Journal of Econometrics 216 (2020) 53–70

Contents lists available at ScienceDirect

Journal of Econometrics

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

Threshold factor models for high-dimensional time series ☆

Xialu Liu [a], Rong Chen [b,*]

[a] Management Information Systems Department, San Diego State University, San Diego, CA 92182, United States of America
[b] Department of Statistics, Rutgers University, Piscataway, NJ 08854, United States of America

ARTICLE INFO

Article history:
Available online 13 February 2020

Keywords:
Factor model
High-dimensional time series
Non-stationary process
Threshold variable

ABSTRACT

We consider a threshold factor model for high-dimensional time series in which the dynamics of the time series is assumed to switch between different regimes according to the value of a threshold variable. This is an extension of threshold modeling to a high-dimensional time series setting under a factor structure. Specifically, within each threshold regime, the time series is assumed to follow a factor model. The regime switching mechanism creates structural changes in the factor loading matrices. It provides flexibility in dealing with situations that the underlying states may be changing over time, as often observed in economic time series and other applications. We develop the procedures for the estimation of the loading spaces, the number of factors and the threshold value, as well as the identification of the threshold variable, which governs the regime change mechanism. The theoretical properties are investigated. Simulated and real data examples are presented to illustrate the performance of the proposed method.

1. Introduction

Professor George C. Tiao is one of the pioneers in the field of jointly modeling multiple time series and has made significant contributions (Tiao and Box, 1981; Tiao and Tsay, 1989). With the recent advances in data collection capability, high-dimensional time series data becomes available in many applications. Extending the work of Professor Tiao and other early researchers in analyzing multivariate time series, researchers have started to investigate models and procedures for analyzing high-dimensional time series. It is a challenging problem due to its complexity and the larger number of parameters involved. Factor analysis is an effective approach to alleviating the problem through effective dimension reduction. Specifically, let $\mathbf{y}_t$ be an observed $p \times 1$ time series $t = 1, \ldots, n$. The general form of a factor model for time series data is

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_t + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{x}_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,k_0})'$ is a set of unobserved (latent) factor time series with dimension k_0 that is much smaller than p, the matrix $\mathbf{A}$ is the loading matrix of the common factors, the term $\mathbf{A}\mathbf{x}_t$ can be viewed as the signal component of the vector time series $\mathbf{y}_t$, and $\boldsymbol{\varepsilon}_t$ is an error process or an idiosyncratic component. The dimension reduction is achieved in the sense that, under the model, the co-movement of the p-dimensional process $\mathbf{y}_t$ is driven by a much lower dimensional process $\mathbf{x}_t$. The loading matrix $\mathbf{A}$ reflects the impact of the common factors $\mathbf{x}_t$ on the observed process $\mathbf{y}_t$.

☆ Chen's research was supported in part by National Science Foundation, United States grants DMS-1513409, DMS-1737857, IIS-1741390 and CCF-1934924. The authors wish to thank the editor, two anonymous referees, James Hamilton and Yixiao Sun for their insightful comments.

* Corresponding author.
E-mail addresses: xialu.liu@sdsu.edu (X. Liu), rongchen@stat.rutgers.edu (R. Chen).

https://doi.org/10.1016/j.jeconom.2020.01.005
0304-4076/

The general dynamic factor model assumes that the latent factor process $\mathbf{x}_t$ possesses certain dynamic structure such as a vector time series structure (Geweke, 1977; Forni and Reichlin, 1998; Forni et al., 2000; Forni and Lippi, 2001; Bai and Ng, 2002; Stock and Watson, 2002; Forni et al., 2003, 2004; Stock and Watson, 2005; Hallin and Liska, 2007). It is commonly assumed that the latent factors should have an impact on most of the series (defined asymptotically). In order to differentiate the signal component from the error process, strong cross-sectional dependence is not allowed for $\{\boldsymbol{\varepsilon}_t\}$. As a consequence, the noise process $\{\boldsymbol{\varepsilon}_t\}$ may have weak serial dependence, i.e. $\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}|E(\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_s)| < C$, where C is a positive constant.

One disadvantage of the above assumptions is that the dynamic component and error process are not separable when the dimension is finite, since both of them have serial dependence. Another setting of factor models for time series data has become more popular in the literature. It assumes that the error process is white noise without serial dependence, i.e., $E(\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_s) = 0$, for $t \neq s$. Consequently the dependence of the observed process $\mathbf{y}_t$ is completely driven by the common factors (Peña and Box, 1987; Peña and Poncela, 2006; Pan and Yao, 2008; Chang et al., 2015; Liu and Chen, 2016). It ensures that the signal component is identifiable when the dimension of the panel time series is finite. In addition, the error process is allowed to have strong cross-sectional correlation. Peña and Poncela (2006) and Lam et al. (2011) developed an approach that takes advantage of information from the autocovariance matrices of the observed process at nonzero leads via eigen-decomposition to estimate the factor loading space, and they established the asymptotic properties as the dimension goes to infinity with sample size. This method is applicable to non-stationary processes, processes with uncorrelated or endogenous regressors, and matrix-valued processes; see Chang et al. (2015) and Wang et al. (2019). In this paper, we adopt these assumptions in developing the estimation procedures and the corresponding theoretical properties.

In many applications it is often observed that the loading matrix of a factor model may vary. For example, the expected market return is an important factor of the expected return of an asset, according to CAPM theory, and its impact (loading) on any individual asset is often observed to change depending on whether the stock market is volatile or stable. In economics, risk-free rate, unemployment, and economic growth are important to all economic activities and decisions. Again, the behavior of these series may vary under different fiscal policies (neutral, expansionary, or contractionary) or in different stages of the economic cycle (expansion, peak, contraction, or trough) (Kim and Nelson, 1998). Liu and Chen (2016) introduced a Markov switching mechanism to the factor model to capture the changes of the loading matrix. Although Markov regime-switching models are widely used in economics to describe the varying structure, it has the drawback of being less interpretable and difficult to forecast.

To address this limitation, we propose a threshold factor model, in which a threshold variable controls the changes of the loadings in different regimes. Such a model enhances the flexibility in modeling the underlying regime switching mechanism, and provides a more interpretable structure and an easier forecasting framework. Threshold models have been extensively studied under the general framework of autoregressive models (Tong and Lim, 1980; Chen, 1995; Tiao and Tsay, 1989; Tsay, 1998; Forbes et al., 1999), nonlinear models (Petruccelli and Davies, 1986; Gourieroux and Monfort, 1992; Tong, 1993), and non-stationary models (Zakoian, 1994; Li and Li, 1996; Balke and Fomby, 1997). In this paper we apply this powerful approach to factor models for high-dimensional time series.

Specifically, we formally introduce a threshold factor model, propose an estimation procedure for the loading spaces and the number of factors based on eigen-analysis of the cross moment matrices of the observed process, develop an objective function for the identification of the threshold value and the threshold variable, and investigate their theoretical properties. It is shown that even when the number of factors is overestimated, our estimators are still consistent. Their asymptotic properties are the same as those when the number of factors is correctly specified.

The paper focuses more on the structural change of the loading matrices and the identification of switching mechanism. We do not impose a specific dynamic structure on the factors. This makes our discussion and theory more general, but also limits the prediction ability of the model. Stationary and nonstationary structures for the factor models can be imposed, similar to those in Peña and Box (1987) and Peña and Poncela (2006), though different estimation procedures will be needed to take advantage of the structure.

The rest of the paper is organized as follows. Section 2 introduces the detailed model setting. In Section 3, estimation procedures are developed and theoretical properties of the proposed estimators are investigated. Section 4 proposes a three-step procedure for searching and identifying the threshold variable. Simulation results are presented in Section 5, and a real example is analyzed in Section 6. The regularity conditions and all detailed proofs are contained in the Appendix.

2. Threshold factor model

We consider the following two-regime threshold factor model for high-dimensional time series here. Let $\mathbf{y}_t$ be an observed $p \times 1$ time series, and $\mathbf{x}_t$ be a $k_0 \times 1$ latent factor process, $t = 1, \ldots, n$.

$$\mathbf{y}_t = \begin{cases} \mathbf{A}_1\mathbf{x}_t + \boldsymbol{\varepsilon}_{t,1} & z_t < r_0, \\ \mathbf{A}_2\mathbf{x}_t + \boldsymbol{\varepsilon}_{t,2} & z_t \geqslant r_0, \end{cases} \quad \text{and } \boldsymbol{\varepsilon}_{t,i} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{t,i}), \quad i = 1, 2, \tag{1}$$

where z_t is a partially known threshold variable, observable at time t, possibly with a small number of unknown parameters. The noise process $\{\boldsymbol{\varepsilon}_{t,1}, \boldsymbol{\varepsilon}_{t,2}\}$ is assumed to be $p \times 1$ uncorrelated noise processes. $\{\boldsymbol{\varepsilon}_{t,1}, \boldsymbol{\varepsilon}_{t,2}\}$ and $\mathbf{x}_t$ are uncorrelated with z_t given $\mathcal{F}_{-\infty}^{t-1}$, where $\mathcal{F}_i^j$ is the σ-field generated by $\{(\mathbf{x}_t, z_t) : i \leqslant t \leqslant j\}$.

The loading matrix is not uniquely defined, since $(\mathbf{A}_i, \mathbf{x}_t)$ in (1) can be replaced by $(\mathbf{A}_i\mathbf{U}_i, \mathbf{U}_i^{-1}\mathbf{x}_t)$ for any $k_0 \times k_0$ non-singular matrix $\mathbf{U}_i$, $i = 1, 2$. However, $\mathcal{M}(\mathbf{A}_i)$, the space spanned by the columns of $\mathbf{A}_i$, is uniquely defined under our assumptions. To estimate the column space $\mathcal{M}(\mathbf{A}_i)$, we will estimate an orthonormal representative of the space, a $p \times k_0$ matrix $\mathbf{Q}_i$, such that

$$\mathbf{Q}_i'\mathbf{Q}_i = \mathbf{I}_{k_0}, \text{ and } \mathbf{A}_i = \mathbf{Q}_i\mathbf{\Gamma}_i, \quad i = 1, 2, \tag{2}$$

where $\mathbf{\Gamma}_i$ is a $k_0 \times k_0$ non-singular matrix that provides the link between $\mathbf{Q}_i$ and $\mathbf{A}_i$. Again, due to ambiguity, $\mathbf{\Gamma}_i$ is not estimable. In any case, we have $\mathcal{M}(\mathbf{Q}_i) = \mathcal{M}(\mathbf{A}_i)$. The columns of $\mathbf{Q}_i$ are k_0 orthonormal vectors, and the column space spanned by $\mathbf{Q}_i$ is the same as the column space spanned by $\mathbf{A}_i$. In (1), we assume $\mathcal{M}(\mathbf{A}_1) \neq \mathcal{M}(\mathbf{A}_2)$. Let $I_{t,i}$ be the indicator function of regime i at time t, i.e. $I_{t,1} = I(z_t < r_0)$ and $I_{t,2} = I(z_t \geqslant r_0)$. Let

$$\mathbf{R}_t = \sum_{i=1}^{2} \mathbf{\Gamma}_i \mathbf{x}_t I_{t,i}, \tag{3}$$

The threshold factor model (1) can be written as

$$\mathbf{y}_t = \sum_{i=1}^{2} (\mathbf{Q}_i\mathbf{R}_t + \boldsymbol{\varepsilon}_{t,i}) I_{t,i}, \tag{4}$$

where $\mathbf{Q}_i$ are orthonormal matrices.

Our aim is to estimate the loading spaces $\mathcal{M}(\mathbf{A}_i)$, $i = 1, 2$, the number of factors k_0, and the threshold value r_0, given the threshold variable. When the threshold variable is unknown, we also propose a procedure for its identification.

Remark 1. We note that Peña and Box (1987), Forni et al. (2000, 2004) and Hallin and Liska (2007) make specific assumptions about the dynamics of the common factors. With a data generating process, forecasting becomes possible. More sophisticated and more accurate estimation procedures can be constructed by fully utilizing the assumed structure. In this paper we do not specify the dynamic structure for the factor process for several reasons. First, the procedure and theory become more general and less restrictive hence no additional model checking procedure is needed for confirming the dynamic structure. Second, we are able to use simple eigen-analysis for estimation. Third, the regimes bring more ambiguity into the model that makes the identification of the dynamic structure of the factors model difficult. Of course, without the dynamic structure, the model loses its forecasting ability, though the results can provide a good start for building a forecasting model if forecasting is the main objective.

Remark 2. The state-space is divided into two regimes, controlled by the threshold variable z_t. We assume z_t is observable at time t. It can be the lag variable of an observed time series. In a more complicate setting, z_t can be partially observable with several unknown parameters. For example, $z_t = \beta_1 z_{1t} + \beta_2 z_{2t}$ where z_{1t} and z_{2t} are observable at time t and β_i's are unknown parameters. Because z_t is observable at time t given the parameters, we know precisely which regime the process is in at time t, given r_0 and β_i's.

