

# IDENTIFYING THE NUMBER OF FACTORS FROM SINGULAR VALUES OF A LARGE SAMPLE AUTO-COVARIANCE MATRIX<sup>1</sup>

BY ZENG LI, QINWEN WANG AND JIANFENG YAO

*University of Hong Kong*

Identifying the number of factors in a high-dimensional factor model has attracted much attention in recent years and a general solution to the problem is still lacking. A promising ratio estimator based on singular values of lagged sample auto-covariance matrices has been recently proposed in the literature with a reasonably good performance under some specific assumption on the strength of the factors. Inspired by this ratio estimator and as a first main contribution, this paper proposes a complete theory of such sample singular values for both the factor part and the noise part under the large-dimensional scheme where the dimension and the sample size proportionally grow to infinity. In particular, we provide an exact description of the phase transition phenomenon that determines whether a factor is strong enough to be detected with the observed sample singular values. Based on these findings and as a second main contribution of the paper, we propose a new estimator of the number of factors which is strongly consistent for the detection of all significant factors (which are the only theoretically detectable ones). In particular, factors are assumed to have the minimum strength above the phase transition boundary which is of the order of a constant; they are thus not required to grow to infinity together with the dimension (as assumed in most of the existing papers on high-dimensional factor models). Empirical Monte-Carlo study as well as the analysis of stock returns data attest a very good performance of the proposed estimator. In all the tested cases, the new estimator largely outperforms the existing estimator using the same ratios of singular values.

**1. Introduction.** Factor models have met a large success in data analysis across many scientific fields such as psychology, economics and signal processing, to name a few. Their attractivity mainly relies on their capability in reducing the generally high dimension of the data to a much lower-dimensional common component. The structure of these models is complex and many different versions have been introduced so far in the long-standing literature on the subject, ranging from *static* to *dynamic* or *generalized dynamic* factor models on one hand, and from *exact* to *approximate* factor models on the other hand. A recent survey of this literature can be found in [Stock and Watson \(2011\)](#). Efforts are however still paid to

---

Received June 2015; revised February 2016.

<sup>1</sup>Supported in part by the HKSAR RGC General Research Fund #17305814.

*MSC2010 subject classifications.* Primary 62M10, 62H25; secondary 15B52.

*Key words and phrases.* High-dimensional factor model, high-dimensional time series, large sample auto-covariance matrices, spiked population model, number of factors, phase transition, random matrices.

the study of these models because their inference is challenging, especially when the cross-sectional dimension  $p$  and the temporal dimension  $T$  are both large.

In such high-dimensional context, the determination of the number  $k$  of common factors in a factor model has a primary importance. Misspecification of this number can deeply affect the quality of the fitted factor model. In this regard, the seminal paper [Bai and Ng \(2002\)](#) provided a consistent estimator of  $k$  for static factor models for the first time. This estimator has attracted much attention and has since been improved or generalized, for example, in [Bai and Ng \(2007\)](#) by the authors themselves, in [Hallin and Liska \(2007\)](#) for dynamic factor models and in [Alessi, Barigozzi and Capasso \(2010\)](#) for approximate factor models. It should be here mentioned that as these developments mainly target at analysis of economic or financial data, the common factors in these models are thought to be *pervasive*, or *strong*, in the sense that their strength is much higher than the strength of the idiosyncratic (error) component. The asymptotic consistency of the factor number estimator depends on this assumption to a large extent. However, some recent studies on factor models suggest the importance for accommodating more factors in these models by including some *weaker factors* which still have a significant explanation power on both cross-sectional and temporal correlations of the data. For example, [Onatski \(2015\)](#) makes a clear distinction between strong factors and weak factors when considering asymptotic approximations of the square loss function from a principal-components-based perspective. A related work allowing weak factors can be found in [Onatski \(2012\)](#).

In this paper, we consider a factor model for high-dimensional time series proposed by [Lam and Yao \(2012\)](#): the observations  $Y$  is a  $p \times T$  matrix with  $p$  cross-sectional units over  $T$  time periods. Let  $y_t$  denote the  $p$ -dimensional vector observed at time  $t$ , then it consists of two components, a low-dimensional common-factor time series  $x_t$  and an idiosyncratic component  $\varepsilon_t$ :

$$(1.1) \quad y_t = Ax_t + \varepsilon_t,$$

where  $A$  is the factor loading matrix of size  $p \times k$  and  $\{\varepsilon_t\}$  is a Gaussian white noise sequence (temporal uncorrelated). The factors in  $(x_t)$  are here loaded contemporaneously; however, this is a time series and its temporal correlation implies that of the observations  $\{y_t\}$ . However, this is the unique source of temporal correlation, and in this aspect, the model is much more restrictive than the general dynamic models as introduced in [Geweke \(1977\)](#), [Sargent and Sims \(1977\)](#) and [Forni et al. \(2000, 2004, 2005\)](#). Nevertheless, there are two advantages in this simplified model. First, since potentially  $(x_t)$  can be any kind of stationary time series of low dimension, the model can already cover a wide range of applications. Second, inference procedures are here more consistently defined and more precise results can be expected, for example, for the determination of the number of factors. The factor model (1.1) can be considered as a reasonable balance between the generality of model coverage and the technical feasibility of underlying inference procedures.

The goal of this paper is to develop a powerful estimator of the number of factors in the model (1.1). Lam and Yao (2012) proposed a ratio-based estimator defined as follows. Let  $\Sigma_y = \text{cov}(y_t, y_{t-1})$  and  $\Sigma_x = \text{cov}(x_t, x_{t-1})$  be the lag-1 auto-covariance matrices of  $y_t$  and  $x_t$ , respectively. Assuming that the factor and the noise are independent, we then have

$$\Sigma_y = A \Sigma_x A',$$

which leads to its symmetric counterpart

$$(1.2) \quad M = \Sigma_y \Sigma_y' = A \Sigma_x \Sigma_x' A'.$$

Notice that here and throughout the paper, the loading matrix  $A$  is normalized by the constraint  $A'A = I_k$  and such constraint is a common set-up in the factor models literature; see, for example, the setting IC2 in Table 1 of Bai and Li (2012). Since the  $k \times k$  matrix  $\Sigma_x$  is generally of full rank  $k$ , the symmetric  $p \times p$  matrix  $M$  has exactly  $k$  nonzero eigenvalues. Moreover, the factor loading space  $\mathcal{M}(A)$ , for example, the  $k$ -dimensional subspace in  $\mathbb{R}^p$  generated by the columns of  $A$ , is spanned by the eigenvectors of  $M$  corresponding to its nonzero eigenvalues  $a_1 \geq \dots \geq a_k > 0$  (factor eigenvalues). Let

$$(1.3) \quad \widehat{M} = \widehat{\Sigma}_y \widehat{\Sigma}_y' \quad \text{where} \quad \widehat{\Sigma}_y = \frac{1}{T} \sum_{t=2}^{T+1} y_t y_{t-1}'$$

be the sample counterparts of  $M$  and  $\Sigma_y$ , respectively. The main observation is that the  $p - k$  null eigenvalues of  $M$  will lead to  $p - k$  “relatively small” sample eigenvalues in  $\widehat{M}$ , while the  $k$  factor eigenvalues ( $a_i$ ) will generate  $k$  “relatively large” eigenvalues in  $\widehat{M}$ . This can be made very precisely in a classical *low-dimensional* framework where we fix the dimension  $p$  and let  $T$  grow to infinity: indeed by law of large numbers,  $\widehat{M} \rightarrow M$  and by continuity, all the eigenvalues  $l_1 \geq l_2 \geq \dots \geq l_p$  (sorted in decreasing order) of  $\widehat{M}$  will converge to the corresponding eigenvalues of  $M$ . In particular, for  $k < i \leq p, l_i \rightarrow 0$  while  $l_i \rightarrow a_i > 0$  for  $1 \leq i \leq k$ . Consider the ratio estimator [Lam and Yao (2012)]:

$$(1.4) \quad \tilde{k} = \arg \min_{1 \leq i < p} l_{i+1}/l_i.$$

As  $l_{k+1}/l_k$  will be the first ratio in this list which tends to zero,  $\tilde{k}$  will be a consistent estimator of  $k$ .

In the high-dimensional context, however,  $\widehat{M}$  will significantly deviate from  $M$  and the spectrum ( $l_i$ ) of  $\widehat{M}$  will not be close to that of  $M$  anymore. In particular, the time for the first minimum of the ratios in (1.4) becomes noisy and can be much different from the target value  $k$ . Notice that the  $k$  nonnull factor eigenvalues ( $a_i$ ) are directly linked to the strength of the factor time series ( $x_t$ ). The precise relationship between the ratios of sample eigenvalues in (1.4) will ultimately depend on a complex interplay between the strength of the factor eigenvalues ( $a_i$ ) (compared to the noise level), the dimension  $p$  and the sample size  $T$ .

Despite of the introduction of a very appealing ratio estimator (1.4), a precise description of the sample ratios  $l_{i+1}/l_i$  is missing in Lam and Yao (2012). Indeed, the authors establish the consistency of the ratio estimator  $\tilde{k}$  by requiring that the factor strengths  $(a_i)$  all explode at a same rate:  $a_i \propto p^\delta$  for all  $1 \leq i \leq k$  and some  $\delta > 0$  as the dimension  $p$  grows to infinity. In other words, the factors are all strong and they have the same asymptotic strength. This limitation is quite severe because factors with different levels of strength cannot all be detected within this framework. For instance, if we have factors with three levels of strength  $p^{\delta_j}$ ,  $j = 1, 2, 3$  where  $\delta_1 > \delta_2 > \delta_3$ , the ratio estimator  $\tilde{k}$  above will correctly identify the group of strongest factors  $a_i \propto p^{\delta_1}$  while all the others will be omitted. In an attempt to correct such undesirable behavior, a two-step estimation procedure is also proposed in Lam and Yao (2012) to successively identify two groups of factors with top two strengths: this means that in the example above, factors of strength  $a_i$  proportional to  $p^{\delta_j}$  with  $j = 1, 2$  will be identified while the others will remain omitted. The issue here is that *a priori*, we do not know how many different levels of strength the factors could have and it is unlikely we could attempt to estimate such different levels as this would lead to a problem that is equally (if not more) difficult than the initial problem of estimating the number of factors.

Inspired by the appropriateness of the ratio estimator  $\tilde{k}$  in the high-dimensional context, the main objective of this paper is to provide a rigorous theory for the estimation of the number of factors based on the ratios  $\{l_{i+1}/l_i\}$  under the high-dimensional setting where  $p$  and  $T$  tend to infinity proportionally.

This paper contains two main contributions. First, we completely characterize the limits of both the factor eigenvalues  $\{l_i\}_{1 \leq i \leq k}$  and the noise eigenvalues  $\{l_i\}_{k < i \leq p}$ . For the noise part, as  $k$  (although unknown) is much smaller than the dimension  $p$ , we prove that the spectral distribution generated by  $\{l_i\}_{k < i \leq p}$  has a limit which coincides with the limit of the spectral distribution generated by the  $p$  eigenvalues of the (unobserved) matrix  $\widehat{M}_\varepsilon = \widehat{\Sigma}_\varepsilon \widehat{\Sigma}'_\varepsilon$  where  $\widehat{\Sigma}_\varepsilon = T^{-1} \sum_{t=2}^{T+1} \varepsilon_t \varepsilon'_{t-1}$ . This limiting distribution has been explored elsewhere in Li, Pan and Yao (2015) and its support is found to be a compact interval  $[a, b]$ . As for the factor part  $\{l_i\}_{1 \leq i \leq k}$ , although it is highly expected that they should have a limit located outside the base interval  $[a, b]$ , we establish a *phase transition phenomenon*: a factor eigenvalue  $l_i$  will tend to a limit  $\lambda_i > b$  (outlier) if and only if the corresponding population factor strength  $a_i$  exceeds some critical value  $\tau$ . In other words, if a factor  $a_i$  is too weak, then the corresponding sample factor eigenvalue  $\lambda_i$  will tend to  $b$ , the (limit of) maximum of the noise eigenvalues and it will be hardly detectable. Moreover, both the outlier limits  $\{\lambda_i\}$  and the critical value  $\tau$  are characterized through the model parameters.

The second main contribution of the paper is the derivation of a new estimator  $\widehat{k}$  of the number of factors based on the finding above. If  $k_0$  denotes the number of significant factors, that is, with factor strength  $a_i > \tau$ , then using an appropriate thresholding interval  $(1 - d_T, 1)$  for the sample ratios  $\{l_{i+1}/l_i\}$ , the derived estimator  $\widehat{k}$  is strongly consistent converging to  $k_0$ . In addition to this well-justified

consistency, the main advantage of the proposed estimator is its robustness against possibly multiple levels of factor strength; in theory, all factors with strength above the constant  $\tau$  are detectable. Therefore, both strong factors and weak factors can be present, and their strengths can have different asymptotic rates with regard to the dimension  $p$  in order to be detected from the observed samples. This is a key difference between the method provided in this paper and most of the existing estimators of the factor number as mentioned previously [the reader is however reminded that the model (1.1) is more restrictive than a general dynamic factor model]. Notice however that these precise results have been obtained at the cost of some drastic simplification of the idiosyncratic component  $\{\varepsilon_t\}$ , namely independence has been assumed both serially and cross-sectionally (over the time and the dimension), and the components are normalized to have a same value of variance (see Assumption 2 in Section 2). These limitations are required by the technical tools employed in this paper and some nontrivial extension of these tools are needed to get rid of these limitations.

