Electronic Journal of Statistics
Vol. 16 (2022) 1891–1951
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS1982

Penalized estimation of threshold auto-regressive models with many components and thresholds*†

Kunhui Zhang[1]
Abolfazl Safikhani[2]
Alex Tank[1]
Ali Shojaie[1,3]

[1] *University of Washington, Department of Statistics, Padelford Hall, W Stevens Way NE, Seattle, WA 98195*
e-mail: zhangkh@uw.edu; alextank@uw.edu; ashojaie@uw.edu

[2] *University of Florida, Department of Statistics, 102 Griffin-Floyd Hall, Gainesville, FL 32611*
e-mail: a.safikhani@ufl.edu

[3] *University of Washington, Department of Biostatistics, Health Sciences Building, 1705 NE Pacific Street, Seattle, WA 98195*

Abstract: Thanks to their simplicity and interpretable structure, auto-regressive processes are widely used to model time series data. However, many real time series data sets exhibit non-linear patterns, requiring non-linear modeling. The threshold Auto-Regressive (TAR) process provides a family of non-linear auto-regressive time series models in which the process dynamics are specific step functions of a thresholding variable. While estimation and inference for low-dimensional TAR models have been investigated, high-dimensional TAR models have received less attention. In this article, we develop a new framework for estimating high-dimensional TAR models, and propose two different sparsity-inducing penalties. The first penalty corresponds to a natural extension of classical TAR model to high-dimensional settings, where the same threshold is enforced for all model parameters. Our second penalty develops a more flexible TAR model, where different thresholds are allowed for different auto-regressive coefficients. We show that both penalized estimation strategies can be utilized in a three-step procedure that consistently learns both the thresholds and the corresponding auto-regressive coefficients. However, our theoretical and empirical investigations show that the direct extension of the TAR model is not appropriate for high-dimensional settings and is better suited for moderate dimensions. In contrast, the more flexible extension of the TAR model leads to consistent estimation and superior empirical performance in high dimensions.

Keywords and phrases: Non-linear time series, high-dimensional time series, fused lasso, threshold estimation.

Received November 2021.

*This work was partially supported by grants DMS-1915855 and DMS-1722246 from the National Science Foundation and grant R01GM133848 from the National Institutes of Health.

†This is an original paper.

1. Introduction

The threshold Auto-Regressive (TAR) model [44, 46] allows regime-specific auto-regressive parameters, where the regimes are governed by a thresholding random variable, typically some previous lag of the time series (see formal definition in Section 2). Thanks to its flexibility, the TAR model has become a popular framework for analyzing non-linear time series from diverse application domains, from economics [27] and finance [11] to genomics [24] and epidemiology [53]. Applications in macroeconomics have been particularly diverse: Enders et al. [15] modeled the U.S. GDP growth, and constructed confidence intervals for the parameters; Juvenal and Taylor [25] explored the validity of the law of one price in nine European countries; and Aslan et al. [1] applied a TAR model to commodity prices, and used it to represent abrupt changes, time-irreversibility, and regime-shifting behavior. See Hansen [20] for a selective review of threshold autoregression in economics.

TAR models have been extensively studied in univariate and fixed-dimensional settings. For example, Chan [8] investigated the asymptotic properties of the least squares estimation for TAR models with two regimes, Chen [12] proposed an estimation procedure when the thresholding variable is unknown, Bruce [6] derived the asymptotic distribution of general TAR models, and Li et al. [29] developed the asymptotic theory of the least squares estimator for a moving average TAR model. In other related work, Chan and Kutoyants [9] proved the consistency of a Bayesian estimator of the TAR model, while Chan et al. [10] proposed a novel modified LASSO approach for threshold estimation and established its consistency in multiple threshold models. Tsay [48] first extended univariate TAR models to multivariate settings, and proposed to use grid search based on the Akaike information criterion (AIC) to select the thresholds. Later, Lo and Zivot [33], Hansen and Seo [21], Dueker et al. [14], Li and Tong [28] used grid search based methods to study the multivariate TAR models assuming either a known number of thresholds or an upper bound on the number of thresholds. However, these approaches may not work in practice, as the number of thresholds is often unknown. More recently, Calderón V and Nieto [7], Orjuela and Villanueva [37] introduced Bayesian methodologies for the estimation of thresholds in multivariate TAR models with an unknown number of thresholds. These methods bypass the assumptions on the number of thresholds, but do not establish the consistency of the number of the estimated thresholds. Another limitation of existing approaches is that they are not applicable in high dimensions. The advantages and limitations of existing approaches are summarized in Table 3 in Appendix 6. See also Tong [45] for a review of threshold models in time series analysis.