Remark 3. Constant terms can be included in model (1) as follows,

$$\mathbf{y}_t = \begin{cases} \boldsymbol{\mu}_1 + \mathbf{A}_1\mathbf{x}_t + \boldsymbol{\varepsilon}_{t,1} & z_t < r_0 \\ \boldsymbol{\mu}_2 + \mathbf{A}_2\mathbf{x}_t + \boldsymbol{\varepsilon}_{t,2} & z_t \geqslant r_0, \end{cases} \text{ and } \boldsymbol{\varepsilon}_{t,i} \sim N(\mathbf{0}, \mathbf{\Sigma}_{t,i}), \quad i = 1, 2. \tag{5}$$

If we combine these terms and loading matrices, model (5) can be written as a threshold factor model with $(k_0 + 1)$ factors. Specifically, in regime i, when $I_{t,i} = 1$,

$$\mathbf{y}_t = \begin{pmatrix} \boldsymbol{\mu}_i & \mathbf{A}_i \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{x}_t \end{pmatrix} + \boldsymbol{\varepsilon}_{t,i}, \text{ and } \boldsymbol{\varepsilon}_{t,i} \sim N(\mathbf{0}, \mathbf{\Sigma}_{t,i}).$$

Hence, model (5) is a special case of model (1), in which one of the common factor is deterministic. In order to accommodate this simplified setting, in the eigen-analysis when performing for loading matrix estimation, we use cross auto-moment matrices, instead of the traditional auto-covariance matrices.

Remark 4. The threshold factor model (1) provides a different approach from the regime switching model in Liu and Chen (2016). A typical regime switching model introduces a random switching mechanism that is not observed. In threshold models, regime switching is observable, given the observable threshold variable z_t and the threshold value r_0. It provides easier estimation, clearer interpretation and better predictability. In all threshold modeling approaches, identification of a suitable threshold variable that drives regime switching is the most important modeling component and is often the most challenging one. We will propose an identification approach that is easy to use and can screen a large number of potential threshold variables in Section 4.

Remark 5. Note that the two-regime threshold factor model (1) can be written as a one regime model as

$$\mathbf{y}_t = \mathbf{A}\mathbf{f}_t + \boldsymbol{\varepsilon}_t, \tag{6}$$

where $\mathbf{A} = [\mathbf{A}_1 \ \ \mathbf{A}_2]$, $\mathbf{f}_t = (\mathbf{x}_t' I_{t,1}, \mathbf{x}_t' I_{t,2})'$, and $\boldsymbol{\varepsilon}_t = \sum_{i=1}^{2} \boldsymbol{\varepsilon}_{t,i} I_{t,i}$.

Although this model seems to be structurally simpler, the switching mechanism is hidden in the factor dynamics. Due to rotation ambiguity, detecting such changes in the factor process is much more difficult. Although in our framework we only impose weak conditions on the factors hence we are able to estimate model (6) even with the hidden structural changes, the results will not be able to reveal the switching mechanism and the underlying threshold variable, which are important features of the time series one would like to know. In addition, model (6) uses a larger number of factors, which makes estimation less accurate and less interpretable.

Another advantage for model (1) is that the factors are allowed to have different 'strengths' across the regimes, while the estimation of model (6) using that in Lam et al. (2011) works better with the factors having the same strength. Strength roughly measures the total squared impact of a factor on the observed time series. A more formal definition is given in Section 3.

Remark 6. The detection of the existence of switching is an interesting but challenging problem. Possible approaches include testing the differences in the loading space, or testing the distance between the two loading spaces is zero. However, under the null hypothesis of no switching, the threshold value parameter becomes latent which poses technical challenges. A complete and rigorous study of the problem is out of scope of this paper.

3. Estimation procedure with a given threshold variable

In this section, we first present a procedure to estimate the loading spaces corresponding to a partition in the form of $I_{t,1}(r_1) = I(z_t < r_1)$ and $I_{t,2}(r_2) = I(z_t \geqslant r_2)$ where $r_1 \leqslant r_2$, and show the asymptotic property of the estimator in the case of $r_1 \leqslant r_0$ and $r_2 \geqslant r_0$, where r_0 is the true threshold value. Then we propose a procedure for estimating r_0 using tentative threshold values to split the data, along its asymptotic properties. The asymptotic properties of the estimated loading spaces using the estimated threshold value are also presented.

Here are some notations. For any matrix $\mathbf{H}$, let $\|\mathbf{H}\|_F$ be the Frobenius norm of $\mathbf{H}$ which is the root of sum squares of singular values of $\mathbf{H}$, and $\|\mathbf{H}\|_2$ be the spectral norm of $\mathbf{H}$ which is the maximum singular value of $\mathbf{H}$. We use $\sigma_i(\mathbf{H})$ to denote the ith largest singular value of $\mathbf{H}$, and $\|\mathbf{H}\|_{\min}$ to denote the square root of minimum nonzero eigenvalue of $\mathbf{H}'\mathbf{H}$. For a square matrix $\mathbf{H}$, $\mathrm{tr}(\mathbf{H})$ denotes its trace. We write $a \asymp b$, if $a = O(b)$ and $b = O(a)$. We use C to denote a positive constant.

3.1. Initial estimation of the loading spaces with tentative threshold values

Define the generalized second cross moment matrices of $\mathbf{x}_t$ and $\mathbf{y}_t$ of lead h in different partitions as

$$\boldsymbol{\Sigma}_{x,i,j}(h, r_1, r_2) = \frac{1}{n-h}\sum_{t=1}^{n-h} \mathrm{E}(\mathbf{x}_t \mathbf{x}_{t+h}' I_{t,i}(r_i) I_{t+h,j}(r_j)),$$
$$\boldsymbol{\Sigma}_{y,i,j}(h, r_1, r_2) = \frac{1}{n-h}\sum_{t=1}^{n-h} \mathrm{E}(\mathbf{y}_t \mathbf{y}_{t+h}' I_{t,i}(r_i) I_{t+h,j}(r_j)),$$

for $i, j = 1, 2$. Here $\boldsymbol{\Sigma}_{y,1,1}(h, r_1, r_2)$ is the cross moment matrix of $\mathbf{y}_t$ and $\mathbf{y}_{t+h}$ when both $\mathbf{y}_t$ and $\mathbf{y}_{t+h}$ are in partition 1 with $\{z_t < r_1\}$, and $\boldsymbol{\Sigma}_{y,1,2}(h, r_1, r_2)$ is that when $\mathbf{y}_t$ is in partition 1 with $\{z_t < r_1\}$ and $\mathbf{y}_{t+h}$ is in partition 2 with $\{z_{t+h} \geqslant r_2\}$. $\boldsymbol{\Sigma}_{y,2,1}(h, r_1, r_2)$ and $\boldsymbol{\Sigma}_{y,2,2}(h, r_1, r_2)$ are similar.

Define the sum of a quadratic version of the cross moment matrices of $\mathbf{y}_t$,

$$\mathbf{M}_i(r_1, r_2) = \sum_{h=1}^{h_0}\sum_{j=1}^{2} \boldsymbol{\Sigma}_{y,i,j}(h, r_1, r_2)\boldsymbol{\Sigma}_{y,i,j}(h, r_1, r_2)', \tag{7}$$

for a pre-fixed maximum lead h_0, and $i = 1, 2$.

Let $\mathbf{q}_{i,k}(r_1, r_2)$ and $-\mathbf{q}_{i,k}(r_1, r_2)$ be the pair of unit eigenvectors of $\mathbf{M}_i(r_1, r_2)$ corresponding to its kth largest eigenvalue. In the following we assume $\mathbf{1}'\mathbf{q}_{i,k}(r_1, r_2) > 0$ (which is uniquely defined) and will use it in all our constructions. Define

$$\mathbf{Q}_i(r_1, r_2) = (\mathbf{q}_{i,1}(r_1, r_2), \ldots, \mathbf{q}_{i,k_0}(r_1, r_2)), \tag{8}$$

for $i = 1, 2$, where k_0 is the number of factors in model (1). For simplicity, in the rest of the paper, if $r_1 = r_2$, notations $\boldsymbol{\Sigma}_{y,i,j}(h, r, r)$, $\mathbf{Q}_i(r, r)$ and $\mathbf{M}_i(r, r)$ are simplified to $\boldsymbol{\Sigma}_{y,i,j}(h, r)$, $\mathbf{Q}_i(r)$, and $\mathbf{M}_i(r)$. Furthermore, we use $\mathbf{Q}_i$ and $\mathbf{M}_i$ to denote $\mathbf{Q}_i(r_0)$ and $\mathbf{M}_i(r_0)$, where r_0 is the true threshold value. Using the sample version of $\boldsymbol{\Sigma}_{y,i,j}(h, r_1, r_2)$ and $\mathbf{M}_i(r_1, r_2)$, we perform eigenanalysis of $\widehat{\mathbf{M}}_i(r_1, r_2)$ to obtain $\widehat{\mathbf{Q}}_i(r_1, r_2)$.

The rationale behind the approach is the following. Let r_0 be the true threshold value. Consider the case that $r_1 \leqslant r_0$ and $r_2 \geqslant r_0$. Then partition 1 with $\{z_t < r_1\}$ is a subset of data in Regime 1 and partition 2 with $\{z_t \geqslant r_2\}$ is a subset of data in Regime 2. For any integer $h > 0$, it follows from model (1) that

$$\boldsymbol{\Sigma}_{y,1,1}(h, r_1, r_2) = \mathbf{A}_1\boldsymbol{\Sigma}_{x,1,1}(h, r_1, r_2)\mathbf{A}_1', \text{ and } \boldsymbol{\Sigma}_{y,1,2}(h, r_1, r_2) = \mathbf{A}_1\boldsymbol{\Sigma}_{x,1,2}(h, r_1, r_2)\mathbf{A}_2',$$

under the white noise assumption. In this case,

$$\mathbf{M}_1(r_1, r_2) = \mathbf{A}_1\left(\sum_{h=1}^{h_0}\sum_{j=1}^{2}\boldsymbol{\Sigma}_{x,1,j}(h, r_1, r_2)\mathbf{A}_j'\mathbf{A}_j\boldsymbol{\Sigma}_{x,1,j}(h, r_1, r_2)'\right)\mathbf{A}_1'.$$

Then $\mathbf{M}_1(r_1, r_2)$ is a positive semi-definite matrix sandwiched by $\mathbf{A}_1$. If there exists at least one $1 \leqslant h \leqslant h_0$ such that $\boldsymbol{\Sigma}_{x,1,1}(h, r_1, r_2)$ or $\boldsymbol{\Sigma}_{x,1,2}(h, r_1, r_2)$ is full rank, then $\mathbf{M}_1(r_1, r_2)$ has rank k_0. Hence the corresponding $\mathbf{Q}_1(r_1, r_2)$ is an orthonormal representative of $\mathcal{M}(\mathbf{A}_1)$. The results also hold for $\mathbf{M}_2(r_1, r_2)$.

Let the sample versions of $\boldsymbol{\Sigma}$ and $\mathbf{M}$ be

$$\widehat{\boldsymbol{\Sigma}}_{y,i,j}(h, r_1, r_2) = \frac{1}{n-h}\sum_{t=1}^{n-h}\mathbf{y}_t\mathbf{y}_{t+h}'I_{t,i}(r_i)I_{t+h,j}(r_j),$$

$$\widehat{\mathbf{M}}_i(r_1, r_2) = \sum_{h=1}^{h_0}\sum_{j=1}^{2}\widehat{\boldsymbol{\Sigma}}_{y,i,j}(h, r_1, r_2)\widehat{\boldsymbol{\Sigma}}_{y,i,j}(h, r_1, r_2)'. \tag{9}$$

Let $\widehat{\mathbf{q}}_{i,k}(r_1, r_2)$ be the eigenvector of $\widehat{\mathbf{M}}_i(r_1, r_2)$ corresponding to its kth largest eigenvalue. $\mathcal{M}(\mathbf{A}_i)$ can be estimated by

$$\widehat{\mathcal{M}(\mathbf{A}_i)} = \mathcal{M}(\widehat{\mathbf{Q}}_i(r_1, r_2)),$$

where $\widehat{\mathbf{Q}}_i(r_1, r_2) = (\widehat{\mathbf{q}}_{i,1}(r_1, r_2), \ldots, \widehat{\mathbf{q}}_{i,k_0}(r_1, r_2))$.

However, when $r_2 \geqslant r_1 > r_0$ or $r_1 \leqslant r_2 < r_0$, the eigenspace of $\mathbf{M}_i(r_1, r_2)$ corresponding to nonzero eigenvalues may not correspond to $\mathcal{M}(\mathbf{A}_i)$. For example, if $r_1 > r_0$, the partition $I_{t,1}(r_1) = 1$ contains observations in both regimes – $r_0 \leqslant z_t < r_1$ for Regime 2 and $z_t < r_0 < r_1$ for Regime 1. Hence $\mathbf{M}_1(r_1, r_2)$ is not sandwiched by $\mathbf{A}_1$. We will show later that, when r_1 and r_2 are sufficiently close the r_0, the estimated space using the sample version of $\mathbf{M}_i(r_1, r_2)$ is still consistent.

