Influence diagnostics for principal factor analysis

By Reinhold Kosfeld

Summary: Outliers can have uncontrolled effects on the results of a factor analysis. A valid interpretation of a factor structure in empirical research therefore requires a control of this special kind of outliers. They are called influential observations and can be identified by adapting Hampel's concept of the influence function. In this paper we apply Hampel's concept for developing influence diagnostics for exploratory factor analysis. We derive alternative empirical versions of influence functions as explorative tools for identifying outliers in principal factor analysis.

Keywords: Factor analysis, influence function, influential observations. JEL C49

1 Introduction

The identification of outliers with a general outlier detection procedure poses many problems regarding application. Since outliers are heterogeneous, a detection method will fail to recognize all outliers. Outliers with respect to location e.g. need not necessarily be outliers with respect to the correlation structure. Although an outlier detection method can perhaps identify both of these outliers in some cases, it cannot be used in identifying others.

Empirical research gives high priority to detecting 'bad data' which have an effect on the results given by a statistical method. These kinds of outliers are important in application because they can invalidate all conclusions deriving from a data analysis. The relevance of an identification of such 'bad data' in factor analysis is emphasized by Huber (1981, p. 199) who states that " all too often a ... factor analysis 'explains' a structure that, ..., has been created by a mere one or two outliers ...". Outliers which have an effect on a special statistical method are called influential observations. It is this kind of deviate data we are concerned with here.

Hampel (1974) has developed the concept of the influence function that provides a basis for identifying influential observations. It always requires a method specific frame, because whether an observation is influential or not can only be judged with reference to a special statistical method. For using this instrument in empirical research it is necessary to make it operational. In dependence on the concrete setting one can obtain alternative empirical versions of the influence function. The valuation of the observations follows from an adequate norm of an empirical version of the influence function which provides influence diagnostics for a statistical method.

In regression analysis there exists well developed influence diagnostics which have proved to be valuable in econometric applications (s. e.g. Belsley, Kuh, and Welch, 1981). Influence functions in principal components were derived by e.g. Critchley (1985) on the basis of the perturbation theory of eigenvalues and eigenvectors (Wilkinson, 1965; Sibson, 1979). In this paper we use the same approach to develop influence diagnostics for principal factor analysis. Although we adopt a covariance-oriented framework for this purpose, according to the perturbational analysis it is based, the influence diagnostics are useful in a correlation-oriented framework as well (see Critchley, 1985, p. 635).

Tanaka and Odaka (1989a) have dealt with this problem in a somewhat different approach of sensitivity analysis (for ML factor analysis see Tanaka and Odaka, 1989b). The authors develop influence functions for the product of the factor matrix and its transpose whereas our focus explicitely lies on the factor matrix. It is this matrix which plays a crucial role in interpreting the results of factor analysis in applications. Castaño-Tostado and Tanaka (1991) discuss two matrix coefficents (especially Escoufier's RV-coefficient) to detect influential observations. Our approach aims to operationalise pertubational effects in a more data-oriented way by studying alternative versions of empirical influence functions. Its efficient use in empirical research stems from the fact that only one eigenanalysis is required for determining influence diagnostics for all observations. The influence diagonstics are not affected by factor rotation.

2 The factor analytic model

Let $\mathbf{x} = (X_1, X_2, \ldots, X_p)'$ be a $p \times 1$ random vector of manifest variables whose distribution function $F_{\boldsymbol{\theta}}(x)$ is an element of a class $\mathcal{F}(\Omega) = \{F_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathrm{P}_{\boldsymbol{\theta}}\}$ of distribution functions with the parameter vector $\boldsymbol{\theta} \in \mathrm{P}_{\boldsymbol{\theta}}.\mathrm{P}_{\boldsymbol{\theta}}$ denotes the parameter space and Ω the sample space. Generally only the first two moments of the distribution of the random vector $\mathbf{x}$ are considered in factor analysis:

$$\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}(F) = [\mu_1(F), \mu_2(F), \ldots, \mu_p(F)]',$$
$$\mathrm{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}(F) = [\sigma_{ij}], \quad i, j = 1, 2, \ldots, p\ .$$

This implies considering the class

$$\mathcal{F}_N(\Omega) = \{F_{\boldsymbol{\mu},\boldsymbol{\Sigma}} : \boldsymbol{\mu} \in \mathrm{P}_{\boldsymbol{\mu}}, \boldsymbol{\Sigma} \in \mathrm{P}_{\boldsymbol{\Sigma}}\} \subset \mathcal{F}(\Omega)$$

of p-variate normal distributions with the parameter spaces $P_{\boldsymbol{\mu}} \subseteq \mathbb{R}^p$ and $P_{\boldsymbol{\Sigma}} \subset \mathbb{R}^{p \times p}$.

The multiple factor model is based on the notion that the p manifest variables $X_j,\ j = 1, 2, \ldots, p$, are generated by a smaller number of common factors $F_k,\ k = 1, 2, \ldots, m$, and p specific or unique factors $U_j,\ j = 1, 2, \ldots, p$. Let $\mathbf{f} = (F_1, F_2, \ldots, F_m)'$ be an $m \times 1$ random vector of the common factors, $\mathbf{u} = (U_1, U_2, \ldots, U_p)'$ a $p \times 1$ random vector of the unique factors and $\boldsymbol{\Lambda}(F) = [\lambda_{jk}(F)], j = 1, 2, \ldots, p;\ k = 1, 2, \ldots, m$, the factor matrix whose elements λ_{jk} are called factor loadings. Then the structural relation between the variables and factors can be written in the form

$$\mathbf{x} = \boldsymbol{\mu}(\mathrm{F}) + \boldsymbol{\Lambda}(\mathrm{F}) \cdot \mathbf{f} + \mathbf{u} \tag{1}$$

under the linearity assumption. Without loss of generality one can set $\mathrm{E}(\mathbf{x}) = \mathbf{0}$ and $\mathrm{E}(\mathbf{u}) = \mathbf{0}$. Furthermore the definitions and assumptions

$$\begin{aligned} \mathrm{Cov}(\mathbf{f}) &= \mathrm{E}(\mathbf{ff}') = \mathbf{\Phi}(F), \\ \mathrm{Cov}(\mathbf{u}) &= \mathrm{E}(\mathbf{uu}') = \mathbf{\Psi}(F) = \mathrm{diag}[\psi_1(F), \psi_2(F), \ldots, \psi_p(F)]' \,, \\ \mathrm{E}(\mathbf{fu}') &= \mathbf{0} \end{aligned}$$

are set (see e.g. Ost, 1996, pp. 641; Seber, 1983, pp. 213; Kosfeld, 1986, pp. 17). The covariance matrix of the common factors, $\mathbf{\Phi}(F)$, is an element of the parameter space $P_{\mathbf{\Phi}} \subset \mathbb{R}^{m \times m}$ and the covariance matrix of the unique factors, $\mathbf{\Psi}(F)$, belongs to the parameter space $P_{\mathbf{\Psi}} \subset \mathbb{R}^{p \times p}$.

The estimation of the factor matrix $\mathbf{\Lambda}(F)$ is of primary importance in factor analysis. Since the common factors are latent variables, it cannot be accomplished directly from the structural relationship. A special case of the analysis of covariance structures shows that the estimation procedure is based on the relation

$$\mathbf{\Sigma}(F) = \mathbf{\Lambda}(F) \cdot \mathbf{\Phi}(F) \cdot \mathbf{\Lambda}'(F) + \mathbf{\Psi}(F) \tag{2}$$

which is called the "fundamental theorem" of factor analysis. To determine a direct solution, the covariance matrix $\mathbf{\Phi}(F)$ of the common factors is made equal to the $m \times m$ unit matrix $\mathbf{I}$. The assumption of uncorrelated factors must not necessary be retained, because the choice of a rotation to a "simple structure" with oblique factors can follow in a later step.

Beside the factor matrix $\mathbf{\Lambda}(F)$ the covariance matrix $\mathbf{\Psi}(F)$ of the unique factors must be estimated on the basis of the fundamental theorem 3.2. The elements of the main diagonal, ψ_j, can be interpreted as parts of the variances of the manifest variables not "explained" by the common factors. Often instead of the unique variances their counterparts, the communalities $h_j^2(F)$,

$$h_j^2(F) = \sigma_{jj} - \psi_j \,, \tag{3}$$

are considered in empirical investigations.

In principal factor analysis the factor extraction is performed from an estimator of the "reduced" covariance matrix

$$\widetilde{\mathbf{\Sigma}}(F) = \mathbf{\Lambda}(F) \cdot \mathbf{\Lambda}'(F) \tag{4}$$

after determining the number of common factors and estimating the communalities. It is generally indicated to iterate the estimators as long as the unique variances remain non-negative. In practice, however, it may be meaningful to perform only a few iterations (see Cureton and D'Agostino, 1983, pp. 139).