High-dimensional time series models have received considerable attention in recent years [2, 26, 19]. In this setting, the ambient dimension is of the same order or larger than the sample size. This poses numerous practical and theoretical challenges. While a number of theoretical results have been established for linear time series models in high dimensions, with few exceptions [e.g., 13, 42], their non-linear counterparts have received less attention. In the context of threshold models, the recent work by Liu and Chen [32] investigates the estimation of threshold factor models with growing number of variables. However, this work assumes a single threshold, which limits the flexibility of the model. Moreover, while the number of time series components is allowed to grow, it is assumed to be smaller than the sample size (see Theorem 1 in [32]). In fact, to the best of our knowledge, methods and theory for high-dimensional TAR models are currently lacking.

Given the paucity of the literature on high-dimensional TAR models, in this paper, we propose two estimators for detecting the (unknown) number and values of thresholds and estimating regime-specific auto-regressive parameters in multivariate TAR models with many components. Both approaches are based on a three-step estimation framework and utilize similar penalized estimation strategies, but they differ in one key aspect. The first approach is a natural extension of the classical TAR model and enforces all auto-regressive parameters to change at the same thresholds. As we discuss in Section 3, this assumption may be too restrictive in high-dimensional settings with many components. In fact, our theoretical and empirical investigations indicate that the extension of the classical TAR is not appropriate for high-dimensional settings and is better suited for moderate dimensions. As such, we refer to this first version as the multivariate TAR (mvTAR) model. To mitigate the limitation of the mvTAR model, we then propose a more flexible high-dimensional TAR model (hdTAR) where different auto-regressive parameters are allowed to change at different thresholds. This flexibility seems to introduce a new challenge, as the model may have many thresholds. However, our theoretical and empirical investigations show that this flexibility is indeed necessary in high dimensions and leads to improved theoretical guarantees and empirical performances. We develop efficient algorithms for both methods and establish the consistency of the thresholds and auto-regressive parameters under certain mixing conditions.

To establish our theoretical results, we address two key challenges that arise in penalized estimation of high-dimensional TAR models. The first challenge involves verifying appropriate concentration inequalities, including two main ingredients in high-dimensional statistics: (1) a restricted eigenvalue condition and (2) a deviation bound condition [34]. These conditions are crucial in deriving consistency results in high-dimensional settings, as hinted in Bickel et al. [4]. The conditions have been previously verified in the setting of i.i.d. observations and, more recently, studied in certain linear time series models [2, 41]. However, extending these results to non-linear TAR models is challenging. This is primarily due to the random ordering of the design matrix based on the threshold (switching) variable (see e.g. Equation (3)). To address this challenge, we develop a bracketing argument [50, 10] specifically designed to handle the threshold-type structure (see Lemmas 5 and 7 in the Appendix). These results are verified under certain mixing conditions (see Assumption A2 in Section 4) and are of independent interest in the context of non-linear high-dimensional time series models. The second challenge concerns our screening step to consistently estimate the number of thresholds. Many theoretical results in the context of TAR models assume that the number of thresholds is known [33, 32]. This assumption may not be realistic in practice; in fact, it is appealing to infer the number of thresholds from data. To that end, the second step of our proposed algorithms utilizes an information criterion that screens candidate thresholds identified in the first step and removes redundant ones. This step successfully resolves the challenge by consistently estimating the number of thresholds with high probability (see Theorem 2).

The rest of the paper is organized as follows. After formally defining the multivariate TAR model in Section 2, we describe our algorithms in Section 3 and establish their theoretical properties in Section 4. In Section 5, we propose data-driven methods to select the hyper-parameters. While the required hyper-parameters are characterized in our asymptotic results, these rates involve unknown constants and cannot be used in practice. The empirical performance of the proposed methods is investigated using both simulated and real data sets, in Section 6 and Section 7, respectively. We conclude with a brief summary in Section 8.

2. Multivariate TAR formulations

The classical TAR model, proposed by Tong and Lim [46], is defined as

$$x_t = a^0_{(j)} + \sum_{k=1}^{K} a^k_{(j)} x_{t-k} + \sigma_{(j)}\epsilon_t, \quad \text{if } r_{j-1} < z_t \le r_j, \tag{1}$$

where m_0 denotes the number of thresholds, r_js are the threshold parameters which partition the time series into m_0+1 regimes, K is the number of lags to be considered in the model, z_t is a switching variable (maybe functions of some components of x_t), σ_js are segment-specific error variances, and $a^0_{(j)}$ and $a^k_{(j)}$ are coefficients in regime j, for $j = 1, \dots, m_0+1$ (they are allowed to be different in each regime). The noise or innovation, ϵ_t, is an i.i.d. sequence of random variables with zero mean and unit variance.