Remark 7. It is tempting not to separate y_{t+h} into two regimes and use

$$\mathbf{M}_i(r_1, r_2) = \sum_{h=1}^{h_0}\boldsymbol{\Sigma}_{y,i}(h, r_1, r_2)\boldsymbol{\Sigma}_{y,i}(h, r_1, r_2)',$$

where

$$\boldsymbol{\Sigma}_{y,1}(h, r_1, r_2) = \frac{1}{n-h}\sum_{t=1}^{n-h}E(\mathbf{y}_t\mathbf{y}_{t+h}'I(z_t < r_1)), \quad \boldsymbol{\Sigma}_{y,2}(h, r_1, r_2) = \frac{1}{n-h}\sum_{t=1}^{n-h}E(\mathbf{y}_t\mathbf{y}_{t+h}'I(z_t \geqslant r_2)).$$

to replace (7). However, if the two loading spaces $\mathcal{M}(\mathbf{Q}_1)$ and $\mathcal{M}(\mathbf{Q}_2)$ are not orthogonal, cancellation may occur, making the rank of $\mathbf{M}_i(r_1, r_2)$ less than the rank of $\mathbf{A}_i$.

Remark 8. Theoretically, $\mathcal{M}(\mathbf{A}_i)$ can be estimated through eigen-decomposition of one of $\{\boldsymbol{\Sigma}_{y,i,j}(h, r_1, r_2)\boldsymbol{\Sigma}_{y,i,j}(h, r_1, r_2)', h = 1, 2, \ldots\}$, as long as it is full rank. Asymptotically they converge at the same rate. The reason for using $\mathbf{M}_i(r_1, r_2)$ in (7) is that by summing over h, we do not need to find a particular h to satisfy the condition. Since the strongest correlation often occurs at smaller leads, a relatively small h_0 is usually adopted. The autocorrelation of each individual $\mathbf{y}_t$ often provides a good indication of the proper h_0 to be used.

To study the asymptotic properties of the estimator, we extend a distance measure of two linear spaces used in Chang et al. (2015) and Liu and Chen (2016). Let $\mathbf{S}_1$ be a $p \times q_1$ matrix with rank q_1 and $\mathbf{S}_2$ be a $p \times q_2$ matrix with rank q_2, where $p \geqslant \max\{q_1, q_2\}$. Let the columns of $\mathbf{O}_i$ be an orthonormal basis of $\mathcal{M}(\mathbf{S}_i)$, for $i = 1, 2$. Define

$$\mathcal{D}(\mathcal{M}(\mathbf{S}_1),\ \mathcal{M}(\mathbf{S}_2)) = \left(1 - \frac{\mathrm{tr}(\mathbf{O}_1\mathbf{O}_1'\mathbf{O}_2\mathbf{O}_2')}{\min\{q_1, q_2\}}\right)^{1/2}, \tag{10}$$

as the distance of the column spaces of $\mathbf{S}_1$ and $\mathbf{S}_2$. It is a quantity between 0 and 1. It is equal to 0 if $\mathcal{M}(\mathbf{S}_1) \subseteq \mathcal{M}(\mathbf{S}_2)$ or $\mathcal{M}(\mathbf{S}_2) \subseteq \mathcal{M}(\mathbf{S}_1)$, and 1 if $\mathcal{M}(\mathbf{S}_1)$ and $\mathcal{M}(\mathbf{S}_2)$ are orthogonal. The distance of two linear spaces with the same dimension is defined in Chang et al. (2015) and Liu and Chen (2016). Here (10) is a modified version and takes into consideration the scenario that the dimensions of two spaces may be different.

For factor models in high-dimensional cases, it is common to assume that the number of factors is fixed and the squared spectral norm of the $p \times k_0$ loading matrix $\mathbf{A}_i$ grows with the dimension p (Bai and Ng, 2002; Doz et al., 2011). The growth rate is called the strength of the factors in Lam et al. (2011), Lam and Yao (2012), Chang et al. (2015) and Liu and Chen (2016). Let

$$\|\mathbf{A}_i\|_2^2 \asymp \|\mathbf{A}_i\|_{\min}^2 \asymp p^{1-\delta_i}, \quad 0 \leqslant \delta_i \leqslant 1.$$

If $\delta_i = 0$, regime i is called a strong regime. If $0 < \delta_i < 1$, the regime is called a weak regime. If $\delta_i = 1$, the regime is called an extremely weak regime. The strength of the regime measures the relative growth rate of the amount of information about the common factors $\mathbf{x}_t$ carried by the observed process $\mathbf{y}_t$, as p increases, comparing to the growth rate of the amount of noise process in regime i. It is seen that in the following theoretical development that the strength of the regime plays an important role in estimation efficiency.

Theorem 1. *Under Conditions 1–6 in Appendix A.1, when $r_1 \leqslant r_0$ and $r_2 \geqslant r_0$, if $p^{\delta_1/2+\delta_2/2}n^{-1/2} = o(1)$ and $\widehat{\mathbf{Q}}_1(r_1, r_2)$ and $\widehat{\mathbf{Q}}_2(r_1, r_2)$ are estimated using the true k_0, it holds that*

$$\mathcal{D}(\mathcal{M}(\widehat{\mathbf{Q}}_i(r_1, r_2)), \mathcal{M}(\mathbf{Q}_i)) = O_p(p^{\delta_i/2+\delta_{\min}/2}n^{-1/2}), \quad \textit{for} \ \ i = 1, 2,$$

as $n, p \to \infty$, where $\delta_{\min} = \min\{\delta_1, \delta_2\}$.

It is clearly not efficient to estimate the loading spaces with only partial observations. In practice, once the threshold value is estimated, we can use the full data set for the estimation of loading spaces. The asymptotic properties of the estimators based on the estimated threshold value will be discussed later in Theorems 3 and 5.

The convergence rates shown in Theorem 1 are the same as those in Liu and Chen (2016). It is worth noting that when the two regimes have different strengths δ_1 and δ_2, the convergence rate of the estimator in the stronger regime is the same as that in the one regime case, but the rate of the weaker regime is faster than that if it is the only regime. In other words, the estimation in the stronger regime is not hurt by the weaker regime, but the weaker regime gains efficiency from the stronger regime due to the switching of the process between the two regimes. We call it the 'helping effect'.

3.2. Estimation of the threshold value

Estimation of the threshold value in a threshold model has been extensively studied in univariate threshold models using least squares or likelihood estimators, including those in Tong and Lim (1980), Tsay (1989), Chan and Tong (1990), Chan (1993), Caner and Hansen (2004), Chen and So (2006) and Wu and Chen (2007). Here, we construct an objective function for the estimation of the threshold value. Since for a given finite sample, the model is not distinguishable for all values between two adjacent observations of z_t as its threshold value, we follow the standard approach and assume that the threshold value takes on a finite number of possible values in the set of all observed z_t. Our method is to traverse all of these possible threshold values and find the best one that optimizes the objective function.

When r is used as the tentative threshold value to split the data into two subsets with $\{z_t < r\}$ and $\{z_t \geqslant r\}$, we define the objective function

$$G(r) = \sum_{i=1}^{2} \left\| \mathbf{B}_i{}' \,\mathbf{M}_i(r)\, \mathbf{B}_i \right\|_2 = \sum_{i=1}^{2} \left\| \sum_{h=1}^{h_0} \sum_{j=1}^{2} \mathbf{B}_i{}' \boldsymbol{\Sigma}_{y,i,j}(h, r) \boldsymbol{\Sigma}_{y,i,j}(h, r)' \mathbf{B}_i \right\|_2, \tag{11}$$

where $\mathbf{B}_i$ is a $p \times (p - k_0)$ matrix for which $(\mathbf{Q}_i, \mathbf{B}_i)$ forms a $p \times p$ orthonormal matrix with $\mathbf{Q}_i'\mathbf{B}_i = \mathbf{0}$ and $\mathbf{B}_i'\mathbf{B}_i = \mathbf{I}_{p-k_0}$. $\mathcal{M}(\mathbf{B}_i)$ is the orthogonal complement space of $\mathcal{M}(\mathbf{Q}_i)$, for $i = 1, 2$. Note that although $\mathbf{B}_i$ is not uniquely defined and subject to any orthogonal transformation, $G(r)$ is invariant under such transformations. The function $G(\cdot)$ measures the sum of the squared norm of the projections of the cross moment matrices $\boldsymbol{\Sigma}_{y,i,j}(h, r)$ onto $\mathcal{M}(\mathbf{B}_i)$ for $h = 1, \ldots, h_0$.

If $r = r_0$ (the true threshold value), the observations in the two subsets identified do belong to the correct regimes. Then $\mathbf{M}_1$ is sandwiched by $\mathbf{A}_1$ and $\mathbf{M}_2$ is sandwiched by $\mathbf{A}_2$. Hence

$$G(r_0) = \sum_{i=1}^{2} \left\| \sum_{h=1}^{h_0} \sum_{j=1}^{2} \mathbf{B}_i{}' \,\mathbf{A}_i \boldsymbol{\Sigma}_{x,i,j}(h, r_0) \mathbf{A}_j' \mathbf{A}_j \boldsymbol{\Sigma}_{x,i,j}(h, r_0) \mathbf{A}_i' \,\mathbf{B}_i \right\|_2 = 0.$$

If $r \neq r_0$, one of two subsets contains observations from both regimes. Then the projection will not be zero. The following proposition formally states that, under mild conditions, we have $G(r) > 0$ for any $r \neq r_0$.

Proposition 1. *Under Conditions 1–9 in Appendix A.1, if $r \neq r_0$, then $G(r) > 0$.*

The proof of the proposition is in Appendix A.2.

To obtain the sample version of $G(r)$, we assume *apriori* that r_0 is in a known region of the support of z_t, $r_0 \in (\eta_1, \eta_2)$. Such an assumption is standard in threshold model estimation. Under this assumption, we can use data corresponding to $z_t \leqslant \eta_1$ and $z_t \geqslant \eta_2$ to obtain estimates for $\mathcal{M}(\mathbf{B}_1)$ and $\mathcal{M}(\mathbf{B}_2)$, respectively. Specifically, $\mathbf{B}_i(\eta_1, \eta_2)$ is estimated by

$$\widehat{\mathbf{B}}_i(\eta_1, \eta_2) = (\widehat{\mathbf{q}}_{i,k_0+1}(\eta_1, \eta_2), \ldots, \widehat{\mathbf{q}}_{i,p}(\eta_1, \eta_2)),$$

for $i = 1, 2$. By Theorem 1, both of them are consistent. Define

$$\widehat{G}(r) = \sum_{i=1}^{2} \left\| \widehat{\mathbf{B}}_i(\eta_1, \eta_2)' \widehat{\mathbf{M}}_i(r) \widehat{\mathbf{B}}_i(\eta_1, \eta_2) \right\|_2.$$

We estimate r_0 by

$$\widehat{r} = \arg \min_{r \in \{z_1, \ldots, z_n\} \bigcap (\eta_1, \eta_2)} \widehat{G}(r). \tag{12}$$

Remark 9. In the above procedure we require that there are sufficient samples corresponding to $z_t \leqslant \eta_1$ and $z_t \geqslant \eta_2$, to ensure the accuracy of the estimated $\mathcal{M}(\mathbf{B}_i)$, for $i = 1, 2$. When the values of η_1 and η_2 are not clear, it is possible to use a sequential procedure based on the ranked sequence of z_t, similar to that in Tsay (1989).

Remark 10. An alternative objective function is the likelihood of $\mathbf{B}_i \mathbf{y}_t$. However, we do not want to involve the structure of the covariance matrices of the noise process which both are $p \times p$ matrices. Here we still take advantage of the whiteness assumption of the noise process, and use the cross moment matrices of $\mathbf{y}_t$ at nonzero leads.

Theorem 2. *Under Conditions 1–9 in Appendix A.1, if $p^{\delta_1/2+\delta_2/2} n^{-1/2} = o(1)$, with true k_0, it holds that, for any $\epsilon > 0$,*

$$P(\widehat{r} < r_0 - \epsilon) \leqslant \frac{C p^{\delta_1/2+\delta_{\min}/2}}{\epsilon n^{1/2}}, \quad P(\widehat{r} > r_0 + \epsilon) \leqslant \frac{C p^{\delta_2/2+\delta_{\min}/2}}{\epsilon n^{1/2}},$$

as $n, p \to \infty$.

Theorem 2 shows that the estimator $\widehat{r}$ in (12) is consistent under some mild conditions. The estimation performance depends on the strength of both regimes. If the two regimes are both strong ($\delta_1 = \delta_2 = 0$), the estimation is immune to the curse of dimensionality. However, if at least one regime is weak, the estimator becomes less efficient as p increases, and would require larger sample size n for consistency. When the two regimes have different strengths, the probability that $\widehat{r}$ falls in the stronger regime is smaller than that in the weaker regime (the one with larger estimation error). Hence the overall rate of convergence of $\widehat{r}$ is determined by the strength of the weaker regime.

3.3. Estimation of the loading spaces with estimated threshold value

The final estimation of $\mathcal{M}(\mathbf{A}_i)$ is obtained using $\widehat{r}$ as the threshold value and the procedure in Section 3.1. Specifically, we define

$$\widehat{\mathbf{Q}}_i(\widehat{r}) = (\widehat{\mathbf{q}}_{i,1}(\widehat{r}), \ldots, \widehat{\mathbf{q}}_{i,k_0}(\widehat{r})),$$

where $\widehat{\mathbf{q}}_{i,k}(\widehat{r})$ is the unit eigenvector of $\widehat{\mathbf{M}}_i(\widehat{r})$ corresponding to its kth largest eigenvalue, and $\widehat{\mathbf{M}}_i(\widehat{r})$ is defined in (9).