3 Influence functions

3.1 The concept of the influence function

Although outliers are somehow deviate and atypical, they are not necessarily always influential with respect to a statistical method. In statistical applications only influential observations are critical. It is those kinds of deviate observations that have strong effects on a statistical analysis. At worst they can invalidate the entire result of a statistical application. A theoretical tool for identifying such observations is the concept of influence function (Hampel, 1974; Hampel et al. 1986, pp. 81–87, 226–227).

In factor analysis the influence function has to be considered as a vector-valued function. Let $\mathbf{T}(F)$ and $\mathbf{T}(\widetilde{F})$ be functionals on the set $\mathcal{F}$ of distribution functions F and $\widetilde{F}$, respectively. Here $\widetilde{F}$ is a contaminated distribution function

$$\widetilde{F} = (1-\varepsilon)\cdot F + \varepsilon \cdot \delta_{\mathbf{x}} \ , \quad 0 < \varepsilon < 1 \ , \tag{5}$$

where $\delta_{\mathbf{x}}$ is the Dirac measure which concentrates the whole probability mass on the point $\mathbf{x}$. The influence function $\mathbf{IF}$ of $\mathbf{T}$ at F is point by point defined as

$$\mathbf{IF}(\mathbf{x};\mathbf{T},F) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}(\mathbf{T}(\widetilde{F}) - \mathbf{T}(F)) \ . \tag{6}$$

Obviously it is the right-side derivative of $\mathbf{T}(\widetilde{F})$ at the point $\varepsilon = 0$. The influence function (6) measures the effect of an infinitesimally small contamination on a functional $\mathbf{T}(F)$ standardized on the portion of this contamination.

To determine the effect of the individual observations on the factor loadings, we consider at first the column vectors $\boldsymbol{\lambda}_k(F)$ of the factor matrix $\boldsymbol{\Lambda}(F)$ separately. In principal factor analysis they have the form

$$\boldsymbol{\lambda}_k(F) = \tau_k^{1/2}\cdot\boldsymbol{\omega}_k \ , \quad k = 1,2,\ldots,m \ , \tag{7}$$

where the scalars $\tau_k, k = 1,2,\ldots,m$, are the largest eigenvalues of the reduced covariance matrix $\widetilde{\boldsymbol{\Sigma}}(F)$ and the vectors $\boldsymbol{\omega}_k$ the corresponding orthonormal eigenvectors (s. Harman, 1976, pp. 136). Hence $\boldsymbol{\Lambda}(F)$ is here conceived as the unrotated factor matrix. Afterwards it will be shown that the influence functions of the rotated factor matrix can immediately be gained from those of the unrotated factor matrix. Furthermore it will turn out that special influence measures are identical for both matrices.

The perturbation theory of eigenvalue problems (Wilkinson, 1965, pp. 65–109) provides a basis for the development of influence functions of the factor vectors $\boldsymbol{\lambda}_k(F)$. Sibson (1979) and Critchley (1985) have also used this framework for robustness studies in other methods of multivariate analysis. In the first place for factor analysis one has to obtain some knowledge about the structure of the perturbed reduced covariance matrix $\widetilde{\boldsymbol{\Sigma}}(\widetilde{F})$:

Theorem 3.1 *The perturbed reduced covariance matrix $\widetilde{\boldsymbol{\Sigma}}(\widetilde{F})$ has the form*

$$\widetilde{\boldsymbol{\Sigma}}(\widetilde{F}) = \widetilde{\boldsymbol{\Sigma}}(F) + \varepsilon\cdot[\tilde{\mathbf{x}}\tilde{\mathbf{x}}' - \widetilde{\boldsymbol{\Sigma}}(F)] - \varepsilon^2\cdot\tilde{\mathbf{x}}\tilde{\mathbf{x}}' \tag{8}$$

$$\textit{with} \qquad \tilde{\mathbf{x}} = \mathbf{x} - \boldsymbol{\mu}(F) - \mathbf{u} \ . \tag{9}$$

Proof. See Appendix. □

Since the factor vectors $\boldsymbol{\lambda}_k(F)$ are determined by the eigenvalues $\tau_k(F)$ and eigenvectors $\boldsymbol{\omega}_k(F)$ of the matrix $\widetilde{\boldsymbol{\Sigma}}(F)$, the effect of the perturbation on these quantities at first requires quantification. Here we suppose that all eigenvalues differ. The perturbed eigenvalues and eigenvectors, $\tau_k(\widetilde{F})$ and $\boldsymbol{\omega}_k(\widetilde{F})$, can be written as Taylor expansions of the forms

$$\tau_k(\widetilde{F}) = \tau_k(F) + \varepsilon \cdot \mathrm{v}_k + \frac{1}{2}\varepsilon^2 \cdot \mathrm{w}_k + O(\varepsilon^3) \ , \tag{10}$$

$$\boldsymbol{\omega}_k(\widetilde{F}) = \boldsymbol{\omega}_k(F) + \varepsilon \cdot \mathbf{g}_k + \frac{1}{2}\varepsilon^2 \cdot \mathbf{h}_k + O(\varepsilon^3) \ , \tag{11}$$

where $O(\varepsilon^3)$ denotes a remainder which is only dependent on cubic or higher powers in ε. The scalars v_k and w_k as well as the $m \times 1$ vector $\mathbf{g}_k$ can be determined on the basis of a lemma of Sibson (1979, p. 219). Since we need also an expression for the $m \times 1$ vector $\mathbf{h}_k$, an extension of this lemma is necessary. First of all, we give a general formulation of the extended lemma, in which a perturbation of a matrix $\mathbf{B}$ not necessarily regular is considered. At this point the concept of the generalized inverse must be introduced. Let $\mathbf{M}$ be a symmetric $p \times p$ matrix with the spectral decomposition

$$\mathbf{M} = \sum_{i=1}^{p} l_i \cdot \mathbf{e}_i \cdot \mathbf{e}_i' \ , \tag{12}$$

where l_i is the ith eigenvalue of $\mathbf{M}$ and $\mathbf{e}_i$ the corresponding eigenvector. Then the matrix

$$\mathbf{M}^+ = \sum_{\substack{i=1 \\ l_i \neq 0}}^{p} l_i^{-1} \cdot \mathbf{e}_i \cdot \mathbf{e}_i{}' \tag{13}$$

with the properties

$$\mathbf{M} \cdot \mathbf{M}^+ \cdot \mathbf{M} = \mathbf{M} \ , \tag{14}$$

$$\mathbf{M}^+ \cdot \mathbf{M} \cdot \mathbf{M}^+ = \mathbf{M}^+ \tag{15}$$

is the generalized inverse of $\mathbf{M}$. According to the properties (14) and (15) $\mathbf{M}$ and $\mathbf{M}^+$ are mutually generalized inverses.

Using these properties Lemma 3.1 (see Appendix) can be established. It is the basis for quantifying the effects of perturbations on the eigenvalues and eigenvectors of the reduced covariance matrix.

Theorem 3.2 *The coefficients* $\mathrm{v_k}$, $\mathbf{g_k}$, $\mathrm{w_k}$ *and* $\mathbf{h_k}$ *of the taylor expansions (10) and (11) are given by*

$$\mathrm{v_k} = \boldsymbol{\omega}_k{}' \cdot \tilde{\mathbf{x}}\tilde{\mathbf{x}}' \cdot \boldsymbol{\omega}_k - \tau_k, \tag{16}$$

$$\mathbf{g}_k = -\sum\nolimits_{(k)} (\tau_h - \tau_k)^{-1} \cdot \boldsymbol{\omega}_h \boldsymbol{\omega}_h{}' \cdot \tilde{\mathbf{x}}\tilde{\mathbf{x}}' \cdot \boldsymbol{\omega}_k \ , \tag{17}$$

$$\mathrm{w_k} = 2\boldsymbol{\omega}_k \cdot \{[\tilde{\mathbf{x}}\tilde{\mathbf{x}}' - \sum \tau_h \cdot \boldsymbol{\omega}_h \boldsymbol{\omega}_h{}'] \cdot \mathbf{g}_k - \tilde{\mathbf{x}}\tilde{\mathbf{x}}' \cdot \boldsymbol{\omega}_k\} \tag{18}$$

$$\mathbf{h}_k = \sum\nolimits_{(k)} (\tau_h - \tau_k)^{-1} \boldsymbol{\omega}_h \boldsymbol{\omega}_h{}' \cdot \{2[\mathrm{v}_k \cdot \mathbf{I} - \tilde{\mathbf{x}}\tilde{\mathbf{x}}' + \sum \tau_j \boldsymbol{\omega}_j \boldsymbol{\omega}_j{}'] \cdot \mathbf{g}_k + (\mathrm{w}_k \cdot \mathbf{I} + 2 \cdot \tilde{\mathbf{x}}\tilde{\mathbf{x}}') \cdot \boldsymbol{\omega}_k\} \ , \tag{19}$$

where $\sum_{(k)}$ *stands for the summation over all terms not having the subscript* k.

