Shrinkage Estimation of Multiple Threshold Factor Models

Chenchen Ma[a] and Yundong Tu[b*]
[a]School of Mathematical Sciences and
Center for Statistical Science, Peking University
[b]Guanghua School of Management and
Center for Statistical Science, Peking University

February 5, 2023

Abstract

This paper proposes a multiple threshold factor model to enhance the flexibility in modeling the underlying regime switching mechanism for high dimensional time series. The factor loadings are assumed to switch between different regimes according to the value of a threshold variable. A novel estimation procedure is proposed to consistently estimate the multiple thresholds with the aid of sorting operation, principal component analysis and shrinkage estimation, which is practically easy-to-implement and computationally efficient. Furthermore, asymptotic properties for the multiple threshold estimators are established, together with other theoretical results. Monte Carlo simulations demonstrate that the procedure works well in finite samples. The U.S. data sets are analyzed with the proposed model to illustrate the threshold effect of economy policy uncertainty on the financial market.

JEL classification: C22; C33; C38; C51; G10.

Keywords: Dimension reduction; Group Lasso; Information criterion; Nonlinear factor model; Principal component analysis.

*Corresponding author. Address: Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, 100871, China. E-mail: yundong.tu@gsm.pku.edu.cn.

1 Introduction

With the increasing availability of large data sets in economics and finance, the large factor model has become one of the most important tools to achieve dimension reduction in the statistical and econometric analysis. To capture the instability caused by economic condition shifts or policy reforms, factor models with structural breaks in the factor loadings are accordingly developed. See Breitung and Eickmeier (2011), Chen et al. (2014), Han and Inoue (2015) and Yamamoto and Tanaka (2015) for testing procedures to detect breaks, and Baltagi et al. (2017, 2021), Su and Wang (2017), Ma and Su (2018), and Ma and Tu (2022) for the estimation theory. These developments facilitate the analysis of large data sets in the presence of important economic, political events and policy changes that induce low-frequency episodes separated in time, such as the European debt crisis, the end of the Cold War or the end of China's one-child policy.

On the other hand, recurring regime shifts that relate to higher frequency recurring fluctuation arise in situation where "history repeats" (Ang and Timmermann, 2012). The behavior of economic variables such as risk-free rate, unemployment and economic growth may vary under different fiscal policies (neutral, expansionary, or contractionary) (Kim and Nelson, 1998; Liu and Chen, 2020). As pointed out by Chan et al. (2017), among others, threshold model is an important strategy to account for recurring regimes. Accordingly, threshold factor models, which allow recurring regime shifts in the factor loadings according to the magnitude of a (continuous) threshold variable, have been developed. For example, Ng and Wright (2013) simulated data from a large dimensional threshold factor model to investigate the effects of nonlinearities on business cycle dynamics. Massacci (2017) studied a large dimensional factor model with regime changes in the loadings governed by the threshold principle. Liu and Chen (2020) also considered a threshold factor model for high-dimensional time series in which the dynamics of the time series was assumed to switch between two regimes. Wu (2021) further extended the analysis of Massacci (2017) to a factor model with multiple threshold-type regime shifts in the factor loadings. Massacci (2021) developed a test for threshold-type regimes in the risk exposures in portfolios with a large number of financial assets whose returns exhibit an approximate factor structure. The threshold factor model not only provides a powerful tool for dimension reduction, but also enhances the flexibility in modeling the recurring regime shifting mechanism driven by the threshold principle, offering an interpretable framework that can easily capture the underlying nonlinearity.

This paper proposes a shrinkage estimation procedure to efficiently and consistently estimate the number and values of thresholds in multiple threshold factor models. The methodology proposed is motivated by Tsay (1989) and Chan et al. (2015), who represented the threshold autoregressive model as an autoregressive model with structural breaks, and Ma and Tu (2022) who used group Lasso (Yuan and Lin, 2006) to estimate factor models with multiple structural breaks. The procedure involves first reformulating the threshold factor model into the factor model with structural breaks, by sorting the observed high dimensional time series according to

the value of the threshold variable in ascending order. This is made possible thanks to the fact that the threshold models are similar to structural break models, except that the breaks occur according to the value of a (continuous) threshold variable rather than the time index. The second step applies the structural break detection method for factor models recently proposed by Ma and Tu (2022) to pin down the break dates, which will in turn offer estimates for the original unknown threshold values. The second step is comprised of two sub-steps: (*i*) ignoring the change points occurred in the factor model, use the information criterion proposed by Bai and Ng (2002) to estimate the number of factors and principal component analysis (PCA) to estimate the (pseudo) factors; (*ii*) use the shrinkage technique to detect the number and locations of the breaks in the factor regression coefficients, where the factor regression is constructed by regressing the estimated factor corresponding to the largest eigenvalue on the remaining ones. The last step is based on the one-to-one correspondence between the breaks in the factor model and the change points in the factor regression coefficients. Consequently, the threshold values are estimated from the locations of estimated change points and the corresponding values of the threshold variable. The details will be outlined in the following section.

Compared with the two-step sequential estimation procedure of Wu (2021), our new shrinkage approach is attractive for at least two aspects. First, the shrinkage estimation is much more appealing in terms of both computational efficiency and estimation accuracy. The procedure of Wu (2021) divides the whole time span into several subintervals and determines the number of thresholds by comparing the number of factors in consecutive intervals, while ours only fits a single factor model for the whole time span and then uses group Lasso to determine the thresholds, which significantly relieves the computational burden, especially when the number of thresholds is large. On the one hand, sample division inevitably introduces extra tuning parameters such as the number of subintervals. On the other hand, it notably reduces the sample size in each regime, which will affect the estimation of factor number, and consequently the estimation accuracy of the thresholds. Second, unlike Wu (2021), our framework allows the factor number to vary across regimes, which includes interesting possibilities, such as emergence of new factors or disappearance of old factors, considered in the recent factor model literature (Cheng et al., 2016; Baltagi et al., 2021; Ma and Tu, 2022). Neither does Massacci (2017) nor Liu and Chen (2020) allow such a scenario in the single threshold factor model. These desirable features are confirmed in the simulation studies.

The theoretical contributions of this paper include the consistency of the group Lasso estimators, and the consistency of the estimated number of factors and factor space in each regime separated by the threshold estimators, among others. The theory incorporates the small threshold effects which can shrink to zero as the time length T increases, a case which has not been entertained in factor models analyzed by Massacci (2017), Liu and Chen (2020) or Wu (2021). We emphasize that threshold estimation in factor models is much more complex than that in the standard regressions, because the (unobserved) factors are not identified (up to rotation) when

the thresholds are unknown. Ignoring the thresholds, only the pseudo factors (up to rotation) instead of the true factors (up to rotation) can be identified in the linear factor model, which will be elaborated in the following section. Besides, the theoretical results of threshold factor models are not simple extensions of those in the change-point framework. Specifically, the factors are no longer stationary mixing after sorting the observations, even if we assume that the true factors are. Therefore, the ordinary large deviation principle for the stationary mixing processes, the key tool used to analyze the properties of change point estimators by Ma and Tu (2022), can not apply in the threshold factor models. To solve the problem, we develop the asymptotic analysis using a bracketing entropy technique inspired by Chan et al. (2015), to derive the convergence rate of the estimators, which brings technical differences and additional challenges as compared to the theoretical study of Ma and Tu (2022).

The rest of the paper is organized as follows. Section 2 introduces the multiple threshold factor model, and discusses the shrinkage estimation procedure for the unknown thresholds. Section 3 contains the asymptotic results, including the consistency of the group Lasso estimators of the thresholds, and a refined two-step procedure to improve the Lasso estimators. Section 4 conducts Monte Carlo simulations to investigate the finite sample performance of the proposed procedure, and demonstrate its superiority over that of Wu (2021). Section 5 applies our method to explore the potential threshold effect of economic policy uncertainty on the financial market. Section 6 concludes the paper with discussions on future research. Due to space limit, the inference for the factor number and factor space in each regime, the empirical study to investigate the influence of economic condition on the macroeconomy and all technical details are relegated to an online Supplementary Material.

A word on notations. Throughout the paper, for a symmetric matrix $\boldsymbol{B}$, we use $\boldsymbol{B} > 0$ to denote that $\boldsymbol{B}$ is positive definite, and denote its Moore-Penrose generalized inverse as $\boldsymbol{B}^+$, and its trace as $tr(\boldsymbol{B})$. For any real matrix $\boldsymbol{A}$, we denote its transpose as $\boldsymbol{A}^\top$, its rank as $\text{rank}(\boldsymbol{A})$, and its Frobenius norm as $\|\boldsymbol{A}\| \equiv [tr(\boldsymbol{A}\boldsymbol{A}^\top)]^{1/2}$. We write $\boldsymbol{v} = O_p(1)$ ($\boldsymbol{v} = o_p(1)$), when $\|\boldsymbol{v}\| = O_p(1)$ ($\|\boldsymbol{v}\| = o_p(1)$) for a finite dimensional vector $\boldsymbol{v}$. For any positive numbers a_n and b_n, let $a_n \gg b_n$ denote $a_n^{-1} b_n = o(1)$. For any constants a and b, let $a \wedge b = \min\{a, b\}$. For any two sets A and B, let $A \backslash B = \{x | x \in A \text{ and } x \notin B\}$, and let $\mathcal{D}(A, B) = \sup_{b \in B} \inf_{a \in A} |a - b|$. The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ signifies convergence in distribution.

2 Methodology

2.1 Threshold factor models

The high dimensional time series $\{x_{it}, i = 1, \ldots, N, t = 1, \ldots, T\}$ is assumed to follow a threshold factor model with multiple regimes, separated by the magnitude of the threshold variable z_t. Specifically, when $\theta_{k-1,0} < z_t \leq \theta_{k,0}$, for $k = 1, \ldots, m+1$, the factor model assumes that the

observed data x_{it} in the k-th regime can be represented as

$$x_{it} = \boldsymbol{f}_{0,t}^{\top}\boldsymbol{\lambda}_{0,i} + \boldsymbol{f}_{-0,t}^{\top}\boldsymbol{\lambda}_{k,i} + e_{it}, \tag{2.1}$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The r dimensional factor $\boldsymbol{f}_t = (\boldsymbol{f}_{0,t}^{\top}, \boldsymbol{f}_{-0,t}^{\top})^{\top}$ and the factor loading $\boldsymbol{\lambda}_{0k,i} = (\boldsymbol{\lambda}_{0,i}^{\top}, \boldsymbol{\lambda}_{k,i}^{\top})^{\top}$ in the k-th regime are decomposed conformably according to whether the loading changes with the regimes or not. That is, the loading vector $\boldsymbol{\lambda}_{0,i}$ associated to the $r-q$ dimensional factor $\boldsymbol{f}_{0,t}$ is not affected by the threshold variable z_t, thus is constant over the whole time span. However, the loading $\boldsymbol{\lambda}_{k,i}$ for the q dimensional factor $\boldsymbol{f}_{-0,t}$ depends on the regime k, such that $\boldsymbol{\lambda}_{k,i} \neq \boldsymbol{\lambda}_{k+1,i}$ for $k = 1, \ldots, m$. The threshold variable $z_t \in \mathcal{R}$ is observed but $\theta_{k,0}$'s $(k = 1, \ldots, m)$ are the unknown threshold values. We adopt the convention that $\theta_{0,0} = -\infty$ (or the lower bound of the support of z_t) and $\theta_{m+1,0} = \infty$ (or the corresponding upper bound). Let $\mathcal{A} = \{\theta_{1,0}, \ldots, \theta_{m,0}\}$ be the set of true thresholds. The number of thresholds m and the number of factors r are both assumed to be fixed.

The model in (2.1) bears similarity to the class of threshold models studied by Tong and Lim (1980), Tsay (1989), Gonzalo and Pitarakis (2002), Li and Ling (2012), Chan et al. (2015), among others. The regime for the loadings prevailing at time t depends on the position of z_t, relative to the unknown threshold values $\theta_{k,0}$, for $k = 1, \ldots, m$. The major difference lies in the fact that the regressors in the above studies are observed, while the factors here are unobserved and the number of factors is unknown. This model nests the large dimensional linear factor models (Bai and Ng, 2002) as a special case when $m = 0$, extends the single threshold factor model $(m = 1)$ recently considered by Massacci (2017) and Liu and Chen (2020) to allow for multiple unknown thresholds, and is similar to that considered by Wu (2021). This general model leads to tremendous difficulties to the identification of thresholds and subsequent inference. Alternative to the two-step sequential procedure of Wu (2021), this paper proposes a simple-to-implement procedure, which involves only principal component analysis and shrinkage estimation, to efficiently estimate all the unknown quantities.