Theorem 3 presents the asymptotics of the estimated loading spaces when the estimated threshold value is used.

Theorem 3. *Under Conditions 1–9 in Appendix A.1, if $p^{\delta_1/2+\delta_2/2} n^{-1/2} = o(1)$ and $\widehat{\mathbf{Q}}_1(\widehat{r})$ and $\widehat{\mathbf{Q}}_2(\widehat{r})$ are estimated with the true k_0, it holds that*

$$\mathcal{D}(\mathcal{M}(\widehat{\mathbf{Q}}_i(\widehat{r})), \mathcal{M}(\mathbf{Q}_i)) = O_p(p^{\delta_i/2+\delta_{\min}/2} n^{-1/2}), \quad \text{for} \quad i = 1, 2,$$

as $n, p \to \infty$.

Theorem 3 shows that the rates of loading space estimators are the same as that when the true threshold value is known.

Let $\mathbf{s}_t$ be the signal (or dynamic) part of $\mathbf{y}_t$, defined as $\mathbf{s}_t = \mathbf{A}_1 \mathbf{x}_t I_{t,1} + \mathbf{A}_2 \mathbf{x}_t I_{t,2}$. Since the column space of $\mathbf{A}_i$ is identifiable only up to a nonsingular transformation across regimes, we cannot estimate the latent process $\mathbf{x}_t$ directly, but we have a natural estimator for $\mathbf{s}_t$ and the latent process $\mathbf{R}_t$ with standardized loadings in (3),

$$\widehat{\mathbf{s}}_t(\widehat{r}) = \sum_{i=1}^{2} \widehat{\mathbf{Q}}_i(\widehat{r}) \widehat{\mathbf{Q}}_i(\widehat{r})' \mathbf{y}_t I_{t,i}(\widehat{r}), \quad \widehat{\mathbf{R}}_t(\widehat{r}) = \sum_{i=1}^{2} \widehat{\mathbf{Q}}_i(\widehat{r})' \mathbf{y}_t I_{t,i}(\widehat{r}). \tag{13}$$

3.4. When the number of factors is unknown

In practice, the number of factors k_0 is usually unknown. This quantity can be estimated through a similar eigenvalue ratio estimator used in Lam et al. (2011). Specifically, again we assume r_0 is in a known interval (η_1, η_2), and let

$$\widehat{k}_i = \arg \min_{1 \leqslant k \leqslant R} \frac{\widehat{\lambda}_{i,k+1}(\eta_1, \eta_2)}{\widehat{\lambda}_{i,k}(\eta_1, \eta_2)}, \quad \text{for} \quad i = 1, 2, \tag{14}$$

where $\widehat{\lambda}_{i,k}(\eta_1, \eta_2)$ is the kth largest eigenvalue of $\widehat{\mathbf{M}}_i(\eta_1, \eta_2)$. We follow Lam and Yao (2012) and use $R = p/2$, under the assumption that $k_0 \ll p$.

Corollary 1. *Under the Conditions 1–9 in Appendix A.1, if $p^{\delta_1/2+\delta_2/2}n^{-1/2} = o(1)$, for $i = 1, 2$, as $n, p \to \infty$, it holds that*

$$\widehat{\lambda}_{i,k+1}(\eta_1, \eta_2)/\widehat{\lambda}_{i,k}(\eta_1, \eta_2) \asymp 1 \ \ \textit{for} \ \ k = 1, \dots, k_0 - 1,$$
$$\widehat{\lambda}_{i,k_0+1}(\eta_1, \eta_2)/\widehat{\lambda}_{i,k_0}(\eta_1, \eta_2) = O_p(p^{\delta_i+\delta_{\min}} n^{-1}).$$

Corollary 1 gives the order of the ratios of the estimated eigenvalues, and shows that the probability of underestimating k_0 with (14) goes to zero asymptotically. For $k > k_0$, the eigenvalues $\lambda_{i,k}$ are theoretically zero hence the property of the ratio $\widehat{\lambda}_{i,k+1}/\widehat{\lambda}_{i,k}$ is difficult to obtain. Though the consistency of (14) cannot been confirmed theoretically (Lam and Yao, 2012), the estimator performs well in numerical experiments (Chang et al., 2015; Liu and Chen, 2016). In Theorems 4 and 5 we will show that even when the numbers of factors are overestimated, the asymptotic properties of the threshold value is the same as that when the numbers of factors is known. The estimated loading spaces with an overestimated $\widehat{k}$, if restricted to the correct dimension is also consistent.

Note that the convergence rate of $\widehat{\lambda}_{i,k_0+1}(\eta_1, \eta_2)/\widehat{\lambda}_{i,k_0}(\eta_1, \eta_2)$ is faster in the stronger regime (smaller δ_i). Since k_0 is common to both regimes, we choose the one identified by the regime with a larger 'strength', reflected by the scale of $\|\widehat{\mathbf{M}}_i(\eta_1, \eta_2)\|_2$ (Liu and Chen, 2016). Hence, we use

$$\widehat{k} = \widehat{k}_{\widehat{\ell}}, \ \text{where } \widehat{\ell} = \arg\max_{\ell=1,2} \|\widehat{\mathbf{M}}_\ell(\eta_1, \eta_2)\|_2. \tag{15}$$

In the following we present the asymptotic properties of the proposed estimators when the number of factors is not correctly estimated. We will show that if k_0 is overestimated, the proposed method can still estimate the threshold value and loading spaces accurately.

Let

$$\widehat{G}_k(r) = \sum_{i=1}^{2} \left\| \widehat{\mathbf{B}}_{i,k}(\eta_1, \eta_2)' \, \widehat{\mathbf{M}}_i(r) \, \widehat{\mathbf{B}}_{i,k}(\eta_1, \eta_2) \right\|_2,$$

where $\widehat{\mathbf{B}}_{i,k}(\eta_1, \eta_2) = (\widehat{\mathbf{q}}_{i,k+1}(\eta_1, \eta_2), \dots, \widehat{\mathbf{q}}_{i,p}(\eta_1, \eta_2))$, for $i = 1, 2$. When k_0 is unknown, r_0 is estimated by

$$\widetilde{r} = \arg \min_{r \in \{z_1,\dots,z_n\} \bigcap (\eta_1,\eta_2)} \widehat{G}_{\widehat{k}}(r).$$

The loading spaces are estimated using $\widehat{k}$ as the number of factors and $\tilde{r}$ as the threshold value. Specifically,

$$\widetilde{\mathbf{Q}}_i(\widehat{k}, \widetilde{r}) = (\widehat{\mathbf{q}}_{i,1}(\widetilde{r}), \dots, \widehat{\mathbf{q}}_{i,\widehat{k}}(\widetilde{r})),$$

where $\widehat{\mathbf{q}}_{i,k}(\widetilde{r})$ is the unit eigenvector of $\widehat{\mathbf{M}}_i(\widetilde{r})$ corresponding to its kth largest eigenvalue, for $i = 1, 2$.

Theorem 4. *Under Conditions 1–10 in Appendix A.1, if $p^{\delta_1/2+\delta_2/2}n^{-1/2} = o(1)$ and $k_0 < \widehat{k} < 2k_0 - \nu$, it holds that*

$$P(\widetilde{r} < r_0 - \epsilon) \leqslant \frac{Cp^{\delta_1/2+\delta_{\min}/2}}{\epsilon n^{1/2}}, \quad P(\widetilde{r} > r_0 + \epsilon) \leqslant \frac{Cp^{\delta_2/2+\delta_{\min}/2}}{\epsilon n^{1/2}},$$

as $n, p \to \infty$, for $\epsilon > 0$, where $\nu = \dim(\mathcal{M}(\mathbf{Q}_1) \cap \mathcal{M}(\mathbf{Q}_2))$.

Theorem 5 will show that the space spanned by the first k_0 columns of $\widetilde{\mathbf{Q}}_i(\widehat{k}, \widetilde{r})$ provides an estimate of $\mathcal{M}(\mathbf{Q}_i)$ which converges as fast as $\mathcal{M}(\widehat{\mathbf{Q}}_i(\widehat{r}))$ in Theorem 3. Define $\widetilde{\mathbf{Q}}_i(\widetilde{r}) = (\widehat{\mathbf{q}}_{i,1}(\widetilde{r}), \dots, \widehat{\mathbf{q}}_{i,k_0}(\widetilde{r}))$, consisting of the first k_0 columns of $\widetilde{\mathbf{Q}}_i(\widehat{k}, \widetilde{r})$.

Theorem 5. *Under Conditions 1–10 in Appendix A.1, if $p^{\delta_1/2+\delta_2/2}n^{-1/2} = o(1)$ and $k_0 < \widehat{k} < 2k_0 - \nu$, it holds that*

$$\mathcal{D}(\mathcal{M}(\widetilde{\mathbf{Q}}_i(\widetilde{r})), \ \mathcal{M}(\mathbf{Q}_i)) = O_p(p^{\delta_i/2+\delta_{\min}/2} n^{-1/2}), \quad \textit{for} \ \ i = 1, 2,$$

as $n, p \to \infty$.

Theorems 4 and 5 state that when the number of factors is overestimated, our estimators for threshold value and the loading spaces are still consistent. Their asymptotic properties are the same as those when the number of factors is correctly estimated. Of course, when k_0 is over estimated, we lose some efficiency.

4. Searching for threshold variable

When there is no prior knowledge on the threshold variable, a data-driven procedure is needed in order to search for a suitable one. In standard univariate threshold models, a typical candidate pool is the lag variables (Tong and Lim, 1980; Tong, 1990; Chan, 1993; Tong, 1993) and identification is often done by using model comparison procedures. However, in the high dimensional setting, the candidate pool can be very large hence the trial-and-error approach can be extremely time consuming, complicated more by the multiple comparison problem at the end. Here we propose a reverse approach that is closely related to the procedure proposed in Wu and Chen (2007). Specifically, we propose to follow a three step procedure: classification, screening and model selection. First, a regime-switching factor model of Liu and Chen (2016) is built to obtain an initial regime identification for each time t, without engaging the threshold mechanism. Then the estimated regime identification is screened against all threshold variable candidates in a possibly very large candidate pool and a small set of candidates is selected by checking whether a candidate variable will produce regime identifications similar to the estimated identifications obtained in the classification step. Lastly model comparison procedure is used to select the most suitable threshold variable among the small subset.

Classification: Following Liu and Chen (2016), a most likely regime identification $\widehat{I_t} \in \{1, 2\}$ is obtained for $t = 1, \ldots, n$, using an iterative procedure of Viterbi algorithm and factor model estimation. For more details, see Liu and Chen (2016). Slightly different from that in Liu and Chen (2016), we assume independent switching instead of Markov switching, with prior probability $P(I_t = 1) = 0.5$. Classification step generates initial regime identifications for the next steps. Although it may not be perfect, it is able to reveal the possible relationship between the observed series and threshold variable candidates, and thus provide guidelines to narrows down a large candidates pool.

Screening: It is noted that if a variable z_t is indeed the threshold variable and the true threshold value is r_0, then the true regime identification I_t satisfies $I_t = 1 + I(z_t \geqslant r_0)$. Let

$$Q(\{z_t\}) = \max_r \left| \sum_{t=1}^{n} 2(I_t - 1.5)(2I(z_t \geqslant r) - 1) \right|.$$

If the variable z_t is the true threshold variable and $r = r_0$ is the true threshold, then $Q(\{z_t\}) = n$, reaching its maximum value. In fact $Q(\{z_t\})$ is the maximum of a binary CUSUM statistic.

Using the estimated $\widehat{I_t}$ from the classification step, we obtain

$$\widehat{Q}(\{z_t\}) = \max_r \left| \sum_{t=1}^{n} 2(\widehat{I_t} - 1.5)(2I(z_t \geqslant r) - 1) \right|.$$

For a set of candidate pool $S = \{z_t^{(\ell)}\}$, we screen each of them by calculating $\widehat{Q}_\ell = \widehat{Q}(\{z_t^{(\ell)}\})$, which can be done very efficiently. The variables with the largest $\widehat{Q}$ values then form a small set of candidates for more careful examinations.

Remark 11. It is also possible to identify a linear combination of several variables as a threshold variable, using a supervised learning algorithm such as support vector machine or classification tree, with $\widehat{I_t}$ obtained from the classification step as the response classification. Since $\widehat{I_t}$ is an estimate with error, the combination of parameters needs to be re-estimated under the original model, similar to our approach of estimating the threshold value r_0, though much more complicated.

Remark 12. Searching among a very large pool of candidates will inevitably find spurious relationships and meaningless variables that happen to match the underlying regime switching dynamics in the observed period. Hence a close inspection of the chosen variables is necessary. Prior knowledge is also important. In certain contexts there are known good candidates, such as composite indexes which summarize contemporaneous information, and their lag variables. For example, in modeling a panel of economic indicators, potential threshold variable candidates can be the recession and expansion indicator of the previous quarter, and its lag variables, since the dynamics of the economy are potentially different in recession or expansion periods. In modeling stock returns, the volatility of the market index and its lag variables can potentially be good threshold variables as stocks often behave differently in markets with different volatility.

Model comparison procedure: With a small set of possible threshold variables, a more careful analysis can be carried out. We use the estimation methods in Section 3 to obtain the cross validated residual sum of squares for model comparison.