Proof. See Appendix. □

In contrast to factor analysis the problem of a generalized inverse does not appear in principal components. The covariance matrix underlying principal components is always positive definite, whereas the reduced covariance matrix underlying a factor analysis is indefinite. Although the manifest variables are generated by the factors, an interpretation of the influence of the observations in dependence on the factor scores cannot immediately be given in factor analysis. Because of their latency, the factor scores can only be estimated from the data. On the other hand with the component scores in principal components one will yield a rather vivid presentation, inasmuch as they arise directly as a linear transformation of the observations.

Indeed the "reduced" observation vector $\tilde{\mathbf{x}}$ occurs with the coefficients of the perturbed eigenvalues and eigenvectors of the reduced covariance matrix $\widetilde{\boldsymbol{\Sigma}}(\widetilde{F})$ in the influence functions of the parameters of the factor analytic model. Since $\tilde{\mathbf{x}}$ must be estimated for the observations from the factor scores, it must nevertheless be recurred on them. But under this genuine estimation aspect the factor scores are solely included in the influence diagnostics.

Theorem 3.3 *The influence function* $\mathbf{IF}(\mathbf{x}; \boldsymbol{\lambda}_k, F)$ *of the vector of unrotated loadings of the kth common factor on the manifest variables,* $\boldsymbol{\lambda}_k(F)$*, has the form*

$$\mathbf{IF}(\mathbf{x}; \boldsymbol{\lambda}_k, F) = \tau_k^{1/2} \cdot (\frac{1}{2}\tau_k^{-1} \cdot \mathrm{v_k} \cdot \boldsymbol{\omega}_k + \mathbf{g}_k) \; . \tag{20}$$

The rotated factor vector $\boldsymbol{\lambda}_k^*(F)$ *has the influence function*

$$\mathbf{IF}(\mathbf{x}; \boldsymbol{\lambda}_k^*, F) = [\tau_1^{1/2}(\frac{1}{2}\tau_1^{-1}\mathrm{v_1}\boldsymbol{\omega}_1 + \mathbf{g}_1), \dots, \tau_m^{1/2}(\frac{1}{2}\tau_m^{-1}\mathrm{v_m}\boldsymbol{\omega}_m + \mathbf{g}_m)] \cdot \mathbf{t}^k \; , \tag{21}$$

in which $\mathbf{t}_k$ *contains the elements of the kth column of the transformation matrix* $\mathbf{T}'$.

Proof. Using the relation (6) $\mathbf{IF}(\mathbf{x}; \boldsymbol{\lambda}_k, F)$ is given by

$$\mathbf{IF}(\mathbf{x}; \boldsymbol{\lambda}_k, F) = \lim_{\varepsilon \to 0} \{\boldsymbol{\lambda}_k[(1-\varepsilon)F + \varepsilon\delta_{\mathbf{x}}] - \boldsymbol{\lambda}_k(F)\}/\varepsilon \tag{22}$$

with expression (7) representing $\boldsymbol{\lambda}_k(F)$. If one inserts the Taylor expansions (10) and (11) for $\tau_k(\widetilde{F})$ and $\boldsymbol{\omega}_k(\widetilde{F})$ in equation (22), one will obtain the expression

$$\mathbf{IF}(\mathbf{x}; \boldsymbol{\lambda}_k, F) = \lim_{\varepsilon \to 0}[\tau_k^{1/2}(1 + \frac{1}{2}\varepsilon \mathrm{v}_k \tau_k^{-1})(\boldsymbol{\omega}_k + \varepsilon \mathbf{g}_k) - \tau_k^{1/2}\boldsymbol{\omega}_k]/\varepsilon \; ,$$

with regard to the formation of the power expansion (s. Bronstein and Semendjajew, 1985, p. 31):

$$[1 + \varepsilon \mathrm{v}_k \tau_k^{-1} + O(\varepsilon^2)]^{1/2} = 1 + \frac{1}{2}\varepsilon \mathrm{v}_k \tau_k^{-1} + O(\varepsilon^2) \; .$$

Hence equation (20) follows as the differential quotient at the point $\varepsilon = 0$.

The influence function $\mathbf{IF}(\mathbf{x}; \boldsymbol{\lambda}_k^*, F)$ of the column vector $\boldsymbol{\lambda}_k^*$ of the rotated factor matrix $\boldsymbol{\Lambda}^*(F)$ is obtained by use of the relation

$$\boldsymbol{\lambda}_k^*(\widetilde{F}) = [\boldsymbol{\lambda}_1(\widetilde{F}), \boldsymbol{\lambda}_2(\widetilde{F}), \dots, \boldsymbol{\lambda}_m(\widetilde{F})] \cdot \mathbf{t}_k{}' \; . \tag{23}$$

□

The interpretation of the influence functions (20) and (21) is not elementary. The first expression $\frac{1}{2}\tau_k^{-1/2}\mathrm{v}_k\boldsymbol{\omega}_k$ in $\mathbf{IF}(\boldsymbol{\lambda}_k)$,

$$\mathbf{IF}(\boldsymbol{\lambda}_k) \equiv \mathbf{IF}(\mathbf{x};\boldsymbol{\lambda}_k, F) \ ,$$

contains only those quantities determining the factor vector $\boldsymbol{\lambda}_k$. But the amount of the contribution of the eigenvalues τ_k and eigenvectors $\boldsymbol{\omega}_k$ is rated by the difference between the matrix of cross products of the "reduced" contaminant $\tilde{\mathbf{x}}$ and the reduced covariance matrix $\widetilde{\boldsymbol{\Sigma}}(F)$. On the other hand, the change of $\boldsymbol{\lambda}_k(F)$ on the second expression $\tau_k^{1/2}\mathbf{g}_k$ does not only dependent on the direction of the other eigenvectors but also on the difference between the kth eigenvector and the other eigenvectors. Inasmuch as the covariance structure of the "reduced" contaminant $\tilde{\mathbf{x}}$ is in accordance with that in the population, its influence on the factor vector $\boldsymbol{\lambda}_k(F)$ is equal to zero.

If one creates on the basis of the influence functions $\mathbf{IF}(\boldsymbol{\lambda}_k)$ and $\mathbf{IF}(\boldsymbol{\lambda}_k^*)$ for the factor matrices $\boldsymbol{\Lambda}(F)$ and $\boldsymbol{\Lambda}^*(F)$ corresponding $p \times m$ "influence matrices", it is easily seen that the relation

$$[\mathbf{IF}(\boldsymbol{\lambda}_1^*), \dots, \mathbf{IF}(\boldsymbol{\lambda}_m^*)] = [\mathbf{IF}(\boldsymbol{\lambda}_1), \dots, \mathbf{IF}(\boldsymbol{\lambda}_m)] \cdot \mathbf{T}'$$

holds. The matrix of influence functions of the rotated factor loadings can be gained by applying the transformation matrix to the corresponding matrices of the unrotated factor loadings. In spite of this simple relation, the situation gets more entangled in case of the influence functions $\mathbf{IF}(\boldsymbol{\lambda}_k^*)$. In that case the influence of a contaminant $\tilde{\mathbf{x}}$ on the factor vector $\boldsymbol{\lambda}_k^*(F)$ is determined by the location of all eigenvectors which correspond to the m largest eigenvalues of the reduced covariance matrix $\widetilde{\boldsymbol{\Sigma}}(F)$ without taking the "correction" $\mathbf{g}_k$ into account.

3.2 Empirical versions of the influence function

3.2.1 Empirical influence functions

Usually F will be unknown, so that one has to refer to the empirical distribution function F_n given by a sample. Particular emphasis is therefore placed on the influence of an observation $\mathbf{x}_\mathrm{i}$ on the estimators $\boldsymbol{\lambda}_k^*(F_n)$ or $\boldsymbol{\lambda}_k(F_n)$. A first approach to this item is established by the *empirical influence function* $\widehat{\mathbf{IF}}(\mathbf{x};\mathbf{T}, F)$.

Definition 3.1 *The empirical influence function* $\hat{\mathbf{IF}}(\mathbf{x};\mathbf{T}, F)$ *results from expression (6) by replacing the distribution function* F *with the empirical distribution function* F_n*:*

$$\widehat{\mathbf{IF}}(\mathbf{x};\mathbf{T}, F) \equiv \mathbf{IF}(\mathbf{x};\mathbf{T}, F_n). \tag{24}$$

□

Thus, one can easily obtain the empirical influence functions of the estimators $\boldsymbol{\lambda}_k(F_n)$ and $\boldsymbol{\lambda}_k^*(F_n)$.