To obtain a form convenient for further analysis, we firstly use the sorting operation to reformulate the threshold factor model as follows. Let $(z_{\pi(1)}, z_{\pi(2)}, \ldots, z_{\pi(T)})^{\top}$ be the order statistics of $(z_1, \ldots, z_T)^{\top}$, ordered from the smallest to the largest. Denote $(\boldsymbol{x}_{\pi(1)}^{\top}, \ldots, \boldsymbol{x}_{\pi(T)}^{\top})^{\top}$ as the sorted observations, where $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{Nt})^{\top}$. With these notations, the model in (2.1) can be expressed alternatively as

$$x_{i,\pi(s)} = \boldsymbol{f}_{0,\pi(s)}^{\top}\boldsymbol{\lambda}_{0,i} + \boldsymbol{f}_{-0,\pi(s)}^{\top}\boldsymbol{\lambda}_{k,i} + e_{i,\pi(s)}, \tag{2.2}$$

for $\tau_{k-1,0} \leq s < \tau_{k,0}$, where $\tau_{k,0} = \min\{s | s \in \{1, \ldots, T\}, z_{\pi(s)} > \theta_{k,0}\}$, $k = 1, \ldots, m$. Here we adopt the convention that $\tau_{0,0} = 1$ and $\tau_{m+1,0} = T + 1$. That is, the threshold factor model with multiple regimes in model (2.1) can be equivalently represented by the factor model with multiple change points in the factor loadings in model (2.2), which has been analyzed by Baltagi et al. (2021) and Ma and Tu (2022).

In matrix form, the model (2.2) can be written as follows:

$$\boldsymbol{X}_k = \boldsymbol{F}_{0k}\boldsymbol{\Lambda}_0^\top + \boldsymbol{F}_{-0k}\boldsymbol{\Lambda}_k^\top + \boldsymbol{E}_k, \quad \text{for} \quad k = 1, \ldots, m+1, \tag{2.3}$$

where $\boldsymbol{X}_k = (\boldsymbol{x}_{\pi(\tau_{k-1,0})}, \ldots, \boldsymbol{x}_{\pi(\tau_{k,0}-1)})^\top$ and $\boldsymbol{E}_k = (\boldsymbol{e}_{\pi(\tau_{k-1,0})}, \ldots, \boldsymbol{e}_{\pi(\tau_{k,0}-1)})^\top$ are both $(\tau_{k,0} - \tau_{k-1,0}) \times N$ matrix. The factor loadings $\boldsymbol{\Lambda}_0 = (\boldsymbol{\lambda}_{0,1}, \ldots, \boldsymbol{\lambda}_{0,N})^\top$ and $\boldsymbol{\Lambda}_k = (\boldsymbol{\lambda}_{k,1}, \ldots, \boldsymbol{\lambda}_{k,N})^\top$ are of dimensions $N \times (r-q)$ and $N \times q$, respectively. The subvector of the true factors $\boldsymbol{F}_k = (\boldsymbol{F}_{0k}, \boldsymbol{F}_{-0k})$ is accordingly decomposed, where $\boldsymbol{F}_{0k}$ and $\boldsymbol{F}_{-0k}$ are of dimensions $(\tau_{k,0} - \tau_{k-1,0}) \times (r-q)$ and $(\tau_{k,0} - \tau_{k-1,0}) \times q$, respectively. Note that our model allows for the emergence of new factors or disappearance of old factors, which is made possible by allowing $\boldsymbol{\Lambda}_k$, for some k, to contain some zero columns that may appear in different locations of $\boldsymbol{\Lambda}_k$ for different k. In this way, the effective number of factors r_k $(< r)$ in the k-th regime could vary with k, in contrast to Massacci (2017), Liu and Chen (2020) and Wu (2021) where the number of factors is assumed to be constant across regimes. It is further noticed that some columns of $\boldsymbol{\Lambda}_k$ are permitted to return to their previous values after two or more breaks. In this representation, with the observed data $\{x_{it}\}$, the objective is to estimate the unknown quantities, including the number of breaks m, the break dates $\tau_{k,0}$ $(k = 1, \ldots, m)$, the common factor $\boldsymbol{f}_t$ and its dimension r, and factor loadings $\boldsymbol{\Lambda}_{0k}$.

The above factor model (2.2) with multiple breaks in the factor loadings can be seen as a standard factor model with stable loading matrix, as mentioned in a few related studies such as Barigozzi et al. (2018), Baltagi et al. (2021) and Ma and Tu (2022). The underlying mechanism is that we can absorb the break information from loading matrix to the pseudo factors. To see how it works, we borrow the definition of Baltagi et al. (2021) to define $\boldsymbol{\Gamma}$ as follows. Set $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}_0$, and start from the first column of $\boldsymbol{\Lambda}_1$, to the last column of $\boldsymbol{\Lambda}_{m+1}$, i.e., traverse all the columns of $\boldsymbol{\Lambda}_k, k = 1, \ldots, m+1$, if it is nonzero and linearly independent with the columns that are already in $\boldsymbol{\Gamma}$, put it in $\boldsymbol{\Gamma}$. Let $\bar{r}$ denote the number of columns in $\boldsymbol{\Gamma}$, the maximum value of which is $(m+1)r$. By the definition of $\boldsymbol{\Gamma}$, it is easily seen that (i) $\boldsymbol{\Gamma}$ has full column rank; and (ii) for any k, we have $\boldsymbol{\Lambda}_{0k} = \boldsymbol{\Gamma}\boldsymbol{R}_k$ for some $\bar{r} \times r$ dimensional matrix $\boldsymbol{R}_k$. Let $\boldsymbol{G}_k = (\boldsymbol{g}_{\pi(\tau_{k-1,0})}, \ldots, \boldsymbol{g}_{\pi(\tau_{k,0}-1)})^\top = \boldsymbol{F}_k\boldsymbol{R}_k^\top$ denote the new (pseudo) factors. It follows that $\boldsymbol{g}_{\pi(s)} = \boldsymbol{R}_k\boldsymbol{f}_{\pi(s)}$ if $\tau_{k-1,0} \le s < \tau_{k,0}$, and (2.3) can be written as

$$\boldsymbol{X}_k = \boldsymbol{F}_k\boldsymbol{\Lambda}_{0k}^\top + \boldsymbol{E}_k = \boldsymbol{F}_k\boldsymbol{R}_k^\top\boldsymbol{\Gamma}^\top + \boldsymbol{E}_k = \boldsymbol{G}_k\boldsymbol{\Gamma}^\top + \boldsymbol{E}_k, \tag{2.4}$$

which is a standard factor model with the new stable factor loadings $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_N]^\top$ and the pseudo factors $\boldsymbol{g}_{\pi(s)}$.

We now discuss the estimation of the pseudo factors in (2.4). We first use the information criterion proposed by Bai and Ng (2002) to estimate the number of pseudo factors $\bar{r}$, the estimator for which is denoted as $\hat{r}$. Then, PCA is used to obtain the first $\hat{r}$ factors. Specifically, let $\hat{\boldsymbol{G}} = [\hat{\boldsymbol{g}}_{\pi(1)}, \ldots, \hat{\boldsymbol{g}}_{\pi(T)}]^\top$ be the estimated (pseudo) factors, which are $\sqrt{T}$ times the $\hat{r}$ eigenvectors corresponding to the $\hat{r}$ largest eigenvalues of the matrix $\boldsymbol{X}\boldsymbol{X}^\top$, with the $T \times N$ matrix $\boldsymbol{X} = [\boldsymbol{x}_{\pi(1)}, \boldsymbol{x}_{\pi(2)}, \ldots, \boldsymbol{x}_{\pi(T)}]^\top$.

To investigate the properties of the estimators, it is necessary to state the technical conditions with the above introduced notations.

Assumption 1 *(i)* $E\|\boldsymbol{f}_t\|^4 < M < \infty$*; (ii)* $E(\boldsymbol{f}_t\boldsymbol{f}_t^\top) = \boldsymbol{\Sigma}_F$ *for all* t*; (iii)* $\boldsymbol{\Sigma}_F$ *is positive definite and* $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f}_t\boldsymbol{f}_t^\top I(\theta_{k-1,0} < z_t \leq \theta_{k,0}) - \boldsymbol{\Sigma}_F = O_p(T^{-1/2})$ *for* $k = 1, \ldots, m+1$.

Assumption 2 *The factor loadings* $\boldsymbol{\Gamma}$*, the error term* $\boldsymbol{E}$ *and the factors* $\boldsymbol{G}$ *satisfy the Assumptions 2-5 and 7 in Ma and Tu (2022).*

Assumption 1 adjusts Assumption 1 in Ma and Tu (2022) to accommodate threshold effects in factor models, which holds under Assumptions C1 and CR(a) in Massacci (2017) or Assumptions 2 and 7(1) in Wu (2021). Conditions imposed in Assumption 2 are standard in the literature on factor model, which are not repeated for brevity. See Baltagi et al. (2021) and Ma and Tu (2022) for detailed discussions of these conditions. Under the above assumptions, it is easy to conclude that the errors $\boldsymbol{E}$, factors $\boldsymbol{G}$ and factor loadings $\boldsymbol{\Gamma}$ in model (2.4) satisfy Assumptions A-D of Bai and Ng (2002). Consequently, it follows immediately from an application of Theorems 1 and 2 of Bai and Ng (2002) that $\hat{r}$ is consistent for $\bar{r}$, and $\hat{\boldsymbol{g}}_{\pi(t)}$ is asymptotically close to $\boldsymbol{D}\boldsymbol{g}_{\pi(t)}$, $t = 1, \ldots, T$, where $\boldsymbol{D}$ is an $\hat{r} \times \bar{r}$ rotation matrix and $\boldsymbol{D} \xrightarrow{p} \boldsymbol{D}^* = \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{\Phi}^\top\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}}^{\frac{1}{2}}$, with $\boldsymbol{V}$ being the diagonal matrix of eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}}^{\frac{1}{2}}\boldsymbol{\Sigma}_G\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}}^{\frac{1}{2}}$, $\boldsymbol{\Phi}$ the corresponding eigenvector matrix, and $\boldsymbol{D}^*$ an $\bar{r} \times \bar{r}$ full rank matrix.

2.2 Group Lasso estimation of multiple thresholds

To estimate the break points in (2.2), Ma and Tu (2022) considered the following time varying factor regression model:

$$\hat{g}_{1,\pi(s)} = c_{2,\pi(s)}\hat{g}_{2,\pi(s)} + \cdots + c_{\hat{r},\pi(s)}\hat{g}_{\hat{r},\pi(s)} + u_{\pi(s)} = \boldsymbol{c}_{\pi(s)}^\top\hat{\boldsymbol{g}}_{-1,\pi(s)} + u_{\pi(s)}, s = 1, \ldots, T, \quad (2.5)$$

where $\hat{g}_{1,\pi(s)}$ denotes the estimated factor corresponding to the largest eigenvalue and $\hat{\boldsymbol{g}}_{-1,\pi(s)} = \left[\hat{g}_{2,\pi(s)}, \ldots, \hat{g}_{\hat{r},\pi(s)}\right]^\top$ collects the remaining $\hat{r}-1$ estimated factors, with the $(\hat{r}-1)$ dimensional coefficient $\boldsymbol{c}_{\pi(s)} = \left[c_{2,\pi(s)}, \ldots, c_{\hat{r},\pi(s)}\right]^\top$. They demonstrated that under some regularity conditions, there is a one-to-one correspondence between the structural break in factor loadings in (2.2) and that in the factor regression coefficients in (2.5). We refer readers to Ma and Tu (2022) for detailed discussions on the regularity condition. The subsequent theoretical analysis will build upon a much more general framework that allows the threshold effect (therefore the break size) to shrink towards zero as the time length T increases, which will be portrayed in Assumption 6. Alternative to (2.5), we can also use the regression $\hat{g}_{\ell,\pi(s)} = \boldsymbol{c}_{\ell,\pi(s)}^\top\hat{\boldsymbol{g}}_{-\ell,\pi(s)} + u_{\pi(s)}$, for $1 \leq \ell \leq \hat{r}$, (i.e., ℓ is not necessarily equal to 1), to perform break detection. The subsequent analysis can be adapted accordingly.

Since neither the number of breaks m nor the break dates are known, and m is typically much smaller than T, this leads us to consider the estimation of $\boldsymbol{c}_{\pi(s)}$'s and the change points via

a variant of fused Lasso (Tibshirani et al., 2005). We propose to estimate $\{\boldsymbol{c}_{\pi(s)}\}$ by minimizing the following penalized least squares objective function:

$$V_{T\lambda}(\{\boldsymbol{c}_{\pi(s)}\}) = \frac{1}{T}\sum_{s=1}^{T}(\hat{g}_{1,\pi(s)} - \boldsymbol{c}_{\pi(s)}^{\top}\hat{\boldsymbol{g}}_{-1,\pi(s)})^2 + \lambda_T \sum_{s=2}^{T}\|\boldsymbol{c}_{\pi(s)} - \boldsymbol{c}_{\pi(s-1)}\|, \tag{2.6}$$

where λ_T is a positive tuning parameter. Note that the objective function in (2.6) is convex in $\{\boldsymbol{c}_{\pi(s)}\}$. As a result, the solution can be obtained efficiently through standard convex optimization algorithms.

