Economics Letters 169 (2018) 63–67

Contents lists available at ScienceDirect

Economics Letters

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

Eigenvalue difference test for the number of common factors in the approximate factor models

Jianhong Wu

College of Mathematics and Science, Shanghai Normal University, Shanghai 200234, China

HIGHLIGHTS

- A eigenvalue difference method is proposed for determining the number of factors.
- The new determination method has more desired finite sample properties.
- Under some mild conditions, the resulting estimator can be proved to be consistent.

ARTICLE INFO

Article history:
Received 14 December 2017
Received in revised form 11 April 2018
Accepted 10 May 2018
Available online 22 May 2018

JEL classification:
C1
C13

Keywords:
Approximate factor model
Dominate factors
Eigenvalue difference test
Number of common factors

ABSTRACT

This paper proposes a new method for determining the number of common factors in the approximate factor models. Firstly, we construct a nonlinear and monotonous function of eigenvalues such that the function values of the first r largest eigenvalues are close to one and the rest are close to zero when both the number of cross-section units (N) and time series length (T) go to infinity, where r is the real value of the number of common factors. Secondly, we obtain the estimator of the number of common factors by maximizing the difference of function values of two adjacent eigenvalues arranged in descending order. Under some mild conditions, the resulting estimator can be proved to be consistent. Monte Carlo simulation study shows that the new estimator has desired performance.

1. Introduction

In the literature, there are various methods on the determination of the number of common factors in the approximate factor models proposed by Bai and Ng (2002, 2007), Hallin and Liska (2007), Pan and Yao (2008), Bathia et al. (2010), Onatski (2010), Alessi et al. (2010) Lam and Yao (2012), Ahn and Horenstein (2013), Caner and Han (2014), Xia et al. (2015), Wu (2016), etc. These methods can be proved to be consistent under certain conditions, however, each of them has its own restrictions and scope of applications. The considered models in the above literature include the static factor models and the dynamic factor models. For the sake of simplicity, this paper focuses mainly on the former and postpones the latter in Section 4 for discussions. For the static factor models, Ahn and Horenstein (2013) compared several methods by Monte Carlo simulation experiments and gave a detailed reviews for various determination methods of the number of common factors. Furthermore, Ahn and Horenstein (2013) proposed two estimators based on the ratio of eigenvalues which had desired properties, especially in the case with cross-sectional dependence. Lam and Yao (2012) also proposed an estimator for the number of factors based on the ratio of eigenvalues in modeling high-dimensional time series. Xia et al. (2015) argued that the ratio-type estimators may suffer from the instability of the 0/0-type ratio values and further proposed a modified ratio-type estimator by adding a so-called ridge parameter to avoid the possible instability issue. Along the same line of the above methods, Wu (2016) proposed a modified determination method for the number of common factors based on the ratio of transformation function of two adjacent eigenvalues arranged in descending order. The simulation results showed that the method of Wu (2016) had better performance than the competitors, especially in the case with dominant factors. Although the mentioned-above ratio-type estimators have been verified to have better finite sample properties than the competitors in the literature, some issues such as the instability of ratio values and the choice of ridge parameter still exist (Xia et al., 2015). These induce us to propose a new estimator as an alternative.

E-mail address: wjhstat1@163.com.

https://doi.org/10.1016/j.econlet.2018.05.009

This paper will propose a so-called difference-type estimator for the number of common factors in the static approximate factor models. Firstly, we construct a nonlinear and monotonous function of eigenvalues such that the function values of the first r largest eigenvalues are close to one and the rest are close to zero when both the number of cross-section units (N) and time series observations (T) go to infinity, where r is the real value of the number of common factors. Secondly, we obtain the estimator of the number of common factors by maximizing the difference of function values of two adjacent eigenvalues arranged in descending order. Under some mild conditions, the resulting estimator can be proved to be consistent. It can be further shown that, compared with the competitors in the existing literature, the new method enjoys some desired properties as follows: it is available for more general factor structures including the case with dominant factors; it does not need the pre-specification of the possible maximum number of factors (e.g. Bai and Ng, 2002; Ahn and Horenstein, 2013); it does not need the ridge parameter used in some methods (e.g. Xia et al., 2015; Wu, 2016). Monte Carlo simulation study shows that the new estimator has better performance than the competitors including the robust estimator of Wu (2016). As the associate editor pointed out, one of the main reasons why the nonlinear transformation of the eigenvalue leads to an improvement of the discriminatory power is the fact that the resulting distance function has desirable properties as follows. The distance function should have a derivative close to zero for small and large eigenvalues, whereas the derivative should be large in the boundary region of small and large eigenvalues. Accordingly, the derivative of the distance function should look like a bell shape. The proposed distance function in simulation study is the cumulated distributed function of a normal distribution with just this property. In contrast, Ahn and Horenstein (2013)'s ratio statistic is equivalent to applying the logarithmic function to the eigenvalue. The latter function is less appealing as the derivative tends to infinity as the eigenvalues get small.

The rest of this paper is organized as follows. In Section 2, we introduce the new determination method of the number of common factors in the approximate factor models and state its asymptotic properties. Section 3 performs some simulation experiments to examine the finite sample performance of the new estimator. Section 4 gives some conclusions and discussions. Technical details on the proofs of Theorem 1 and Corollary 1 are presented in the Appendix.