From a methodological point of view, our approach is based on recent advances in random matrix theory, specifically on the so-called spiked population models or more generally on finite-rank perturbations of large random matrices. We start by identifying the sample matrix  $\widehat{M}$  as a finite-rank perturbation of the base matrix  $\widehat{M}_\varepsilon$  associated to the noise. In a recent paper [Li, Pan and Yao \(2015\)](#), the limiting spectral distribution of the eigenvalues of  $\widehat{M}_\varepsilon$  has been found and the base interval  $[a, b]$  characterized. By developing the mentioned perturbation theory for the auto-covariance matrix  $\widehat{M}$ , we find the characterization of the limits of its eigenvalues  $\{l_i\}$ .

For the strong consistency of the proposed ratio estimator  $\widehat{k}$ , a main ingredient is the almost sure convergence of the largest eigenvalue of the base matrix  $\widehat{M}_\varepsilon$  to the right edge  $b$ , recently established in [Wang and Yao \(2014\)](#). This result serves as the cornerstone for distinguishing between significant factors and noise components.

It is worth mentioning a related paper [Onatski \(2010\)](#) where the author stands from a similar perspective with the method in this paper. However, that paper addresses static approximate factor models without time series dependence and more importantly, the assumption of explosion of all factor eigenvalues is still required which, on the contrary, is released in this paper.

The rest of the paper is organized as follows. In Section 2, after introduction of the model assumptions we develop our first main result regarding spectral limits of  $\widehat{M}$ . The new estimator  $\widehat{k}$  is then introduced in Section 3 and its strong convergence to the number of significant factors  $k_0$  established. In Section 4, detailed Monte-Carlo experiments are conducted to check the finite-sample properties of the proposed estimator and to compare it with the ratio estimator  $\widehat{k}$  (1.4) from [Lam and Yao \(2012\)](#). Both estimators  $\widehat{k}$  and  $\widehat{k}$  are then tested in Section 5 on a real data set from Standard & Poor stock returns and compared in detail. Notice that some technical lemmas used in the main proofs are gathered in a companion paper of supplementary material [[Li, Wang and Yao \(2016\)](#)].

**2. Large-dimensional limits of noise and factor eigenvalues.** The static factor model (1.1) is further specified to satisfy the following assumptions.

ASSUMPTION 1. The factor  $(x_t)$  is a  $k$ -dimensional ( $k \ll p$  fixed) stationary time series whose  $p$  components are independent linear processes of form

$$x_{it} = \sum_{l=0}^{\infty} \phi_{il} \eta_{i,t-l}, \quad i = 1, \dots, k, t = 1, \dots, T + 1.$$

Here, for each  $i$ ,  $(\eta_{ik})$  is a real-valued and weakly stationary white noise with mean 0 and variance  $\sigma_i^2$ . The  $i$ th time series  $\{x_{it}\}_{t \geq 1}$  has variance  $\gamma_0(i)$  and lag-1 auto-covariance  $\gamma_1(i)$ . Moreover, the variance  $\gamma_0(i)$  will be hereafter referred to as the *strength* of the  $i$ th factor time series  $\{x_{it}\}$ .

ASSUMPTION 2. The idiosyncratic component  $(\varepsilon_t)$  is independent of  $(x_t)$ . Each  $\varepsilon_t$  is a  $p$ -dimensional real valued random vector with independent entries  $\varepsilon_{it}, i = 1, \dots, p$  and the whole array of variables  $\{\varepsilon_{it}\}$  are independent and satisfies the conditions:

$$\mathbb{E}(\varepsilon_{it}) = 0, \quad \mathbb{E}(\varepsilon_{it}^2) = \sigma^2,$$

and for any  $\eta > 0$ ,

$$(2.1) \quad \frac{1}{\eta^4 p T} \sum_{i=1}^p \sum_{t=1}^{T+1} \mathbb{E}(|\varepsilon_{it}|^4 I_{(|\varepsilon_{it}| \geq \eta T^{1/4})}) \longrightarrow 0 \quad \text{as } (pT) \rightarrow \infty.$$

ASSUMPTION 3. The dimension  $p$  and the sample size  $T$  tend to infinity proportionally:  $p \rightarrow \infty, T = T(p) \rightarrow \infty$  and  $p/T \rightarrow y > 0$ .

Assumption 1 defines the static factor model considered in this paper. Assumption 2 details the moment condition and the independent structure of the noise. In particular, (2.1) is a Lindeberg-type condition commonly used in random matrix theory. In particular, if the fourth moments of the variables  $\{\varepsilon_{it}\}$  are uniformly bounded, the Lindeberg condition is satisfied. Assumption 3 defines the high-dimensional setting where both the dimension and the sample size can be large with comparable magnitude.

First, we have

$$\begin{aligned} \widehat{\Sigma}_y &= \frac{1}{T} \sum_{t=2}^{T+1} y_t y_{t-1}' = \frac{1}{T} \sum_{t=2}^{T+1} (Ax_t + \varepsilon_t)(Ax_{t-1} + \varepsilon_{t-1})' \\ &= \frac{1}{T} \sum_{t=2}^{T+1} Ax_t x_{t-1}' A' + \frac{1}{T} \sum_{t=2}^{T+1} (Ax_t \varepsilon_{t-1}' + \varepsilon_t x_{t-1}' A') + \frac{1}{T} \sum_{t=2}^{T+1} \varepsilon_t \varepsilon_{t-1}' \\ &:= P_A + \widehat{\Sigma}_\varepsilon. \end{aligned}$$

The matrix  $\widehat{\Sigma}_\varepsilon = T^{-1} \sum_t \varepsilon_t \varepsilon'_{t-1}$  is the analogous sample auto-covariance matrix associated to the noise  $(\varepsilon_t)$ . Since  $A$  has rank  $k$ , the rank of the matrix  $P_A$  is bounded by  $2k$  (we will see in fact that asymptotically, the rank of  $P_A$  will be eventually  $k$ ). Therefore, the auto-covariance matrix of interest  $\widehat{\Sigma}_y$  is seen as a finite-rank perturbation of the noise auto-covariance matrix  $\widehat{\Sigma}_\varepsilon$ . Since the matrix  $\widehat{\Sigma}_y$  is not symmetric, we consider its singular values, that is, the square roots of the positive eigenvalues of  $\widehat{M} := \widehat{\Sigma}_y \widehat{\Sigma}'_y$ . Therefore, the study of the singular values of  $\widehat{\Sigma}_y$  reduces to the study of the eigenvalues of  $\widehat{M}$ , which is also a finite rank perturbation of the base component  $\widehat{M}_\varepsilon := \widehat{\Sigma}_\varepsilon \widehat{\Sigma}'_\varepsilon$ .

Finite-rank perturbations of random matrices have been actively studied in recent years and the theory is much linked to the *spiked population models* well known in high-dimensional statistics literature. For some recent accounts on this theory, we refer to Bai and Yao (2008), Baik and Silverstein (2006), Benaych-Georges and Nadakuditi (2011), Johnstone (2001), Passemier and Yao (2012) and the references therein. A general picture from this theory is that first, the eigenvalues of the base matrix will converge to a limiting spectral distribution (LSD) with a compact support, say an interval  $[a, b]$ ; and second, for the eigenvalues of the perturbed matrices; most of them (base eigenvalues) will converge to the same LSD independently of the perturbation while a small number among the largest ones will converge to a limit outside the support of the LSD (outliers). However, all the existing literature cited above concern the finite rank perturbation of large-dimensional sample covariance matrices or Wigner matrices. As a theoretic contribution of this paper, we extend this theory to the case of a *perturbed auto-covariance matrix* by giving exact conditions under which the aforementioned dichotomy between base eigenvalues and outliers still hold. Specifically, we prove in this section that once the  $k$  factor strengths  $(a_i)$  are not “too weak”, they will generate exactly  $k$  outliers, while the remaining  $p - k$  eigenvalues will behave as the eigenvalues of the base  $\widehat{M}_\varepsilon$ , which converges to a compactly supported LSD. It is then apparent that under such dichotomy and by “counting” the outliers outside the interval  $[a, b]$ , we will be able to obtain a consistent estimator of the number of factors  $k$ .

In what follows, we first recall two existing results on the limits of the singular values of  $\widehat{\Sigma}_\varepsilon$ . Then we develop our theory on the limits of largest (outliers) and base singular values of  $\widehat{\Sigma}_y$ .

2.1. *LSD of  $\widehat{M}_\varepsilon$ .* We first recall two useful results on the base matrix  $\widehat{M}_\varepsilon$ . First, the LSD of the matrix  $\widehat{M}_\varepsilon$  has been obtained in a recent paper of Li, Pan and Yao (2015). Write

$$\begin{aligned} \widehat{M}_\varepsilon &= \left( \frac{1}{T} \sum_{t=2}^{T+1} \varepsilon_t \varepsilon'_{t-1} \right) \left( \frac{1}{T} \sum_{t=2}^{T+1} \varepsilon_t \varepsilon'_{t-1} \right)' \\ &= \frac{1}{T^2} X Y' Y X', \end{aligned}$$

with the data matrices

$$X = \begin{pmatrix} \varepsilon_{12} & \cdots & \varepsilon_{1,T+1} \\ \vdots & \ddots & \vdots \\ \varepsilon_{p2} & \cdots & \varepsilon_{p,T+1} \end{pmatrix}, \quad Y = \begin{pmatrix} \varepsilon_{11} & \cdots & \varepsilon_{1T} \\ \vdots & \ddots & \vdots \\ \varepsilon_{p1} & \cdots & \varepsilon_{pT} \end{pmatrix}.$$

Furthermore, let  $\mu$  be a measure on the real line supported on an interval  $[\alpha, \beta]$  (the end points can be infinity), with its Stieltjes transform defined as

$$m(z) = \int \frac{1}{t - z} d\mu(t) \quad \text{for } z \in \mathbb{C} \setminus \text{supp}(\mu),$$

and its  $T$ -transform as

$$T(z) = \int \frac{t}{z - t} d\mu(t) \quad \text{for } z \in \mathbb{C} \setminus \text{supp}(\mu).$$

Notice here that the  $T$ -transform is a decreasing homeomorphism from  $(-\infty, \alpha)$  onto  $(T(\alpha^-), 0)$  and from  $(\beta, +\infty)$  onto  $(0, T(\beta^+))$ . These two transforms are related each other by the following equation:

$$T(z) = -1 - zm(z).$$

**PROPOSITION 2.1** [Li, Pan and Yao (2015)]. *Suppose that Assumptions 2 and 3 hold with  $\sigma^2 = 1$ . Then the empirical spectral distribution of  $B := 1/T^2 Y' Y X' X$  (which is the companion matrix of  $\widehat{M}_\varepsilon$ ) converges a.s. to a nonrandom limit  $F$ , whose Stieltjes transform  $m = m(z)$  satisfies the equation*

$$(2.2) \quad z^2 m^3(z) - 2z(y - 1)m^2(z) + (y - 1)^2 m(z) - zm(z) - 1 = 0.$$

*In particular, this LSD is supported on the interval  $[a\mathbf{1}_{\{y \geq 1\}}, b]$  whose end points are*

$$(2.3) \quad a = (-1 + 20y + 8y^2 - (1 + 8y)^{3/2})/8,$$

$$(2.4) \quad b = (-1 + 20y + 8y^2 + (1 + 8y)^{3/2})/8.$$

Notice that the companion matrix  $B$  is  $T \times T$  and it shares the same  $p \wedge T$  non-null eigenvalues with  $\widehat{M}_\varepsilon$ . Therefore, the support of  $\widehat{M}_\varepsilon$  is also  $[a, b]$ . The LSD  $F$  of  $B$  and the LSD  $F^*$  of  $\widehat{M}_\varepsilon$  are linked by the relationship

$$yF^* - F = (y - 1)\delta_0,$$

where  $\delta_0$  is the Dirac mass at the origin.

**REMARK 2.1.** The equation (2.2) can be expressed using the  $T$ -transform:

$$(2.5) \quad (T(z) + 1)(T(z) + y)^2 = zT(z).$$

The second result is about the convergence of the largest eigenvalue of  $\widehat{M}_\varepsilon$ .



PROPOSITION 2.2 [Wang and Yao (2014)]. *Suppose that Assumptions 2 and 3 hold with  $\sigma^2 = 1$ . Then the largest eigenvalue of  $\widehat{M}_\varepsilon$  converges a.s. to the right end point  $b$  of its LSD given in (2.4).*

Combining Propositions 2.1 and 2.2, we have the following corollary.

