Journal of Multivariate Analysis 150 (2016) 125–151

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Exact and asymptotic tests on a factor model in low and large dimensions with applications

Taras Bodnar [a,*], Markus Reiß [b]

[a] Department of Mathematics, Stockholm University, Roslagsvägen 101, SE-10691 Stockholm, Sweden
[b] Department of Mathematics, Humboldt-Universität of Berlin, Unter den Linden 6, D-10099 Berlin, Germany

ARTICLE INFO

Article history:
Received 22 January 2015
Available online 2 June 2016

JEL classification:
C12
C38
C52
C55
C58
C65
G12

AMS 2010 subject classifications:
91G70
62H25
62H15
62H10
62E15
62E20
60B20

Keywords:
Factor model
Exact test
Asymptotic test
High-dimensional test
High-dimensional asymptotics
Precision matrix
Inverse Wishart distribution
Random matrix theory

ABSTRACT

In the paper, we suggest three tests on the validity of a factor model which can be applied for both, small-dimensional and large-dimensional data. The exact and asymptotic distributions of the resulting test statistics are derived under classical and high-dimensional asymptotic regimes. It is shown that the critical values of the proposed tests can be calibrated empirically by generating a sample from the inverse Wishart distribution with identity parameter matrix. The powers of the suggested tests are investigated by means of simulations. The results of the simulation study are consistent with the theoretical findings and provide general recommendations about the application of each of the three tests. Finally, the theoretical results are applied to two real data sets, which consist of returns on stocks from the DAX index and on stocks from the S&P 500 index. Our empirical results do not support the hypothesis that all linear dependencies between the returns can be entirely captured by the factors considered in the paper.

1. Introduction

Factor models are widely spread in different fields of science, especially, in economics and finance where this type of models have been increasing in popularity recently. They are often used in forecasting mean and variance (see, e.g., Stock and Watson, [80,81], Marcellino et al. [63], Artis et al. [6], Boivin and Ng [25], Anderson and Vahid [5] and the references

* Corresponding author.
E-mail address: taras.bodnar@math.su.se (T. Bodnar).

http://dx.doi.org/10.1016/j.jmva.2016.05.011

therein), in macroeconomic analysis (see, Bernanke and Boivin [18], Favero et al. [46], Giannone et al. [48]), in portfolio theory (see, Ross [72,73], Engle and Watson [38], Chamberlain [31], Chamberlain and Rothschild [32], Diebold and Nerlove [36], Fama and French [40,41], Aguilar and West [3], Bai [7], Ledoit and Wolf [59]). Factor models are also popular in physics, psychology, biology (e.g., Rubin and Thayer [74], Carvalho et al. [29]) as well as in multiple testing theory (e.g., Friguet et al. [47], Dickhaus [35], Fan et al. [43]).

Another stream of research related to factor models deals with the estimation of high-dimensional covariance and precision matrices. This approach is motivated by a rapid development of high-dimensional factor models during the last years [10,11,9,8]. Fan et al. [42], Fan et al. [45], Fan et al. [44] among others have suggested several methods for estimating the covariance and precision matrices based on factor models in high dimensions and applied their results to portfolio theory, whereas Ledoit and Wolf [59] have proposed to combine the sample covariance matrix with the single-factor model based estimator in order to improve the estimate of the covariance matrix. Here, they use the capital asset pricing model (CAPM) as a single-factor model. Ross [72,73] argues that the empirical success of the CAPM can be explained by the validity of the following three assumptions: (i) there are many assets; (ii) the market permits no arbitrage opportunity; (iii) asset returns have a factor structure with a small number of factors. He also presents a heuristic argument that if an infinite number of assets are present on the market, then it is possible to construct sufficiently many riskless portfolios. In Chamberlain [31], conditions are derived under which this heuristic argument of Ross is justified. Furthermore, Chamberlain and Rothschild [32] suggest the so-called approximate K-factor structure model where the number of assets is assumed to be infinite, while Fan et al. [42] and Li et al. [60] extend this model by considering an asymptotically infinite number of known and unknown factors, respectively.