The original TAR model was restricted to univariate time series, but can be extended to multivariate settings, as described in [47]. Formally, a multivariate time series $\{\boldsymbol{x}_t\}$ follows TAR model with one switching variable if

$$\boldsymbol{x}_t = \sum_{k=1}^{K} \boldsymbol{A}^{(k,j)} \boldsymbol{x}_{t-k} + \boldsymbol{\Sigma}_j^{1/2}\boldsymbol{\epsilon}_t, \quad \text{if } r_{j-1} < z_t \le r_j, \tag{2}$$

where $\boldsymbol{x}_t = (x_{(t,1)}, x_{(t,2)}, \dots, x_{(t,p)})'$ is the observed process in $\mathbb{R}^p$ at time t, p is the number of time series components, and K is the number of lags considered in the model. Here $\boldsymbol{\epsilon}_t = (\epsilon_{(t,1)}, \epsilon_{(t,2)}, \dots, \epsilon_{(t,p)})' \in \mathbb{R}^p$ is a multivariate i.i.d. sequence with zero mean in all components. The covariance matrix $\boldsymbol{\Sigma}_j$ for the j-th regime, $\boldsymbol{\Sigma}_j$, is allowed to be different in each regime. To simplify the notations, when there is no ambiguity, we simply denote the error term by $\boldsymbol{\epsilon}_t$ instead of $\boldsymbol{\Sigma}_j^{1/2}\boldsymbol{\epsilon}_t$. The transition matrices $\boldsymbol{A}^{(k,j)} \in \mathbb{R}^{p\times p}$ is the coefficient matrix corresponding to the k-th lag of a TAR process in regime j. More specifically, similar to the modeling framework of [10], we assume there exist m_0 threshold values $-\infty < r_1 < r_2 < \ldots < r_{m_0} < +\infty$ with $r_0 = -\infty$ and $r_{m_0+1} = +\infty$ which partition the process into $m_0 + 1$ regimes. For each regime, the total transition matrices $\boldsymbol{A}^{(.,j)} = \left(\boldsymbol{A}^{(1,j)}, \boldsymbol{A}^{(2,j)}, \ldots, \boldsymbol{A}^{(K,j)}\right) \in \mathbb{R}^{p \times pK}$ are fixed where $r_{j-1} < z_t \leq r_j$ for $j = 1, \ldots, m_0 + 1$.

Our goal is to estimate the number of thresholds, i.e. m_0, together with the threshold values, r_j, and the auto-regressive parameters in each regime.

Next, we introduce some additional notations. For a symmetric matrix $\mathbf{X}$, let $\lambda_{\min}(\mathbf{X})$ and $\lambda_{\max}(\mathbf{X})$ denote its minimum and maximum eigenvalues. Let the h-th row of $\boldsymbol{A}^{(.,j)}$ be $\boldsymbol{A}_h^{(.,j)}$, and set the number of non-zero elements in $\boldsymbol{A}_h^{(.,j)}$ to $d_{h,j}$ for $h = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, m_0 + 1$. Denote the total sparsity of the model by $d_n^* = \sum_{j=1}^{m_0+1} \sum_{h=1}^{p} d_{h,j}$. Further, let $\mathcal{I}_{h,j}$ represent the set of all column indexes of $\boldsymbol{A}_h^{(.,j)}$, $\mathcal{I} = \cup_{h,j} \mathcal{I}_{h,j}$ and define $d_n = \max_{1 \leq h \leq p, 1 \leq j \leq 1+m_0} |\mathcal{I}_{h,j}|$. Note that p, m_0 and the sparsity may increase with the number of time points, T, specifically, $p \equiv p(n)$ and $m_0 \equiv m_0(n)$ and $d_{h,j} \equiv d_{h,j}(n)$, where $n = T - K$. For simplicity, we suppress the n-index. Finally, let $\epsilon_{t,l}$ be error term of l-th time series, and recall that $\boldsymbol{\epsilon}_t = \left(\epsilon_{(t,1)}, \epsilon_{(t,2)}, \ldots, \epsilon_{(t,p)}\right)'$. Throughout the paper, positive constants $C, C_1, C_2, \ldots$ are used to denote universal constant, $\boldsymbol{A}'$ denotes the transpose of a matrix $\boldsymbol{A}$, and $\|\boldsymbol{A}\|_1$ and $\|\boldsymbol{A}\|_2$ denotes its ℓ_1 and Frobenius norms, respectively. We denote the ℓ_1 and ℓ_2 norms of a vector v by $\|v\|_1$ and $\|v\|$, respectively.