COROLLARY 2.1. *Under the same conditions as in Proposition 2.2, if  $(\beta_j)$  are sorted eigenvalues of  $\widehat{M}_\varepsilon$ , then for any fixed  $m$ , the  $m$  largest eigenvalues  $\beta_1 \geq \beta_2 \geq \dots \geq \beta_m$  all converge to  $b$  almost surely.*

PROOF. For any  $\delta > 0$ , almost surely the number of sample eigenvalues of  $\beta_j$  falling into the interval  $(b - \delta, b)$  grows to infinity due to the fact the density of the LSD is positive and continuous on this interval. Then for fixed  $m$ , a.s.  $\liminf_{p \rightarrow \infty} \beta_m \geq b - \delta$ . By letting  $\delta \rightarrow 0$ , we have a.s.  $\liminf_{p \rightarrow \infty} \beta_m \geq b$ . Obviously,  $\limsup_{p \rightarrow \infty} \beta_m \leq \limsup_{p \rightarrow \infty} \beta_1 = b$ , that is, a.s.  $\lim_{p \rightarrow \infty} \beta_m = b$ .  $\square$

2.2. *Convergence of the largest eigenvalues of the sample auto-covariance matrix  $\widehat{M}$ .* The following main result of the paper characterizes the limits of the  $k$ -largest eigenvalues of the sample auto-covariance matrix  $\widehat{M}$ . Notice that Propositions 2.1 and 2.2 hold for general white noise  $\varepsilon_t$ , for technical reasons; our main results (Theorems 2.1 and 3.1 below) are established under the Gaussian assumption.

THEOREM 2.1. *Suppose that the model (1.1) satisfies Assumptions 1, 2 and 3. The noise  $\{\varepsilon_t\}$  are normally distributed and the loading matrix  $A$  is normalized as  $A'A = I_k$ . Let  $l_i (1 \leq i \leq k)$  denote the  $k$  largest eigenvalue of  $\widehat{M}$ . Then for each  $1 \leq i \leq k$ ,  $l_i/\sigma^4$  converges almost surely to a limit  $\lambda_i$ . Moreover,*

$$\lambda_i = b \quad \text{when } T_1(i) \geq T(b^+),$$

where

$$(2.6) \quad T_1(i) = \left( 2y\sigma^2\gamma_0(i) + \gamma_1(i)^2 - \sqrt{(2y\sigma^2\gamma_0(i) + \gamma_1(i)^2)^2 - 4y^2\sigma^4(\gamma_0(i)^2 - \gamma_1(i)^2)} \right) / (2\gamma_0(i)^2 - 2\gamma_1(i)^2).$$

Otherwise, that is,  $T_1(i) < T(b^+)$ ,  $\lambda_i > b$  and its value is characterized by the fact that the  $T$ -transform  $T(\lambda_i)$  is the solution to the equation:

$$(2.7) \quad (y\sigma^2 - \gamma_0(i)T(\lambda_i))^2 = \gamma_1(i)^2 T(\lambda_i)(1 + T(\lambda_i)).$$

The theorem establishes a *phase transition phenomenon* for the  $k$  sample factor eigenvalues  $(l_i)$ . Define the *number of significant factors*:

$$(2.8) \quad k_0 = \#\{1 \leq i \leq k : T_1(i) < T(b^+)\}.$$

Therefore, for each of the  $k_0$  significant factors, the corresponding sample eigenvalue  $l_i$  will converge to a limit  $\lambda_i$  outside the base support interval  $[a, b]$ . In contrary, for the  $k - k_0$  factors for which  $T_1(i) \geq T(b^+)$ , they are too weak in the sense that the corresponding sample eigenvalue  $l_i$  will converge to  $b$ , which is also the limit of the largest noise eigenvalues  $l_{k+1}, \dots, l_{k+m}$  ( $m$  is a fixed number here). Therefore, these weakest factors will be merged into the noise component and their detection becomes nearly impossible.

Later in Section 2.3, it will be established that for the  $i$ th factor time series to be significant, the phase transition condition  $T_1(i) < T(b^+)$  essentially requires its strength  $\gamma_0(i)$  be large enough.

**PROOF OF THEOREM 2.1.** The proof consists of four steps where some technical lemmas are to be found in the companion paper of supplementary material [Li, Wang and Yao (2016)].

*Step 1. Simplification of variance  $\sigma^2$  of white noise  $\{\varepsilon_{it}\}$ .* To start with, we reduce the variance of the white noise from  $\sigma^2$  to 1. Indeed, the model (1.1) is equivalent to

$$\frac{y_t}{\sigma} = A \frac{x_t}{\sigma} + \frac{\varepsilon_t}{\sigma}.$$

And if we denote  $\tilde{y}_t = y_t/\sigma$ ,  $\tilde{x}_t = x_t/\sigma$  and  $\tilde{\varepsilon}_t = \varepsilon_t/\sigma$ , then we are dealing with the model

$$(2.9) \quad \tilde{y}_t = A\tilde{x}_t + \tilde{\varepsilon}_t,$$

where the white noise  $\tilde{\varepsilon}$  has mean zero and unit variance and the variance and auto-covariance of the factor process  $\{\tilde{x}_t\}$  satisfies

$$(2.10) \quad \tilde{\gamma}_0(i) = \gamma_0(i)/\sigma^2, \quad \tilde{\gamma}_1(i) = \gamma_1(i)/\sigma^2,$$

in which  $\gamma_0(i)$  and  $\gamma_1(i)$  are the variance and auto-covariance of the original factor process  $\{x_t\}$ . Therefore, in all the following, we just consider the standardized Model (2.9). For convenience, we use the notation of the original model (1.1) and set  $\sigma^2 = 1$  to investigate Model (2.9). At the end of the proof, we will replace the value of  $\gamma_0(i)$  and  $\gamma_1(i)$  with  $\tilde{\gamma}_0(i)$  and  $\tilde{\gamma}_1(i)$  to recover the corresponding results for Model (1.1).

*Step 2. Simplification of matrix  $A$ .* Here, we argue that it is enough to consider the case where the loading matrix  $A$  has the canonical form

$$A = \begin{pmatrix} I_k \\ \mathbf{0}_{p-k} \end{pmatrix}.$$

Suppose  $A$  is not in this canonical form. Since by assumption  $A'A = I_k$ , we can complete  $A$  to an orthogonal matrix  $Q = (A, C)$  by adding appropriate orthonormal columns. From the model equation (1.1), we have

$$Q'y_t = Q'Ax_t + Q'\varepsilon_t = \begin{pmatrix} A' \\ C' \end{pmatrix} Ax_t + Q'\varepsilon_t = \begin{pmatrix} I_k \\ \mathbf{0}_{p-k} \end{pmatrix} x_t + Q'\varepsilon_t.$$

Since  $\varepsilon_t \sim \mathcal{N}(0, I_p)$  and  $Q'$  is orthogonal,  $Q'\varepsilon_t \sim \mathcal{N}(0, I_p)$ . Let  $z_t := Q'y_t$ , then  $z_t$  satisfies the model equation (1.1) with a canonical loading matrix. What happens is that the singular values of the two lag-1 auto-covariance matrices

$$\frac{1}{T} \sum_{t=2}^{T+1} z_t z'_{t-1}, \quad \frac{1}{T} \sum_{t=2}^{T+1} y_t y'_{t-1}$$

are the same: this is simply due to fact that

$$\begin{aligned} & \left( \frac{1}{T} \sum_{t=2}^{T+1} y_t y'_{t-1} \right) \left( \frac{1}{T} \sum_{t=2}^{T+1} y_t y'_{t-1} \right)' \\ &= \left( \frac{1}{T} \sum_{t=2}^{T+1} Q'z_t \cdot (Q'z_{t-1})' \right) \left( \frac{1}{T} \sum_{t=2}^{T+1} Q'z_t \cdot (Q'z_{t-1})' \right)' \\ &= Q' \left( \frac{1}{T} \sum_{t=2}^{T+1} z_t z'_{t-1} \right) \left( \frac{1}{T} \sum_{t=2}^{T+1} z_t z'_{t-1} \right)' Q. \end{aligned}$$

*Step 3. Derivation of the main equation (2.7).* From now on, we assume that  $A$  is in its canonical form. By the definition of  $y_t$ , we have

$$\begin{aligned} \widehat{\Sigma}_y &= \frac{1}{T} \sum_{t=2}^{T+1} y_t y'_{t-1} \\ &= \frac{1}{T} \sum_{t=2}^{T+1} \begin{pmatrix} x_{1t} + \varepsilon_{1t} \\ \vdots \\ x_{kt} + \varepsilon_{kt} \\ \varepsilon_{k+1t} \\ \vdots \\ \varepsilon_{pt} \end{pmatrix} \\ (2.11) \quad & \times (x_{1t-1} + \varepsilon_{1t-1} \quad \cdots \quad x_{kt-1} + \varepsilon_{kt-1} \quad \varepsilon_{k+1t-1} \quad \cdots \quad \varepsilon_{pt-1}) \\ & := \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \end{aligned}$$

where we use  $A, B, C$  and  $D$  to denote the four blocks. Define

$$\begin{aligned}
 X_0 &:= \frac{1}{\sqrt{T}} \begin{pmatrix} x_{11} + \varepsilon_{11} & \cdots & x_{k1} + \varepsilon_{k1} \\ \vdots & \ddots & \vdots \\ x_{1T} + \varepsilon_{1T} & \cdots & x_{kT} + \varepsilon_{kT} \end{pmatrix}, \\
 X_1 &:= \frac{1}{\sqrt{T}} \begin{pmatrix} x_{12} + \varepsilon_{12} & \cdots & x_{k2} + \varepsilon_{k2} \\ \vdots & \ddots & \vdots \\ x_{1T+1} + \varepsilon_{1T+1} & \cdots & x_{kT+1} + \varepsilon_{kT+1} \end{pmatrix}, \\
 E_1 &:= \frac{1}{\sqrt{T}} \begin{pmatrix} \varepsilon_{k+11} & \cdots & \varepsilon_{p1} \\ \vdots & \ddots & \vdots \\ \varepsilon_{k+1T} & \cdots & \varepsilon_{pT} \end{pmatrix}, \\
 E_2 &:= \frac{1}{\sqrt{T}} \begin{pmatrix} \varepsilon_{k+12} & \cdots & \varepsilon_{p2} \\ \vdots & \ddots & \vdots \\ \varepsilon_{k+1T+1} & \cdots & \varepsilon_{pT+1} \end{pmatrix}.
 \end{aligned}$$

Then we have

$$(2.12) \quad A = X_1' X_0, \quad B = X_1' E_1, \quad C = E_2' X_0, \quad D = E_2' E_1.$$

Since  $l$  is the extreme large eigenvalue of  $\widehat{\Sigma}_y \widehat{\Sigma}'_y$ ,  $\sqrt{l}$  is the extreme large singular value of  $\widehat{\Sigma}_y$ , or equivalently,  $\sqrt{l}$  is the positive eigenvalue of the  $2p \times 2p$  matrix

$$(2.13) \quad \begin{pmatrix} 0 & \widehat{\Sigma}_y \\ \widehat{\Sigma}'_y & 0 \end{pmatrix}.$$

When the block expression (2.11) is combined with the definition of each block in (2.12), (2.13) is equivalent to

$$(2.14) \quad \begin{pmatrix} 0 & 0 & X_1' X_0 & X_1' E_1 \\ 0 & 0 & E_2' X_0 & E_2' E_1 \\ X_0' X_1 & X_0' E_2 & 0 & 0 \\ E_1' X_1 & E_1' E_2 & 0 & 0 \end{pmatrix}.$$

If we interchange the second and third row block and column block in (2.14), its eigenvalues remain the same. Therefore,  $\sqrt{l}$  should satisfy the following equa-

tion:

$$(2.15) \quad \left| \begin{pmatrix} \sqrt{l} & -X'_1 X_0 & 0 & -X'_1 E_1 \\ -X'_0 X_1 & \sqrt{l} & -X'_0 E_2 & 0 \\ 0 & -E'_2 X_0 & \sqrt{l} & -E'_2 E_1 \\ -E'_1 X_1 & 0 & -E'_1 E_2 & \sqrt{l} \end{pmatrix} \right| = 0.$$

Then for block matrix, we have the identity  $\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det D \cdot \det(A - BD^{-1}C)$  when  $D$  is invertible, then (2.15) is equivalent to

$$(2.16) \quad \left| \begin{pmatrix} \sqrt{l} & -X'_1 X_0 \\ -X'_0 X_1 & \sqrt{l} \end{pmatrix} - \begin{pmatrix} 0 & -X'_1 E_1 \\ -X'_0 E_2 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{l} & -E'_2 E_1 \\ -E'_1 E_2 & \sqrt{l} \end{pmatrix}^{-1} \begin{pmatrix} 0 & -E'_2 X_0 \\ -E'_1 X_1 & 0 \end{pmatrix} \right| = 0.$$

Here the inverse matrix exists because  $\sqrt{l}$  is an extreme singular value, so that

$$\left| \begin{pmatrix} \sqrt{l} & -E'_2 E_1 \\ -E'_1 E_2 & \sqrt{l} \end{pmatrix} \right| \neq 0.$$