Corollary 3.1 *Let*

$$\hat{\mathrm{v}}_{\mathrm{k}} = \hat{\boldsymbol{\omega}}_k' \cdot \hat{\hat{\mathbf{x}}}\hat{\hat{\mathbf{x}}}' \cdot \hat{\boldsymbol{\omega}}_k - \hat{\tau}_k \tag{25}$$

$$\text{and} \qquad \hat{\mathbf{g}}_k = -\sum\nolimits_{(k)} (\hat{\tau}_h - \hat{\tau}_k)^{-1} \cdot \hat{\boldsymbol{\omega}}_h \hat{\boldsymbol{\omega}}_h' \cdot \hat{\hat{\mathbf{x}}}\hat{\hat{\mathbf{x}}}' \cdot \hat{\boldsymbol{\omega}}_k \tag{26}$$

$$\text{with} \qquad \hat{\hat{\mathbf{x}}} = \mathbf{x} - \boldsymbol{\mu}(F_n) - \hat{\mathbf{u}} \tag{27}$$

be estimators for the coefficients v_k *and* $\mathbf{g}_k$*, respectively. The empirical influence function of the unrotated factor vector* $\boldsymbol{\lambda}_k(F)$ *is then given by*

$$\widehat{\mathbf{IF}}(\mathbf{x}; \boldsymbol{\lambda}_k, F) = \hat{\tau}_k^{-1/2}(\frac{1}{2} \cdot \tau_k^{-1} \cdot \hat{\mathrm{v}}_{\mathrm{k}} \cdot \hat{\boldsymbol{\omega}}_k + \hat{\mathbf{g}}_k) \tag{28}$$

and the function of the rotated factor vector $\boldsymbol{\lambda}_k^*(F)$ *by*

$$\widehat{\mathbf{IF}}(\mathbf{x}; \boldsymbol{\lambda}_k^*, F) = [\hat{\tau}_1^{1/2}(\frac{1}{2}\hat{\tau}_1^{-1}\hat{\mathrm{v}}_1\hat{\boldsymbol{\omega}}_1 + \hat{\mathbf{g}}_1), \ldots, \hat{\tau}_m^{1/2}(\frac{1}{2}\hat{\tau}_m^{-1}\hat{\mathrm{v}}_{\mathrm{m}}\hat{\boldsymbol{\omega}}_m + \hat{\mathbf{g}}_m)] \cdot \mathbf{t}_k \ . \tag{29}$$

Proof. The equations (28) and (29) follow directly from theorem 3.2 in consideration of definition 3.1 with F_n rather than F. □

The values of the empirical influence function at the points $\mathbf{x} = \mathbf{x_i}$ are of particular interest:

$$\widehat{\mathbf{IF}}_i(\boldsymbol{\lambda}_k^*) \equiv \widehat{\mathbf{IF}}(\mathbf{x}_i; \boldsymbol{\lambda}_k^*, F) \ .$$

They approximately indicate the effect which observation $\mathbf{x_i}$ exerts on the estimator $\boldsymbol{\lambda}_k^*(F)$. More precisely, $\widehat{\mathbf{IF}}_i(\boldsymbol{\lambda}_k^*)$ reflects the rate of change for the loadings of the kth common factor in the p manifest variables which result from an addition of an observation vector $\mathbf{x_i}$ under $F = F_n$. The sample size here is not taken into account.

3.2.2 Deleted empirical influence functions

The effect of a deletion of the ith observation on the estimators $\boldsymbol{\lambda}_k(F_n)$ and $\boldsymbol{\lambda}_k^*(F_n)$ can be measured by means of the *deleted empirical influence function*. This pertains to a multitude of functions of which values are of interest at one point in particular: the reason is that an evaluation of the influence of the ith observation on the estimation the function value of the ith deleted influence function at the point $\mathbf{x} = \mathbf{x_i}$ is relevant. The fiction of a sample size growing over all limits is therein still retained.

Definition 3.2 *Let*

$$F_{(i)} \equiv [1 + (n-1)^{-1}] \cdot F_n - (n-1)^{-1} \cdot \delta_i \tag{30}$$

with

$$\delta_i = \begin{cases} 1 & \textit{if } \mathbf{x} = \mathbf{x_i} \\ 0 & \textit{otherwise} \end{cases}$$

be the empirical distribution function with an omission of the ith observation. The deleted empirical influence functions $\widehat{\mathbf{IF}}_{(i)}(\mathbf{x};\mathbf{T},F)$, $i=1,2,\ldots,n$, are then obtained by substitution of F with $F_{(i)}$ in equation (6), respectively:

$$\widehat{\mathbf{IF}}_{(i)}(\mathbf{x};\mathbf{T},F)\equiv\mathbf{IF}(\mathbf{x};\mathbf{T},F_{(i)})\ ,\quad i=1,2,\ldots,n\ . \tag{31}$$

□

Consequently the deleted empirical influence functions for the factor vector $\boldsymbol{\lambda}_k(F)$ take the form

$$\widehat{\mathbf{IF}}_{(i)}(\mathbf{x};\boldsymbol{\lambda}_k,F)=\hat{\tau}_{(i)k}^{1/2}\cdot(\frac{1}{2}\cdot\hat{\tau}_{(i)k}^{-1}\cdot\hat{\mathrm{v}}_{\mathrm{(i)k}}\cdot\hat{\boldsymbol{\omega}}_{(i)k}+\hat{\mathbf{g}}_{(i)k}) \tag{32}$$

for $i=1,2,\ldots,n$. Here, the subscript (i) indicates the deletion of the ith observation. $\hat{\mathrm{v}}_{\mathrm{(i)k}}$ and $\hat{\mathbf{g}}_{(i)k}$ can be determined by replacing $\hat{\tau}_{(i)k}$, $\hat{\boldsymbol{\omega}}_{(i)k}$, and $\hat{\boldsymbol{\mu}}_{(i)k}$ for τ_k, $\hat{\boldsymbol{\omega}}_k$, and $\boldsymbol{\mu}(F_n)$ in the equations (25) and (26) with

$$\hat{\boldsymbol{\mu}}_{(i)}\equiv\boldsymbol{\mu}(F_{(i)})=\boldsymbol{\mu}(F_n)-(n-1)^{-1}[\mathbf{x}_i-\boldsymbol{\mu}(F_n)]\ . \tag{33}$$

In principle the function value of the ith deleted empirical influence function $\widehat{\mathbf{IF}}_{(i)}(\boldsymbol{\lambda}_k)$,

$$\widehat{\mathbf{IF}}_{(i)}(\boldsymbol{\lambda}_k)\equiv\widehat{\mathbf{IF}}_{(i)}(\mathbf{x};\boldsymbol{\lambda}_k,F)\ ,$$

at the point $\mathbf{x}=\mathbf{x}_i$ can be determined by again solving the corresponding eigenvalue problem. But already at a moderate sample size a standard-like execution of this kind of influence analysis is scarcely practicable because of the computational expense. However, it can be shown that the function values of the deleted empirical influence functions $\widehat{\mathbf{IF}}_{(i)}(\boldsymbol{\lambda}_k)$ can approximately be computed from the statistics of the spectral decomposition of $\widetilde{\boldsymbol{\Sigma}}(F_n)$. Our derivations are based on the general approximation formulae which are given by Critchley (1985) for any functionals $\mathbf{T}(F_n)$.

Proposition 3.1 (see Critchley, 1985, p. 631)
Let

$$\eta=-\,(n-1)^{-1} \tag{34}$$

$$\textit{and}\qquad \mathbf{q}(\varepsilon)=\mathbf{T}[F_n+\varepsilon(\delta_i-F_n)]\ ,\quad j=1,2,\ldots,n \tag{35}$$

for all $\varepsilon\geq\eta$. It is assumed that there exist a continuous second derivative in $[\eta;0]$ and a third derivative in $(\eta;0)$ for $\mathbf{q}(.)$. Furthermore, it is assumed that $\mathbf{q}'(0)$, $\mathbf{q}''(0)$, and $\mathbf{q}'''(\varepsilon)$ for $\eta<\varepsilon<0$ do not depend on η. Then the relations

$$\widehat{\mathbf{IF}}(\mathbf{x}_i;\mathbf{T},F)=\mathbf{q}'(0)\ , \tag{36}$$

$$\widehat{\mathbf{IF}}_{(i)}(\mathbf{x};\mathbf{T},F)=\mathbf{q}'(0)+\eta[\mathbf{q}''(0)-\mathbf{q}'(0)]+O(\eta^2) \tag{37}$$

hold. □

While $\mathbf{q}'(0)$ indicates the rate of change of the functional $\mathbf{T}$ for F in direction of the point mass δ_i, $\mathbf{q}''(0)$ measures its curvature. $\widehat{\mathbf{IF}}_{(i)}(\mathbf{x}_i;\mathbf{T},F)$ is the slightly extended gradient of $\mathbf{q}$ at the point η.