Specifically, for each threshold variable candidate z_t, we estimate the loading matrices and the threshold value using data $\{\mathbf{y}_1, \ldots, \mathbf{y}_{t_0}\}$. With those estimates, we calculate the residual sum of squares for the remaining data $\{\mathbf{y}_{t_0+1}, \ldots, \mathbf{y}_n\}$,

$$E = \sum_{t=t_0+1}^{n} \sum_{i=1}^{2} \left(\widehat{\mathbf{B}}_i(\widehat{r})'\mathbf{y}_t\right)' \left(\widehat{\mathbf{B}}_i(\widehat{r})'\,\mathbf{y}_t\right) I_{t,i}(\widehat{r}). \tag{16}$$

If the threshold variable is correctly identified and r_0 is given, then $\mathbf{B}_i'\mathbf{y}_t I_{t,i}(r_0) = \mathbf{B}_i'\boldsymbol{\varepsilon}_{t,i} I_{t,i}(r_0)$. It measures the residual sum of squares after we extract the common factor process. The preferred model is the one with minimum E.

Remark 13. Although we use all samples to build a Markov regime switching model and use it for the purpose of screening, we use a cross validation type of criterion in (16) for model comparison among a small group of candidate threshold variables. The criterion in (16) is used simply for computational consideration. More sophisticated approaches such as m-fold cross-validation can be used, with additional computational cost.

Remark 14. When calculating E, the number of factors is needed. For threshold factor models, even when the number of factors is overestimated, we still can estimate the threshold value and loading spaces as shown in Theorems 4 and 5. Hence, we begin with a one-regime factor model, and estimate the largest possible value for k_0. This estimate can be used to compare different threshold variable candidates.

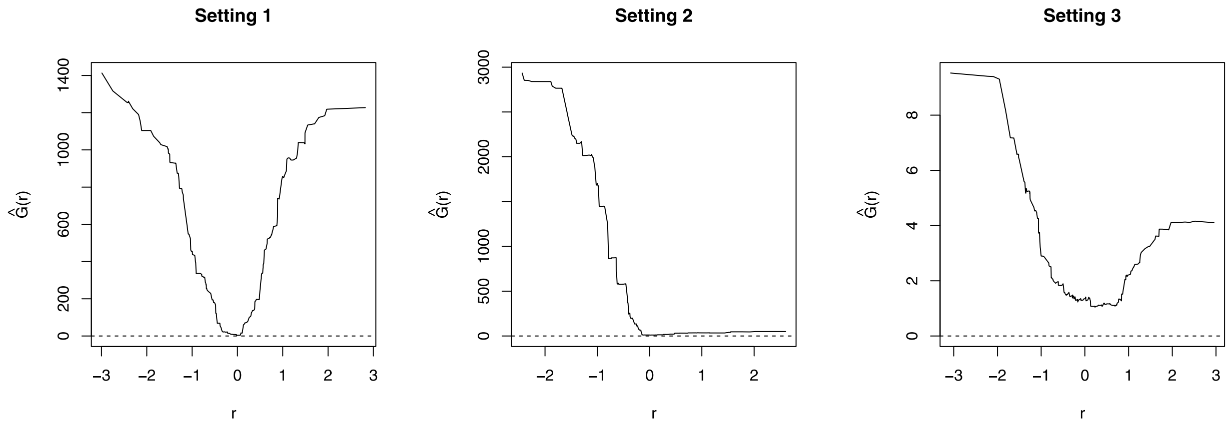

Fig. 1. Plots of $\widehat{G}(r)$ under three settings for a typical data set of Example 1, $n = 200$, $p = 20$.

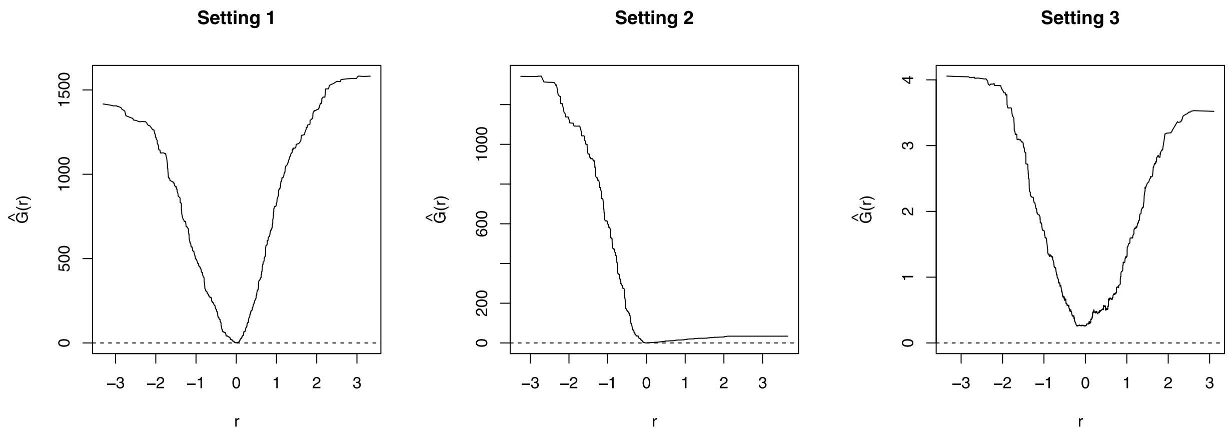

Fig. 2. Plots of $\widehat{G}(r)$ under three settings for a typical data set of Example 1, $n = 1000$, $p = 20$.

5. Simulation

In this section, we demonstrate the performance of the proposed estimators with numerical experiments, and compare the estimation errors under different settings.

In all the examples, we use $h_0 = 1$. The idiosyncratic noise process $\{\boldsymbol{\varepsilon}_{t,1}, \boldsymbol{\varepsilon}_{t,2}\}$ are independent vector white noise processes whose covariance matrix has 1 on the diagonal and 0.5 for all off-diagonal entries. For estimation of the threshold value, we use the 30th and 70th quantiles of $\{z_t\}$ as η_1 and η_2. Estimation error of $\widehat{\mathcal{M}(\mathbf{A}_i)}$ is measured by $\mathcal{D}(\widehat{\mathcal{M}(\mathbf{A}_i)}, \mathcal{M}(\mathbf{A}_i))$ where $\mathcal{D}$ is defined in (10). The error of $\widehat{r}$ is measured by $|\widehat{r} - r_0|$. Sample sizes used are $n = 200$ and 1000, and the dimensions considered are $p = 20, 40, 100$. For each setting, we repeated the simulation 100 times.

In Examples 1 to 3, we consider three settings with different regime strengths. In Setting 1, both regimes are strong, with $\delta_1 = \delta_2 = 0$. In Setting 2, one regime is strong, and the other one is weak. In Setting 3, both regimes are weak. All $p \times k_0$ entries in $\mathbf{A}_i$ were generated independently from the uniform distribution on $[-p^{-\delta_i/2}, p^{-\delta_i/2}]$ to ensure that the strength of $\mathbf{A}_i$ is δ_i. In Example 4, we discuss the impact of the distance $\mathcal{D}(\mathcal{M}(\mathbf{Q}_1), \mathcal{M}(\mathbf{Q}_2))$ of the loading spaces of the two regimes on the estimation accuracy.

Example 1. In this experiment, we use one factor process ($k_0 = 1$) following an AR(1) model with autoregressive coefficient 0.9 and N(0, 4) noise process. Weak regimes in Settings 2 and 3 are extremely weak with strength 1, which means that as p increases, we only collect noise and no more useful signal. The threshold variable z_t is independent of $\mathbf{x}_t$ and $\mathbf{y}_t$, following an AR(1) process with AR coefficient 0.3 and $N(0, 1)$ noise process. The threshold value used is $r_0 = 0$. We assume that $k_0 = 1$ is known.

Figs. 1 and 2 show the function $\widehat{G}(r)$ for three typical data sets of Example 1, one for each of the three settings, with $n = 200$ and $n = 1000$, respectively. Note that the curves are not smooth since $\widehat{G}(r)$ is only evaluated at the discrete set of all observed values of the threshold variable. We can see that $\widehat{G}(r)$ approaches the theoretical minimum value 0 of

Table 1
The relative frequency that $\widehat{r} < r_0$ when k_0 is known for Example 1.

n	200			1000		
p	20	40	100	20	40	100
Setting 1	0.48	0.48	0.47	0.52	0.49	0.53
Setting 2	0.37	0.38	0.41	0.32	0.35	0.30
Setting 3	0.44	0.50	0.52	0.47	0.51	0.54

Table 2
Average estimation errors $|\widehat{r} - r_0|$ when k_0 is known for Example 1.

n		200			1000		
p		20	40	100	20	40	100
Setting 1	$\widehat{r} < r_0$	0.046	0.070	0.049	0.020	0.017	0.021
	$\widehat{r} > r_0$	0.063	0.056	0.060	0.020	0.018	0.014
Setting 2	$\widehat{r} < r_0$	0.089	0.099	0.095	0.029	0.034	0.035
	$\widehat{r} > r_0$	0.223	0.230	0.276	0.094	0.117	0.153
Setting 3	$\widehat{r} < r_0$	0.391	0.546	0.537	0.178	0.291	0.587
	$\widehat{r} > r_0$	0.413	0.507	0.528	0.162	0.331	0.681

Table 3
Average estimation errors $\mathcal{D}(\widehat{\mathcal{M}(\mathbf{A}_i)}, \mathcal{M}(\mathbf{A}_i))$, when k_0 is known and $\widehat{r} < r_0$ for Example 1.

n		200			1000		
p		20	40	100	20	40	100
Setting 1	$\mathcal{M}(\mathbf{A}_1)$ $(\delta_1 = 0)$	0.044	0.046	0.043	0.019	0.018	0.019
	$\mathcal{M}(\mathbf{A}_2)$ $(\delta_2 = 0)$	0.054	0.052	0.050	0.022	0.021	0.023
Setting 2	$\mathcal{M}(\mathbf{A}_1)$ $(\delta_1 = 0)$	0.056	0.060	0.067	0.022	0.026	0.030
	$\mathcal{M}(\mathbf{A}_2)$ $(\delta_2 = 1)$	0.302	0.445	0.572	0.144	0.193	0.314
Setting 3	$\mathcal{M}(\mathbf{A}_1)$ $(\delta_1 = 1)$	0.318	0.637	0.921	0.116	0.165	0.681
	$\mathcal{M}(\mathbf{A}_2)$ $(\delta_2 = 1)$	0.362	0.557	0.927	0.153	0.250	0.603

Table 4
Average estimation errors $\mathcal{D}(\widehat{\mathcal{M}(\mathbf{A}_i)}, \mathcal{M}(\mathbf{A}_i))$, when k_0 is known and $\widehat{r} > r_0$ for Example 1.

n		200			1000		
p		20	40	100	20	40	100
Setting 1	$\mathcal{M}(\mathbf{A}_1)$ $(\delta_1 = 0)$	0.055	0.052	0.054	0.023	0.022	0.021
	$\mathcal{M}(\mathbf{A}_2)$ $(\delta_2 = 0)$	0.049	0.040	0.046	0.019	0.018	0.018
Setting 2	$\mathcal{M}(\mathbf{A}_1)$ $(\delta_1 = 0)$	0.069	0.073	0.071	0.032	0.032	0.033
	$\mathcal{M}(\mathbf{A}_2)$ $(\delta_2 = 1)$	0.284	0.376	0.527	0.111	0.161	0.264
Setting 3	$\mathcal{M}(\mathbf{A}_1)$ $(\delta_1 = 1)$	0.361	0.590	0.931	0.151	0.260	0.611
	$\mathcal{M}(\mathbf{A}_2)$ $(\delta_2 = 1)$	0.342	0.641	0.949	0.093	0.195	0.722

$G(r)$ around $r = 0$ when at least one regime is strong (Settings 1 and 2). When two regimes are both extremely weak, the range of $\widehat{G}(r)$ is very small and the minimum value is above 0, but also occurs around $r = 0$. In Setting 2 where two regimes have different levels of strength, $\widehat{G}(r)$ is much larger in the stronger regime $r < r_0$ than that in the weaker regime $r > r_0$, a property shown in Lemma 5 in Appendix A.2.

Table 1 reports the relative frequency that $\widehat{r} < r_0$ for different settings when k_0 is known. In Setting 2, the frequencies to underestimate r_0 are much lower than these to overestimate r_0. The results are in line with our conclusions in Theorem 2.

Tables 2 to 4 show the estimation errors of the threshold value and loading spaces. We report the results under $\widehat{r} < r_0$ and $\widehat{r} > r_0$ cases separately to highlight the impact of over- and under-estimate the threshold value. Results for threshold value estimation and loading space estimation share many common characteristics. It is seen that as sample size increases and as the strength increases, estimation improves almost in all settings. Regarding the impact of the dimension, the estimate accuracy suffers from the curse of dimensionality when one regime is weak (in Setting 2 and Setting 3); when two regimes are both strong, the accuracy does not change much with different values of p (in Setting 1).

