0.090)	0.108 (0.066)	0.283 (0.182)	0.205 (0.155)	0.254 (0.193)	0.221 (0.170)
Average diff. A_{51} Stand.dev. diff. A_{51}	0.137 (0.084)	0.111 (0.069)	0.231 (0.181)	0.204 (0.156)	0.246 (0.194)	0.223 (0.176)
Average diff. A_{12} Stand.dev. diff. A_{12}	0.139 (0.087)	0.092 (0.057)	0.4293 (0.289)	0.436 (0.284)	0.338 (0.242)	0.301 (0.213)
Average diff. A_{22} Stand.dev. diff. A_{22}	0.149 (0.094)	0.100 (0.063)	0.434 (0.291)	0.454 (0.230)	0.337 (0.233)	0.304 (0.226)
Average diff. A_{32} Stand.dev. diff. A_{32}	0.136 (0.086)	0.099 (0.064)	0.443 (0.290)	0.445 (0.281)	0.341 (0.245)	0.296 (0.209)
Average diff. A_{42} Stand.dev. diff. A_{42}	0.139 (0.081)	0.091 (0.057)	0.433 (0.292)	0.459 (0.294)	0.348 (0.240)	0.306 (0.222)
Average diff. A_{52} Stand.dev. diff. A_{52}	0.142 (0.087)	0.099 (0.063)	0.444 (0.280)	0.443 (0.289)	0.348 (0.237)	0.308 (0.224)

Table 2, *Five characteristics of entrepreneurs in the small and medium-sized enterprises sector*

VARIABLE 1: Commitment and determination
VARIABLE 2: Persistence in problem solving
VARIABLE 3: Creativity
VARIABLE 4: Team spirit and motivational capacities
VARIABLE 5: Client orientation

Table 3, *The estimates of the factor loadings for the ESI data* (corresponding standard deviation between brackets)

Factor loadings	Estimates	Factor loadings	Estimates
A_{11}	0.819 (0.053)	A_{12}	-0.189 (0.022)
A_{21}	0.846 (0.046)	A_{22}	-0.180 (0.012)
A_{31}	0.731 (0.031)	A_{32}	-0.120 (0.010)
A_{41}	0.650 (0.060)	A_{42}	0.303 (0.027)
A_{51}	0.215 (0.033)	A_{52}	0.924 (0.054)
Explained variance	46.1 %	Explained variance	20.6%

Df=9
χ^2 =8.0

References.

Anderson, T.W. and Rubin H. (1956). Statistical inference in factor analysis. Proceedings of the third Berkeley symposium on mathematical statistics and probability, 5, 111-150.

Bentler, P.M. (1983). Simultaneous equation systems as moments structure models: With an introduction to latent variable models. Journal of Econometrics, 22, 13-42.

De Boor, C. (1978). A practical guide to splines. Berlin, Springer.

Browne, M.W. (1984). Asymptotic distribution-free methods for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 37, 62-83.

De Leeuw, J. and J. Van Rijckevorsel (1981). Non linear principal component analysis with B-splines. Research Note RR-81-02, Erasmus University Rotterdam.

Feuerverger, A. and P.McDunnough (1981). On the Efficiency of Empirical Characteristic Function Procedures. Journal of the Royal Statistical Society B, 43, 1, 20-27.

Ichikawa, M. and Konishi (1995). Application of the bootstrap methods in factor analysis. Psychometrika, 60, 77-93.

Krijnen, W.P., Dijkstra, T.K. and R.D. Gill (1998). Conditions for factor (in)determinancy in factor analysis. Psychometrika, 63, 4, 359-367.

Lewis-Beck, S.L. (1994). Factor Analysis and Related Techniques. Sage Publication, London.

Meulman, J.J. (1986). A distance approach to nonlinear multivariate analysis, Leiden, DSWO Press.

Mooijaart, A. (1985). Factor analysis for non-normal variables. Psychometrika, 50, 323-342.

Muthen, B. (1984). A general structural equation model with dichotomous, ordered, categorical, and continuous latent variable indicators. Psychometrika, 49, 1, 115-132.

Powell, M. (1981). Approximate theory and methods. Cambridge, Cambridge University Press.

Ralston, A. and P. Rabinowitz (1978). A first course in numerical analysis. McGraw-Hill International Editions.

Schwetlick, H. (1979). Numerische Lösung nichtlinearer Gleichungen. Berlin, VEB Deutscher Verlag der Wissenschaft.

Van Rijckevoorsel, J. (1987). The application of fuzzy coding and horsehoes in multiple correspondence analysis, DSWO Press, Leiden.

Winsberg, S. and J.O. Ramsay (1980). Monotonic transformations to additivity using splines. Biometrika, 67, 3, 669-674.

Wold, S. (1974). Spline functions in data analysis. Technometrics, vol. 16, no. 1, 1-11.

Wold, H. (1982). Soft modelling: the basic design and some extensions. In K.G. Jøreskog and H. Wold (eds.), Systems under indirekt observation. Causality, structure, prediction. Amsterdam, North Holland Publishing Co.

Yuan, K.-H., Marshall, L.L. and P.M. Bentler (2002). An unified approach for exploratory factor analysis with missing data, nonnormal data, and in the presence of outliers. Psychometrika, 67, 95-122.

Young, F.W. (1981). Quantitative analysis of qualitative data. Pschychometrika, 46, 357-388.

Zangwill, W.I. (1969). Nonlinear programming. A unified approach. Englewood Cliffs, Prentice Hall.