Note further that the penalty is imposed on vectors, therefore it is natural to connect it with the group Lasso of Yuan and Lin (2006) by reconstructing the design matrix. Let $\boldsymbol{d}_{\pi(1)} = \boldsymbol{c}_{\pi(1)}$, $\boldsymbol{d}_{\pi(s)} = \boldsymbol{c}_{\pi(s)} - \boldsymbol{c}_{\pi(s-1)}$, $s = 2, \ldots, T$. Let $\boldsymbol{c} = (\boldsymbol{c}_{\pi(1)}^{\top}, \ldots, \boldsymbol{c}_{\pi(T)}^{\top})^{\top}$, $\boldsymbol{d} = (\boldsymbol{d}_{\pi(1)}^{\top}, \ldots, \boldsymbol{d}_{\pi(T)}^{\top})^{\top}$, $\boldsymbol{Y} = (\hat{g}_{1,\pi(1)}, \hat{g}_{1,\pi(2)}, \ldots, \hat{g}_{1,\pi(T)})^{\top}$, and $\boldsymbol{u} = (u_{\pi(1)}, \ldots, u_{\pi(T)})^{\top}$. Define

$$\boldsymbol{X}^* = \begin{bmatrix} \hat{\boldsymbol{g}}_{-1,\pi(1)}^{\top} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \hat{\boldsymbol{g}}_{-1,\pi(2)}^{\top} & \hat{\boldsymbol{g}}_{-1,\pi(2)}^{\top} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \hat{\boldsymbol{g}}_{-1,\pi(3)}^{\top} & \hat{\boldsymbol{g}}_{-1,\pi(3)}^{\top} & \hat{\boldsymbol{g}}_{-1,\pi(3)}^{\top} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\boldsymbol{g}}_{-1,\pi(T)}^{\top} & \hat{\boldsymbol{g}}_{-1,\pi(T)}^{\top} & \hat{\boldsymbol{g}}_{-1,\pi(T)}^{\top} & \cdots & \hat{\boldsymbol{g}}_{-1,\pi(T)}^{\top} \end{bmatrix},$$

which is a $T \times (T(\hat{r}-1))$ design matrix. Then (2.5) can be rewritten in an alternative form as $\boldsymbol{Y} = \boldsymbol{X}^*\boldsymbol{d} + \boldsymbol{u}$, and minimizing (2.6) is equivalent to minimizing the following group Lasso criterion function

$$\begin{aligned} \bar{V}_{T\lambda}(\{\boldsymbol{d}_{\pi(s)}\}) &= \frac{1}{T}\|\boldsymbol{Y} - \boldsymbol{X}^*\boldsymbol{d}\|^2 + \lambda_T \sum_{s=2}^{T}\|\boldsymbol{d}_{\pi(s)}\| \\ &= \frac{1}{T}\sum_{s=1}^{T}\left(\hat{g}_{1,\pi(s)} - \hat{\boldsymbol{g}}_{-1,\pi(s)}^{\top}\sum_{q=1}^{s}\boldsymbol{d}_{\pi(q)}\right)^2 + \lambda_T \sum_{s=2}^{T}\|\boldsymbol{d}_{\pi(s)}\|. \end{aligned} \tag{2.7}$$

The penalty is imposed on the sum of Frobenius norm of $\boldsymbol{d}_{\pi(s)}$. For some large magnitude of λ_T, the entire group $\boldsymbol{d}_{\pi(s)}$ may drop out of the model, since this procedure encourages sparsity at both the group and individual levels by the fact that the Frobenius norm of a vector is zero only if all of its components are zero. This coincides with our starting point that the whole entry of $\boldsymbol{d}_{\pi(s)}$ is a zero vector unless there is a break at time $\pi(s)$.

At the first glance, the above loss function is similar to that (eq. (2.6)) in Ma and Tu (2022) for structural break detection in factor models. However, to perform threshold detection, the loss function here is constructed based on estimated factors that are sorted according to the value of the threshold variable z_t instead. This does not cause any additional challenge from the computational aspect, because the standard algorithm to solve the group Lasso problem can

simply apply to (2.7), illustrating the practical advantage of the proposed method. Nevertheless, the new feature in the loss function brings essential technical difficulties in analyzing the theoretical properties of the resulting estimators, as will be detailed in the following section.

Let $\{\hat{\boldsymbol{d}}_{\pi(s)}\}$ be a solution of (2.7), which would exhibit certain sparsity due to the shrinkage estimation. For convenience, define the index set of the non-zero components in $\{\hat{\boldsymbol{d}}_{\pi(s)}\}_{s=2}^{T}$, arranged in ascending order, as $\mathcal{I} = \{\hat{\tau}_1, \ldots, \hat{\tau}_{\hat{m}}\}$, where $\hat{m}$ is the cardinality of the set $\mathcal{I}$. That is to say, $\hat{\boldsymbol{d}}_{\pi(\hat{\tau}_k)} \neq \mathbf{0}$ for $k = 1, \ldots, \hat{m}$, and $\hat{\boldsymbol{d}}_{\pi(s)} = \mathbf{0}$ when $s \in \{2, \ldots, T\} \setminus \mathcal{I}$. Then, $\hat{m}$ and $\mathcal{I}$ are the estimated number of breaks and estimated set of break points, respectively. Moreover, by the relationship between $\{\boldsymbol{d}_{\pi(s)}\}$ and $\{\boldsymbol{c}_{\pi(s)}\}$, we compute $\hat{\boldsymbol{c}}_{\pi(s)} = \sum_{t=1}^{s} \hat{\boldsymbol{d}}_{\pi(t)}$ for $s = 1, \ldots, T$.

We next derive the threshold estimates from the estimated set of break points $\mathcal{I}$. From the representations in (2.1) and (2.2), it is clear that $\theta_{k,0} < z_{\pi(\tau_{k,0})} \leq \theta_{k+1,0}$, for $k = 1, \ldots, m$. Consequently, the thresholds $\theta_{k,0}$'s, $k = 1, \ldots, m$, are estimated from $z_{\pi(\hat{\tau}_k - 1)}$'s, $k = 1, \ldots, \hat{m}$. This leads to the estimate of the set of thresholds by

$$\mathcal{A}_T = \{z_{\pi(\hat{\tau}_k - 1)}, k = 1, \ldots, \hat{m}.\}. \tag{2.8}$$

3 Asymptotic properties

This section presents the asymptotic properties of the group Lasso procedure. In particular, the consistency of group Lasso estimators for the thresholds, and that of the refined two-step estimators will be investigated. The theoretical results of estimators of the factor space separated by the threshold estimates can be found in the Supplementary Material.

We emphasize that the statistical analyses of threshold factor models are not simple extensions of those in the change-point framework. As seen in the structural break representation in (2.2) and the group Lasso objective function in (2.7), the theoretical development will involve factors sorted according to the threshold variable. It is worth noting that the factors are no longer stationary mixing after sorting the observations, even if we assume that the true factors are. Therefore, the ordinary large deviation principle for the stationary mixing processes, the key tool used to analyze the properties of change point estimators by Ma and Tu (2022), can not apply in the threshold factor models. In contrast, the derivation involves large deviation probability bounds for quantities such as $\max_{x \in \mathcal{R}} \frac{1}{\sqrt{T}} \| \sum_{t=1}^{T} \boldsymbol{f}_t u_t I(z_t \leq x) \|$. Inspired by Chan et al. (2015), a bracketing entropy technique, the maximal inequality (Van der Vaart, 1998, Lemma 19.34), the Bernstein and Cauchy-Schwarz inequalities are applied instead to establish the large deviation bound (the details are contained in the proofs of Lemmas A.3 and A.4 in the Supplementary Material). This brings important technical differences and additional challenges as compared to the theoretical study of Ma and Tu (2022). A second type of technical challenge, compared to the development in Chan et al. (2015), is to eliminate the effect of factor estimation error in the derivations, because the group Lasso is implemented based on estimated factors instead of observed data. The technical conditions imposed below are shaped accordingly for

these new theoretical developments.

3.1 Consistency of group Lasso estimators

To study the consistency of group Lasso estimators, we first impose the following assumptions.

Assumption 3

(a) $E\left(\|\boldsymbol{f}_t\|^{2+\ell}|z_t=\theta\right)<\infty$ *for all* θ *and some* $\ell>2$*;*

(b) $\{\boldsymbol{f}_t, z_t\}$ *is a stationary and* α*-mixing process with a geometric decaying rate;*

(c) $\{\boldsymbol{f}_t\}$ *and* $\{z_t\}$ *have bounded, positive and continuous densities.*

Assumption 4 *There exists an* $M<\infty$ *such that:*

(a) $E\left(\sup_{\theta_{0,0}\leq\theta_1<\theta_2\leq\theta_{m+1,0}} \frac{1}{L}\sum_{t=1}^{T}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it}-E(e_{is}e_{it})]\right| I(\theta_1<z_t\leq\theta_2)\right)\leq M$ *for all* s*;*

(b) $E\left(\sup_{\theta_{0,0}\leq\theta_1<\theta_2\leq\theta_{m+1,0}} \frac{1}{L}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\gamma}_i e_{it} I(\theta_1<z_t\leq\theta_2)\right\|\right)\leq M,$

where $L=\sum_{t=1}^{T} I(\theta_1<z_t\leq\theta_2)$.

Assumption 5 *Define* $\boldsymbol{\epsilon}_t=vech(\boldsymbol{f}_t\boldsymbol{f}_t^\top-\boldsymbol{\Sigma}_F)$.

(a) The Hajek-Renyi inequality can be applied to the processes $\{\boldsymbol{\epsilon}_t I(\theta_{k-1,0}<z_t\leq\theta_{k,0})\}$ *for each* $k=1,\ldots,m+1$*;*

(b) There exist $\delta>0$ *and* $M<\infty$ *such that* $E\left(\sup_{\theta_{k-1,0}\leq\theta_1<\theta_2\leq\theta_{k,0}}\|\frac{1}{\sqrt{L}}\sum_{t=1}^{T}\boldsymbol{\epsilon}_t I(\theta_1<z_t\leq\theta_2)\|^{4+\delta}\right)<M$*, for* $k=1,\ldots,m+1$*, where* $L=\sum_{t=1}^{T} I(\theta_1<z_t\leq\theta_2)$.

Assumption 3 is a regularity condition in the literature for standard asymptotic properties, which restricts the memory of the sequence $\{\boldsymbol{f}_t, z_t\}$, excludes trends and integrated processes (Li and Ling, 2012; Chan et al., 2015), and imposes the existence of conditional moment of factors given the threshold variable. This assumption is less restrictive than those imposed by Massacci (2017) and Wu (2021), which require instead that $\{\boldsymbol{f}_t, z_t\}$ be strictly stationary, ergodic and ρ-mixing. Assumption 4 imposes further conditions on the errors in the factor model, which are natural extensions of Assumption 9 in Ma and Tu (2022) to incorporate threshold effect. To provide some intuition for Assumption 4, let $\xi_{s,t}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[e_{is}e_{it}-E(e_{is}e_{it})]$. Note that $E(\sup_{\theta_{0,0}\leq\theta_1<\theta_2\leq\theta_{m+1,0}} 1/L\sum_{t=1}^{T}|\xi_{s,t}|I(\theta_1<z_t\leq\theta_2))\leq E(\sup_t|\xi_{s,t}|)$, then $E(\sup_t|\xi_{s,t}|)\leq M$, for all s, is a sufficient condition for Assumption 4(a). This rules out the case where the expectation of $\sup_t|\xi_{s,t}|$ does not exist, and imposes restrictions on the tail probability of $\xi_{s,t}$.

This condition is needed to facilitate the derivation of the uniform bound of the differences between the estimated factors and the rotation of the pseudo factors. The fact that $\xi_{s,t}$ is a demeaned sample average of $e_{is}e_{it}$ shows that, when $\xi_{s,t}$ is asymptotically normal under a central limit theorem, the assumption is not as strong as it might appear. Assumption 5, similar to Assumption 7 of Baltagi et al. (2021), imposes a further requirement on the factor process. It is emphasized that the estimated pseudo factors used in the factor regression contain estimation errors, the treatment of which is crucial in deriving the asymptotic properties of the threshold estimates. Here we require the Hajek-Renyi inequality be applicable to the second moment of the factors, which facilitates to pin down the order of the estimation error in the estimated thresholds.