3. Regularized estimation of high-dimensional TARs

The number of parameters in the TAR model (2), $(m_0 + 1)(Kp^2)$, increases with the number of time series p and the number of thresholds m_0. Estimating these parameters becomes especially challenging when the model has more than one threshold, i.e. $m_0 > 1$, and the number of thresholds is unknown. This is because identifying the thresholds would require a search over all possible values of threshold levels z_t, which is infeasible.

To overcome the above challenges, in Section 3.1 we first reformulate the TAR estimation problem via a non-parametric model with $(T - K)p^2K$ parameters. This over-parameterization allows us to use regularized estimation strategies to efficiently obtain an initial estimate of the thresholds by solving a penalized least squares estimation problem. In particular, we use a total variation penalty [43] to obtain piecewise constant estimates of $\boldsymbol{A}^{(k,j)}$ for regime j with respect to the threshold variable z_t.

The classical multivariate TAR model (2) requires the parameters of transition matrices $\boldsymbol{A}^{(k,j)}$ to change at the same threshold values z_t. To obtain such an estimate, we consider a grouped fused lasso penalty in Section 3.2. The resulting estimate, referred to as mvTAR, is suitable for low-to-moderate-dimensional problems, where p is fixed or small compared to the number of observations T. However, for problems with large p, especially when $p \gg T$, requiring that all transition matrix parameters change at the same threshold value becomes

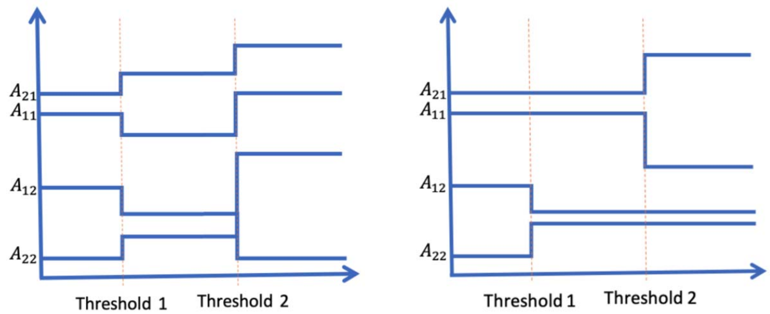


Fig 1. *Example of changes of transition matrices. The left panel depicts the situation in which the classical TAR multivariate TAR model (mvTAR) in which all elements of the transition matrices change together at all threshold values. The right panel illustrates the proposed flexible TAR model for high dimensions (hdTAR) in which different elements of the transition matrices would not change at some threshold values.*

restrictive. Moreover, the theoretical advantages of the group lasso penalty dissipate when grouped parameters do not follow the same sparsity pattern [23]. These limitations are reflected in our theoretical and numerical analyses in Sections 4 and 6. To achieve efficient estimation in high-dimensions, in Section 3.2 we propose a more flexible high-dimensional TAR model, named hdTAR, in which transition matrix parameters are allowed to change at different thresholds. As we show, this flexibility results in theoretical and empirical advantages. The difference between the flexible TAR model and the original version is illustrated in Figure 1.

Both our group and regular fused lasso penalties overestimate the number of thresholds. This is because a key requirement for consistency of ℓ_1-regularized estimation strategies, namely the restricted eigenvalue property [4] is not guaranteed to hold in our setting (see Section 4). To remove the redundant selected thresholds, we introduce a screening criterion in Section 3.3 that consistently estimates the (many) unknown thresholds. In Section 3.4, we obtain consistent estimates of high-dimensional auto-regressive parameters within each estimated regime.

3.1. Reparametrization of the TAR model

In this section, we reparametrize the TAR model (2) by considering n transition matrices for each value of the ordered switching variable z_t (assuming, without loss of generality, that z_t assumes unique values).