From Table 2, it is seen that misclassification and unbalanced regime strength do have impacts on the estimation of threshold value. Estimates are more accurate when $\widehat{r} < r_0$ than when $\widehat{r} > r_0$ for Setting 2. The reason is that the order of $G(r)$ is higher in the stronger regime $(r < r_0)$ than in the weaker regime $(r > r_0)$, resulting in a much flatter curve in the region $r > r_0$ (See middle panels in Figs. 1 and 2 and Lemma 5 in Appendix A.2). Hence, it is more likely to overestimate r_0 as shown in Table 1, and even when the threshold value is underestimated, the error is much smaller as shown in Table 2.

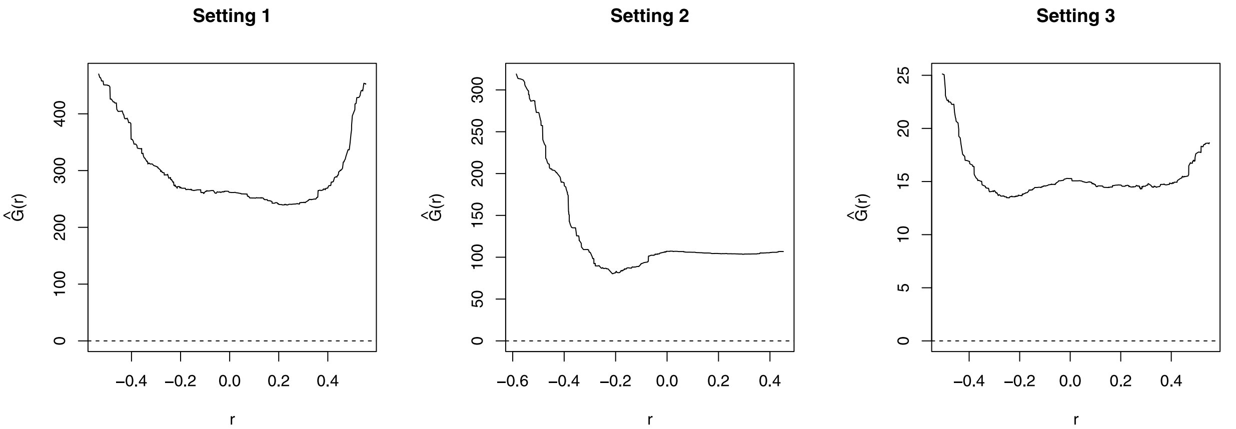

Fig. 3. Plots of $\widehat{G}(r)$ of a typical data set under each of the three settings for Example 2, when $n = 1000$, $p = 20$, and the number of factors is underestimated ($\widehat{k} = 2$).

Table 5
The relative frequency that $\widehat{k} = k_0$ for Example 2.

$\widehat{k}$	$n = 200$			$n = 1000$		
p	20	40	100	20	40	100
Setting 1	0.66	0.75	0.74	0.97	0.99	1.00
Setting 2	0.77	0.84	0.83	0.99	0.99	1.00
Setting 3	0.33	0.24	0.18	0.90	0.82	0.75

The estimation results for loading spaces also show some properties that do not apply to those for threshold value. Comparing Regime 1 in Setting 2 to both regimes in Setting 1, we see that the estimate accuracy of $\mathcal{M}(\mathbf{A}_1)$ of the stronger regime does not change much after introducing a weak regime in Tables 3 and 4. Comparing the estimation of $\mathcal{M}(\mathbf{A}_2)$ in Setting 2 to both $\mathcal{M}(\mathbf{A}_1)$ and $\mathcal{M}(\mathbf{A}_2)$ in Setting 3, we can see that the estimation of $\mathcal{M}(\mathbf{A}_2)$ of the weak regime benefits from the existence of a strong regime, especially when p is large. There is indeed a 'helping effect' for the weak regime after adding a strong regime. These observations are in line with the observations shown in Liu and Chen (2016).

Example 2. In this experiment, we investigate the performance of the proposed estimator for the number of factors k_0, and study the estimators of loading spaces and threshold value when k_0 is not correctly estimated. We also consider the case when the threshold variable is not correctly identified. The number of factors here is set to 3. The factor process is set to be three independent AR(1) processes with N(0, 4) noises process and AR coefficients 0.9, −0.7, and 0.8. The threshold variable z_t is independent of $\mathbf{x}_t$ and $\mathbf{y}_t$, following an AR(1) process with AR coefficient −0.7 and $N(0, 1)$ noise process. The threshold value is $r_0 = 0$. The strength for the weak regimes in Setting 2 and Setting 3 is set to be 0.5.

Table 5 shows the relative frequencies that $\widehat{k} = k_0$, when the true threshold variable is chosen but the threshold value r_0 is unknown, and only partial data with $\{z_t \leqslant \eta_1\}$ and $\{z_t \geqslant \eta_2\}$ are used. As n increases from 200 to 1000, the estimates improve in all settings. For the impact of regime strength, the results show that the existence of a strong regime (Setting 1 and Setting 2) results in much more accurate estimates for the number of factors k_0. Regarding the impact of p, it is seen that the estimation performance remains about the same as p increases, when one or more strong regimes exist, benefiting from a 'blessing of dimensionality'; see Lam et al. (2011). However, when both regimes are weak, the number of correct estimations may decrease as p increases. This is because the signal to noise ratio in the system decreases as p increases.

Figs. 3 and 4 plot $\widehat{G}(r)$ of a typical data set under each of the three settings for Example 2, when $\widehat{k}$ is underestimated ($\widehat{k} = k_0 - 1$) and overestimated ($\widehat{k} = k_0 + 1$), respectively, with sample size $n = 1000$. It is seen that, when $\widehat{k} = 2$, $\widehat{G}(r)$ does not show a sharp V-shape, and the minimum value is far above 0 in all panels of Fig. 3. When $\widehat{k} = 4$, $\widehat{G}(r)$ in Fig. 4 reaches its minimum value around $r = 0$ for all the settings.

Tables 6 and 7 show the estimation errors of the threshold value and loading spaces. When the number of factors is overestimated as $\widehat{k} = 4$, we can obtain consistent estimators for threshold value and loading spaces. It confirms the conclusions in Theorems 4 and 5. However, when the number of factors is underestimated, the performance of $\widehat{\mathbf{Q}}_i(\widehat{k}, \widetilde{r})$ and $\widetilde{r}$ is rather poor. This is because when defining $G(\cdot)$ we take advantage of the complement of loading spaces, $\mathcal{M}(\mathbf{B}_i)$. Smaller number of factors makes the complement space to be estimated larger than it is. As a result, the estimates are much worse when $\widehat{k} < k_0$ than that when $\widehat{k} \geqslant k_0$.

To demonstrate the properties of the threshold variable selection procedure, four candidates $\{z_{t-\ell} \mid \ell = 0, \ldots, 3\}$ are considered in this example. Table 8 reports the performance of threshold variable selection when 1000 and $t_0 = n/2$. The

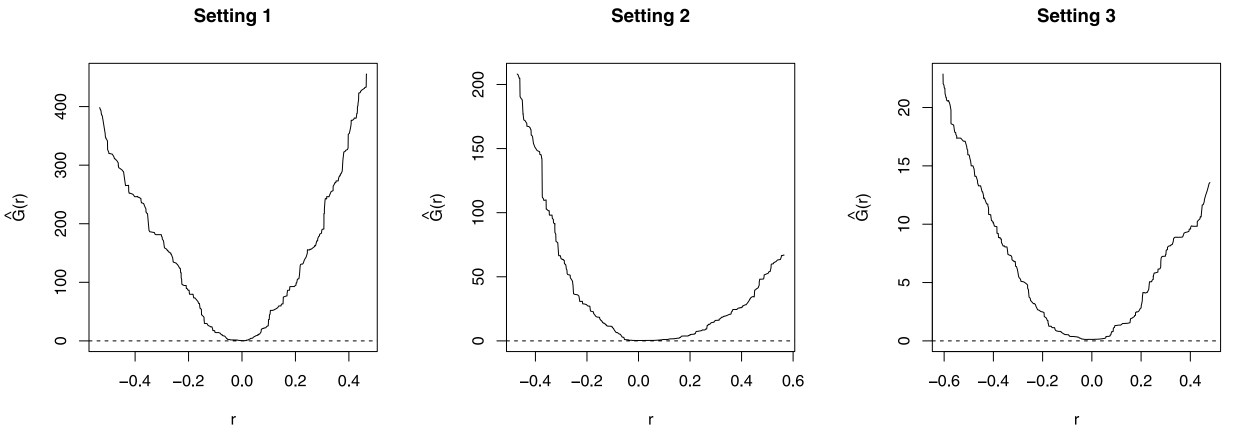

Fig. 4. Plots of $\widehat{G}(r)$ of a typical data set under each of the three settings for Example 2, when $n = 1000$, $p = 20$, and the number of factors is overestimated ($\widehat{k} = 4$).

Table 6
Average estimation error $|\widehat{r} - r_0|$ when k_0 is unknown for Example 2, $n = 1000$.

$\widehat{k}$	2 (underestimated)			3 (correctly specified)			4 (overestimated)		
p	20	40	100	20	40	100	20	40	100
Setting 1	0.440	0.458	0.474	0.035	0.036	0.039	0.012	0.012	0.012
Setting 2	0.363	0.356	0.362	0.067	0.075	0.104	0.021	0.022	0.030
Setting 3	0.446	0.441	0.471	0.090	0.117	0.184	0.026	0.029	0.031

Table 7
Average estimation errors $\mathcal{D}(\widehat{\mathcal{M}(\mathbf{A}_i)}, \mathcal{M}(\mathbf{A}_i))$ when k_0 is unknown for Example 2, $n = 1000$.

	$\widehat{k}$	2 (underestimated)			3 (correctly specified)			4 (overestimated)		
	p	20	40	100	20	40	100	20	40	100
Setting 1	$\mathcal{M}(\mathbf{A}_1)$ ($\delta_1 = 0$)	0.134	0.144	0.148	0.038	0.036	0.036	0.033	0.033	0.032
	$\mathcal{M}(\mathbf{A}_2)$ ($\delta_1 = 0$)	0.160	0.166	0.180	0.036	0.036	0.034	0.032	0.032	0.032
Setting 2	$\mathcal{M}(\mathbf{A}_1)$ ($\delta_1 = 0$)	0.039	0.034	0.034	0.044	0.043	0.043	0.037	0.038	0.038
	$\mathcal{M}(\mathbf{A}_2)$ ($\delta_1 = 0.5$)	0.466	0.582	0.695	0.091	0.110	0.150	0.083	0.096	0.131
Setting 3	$\mathcal{M}(\mathbf{A}_1)$ ($\delta_1 = 0.5$)	0.161	0.193	0.190	0.087	0.105	0.164	0.074	0.087	0.106
	$\mathcal{M}(\mathbf{A}_2)$ ($\delta_1 = 0.5$)	0.174	0.179	0.215	0.084	0.105	0.158	0.072	0.083	0.104

Table 8
The relative frequency to correctly identify the threshold variable when k_0 is unknown and $n = 1000$ for Example 2.

$\widehat{k}$	2 (underestimated)			3 (correctly specified)			4 (overestimated)		
p	20	40	100	20	40	100	20	40	100
Setting 1	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
Setting 2	0.30	0.17	0.05	1.00	1.00	1.00	1.00	1.00	1.00
Setting 3	0.90	0.96	0.85	1.00	1.00	1.00	1.00	1.00	1.00

results show that our method can identify the threshold variable correctly when k_0 is correctly specified or overestimated. However, underestimating k_0 may lead to poor results, especially in Setting 2.

Fig. 5 plots $\widehat{G}(r)$ of a typical data set under each of the three settings of Example 2, when z_{t-1} is used as the threshold variable. When the threshold variable is not correctly specified, $\widehat{G}(r)$ does not show a clear V-shape curve. There may be multiple minimum points, and the minimum value is significantly larger than 0.

Example 3. In this experiment, we examine the estimation performance in a more complicated context, where the threshold variable $\mathbf{z}_t$ is correlated to the lag variable of $\mathbf{y}_t$. Specifically, the threshold variable used is the cross-sectional standard deviation of $\mathbf{y}_{t-1}$,

$$z_t = \sqrt{\frac{1}{p-1}\sum_{q=1}^{p}(y_{t-1,q} - \bar{y}_{t-1})^2},$$

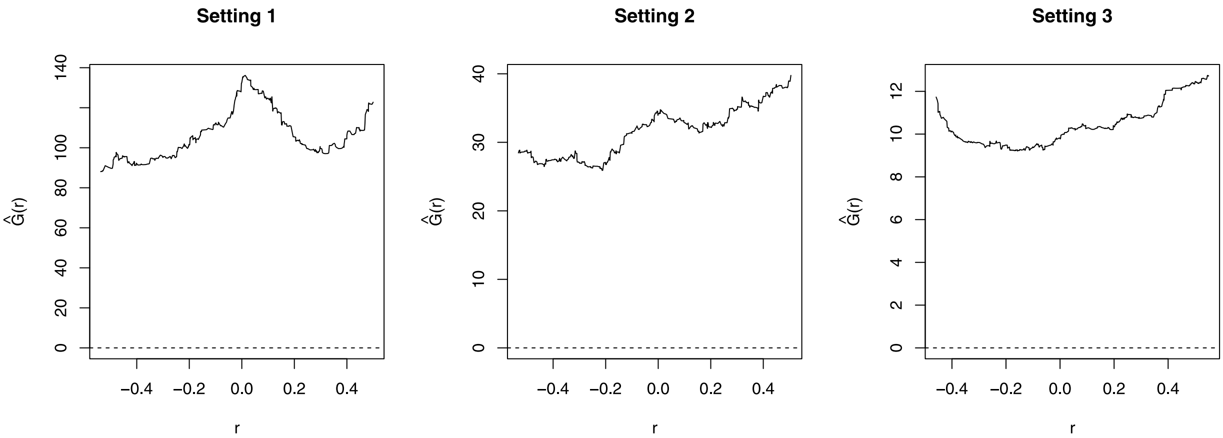

Fig. 5. Plots of $\widehat{G}(r)$ of a typical data set under each of the three settings of Example 2, when $n = 1000$, $p = 20$, and z_{t-1} is used as the threshold variable instead of true threshold variable z_t.