Let $\boldsymbol{\beta}_k$ be the factor regression coefficient of (2.5) in the k-th regime. The next assumption portrays the size of the threshold effects allowed in the factor model in terms of the changes from $\boldsymbol{\beta}_k$ to $\boldsymbol{\beta}_{k+1}$. To proceed, let $J_{\min} = \min_{1\le k\le m} \|\boldsymbol{D}_{-1}^{*}\boldsymbol{R}_k\boldsymbol{R}_k^{\top}\boldsymbol{D}_{-1}^{*\top}(\boldsymbol{\beta}_{k+1}-\boldsymbol{\beta}_k)\|$, where $\boldsymbol{D}^{*\top} = (\boldsymbol{D}_1^{*}, \boldsymbol{D}_{-1}^{*\top})$ is the probability limit of the rotation matrix $\boldsymbol{D} = (\boldsymbol{D}_1, \boldsymbol{D}_{-1}^{\top})$ and $\boldsymbol{D}_1$ is an $\bar{r}-1$ vector and $\boldsymbol{D}_{-1}^{\top}$ is an $(\hat{r}-1)\times\bar{r}$ matrix. Furthermore, define $I_{\min} = \min_{2\le k\le m}|\theta_{k,0}-\theta_{k-1,0}|$.

Assumption 6

(a). It holds that $T^{1/2-w}J_{\min}\to\infty$ *for* $0\le w<1/2$, *and* $\gamma_T J_{\min}/\lambda_T\to\infty$.

(b). There exist constants l, u *such that* $\theta_{k,0}\in[l,u], 1\le k\le m$ *and* $I_{\min}/\gamma_T\to\infty$ *for some* $\gamma_T\to 0$, *where* $\gamma_T^{-1}(T\gamma_T)^{-\ell/2}(\log T)^{2(2+\ell)}\to 0$ *and* $\gamma_T/\lambda_T\to\infty$ *as* $T\to\infty$.

Assumption 6(a) characterizes the specific conditions on $\gamma_T, J_{\min}$ and λ_T. In particular, the minimum size of structural break $J_{\min}$ is allowed to diminish toward zero at an order no smaller than $\max\{\lambda_T/\gamma_T, T^{-1/2}\}$, the magnitude of which is roughly equivalent to that of the threshold effect allowed in the factor loadings. This assumption describes the types of threshold effects that are allowed and rules out cases where two adjacent factor regression coefficients are the same, for example, there are two factors and one threshold, and the factors exchange their loadings in the two regimes. This special case rarely occurs in reality and is also excluded by Liu and Chen (2020) and Wu (2021) via the condition that the column spaces of two adjacent loading matrices should be different. However, their condition is more stringent than our Assumption 6(a). To see this, Example 2.1 in Ma and Tu (2022) offers a case where the column spaces spanned by the adjacent loading matrices are the same but the break detection method adopted in this paper remains effective. Further, note that Massacci (2017, Assumption I) and Wu (2021, Assumption 1) portrayed another form of weak threshold effects, i.e., the changes in the loadings are fixed but only $O(N^{\alpha_k})$ variables of x_{it}'s undergo the changes, for $0.5<\alpha_k\le 1$ and $k=1,\ldots,m$. To ensure identification of the thresholds, they required that the order of the largest eigenvalue of $\Delta_k\Delta_k^{\top}/(NT)$, $O_p(N^{\alpha_k-1})$, should be larger than that of the error second moment δ_{NT}^{-2}, where $\Delta_k = (\boldsymbol{\Lambda}_{0k}-\boldsymbol{\Lambda}_{0k-1})\boldsymbol{F}_k$ and $\delta_{NT} = \min\{\sqrt{N},\sqrt{T}\}$. This implies that

$N^{1-\alpha_k} = o(T)$ for $k = 1, \ldots, m$. To entertain the small magnitude of changes in the loadings, this condition ought to be changed to $N^{1-\alpha_k} J_{\min}^{-2} = o(\delta_{NT}^2)$ accordingly. The detailed investigation along this direction can be carried out in a way similar to the analysis below, but is beyond the scope of the current study due to the extra complexity involved. This leaves us to focus on the case where $\alpha_k = 1$ subsequently for the sake of exposition. Assumption 6(b) is similar to Chan et al. (2015) and requires that the distance between two consecutive thresholds should be much larger than a diminishing sequence γ_T, which in turn should be larger than the regularization parameter λ_T used in the group Lasso. This relaxes the condition $\min_{2 \le k \le m} |\theta_{k,0} - \theta_{k-1,0}| > \delta$ for some $\delta > 0$ considered in most of the studies on threshold regression.

The following theorem presents the consistency for the estimators of the thresholds, when the estimated number of thresholds equals to m.

Theorem 3.1 *Let $\hat{\theta}_k = z_{\pi(\hat{\tau}_k - 1)}$ be the k-th estimated threshold given in (2.8). Suppose that Assumptions 1-6 hold. Then, if $|\mathcal{A}_T| = m$, we have*

$$\Pr\left(\max_{1 \le k \le m} |\hat{\theta}_k - \theta_{k,0}| \le \gamma_T \right) \to 1, \quad as \quad N, T \to \infty. \tag{3.1}$$

Theorem 3.1 requires that the cardinality of $\mathcal{A}_T$ be equal to the true number m, which does not account for the uncertainty in the estimation of the number of thresholds. Therefore, this result is not practically meaningful, but serves as a key intermediate step toward results to be presented below. Theorem 3.2 shows that, the true number of thresholds m will tend to be overestimated. However, for every true threshold in $\mathcal{A}$, there exists one estimate in $\mathcal{A}_T$ that can identify it within a γ_T neighborhood.

Theorem 3.2 *Under Assumptions 1-6, as $N, T \to \infty$, we have*

$$\Pr(|\mathcal{A}_T| \ge m) \to 1, \tag{3.2}$$

and

$$\Pr\left(\mathcal{D}(\mathcal{A}_T, \mathcal{A}) \le \gamma_T\right) \to 1. \tag{3.3}$$

Remark 3.1 *(i) Theorem 3.1 only requires that N and T go to infinity jointly, without imposing any condition on the N-T ratio. This is similar to that found in Baltagi et al. (2021) and Ma and Tu (2022) under the setting of factor model with multiple structural breaks, but different from the requirement in factor-augmented forecast and vector autoregression, such as Bai and Ng (2006) where $\sqrt{T}/N \to 0$ is needed to eliminate the estimation effect of the factors. This is due to the uniqueness of the threshold estimation, which relies on observations lying only in a local region of the true threshold (i.e., change points in (2.2)) and consequently the estimation error brought by the estimated factors will not accumulate as $T \to \infty$. Therefore, the estimation error asymptotically has no effect on the estimated thresholds. Consequently, no N-T ratio condition*

is needed. In contrast, the estimation in factor-augmented forecasting and vector autoregression depends on all observations and consequently the estimation error will accumulate as $T \to \infty$.

(ii) The overestimation of the number of thresholds is related to the violation of the restricted eigenvalue conditions (e.g., Bickel et al. (2009)) in the group Lasso problem defined in (2.7). These conditions require that any subset of the explanatory variables of a given size should form a positive definite design matrix, and they are essential for establishing the consistency properties of Lasso. However, any two adjacent columns of the design matrix $\boldsymbol{X}^*$ *only differ by one entry, which fails such a requirement. Consequently, the consistency argument used for ordinary regression problem does not apply here. Instead, the theorem is proved by contradiction. It is shown that for any true threshold value, there exists one estimator that can identify it within a* γ_T *neighborhood. Otherwise, the Karush-Kuhn-Tucker condition can not be satisfied.*

Remark 3.2 *(i) The quantity* γ_T *can be regarded as the convergence rate of the threshold estimator* $\hat{\theta}_k$. *The optimal rate of convergence for the threshold estimator in regression setting is* $O(1/T)$ *when the number of thresholds* m *is known and the true thresholds* $\theta_{k,0}$*'s are fixed constants (Li and Ling, 2012). In our case, from Assumption 6,* γ_T *can be taken as* $O(\log^{4+c} T/T^{\frac{\ell}{2+\ell}})$ *for some constant* $c > 0$. *When* ℓ *is arbitrarily large, the order of* γ_T *can be made close to* $O(1/T)$ *(except for a factor of* $\log T$*). Accordingly, the tuning parameter choice* $\lambda_T = \log T/T$ *ensures that Assumption 6 holds, and is found proper in the simulation studies. Alternatively,* λ_T *can also be chosen by an information criterion as in Qian and Su (2016) and Ma and Tu (2022).*

(ii) For the single threshold factor model, under the presumption that the number of true factors is known, Massacci (2017) showed that the threshold parameter estimator is consistent at a rate $O_p(1/(N^{\alpha^0}T))$, *where* $0.5 < \alpha^0 \leq 1$, *with a requirement that at least a fraction of* $O(N^{\alpha^0})$ *of the* N *cross-sectional units experience a regime shift in the loadings. This convergence rate is much faster than our convergence rate* γ_T. *The faster convergence rate, especially the* $1/N^{\alpha^0}$ *multiplier, is a direct consequence of applying the concentrated least squares and PCA, which makes full use of all the information across the cross-sectional units. In contrast, our detection of the thresholds relies on the sparsity in the factor regression coefficients over time, to build which the cross-sectional unit information has been first aggregated to obtain consistent (pseudo) factor estimators (up to rotation). For the multiple threshold factor model, the concentrated least squares may be also married with PCA to form a joint estimation of the thresholds, similar to the extensions of Gonzalo and Pitarakis (2002) and Li and Ling (2012) to tackle multiple threshold regressions. However, the computational burden could become overwhelming when the number of thresholds is more than two, especially with a long time series. It is in this sense that our proposed procedure becomes practically advantageous. A more efficient procedure that combines our estimator with that of Massacci (2017) will be discussed in Remark 3.3 below.*

(iii) Liu and Chen (2020) approached the estimation of the single threshold factor model by using a sum of squared norm of the projection cross moment matrices. The convergence rate

they obtained for the single threshold estimator is at most $T^{-1/2}$ (i.e., in the strong regimes case), which is slower than the fastest possible rate that can be achieved for our estimator.

(iv) For multiple threshold factor model, Wu (2021) showed that the convergence rate of the k-th threshold can achieve $N^{\alpha_k^0}T$ with α_k^0 defined similar to α^0 in (ii). The faster consistency is a consequence of the procedure using the concentrated least squares proposed by Massacci (2017), as demonstrated in (ii). However, the division of the data into subintervals significantly reduces the available sample size for estimation of each threshold, which undermines the estimation accuracy in finite samples as confirmed in the simulation results reported later on.

3.2 Two-step estimation procedure

From Theorem 3.2, we have that with probability approaching 1, all the true thresholds can be identified by $\mathcal{A}_T$ within a γ_T neighborhood. This leads to a set of threshold candidates, whose size is larger than that of the true set $\mathcal{A}$. Therefore, one practical problem is how to better estimate m, or equivalently, how to choose the best subset of threshold estimates from $\mathcal{A}_T$.

To this aim, we construct an information criterion based on the post-Lasso estimators to select the optimal subset from $\mathcal{A}_T$. Given $\hat{m}$ and the threshold set $\mathcal{A}_T = (\hat{\theta}_1, \ldots, \hat{\theta}_{\hat{m}})$, let $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1^\top, \ldots, \tilde{\boldsymbol{\beta}}_{\hat{m}+1}^\top)^\top$ be the post-Lasso least squares estimates and $S_T(\hat{\theta}_{k-1}, \hat{\theta}_k) = \sum_{t=1}^{T}(\hat{g}_{1,t} - \hat{\boldsymbol{g}}_{-1,t}^\top \tilde{\boldsymbol{\beta}}_k)^2 I(\hat{\theta}_{k-1} < z_t \leq \hat{\theta}_k)$ be the residual sum of squares for the k-th regime. Consider a general information criterion of the form

$$IC(\hat{m}, \mathcal{A}_T) = S_T(\hat{\theta}_1, \ldots, \hat{\theta}_{\hat{m}}) + \hat{m}\omega_T, \tag{3.4}$$

where the least squares criterion $S_T(\hat{\theta}_1, \ldots, \hat{\theta}_{\hat{m}}) = \sum_{k=1}^{\hat{m}+1} S_T(\hat{\theta}_{k-1}, \hat{\theta}_k)$ is the goodness-of-fit measure and ω_T is the penalty term. We estimate the number and locations of the thresholds by solving

$$(\tilde{m}, \tilde{\mathcal{A}}_T) = \arg \min_{\substack{n \in (0,1,\ldots,|\mathcal{A}_T|) \\ \boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \subset \mathcal{A}_T}} IC(n, \boldsymbol{\theta}). \tag{3.5}$$

The following theorem shows that $\tilde{m}$ and $\tilde{\mathcal{A}}_T$ in (3.5) are consistent for m and the true threshold values, respectively.