Next, by expanding

$$\begin{pmatrix} \sqrt{l} & -E'_2 E_1 \\ -E'_1 E_2 & \sqrt{l} \end{pmatrix}^{-1},$$

we find that (2.16) is equivalent to

$$\left| \begin{pmatrix} \sqrt{l}I_k - \sqrt{l}X'_1 E_1 (II - E'_1 E_2 E'_2 E_1)^{-1} E'_1 X_1 & -X'_1 (I + E_1 E'_1 E_2 (II - E'_2 E_1 E'_1 E_2)^{-1} E'_2) X_0 \\ -X'_0 (I + E_2 E'_2 E_1 (II - E'_1 E_2 E'_2 E_1)^{-1} E'_1) X_1 & \sqrt{l}I_k - \sqrt{l}X'_0 E_2 (II - E'_2 E_1 E'_1 E_2)^{-1} E'_2 X_0 \end{pmatrix} \right| = 0,$$

and using the simple fact that

$$A(II - BA)^{-1} = (II - AB)^{-1}A$$

leads to

$$(2.17) \quad \left| \begin{pmatrix} \sqrt{l}I_k - \sqrt{l}X'_1 (II - E_1 E'_1 E_2 E'_2)^{-1} E_1 E'_1 X_1 & -X'_1 (I + (II - E_1 E'_1 E_2 E'_2)^{-1} E_1 E'_1 E_2 E'_2) X_0 \\ -X'_0 (I + (II - E_2 E'_2 E_1 E'_1)^{-1} E_2 E'_2 E_1 E'_1) X_1 & \sqrt{l}I_k - \sqrt{l}X'_0 (II - E_2 E'_2 E_1 E'_1)^{-1} E_2 E'_2 X_0 \end{pmatrix} \right| = 0.$$

Taking Lemmas 1.3 and 1.4 given in Li, Wang and Yao (2016) into consideration, the matrix in (2.17) tends to a block matrix:

$$\left( \begin{array}{ccc|ccc} \frac{\sqrt{\lambda}(y - \gamma_0(1)T(\lambda))}{y + T(\lambda)} & \cdots & 0 & -(1 + T(\lambda))\gamma_1(1) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\sqrt{\lambda}(y - \gamma_0(k)T(\lambda))}{y + T(\lambda)} & 0 & \cdots & -(1 + T(\lambda))\gamma_1(k) \\ \hline -(1 + T(\lambda))\gamma_1(1) & \cdots & 0 & \frac{\sqrt{\lambda}(y - \gamma_0(1)T(\lambda))}{y + T(\lambda)} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -(1 + T(\lambda))\gamma_1(k) & 0 & \cdots & \frac{\sqrt{\lambda}(y - \gamma_0(k)T(\lambda))}{y + T(\lambda)} \end{array} \right),$$

so  $\lambda$  should make the determinant of this matrix equal to 0. If we interchange the first and second column block, the matrix becomes the following:

$$\left( \begin{array}{ccc|ccc} -(1 + T(\lambda))\gamma_1(1) & \cdots & 0 & \frac{\sqrt{\lambda}(y - \gamma_0(1)T(\lambda))}{y + T(\lambda)} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -(1 + T(\lambda))\gamma_1(k) & 0 & \cdots & \frac{\sqrt{\lambda}(y - \gamma_0(k)T(\lambda))}{y + T(\lambda)} \\ \hline \frac{\sqrt{\lambda}(y - \gamma_0(1)T(\lambda))}{y + T(\lambda)} & \cdots & 0 & -(1 + T(\lambda))\gamma_1(1) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\sqrt{\lambda}(y - \gamma_0(k)T(\lambda))}{y + T(\lambda)} & 0 & \cdots & -(1 + T(\lambda))\gamma_1(k) \end{array} \right).$$

Since the diagonal block is

$$\left| \left( \begin{array}{ccc} -(1 + T(\lambda))\gamma_1(1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & -(1 + T(\lambda))\gamma_1(k) \end{array} \right) \right| \neq 0,$$

we can use again the identity for determinant of a block matrix and find that

$$\lambda(y - \gamma_0(i)T(\lambda))^2 - \gamma_1(i)^2(1 + T(\lambda))^2(y + T(\lambda))^2 = 0, \quad i \in [1, \dots, k].$$

Combining this equation with (2.5) and replacing  $\gamma_0(i)$ ,  $\gamma_1(i)$  with  $\gamma_0(i)/\sigma^2$ ,  $\gamma_1(i)/\sigma^2$  leads to the equation (2.7).

*Step 4. Derivation of the condition  $T_1(i) < T(b^+)$ .* We now look at the solution of the main equation (2.7). The equation reduces to

$$(2.18) \quad [\gamma_0(i)^2 - \gamma_1(i)^2] \cdot T^2(\lambda_i) - [\gamma_1(i)^2 + 2y\sigma^2\gamma_0(i)] \cdot T(\lambda_i) + \sigma^4y^2 = 0.$$

Since the part  $\gamma_0(i)^2 - \gamma_1(i)^2 > 0$  and  $\gamma_1(i)^2 + 2y\sigma^2\gamma_0(i) > 0$ , the equation (2.18) has two positive roots

$$(2.19) \quad \begin{cases} T_1(i) = \frac{2y\sigma^2\gamma_0(i) + \gamma_1(i)^2 - \sqrt{(2y\sigma^2\gamma_0(i) + \gamma_1(i)^2)^2 - 4y^2\sigma^4(\gamma_0(i)^2 - \gamma_1(i)^2)}}{2\gamma_0(i)^2 - 2\gamma_1(i)^2}, \\ T_2(i) = \frac{2y\sigma^2\gamma_0(i) + \gamma_1(i)^2 + \sqrt{(2y\sigma^2\gamma_0(i) + \gamma_1(i)^2)^2 - 4y^2\sigma^4(\gamma_0(i)^2 - \gamma_1(i)^2)}}{2\gamma_0(i)^2 - 2\gamma_1(i)^2}. \end{cases}$$

Recall the definition of the  $T$ -transform that

$$T(z) = \int \frac{t}{z-t} d\mu(t),$$

taking derivatives with respect to  $z$  on both side leads to

$$T'(z) = - \int \frac{t}{(z-t)^2} d\mu(t) < 0.$$

So between the two solutions  $T_1(i)$  and  $T_2(i)$ , only  $T_1(i)$  satisfies this condition. And due to the fact that  $\lambda_i > b$ , the region of  $T(\lambda_i)$  is  $(0, T(b^+))$ , therefore, the condition that there exists a unique solution in the region of  $(0, T(b^+))$  is  $T_1(i) \in (0, T(b^+))$ .

The proof of the theorem is complete.  $\square$

REMARK 2.2. The normal assumption in Theorem 2.1 is used to reduce an arbitrary loading matrix  $A$  satisfying  $A'A = I_k$  to its canonical form as explained in Step 2 of the proof. If the loading matrix is assumed to have the canonical form, this normal assumption is no more necessary.

2.3. *On the phase transition condition  $T_1(i) < T(b^+)$ .* In this section, we detail the phase transition condition  $T_1(i) < T(b^+)$  that defines the detection frontier of the factors. Unlike similar phenomenon observed for large sample covariance matrices as exposed in Baik and Silverstein (2006) and Bai and Yao (2012), this transition condition for the auto-covariance matrix has a more complex nature involving the three parameters: the limiting ratio  $y$ , the two signal-to-noise ratios (SNR)  $\gamma_0(i)/\sigma^2$  and  $\gamma_1(i)/\sigma^2$  involving the variance and lag-1 auto-covariance of the  $i$ th factor time series  $(x_{it})$ .

To begin with, we observe that the condition can be reduced to

$$(2.20) \quad \begin{aligned} & 2y \frac{\gamma_0(i)}{\sigma^2} + \left(\frac{\gamma_1(i)}{\sigma^2}\right)^2 - \left(2\left(\frac{\gamma_0(i)}{\sigma^2}\right)^2 - 2\left(\frac{\gamma_1(i)}{\sigma^2}\right)^2\right)T(b^+) \\ & < \sqrt{\left(2y \frac{\gamma_0(i)}{\sigma^2} + \left(\frac{\gamma_1(i)}{\sigma^2}\right)^2\right)^2 - 4y^2\left(\left(\frac{\gamma_0(i)}{\sigma^2}\right)^2 - \left(\frac{\gamma_1(i)}{\sigma^2}\right)^2\right)}, \end{aligned}$$

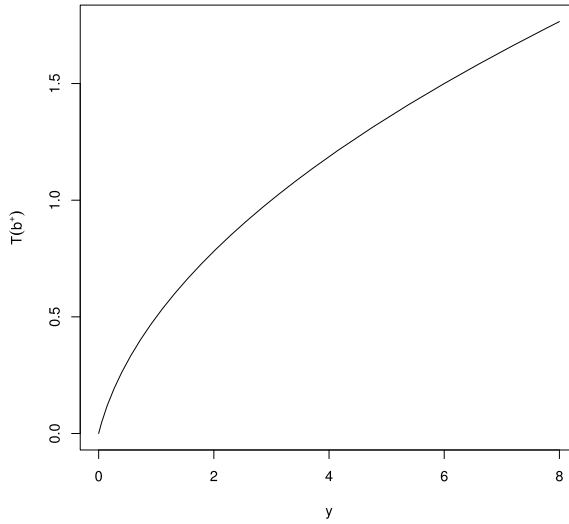


FIG. 1. Values of  $T(b^+)$  as a function of the limiting ratio  $y$ .

which leads to the following two possibilities:

$$(2.21) \quad \begin{cases} 2y \frac{\gamma_0(i)}{\sigma^2} + \left(\frac{\gamma_1(i)}{\sigma^2}\right)^2 - \left(2\left(\frac{\gamma_0(i)}{\sigma^2}\right)^2 - 2\left(\frac{\gamma_1(i)}{\sigma^2}\right)^2\right) T(b^+) > 0, \\ \left(\frac{\gamma_0(i)}{\sigma^2} T(b^+) - y\right)^2 < \left(\frac{\gamma_1(i)}{\sigma^2}\right)^2 (T^2(b^+) + T(b^+)), \end{cases}$$

or

$$(2.22) \quad 2y \frac{\gamma_0(i)}{\sigma^2} + \left(\frac{\gamma_1(i)}{\sigma^2}\right)^2 - \left(2\left(\frac{\gamma_0(i)}{\sigma^2}\right)^2 - 2\left(\frac{\gamma_1(i)}{\sigma^2}\right)^2\right) T(b^+) \leq 0.$$

First, we see the value of  $T(b^+)$  can be derived using (2.5), with the value of  $b$  given in (2.4) as a function of  $y$ , which is presented in Figure 1. When  $y$  increases from zero to infinity, the value of  $T(b^+)$  also increases from zero to infinity. Observe also that the slope at the origin is infinity:  $\lim_{y \rightarrow 0^+} T(b^+)/y = \infty$ .

Once  $p$  and  $T$  are given ( $y$  is fixed), the value of  $T(b^+)$  is fixed, then the conditions (2.21) and (2.22) can be considered as the restriction on the two parameters  $\gamma_0(i)/\sigma^2$  and  $\gamma_1(i)/\sigma^2$ . This defines a complex region in the  $\gamma_0/\sigma^2 - \gamma_1/\sigma^2$  plane as depicted in Figure 2. The dashed curve in Figure 2 stands for the equality

$$2y \frac{\gamma_0(i)}{\sigma^2} + \left(\frac{\gamma_1(i)}{\sigma^2}\right)^2 - \left(2\left(\frac{\gamma_0(i)}{\sigma^2}\right)^2 - 2\left(\frac{\gamma_1(i)}{\sigma^2}\right)^2\right) T(b^+) = 0,$$

and the area inside this curve (the darker region) is the condition (2.22), while outside (the lighter region) stands for condition (2.21). The dotted lines stand for

$$\left(\frac{\gamma_0(i)}{\sigma^2} T(b^+) - y\right)^2 = \left(\frac{\gamma_1(i)}{\sigma^2}\right)^2 (T^2(b^+) + T(b^+)),$$



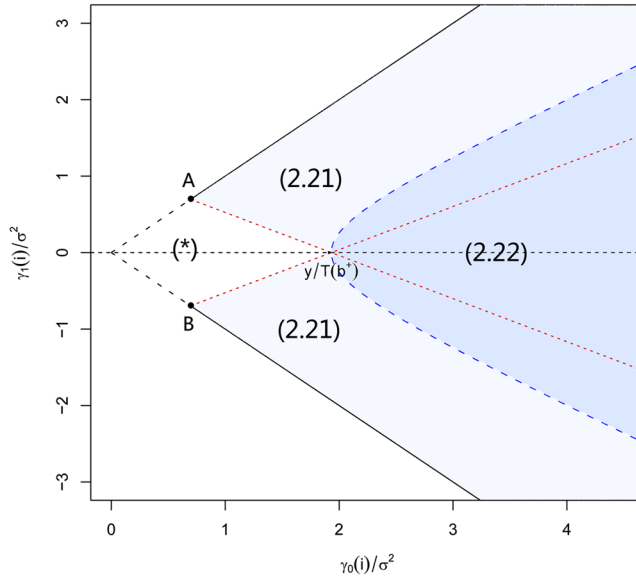


FIG. 2. Region of  $\gamma_0(i)/\sigma^2$  and  $\gamma_1(i)/\sigma^2$  that will lead to significant factors.

and the upper and lower boundaries in solid lines are due to the fact that we have always  $|\gamma_1(i)| \leq \gamma_0(i)$  (by Cauchy–Schwarz inequality). These solid and dotted lines intersect at points  $A = (\tau_0, \tau_0)$  and  $B = (\tau_0, -\tau_0)$  where

$$(2.23) \quad \tau_0 = \frac{y}{T(b^+) + \sqrt{T^2(b^+) + T(b^+)}}$$

In other words, except for the quadrilateral region (\*), our conditions (2.21) and (2.22) will hold true, which means that the corresponding factors are significant (and thus asymptotically detectable). The quadrilateral region (\*) thus defines the phase transition boundary for the significance of the factors.