Let $n = T - K$ and let $\pi(i)$ be the time index of the i-th smallest element of z_t for $i = 1, \ldots, n$. Then the TAR model (2) with lag K can be written as

$$\begin{pmatrix} \boldsymbol{x}'_{\pi(1)} \\ \boldsymbol{x}'_{\pi(2)} \\ \vdots \\ \boldsymbol{x}'_{\pi(n)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}'_{\pi(1)-1} & \cdots & \boldsymbol{x}'_{\pi(1)-K} & & \mathbf{0} & & \cdots & & \mathbf{0} & \\ \boldsymbol{x}'_{\pi(2)-1} & \cdots & \boldsymbol{x}'_{\pi(2)-K} & \boldsymbol{x}'_{\pi(2)-1} & \cdots & \boldsymbol{x}'_{\pi(2)-K} & \cdots & & \mathbf{0} & \\ & \vdots & & & & & \ddots & & & \\ \boldsymbol{x}'_{\pi(n)-1} & \cdots & \boldsymbol{x}'_{\pi(n)-K} & \boldsymbol{x}'_{\pi(n)-1} & \cdots & \boldsymbol{x}'_{\pi(n)-K} & \cdots & \boldsymbol{x}'_{\pi(n)-1} & \cdots & \boldsymbol{x}'_{\pi(n)-K} \end{pmatrix} \times \begin{pmatrix} \boldsymbol{\theta}'_1 \\ \boldsymbol{\theta}'_2 \\ \vdots \\ \boldsymbol{\theta}'_n \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}'_{\pi(1)} \\ \boldsymbol{\epsilon}'_{\pi(2)} \\ \vdots \\ \boldsymbol{\epsilon}'_{\pi(n)} \end{pmatrix}. \tag{3}$$

Let $\boldsymbol{\theta}_1 = \left(\boldsymbol{A}^1_{\pi(1)}, \dots, \boldsymbol{A}^K_{\pi(1)}\right) \in \mathbb{R}^{p \times pK}$, and $\boldsymbol{\theta}_i = \left(\boldsymbol{A}^1_{\pi(i+1)} - \boldsymbol{A}^1_{\pi(i)}, \dots, \boldsymbol{A}^K_{\pi(i+1)} - \boldsymbol{A}^K_{\pi(i)}\right)$, where $\boldsymbol{A}^k_{\pi(i)}$ is the transition matrix for i-th ordered observation at lag k. Denote the response matrix, the design matrix, the model parameters and the error term in Equation (3) by $\mathcal{Y}, \mathcal{X}, \Theta$ and E, respectively. Then, (3) can be written as $\mathcal{Y} = \mathcal{X}\Theta + E$. Moreover, letting $\boldsymbol{Y} = \text{vec}(\mathcal{Y}) \in \mathbb{R}^{np \times 1}$, $\boldsymbol{\Theta} = \text{vec}(\Theta) \in \mathbb{R}^{np^2K \times 1}$, $\boldsymbol{E} = vec(E) \in \mathbb{R}^{np \times 1}$, and $\boldsymbol{Z} = I_p \otimes \mathcal{X} \in \mathbb{R}^{np \times np^2K}$ with $\otimes$ denoting the tensor product, (3) can be written in vector form as

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\Theta} + \boldsymbol{E}. \tag{4}$$

While redundant, the over-parametrization in (3) has an important benefit: $\boldsymbol{\theta}'_i \neq 0$ if and only if the auto-regressive coefficients change in TAR process at time $\pi(i)$. Thus, finding the thresholds is equivalent to finding non-zero $\boldsymbol{\theta}_i$s for $i > 1$. In other words, the problem of threshold estimation can be translated to a high-dimensional variable selection problem in (4).

3.2. Penalties for moderate- and high-dimensional TARs

Sparsity-inducing penalties, such as lasso, are particularly suitable for estimating $\boldsymbol{\Theta}$ in (4): A sparse estimate $\hat{\boldsymbol{\theta}}_1$ gives an interpretable estimate of the transition matrices for the smallest value of z_t, while sparsity in $\hat{\boldsymbol{\theta}}_i$ for $i > 1$ would imply no changes in the transition matrices over z_t. Such a strategy corresponds to a fused lasso, or total variation, penalty [43, 38]. In this paper, we consider a similar strategy and obtain an estimate of $\boldsymbol{\Theta}$ by solving

$$\hat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta}} \|\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{\Theta}\|_2^2 + \lambda_1 \|\boldsymbol{\Theta}\|_\diamond + \lambda_2 \sum_{i=1}^{n} \left\| \sum_{i'=1}^{i} \boldsymbol{\theta}_{i'} \right\|_1, \tag{5}$$

The first penalty in (5), $\|\cdot\|_\diamond$, encodes either an ℓ_2, or grouped fused lasso penalty, $\|\cdot\|_2$, or an ℓ_1, or fused lasso penalty, $\|\cdot\|_1$. The group fused lasso penalty encourages all entries of the transition matrices to change at the same threshold values. In contrast, the fused lasso penalty provides a more flexible TAR model in which different transition matrix parameters are allowed to change at different thresholds. As discussed earlier, the group fused lasso penalty is only suitable for low to moderate-dimensional problems (where p is allowed to grow, but $p < T$), whereas the more flexible fused lasso penalty is appropriate for both moderate- and high-dimensional problems (where $p \gg T$); see also Figure 1. In both cases, the magnitude of the penalty is controlled by the tuning parameter λ_1, which is chosen data-adaptively via cross validation; see Section 5 for more details.