Theorem 3.3 *If ω_T satisfies that $\lim_{T\to\infty} \omega_T/(T \min_{1\leq k\leq m} |\theta_{k,0} - \theta_{k-1,0}|) = 0$ and $\lim_{T\to\infty} \omega_T/(mT\gamma_T) \geq C$ for some constant $C > 3$, then under Assumptions 1-6, as $N, T \to \infty$, we have*

$$\Pr(\tilde{m} = m) \to 1, \tag{3.6}$$

and there exists a constant $B > 0$ such that

$$\Pr\left(\mathcal{D}(\tilde{\mathcal{A}}_T, \mathcal{A}) \leq B\gamma_T\right) \to 1. \tag{3.7}$$

The optimal penalty term ω_T in Theorem 3.3 is difficult to choose since the constant C is hard to determine in practice. Actually, it suffices to choose an ω_T such that $\omega_T/(T\gamma_T) \to \infty$ as $T \to \infty$. Since γ_T can be nearly optimal (γ_T is of the order $O(1/T)$) when ℓ in Assumptions 3(a) and 6 can be arbitrarily large, $T\gamma_T$ is nearly $O(1)$, as noted in Remark 3.2(i). Therefore, the usual consistent model selection criteria such as BIC ($\omega_T = \log T$) can be applied (Chan et al., 2015). The other choice of $\omega_T = \sqrt{T}$ also satisfies all the required conditions even when γ_T is not optimal.

The practical implementation of the above method involves the evaluation of all possible subsets of $\mathcal{A}_T$ to yield the threshold estimates. This becomes very computationally demanding, especially when $|\mathcal{A}_T|$ is large. To expedite the computation, we follow Chan et al. (2015) to use the well known backward elimination algorithm (BEA) to help us choose the best subset. We start with the whole set of thresholds $\mathcal{A}_T$ obtained from the previous step, then remove the most redundant threshold which corresponds to the largest reduction of IC defined in (3.4). Then we repeat the above process until the IC can not be reduced any more. More specifically, the BEA goes as follows.

Backward Elimination Algorithm (BEA):

Step (i). Set $K = |\mathcal{A}_T|$, $\boldsymbol{\theta}_K := (\theta_{K,1}, \ldots, \theta_{K,K}) = \mathcal{A}_T$ and $V_K^\star = IC(K, \mathcal{A}_T)$.

Step (ii). For $k = 1, \ldots, K$, compute $V_{K,k} = IC(K-1, \boldsymbol{\theta}_K \setminus \{\theta_{K,k}\})$. Set $V_{K-1}^\star = \min_k V_{K,k}$.

Step (iii).
- If $V_{K-1}^\star > V_K^\star$, then the estimated set of thresholds is $\mathcal{A}_T^\star = \boldsymbol{\theta}_K$.
- If $V_{K-1}^\star \le V_K^\star$ and $K = 1$, then $\mathcal{A}_T^\star = \emptyset$. There is no threshold in the factor model.
- If $V_{K-1}^\star \le V_K^\star$ and $K > 1$, then set $j = \arg\min_k V_{K,k}$, $\boldsymbol{\theta}_{K-1} = \boldsymbol{\theta}_K \setminus \{\theta_{K,j}\}$ and $K = K - 1$. Go to step (ii).

For the estimate $\mathcal{A}_T^\star =: (\hat{\theta}_1^\star, \ldots, \hat{\theta}_{|\mathcal{A}_T^\star|}^\star)$ obtained from BEA, we have the same consistency results as those in Theorem 3.3.

Theorem 3.4 *Under the same conditions of Theorem 3.3, as $N, T \to \infty$,*

$$\Pr(|\mathcal{A}_T^\star| = m) \to 1, \tag{3.8}$$

and there exists a constant $B > 0$ such that

$$\Pr\left(\mathcal{D}(\mathcal{A}_T^\star, \mathcal{A}) \le B\gamma_T\right) \to 1. \tag{3.9}$$

To sum up, the one-step group Lasso estimate $\mathcal{A}_T$ contains a subset $\mathcal{A}_T^\star$ that can consistently estimate the number and values of true thresholds. Theorem 3.3 and Theorem 3.4 ensure that the two-step procedure estimates the true number of thresholds m with probability approaching 1, and estimates the true thresholds consistently.

Remark 3.3 *With consistent estimation of the threshold number ($|\mathcal{A}_T^\star|$) and values ($\hat{\theta}_k^\star$'s) via the two-step procedure, it is possible to obtain super-consistent threshold estimators as in Massacci (2017) through a further re-estimation step. This is to re-estimate the regime specific loadings and factors and threshold values sequentially by using the principal components and concentrated least squares. Specifically, to estimate $\theta_{k,0}$, we collect the sample at all time t such that $\hat{\theta}_{k-1}^\star + \epsilon < z_t < \hat{\theta}_{k+1}^\star - \epsilon$, for a small positive number ϵ. For such a selected sample, there is only one threshold in the factor representation, so that the method of Massacci (2017) could be applied to obtain an estimator for $\theta_{k,0}$. The sequential procedure continues until all the $|\mathcal{A}_T^\star|$ thresholds are estimated. It can be shown that such estimators are converging at the same rate as that of Massacci (2017), if the number of thresholds is fixed and does not grow with the sample size. This further re-estimation is also more convenient than that proposed by Wu (2021), since we only need to search possible threshold values around our consistent two-step estimators ($\hat{\theta}_k^*$), which relieves significantly the computational burden. The detailed investigation is beyond the scope of this study. We emphasize that the above re-estimation significantly downgrades the computational challenge, in comparison to the joint search of all the thresholds from the start using the concentrated least squares as mentioned in Remark 3.2(ii).*

4 Simulation

In this section, a set of Monte Carlo experiments are conducted to evaluate the finite sample performance of our proposed method (MT hereafter), with a comparison to that of Wu (2021) (Wu hereafter). The data generating processes contain factor models with no threshold, one threshold, and multiple thresholds.

4.1 Data generating processes

DGP0 (No threshold) When there is no threshold effect, we consider the linear factor model

$$x_{it}^s = \lambda_{11i} f_{1t}^s + \lambda_{12i} f_{2t}^s + e_{it}^s,$$

for $i = 1, \ldots, N,\ t = 1, \ldots, T$, where $s = 1, \ldots, S$ refers to the replication and S is the total number of replications. The factor loadings are generated as $\lambda_{11i} \sim N(1,1)$ and $\lambda_{12i} \sim N(1,1)$, which are fixed in the repeated replications.

DGP1 (One threshold) For the factor model with one threshold, we consider:

$$\begin{aligned} x_{it}^s &= I(z_t^s \le \theta_{1,0})(\lambda_{11i} f_{1t}^s + \lambda_{12i} f_{2t}^s) \\ &\quad + I(z_t^s > \theta_{1,0})(\lambda_{21i} f_{1t}^s + \lambda_{22i} f_{2t}^s) + e_{it}^s, \end{aligned}$$

for $i = 1, \ldots, N,\ t = 1, \ldots, T$, where $s = 1, \ldots, S$ refers to the replication. We fix the factor loadings $\lambda_{11i}, \lambda_{12i}$ and $\lambda_{21i}, \lambda_{22i}$ and the threshold parameter $\theta_{1,0}$ throughout the replications,

with $\lambda_{11i} \sim N(1,1)$, $\lambda_{12i} \sim N(1,1)$, $\lambda_{21i} \sim N(1+\delta,1)$ and $\lambda_{22i} \sim N(1+\delta,1)$ for $i = 1,\ldots,N$, and set $\theta_{1,0} = 2$. We generate z_t^s according to

$$z_t^s = \mu_z(1-\rho_z) + \rho_z z_{t-1}^s + (1-\rho_z^2)^{1/2}\epsilon_{zt}^s, \quad z_{-50}^s = \mu_z,$$

for $t = -49,\ldots,0,\ldots,T$, where μ_z and $\rho_z \sim U(0.05, 0.95)$ are fixed in repeated samples, and $\epsilon_{zt}^s \sim$ i.i.d. $N(0,1)$. In this way, $E(z_t^s) = \mu_z$ and $Var(z_t^s) = 1$. We let $\pi^0 = P(z_t^s \leq \theta_{1,0}) = P(z_t^s - \mu_z \leq \theta_{1,0} - \mu_z) = \Phi(\theta_{1,0} - \mu_z) = 0.50$ and obtain $\mu_z = \theta_{1,0} - \Phi^{-1}(\pi^0) = 2$. The choice $\pi^0 = 0.50$ is consistent with that of Massacci (2017).

DGP2-5 (Multiple thresholds) We next consider a factor model with three-regime factor loadings separated by 2 thresholds (DGP2):

$$\begin{aligned} x_{it}^s &= I(z_t^s \leq \theta_{1,0})(\lambda_{11i} f_{1t}^s + \lambda_{12i} f_{2t}^s) \\ &\quad + I(\theta_{1,0} < z_t^s \leq \theta_{2,0})(\lambda_{21i} f_{1t}^s + \lambda_{22i} f_{2t}^s) \\ &\quad + I(z_t^s > \theta_{2,0})(\lambda_{31i} f_{1t}^s + \lambda_{32i} f_{2t}^s) + e_{it}^s, \end{aligned} \tag{4.1}$$

with $\lambda_{11i} \sim N(1,1), \lambda_{12i} \sim N(1,1)$, $\lambda_{21i} \sim N(1+\delta,1), \lambda_{22i} \sim N(1+\delta,1)$ and $\lambda_{31i} \sim N(1+2\delta,1), \lambda_{32i} \sim N(1+2\delta,1)$ for $i = 1,\ldots,N$. We set $\theta_{1,0} = 1.57$ and $\theta_{2,0} = 2.43$, which are $1/3, 2/3$ quantiles of $N(2,1)$, respectively. The factor loadings and the threshold values are fixed in the repeated samples.

We also consider the cases where there are $m = 3,4,5$ thresholds. For DGPm, the whole time span is separated into $m+1$ regimes based on the quantile statistics of the observations $\{z_1,\ldots,z_T\}$. That is, the $k/(m+1)$-th quantile of $\{z_1,\ldots,z_T\}$ is set as the true value of the k-th threshold, for $k = 1,\ldots,m$. Finally, the factor loadings are generated similarly to DGP2. More specifically, λ_{k1i} and λ_{k2i} are i.i.d. $N(1+(k-1)\times\delta,1)$ in the k-th regime, where $k = 1,\ldots,m+1, i = 1,\ldots,N$, for DGP$m$, with $m = 3,4,5$.

Our setup incorporates the scenario that allows the number of factors to change over time, which is important because new (old) factor(s) can arise (disappear) in empirical applications. For example, a new factor may appear to capture financial co-movement emerges at the beginning of the Great Recession (Cheng et al., 2016). To study the finite sample performance of our proposed method when the number of factors is varying across regimes, we consider the following DGP.

DGP2.1 **(Varying factor numbers)** We simulate the data separated by 2 thresholds as model (4.1) with $\lambda_{11i} \sim N(1,1), \lambda_{12i} \sim N(1,1)$, $\lambda_{21i} \sim N(1+\delta,1), \lambda_{22i} = 0$ and $\lambda_{31i} \sim N(1+2\delta,1), \lambda_{32i} \sim N(1+2\delta,1)$ for $i = 1,\ldots,N$. We set $\theta_{1,0}$ and $\theta_{2,0}$ as the 60% and 70% quantile statistics of $\{z_1,\ldots,z_T\}$ to examine the efficacy of both methods when the distance between the two thresholds is relatively small. By setting $\lambda_{22i} = 0$ for $i = 1,2,\ldots,N$, the second factor that appears in the first regime disappears in the second regime and emerges again in the last regime. Thus the numbers of factors in the three regimes are 2, 1, 2, respectively, which results in the number of pseudo factors to be 5.

For all the above data generating processes, we set $N = 100, 250$, $T = 100, 250, 500$ for DGP1-2 and $T = 250, 500$ for DGP3-5, and the number of replications as $S = 1,000$. We generate the factor f_{1t}^s as $f_{1t}^s = \rho_{f1} f_{1,t-1}^s + u_{1t}$, u_{1t}'s are i.i.d. $N(0, 1-\rho_{f1}^2)$, and $f_{2t}^s = \rho_{f2} f_{2,t-1}^s + u_{2t}$, u_{2t}'s are i.i.d. $N(0, 1-\rho_{f2}^2)$, where $\rho_{f1} \sim U(0.05, 0.55)$ and $\rho_{f2} \sim U(0.55, 0.95)$. Note that ρ_{f1} and ρ_{f2} are generated once and fixed in all the replications. The error terms e_{it}^s's are generated in four cases. (C1) independent and identically errors: e_{it}^s's are i.i.d. $N(0, 2)$, (C2) cross-sectionally heterogeneous errors: $e_{it}^s = \sigma_i^s v_{it}^s$, where σ_i^s's are i.i.d. $U(0.5, 1.5)$, v_{it}^s's are i.i.d. $N(0, 2)$, (C3) serially correlated errors: $e_{it}^s = 0.2 e_{i,t-1}^s + u_{it}^s$, where u_{it}^s's are i.i.d. $N(0, 2(1-0.2^2))$, and (C4) cross-sectionally dependent errors: $e_{it} = \eta_{it}/\sqrt{1+2d\beta^2}$, where $\eta_{it} = (1-\beta)v_{it} + \sum_{l=\max\{i-d,1\}}^{\min\{i+d,N\}} \beta v_{lt}$, $d = \max\{10, N/20\}$ and $\beta = 0.1$. The threshold variable z_t^s in DGP2-5 and DGP2.1 is generated the same way as that in DGP1. Further, we control the magnitude of the threshold effect by setting $\delta = 1, 2$.