We summarize the above findings as follows.

**COROLLARY 2.2.** *Under the same conditions as in Theorem 2.1, the  $i$ th time series  $(x_{it})$  will generate a significant factor in the sense that  $T_1(i) < T(b^+)$  if and only if either*

$$(2.24) \quad \frac{|\gamma_1(i)|}{\sigma^2} > \tau_0,$$

or

$$(2.25) \quad \frac{|\gamma_1(i)|}{\sigma^2} \leq \tau_0 \quad \text{and} \quad \frac{\gamma_0(i)}{\sigma^2} > \frac{y - \sqrt{T^2(b^+) + T(b^+)}|\gamma_1(i)|/\sigma^2}{T(b^+)},$$

where the constant  $\tau_0$  is given in (2.23).

We now introduce some important comments on the meaning of these conditions.

1. The essential message from these conditions is that *the  $i$ th factor time series is a significant factor once its strength  $\gamma_0(i)$ , or more exactly, its SNR  $\gamma_0(i)/\sigma^2$  exceeds a certain level  $\tau$* . A sufficient value for this level  $\tau$  is  $\tau_1 = y/T(b^+)$  as shown in Figure 2. Meanwhile, the SNR should at least equal to  $\tau_0$  given in (2.23). See Point A on the figure with coordinates  $(\tau_0, \tau_0)$ . When  $\tau_0 < \gamma_0(i)/\sigma^2 \leq \tau_1$ , the exact condition also depends on the lag-1 SNR  $|\gamma_1(i)|/\sigma^2$  as given in equations (2.24)–(2.25).

This is much in line with what is known for the phase transition phenomenon for large sample covariance matrices as exposed in Baik and Silverstein (2006) and Bai and Yao (2012).

2. As said in the Introduction, in most of existing literature on high-dimensional factor models, the factor strengths are assumed to grow to infinity with the dimension  $p$ . Clearly, such *pervasive factors* are highly significant in our scheme, that is,  $k_0 = k$ , since they will exceed the upper limit  $\tau_1$  very quickly as the dimension  $p$  grows.

3. Assume that  $y \rightarrow 0_+$ , that is, the sample size  $T$  is much larger than the dimension  $p$ . Then it can be checked that both the quantities  $\tau_0$  and  $\tau_1$  will vanish. Therefore, when  $p/T$  is small enough, any factor time series will generate a significant sample factor eigenvalue. In other words, we have recovered the classical low-dimensional situation where  $p$  is hold fixed and  $T \rightarrow \infty$  for which all the  $k$  factor time series can be consistently detected and identified.

**3. Estimation of the number of factors.** Let  $l_1, \dots, l_p$  be the eigenvalues of  $\widehat{M} = \widehat{\Sigma}_y \widehat{\Sigma}'_y$ , sorted in decreasing order. Assume that among the  $k$  factors, the first  $k_0$  are significant which satisfy the phase transition condition  $T_1(i) < T(b^+)$ ; see equation (2.8). Following Theorem 2.1, the  $k$  largest sample eigenvalue  $(l_i/\sigma^4)_{1 \leq i \leq k}$  converges respectively to a limit  $(\lambda_i)$ , which is larger than the right edge  $b$  of the limiting spectral distribution for  $1 \leq i \leq k_0$ , and equal to  $b$  for  $k_0 < i \leq k$ .

It will be proven below that the largest noise sample eigenvalues of a given finite number all converge to  $b$ , that is, for any fixed range  $m > 0$ ,

$$(3.1) \quad l_{k+1}/\sigma^4 \rightarrow b, \dots, l_{k+m}/\sigma^4 \rightarrow b \quad \text{almost surely.}$$

Consider the sequence of ratios

$$(3.2) \quad \theta_j := \frac{l_{j+1}/\sigma^4}{l_j/\sigma^4} = \frac{l_{j+1}}{l_j}, \quad j \geq 1.$$

By definition  $\theta_j \leq 1$ . Therefore, we have almost surely,

$$\theta_j \rightarrow \frac{\lambda_{j+1}}{\lambda_j} < 1, \quad j < k_0,$$

$$\begin{aligned}
 (3.3) \quad & \theta_{k_0} \rightarrow \frac{b}{\lambda_{k_0}/\sigma^4} < 1, \\
 & \theta_j \rightarrow \frac{b}{b} = 1, \quad k_0 < j \leq k, \\
 & \theta_{k+1}, \dots, \theta_{k+m} \rightarrow \frac{b}{b} = 1 \quad \text{for all fixed } m.
 \end{aligned}$$

REMARK 3.1. Note that the value of  $\theta_j$  is independent of  $\sigma^2$ . In other words, we do not need any estimate of  $\sigma^2$  for estimating the number of factors.

Let  $0 < d_T < 1$  be a positive constant and we introduce the following estimator for the number of factors  $k$ :

$$(3.4) \quad \widehat{k} = \{\text{first } j \geq 1 \text{ such that } \theta_j > 1 - d_T\} - 1.$$

THEOREM 3.1. Consider the factor model (1.1) and assume that the same conditions as in Theorem 2.1 are satisfied. Let  $k_0$  be the number of significant factors defined in equation (2.8) and a threshold constant  $d_T$  be chosen such that

$$(3.5) \quad \max_{1 \leq j \leq k_0} \lambda_{j+1}/\lambda_j < 1 - d_T < 1.$$

Then  $\widehat{k} \xrightarrow{\text{a.s.}} k_0$ .

This theorem thus formally establishes the fact that the ratio estimator  $\widehat{k}$  is able to detect all the significant factors that satisfy the phase transition condition given in Theorem 2.1 and detailed in equations (2.24)–(2.25).

PROOF OF THEOREM 3.1. As  $\theta_j \xrightarrow{\text{a.s.}} \lambda_{j+1}/\lambda_j$  for  $1 \leq j \leq k_0$  and by assumption (3.5), almost surely, it will happen eventually that  $\widehat{k} > k_0$ . Next, under the claim (3.1) and following the limits given in (3.3),

$$(3.6) \quad \theta_j \xrightarrow{\text{a.s.}} 1 \quad \text{for } j > k_0.$$

Consequently, we will eventually have  $\widehat{k} \leq k_0$  almost surely. When this is combined with the conclusion above, the almost sure convergence of  $\widehat{k}$  to  $k_0$  is proven.

It remains to prove the claim (3.1). Since  $\theta_j$  is independent of the choice of  $\sigma^2$ , we can assume w.l.o.g. that  $\sigma^2 = 1$  as before. Recall that in the proof of Theorem 2.1, it has been proved in equations (2.13)–(2.14) that if  $l$  is a eigenvalue of  $\widehat{M}$ , then  $\sqrt{l}$  is a positive eigenvalue of the matrix

$$\Gamma = \begin{pmatrix} 0 & X'_1 X_0 & 0 & X'_1 E_1 \\ X'_0 X_1 & 0 & X'_0 E_2 & 0 \\ 0 & E'_2 X_0 & 0 & E'_2 E_1 \\ E'_1 X_1 & 0 & E'_1 E_2 & 0 \end{pmatrix},$$

which is obtained after permutation of the second and third row block and column block in (2.14) without modifying the eigenvalues. Now  $\Gamma$  is a symmetric block matrix and the positive eigenvalues of the lower diagonal block

$$\begin{pmatrix} 0 & E'_2 E_1 \\ E'_1 E_2 & 0 \end{pmatrix},$$

are associated to the eigenvalues of the matrix  $DD' = E'_2 E_1 E'_1 E_2$  which is of dimension  $p - k$  (for the definition of these matrices, see that proof). Let  $\beta_1 \geq \dots \geq \beta_{p-k}$  be the eigenvalues of  $DD'$ . By the Cauchy interlacing theorem, we have

$$\beta_{k+1} \leq l_{k+1} \leq \beta_1.$$

Observing that  $D$  is distributed as  $\widehat{\Sigma}_\varepsilon$  except that the dimension is changed from  $p$  to  $p - k$ . Therefore, the global limit of the eigenvalues of  $DD'$  are the same as for the matrix  $\widehat{\Sigma}_\varepsilon \widehat{\Sigma}'_\varepsilon$ ; in particular, according to Corollary 2.1, both  $\beta_{k+1}$  and  $\beta_1$  converge to  $b$  almost surely. This proves the fact that  $l_{k+1} \xrightarrow{\text{a.s.}} b$ . Using similar arguments, we can establish the same fact for  $l_{k+j} \xrightarrow{\text{a.s.}} b$  for any fixed index  $j \geq 1$ . The claim (3.1) is thus established.  $\square$

3.1. *Calibration of the tuning parameter  $d_T$ .* For the estimator  $\widehat{k}$  in (3.4) to be practically useful, we need to set up an appropriate value of the tuning parameter  $d_T$ . Although any vanishing sequence  $d_T \rightarrow 0$  will theoretically guarantee the consistency of  $\widehat{k}$ , it is preferable to have an indicated and practically useful sequence  $(d_T)$  for real-life data analysis. Here, we propose an *a priori* calibration of  $d_T$  based on some knowledge from random matrix theory on the largest eigenvalues of sample covariance matrices and of their perturbed versions. The most important property we will use is that according to such recent results on finite rank perturbations of symmetric random matrices; see, for example, Benaych-Georges, Guionnet and Maida (2011). It is very likely that the asymptotic distribution of  $T^{\frac{2}{3}} \left( \frac{l_{k+2}}{l_{k+1}} - 1 \right)$  is the same as that of  $T^{\frac{2}{3}} \left( \frac{v_2}{v_1} - 1 \right)$ , where  $v_1, v_2$  are the two largest eigenvalues of the base noise matrix  $\widehat{M}_\varepsilon$ . Using this similarity, we calibrate  $d_T$  by simulation: for any given pair  $(p, T)$ , the distribution of  $T^{\frac{2}{3}} \left( \frac{v_2}{v_1} - 1 \right)$  is sampled using a large number (in fact 2000) of independent replications of standard Gaussian vectors  $\varepsilon_T \sim N(\mathbf{0}, \mathbf{I}_p)$  and its lower 0.5% quantile  $q_{p,T,0.5\%}$  is obtained (notice that the quantile is negative). Using the approximation

$$P \left\{ T^{\frac{2}{3}} \left( \frac{l_{k+2}}{l_{k+1}} - 1 \right) \leq q_{p,T,0.5\%} \right\} \simeq P \left\{ T^{\frac{2}{3}} \left( \frac{v_2}{v_1} - 1 \right) \leq q_{p,T,0.5\%} \right\} = 0.5\%,$$

we calibrate  $d_T$  at the value  $d_T = |q_{p,T,0.5\%}|/T^{2/3}$ . Notice that  $d_T$  vanishes at a rate of  $T^{-2/3}$ . Overall, this tuned value of  $d_T$  is used for all the given pairs of  $(T, p)$  in the simulation experiments in Section 4 and in the data analysis reported in Section 5.

**4. Monte-Carlo experiments.** In this section, we report some simulation results to show the finite-sample performance of our estimator. For the sake of robustness, we will consider a reinforced estimator  $\hat{k}^*$  defined as

$$(4.1) \quad \hat{k}^* = \{\text{first } j \geq 1 \text{ such that } \theta_j > 1 - d_T \text{ and } \theta_{j+1} > 1 - d_T\} - 1.$$

Clearly,  $\hat{k}^*$  is asymptotically equivalent to the initial estimator  $\hat{k}$  which uses only one single test value  $j$ . As for the factor model, we adopt the same settings as in Lam and Yao (2012) where

$$\begin{aligned} y_t &= Ax_t + \varepsilon_t, & \varepsilon_t &\sim N_p(\mathbf{0}, \mathbf{I}_p), \\ x_t &= \Theta x_{t-1} + e_t, & e_t &\sim N_k(\mathbf{0}, \Gamma), \end{aligned}$$

where  $A$  is a  $p \times k$  matrix, w.l.o.g., we set the variance  $\sigma^2$  of the white noise  $\varepsilon_t$  to 1.

In Lam and Yao (2012), the factor loading matrix  $A$  are independently generated from uniform distribution on the interval  $[-1, 1]$  first and then divided by  $p^{\delta/2}$  where  $\delta \in [0, 1]$ . The induced  $k$  factor strengths are thus of order  $O(p^{1-\delta})$ . Their estimator of the number of factors is recalled in (1.4). Cases where three factors are either all very strong with  $\delta = 0$  or all moderately strong with  $\delta = 0.5$  are discussed in detail in that paper. The results show that  $\tilde{k}$  performs better when factors are stronger. An experimental setting with a combination of two strong factors and one moderate factor indicates that a two-step estimation procedure needs to be employed in order to identify all three factors. In each step, only factors with the highest level of strength can be detected. While in our case, the coefficient matrix  $A$  satisfies  $A'A = I_k$ . Considering the eigenvalues of  $\widehat{M}$  are invariant under orthogonal transformation (See Step 2 in the proof of Theorem 2.1), we fix

$$A = \begin{pmatrix} I_k \\ \mathbf{0}_{p-k} \end{pmatrix}.$$

Then we manipulate the factor strength by adjusting the value of  $\Theta$  and  $\Gamma$ . To ensure the stationarity of the  $\{y_t\}$  process and the independence among the components of the factor process  $\{x_t\}$ ,  $\Theta$  and  $\Gamma$  are both diagonal and the diagonal elements of  $\Theta$  belong to  $(-1, 1)$ . To keep pace with the settings in Lam and Yao (2012), we multiply the diagonal entries of  $\Gamma$  by  $p^{\frac{1-\delta}{2}}$  in order to adjust the corresponding factor strength. It can be seen that when  $\delta = 0$ , the factor is strongest while with  $\delta = 1$ , the factor is weakest.