Theorem 3.4 *The deleted empirical influence functions for the estimator* $\boldsymbol{\lambda}_k(F)$ *have the form*

$$\widehat{\mathbf{IF}}_{(i)}(\mathbf{x};\boldsymbol{\lambda}_k,F) = (1-\eta)\cdot\widehat{\mathbf{IF}}(\boldsymbol{\lambda}_k) + \eta\cdot\hat{\tau}^{1/2}\cdot\{\frac{1}{2}\hat{\tau}_k^{-1}\cdot \cdot[(\frac{1}{2}\cdot\hat{\mathrm{w}}_{\mathrm{k}} - \frac{1}{4}\cdot\hat{\mathrm{v}}_k^2\cdot\hat{\tau}_k^{-1})\cdot\hat{\boldsymbol{\omega}}_k + \hat{\mathrm{v}}_k\cdot\hat{\mathbf{g}}_k] + \hat{\mathbf{h}}_k\} + O(\eta^2) \quad (38)$$

and the function for the estimator $\boldsymbol{\lambda}_k^*(F_n)$ *is given by*

$$\widehat{\mathbf{IF}}_{(i)}(\mathbf{x};\boldsymbol{\lambda}_k^*,F) = (1-\eta)\cdot\widehat{\mathbf{IF}}(\boldsymbol{\lambda}_k^*) + \eta\cdot\hat{\tau}_k^{1/2}\cdot\{\frac{1}{2}\cdot\hat{\tau}_k^{-1}\cdot \cdot[(\frac{1}{2}\cdot\hat{\mathrm{w}}_k - \frac{1}{4}\cdot\hat{\mathrm{v}}_k^2\cdot\hat{\tau}_k^{-1})\cdot\hat{\boldsymbol{\omega}}_k + \hat{\mathrm{v}}_k\cdot\hat{\mathbf{g}}_k] + \hat{\mathbf{h}}_k\}\cdot\mathbf{t}_k + O(\eta^2) \quad (39)$$

for $i = 1, 2, \ldots, n$.
Proof. See Appendix. □

3.2.3 Sample influence functions

For judging the resistence of one-dimensional estimators, Tukey (1977) has developed an indicator measuring the influence of an additional observation. At first it is the sensitivity curve, which draws attention to the given sample and therefore removes the fiction of an infinite sample size. In empirical work of course it is disadvantageous to insinuate a new observation. Usually one wants to know the effect of a deletion of an observation on an estimator by explicit taking into account of the sample size. This effect can be measured by the *sample influence function*.

Definition 3.3 *The sample influence function* $\widetilde{\mathbf{IF}}(\mathbf{x}_i;\mathbf{T},F)$ *is defined by*

$$\widetilde{\mathbf{IF}}(\mathbf{x}_i;\mathbf{T},F) = (n-1)\cdot[\mathbf{T}(F_n) - \mathbf{T}(F_{(i)})] \quad (40)$$

with $F_{(i)}$ *according to equation (30).* □

$\widetilde{\mathbf{IF}}(\mathbf{x}_i;\mathbf{T},F)$ can be determined from the Taylor expansion of $\mathbf{q}(.)$ around 0.

Proposition 3.2 (see Critchley, 1985, p. 631)
Under the assumptions given in proposition 3.1

$$\widetilde{\mathbf{IF}}(\mathbf{x}_i;\mathbf{T},F) = \mathbf{q}'(0) + \frac{1}{2}\eta\mathbf{q}''(0) + O(\eta^2) \quad (41)$$

holds. □

The presentation

$$\widetilde{\mathbf{IF}}(\mathbf{x}_i;\mathbf{T},F) = \eta^{-1}[\mathbf{q}(\eta) - \mathbf{q}(0)]$$

suggests an interpretation of the sample influence function as the gradient of the line between the points $(0, \mathbf{q}(0))$ and $(\eta, \mathbf{q}(\eta))$. As a consequence of proposition 3.2, the sample influence function of the vector of loadings on the kth common factor, $\boldsymbol{\lambda}_k(F)$, can immediately be given as a function of the statistics of the eigenvalue problem of $\widetilde{\boldsymbol{\Sigma}}(F_n)$.

Theorem 3.5 *The sample influence function for the unrotated factor vector* $\boldsymbol{\lambda}_k(F)$ *is*

$$\widetilde{\mathbf{IF}}(\mathbf{x}_i;\boldsymbol{\lambda}_k,F) = \widehat{\mathbf{IF}}(\boldsymbol{\lambda}_k) + \frac{1}{2}\eta\hat{\tau}_k^{1/2}\{\hat{\tau}_k^{-1}[(\frac{1}{2}\hat{\mathrm{w}}_{\mathrm{k}} - \frac{1}{4}\hat{\mathrm{v}}_{\mathrm{k}}^2\hat{\tau}_k^{-1})\cdot\hat{\boldsymbol{\omega}}_k + \hat{\mathrm{v}}_{\mathrm{k}}\cdot\hat{\mathbf{g}}_k + \hat{\mathbf{h}}_k\} + O(\eta^2) \quad (42)$$

and that of the rotated factor vector

$$\widetilde{\mathbf{IF}}(\mathbf{x}_i;\boldsymbol{\lambda}_k^*,F) = \widetilde{\mathbf{IF}}(\boldsymbol{\lambda}_k^*) + \frac{1}{2}\eta\hat{\tau}_k^{1/2}\{\hat{\tau}_k^{-1}[(\frac{1}{2}\hat{\mathrm{w}}_{\mathrm{k}} - \frac{1}{4}\hat{\mathrm{v}}_{\mathrm{k}}^2\hat{\tau}_k^{-1})\cdot\hat{\boldsymbol{\omega}}_k + \hat{\mathrm{v}}_{\mathrm{k}}\cdot\hat{\mathbf{g}}_k + \hat{\mathbf{h}}_k\}\cdot\mathbf{t}_k + O(\eta^2)\ . \quad (43)$$

Proof. Equation (42) results from equation (41) with the aid of equation (28) for $\mathbf{q}'(0)$ and equation (64) for $\mathbf{q}''(0)$ after an arrangement and simplification of some terms. Analogously equation (43) can be shown with regard to the relation (23) between $\boldsymbol{\lambda}_k(F)$ and $\boldsymbol{\lambda}_k^*(F)$. □

The sample influence functions

$$\widetilde{\mathbf{IF}}_i(\boldsymbol{\lambda}_k) = \widetilde{\mathbf{IF}}(\mathbf{x}_i;\boldsymbol{\lambda}_k,F)\ ,$$
$$\widetilde{\mathbf{IF}}_i(\boldsymbol{\lambda}_k^*) = \widetilde{\mathbf{IF}}(\mathbf{x}_i;\boldsymbol{\lambda}_k^*,F)$$

are obtained by adding a "correction term" to the corresponding empirical influence functions. This term, which consists of estimators of quantities determining the eigenvalues and eigenvectors of the perturbed reduced covariance matrix $\widetilde{\boldsymbol{\Sigma}}(\widetilde{F})$, can be relevant for small and moderate sample sizes. With an increase in sample size the differences between the empirical influence function and the sample influence function will be diminished.

3.3 Reduced observations and norms of the influence functions

The reduced observation vectors $\tilde{\mathbf{x}}_i$ are given along with an estimation of the vectors $\mathbf{f_i}$, $i = 1, 2, \ldots, n$, of the factor values. Certainly an estimation of the factor values is not without difficulty inasmuch as $\mathbf{f}$ is not a parameter vector but a random vector. A feasible solution consists in regarding $\mathbf{f}$ as fixed at the estimation. According to their properties it is advisable to take the Barlett estimator

$$\mathbf{f}_B = [\boldsymbol{\Lambda}'(F)\cdot\boldsymbol{\Psi}(F)^{-1}\cdot\boldsymbol{\Lambda}(F)]^{-1}\cdot\boldsymbol{\Lambda}'(F)\cdot[\boldsymbol{\Psi}(F)]^{-1}\cdot[\mathbf{x}-\boldsymbol{\mu}(F)] \quad (44)$$

and the Thomson estimator

$$\mathbf{f}_T = \boldsymbol{\Lambda}'(F)\cdot[\boldsymbol{\Sigma}(F)]^{-1}\cdot[\mathbf{x}-\boldsymbol{\mu}(F)] \quad (45)$$

into consideration as estimators for the vectors $\mathbf{f}_i$. While $\mathbf{f}_B$ is a (conditional) blu-estimator, $\mathbf{f}_T$ is indeed biased, but its mean square error (MSE) is lower than that of the former estimator (see Lawley and Maxwell, 1971, pp. 106–111). Seber (1984, p. 221) has shown that $\mathbf{f}_T$ can be interpreted as a ridge estimator. In the past, preference was given to the Barlett estimator which is now to be viewed as an overvaluation of the blu property. In broad fields of statistics the MSE has turned out to be the better overall-criterion which should be also taken into account in factor analysis.