DGP2.2 **(Small threshold effects)** We finally consider a factor model with shrinking threshold effects. To this end, we generate the data according to DGP2, except that we set $\delta = T^{-3/7}$.

4.2 Results

For DGP0 with no threshold, the false detection proportions (%) for thresholds and the average running time (in seconds) for both MT and Wu's methods are shown in Table 1. We only report the results for $N = 100$ here and those for $N = 250$ are shown in the Supplementary Material for space consideration. For the model complexity term ω_T, we experiment with $\log T$ (BIC), $2\log T$ (2BIC) and $3\log T$ (3BIC) in the simulations for MT method. There are several interesting findings which are summarized as follows. First, it is observed that for different error settings, the false detection proportions of our method (MT) are quite small and would decrease as the time length T increases. However, Wu only works for the first three error cases, the FDPs of which in the cross-sectional dependent error case (C4) can reach 100%, due to the poor estimation of the number of factors in each subinterval. Furthermore, the FDPs of Wu are also quite large for small sample size ($T = 100$). Second, MT is less time demanding, especially for large sample size, which demonstrates the computational advantages of our method in the presence of large data sets.

For data generating processes with one or more thresholds, we use the percentages of correct estimation (PCE) for the number of thresholds and the accuracy of threshold estimation to evaluate the performance of finite sample properties (conditional on the correct estimation of m). More specifically, let $\hat{m}^s$ denote the estimated number of thresholds in the s-th replication. The PCE of $\hat{m}$ is defined as $PCE = \sum_{s=1}^S 1\{\hat{m}^s = m\}/S$, where the indicator function $1\{\hat{m}^s = m\} = 1$ if $\hat{m}^s = m$, and 0 otherwise. Conditional on the correct estimation of m, the accuracy of thresholds estimation will be measured by the Hausdorff distance (HD) between the set of the estimated thresholds and that of the true thresholds, which is then averaged over all the replications. Recall that the Hausdorff distance between any two sets A and B is defined as

$HD = \max\{\mathcal{D}(A,B), \mathcal{D}(B,A)\}$, where $\mathcal{D}(A,B) \equiv \sup_{b\in B} \inf_{a\in A} |a-b|$. The estimation accuracy is also measured by the bias and root mean squared error (RMSE) of each threshold for each case:

$$\text{Bias}_k = S^{-1}\sum_{s=1}^{S}(\hat{\theta}_k^s - \theta_{k,0}), \quad \text{RMSE}_k = \sqrt{S^{-1}\sum_{s=1}^{S}(\hat{\theta}_k^s - \theta_{k,0})^2}, \text{ for } k = 1, \ldots, m.$$

To compare the computational efficiency between the two methods, we calculate the average running time for each replication, i.e., $S^{-1}\sum_{s=1}^{S} time_s$, where $time_s$ denotes the running time (in seconds) in the s-th replication. We set the number of subintervals $J+1 = 10$ for the method of Wu (2021), so that the sample size in each subinterval is about $10\%T$. Due to space limit, the subsequent experiments only report the results with BIC, i.e., $\omega_T = \log T$, while the detailed results for other choices of ω_T, including $2\log T, 3\log T, \sqrt{T}$, are quite similar for large sample sizes and relegated to the Supplementary Material.

We report the estimation accuracy, as measured by the PCE, HD, bias and RMSE in Table 2 for DGP3 with $\delta = 1$ and $N = 100$. The results for DGP1, 2 and 4, and other settings for DGP3 are similar and relegated to the online Supplementary Material to save space. Several interesting findings are as follows. First, it is observed that the PCEs of MT are significantly larger than the corresponding ones of Wu, which means MT can detect the number of unknown multiple thresholds more accurately. Furthermore, compared with the results for DGP1 in Table D.2 in the Supplementary Material, the performance of MT is not much different, while for Wu, the PCEs for DGP3 are apparently smaller, which demonstrates that our method is more robust for detecting multiple thresholds in the factor models. Second, the RMSEs and HDs of MT are apparently smaller than those of Wu, which indicates that in terms of estimation accuracy, our estimators dominate for almost all cases in finite samples. Third, the results of MT are robust to the error dependences, while the cross-sectional dependence in the idiosyncratic error, i.e., case $(C4)$, greatly impairs the results of Wu. Specifically, the PCEs of Wu are quite small for C4 and the method almost fails to detect the correct number of thresholds even for sample size as large as $T = 500$. This finding is consistent with that reported in Wu (2021), which is perhaps due to the fact that the unsatisfactory performance of IC_{p1} of Bai and Ng (2002) under cross-sectional error dependence jeopardises the performance of Wu. Lastly, Wu is much more computationally time-consuming than MT, especially when the sample sizes N, T and the number of thresholds are large, as also confirmed by the results in Tables D.2-D.5 in the Supplementary Material.

We further present the estimation results (PCE, HD, RMSE and running time) in Figure 4.1 for DGP5 with five thresholds when $\delta = 1$. The results for $\delta = 2$ can be found in the Supplementary Material. In Figure 4.1, RMSE denotes the aggregated root mean squared errors of all the five threshold estimators. It is seen that Wu fails to detect the correct number of thresholds, while our PCEs can be very close to 100%. The estimation accuracy of our estimators is almost unaffected by the the number of thresholds. In addition, the HDs and

RMSEs of our estimated thresholds are smaller than the corresponding ones of Wu, conditional on that the threshold number is correctly detected. Furthermore, the running time of MT is also much less than that of Wu for each set of parameter specification, which highlights the computational efficiency of our proposed method.

We next turn to the case where the number of factors is changing in adjacent regimes, i.e., DGP2.1, the results for which are shown in Table 3 for both MT and Wu, with $\delta = 1, N = 250$. The results for other parameter specifications are similar and can be found in the Supplementary Material. We summarize some new findings here. Although the PCEs of MT are small for $T = 100$, they increase quickly to 95% as T increases. However, Wu can not detect the correct threshold number even when $T = 500$, and the RMSEs and HDs of Wu are extremely large. Therefore, the approach of Wu is not effective in the presence of varying number of factors and closely distributed thresholds. Finally, the results for DGP2.2 with shrinking threshold effects are presented in Table 4. The main findings are not much different from those discussed above, except that the PCEs of both our method and Wu become smaller, especially for small T. These findings highlight the wide applicability of our method in practice.

Table 1: False detection proportions (%) and average running time (in seconds) for DGP0, $N = 100$.

		$T = 100$		$T = 250$		$T = 500$	
		FDP	Time	FDP	Time	FDP	Time
C1	MT (BIC)	3.9	0.03	0.8	0.08	0	0.32
	MT (2BIC)	0	0.03	0	0.08	0	0.21
	MT (3BIC)	0	0.03	0	0.08	0	0.19
	Wu	100	0.13	0	0.26	0	0.70
C2	MT (BIC)	6.5	0.03	1.6	0.09	0	0.28
	MT (2BIC)	0	0.03	0	0.09	0	0.26
	MT (3BIC)	0	0.03	0	0.09	0	0.26
	Wu	100	0.13	0.4	0.26	0	0.69
C3	MT (BIC)	5.7	0.03	2.0	0.11	0	0.24
	MT (2BIC)	0.3	0.03	0	0.11	0	0.24
	MT (3BIC)	0	0.03	0	0.1	0	0.24
	Wu	100	0.13	0	0.26	0	0.70
C4	MT (BIC)	1.2	0.03	0	0.30	0.1	0.45
	MT (2BIC)	0	0.03	0	0.13	0	0.43
	MT (3BIC)	0	0.02	0	0.13	0	0.43
	Wu	100	0.13	100	2.13	100	2.16

Table 2: Percentages of correct estimation (%) of the threshold number, estimation accuracy of the threshold estimators ($\times 100$) and average running time (in seconds) for DGP3, $\delta = 1, N = 100$.

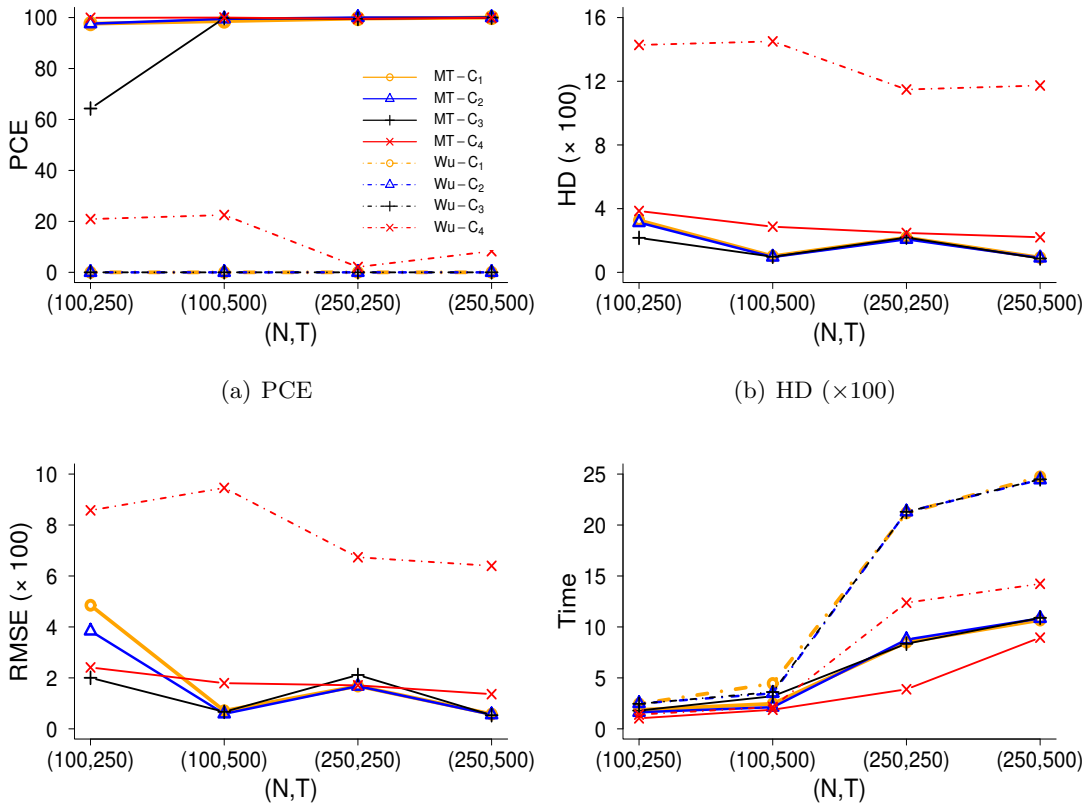
T			PCE	HD	Bias_1	RMSE_1	Bias_2	RMSE_2	Bias_3	RMSE_3	Time
250	C1	MT	97.6	2.64	-1.22	2.99	-1.06	2.64	-1.37	3.71	1.03
		Wu	73.3	3.79	-0.45	4.47	-0.19	7.21	-0.26	4.10	1.42
	C2	MT	98.6	2.42	-1.19	2.15	-0.98	1.49	-1.22	1.97	1.14
		Wu	76.2	3.13	-0.31	3.40	-0.21	2.44	-0.02	3.52	1.45
	C3	MT	98.9	2.35	-1.20	1.99	-1.00	1.50	-1.20	1.86	1.10
		Wu	83.0	3.32	-0.53	3.34	-0.01	4.76	0.06	3.33	1.48
	C4	MT	100	3.13	-1.73	2.51	-1.35	1.92	-1.83	2.78	0.71
		Wu	36.8	26.7	-10.2	34.1	-6.57	27.9	-3.22	20.0	1.42
500	C1	MT	97.7	1.12	-0.65	0.94	-0.48	0.73	-0.62	0.90	1.21
		Wu	88.4	1.77	-0.44	2.10	-0.13	1.39	-0.19	1.68	2.26
	C2	MT	98.6	1.20	-0.52	2.33	-0.43	2.26	-0.61	0.97	1.38
		Wu	82.4	1.91	-0.44	2.78	-0.13	1.89	-0.21	1.57	2.21
	C3	MT	98.9	1.14	-0.59	0.96	-0.47	0.75	-0.61	0.94	1.29
		Wu	89.5	2.00	-0.40	2.23	-0.03	2.18	-0.10	1.67	2.29
	C4	MT	100	2.07	-1.23	1.81	-0.83	1.22	-1.19	1.72	1.46
		Wu	31.5	27.7	-6.60	30.2	-6.58	27.9	-6.19	20.9	2.17

Table 3: Percentages of correct estimation (%) of the threshold number, estimation accuracy of the threshold estimators ($\times 100$) and average running time (in seconds) for DGP2.1, $\delta = 1, N = 250$.