The simulation study comprises four parts corresponding to the four scenarios defined as follows:

- (I) Two very strong factors with  $\delta_1 = 0.5$  and  $\delta_2 = 0.8$  and

$$\Theta = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.5 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 4 \times p^{\frac{1-\delta_1}{2}} & 0 \\ 0 & 4 \times p^{\frac{1-\delta_2}{2}} \end{pmatrix}.$$

(II) Four weak factors with same strength level  $\delta = 1$ ; three of them are significant with their theoretical limits  $\lambda_1, \lambda_2, \lambda_3$  all keeping a moderate distance from the right edge  $b$  of the noise eigenvalues while the fourth factor is insignificant with its theoretical limit  $\lambda_4$  equal to  $b$ . Precisely,

$$\Theta = \begin{pmatrix} 0.6 & 0 & 0 & 0 \\ 0 & -0.5 & 0 & 0 \\ 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0.2 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

(III) Three weak factors with  $\delta = 1$  and  $\lambda_3$  stays very close to  $b$  and

$$\Theta = \begin{pmatrix} 0.6 & 0 & 0 \\ 0 & -0.5 & 0 \\ 0 & 0 & 0.3 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

(IV) A mixed case with two strong factors with  $\delta_1 = 0.5, \delta_2 = 0.8$ , and five weak factors with  $\delta = 1$ , and

$$\Theta = \text{diag}(0.6, 0.5, 0.6, -0.5, 0.3, 0.6, -0.5),$$

$$\Gamma = \text{diag}(4 \times p^{\frac{1-\delta_1}{2}}, 4 \times p^{\frac{1-\delta_2}{2}}, 4, 4, 4, 2, 2).$$

Recall that for the estimator  $\widehat{k}^*$ , the critical value  $d_T$  is calibrated as explained in Section 3.1 using the simulated empirical 0.5% lower quantile. We set  $p = 100, 300, 500, 1000, 1500, T = 0.5p, 2p$ , that is,  $y = 2, 0.5$ . It will be seen below that in general, the cases with  $T = 0.5p$  will be harder to deal with than the cases with  $T = 2p$ . We repeat this 1000 times to calculate the empirical frequencies of the different decisions  $(\widehat{k}^* = k_0), (\widehat{k}^* = k_0 \pm 1)$  and  $(|\widehat{k}^* - k_0| > 1)$ . The results are as follows.

(I) In Scenario I, we have two very strong factors with  $\delta_1 = 0.5$  and  $\delta_2 = 0.8$  and their strengths grow to infinity with  $p$ . Thus,  $k_0 = k = 2$  and the two factors must be easily detectable. As seen from Table 1, our estimator  $\widehat{k}^*$  quickly converges to the true number of factors. On the other hand, the one-step estimator  $\tilde{k}$  of Lam and Yao (2012) tends to detect only one factor in each step due to the fact that the two factors are of different strength.

(II) In Scenario II, we have four weak factors of same strength level  $\delta = 1$ . The theoretical limits related to Theorem 2.1 are displayed in Table 2. Figure 3 for  $T = 2p$  and Figure 4 for  $T = 0.5p$  depict the position of these four factors (numbered from 1 to 4) in the phase transition diagram defined in Corollary 2.2, and we see three among the four lying inside the detectable area in both situations. It can be seen from the table that for both combinations of  $T = 2p$  and  $T = 0.5p$ , the first three limits  $\lambda_i$  are far from the right edge  $b$  and the fourth limit  $\lambda_4$  equals to  $b$ . We thus have three significant factors ( $k_0 = 3$ ) which are detectable while the fourth one is too weak for the detection. Results in Table 3 show that both the estimators  $\tilde{k}$  (one-step) and  $\widehat{k}^*$  are consistent with a much higher convergence speed for  $\widehat{k}^*$ .

TABLE 1  
Scenario I with two strong factors ( $k_0 = k = 2$ )

$p$	100	300	500	1000	1500	$p$	100	300	500	1000	1500
$T = 2p$	200	600	1000	2000	3000	$T = 2p$	200	600	1000	2000	3000
$\tilde{k} = 1$	0.343	0.294	0.257	0.287	0.317	$\hat{k}^* = 1$	0	0	0	0	0
$\tilde{k} = k_0$	<b>0.657</b>	<b>0.706</b>	<b>0.743</b>	<b>0.713</b>	<b>0.683</b>	$\hat{k}^* = k_0$	<b>0.974</b>	<b>0.984</b>	<b>0.993</b>	<b>0.996</b>	<b>0.998</b>
$\tilde{k} \geq 3$	0	0	0	0	0	$\hat{k}^* \geq 3$	0.026	0.016	0.007	0.004	0.002
$p$	100	300	500	1000	1500	$p$	100	300	500	1000	1500
$T = 0.5p$	50	150	250	500	750	$T = 0.5p$	50	150	250	500	750
$\tilde{k} = 1$	0.786	0.801	0.876	0.96	0.992	$\hat{k}^* = 1$	0.086	0	0	0	0
$\tilde{k} = k_0$	<b>0.21</b>	<b>0.199</b>	<b>0.124</b>	<b>0.04</b>	<b>0.008</b>	$\hat{k}^* = k_0$	<b>0.771</b>	<b>0.882</b>	<b>0.896</b>	<b>0.881</b>	<b>0.881</b>
$\tilde{k} \geq 3$	0.004	0	0	0	0	$\hat{k}^* \geq 3$	0.143	0.118	0.104	0.119	0.119

(III) Theoretical limits and empirical results for Scenario III are presented in Table 4, Figures 3 and 4 and Table 5. For both situations of  $T = 0.5p$  and  $T = 2p$ , the model has three significant factors ( $k_0 = k = 3$ ). Notice, however, that when  $T = 0.5p$ , the third factor is quite weak and the corresponding limit  $\lambda_3 = 17.95$  is very close to the right edge  $b = 17.64$  so that this factor would be detectable only in theory (or with very large sample sizes). This is also easily verified in Figure 4 that the point (3) corresponding to the weakest factor lies very close to the boundary of the detectable region. As for the empirical values in Table 5, the estimator  $\hat{k}^*$  converges quickly when  $T = 2p$  and much more slowly when  $T = 0.5p$ . Meanwhile, the estimator  $\tilde{k}$  (with one-step) seems inconsistent even in the easier case of  $T = 2p$ .

(IV) Scenario IV is the most complex case with two very strong factors and five weak factors. As predicted by the theory, the two largest factor eigenvalues  $l_1, l_2$  of  $\bar{M}$  blow up to infinity while the following 5 factor eigenvalues  $l_3 \sim l_7$  converge to a  $\lambda_i > b$ . The corresponding theoretical limits for the five weak factors are given in Table 6 and their SNRs are depicted in Figures 3 and 4. Meanwhile, all the  $k_0 = k = 7$  factors are significant. Clearly, in this scenario, the performance of the

TABLE 2  
Scenario II—Theoretical limits ( $k_0 = 3, k = 4$ )

No.	$\Theta$	$\Gamma$	$\gamma_0(i)$	$\gamma_1(i)$	$T = 2p$				$T = 0.5p$			
					$T_1(i)$	$T(b^+)$	$\lambda_i$	$b$	$T_1(i)$	$T(b^+)$	$\lambda_i$	$b$
(1)	0.6	4	6.25	3.75	0.0125	0.3076	21.2	2.7725	0.1102	0.7775	44.8	17.6366
(2)	-0.5	4	5.33	-2.67	0.021	0.3076	13.1	2.7725	0.1596	0.7775	33.85	17.6366
(3)	0.3	4	4.3956	1.3187	0.047	0.3076	6.65	2.7725	0.2767	0.7775	23.92	17.6366
(4)	0.2	1	1.042	0.2083	0.3446	0.3076	<b>2.7725</b>	<b>2.7725</b>	1.5296	0.7775	<b>17.6366</b>	<b>17.6366</b>

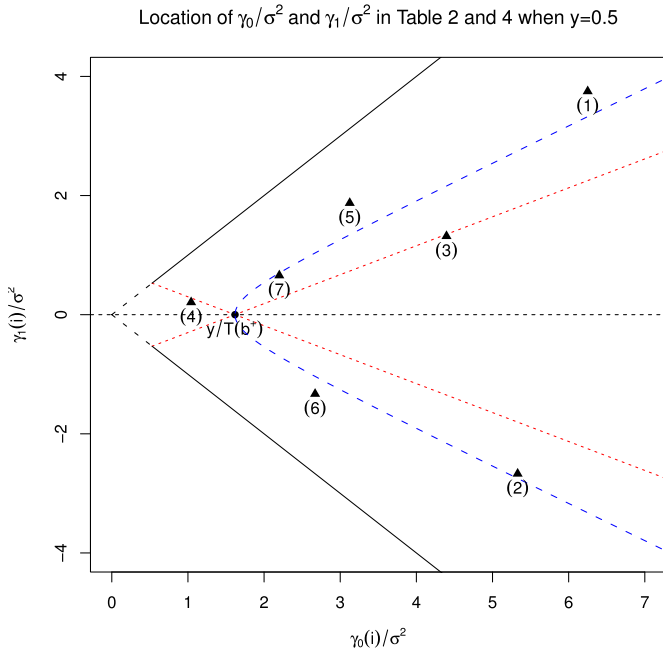


FIG. 3. Locations of factor SNR's  $(\gamma_0/\sigma^2, \gamma_1/\sigma^2)$  from Tables 2 (points numbered from 1 to 4), 4 (points numbered from 5 to 7), and 6 (points numbered 1-2-3-5-6) with  $T = 2p$  ( $y = 0.5$ ).

one-step estimator  $\tilde{k}$ , denoted as  $\tilde{k}^{(1)}$ , is quite limited and in order to make a closer comparison with our estimator  $\hat{k}^*$ , we have also run the two-step and the three-step versions of the estimator  $\tilde{k}$ . Among these two versions, we report the best results obtained by the three-step version (denoted as  $\tilde{k}^{(3)}$ ). It can be seen from Table 7 that our estimator is able to detect the 7 factors with multi-level strength in a single step while  $\tilde{k}$  can only identify one factor in each step, that is,  $\tilde{k}^{(1)} \rightarrow 1$  and  $\tilde{k}^{(3)} \rightarrow 3$ .

**5. An example of real data analysis.** We analyze the log returns of 100 stocks (denoted by  $y_t$ ), included in the S&P500 during the period from 2005-01-03 to 2011-09-16. We have in total  $T = 1689$  observations with  $p = 100$ . Thorough eigenvalue analysis is applied to the lag-1 sample auto-covariance matrix  $\widehat{M} = \widehat{\Sigma}_y \widehat{\Sigma}'_y$  of  $y_t$ . The largest eigenvalue of  $\widehat{M}$  is  $\lambda_1(\widehat{M}) = 38.69$ . The second to the 30th largest eigenvalues and their ratios are plotted in Figure 5.

To estimate the number of factors, we first adopt the two-step procedure investigated by Lam and Yao (2012) since the ratio plot in Figure 5 is exhibiting at least two different levels of factor strength. Obviously, in the first step,

$$\hat{r}_1 = \arg \min_{1 \leq i \leq 99} \lambda_{i+1}/\lambda_i = 1,$$



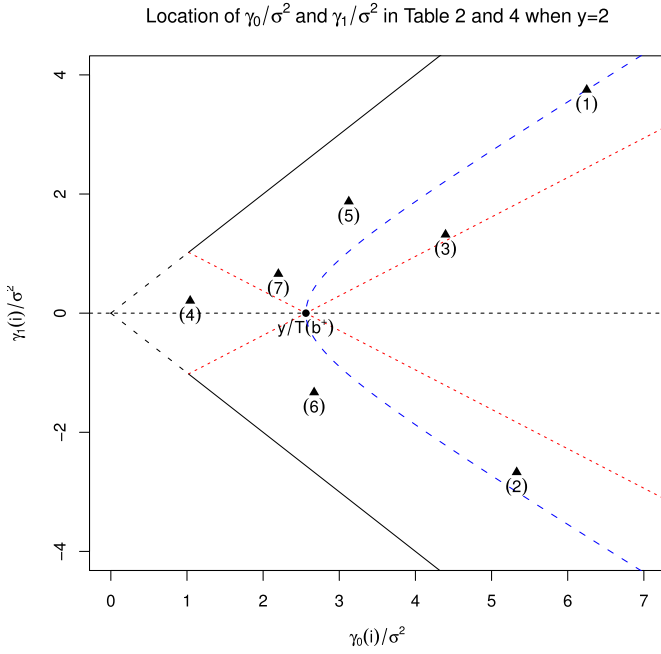


FIG. 4. Locations of factor SNRs  $(\gamma_0, \gamma_1)/\sigma^2$  from Tables 2 (points numbered from 1 to 4), 4 (points numbered from 5 to 7), and 6 (points numbered 1-2-3-5-6) with  $T = 0.5p$  ( $y = 2$ ).

