Factor analysis with a single common factor

Chen, Siyan and Desiderio, Saul

2018

Factor analysis with a single common factor

Siyan Chen[a] Saul Desiderio[a, b]

Abstract

In this paper we present a simple approach to factor analysis to estimate the true correlations between observable variables and a single common factor. We first provide the exact formula for the correlations under the orthogonality conditions, and then we show how to consistently estimate them using a random sample and a proper instrumental variable.

Keywords : Factor analysis, correlation, instrumental variable estimation

1 Introduction

Factor analysis (FA) is a technique employed in multivariate statistical analysis[1]. Its purpose is to determine to what extent a set of k observable variables y depends on a smaller set of $p < k$ underlying (unobservable) common variables x, called "common factors" or, more simply, "factors". Each observed variable y_j is supposed to be modeled as a linear combination of the unobserved factors (generally assumed to be orthogonal with each other) plus an idiosyncratic term u_j:

$$y_j = \alpha_{j1}x_1 + \alpha_{j2}x_2 + ... + \alpha_{jp}x_p + u_j, \tag{1}$$

where the α's are unknown coefficients called "loadings".

In general, the researcher's main interest is to determine the correlation coefficients between y_j and each factor x. To achieve this goal, factors and loadings need to be estimated first. However, estimation of both factors and loadings involves a high degree of arbitrariness, especially in the decomposition of the correlation matrix. In addition, the solution (i.e. the set of estimated factors and loadings) is unique only up to an orthogonal transformation. What the researcher ends up with, therefore, is not clear.

The aim of this paper is to show how to consistently estimate the *true* correlations for the case of a single common factor when the orthogonality conditions hold. The idea is to directly estimate the correlation coefficients between the observed variables and the factor without going through the intermediate phase of estimating the factor. As we will show, the solution is unique.

The limitation of being confined to the case of a single common factor does not prevent our approach from having useful practical applications. If in fact we restrict our attention to the case of two observable variables, it is easy to

[a]Business School, Shantou University, Daxue Road 243, Shantou, Guangdong, P. R. China. Tel.:+86 754 86502786; fax:+86 754 86503442

[b]Corresponding author. E-mails: saul@stu.edu.cn, saul1979@libero.it

[1]On what follows we refer to any introductory textbook of factor analysis.

verify that under the usual orthogonality conditions only one common factor is admissible.

The remainder of the paper is organized as follows. In Section 2 the population model is described, with Theorem 1 giving the exact formula for the correlations. Section 3 shows how to consistently estimate the correlations using a random sample and an instrumental variable. Section 4 concludes.

2 The population model

Let us consider two correlated variables y_1 and y_2, both depending on a common factor x and on idiosyncratic terms[2]:

$$\begin{cases} y_1 = \alpha_1 x + u_1 \\ y_2 = \alpha_2 x + u_2. \end{cases} \tag{2}$$

Loading α_1 can always be normalized to 1 without loss of generality. In fact, by defining $\alpha_1 x = x'$ we get

$$y_1 = x' + u_1 \tag{3}$$

and

$$y_2 = \frac{\alpha_2}{\alpha_1} x' + u_2 \equiv \beta x' + u_2. \tag{4}$$

Hence, system (2) can always be written as

$$\begin{cases} y_1 = x + u_1 \\ y_2 = \beta x + u_2, \end{cases} \tag{5}$$

where β has now a relative interpretation.

Following FA literature, the orthogonality conditions for our system are

$$\mathrm{E}(u_1) = \mathrm{E}(u_2) = 0;\ \mathrm{Cov}(u_1, u_2) = 0;\ \mathrm{Cov}(u_i, x) = 0 \ \ \forall i = 1, 2. \tag{6}$$

Thus, given above setup, we can now state the following:

Theorem 1 *Given System (5) and the orthogonality conditions (6), then*

$$Corr(y_1, x) = \sqrt{\frac{Cov(y_1, y_2)}{\beta\, Var(y_1)}}.$$

Proof. As now $\mathrm{Cov}(y_1, y_2) = \mathrm{Cov}(y_2, x) = \beta \mathrm{Var}(x)$ and $\mathrm{Cov}(y_1, x) = \mathrm{Var}(x)$, we can define

$$\mathrm{Corr}(y_1, x) = \frac{\mathrm{Var}(x)}{\sqrt{\mathrm{Var}(y_1)}\sqrt{\mathrm{Var}(x)}} = \frac{\mathrm{std}(x)}{\mathrm{std}(y_1)}, \tag{7}$$

$$\mathrm{Corr}(y_2, x) = \frac{\beta \mathrm{Var}(x)}{\sqrt{\mathrm{Var}(y_2)}\sqrt{\mathrm{Var}(x)}} = \frac{\beta \mathrm{std}(x)}{\mathrm{std}(y_2)} \tag{8}$$