T			PCE	HD	Bias_1	RMSE_1	Bias_2	RMSE_2	Time
100	C1	MT	64.2	3.88	-0.62	2.80	-2.87	7.14	0.04
		Wu	0	-	-	-	-	-	0.28
	C2	MT	69.7	3.23	-0.34	3.58	-1.87	5.64	0.04
		Wu	0	-	-	-	-	-	0.28
	C3	MT	56.4	4.75	-0.01	4.83	-3.19	7.76	0.04
		Wu	0	-	-	-	-	-	0.27
	C4	MT	74.3	2.22	-0.74	2.44	-1.11	2.43	0.05
		Wu	0	-	-	-	-	-	0.28
250	C1	MT	99.7	1.00	-0.19	0.93	-0.49	1.33	0.53
		Wu	0.1	145	-145	145	21.4	21.4	5.15
	C2	MT	98	1.19	-0.22	0.87	-0.63	2.40	0.57
		Wu	0	-	-	-	-	-	5.13
	C3	MT	97.1	1.05	-0.19	0.78	-0.59	2.13	0.54
		Wu	0.2	88.7	-88.7	91.6	15.6	18.4	5.14
	C4	MT	100	0.83	-0.36	0.74	-0.45	0.96	0.44
		Wu	25.1	81.5	-69.9	92.3	27.1	41.0	5.95
500	C1	MT	100	0.69	-0.03	0.52	-0.15	1.15	2.13
		Wu	0	-	-	-	-	-	6.51
	C2	MT	100	0.64	-0.06	0.54	-0.14	0.99	2.34
		Wu	0	-	-	-	-	-	6.49
	C3	MT	99	0.76	0.01	0.63	-0.28	1.23	2.02
		Wu	0	-	-	-	-	-	6.48
	C4	MT	100	0.59	-0.18	0.58	-0.17	0.70	2.15
		Wu	35.5	112	-107	124	38.1	49.7	13.0

Table 4: Percentages of correct estimation (%) of the threshold number, estimation accuracy of the threshold estimators (×100) and average running time (in seconds) for DGP2.2.

T			$N = 100$							$N = 250$						
			PCE	HD	Bias_1	RMSE_1	Bias_2	RMSE_2	Time	PCE	HD	Bias_1	RMSE_1	Bias_2	RMSE_2	Time
100	C1	MT	97.8	4.68	-2.59	5.03	-2.68	5.69	0.10	78.8	16.6	-1.25	17.4	-11.6	29.1	0.15
		Wu	0	-	-	-	-	-	0.14	0	-	-	-	-	-	0.29
	C2	MT	75.3	19.0	0.76	21.3	-12.1	30.0	0.08	85.8	12.1	-2.94	6.30	-9.94	26.1	0.17
		Wu	0	-	-	-	-	-	0.14	0	-	-	-	-	-	0.29
	C3	MT	89.5	10.7	-2.74	7.18	-7.99	22.8	0.10	87.8	10.8	-2.29	8.05	-8.38	23.0	0.17
		Wu	0	-	-	-	-	-	0.14	0	-	-	-	-	-	0.29
	C4	MT	87.4	11.0	-3.30	4.99	-9.07	23.0	0.08	86.2	10.2	-2.74	4.05	-8.71	23.5	0.17
		Wu	0	-	-	-	-	-	0.14	0	-	-	-	-	-	0.29
250	C1	MT	100	1.71	-1.08	1.63	-1.02	1.58	0.85	98.5	1.47	-0.60	1.42	-1.05	1.09	1.33
		Wu	97.3	2.59	-0.94	3.65	0.68	3.34	1.18	95.6	2.85	-1.18	4.85	0.46	3.70	8.50
	C2	MT	99.8	1.75	-1.07	1.68	-1.04	1.60	0.84	98.5	2.11	-1.10	1.59	-1.42	6.24	1.48
		Wu	96.6	2.48	-0.92	3.57	0.33	3.14	1.18	97.7	2.53	-1.10	4.56	0.37	2.94	8.94
	C3	MT	99.1	1.67	-1.00	1.57	-1.06	1.60	0.95	98.7	2.07	-1.16	1.67	-1.40	5.64	1.63
		Wu	92.5	3.39	-1.38	8.98	0.54	3.86	1.14	97.1	2.79	-1.23	5.72	0.52	3.48	8.87
	C4	MT	99.5	2.97	-1.54	2.48	-1.89	4.02	0.78	94.9	2.31	0.09	1.98	-1.83	2.73	1.03
		Wu	9.3	57.3	-43.2	64.5	27.1	48.8	1.41	83.9	11.9	-7.95	28.2	2.34	16.4	8.93
500	C1	MT	100	0.83	-0.50	0.76	-0.55	0.83	1.46	99.9	0.87	-0.55	0.85	-0.57	0.84	2.65
		Wu	99.3	1.34	-0.52	1.58	0.23	1.53	1.89	99.4	1.77	-0.80	3.38	0.19	2.32	10.9
	C2	MT	100	0.86	-0.51	0.78	-0.54	0.81	1.47	100	0.84	-0.56	0.80	-0.53	0.79	2.62
		Wu	99.7	1.73	-0.58	1.55	0.04	2.97	1.90	99.8	1.86	-0.82	3.84	0.33	2.57	11.1
	C3	MT	99.6	0.86	-0.54	0.83	-0.52	0.80	1.48	100	0.81	-0.55	0.78	-0.52	0.75	2.38
		Wu	98.8	1.56	-0.65	2.19	0.21	1.74	1.88	99.4	1.57	-0.59	2.11	0.20	2.20	11.0
	C4	MT	98.6	2.42	-0.91	1.67	-1.74	5.89	1.27	97.8	4.45	-0.93	1.45	-3.89	14.9	2.82
		Wu	2.5	72.6	-46.4	69.4	42.1	62.7	2.22	36.4	60.0	-38.5	65.9	27.7	53.6	13.9



(a) PCE

(b) HD (×100)

(c) RMSE (×100)

(d) Time

Figure 4.1: Plots of the estimation accuracy (PCE, HD, RMSE) and the average running time for DGP5, $\delta = 1$.

5 Empirical application

Ang and Timmermann (2012) pointed out that financial markets often undergo recurring regime shifts when "history repeats," and such phenomenon was described through a threshold factor specification by Massacci (2017) and Wu (2021). This section extends this line of research and shows how our framework is used to portray this nonlinear characteristic with a focus on financial variables. As argued in Baker et al. (2016), the index of economic policy uncertainty is associated with stock price volatility and employment in policy-sensitive sectors like defense, health care, finance, and infrastructure construction. The behaviors of the financial markets are different when the policy uncertainty differs in turn, which means that economic policy uncertainty has certain guiding effect on the behaviors of financial market. In the following, we first study how economic policy uncertainty affects the financial market by using our proposed threshold factor models, measure the connectedness in multivariate nonlinear dynamics system (Massacci, 2017), and then evaluate the importance of the common factors jointly in each regime. Finally, the

effectiveness of the threshold factor model contrasted with the structural break factor model is presented.

The monthly financial dataset used in Jurado et al. (2015) and Massacci (2017) consists of 147 series ($\boldsymbol{x}_t$) related to the U.S. financial markets, which are used to measure the behaviors of stock market and asset returns. The dataset covers the period ranging from January 1985 to December 2011, with a total of $T = 324$ observations, due to data availability issues. The threshold variable z_t is set as the lagged index of economic policy uncertainty (EPU) proposed by Baker et al. (2016), with a higher index value signifying higher uncertainty. See Jurado et al. (2015) and Baker et al. (2016) for detailed data description.

Connectedness is important for risk measurement and management, which is measured based on several different underlying mechanisms (Massacci, 2017). To tightly relate to our proposed model, we follow Massacci (2017) to define a connectedness measure that accommodates regime shifts and relies on the optimally selected number of eigenvalues. To be precise, let $\hat{\boldsymbol{\Sigma}}_{k\boldsymbol{x}} = [(NT)^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t\boldsymbol{x}_t^\top I(\hat{\theta}_{k-1}^\star < z_t \leq \hat{\theta}_k^\star)]$, where $\hat{\theta}_k^\star$'s denote the estimated thresholds from BEA, for $k = 1, \ldots, |\mathcal{A}_T^\star| + 1$, and $\{\omega_{ki}\}_{i=1}^N$ be the sequence of eigenvalues of the $N \times N$ covariance matrix $\hat{\boldsymbol{\Sigma}}_{k\boldsymbol{x}}$ arranged in descending order. Since the number of factors is allowed to vary in different regimes, we measure connectedness through

$$C_k(\hat{r}_k) = \frac{\sum_{i=1}^{\hat{r}_k}\omega_{ki}}{\sum_{i=1}^{N}\omega_{ki}}, \ k = 1, \ldots, |\mathcal{A}_T^\star| + 1.$$

Note that by considering the estimated number of factors in each regime, $C_k(\hat{r}_k)$ quantifies regime specific connectedness. A higher $C_k(\hat{r}_k)$ denotes higher connectedness amongst the financial variables.

To measure the importance of the factors jointly (i.e., the common component of the data), we compute the cross-sectional average of standard deviation of the common component $\hat{x}_{it} = \hat{\boldsymbol{\lambda}}_i^\top \hat{\boldsymbol{f}}_t$ in each separated regime. Ng and Wright (2013), who presented a factor view of business cycles, found that the factors estimated from a monthly panel data are noticeably more variable during recession years than the nonrecession years. In addition, they suggested that uncertainty should be a contributing factor to the severity of the Great Recession, which was later confirmed by Basu and Bundick (2017) that increased uncertainty about fiscal policy and demand may lead to an economic slowdown. Therefore, uncertainty may be the real driver of the variation in the factor importance. This leads us to check whether factor variability is related to the EPU.

We first build our multiple threshold factor model with the available data. After rearranging the data $\boldsymbol{x}_t$ according to the value of z_t, we fit a linear factor model with $\hat{r} = 8$, which is suggested by the information criterion IC_{p1} of Bai and Ng (2002). Then we use the two-step estimation procedure to help detect the thresholds. There are two thresholds found, i.e., $\hat{\theta}_1^\star = 75.65$ and $\hat{\theta}_2^\star = 139.8$, which separate the whole sample into three regimes, with frequencies equal to $\hat{\pi}_1 = 0.329$, $\hat{\pi}_2 = 0.512$ and $\hat{\pi}_3 = 0.158$, respectively. These three regimes correspond to the low, median, and high economic policy uncertainty. Note that the second estimated threshold

$\hat{\theta}_2^\star = 139.8$ is close the single threshold 131.4 discovered in Massacci (2017), indicating that the low policy uncertainty regime in his setting is further separated into two regimes in our model. The time series plot of the threshold variable EPU and the locations of the estimated thresholds $\hat{\theta}_1^\star$ and $\hat{\theta}_2^\star$ are shown in Figure 5.1.

We next calculate the estimated number of factors, connectedness, and the standard deviation of common components in each regime, with the results summarized in Table 5. The information criterion suggests that the numbers of factors in the three regimes are 3, 7 and 5, respectively. The connectedness measures of the three regimes are 0.779, 0.806 and 0.883, respectively, which shows an increasing trend as consistent with the finding of Massacci (2017). This growing trend of connectedness indicates that the variables are becoming more interrelated as the EPU increases, likely increasing the level of systemic risk, which corroborates the earlier discovery by Billio et al. (2012). This insight may be beneficial to risk measurement and management. Moreover, the standard deviations of common components in the three regimes are 0.862, 0.887 and 0.927, respectively. This means that the common factors are more important at higher economic policy uncertainty.

As pointed out by Massacci (2021), financial variables are likely to be heteroskedastic, which could invalidate the time invariance assumption imposed on the second moment of factors in financial markets. Consequently, the identified threshold effects could be driven by the changes in the factor dynamics. To disentangle breaks in loadings from those in factors, we next adapt the identification strategy of Chen et al. (2014) to the current more general framework that allows the number of factors to vary across regimes. The procedure is described as follows. We first estimate the number of factors in each regime separated by the estimated thresholds. Then we consider sequentially each pair of adjacent regimes that have the same number of factors (i.e., $\hat{r}_1 = \hat{r}_2$), merge the two subsamples into one subsample and re-estimate the number of factors in the merged subsample, denoted as $\hat{r}_{12}$. Finally, it is clear from the discussion of this paper that if the break comes from the loadings, then it holds that $\hat{r}_{12} > \hat{r}_1$; on the other hand, when $\hat{r}_{12} = \hat{r}_1 = \hat{r}_2$, the break comes from the factor heteroskedasticity. It is worth noting that, for the current study, the estimated factor numbers for the three regimes are 3, 7 and 5, respectively, indicating that there should be either emergence of new factors or disappearance of old factors and the changes in factor loadings occur. In this case, it is not clear whether factors are heteroskedastic or not.