The second penalty in (5), controlled by tuning parameter λ_2, further encourages the overall sparsity of the estimated transition matrices by penalizing changes in transition matrices after each potential threshold index i. While often not needed in practice, this additional sparsity results in improved estimation and allows us to obtain better rates of convergence for the proposed estimator in Section 4.

With either ℓ_2 or ℓ_1 penalties, the optimization problem in (5) is convex and can be solved efficiently. With the ℓ_2 penalty, the problem can be solved using a sub-gradient descent algorithm. However, the problem further simplifies when $\lambda_2 = 0$ and we can instead use a more efficient proximal gradient descent algorithm; see Algorithm 2 in the Appendix. With the ℓ_1 penalty, the problem is easy to solve efficiently using a path-wise coordinate descent algorithm [17] regardless of the value of λ_2. This is because, by Proposition 1 in [17], it suffices to first find the solution for $\lambda_2 = 0$, and then apply an element-wise soft thresholding operator; see Algorithm 1 in the Appendix.

3.3. Threshold selection

Using Equation (5), we can define a set of candidates threshold estimates as

$$\hat{\mathcal{A}}_n = \left\{ z_{\pi(i-1)} : \|\hat{\boldsymbol{\theta}}_i\|_2 \neq 0 \ , i \geq 2 \right\}. \tag{6}$$

Let $\hat{r}_j$ be the j-th sorted (from the lowest to the highest) estimated threshold in the set $\hat{\mathcal{A}}_n$, and let $\hat{m}$ be the cardinality of the set $\hat{\mathcal{A}}_n$. As we show in Section 4, it is likely for the fused lasso to over-estimate the number of thresholds [22]. Thus, we need to remove the redundant thresholds. In our screening step, we aim to keep exactly m_0 points in $\hat{\mathcal{A}}_n$ which are close enough to the true threshold values. To that end, we develop an *information criterion* by modifying the screening procedure of [41] to make it more suitable for the threshold structure of model (2). Essentially, this step consists of estimating the transition parameters within each estimated regime $\{t : \hat{r}_j < z_t \leq \hat{r}_{j+1}\}$ for $j = 0, 1, \dots, \hat{m}$ with $\hat{r}_0 = -\infty$ and $\hat{r}_{\hat{m}+1} = +\infty$ and comparing the total sum of squared error (SSE) before and after excluding a certain estimated threshold $\hat{r}_j$. The basic idea is to keep the estimated thresholds for which the value of SSE increases significantly if we remove them. More specifically, for a given set of estimated thresholds $\{-\infty, s_1, s_2, \dots, s_m, +\infty\}$ with $1 \leq m \leq \hat{m}$, and for j-th estimated threshold s_j, denote by $\mathcal{T}_{(s_{j-1}, s_j)} = \left\{ i : s_{j-1} < z_{\pi(i)} \leq s_j \right\}$ the set of orders of z_ts for which their corresponding ordered switching variable $z_{\pi(i)}$s fall into the interval $[s_{j-1}, s_j]$. Now, given a fixed number of thresholds m, we obtain the estimator $\hat{\boldsymbol{\theta}}_{s_1,s_2,\ldots,s_m}$ of $\boldsymbol{\theta}_{s_1,s_2,\ldots,s_m}$ by minimizing the following penalized regression problem

$$\sum_{j=1}^{m+1} \frac{1}{\left|\mathcal{T}_{(s_{j-1},s_j)}\right|} \sum_{i\in\mathcal{T}_{(s_{j-1},s_j)}} \left\|\boldsymbol{x}_{\pi(i)} - \boldsymbol{\theta}_{(s_{j-1},s_j)}\boldsymbol{Y}_{\pi(i)}\right\|_2^2 + \eta_{(s_{j-1},s_j)}\|\boldsymbol{\theta}_{(s_{j-1},s_j)}\|_1, \tag{7}$$