TABLE 3  
Scenario II with three weak yet significant factors among four ( $k_0 = 3, k = 4$ )

$p$	100	300	500	1000	1500	$p$	100	300	500	1000	1500
$T = 2p$	200	600	1000	2000	3000	$T = 2p$	200	600	1000	2000	3000
$\tilde{k} = 1$	0.152	0.074	0.045	0.01	0.001	$\hat{k}^* = 1$	0.005	0	0	0	0
$\tilde{k} = 2$	0.402	0.344	0.276	0.194	0.126	$\hat{k}^* = 2$	0.026	0	0	0	0
$\tilde{k} = k_0$	<b>0.446</b>	<b>0.582</b>	<b>0.679</b>	<b>0.796</b>	<b>0.873</b>	$\hat{k}^* = k_0$	<b>0.928</b>	<b>0.967</b>	<b>0.953</b>	<b>0.96</b>	<b>0.966</b>
$\tilde{k} = 4$	0	0	0	0	0	$\hat{k}^* = 4$	0.04	0.033	0.046	0.04	0.033
$\tilde{k} \geq 5$	0	0	0	0	0	$\hat{k}^* \geq 5$	0.001	0	0.001	0	0.001
$p$	100	300	500	1000	1500	$p$	100	300	500	1000	1500
$T = 0.5p$	50	150	250	500	750	$T = 0.5p$	50	150	250	500	750
$\tilde{k} = 1$	0.479	0.368	0.344	0.284	0.289	$\hat{k}^* = 1$	0.376	0.02	0.003	0	0
$\tilde{k} = 2$	0.406	0.432	0.454	0.495	0.514	$\hat{k}^* = 2$	0.456	0.221	0.048	0.001	0
$\tilde{k} = k_0$	<b>0.105</b>	<b>0.199</b>	<b>0.202</b>	<b>0.221</b>	<b>0.197</b>	$\hat{k}^* = k_0$	<b>0.16</b>	<b>0.73</b>	<b>0.915</b>	<b>0.986</b>	<b>0.982</b>
$\tilde{k} = 4$	0.006	0.001	0	0	0	$\hat{k}^* = 4$	0.008	0.029	0.03	0.013	0.017
$\tilde{k} \geq 5$	0.004	0	0	0	0	$\hat{k}^* \geq 5$	0	0	0.004	0	0.001

TABLE 4  
Scenario III—Theoretical limits ( $k_0 = k = 3$ )

No.	$\Theta$	$\Gamma$	$\gamma_0(i)$	$\gamma_1(i)$	$T = 2p$				$T = 0.5p$			
					$T_1(i)$	$T(b^+)$	$\lambda_i$	$b$	$T_1(i)$	$T(b^+)$	$\lambda_i$	$b$
(5)	0.6	2	3.125	1.875	0.0391	0.3076	7.65	2.7725	0.2845	0.7775	23.79	17.6366
(6)	-0.5	2	2.67	-1.33	0.0607	0.3076	5.48	2.7725	0.3852	0.7775	20.45	17.6366
(7)	0.3	2	2.20	0.659	0.1183	0.3076	3.61	2.7725	0.6116	0.7775	<b>17.95</b>	<b>17.6366</b>

TABLE 5  
Scenario III with three weak yet insignificant factors ( $k_0 = k = 3$ )

$p$	100	300	500	1000	1500	$p$	100	300	500	1000	1500
$T = 2p$	200	600	1000	2000	3000	$T = 2p$	200	600	1000	2000	3000
$\tilde{k} < 2$	0.403	0.322	0.327	0.302	0.308	$\hat{k}^* < 2$	0.074	0	0	0	0
$\tilde{k} = 2$	0.454	0.587	0.598	0.653	0.669	$\hat{k}^* = 2$	0.441	0.047	0.005	0	0
$\tilde{k} = k_0$	<b>0.143</b>	<b>0.091</b>	<b>0.075</b>	<b>0.045</b>	<b>0.023</b>	$\hat{k}^* = k_0$	<b>0.48</b>	<b>0.945</b>	<b>0.991</b>	<b>0.996</b>	<b>0.999</b>
$\tilde{k} \geq 4$	0	0	0	0	0	$\hat{k}^* \geq 4$	0.005	0.008	0.004	0.004	0.001
$p$	100	300	500	1000	1500	$p$	100	300	500	1000	1500
$T = 0.5p$	50	150	250	500	750	$T = 0.5p$	50	150	250	500	750
$\tilde{k} < 2$	0.548	0.57	0.589	0.548	0.547	$\hat{k}^* < 2$	0.886	0.639	0.435	0.114	0.049
$\tilde{k} = 2$	0.264	0.359	0.371	0.437	0.447	$\hat{k}^* = 2$	0.107	0.338	0.508	0.718	0.745
$\tilde{k} = k_0$	<b>0.08</b>	<b>0.053</b>	<b>0.036</b>	<b>0.015</b>	<b>0.006</b>	$\hat{k}^* = k_0$	<b>0.006</b>	<b>0.022</b>	<b>0.057</b>	<b>0.167</b>	<b>0.205</b>
$\tilde{k} \geq 4$	0.108	0.018	0.004	0	0	$\hat{k}^* \geq 4$	0.001	0.001	0	0.001	0.001

TABLE 6  
Scenario IV—Theoretical limits ( $k_0 = k = 7$ )

No.	$\Theta$	$\Gamma$	$\gamma_0(i)$	$\gamma_1(i)$	$T = 2p$				$T = 0.5p$			
					$T_1(i)$	$T(b^+)$	$\lambda_i$	$b$	$T_1(i)$	$T(b^+)$	$\lambda_i$	$b$
(1)	0.6	4	6.25	3.75	0.0125	0.3076	21.2	2.7725	0.1102	0.7775	44.8	17.6366
(2)	-0.5	4	5.33	-2.67	0.021	0.3076	13.1	2.7725	0.1596	0.7775	33.85	17.6366
(3)	0.3	4	4.3956	1.3187	0.047	0.3076	6.65	2.7725	0.2767	0.7775	23.92	17.6366
(5)	0.6	2	3.125	1.875	0.0391	0.3076	7.65	2.7725	0.2845	0.7775	23.79	17.6366
(6)	-0.5	2	2.67	-1.33	0.0607	0.3076	5.48	2.7725	0.3852	0.7775	20.45	17.6366

the factor loading estimator of the first factor  $\hat{A}$  is the eigenvector of  $\hat{M}$  which corresponds to the largest eigenvalue  $\lambda_1$ . The resulting residuals after eliminating the effect of the first factor is

$$\hat{\varepsilon}_t = (\mathbf{I}_{100} - \hat{A}\hat{A}')y_t.$$

TABLE 7  
*Scenario IV with seven factors of multiple strength levels ( $k_0 = k = 7$ )*

$p$	100	300	500	1000	1500	$p$	100	300	500	1000	1500
$T = 2p$	200	600	1000	2000	3000	$T = 0.5p$	50	150	250	500	750
$\tilde{k}^{(1)} = 1$	0.696	0.858	0.949	0.995	1	$\tilde{k}^{(1)} = 1$	0.73	0.812	0.881	0.95	0.986
$\tilde{k}^{(1)} = 2$	0.244	0.137	0.051	0.005	0	$\tilde{k}^{(1)} = 2$	0.211	0.177	0.118	0.05	0.014
$\tilde{k}^{(1)} = 3$	0.033	0.004	0	0	0	$\tilde{k}^{(1)} = 3$	0.039	0.011	0.001	0	0
$\tilde{k}^{(1)} = 4$	0.019	0.001	0	0	0	$\tilde{k}^{(1)} = 4$	0.015	0	0	0	0
$\tilde{k}^{(1)} = 5$	0.005	0	0	0	0	$\tilde{k}^{(1)} = 5$	0.004	0	0	0	0
$\tilde{k}^{(1)} = 6$	0.002	0	0	0	0	$\tilde{k}^{(1)} = 6$	0.001	0	0	0	0
$\tilde{k}^{(1)} = k_0$	<b>0.001</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	$\tilde{k}^{(1)} = k_0$	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
$\tilde{k}^{(1)} \geq 8$	0	0	0	0	0	$\tilde{k}^{(1)} \geq 8$	0	0	0	0	0
$p$	100	300	500	1000	1500	$p$	100	300	500	1000	1500
$T = 2p$	200	600	1000	2000	3000	$T = 0.5p$	50	150	250	500	750
$\tilde{k}^{(3)} = 1$	0	0	0	0	0	$\tilde{k}^{(3)} = 1$	0	0	0	0	0
$\tilde{k}^{(3)} = 2$	0	0	0	0	0	$\tilde{k}^{(3)} = 2$	0	0	0	0	0
$\tilde{k}^{(3)} = 3$	0.691	0.875	0.945	0.998	0.999	$\tilde{k}^{(3)} = 3$	0.71	0.802	0.862	0.955	0.982
$\tilde{k}^{(3)} = 4$	0.002	0	0	0	0	$\tilde{k}^{(3)} = 4$	0	0	0	0	0
$\tilde{k}^{(3)} = 5$	0	0	0	0	0	$\tilde{k}^{(3)} = 5$	0.001	0	0	0	0
$\tilde{k}^{(3)} = 6$	0.244	0.125	0.055	0.002	0.001	$\tilde{k}^{(3)} = 6$	0.212	0.192	0.135	0.045	0.018
$\tilde{k}^{(3)} = k_0$	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	$\tilde{k}^{(3)} = k_0$	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
$\tilde{k}^{(3)} \geq 8$	0.063	0	0	0	0	$\tilde{k}^{(3)} \geq 8$	0.077	0.006	0.003	0	0
$p$	100	300	500	1000	1500	$p$	100	300	500	1000	1500
$T = 2p$	200	600	1000	2000	3000	$T = 0.5p$	50	150	250	500	750
$\hat{k}^* = 1$	0.012	0	0	0	0	$\hat{k}^* = 1$	0.151	0.01	0	0	0
$\hat{k}^* = 2$	0.031	0.001	0	0	0	$\hat{k}^* = 2$	0.25	0.038	0.01	0.001	0
$\hat{k}^* = 3$	0.034	0.002	0	0	0	$\hat{k}^* = 3$	0.28	0.065	0.027	0.003	0
$\hat{k}^* = 4$	0.062	0.015	0.006	0.001	0	$\hat{k}^* = 4$	0.254	0.227	0.107	0.022	0.007
$\hat{k}^* = 5$	0.049	0	0	0	0	$\hat{k}^* = 5$	0.06	0.384	0.295	0.035	0.002
$\hat{k}^* = 6$	0.185	0	0	0	0	$\hat{k}^* = 6$	0.005	0.231	0.414	0.34	0.138
$\hat{k}^* = k_0$	<b>0.597</b>	<b>0.939</b>	<b>0.958</b>	<b>0.95</b>	<b>0.959</b>	$\hat{k}^* = k_0$	<b>0</b>	<b>0.044</b>	<b>0.142</b>	<b>0.557</b>	<b>0.783</b>
$\hat{k}^* \geq 8$	0.03	0.043	0.036	0.049	0.041	$\hat{k}^* \geq 8$	0	0.001	0.005	0.042	0.07

Repeating the procedure in step one, we treat  $\hat{\varepsilon}$  as the original sequence  $y_t$  and get the eigenvalues  $\lambda_i^*$ s of the lag-1 sample auto-covariance matrix  $\widehat{M}^{(1)} = \widehat{\Sigma}_{\hat{\varepsilon}} \widehat{\Sigma}_{\hat{\varepsilon}}'$ . The 30 largest eigenvalues and their ratios are plotted in Figure 6.

It can be seen from the second step that

$$\widehat{r}_2 = \arg \min_{1 \leq i \leq 99} \lambda_{i+1}^* / \lambda_i^* = 2,$$

the factor loading estimator of the second level factors  $\widehat{A}^*$  are the orthonormal eigenvectors of  $\widehat{M}^{(1)}$  corresponding to the first two largest eigenvalues.

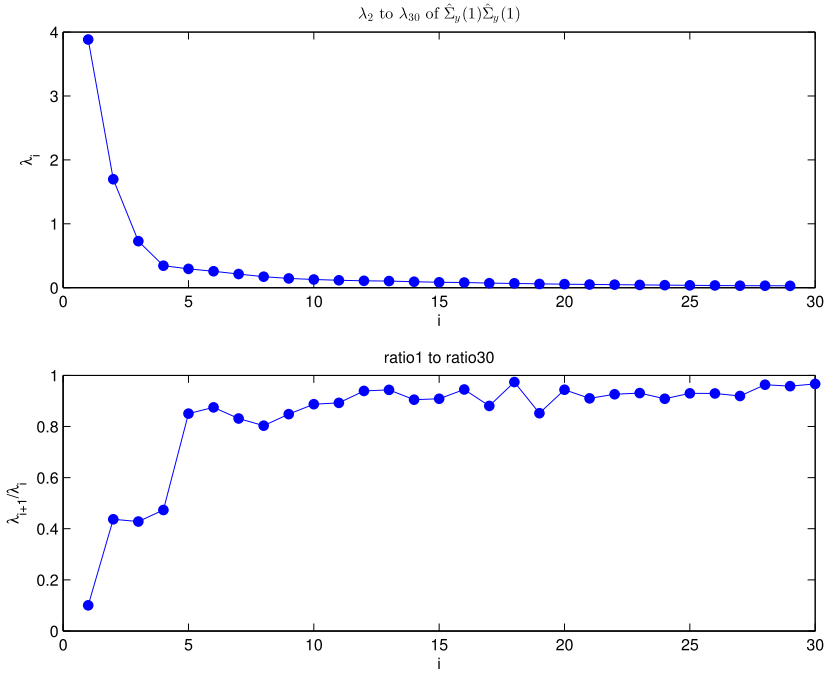


FIG. 5. Eigenvalues of  $\widehat{M}$ .