The substitution of equation (1) in equation (9) gives

$$\tilde{\mathbf{x}} = \mathbf{\Lambda}(F) \cdot \mathbf{f} \ , \tag{46}$$

so that the reduced observation vector of the ith unit can be estimated by using the estimator $\mathbf{\Lambda}(F_n)$ according to equation (7) and an estimator $\hat{\mathbf{f}}$ for $\mathbf{f}$ by

$$\hat{\tilde{\mathbf{x}}}_i = \mathbf{\Lambda}(F_n) \cdot \hat{\mathbf{f}}_i \ . \tag{47}$$

To assess the influence of a unit on the factor matrix $\mathbf{\Lambda}(F_n)$ one needs a measure which aggregates the information provided by the empirical version of the influence functions $\mathbf{IF}(\mathbf{x}; \boldsymbol{\lambda}_\mathbf{k}, \mathbf{F})$, $k = 1, 2, \ldots, m$. Let

$$\mathbf{if} = [\mathbf{IF}'(\boldsymbol{\lambda}_1), \mathbf{IF}'(\boldsymbol{\lambda}_2), \ldots, \mathbf{IF}'(\boldsymbol{\lambda}_m)]' \tag{48}$$

be the $pm \times 1$ vector of the components of the vector-valued influence functions of $\boldsymbol{\lambda}_1(F), \boldsymbol{\lambda}_2(F), \ldots, \boldsymbol{\lambda}_m(F)$. The mapping

$$\|\mathbf{if}\| : \mathbb{R}^{pm} \to \mathbb{R}$$

is a norm if the relations

$$\|\mathbf{if}\| \geq 0, \quad \|k \cdot \mathbf{if}\| = |k| \cdot \|\mathbf{if}\|, \quad \|\mathbf{if} + \mathbf{if}'\| \leq \|\mathbf{if}\| + \|\mathbf{if}'\|$$

hold for all $k \in \mathbb{R}$ and $\mathbf{if}, \mathbf{if}' \in \mathbb{R}^{pm}$. Cook and Weisberg (1982) differentiate outer norms, which can be based on a statistical model, and inner norms, which take no distributional assumptions. In the latter class especially the Hölder r norm is of importance:

$$\|\mathbf{if}\|_r = \Big[\sum_{l=1}^{pm} (|\mathrm{if}_l|^r) \Big]^{1/r} . \tag{49}$$

In this case, the quantity if_l is the lth component of the vector $\mathbf{if}$.

Notice that the Euclidean norm is included in equation (49) as the special case $r = 2$. For reasons of simplicity and clearness, we suggest taking the Euclidean norm into account for an influence diagnostic in factor analysis. In addition

$$\|\mathbf{if}\|_2 = \|\mathbf{if}^*\|_2 \tag{50}$$

holds when $\mathbf{if}^*$ is a $pm \times 1$ influence vector for the rotated factor loadings. This means that on the whole an influential observation changes the rotated factor matrix to the same extend as the unrotated factor matrix.

In empirical research all three empirical versions of the influence functions can be used for assessing the effect of an observation on the estimated factor matrix. In most applications, according to the nearness to data, the sample influence function will be the concept of choice.

4 An example

An example should illustrate the performance of the alternative empirical versions of the influence function for real data. Moreover, an intuition should be given regarding the application of influence diagnostics in factor analysis. For these purposes we use Kendall's data of a sample of 48 applicants for a position in a firm (Kendall, 1975, p. 33). Kendall extracts four factors on the basis of a correlation matrix of the 15 variables with eigenvalues that are greater than one. The same number of factors is indicated by the scree test when the extraction is accomplished on the basis of the covariance matrix of manifest variables. The factors "explain" successively 54.2%, 14.8%, 8.7%, and 5.4% of the variance of all variables. All together 83.1% of the variability of the manifest variables can be attributed to the four extracted factors.

Despite some inductively developed bounds for the factor loadings to be significant are known (see e.g. Cliff and Hamburger, 1967; Pennell, 1968; Jennrich, 1974), factor loadings are often interpreted as meaningful by a rule of thumb in empirical research. When using standardized data for a factor loading to be valued "practical" significant, it is often demanded, that its absolute value has to exceed a lower bound of 0.5 (see e.g. Ost, 1996, p. 682; Backhaus et al., 2000, p.292; for a more sophisticated rule of thumb see Boriz, 1999, pp. 534; Guadagnoli and Velicer, 1988). In this case at least 25% of the variance of a manifest variable can be "explained" by the corresponding common factor. This rule can be used for interpretation of the factor structure of a covariance-oriented factor analysis after an adequate rotation of the factor matrix.

The factor pattern of Kendall's data according to a varimax rotation of the factor matrix is shown in Table 4.1. To relieve the interpretation the significant factor loadings are underlined. Up to two cases the manifest variables can be attached quite clearly to a common factor.

Table 4.1: Factor pattern for applicants

	Factor 1	Factor 2	Factor 3	Factor 4
Application	0.261	0.281	<u>1.880</u>	0.350
Appearance	-0.222	0.552	0.333	0.848
Academic ability	<u>-1.369</u>	0.000	0.258	0.137
Likeability	0.198	<u>2.387</u>	0.655	0.636
Self-confidence	0.187	0.386	-0.205	<u>2.202</u>
Lucidity	-0.125	0.919	0.361	<u>2.660</u>
Honesty	0.036	<u>1.870</u>	-0.499	0.606
Salesmanship	0.262	0.273	0.813	<u>3.133</u>
Experience	-0.625	0.147	<u>2.649</u>	0.262
Drive	0.154	0.516	1.110	<u>2.249</u>
Ambition	0.182	0.314	0.525	<u>2.623</u>
Grasp	-0.518	1.076	0.809	<u>2.404</u>
Potential	-0.790	1.402	1.082	<u>2.326</u>
Keenness	<u>1.559</u>	<u>1.463</u>	1.004	1.122
Suitability	-0.309	0.211	<u>2.620</u>	1.211

The variable *keenness* shows high loadings on the first two factors, whereas the contributions of the single factors to the variance of the variable *appearance* remain lower than 20%. The unique variance of this variable has a portion of more than 70% of its whole variance. More than anywhere else, the variable *appearance* can be attached to the fourth factor. Only the variable *academic ability* continues to have a low proportion of explanation of slightly less than 50%. However, the first factor clearly dominates the factor pattern of this variable.

Kendall (1980, pp. 54) assigns the factor 1 (there: factor 4) entirely to the variable *academic ability* which has the absolutely highest loading in correlation-oriented factor analysis. Only the variable *keenness* has an absolutely high loading on this factor with a reversed sign yet. In our covariance-oriented factor analysis the last loading still exceeds the former, whereby the variables again are polarized oppositely. Kendall interprets factor 2 (there: factor 3) as a general likeability of a person and factor 3 (there: factor 2) as experience. Finally, factor 4 (there: factor 1) comprises a composition of properties and abilities such as self-confidence, salesmanship, ambition, and lucidity.

But how stable is the extracted factor pattern? In other words: Are there any observations which determine the factor pattern, so that it is changed when they are deleted? An answer can be found by calculating influence diagnostics as norms of the empirical versions of the influence function. In table 4.2 values of the Euclidean norm are shown for all 48 applicants for the empirical influence function, the empirical deleted influence function, and the sample influence function. In this application there is no substantial difference in the ranking of the observations according to their influence between the three concepts. The most striking observations concern the applicants 41 and 42 whose influence values exceed those of the other observations to a large extent. Indeed a deletion of one of both observations leads to a different interpretation of the factor pattern. Furthermore, without observation 41 the variable *keenness* does not load significantly on factor 2. On the contrary, this variable has then a substantial loading on factor 3. Similar to observation 41, a deletion of observation 42 will lead to the consequence that the variable *keenness* will tend to be attached to the third factor rather than to the second factor. In any case, this variable does not provide any more a significant loading on the factor 2.

Indeed no other observation has such a great influence on the personality structure of the applicants as the observations 41 and 42 which have been identified by the influence diagnostics. Not only the values of the factor matrix are influenced by the observations 41 and 42. Also the interpretation of the factor pattern is changed by a deletion of one of both observations. Apart from observation 11 which points to a possible assignment of the variable *potential* to the second factor, no other observation has such an influence on the analysis with respect to a change of the interpretable factor pattern.

In this way, the influence diagnostics can be used to validate the results of a factor analysis in empirical research. When the sample size is moderate or low, larger differences between the empirical versions of the influence function are to be expected.