where $\boldsymbol{Y}_{\pi(i)} = \begin{pmatrix}\boldsymbol{x}'_{\pi(i)-1} & \cdots & \boldsymbol{x}'_{\pi(i)-K}\end{pmatrix}'$, $\boldsymbol{\theta}_{s_1,s_2,\ldots,s_m} = \left(\boldsymbol{\theta}'_{(s_0,s_1)}, \boldsymbol{\theta}'_{(s_1,s_2)}, \ldots, \boldsymbol{\theta}'_{(s_{m-1},s_m)}\right)$, and tuning parameters $\boldsymbol{\eta}_n = \left(\eta_{(-\infty,s_1)}, \ldots, \eta_{(s_m,+\infty)}\right)$. The `glmnet` package [18] readily solves the problem.

Denoting

$$\begin{aligned} L_n(s_1, s_2, \ldots, s_m\,; \boldsymbol{\eta}_n) = &\sum_{j=1}^{m+1} \sum_{i\in\mathcal{T}_{(s_{j-1},s_j)}} \left\|\boldsymbol{x}_{\pi(i)} - \boldsymbol{\theta}_{(s_{j-1},s_j)}\boldsymbol{Y}_{\pi(i)}\right\|_2^2 \\ &+ \sum_{j=1}^{m+1} \eta_{(s_{j-1},s_j)} \left\|\boldsymbol{\theta}_{(s_{j-1},s_j)}\right\|_1, \end{aligned} \tag{8}$$

we construct our information criterion as

$$\mathrm{IC}\left(s_1, s_2, \ldots, s_m\,; \boldsymbol{\eta}_n\right) = L_n(s_1, s_2, \ldots, s_m\,; \boldsymbol{\eta}_n) + m\omega_n, \tag{9}$$

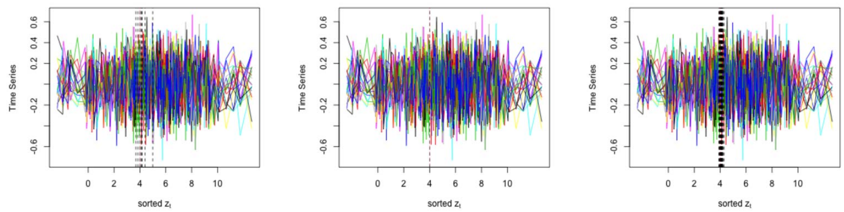
where ω_n is a carefully chosen sequence defined in Section 4. We then select a subset of the initial $\hat{m}$ candidate threshold values by solving

$$(\tilde{m}, \tilde{r}_1, \tilde{r}_1, \ldots, \tilde{r}_{\tilde{m}}) = \mathrm{argmin}_{0\le m\le\hat{m}, \boldsymbol{s}=(s_1,s_2,\ldots,s_m)\in\hat{\mathcal{A}}_n} \mathrm{IC}\left(\boldsymbol{s}; \boldsymbol{\eta}_n\right). \tag{10}$$

Practical choices for tuning parameters $\boldsymbol{\eta}_n$ and ω_n are discussed in Section 5.

The over-estimation of the thresholds and the effect of the screening step are illustrated in Figure 2. The left panel of Figure 2 — which is obtained for one replicate of simulation Scenario 1 in Section 6 — clearly shows that the first step of our procedure detects more threshold values. The middle panel shows that second step successfully screens out the extra threshold estimates and keep a single value which is very close to the true threshold (here, the true threshold value is 4). The right panel of Figure 2 confirms that the final estimated thresholds across all 200 replicates are indeed close to the true thresholds.

When the number of estimated thresholds selected in Step 1 is large, it might be computationally demanding to find the minimizer of the IC. In such cases, we propose to approximate the optimal thresholds using the backward elimination algorithm (BEA) proposed in [41]. Starting with the set of initial thresholds $\hat{\mathcal{A}}_n$, the algorithm reduces the computational cost by removing one threshold at a time until IC does not reduce any further.



(a) First Step. (b) Second Step. (c) Second Step for 200 replicates.