2. Methods and asymptotic properties

Consider the static approximate factor model as follows

$$x_{it} = \lambda_i' f_t + e_{it}, \quad i = 1, 2, \ldots, N, t = 1, 2, \ldots, T, \tag{1}$$

where $f_t = (f_{1t}, f_{2t}, \ldots, f_{rt})'$ and $\lambda_i = (\lambda_{1i}, \lambda_{2i}, \ldots, \lambda_{ri})'$ are respectively r-dimensional unknown common factors and factor loadings, and e_{it} is the idiosyncratic error. For the sake of convenience in statements, denote $m = \min\{N, T\}$, $M = \max\{N, T\}$, $F = (f_1, f_2, \ldots, f_T)'$, $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)'$, X is a $N \times T$ observation matrix with element x_{it}, and E is a $N \times T$ idiosyncratic error matrix with element e_{it}, $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$. Model (1) can be rewritten as the following matrix form

$$X = \Lambda F' + E. \tag{2}$$

Denote the eigenvalues of $\frac{X'X}{NT}$ arranged in descending order $\varphi_i(\frac{X'X}{NT}) = \tilde{\mu}_{NT,i}, i = 1, 2, \ldots, T$, where $\varphi_i(\frac{X'X}{NT})$ means the ith largest eigenvalue of $\frac{X'X}{NT}$. Note that the first r eigenvalues are $O_p(1)$ and the rest are at most $O_p(\frac{1}{m})$ under some conditions, see, e.g. Ahn and Horenstein (2013). Accordingly, some methods are proposed to determine the number of common factors such as the eigenvalues-based ratio-type methods of Ahn and Horenstein (2013) and Wu (2016). As argued by Xia et al. (2015), the ratio-type method of Ahn and Horenstein (2013) has the instability issue of 0/0-type ratio values. Although the instability can be avoided by the ridge parameter, the choice of the parameter can also be thought to be arbitrary. In the following, we suggest a new method to determine the number of common factors in the approximate factor models based on the so-called eigenvalue difference method.

A natural idea is to find a monotonous function so that the function values of the first r largest eigenvalues are very close to one and the rest are very close to zero, and then the number (r) of common factors can be detected easily. That is

$$\begin{aligned} &G(\tilde{\mu}_{NT,i}) \to 1, i = 1, 2, \ldots, r, \quad \text{and} \\ &G(\tilde{\mu}_{NT,i}) \to 0, i = r+1, r+2, \ldots, \end{aligned} \tag{3}$$

and then we can consider the following estimator

$$\hat{r}_{ED} = \arg \max_{1 \le i \le m-1} \{G(\tilde{\mu}_{NT,i}) - G(\tilde{\mu}_{NT,i+1})\}. \tag{4}$$

Clearly, there are many different choices on the functions such as

$$G(\tilde{\mu}_{NT,i}) = \frac{2}{\pi} \arctan(\gamma \ln(m) \tilde{\mu}_{NT,i}),$$

and

$$G(\tilde{\mu}_{NT,i}) = 2\Phi(\gamma \ln(m) \tilde{\mu}_{NT,i}) - 1 =: Pr(|\xi| \le \gamma \ln(m) \tilde{\mu}_{NT,i}),$$

where $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard normal distribution, ξ is a random variable of the standard normal distribution, γ is an user-specified constant.

Let c_0 be a large positive constant. Our theoretical results are derived based on the following assumptions.

Assumption A. Let $\mu_{NT,i} = \varphi_i(\frac{\Lambda'\Lambda}{N}\frac{F'F}{T})$ for $i = 1, 2, \ldots, r$. Then, $plim_{m\to\infty} \mu_{NT,i} = \mu_i \in (0, +\infty)$, $i = 1, 2, \ldots, r$. Moreover, the number r of common factors is assumed to be finite.

Assumption B. (i) $\mathbb{E}\|f_t\|^4 \le c_0$, $\mathbb{E}\|\lambda_i\|^4 \le c_0$.
(ii) $\mathbb{E}\big(\|N^{-\frac{1}{2}}\sum_{i=1}^N e_{it}\lambda_i\|^2\big) \le c_0$.
(iii) $\mathbb{E}\big(N^{-1}\sum_{i=1}^N \|T^{-\frac{1}{2}}\sum_{t=1}^T e_{it} f_t\|^2\big) = \mathbb{E}\big[(NT)^{-1}\|E'F\|^2\big] \le c_0$.

Assumption C. $\varphi_1(\frac{E'E}{M}) = O_p(1)$, where E is a $N \times T$ idiosyncratic error matrix with element e_{it}.

Theorem 1. *Suppose that Assumptions A–C hold, we then have*

$$\lim_{m\to\infty} Pr(\hat{r}_{ED} = r) = 1.$$

In the case with no factors, i.e. $r = 0$, we can slightly modify the above estimator by the same method as that of Ahn and Horenstein (2013). Specifically, we can define a mock eigenvalue, e.g. $\tilde{\mu}_{NT,0} = \min\{\frac{1}{\sqrt{\ln(m)}}, \frac{m}{\sqrt{\ln(m)}}\tilde{\mu}_{NT,1}\}$ such that $\tilde{\mu}_{NT,0} \to 0$ and $\gamma \ln(m)\tilde{\mu}_{NT,0} \to \infty$ as $m \to \infty$. In the following, we redefine the estimator $\hat{r}_{ED}$ using $\tilde{\mu}_{NT,0}$ so that the possible case with $r = 0$ can be covered,

$$\hat{r}_{ED} = \arg \max_{0 \le i \le m-1} \{G(\tilde{\mu}_{NT,i}) - G(\tilde{\mu}_{NT,i+1})\}, \tag{5}$$

and then state a corollary as follows.

Corollary 1. *Suppose that Assumptions A–C hold, we have*

$$\lim_{m\to\infty} Pr(\hat{r}_{ED} = r) = 1.$$