Table 4.2: Influence measures for applicants

Person	$\|\mathbf{if}_e\|$	$\|\mathbf{if}_d\|$	$\|\mathbf{if}_s\|$
1	9.01752080	8.90236472	8.86168310
2	3.91660974	4.17023944	4.00016448
3	4.64850911	4.69944970	4.62333635
4	5.36386615	5.46496673	5.35712744
5	15.97921841	16.65358880	16.11617418
6	4.61516047	4.65779242	4.58736158
7	7.64978243	8.24710487	7.86506002
8	9.22277460	10.18268519	9.60196654
9	7.77424474	8.50405019	8.05322630
10	16.99680513	22.54297151	19.47433273
11	20.39719980	30.08333649	24.80641730
12	20.82963242	23.53698343	21.88327990
13	8.73925580	8.95439553	8.73751539
14	8.69861986	8.89441659	8.70259054
15	8.79636540	8.67573728	8.63998875
16	4.43136578	4.59474197	4.46479300
17	4.18098321	4.20901178	4.15045061
18	5.35278008	5.39261400	5.31532192
19	4.87314972	4.94839472	4.85873397
20	5.99510704	6.23934632	6.05112795
21	4.70846022	4.78229569	4.69502700
22	5.81831172	6.28302942	5.98748799
23	8.19024640	9.02776159	8.52024329
24	7.73118002	8.48602975	8.02343694
25	3.78387027	3.84362275	3.77283671
26	5.50063629	5.65203274	5.51722726
27	6.91593665	7.21016233	6.98842956
28	11.62005342	12.81004637	12.08863201
29	12.75039342	14.02028774	13.24460569
30	16.26874207	17.62254347	16.76572833
31	14.00224880	14.70068215	14.19604563
32	6.79137773	7.31827455	6.97897808
33	7.87015188	8.77650929	8.23265055
34	5.63877301	6.13381935	5.82494304
35	7.92889878	8.74091801	8.24897843
36	4.67152690	4.76445559	4.66776797
37	30.04156041	30.79973381	30.07212710
38	17.58436065	20.18379226	18.68982781
39	11.40873190	12.82078873	11.99007305
40	11.04688243	12.38294677	11.59423026
41	45.00916393	68.32826921	55.69592955
42	53.17996414	85.51808647	67.87832595
43	10.84341981	12.11134495	11.35640762
44	4.07364933	4.13316525	4.05956042
45	9.39377526	8.77275241	8.96586077
46	11.13586233	10.53836209	10.68087577
47	18.60603607	21.34515641	19.76783661
48	19.24155806	22.23416317	20.52138470

Because of its close proximity to the underlying data, the sample influence function is a particularly adequate tool as an influence measure in factor analysis. In empirical work it may be advantageous to use this concept graphically in form of a ‘case plot’.

Appendix

Proof of Theorem 3.1 When using the expectation operator E, the perturbed reduced covariance matrix $\widetilde{\mathbf{\Sigma}}(\widetilde{F})$ can be written in the form

$$\widetilde{\mathbf{\Sigma}}(\widetilde{F}) = \mathrm{E}_{\tilde{F}}(\tilde{\mathbf{x}}\tilde{\mathbf{x}}') - \{\mathrm{E}_{\tilde{F}}(\tilde{\mathbf{x}}) \cdot [\mathrm{E}_{\tilde{F}}(\tilde{\mathbf{x}})]'\} \ . \tag{51}$$

With regard to

$$\mathrm{E}_{\tilde{F}}(\tilde{\mathbf{x}}) = (1-\varepsilon) \cdot \mathrm{E}_F(\tilde{\mathbf{x}}) + \varepsilon \cdot \tilde{\mathbf{x}} \qquad \textit{and} \qquad \mathrm{E}_F(\tilde{\mathbf{x}}) = \mathbf{0} \ ,$$

the equation (51) can be simplified to

$$\widetilde{\mathbf{\Sigma}}(\widetilde{F}) = (1-\varepsilon) \cdot \mathrm{E}_F(\tilde{\mathbf{x}}\tilde{\mathbf{x}}') + \varepsilon \cdot \tilde{\mathbf{x}}\tilde{\mathbf{x}}' - \varepsilon^2 \cdot \tilde{\mathbf{x}}\tilde{\mathbf{x}}' \ . \tag{52}$$

Because of $\widetilde{\mathbf{\Sigma}}(F) = \mathrm{E}_F(\tilde{\mathbf{x}}\tilde{\mathbf{x}}')$ one can obtain the relation (8) from equation (52) after a suitable compilation. □

Lemma 4.1 *Let* $\mathbf{B}$*,* $\mathbf{C}$*, and* $\mathbf{D}$ *be symmetric* $p \times p$ *matrices forming the Taylor expansion*

$$\mathbf{B}(\varepsilon) = \mathbf{B} + \varepsilon \cdot \mathbf{C} + \frac{1}{2}\varepsilon^2 \cdot \mathbf{D} + O(\varepsilon^3) \tag{53}$$

for the perturbed matrix $\mathbf{B}(\varepsilon)$*. The perturbations of an eigenvalue* l *and the corresponding eigenvector* $\mathbf{e}$ *of* $\mathbf{B}$ *are described by the Taylor expansion*

$$l(\varepsilon) = l + \varepsilon \cdot \mathrm{v} + \frac{1}{2}\varepsilon^2 \cdot \mathrm{w} + O(\varepsilon^3) \ , \tag{54}$$

$$\mathbf{e}(\varepsilon) = \mathbf{e} + \varepsilon \cdot \mathbf{g} + \frac{1}{2}\varepsilon^2 \cdot \mathbf{h} + O(\varepsilon^3) \ . \tag{55}$$

Subsequently, the scalars v *and* w *are given by the relations*

$$\mathrm{v} = \mathbf{e}'\mathbf{C}\mathbf{e} \ , \tag{56}$$

$$\mathrm{w} = \mathbf{e}'(\mathbf{2}\mathbf{C}\mathbf{g} + \mathbf{D}\mathbf{e}). \tag{57}$$

For the $p \times 1$ *vectors* $\mathbf{g}$ *and* $\mathbf{h}$ *the relations*

$$\mathbf{g} = -(\mathbf{B} - l{\cdot}\mathbf{I})^{+}{\cdot}\mathbf{C}\mathbf{e} \ , \tag{58}$$

$$\mathbf{h} = (\mathbf{B} - l{\cdot}\mathbf{I})^{+}{\cdot}[2(\mathrm{v}{\cdot}\mathbf{I} - \mathbf{C}) \cdot \mathbf{g} + (\mathrm{w} \cdot \mathbf{I} - \mathbf{D}) \cdot \mathbf{e}] \tag{59}$$

hold.

Proof. The proofs of the expressions (56), (57), and (58) are already given by Sibson (1979, p. 219). In relation (57) w is merely written in a simpler form. Hence, only equation (59) remains to be proved.

Equating the coefficients of $\frac{1}{2}\varepsilon^2$ of the eigenvalue problem

$$\mathbf{B}(\varepsilon) \cdot \mathbf{e}(\varepsilon) = l(\varepsilon) \cdot \mathbf{e}(\varepsilon)$$

will lead to the relation

$$\mathbf{B} \cdot \mathbf{h} + 2 \cdot \mathbf{C} \cdot \mathbf{g} + \mathbf{D} \cdot \mathbf{e} = l \cdot \mathbf{h} + 2 \cdot \mathrm{v} \cdot \mathbf{g} + \mathrm{w} \cdot \mathbf{e}$$

with the aid of the equations (53), (54), and (55). From this equation the expression

$$(\mathbf{B} - l \cdot \mathbf{I}) \cdot \mathbf{h} = 2(\mathrm{v} \cdot \mathbf{I} - \mathbf{C}) \cdot \mathbf{g} + (\mathrm{w} \cdot \mathbf{I} - \mathbf{D}) \cdot \mathbf{e}$$

follows. With regard to

$$(\mathbf{B} - l \cdot \mathbf{I})^+ \cdot (\mathbf{B} - l \cdot \mathbf{I}) \cdot \mathbf{h} = \mathbf{h}$$

then one obtains then the expression (59). □

Proof of Theorem 3.2 According to equation (53) we view $\mathbf{B}(\varepsilon)$ as being equal to $\widetilde{\boldsymbol{\Sigma}}(\widetilde{F})$ as given in expression (8). With regard to equation (55), relation (56) can be written in the form

$$\mathrm{v_k} = \boldsymbol{\omega}_k' \cdot [\tilde{\mathbf{x}}\tilde{\mathbf{x}}' - \widetilde{\boldsymbol{\Sigma}}(F)] \cdot \boldsymbol{\omega}_k \tag{60}$$

$$\text{with} \qquad \mathbf{C} = \tilde{\mathbf{x}}\tilde{\mathbf{x}}' - \widetilde{\boldsymbol{\Sigma}}(F) \; . \tag{61}$$

Using the identity

$$\boldsymbol{\omega}_k' \cdot \boldsymbol{\Sigma}(F) \cdot \boldsymbol{\omega}_k = \tau_k$$

the expression (60) is immediately simplified to equation (16).

Due to equation (58) and expression (61), one obtains

$$\mathbf{g}_k = -[\widetilde{\boldsymbol{\Sigma}}(F) - \tau_k \cdot \mathbf{I}]^+ \cdot [\tilde{\mathbf{x}}\tilde{\mathbf{x}}' - \widetilde{\boldsymbol{\Sigma}}(F)] \cdot \boldsymbol{\omega}_k \; ,$$

which leads to the relation

$$\mathbf{g}_k = -\sum\nolimits_{(k)} (\tau_h - \tau_k)^{-1} \boldsymbol{\omega}_h {\boldsymbol{\omega}_h}' (\tilde{\mathbf{x}}\tilde{\mathbf{x}}' - \tau_k \cdot \mathbf{I}) \cdot \boldsymbol{\omega}_k$$

with respect to the spectral decompositions (12) and (13). Because ${\boldsymbol{\omega}_h}'\boldsymbol{\omega}_k = 0$ for $h \neq k$, equation (17) ensues.