To provide an intuitive interpretation of the estimated factors, we identify the series that load most heavily on each of the factors as done by Ng and Wright (2013). In the first regime with low EPU, the three factors represent the value weighted security price, the small stock value spread $R15 - R11$ and portfolios sorted on (size, book to market), respectively (Jurado et al., 2015). In the second regime with moderate EPU, 7 series mainly contain the group of variables called "risk-factors", since they have been used in cross-sectional or time-series studies to uncover variation in the market risk-premium. The risk factors include the three Fama and French (1992)

risk factors, namely the excess return on the market, the "small-minus-big" (SMB) and "high-minus-low" (HML) portfolio returns, the momentum factor "up-minus-down" (UMD), and the bond risk premia factor of Cochrane and Piazzesi (2005). In addition to these risk factors, the 7 series are also related to food industry portfolio and dividends. In the last regime with high EPU, the 5 factors capture market portfolio under reinvestment of dividends, risk factors (UMD, SMB), dividends, and coal industry portfolio, respectively. Factors representing market risk-premium appear in all three regimes, which include $R15-R11$ in the first regime, risk factors mentioned above in the second, and UMD, SMB in the last. This indicates that the variables related to market risk play a pivotal role in financial market, especially in the second regime where 5 of 7 factors are bound with risk. Furthermore, there exist some industrial portfolios, such as food and coal industries, which occupy an increasingly important position in financial market as EPU increases.

We finally contrast the performance of the proposed threshold factor models with the structural break factor models. The EPU plotted in Figure 5.1 is relatively high in the recession years (the National Bureau of Economic Research (NBER) recession periods are in shaded areas), consistent with the findings in Basu and Bundick (2017). This points to the possibility that both types of structural instabilities could be appropriate for the observed data. As a result, we next fit a factor model with change points in the loadings following the procedure proposed by Ma and Tu (2022). Two structural breaks are found, i.e., $\hat{\tau}_1$ =1999M2 and $\hat{\tau}_2$ =2008M1, which also split the time series into three regimes, as shown in Figure 5.1. The two change points are closely related to the two recessions in 2000 and 2008. As can be seen from Figure 5.1, the structural break effect induces low-frequency episodes in time, while the threshold effect generates recurring regime shifts at much higher frequency, although both types of instability in factor loadings separate the time series into three regimes. To evaluate the two types of models, we follow the strategy used by Massacci (2021) to select the threshold variable that adopts the following information criterion $IC = \frac{1}{3}\sum_{k=1}^{3}\left[IC_{p1}(\hat{r}_k)\right]$, where IC_{p1} is the information criterion proposed by Bai and Ng (2002) to estimate the number of factors and $\hat{r}_k$ is the selected number of factor in the k-th regime. The IC calculated from the threshold factor model is 3.688, while that from the estimated factor model with two structural breaks is 3.695. Although the two types of models are non-nested, the slightly lower IC value from our threshold factor model could indicate its superiority as a better fit to the data. A rigorous theoretical study to compare these two types of models would provide more insight here but is beyond the scope of this paper due to space consideration.

6 Concluding remarks

This paper proposes a simple-to-implement procedure to estimate the multiple unknown thresholds in high-dimensional threshold factor model. The procedure is comprised of sorting the

Table 5: Empirical influence of economic policy uncertainty on the financial market.

Thresholds		$\hat{\theta}_1^\star = 75.65$	$\hat{\theta}_2^\star = 139.8$
Regime	1st	2nd	3rd
Frequency	0.329	0.512	0.158
Number of Factors	3	7	5
Connectedness	0.779	0.806	0.883
Standard deviation	0.862	0.887	0.927

original data according to the value of the threshold variable, one factor estimation using PCA, and one group Lasso to find breaks in the factor regression coefficients, which is quite computationally efficient. It is shown that the number and the locations of the thresholds, and the factor space can be consistently estimated. Simulation studies demonstrate superior performance of the proposed method in finite samples. Applications to the U.S. macroeconomic and financial market data set further illustrate the merits of our method.

The current study could be extended in several aspects. First, it is theoretically interesting to extend the current results to the case where the number of thresholds and the number of factors diverge along N and T. The theoretical setup of Li et al. (2017) could be inspiring in this regard. Second, after representing the threshold factor model in the change-point form, there are several alternative methods that can be applied to detect the structural breaks, such as wild binary segmentation, MOSUM, etc. The asymptotic properties of the corresponding threshold estimators worth detailed investigation. Third, the threshold variable could be endogenous in practice, which has been incorporated in our model as long as Assumption 6(a) holds. Our treatment on endogeneity without the need for instrumental variable has been similarly used by Qian and Su (2014) when estimating structural changes in the presence of endogenous regressors. Yu and Phillips (2018) showed that instrumentation can improve efficiency under the framework of endogeneous threshold regression, which may shed light on the threshold factor setup. Fourth, the present work can be further extended to allow both structural breaks and threshold effects simultaneously in the factor models. In this way, a theory of the information criterion used in Section 5 to select between structural breaks and threshold effects in factor models can be developed. These new topics involve quite challenging technical difficulties, therefore deserve separate future effort.

Acknowledgments

The authors would like to thank the Co-Editor, Serena Ng, and four anonymous referees for constructive suggestions that help improve the paper significantly, and acknowledge support from National Natural Science Foundation of China (Grant 72073002, 12026607, 92046021),

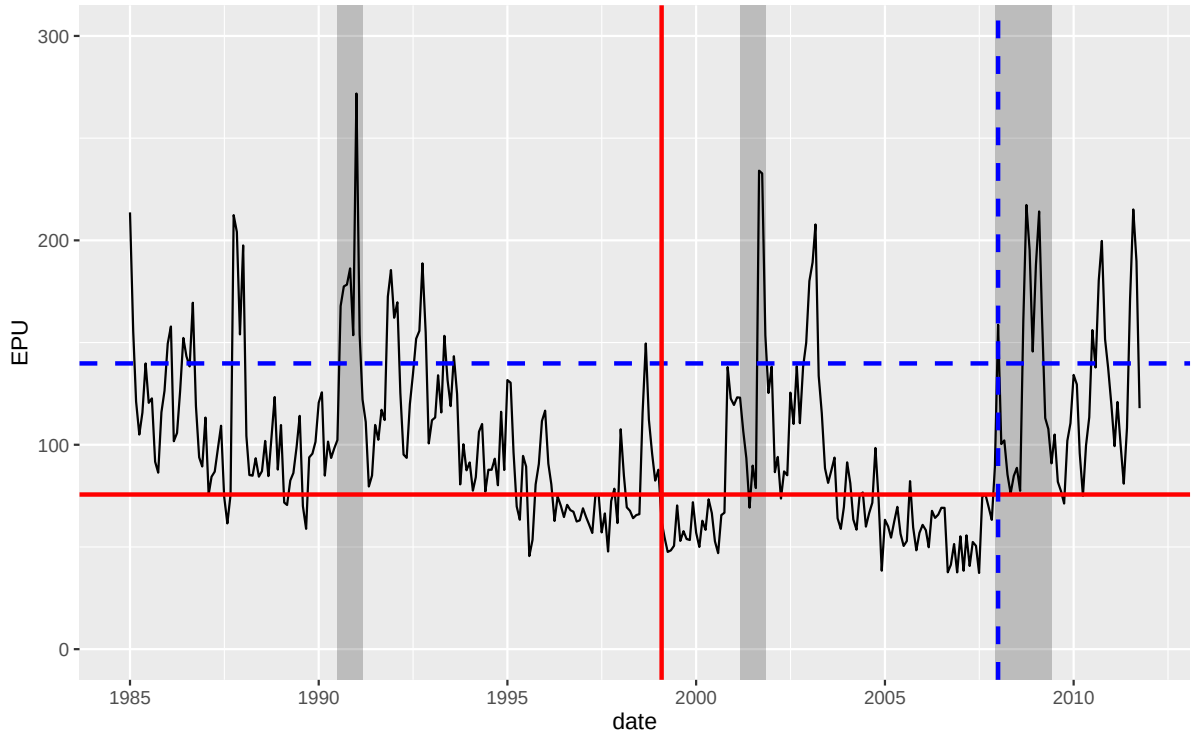


Figure 5.1: Time series plot of the economic policy uncertainty, with two thresholds $\hat{\theta}_1^\star = 75.65$ (red solid horizontal line) and $\hat{\theta}_2^\star = 139.8$ (blue dashed horizontal line), and two structural breaks $\hat{\tau}_1 = 1999M2$ (red solid vertical line) and $\hat{\tau}_2 = 2008M1$ (blue dashed vertical line). The shaded areas in gray denote the NBER recession periods.

the Center for Statistical Science at Peking University, and Key Laboratory of Mathematical Economics and Quantitative Finance (Peking University), Ministry of Education.

Appendix

The inference for the factor number and factor space in each regime, empirical investigation of the influence of economic condition on the macroeconomy, technical proofs and some additional simulation results are contained in an online Supplementary Material.

References

Ang, A. and Timmermann, A. (2012). Regime changes and financial markets. *Annual Review of Financial Economics*, 4(1):313–337.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Baltagi, B. H., Kao, C., and Wang, F. (2017). Identification and estimation of a large factor model with structural instability. *Journal of Econometrics*, 197(1):87–100.

Baltagi, B. H., Kao, C., and Wang, F. (2021). Estimating and testing high dimensional factor models with multiple structural changes. *Journal of Econometrics*, 220(2):349–365.

Barigozzi, M., Cho, H., and Fryzlewicz, P. (2018). Simultaneous multiple change-point and factor analysis for high-dimensional time series. *Journal of Econometrics*, 206(1):187–225.

Basu, S. and Bundick, B. (2017). Uncertainty shocks in a model of effective demand. *Econometrica*, 85(3):937–958.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37(4):1705 – 1732.

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559. Market Institutions, Financial Market Risks and Financial Crisis.

Breitung, J. and Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, 163(1):71–84.

Chan, K.-S., Hansen, B. E., and Timmermann, A. (2017). Guest editors introduction: Regime switching and threshold models. *Journal of Business & Economic Statistics*, 35(2):159–161.

Chan, N. H., Yau, C. Y., and Zhang, R. (2015). Lasso estimation of threshold autoregressive models. *Journal of Econometrics*, 189(2):285–296.

Chen, L., Dolado, J. J., and Gonzalo, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics*, 180(1):30–48.

Cheng, X., Liao, Z., and Schorfheide, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies*, 83(4):1511–1543.

Cochrane, J. H. and Piazzesi, M. (2005). Bond risk premia. *American Economic Review*, 95(1):138–160.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465.

Gonzalo, J. and Pitarakis, J.-Y. (2002). Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics*, 110(2):319–352.

Han, X. and Inoue, A. (2015). Tests for parameter instability in dynamic factor models. *Econometric Theory*, 31(5):1117–1152.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *The American Economic Review*, 105(3):1177–1216.

Kim, C.-J. and Nelson, C. (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *The Review of Economics and Statistics*, 80(2):188–201.

Li, D. and Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models. *Journal of Econometrics*, 167(1):240–253.

Li, H., Li, Q., and Shi, Y. (2017). Determining the number of factors when the number of factors can increase with sample size. *Journal of Econometrics*, 197(1):76–86.

Liu, X. and Chen, R. (2020). Threshold factor models for high-dimensional time series. *Journal of Econometrics*, 216(1):53–70.

Ma, C. and Tu, Y. (2022). Group fused lasso for large factor models with multiple structural breaks. *Journal of Econometrics*. Online https://doi.org/10.1016/j.jeconom.2022.02.003.

Ma, S. and Su, L. (2018). Estimation of large dimensional factor models with an unknown number of breaks. *Journal of Econometrics*, 207(1):1–29.

Massacci, D. (2017). Least squares estimation of large dimensional threshold factor models. *Journal of Econometrics*, 197(1):101–129.

Massacci, D. (2021). Testing for regime changes in portfolios with a large number of assets: A robust approach to factor heteroskedasticity. *Journal of Financial Econometrics*, forthcoming.

Ng, S. and Wright, J. H. (2013). Facts and challenges from the great recession for forecasting and macroeconomic modeling. *Journal of Economic Literature*, 51(4):1120–1154.

Qian, J. and Su, L. (2014). Structural change estimation in time series regressions with endogenous variables. *Economics Letters*, 125(3):415–421.

Qian, J. and Su, L. (2016). Shrinkage estimation of regression models with multiple structural changes. *Econometric Theory*, 32(6):1376–1433.

Su, L. and Wang, X. (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics*, 198(1):84–101.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 42(3):245–292.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, 84(405):231–240.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wu, J. (2021). Estimation of high dimensional factor model with multiple threshold-type regime shifts. *Computational Statistics and Data Analysis*, 157:107–153.

Yamamoto, Y. and Tanaka, S. (2015). Testing for factor loading structural change under common breaks. *Journal of Econometrics*, 189(1):187–206.

Yu, P. and Phillips, P. C. (2018). Threshold regression with endogeneity. *Journal of Econometrics*, 203(1):50–68.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.