Let X_{it} be the observation data for the ith cross-section unit at time t. For instance, in the case of portfolio theory, X_{it} represents the return of the ith asset at time t. Let $\mathbf{X}_t = (X_{1t}, \ldots, X_{pt})^\top$ be the observation vector at time t and let $\mathbf{f}_t$ be a K-dimensional vector of common observable factors at time t. Then the factor model in vector form is expressed as

$$\mathbf{X}_t = \mathbf{B}\mathbf{f}_t + \mathbf{u}_t \tag{1}$$

where $\mathbf{B}$ is the matrix of factor loadings and $\mathbf{u}_t$, $t = 1, \ldots, T$, are independent errors with covariance matrix $\boldsymbol{\Sigma}_u$. It is also assumed that $\mathbf{f}_t$ are independent in time as well as independent of $\mathbf{u}_t$. The estimation of the factor model or the covariance matrix resulting from the factor model with observable factors is considered by Fan et al. [42], whereas Bai [7], Bai and Li [9], Fan et al. [44] present the results under the assumption that the factors are unobservable. Moreover, Bai and Ng [10], Hallin and Liška [53], Kapetanios [58], Onatski [68], Ahn and Horenstein [4] among others deal with the problem of determining the number of factors K used in (1). Note that not in all models the factors are observable. For example, this is not the case in many applications in psychology or in multiple testing theory, and, consequently, the results derived in the present paper cannot be directly applied. On the other hand, factor models with observable variables are usually considered in economics and finance where we also provide two empirical illustrations of the obtained theoretical results.

Under the generic assumption that $\boldsymbol{\Sigma}_u$ is a diagonal matrix, the dependence between the elements of $\mathbf{X}_t$ is fully determined by the factors $\mathbf{f}_t$. This means that the precision matrix of $\mathbf{Y}_t = (\mathbf{X}_t^\top, \mathbf{f}_t^\top)^\top$ has the following structure

$$\boldsymbol{\Omega} = \{\mathrm{cov}(\mathbf{Y}_t)\}^{-1} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}, \tag{2}$$

where $\boldsymbol{\Omega}_{21} = \boldsymbol{\Omega}_{12}^\top$ is a $p \times K$ matrix and $\boldsymbol{\Omega}_{11}$ is a diagonal $p \times p$ matrix, if the factor model (1) is true, i.e., if all linear dependencies among the components of $\mathbf{X}_t$ are fully captured by the factor vector $\mathbf{f}_t$. As a result, the test on the validity of the factor model (1) is equivalent to testing

$$H_0 : \boldsymbol{\Omega}_{11} = \mathrm{diag}(\omega_{11}, \ldots, \omega_{pp}) \quad \text{versus} \quad H_1 : \boldsymbol{\Omega}_{11} \neq \mathrm{diag}(\omega_{11}, \ldots, \omega_{pp}) \tag{3}$$

for some positive constants $\omega_{11}, \ldots, \omega_{pp}$.

We contribute to the existing literature on factor models by deriving exact and asymptotic tests on the validity of the factor model which are based on testing (3). Furthermore, the distributions of the suggested test statistics are obtained under both hypotheses and also they are analyzed in detail when the dimension of the factor model tends to infinity as the sample size increases such that $p/(T-K) \longrightarrow c \in (0, 1]$. This asymptotic regime is known in the statistical literature as double asymptotic regime or high-dimensional asymptotics.

Alternatively to the test (3), one can apply the classical goodness-of-fit test which is based on the estimated residuals given by

$$\hat{\mathbf{u}}_t = \mathbf{X}_t - \widehat{\mathbf{B}}\mathbf{f}_t,$$

where $\widehat{\mathbf{B}}$ is an estimate of the factor loading matrix. This approach, however, does not always lead to reliable results. To see this, let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_T)$, $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_T)$, and $\widehat{\mathbf{U}} = (\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_T)$. If $\mathbf{B}$ is estimated by applying the least square method, i.e., $\widehat{\mathbf{B}} = \mathbf{X}\mathbf{F}^\top(\mathbf{F}\mathbf{F}^\top)^{-1}$, then

$$\widehat{\mathbf{U}} = \mathbf{X} - \widehat{\mathbf{B}}\mathbf{F} = \mathbf{X}(\mathbf{I}_T - \mathbf{F}^\top(\mathbf{F}\mathbf{F}^\top)^{-1}\mathbf{F}),$$

where $\mathbf{I}_T$ is the T-dimensional identity matrix. Under the assumption of normality it holds that $\mathbf{U}|\mathbf{F} \sim \mathcal{N}_{p,n}(\mathbf{0}, \boldsymbol{\Sigma}_u \otimes \mathbf{I}_n)$ ($p \times n$ dimensional matrix variate normal distribution with zero mean matrix and covariance matrix $\boldsymbol{\Sigma}_u \otimes \mathbf{I}_n$) and, consequently,