[2] What follows can be straightforwardly extended to any number of covariates y with a single factor in common.

and

$$\mathrm{Corr}(y_1, y_2) = \frac{\beta \mathrm{Var}(x)}{\sqrt{\mathrm{Var}(y_1)}\sqrt{\mathrm{Var}(y_2)}} = \mathrm{Corr}(y_1, x)\,\mathrm{Corr}(y_2, x). \tag{9}$$

As from Eqs. (7) and (8) we can write

$$\mathrm{Corr}(y_2, x) = \frac{\beta \mathrm{std}(x)}{\mathrm{std}(y_2)} = \frac{\beta \mathrm{std}(x)}{\mathrm{std}(y_2)} \frac{\mathrm{std}(y_1)}{\mathrm{std}(y_1)} =$$

$$= \beta \mathrm{Corr}(y_1, x) \frac{\mathrm{std}(y_1)}{\mathrm{std}(y_2)}, \tag{10}$$

equation (9) becomes

$$\mathrm{Corr}(y_1, y_2) = \beta \mathrm{Corr}(y_1, x) \mathrm{Corr}(y_1, x) \frac{\mathrm{std}(y_1)}{\mathrm{std}(y_2)}, \tag{11}$$

whereby

$$\mathrm{Corr}(y_1, x) = \pm\sqrt{\frac{\mathrm{std}(y_2)\mathrm{Corr}(y_1, y_2)}{\beta \mathrm{std}(y_1)}} = \pm\sqrt{\frac{\mathrm{Cov}(y_1, y_2)}{\beta \mathrm{Var}(y_1)}}, \tag{12}$$

which must be taken with positive sign because the covariance between x and y_1 is by construction equal to $\mathrm{Var}(x)$. ■

3 Estimation

All the population moments and parameter β need to be estimated. If we assume to have available a random sample for both y_1 and y_2, then we can use it to consistently estimate the population moments by their corresponding sample moments[3]. Hence we have

$$\mathrm{plim}\, \mathrm{SCov}(y_1, y_2) = \mathrm{Cov}(y_1, y_2),\ \mathrm{plim}\, \mathrm{SVar}(y_1) = \mathrm{Var}(y_1). \tag{13}$$

To estimate β we can substitute y_1 for x in the definition of y_2 to obtain:

$$y_2 = \beta y_1 + (u_2 - \beta u_1) \tag{14}$$

and estimate Eq. (14) by OLS to get $\hat{\beta}$. However, this estimator suffers from attenuation bias. In fact, as by assumption $\mathrm{E}(u_1) = 0$, we have

$$\mathrm{Cov}(y_1, u_2 - \beta u_1) = -\beta \mathrm{Var}(u_1), \tag{15}$$

and therefore

$$\mathrm{plim}\, \hat{\beta} = \beta + \frac{\mathrm{Cov}(y_1, u_2 - \beta u_1)}{\mathrm{Var}(y_1)} =$$

$$= \beta \frac{\mathrm{Var}(x)}{\mathrm{Var}(y_1)} \tag{16}$$

as $\mathrm{Var}(y_1) - \mathrm{Var}(u_1) = \mathrm{Var}(x)$.

[3]By SM we will indicate the sample counterpart of population moment M.

To overcome this problem we can resort to an instrumental variable estimation by finding another variable which is correlated with y_1 but not with the idiosyncratic errors u_1 and u_2. A natural candidate would be another variable y_3 sharing the common factor x:

$$y_3 = \gamma x + u_3. \tag{17}$$

Under the usual orthogonality assumptions already postulated for y_1 and y_2, y_3 is a valid instrument for the endogenous regressor y_1 in Eq. (14)[4]. Thus, given a consistent IV estimator $\tilde{\beta}_{IV}$, correlation (12) is consistently estimated by

$$\text{SCorr}(y_1, x) = \sqrt{\frac{\text{SCov}(y_1, y_2)}{\tilde{\beta}_{IV}\text{SVar}(y_1)}}. \tag{18}$$

Finally, using Eqs. (9) and (18) the correlation between y_2 and x can be estimated by

$$\text{SCorr}(y_2, x) = \text{SCorr}(y_1, y_2)/\text{SCorr}(y_1, x). \tag{19}$$

4 Conclusive remarks

Factor analysis is a statistical technique characterized by many issues. In particular, solutions are not unique and estimation procedures involve a high degree of arbitrariness by the researcher. In this paper we provided a simpler approach based on typical orthogonality assumptions that can be applied to the case of a single common factor. Basically, it consists in estimating directly the correlations between observable variables and the common factor skipping the intermediate step of factor estimation. We first describe the population model and derive the exact formula of the correlations, and then we provide estimators to consistently estimate them using a random sample. Although limited to the case of a single common factor, our approach solves the above-mentioned issues characterizing factor analysis.

[4]In the case of n covariates y, one of them must be sacrificed as instrumental variable.