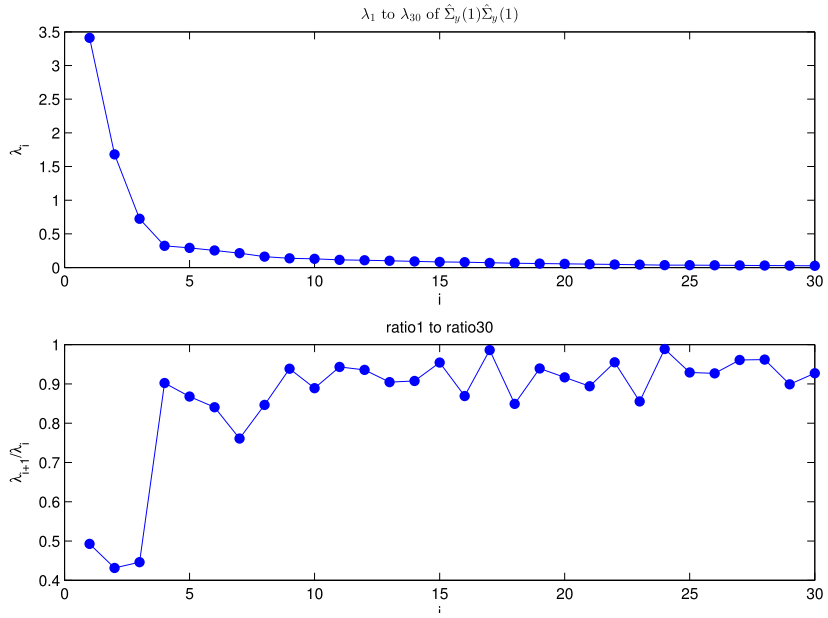


FIG. 6. Eigenvalues of  $\widehat{M}^{(1)}$ .

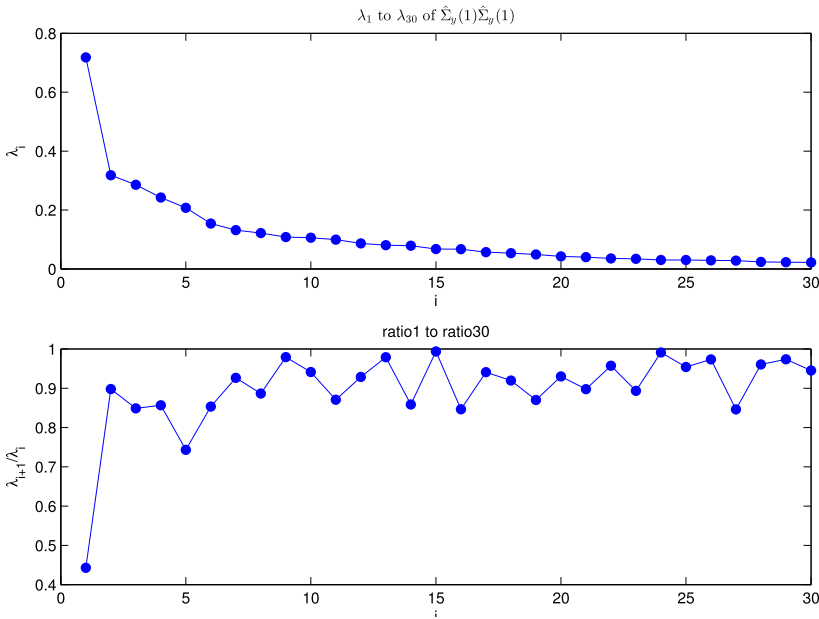


FIG. 7. Eigenvalues of  $\widehat{M}^{(2)}$ .

In conclusion, the two-step procedure proposed by Lam and Yao (2012) identifies three factors in total with two different levels of factor strength. The eigenvalues of the lag-1 sample auto-covariance matrix  $\widehat{M}^{(2)}$  of residuals after subtracting the three factors detected previously are shown in Figure 7.

There is still one isolated eigenvalue in the eigenvalues plot. If we go one step further and treat it as an extra factor with the weakest strength, then the eigenvalue plot of the lag-1 sample auto-covariance matrix  $\widehat{M}^{(3)}$  of residuals after eliminating four factors looks like that in Figure 8.

A major problem of the methodology in Lam and Yao (2012) is that it does not provide a clear criterion to stop this two or multi-step procedure. Clearly, this method can only detect factors with one level of strength at each step and can hardly handle problems with factors of multi-level strengths due to the lack of stopping criterion in multi-step detection.

In the following, we use the estimator  $\widehat{k}^*$  (4.1) of this paper to estimate the number of factors. First, the tuning parameter  $d_T$  is calibrated with  $(p, T) = (100, 1689)$  using the simulation method indicated in Section 3.1; the value found is  $d_T = 0.1713$  in this case. The eigenvalue ratios of the sample matrix  $\widehat{M}$  are shown in Figure 9 (already displayed in the lower panel of Figure 5) where the detection line of value  $1 - d_T = 0.8287$  is also drawn. As displayed, we found  $\widehat{k}^* = 4$  factors.

In conclusion, for this data set of  $p = 100$  stocks, our estimator proposes 4 significant factors while the estimator  $\tilde{k}$  from Lam and Yao (2012) indicates 1, 3 and

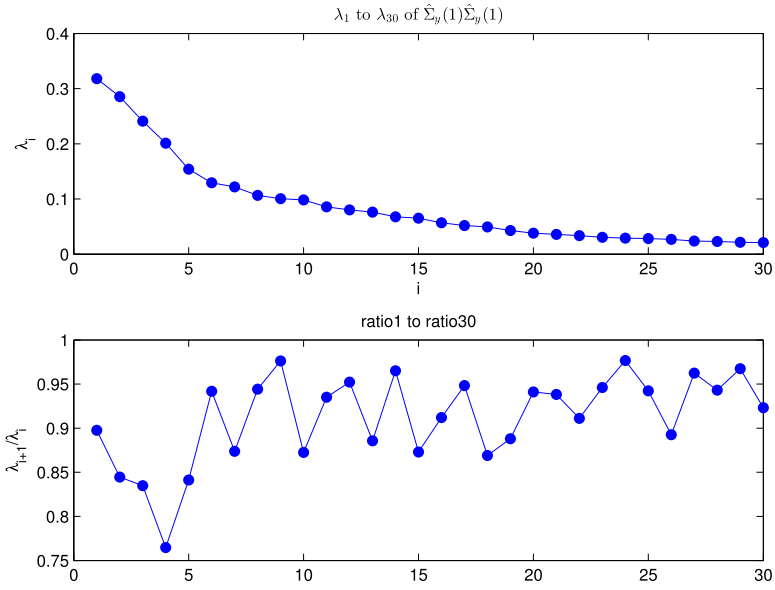


FIG. 8. Eigenvalues of  $\widehat{M}^{(3)}$ .

4 factors when one step, two steps and three steps are used, respectively. It appears again that multiple steps are needed for the use of the estimator  $\widehat{k}$  in real data analysis. However, it remains unclear how to decide the number of necessary steps. On the contrary, our estimator is able to simultaneously identify all significant factors. The procedure is independent of the number of different levels of the factor strengths.

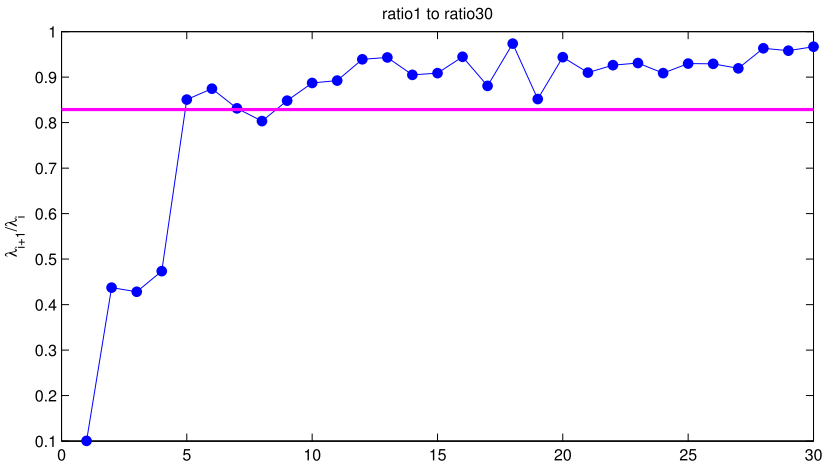


FIG. 9. Eigenvalues of  $\widehat{M}$ .

## SUPPLEMENTARY MATERIAL

**Supplement to “Identifying the number of factors from singular values of a large sample auto-covariance matrix”** (DOI: [10.1214/16-AOS1452SUPP](https://doi.org/10.1214/16-AOS1452SUPP); .pdf). A supplementary file [Li, Wang and Yao (2016)] collects several technical proofs used in the paper.

## REFERENCES

- ALESSI, L., BARIGOZZI, M. and CAPASSO, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statist. Probab. Lett.* **80** 1806–1813. [MR2734245](#)
- BAI, J. and LI, K. (2012). Statistical analysis of factor models of high dimension. *Ann. Statist.* **40** 436–465. [MR3014313](#)
- BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. [MR1926259](#)
- BAI, J. and NG, S. (2007). Determining the number of primitive shocks in factor models. *J. Bus. Econom. Statist.* **25** 52–60. [MR2338870](#)
- BAI, Z. and YAO, J. (2008). Central limit theorems for eigenvalues in a spiked population model. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 447–474. [MR2451053](#)
- BAI, Z. and YAO, J. (2012). On sample eigenvalues in a generalized spiked population model. *J. Multivariate Anal.* **106** 167–177. [MR2887686](#)
- BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. [MR2279680](#)
- BENAYCH-GEORGES, F., GUIONNET, A. and MAIDA, M. (2011). Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electron. J. Probab.* **16** 1621–1662. [MR2835249](#)
- BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. [MR2782201](#)
- FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Rev. Econ. Stat.* **82** 540–554.
- FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2004). The generalized dynamic factor model consistency and rates. *J. Econometrics* **119** 231–255. [MR2057100](#)
- FORNI, M., HALLIN, M., LIPPI, M. and REICHLIN, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *J. Amer. Statist. Assoc.* **100** 830–840. [MR2201012](#)
- GEWEKE, J. (1977). The dynamic factor analysis of economic time series. In *Latent Variables in Socio-Economic Models* (D. J. Aigner and A. S. Goldberger, eds.). North-Holland, Amsterdam.
- HALLIN, M. and LIŠKA, R. (2007). Determining the number of factors in the general dynamic factor model. *J. Amer. Statist. Assoc.* **102** 603–617. [MR2325115](#)
- JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. [MR1863961](#)
- LAM, C. and YAO, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Statist.* **40** 694–726. [MR2933663](#)
- LI, Z., PAN, G. and YAO, J. (2015). On singular value distribution of large-dimensional autocovariance matrices. *J. Multivariate Anal.* **137** 119–140. [MR3332802](#)
- LI, Z., WANG, Q. W. and YAO, J. (2016). Supplement to “Identifying the number of factors from singular values of a large sample auto-covariance matrix.” DOI:[10.1214/16-AOS1452SUPP](https://doi.org/10.1214/16-AOS1452SUPP).
- ONATSKI, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Stat.* **92** 1004–1016.
- ONATSKI, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *J. Econometrics* **168** 244–258. [MR2923766](#)

- ONATSKI, A. (2015). Asymptotic analysis of the squared estimation error in misspecified factor models. *J. Econometrics* **186** 388–406. [MR3343793](#)
- PASSEMIER, D. and YAO, J.-F. (2012). On determining the number of spikes in a high-dimensional spiked population model. *Random Matrices Theory Appl.* **1** 1150002, 19. [MR2930380](#)
- SARGENT, T. J. and SIMS, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. In *New Methods in Business Cycle Research, Vol. 1* 45–109. Federal Reserve Bank of Minneapolis, Minneapolis.
- STOCK, J. H. and WATSON, M. W. (2011). Dynamic factor models. In *The Oxford Handbook of Economic Forecasting* 35–59. Oxford Univ. Press, Oxford. [MR3204189](#)
- WANG, Q. W. and YAO, J. (2016). Moment approach for singular values distribution of a large auto-covariance matrix. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1641–1666. [MR3573290](#)

DEPARTMENT OF STATISTICS  
AND ACTUARIAL SCIENCE  
UNIVERSITY OF HONG KONG  
POKFULAM, HONG KONG SAR  
CHINA  
E-MAIL: [u3001205@hku.hk](mailto:u3001205@hku.hk)  
[wqw8813@gmail.com](mailto:wqw8813@gmail.com)  
[jeff Yao@hku.hk](mailto:jeff Yao@hku.hk)