Table 9
Average estimation errors $|\widehat{r} - r_0|$ when k_0 is known for Example 3.

n	200			1000		
p	20	40	100	20	40	100
Setting 1	0.174	0.162	0.187	0.056	0.052	0.047
Setting 2	0.085	0.072	0.094	0.037	0.034	0.032
Setting 3	0.093	0.104	0.128	0.042	0.057	0.076

Table 10
Average estimation errors $\mathcal{D}(\widehat{\mathcal{M}(\mathbf{A}_i)}, \mathcal{M}(\mathbf{A}_i))$ when k_0 is known for Example 3.

n		200			1000		
p		20	40	100	20	40	100
Setting 1	$\mathcal{M}(\mathbf{A}_1)$ ($\delta_1 = 0$)	0.142	0.143	0.143	0.053	0.051	0.052
	$\mathcal{M}(\mathbf{A}_2)$ ($\delta_2 = 0$)	0.032	0.032	0.032	0.013	0.013	0.013
Setting 2	$\mathcal{M}(\mathbf{A}_1)$ ($\delta_1 = 0$)	0.113	0.096	0.079	0.052	0.045	0.037
	$\mathcal{M}(\mathbf{A}_2)$ ($\delta_2 = 0.5$)	0.007	0.085	0.139	0.027	0.033	0.054
Setting 3	$\mathcal{M}(\mathbf{A}_1)$ ($\delta_1 = 0.5$)	0.279	0.323	0.436	0.101	0.118	0.118
	$\mathcal{M}(\mathbf{A}_2)$ ($\delta_2 = 0.5$)	0.071	0.089	0.142	0.029	0.040	0.070

where $y_{t,q}$ is the qth entry in $\mathbf{y}_t$ and $\bar{y}_t = \sum_{q=1}^{p} y_{t,q}/p$, for $t = 1, \ldots, n$. The factor process is generated from an AR(1) process with AR coefficient 0.9 and N(0, 4) noise process. The strength of the weak regimes in Setting 2 and Setting 3 is set to be 0.5. The threshold values used are $r_0 = 1.5$, 1.2 and 1 for Settings 1 to 3, respectively.

Tables 9 and 10 present the estimation errors for threshold value and loadings spaces. They show a similar pattern to the results in Example 1. Note that the three settings are not comparable in this case because each setting yields different sample size in each regime due to the dependency of z_t and $\mathbf{y}_{t-1}$.

Example 4. We present the influence of the distance of loading spaces $\mathcal{D}(\mathcal{M}(\mathbf{Q}_1), \mathcal{M}(\mathbf{Q}_2))$ on the estimation accuracy. One factor ($k_0 = 1$) is used since the distance between two one-dimensional spaces is easier to interpret. For simplicity, strong regimes $\delta_1 = \delta_2 = 0$ are used. The latent factor $\{x_t\}$ follows an AR(1) process with AR coefficient 0.9 and $N(0, 1)$ noise process. $\{z_t\}$ follows i.i.d $N(0, 1)$, and threshold value is 0. The ith entries in $\mathbf{A}_1$ and $\mathbf{A}_2$ are generated from a bivariate normal distribution with unit variances and covariance $(1 - d^2)$ so that $\mathcal{D}(\mathcal{M}(\mathbf{A}_1), \mathcal{M}(\mathbf{A}_2)) \approx |d|$.

Table 13 reports the average estimation errors for threshold value. It is not surprising that when the two loading spaces are well apart, the estimates are more accurate. For the case of $d = 0$ (there is no switching mechanism), the threshold value r_0 is unavailable ($r_0 = 0$ is used in the calculation in the table). The forcedly estimated $\widehat{r}$ is random in the middle of the range of $z_t \sim N(0, 1)$.

Table 11 reports the average estimation errors of the loading spaces. As expected, the accuracy does not change as p increases since both regimes are strong. Larger sample size n improves the accuracy. It is noted that the distance between the two loading spaces has almost no effect on the estimation accuracy of the loading spaces, even the estimated threshold values are not accurate when the distance is small as shown in Table 13. One possible reason may be that the distortion

Table 11
Average estimation errors $\mathcal{D}(\widehat{\mathcal{M}(\mathbf{A}_i)}, \mathcal{M}(\mathbf{A}_i))$ when k_0 is known for Example 4.

n		200			1000		
p		20	40	100	20	40	100
$d=1$	Regime 1	0.054	0.054	0.054	0.024	0.024	0.023
	Regime 2	0.052	0.054	0.050	0.023	0.023	0.024
$d=0.7$	Regime 1	0.052	0.053	0.049	0.023	0.022	0.022
	Regime 2	0.052	0.051	0.052	0.022	0.021	0.021
$d=0.3$	Regime 1	0.050	0.052	0.051	0.022	0.021	0.020
	Regime 2	0.052	0.049	0.051	0.021	0.022	0.022
$d=0.2$	Regime 1	0.051	0.051	0.051	0.024	0.021	0.020
	Regime 2	0.052	0.051	0.048	0.022	0.022	0.022
$d=0.1$	Regime 1	0.050	0.048	0.051	0.023	0.023	0.021
	Regime 2	0.049	0.048	0.052	0.022	0.021	0.021
$d=0$	Regime 1	0.049	0.048	0.049	0.023	0.021	0.022
	Regime 2	0.054	0.049	0.048	0.022	0.031	0.021

Table 12
Average and standard deviation (in brackets) of $\mathcal{D}(\widehat{\mathcal{M}(\mathbf{A}_1)}, \widehat{\mathcal{M}(\mathbf{A}_2)})$ for Example 4.

n	$n=200$			$n=1000$		
p	20	40	100	20	40	100
$d=1$	0.972(0.041)	0.986(0.018)	0.993(0.007)	0.9723(0.036)	0.990(0.013)	0.995(0.007)
$d=0.7$	0.685(0.117)	0.696(0.083)	0.688(0.051)	0.701(0.100)	0.696(0.083)	0.697(0.042)
$d=0.3$	0.290(0.066)	0.291(0.055)	0.304(0.031)	0.300(0.068)	0.301(0.041)	0.296(0.029)
$d=0.2$	0.199(0.045)	0.202(0.035)	0.199(0.026)	0.190(0.047)	0.192(0.033)	0.195(0.019)
$d=0.1$	0.118(0.037)	0.117(0.028)	0.116(0.018)	0.099(0.025)	0.099(0.016)	0.100(0.012)
$d=0$	0.075(0.034)	0.074(0.032)	0.069(0.027)	0.029(0.011)	0.031(0.011)	0.040(0.097)

Table 13
Average estimation errors $|\hat{r} - r_0|$ when k_0 is known for Example 4.

n	$n=200$			$n=1000$		
p	20	40	100	20	40	100
$d=1$	0.043	0.048	0.050	0.020	0.019	0.019
$d=0.7$	0.057	0.061	0.047	0.020	0.018	0.016
$d=0.3$	0.099	0.086	0.086	0.035	0.033	0.033
$d=0.2$	0.097	0.114	0.090	0.052	0.042	0.041
$d=0.1$	0.180	0.148	0.161	0.092	0.079	0.084
$d=0$	0.278	0.190	0.187	0.174	0.192	0.180

created by the misclassified observations (due to the misplaced threshold value) is relatively small, as the two loading spaces are close. In the case of $d = 0$, there is no distortion.

Table 12 shows the average and sample standard deviation of the distance between the estimated loading spaces. It is seen that the distances of the estimated loading spaces are close to the true distance with small variation, since the estimated loading spaces are estimated accurately as shown in Table 11. The variation of the estimated space distance is smaller for larger p, as in this case the distance is related to the sample correlation of two vectors ($\widehat{\mathbf{Q}}_1$ and $\widehat{\mathbf{Q}}_2$) of length p. The variation does not change much for larger n.

6. Real example

We applied the proposed approach to the daily returns of 123 stocks from January 2, 2002 to July 11, 2008. These stocks were selected among those included in the S&P 500 and traded every day during the period. The returns were calculated in percentages based on daily closing prices. This data set was analyzed by Lam and Yao (2012), Chang et al. (2015) and Liu and Chen (2016). The sample size n is 1642 and the dimension p is 123. We use $h_0 = 1$ in this analysis.

Lam and Yao (2012) and Chang et al. (2015) used a factor model (with no switching) to analyze the data. The estimated number of factors is 2. Liu and Chen (2016) used a Markov switching factor model on the same data set and found that there are two regimes, with one factor in each regime. Here we analyze the data using a 2-regime threshold factor model. We consider the lag cross-sectional standard deviation of $\mathbf{y}_{t-\ell}$ and the lag of squared S&P 500 return $r_{t-\ell}$ as the candidates for the threshold variable ($\ell = 1, \ldots, 8$). Since an overestimated number of factors still can identify the threshold variable, $\hat{k} = 2$ is tentatively used when calculating E defined in (16). Table 14 shows the value of E for each candidate with $t_0 = n/2$. In the following we use the cross-sectional standard deviation of $\mathbf{y}_{t-6}$ as the threshold variable, since it minimizes E.

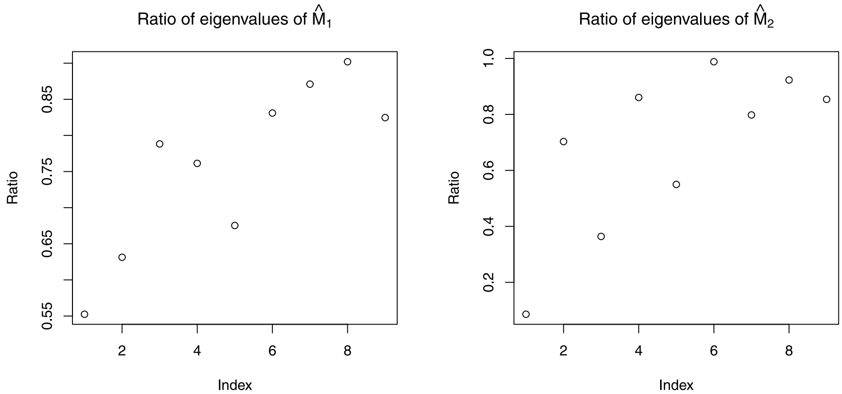

Fig. 6. Ratios of eigenvalues for real data analysis; left panel: ratio of eigenvalues of $\widehat{\mathbf{M}}_1(\eta_1, \eta_2)$; right panel: ratio of eigenvalues of $\widehat{\mathbf{M}}_2(\eta_1, \eta_2)$.

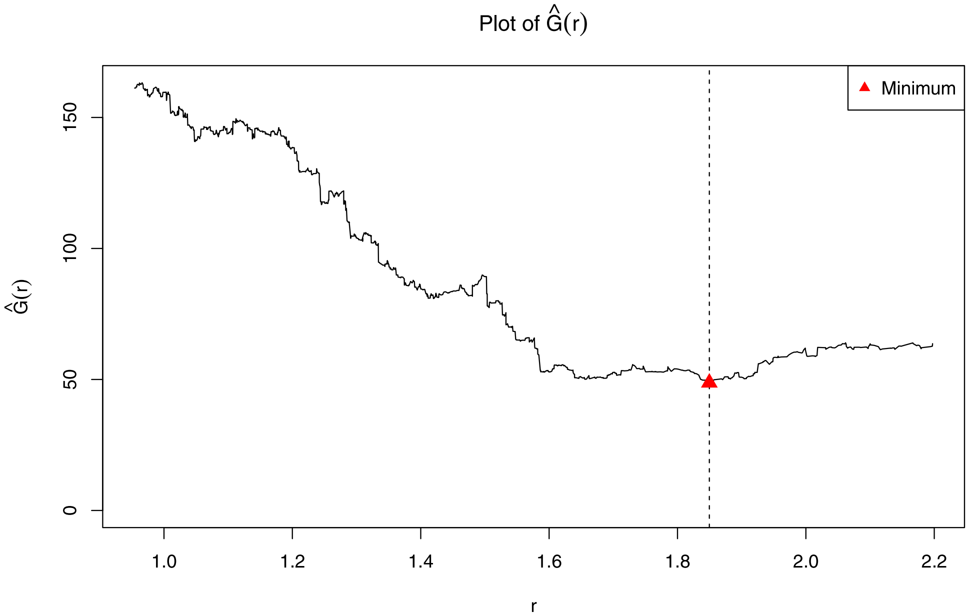

Fig. 7. Plot of $\widehat{G}(r)$ for real data analysis.

Table 14
E divided by 10^5 for all threshold variable candidates for real data analysis.