For w_k the expression

$$\mathrm{w_k} = 2{\boldsymbol{\omega}_k}' \cdot \{[\tilde{\mathbf{x}}\tilde{\mathbf{x}}' - \widetilde{\boldsymbol{\Sigma}}(F)] \cdot \mathbf{g}_k - \tilde{\mathbf{x}}\tilde{\mathbf{x}}' \cdot \boldsymbol{\omega}_k\}$$

is derived from equation (57) by using expression (61) and

$$\mathbf{D} = -2\tilde{\mathbf{x}}\tilde{\mathbf{x}}' \; . \tag{62}$$

In consideration of equation (12) the relation (18) holds.

Finally, $\mathbf{h}_\mathrm{k}$ according to equation (19) is obtained from equation (59) with regard to the spectral decompositions (12) and (13) by using the expressions (61) and (62). □

Proof of Theorem 3.4 The vector-valued function $\mathbf{q}(\varepsilon)$ can be written as

$$\mathbf{q}(\varepsilon) = \hat{\boldsymbol{\lambda}}_k(\widetilde{F}) = [\hat{\tau}_k(\widetilde{F})]^{1/2} \hat{\boldsymbol{\omega}}_k(\widetilde{F}) \tag{63}$$

for $\mathbf{T}(F) = \boldsymbol{\lambda}_k(F)$ with regard to equation (7). Substitution of the Taylor expansions (54) and (55) for the estimators of the eigenvalues and eigenvectors

τ_k and $\boldsymbol{\omega}_k$ as well as for the coefficients $\mathrm{v_k}$, $\mathrm{w_k}$, $\mathbf{g}_k$, and $\mathbf{h}_k$, gives an analogous presentation for the perturbed eigenvalues and eigenvectors of $\widetilde{\boldsymbol{\Sigma}}(F_n)$, so that equation (63) can be transformed into the relation

$$\mathbf{q}(\varepsilon) = \hat{\tau}_k^{1/2} \cdot [1 + \frac{1}{2}\varepsilon \hat{\mathrm{v}}_\mathrm{k} \hat{\tau}_k^{-1} + \frac{1}{2}\varepsilon^2 \hat{\tau}_k^{-1}(\hat{\mathrm{w}}_\mathrm{k} - \frac{1}{2}\hat{\mathrm{v}}_\mathrm{k}^2 \hat{\tau}_k^{-1})] \cdot (\hat{\boldsymbol{\omega}}_k + \varepsilon \hat{\mathbf{g}}_k + \frac{1}{2}\varepsilon^2 \hat{\mathbf{h}}_k) + O(\varepsilon^2) \ .$$

From this we can obtain the derivatives

$$\mathbf{q}'(\varepsilon) = \hat{\tau}_k^{1/2}\{\frac{1}{2}\hat{\tau}_k^{-1}[\hat{\mathrm{v}}_\mathrm{k} + \varepsilon(\hat{\mathrm{w}}_\mathrm{k} - \frac{1}{2}\hat{\mathrm{v}}_\mathrm{k}^2\hat{\tau}_k^{-1})] \cdot \hat{\boldsymbol{\omega}}_k + (1 + \varepsilon \hat{\mathrm{v}}_\mathrm{k}\hat{\tau}_k^{-1})\hat{\mathbf{g}}_k + \varepsilon \hat{\mathbf{h}}_k\} + O(\varepsilon^3)$$

$$\mathbf{q}''(\varepsilon) = \hat{\tau}_k^{1/2}[\frac{1}{2}\tau_k^{-1}(\hat{\mathrm{w}}_\mathrm{k} - \frac{1}{2}\hat{\mathrm{v}}_\mathrm{k}^2\hat{\tau}_k^{-1})\hat{\boldsymbol{\omega}}_k + \hat{\mathrm{v}}_\mathrm{k}\hat{\tau}_k^{-1}\hat{\mathbf{g}}_k + \hat{\mathbf{h}}_k] + O(\varepsilon^2) \ .$$

If one now substitutes $\widehat{\mathbf{IF}}(\mathbf{x};\boldsymbol{\lambda}_k, F)$ given by equation (28) for $\mathbf{q}'(0)$ [because of equation (36)] into equation (37), with regard to

$$\mathbf{q}''(0) = \hat{\tau}_k^{1/2}[\frac{1}{2}\hat{\tau}_k^{-1}(\hat{\mathrm{w}}_\mathrm{k} - \frac{1}{2}\hat{\mathrm{v}}_\mathrm{k}^2\tau_k^{-1})\hat{\boldsymbol{\omega}}_k + \hat{\mathrm{v}}_\mathrm{k}\hat{\tau}_k^{-1}\hat{\mathbf{g}}_k + \hat{\mathbf{h}}_k] \ , \tag{64}$$

equation (39) follows after an appropriate arrangement and simplification of the terms. □

Acknowledgement

I like to thank the two anonymous referees for their helpful comments.

References

Backhaus, K., Erichson, B., Plinke, V., and Weber, R. (2000), Multivariate Analysemethoden, 9. Aufl., Springer Verlag, Berlin.

Belsley, D.A., Kuh, E., and Welch, R.E. (1980), Regression diagnostics: identifying influential data and sources of collinearity, Wiley & Sons, New York.

Bortz, J. (1999), Statistik für Sozialwissenschaftler, 5. Aufl., Springer Verlag, Berlin.

Bronstein, I.N. and Semendjajew, A. (1985), Taschenbuch der Mathematik, 22. Aufl., Verlag Harri Deutsch, Thun, Frankfurt a. M.

Cliff, N. and Hamburger, C.D. (1967), The study of sampling errors in factor analysis by means of artificial experiments, Psychological Bulletin 68, 605–620.

Cook, R.D. and Weisberg, S. (1982), Residuals and influence in regression, Chapman Hall, London.

Costaño–Tostado, E. and Tanaka, Y. (1991), Sensitivity measures of influence on the loading matrix in exploratory factor analysis, Communications in Statistics-Theory and Methods 20, 1329–1943.

Critchley, F. (1985), Influence in principal components, Biometrika 72, 627–636.

Cureton, E.E. and D'Agostino, R.B. (1983), Factor analysis: an applied approach, Erlbaum, London.

Fahrmeir, L., Hamerle, A. and Tutz, G. (eds.) (1996), Multivariate statistische Verfahren, 2nd ed., de Gruyter, Berlin, New York.

Guadagnoli, E. and Velicer, W.E. (1988) Relation of sample size to the stability of component patterns, Psychological Bulletin 103, 265–275.

Hampel, F.R. (1974), The influence curve and its role in robust estimation, Journal of the American Statistical Association 69, 383–393.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986), Robust statistics. The approach based on influence functions, Wiley & Sons, New York.

Harman, H.H. (1976), Modern factor analysis, University of Chicago Press, Chicago.

Huber, P. (1981), Robust statistics, Wiley & Sons, New York.

Jennrich, R.I. (1974), Simplified formulae for standard errors in maximum likelihood factor analysis, Psychometrika 38, 571–580.

Kendall, M. (1975), Multivariate analysis, Griffin, London.

Kosfeld, R. (1986), Theoretische und numerische Aspekte in der Maximum-Likelihood-Faktorenanalyse, Campus Verlag, Frankfurt a. M.

Lawley, D.N. and Maxwell, A.E. (1971), Factor analysis as a statistical method, Butterworth, London.

Ost, F. (1996), Faktorenanalyse, in: Fahrmeir, L., Hamerle, A. and Tutz, G. (eds.), Multivariate statistische Verfahren, de Gruyter, Berlin, New York.

Pennell, R. (1968), The influence of communality and n on the sampling distributions of factor loadings, Psychometrika 33, 423–440.

Seber, G.A.F. (1984), Multivariate Observations, Wiley & Sons, New York.

Sibson, R. (1979), Studies in the robustness of multidimensional scaling: perturbational analysis of classical scaling, Journal of the Royal Statistical Society, Ser. B 41, 217–229.

Tanaka, Y. and Odaka, Y. (1989a), Influential observations in principal factor analysis, Psychometrika, 54, 475–485.

Tanaka, Y. and Odaka, Y. (1989b), Sensitivity analysis in maximum likelihood factor analysis, Communications in Statistics-Theory and Methods 18, 4067–4084.

Tukey, J.W. (1977), Exploratory Data Analysis, Addison-Wesley, Reading, Mass.

Wilkinson, J.H. (1965), The algebraic eigenvalue problem, Clarendon Press, Oxford.

Reinhold Kosfeld
Fachbereich Wirtschaftswissenschaften
Universität Kassel
D-34109 Kassel