Lag	z_{t-1}	z_{t-2}	z_{t-3}	z_{t-4}	z_{t-5}	z_{t-6}	z_{t-7}	z_{t-8}
Lag of cross-section standard deviation	2.097	2.051	2.060	1.984	2.091	1.975	2.067	1.993
Lag of squared S&P return	2.069	2.053	2.069	2.164	2.043	2.069	2.086	2.055

We use 10th and 90th percentiles of the threshold variable as η_1 and η_2 to estimate the number of factors. The left and right panels in Fig. 6 display the ratio of eigenvalues of $\widehat{\mathbf{M}}_1(\eta_1, \eta_2)$ and $\widehat{\mathbf{M}}_2(\eta_1, \eta_2)$, respectively, where both the ratios reach their minimum values at 1. It yields that $\widehat{k} = 1$.

The above results indicate that only one factor drives the 123 stocks, but the factor loadings switch between regimes according to the cross-sectional standard deviation of all stock returns 6 trading days before. Ignoring switching structure as in Lam and Yao (2012), it would appear that there are two different factors. Introducing a threshold variable reduces the number of factor by 1.

Fig. 7 plots $\widehat{G}(r)$, showing a V-shape curve with a relatively flat bottom. By minimizing $\widehat{G}(r)$, we have that $\widehat{r} = 1.850$.

Note that $\mathbf{R}_t$ is the common factor process with standardized loading matrix defined in (3), and can be estimated through (13). Fig. 8 displays the time series plot of $\widehat{\mathbf{R}}_t(\widehat{r})$ on the top panel and the daily returns of S&P 500 index on the bottom panel. The estimated signal $\widehat{\mathbf{R}}_t(\widehat{r})$ in this period was closely correlated with S&P 500 index except for several days

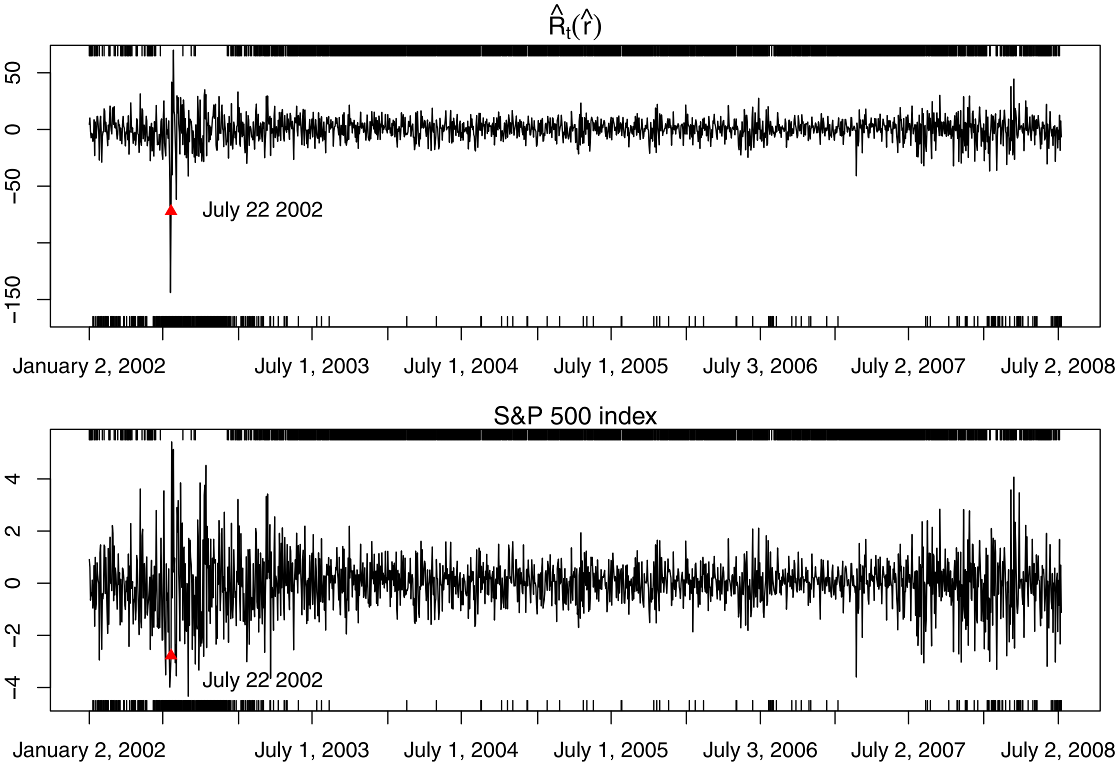

Fig. 8. Time series plots of $\widehat{\mathbf{R}}_t(\widehat{r})$ (top panel) and the daily return of the S&P 500 index (bottom panel) in the same period. Indicators of the estimated regimes of the observations $I_{t,i}(\widehat{r})$ for $i = 1, 2$, are shown in the rug plots, on the top for Regime 1 and at the bottom for Regime 2.

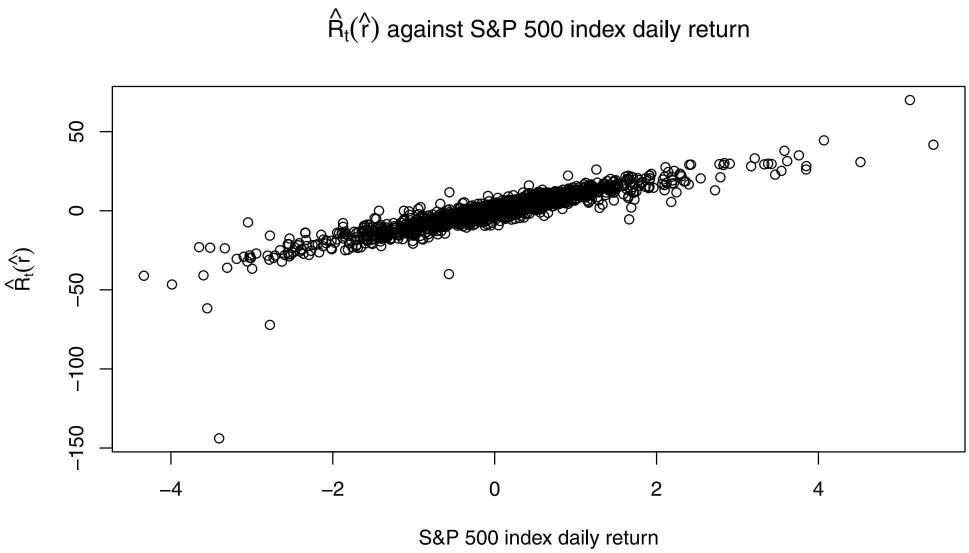

Fig. 9. Plot of $\widehat{\mathbf{R}}_t(\widehat{r})$ against the daily return of the S&P 500 index.

around July 22, 2002. Fig. 9 plots $\widehat{\mathbf{R}}_t(\widehat{r})$ against daily returns of the S&P 500 index, and the correlation is 0.910. Hence, this factor can be regarded as a representation of market performance, which is in line with results in Liu and Chen (2016).

The estimated indicator functions for regimes $\{I_{t,i}(\widehat{r}), i = 1, 2\}$, are shown in the rug plots of both panels in Fig. 8. Regime 1 is plotted on the top and Regime 2 at the bottom. When the market was volatile in 2002, 2003, and 2008 due

Table 15
In-sample residual sum of squares of different models for real data analysis.

Model	Residual sum of squares
Threshold factor models	606 256.9
Regime-switching factor models	713 742.6

to the internet bubble, the invasion of Iraq, and the subprime crisis, respectively, the observations are more likely to belong to Regime 1; when the market was stable in 2004–2007, the observations tend to be in Regime 2.

The distance between the two estimated loading spaces defined in (10) is 0.763. Since the loading spaces are one-dimensional (lines), the distance measures the absolute value of sine of the angle between the two lines. Although we do not have asymptotic distribution results for the estimated distance, the variation of estimated distance in Example 4 partially suggests that the two estimated spaces are quite different.

To compare threshold factor models with regime-switching factor models (Liu and Chen, 2016), we find the optimal path for the regime-switching factor model, calculate the in-sample residual sums of squares defined in (16) for both models, and report in Table 15. It shows that threshold factor models outperform the other.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2020.01.005.

References

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.
Balke, N., Fomby, T., 1997. Threshold cointegration. Int. Econ. Rev. 38, 627–645.
Caner, M., Hansen, B., 2004. Instrumental variable estimation of a threshold model. Econometric Theory 20, 813–843.
Chan, K., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. Ann. Statist. 21, 520–533.
Chan, K., Tong, H., 1990. On likelihood ratio tests for threshold autoregression. J. R. Stat. Soc. Ser. B Stat. Methodol. 52, 469–476.
Chang, J., Guo, B., Yao, Q., 2015. High dimensional stochastic regression with latent factors, endogeneity and nonlinearity. J. Econometrics 189, 297–312.
Chen, R., 1995. Threshold variable selection in open-loop threshold autoregressive models. J. Time Series Anal. 16, 461–481.
Chen, C., So, M., 2006. On a threshold heteroscedastic model. Int. J. Forecast. 22, 73–89.
Doz, C., Giannone, D., Reichlin, L., 2011. A two-step estimator for large approximate dynamic factor models based on kalman filtering. J. Econometrics 164, 188–205.
Forbes, C., Kalb, G., Kofhiar, P., 1999. Bayesian arbitrage threshold analysis. J. Bus. Econom. Statist. 17, 364–372.
Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic factor model: identification and estimation. Rev. Econ. Stat. 82, 540–554.
Forni, M., Lippi, M., 2001. The generalized factor model: representation theory.. Econometric Theory 17, 1113–1141.
Forni, M., Lippi, M., Reichlin, L., 2003. Do financial variables help forecasting inflation and real activity in the euro area?. J. Monetary Econ. 50, 1243–1255.
Forni, M., Lippi, M., Reichlin, L., 2004. The generalized dynamic factor model: consistency and rates. J. Econometrics 119, 231–255.
Forni, M., Reichlin, L., 1998. Let's get real: a factor-analytical approach to disaggregated business cycle dynamics. Rev. Econom. Stud. 65, 453–473.
Geweke, J., 1977. In: Aigner, D., Goldberg, A. (Eds.), Latent variable in socio-economic models. In: The Dynamic Factor Analysis of Economic Time Series, North Holland, Amsterdam, pp. 365–383.
Gourieroux, C., Monfort, A., 1992. Qualitative threshold ARCH models. J. Econometrics 52, 159–199.
Hallin, M., Liska, R., 2007. Determining the number of factors in the general dynamic factor model.. J. Amer. Statist. Assoc. 102, 603–617.
Kim, C.J., Nelson, C.R., 1998. Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. Rev. Econ. Stat. 80, 188–201.
Lam, C., Yao, Q., 2012. Factor modeling for high-dimensional time series: inference for the number of factors. Ann. Statist. 40 (2), 694–726.
Lam, C., Yao, Q., Bathia, N., 2011. Estimation of latent factors for high-dimensional time series. Biometrika 98 (4), 901–918.
Li, C.W., Li, W.K., 1996. On a double-threshold autoregressive heteroscedastic time series model. J. Appl. Econometrics 11, 253–274.
Liu, X., Chen, R., 2016. Regime-switching factor models for high-dimensional time series. Statist. Sinica 26, 1427–1451.
Pan, J., Yao, Q., 2008. Modelling multiple time series via common factors. Biometrika 95, 365–379.
Peña, D., Box, G.E.P., 1987. Identifying a simplifying structure in time series. J. Amer. Statist. Assoc. 82, 836–843.
Peña, D., Poncela, P., 2006. Nonstationary dynamic factor analysis. J. Statist. Plann. Inference 136, 1237–1257.
Petruccelli, J., Davies, N., 1986. A portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. Biometrika 73, 687–694.
Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indices. J. Bus. Econom. Statist. 20, 147–162.
Stock, J.H., Watson, M.W., 2005. Implications of Dynamic Factor Models for VAR Analysis. Working Paper 11467, National Bureau of Economic Research.
Tiao, G., Box, G.E.P., 1981. Modeling multiple time series with applications. J. Amer. Statist. Assoc. 76, 802–816.
Tiao, G., Tsay, R., 1989. Model specification in multivariate time series. J. R. Stat. Soc. Ser. B Stat. Methodol. 51, 157–213.
Tong, H., 1990. Nonlinear Time Series: A Dynamic System Approach. Clarendon Press, Oxford.
Tong, H., 1993. Threshold Models in Non-linear Time Series Analysis. Springer-Verlag, New York.
Tong, H., Lim, K., 1980. Threshold autoregression, limit cycles and cyclical data. J. R. Stat. Soc. Ser. B Stat. Methodol. 42 (3), 245–292.
Tsay, R., 1989. Testing and modeling threshold autoregressive process. J. Amer. Statist. Assoc. 84, 231–240.
Tsay, R., 1998. Testing and modeling multivariate threshold models. J. Amer. Statist. Assoc. 93, 1188–1202.
Wang, D., Liu, X., Chen, R., 2019. Factor models for matrix-valued high-dimensional time series. J. Econometrics 208, 231–248.
Wu, S., Chen, R., 2007. Threshold variable selection and threshold variable driven switching autoregressive models. Statist. Sinica 17, 241–264.
Zakoian, J., 1994. Threshold heteroskedastic models. J. Econ. Dyn. control 18, 931–955.