Fig 2. *Estimated thresholds in Simulation Scenario 1 with hdTAR. On average around 8 points are selected in the first step, and Figure 2a shows the result of one single run in first step. Figure 2b shows the results of final selected threshold estimates for single simulation in Figure 2a, and Figure 2c shows the final selected threshold estimates all 200 simulation runs.*

3.4. Estimation of auto-regressive parameters

Given the estimated thresholds, we simply take each estimated regime $\mathcal{T}_{(\tilde{r}_{j-1},\tilde{r}_j)} = \left\{i : \tilde{r}_{j-1} < z_{\pi(i)} \le \tilde{r}_j\right\}$ with $\tilde{r}_0 = -\infty$ and $\tilde{r}_{\tilde{m}+1} = -\infty$ for $j = 1, \dots, \tilde{m}+1$, and estimate the transition matrices in each regime separately. More specifically, for a fixed $j = 1, \dots, \tilde{m} + 1$ we solve

$$\hat{\boldsymbol{\beta}}^{(.,j)} = \arg\min_{\boldsymbol{\beta}} \left(\sum_{i \in \mathcal{T}_{(\tilde{r}_{j-1},\tilde{r}_j)}} \left\| \boldsymbol{x}_{\pi(i)} - \boldsymbol{\beta} \boldsymbol{Y}_{\pi(i)} \right\|_2^2 + \alpha_j \left\| \boldsymbol{\beta} \right\|_1 \right), \tag{11}$$

where α_j is the tuning parameter for the j-th regime for $j = 1, 2, \dots, \tilde{m}$. It can be solved efficiently using existing software and HBIC can be used to select α_j. As an alternative to the separate estimation in (11), if the distances between consecutive threshold values are of the same order, the auto-regressive parameters can also be jointly estimated [40].

4. Theoretical properties

In this section, we establish the consistency of our procedure proposed in Section 3.2. Recall that in the first step of our procedure we use either the ℓ_2 or the ℓ_1 penalty in Equation (5), corresponding to classical (mvTAR) and flexible (hdTAR) TAR models. More specifically, in the following, $\hat{\boldsymbol{\Theta}}$ is the estimator defined in Equation (5) with either the ℓ_1 penalty or the ℓ_2 penalty, $\hat{\boldsymbol{\theta}}_{s_1,s_2,\dots,s_m}$ is the estimator defined in Equation (7), and finally, $\hat{\boldsymbol{\beta}}^{(.,j)}$ is the estimator defined in Equation (11). We make the following assumptions.

Assumption A1. $\{\boldsymbol{\epsilon}_t\}$ *is a sequence of i.i.d. sub-Weibull random variables with bounded continuous and positive density and sub-Weibull constant* K_ϵ *and sub-Weibull parameter* $\varkappa_c > 0$*; specifically, there exist constants* K_ϵ *and* $\varkappa_c > 0$ *such that* $\|\boldsymbol{\epsilon}_t\|_\psi \le K_\epsilon$ *where* $\|\boldsymbol{\epsilon}_t\|_\psi := \sup_{c \ge 1 \wedge \varkappa_c} c^{-\frac{1}{\varkappa_c}} \left(\mathbb{E}\left|\boldsymbol{\epsilon}_t\right|^c\right)^{1/c}$.

Assumption A2. *For each* $j = 1, 2, \ldots, m_0 + 1$*, the process*

$$\boldsymbol{x}_t = \sum_{k=1}^{K} \boldsymbol{A}^{(k,j)} \boldsymbol{x}_{t-k} + \boldsymbol{\epsilon}_t$$

is sub-Weibull with sub-Weibull parameter $\varkappa_1 > 0$ *and* β*-mixing stationary with a geometrically decaying mixing coefficient* b_n*; specifically, there exist constants* $c_b > 0$ *and* $\varkappa_2 > 0$ *such that for all* $n \in \mathbb{N}$, $b(n) \leq \exp(-c_b n^{\varkappa_2})$ *and for all* $t, \tau > 0$, $(\boldsymbol{x}_t, \ldots, \boldsymbol{x}_{t+n}) \overset{d}{=} (\boldsymbol{x}_{t+\tau}, \ldots \boldsymbol{x}_{t+\tau+n})$*, where* $\overset{d}{=}$ *denotes equality in distribution. Moreover,* $\mathbb{E}[\boldsymbol{x}_t] = \boldsymbol{0}_{p\times 1}$*. In addition, assume* $2/3 \leq \varkappa_0 < 1$*, where* $\varkappa_0 := \left(\frac{2}{\varkappa_1} + \frac{1}{\varkappa_2}\right)^{-1}$.

Assumption A3. *The matrices* $\boldsymbol{A}^{(.,j)}$ *are sparse for* $j = 1, \ldots, m_0 + 1$*. More specifically, for all* $h = 1, 2, \ldots p$ *and* $j = 1, 2, \ldots m_0 + 1, d_{hj} \ll p$*, i.e.,* $d_{kj}/p = o(1)$*. Moreover, there exists a positive constant* $M_A > 0$ *such that*

$$\max_{1 \leq j \leq m_0+1} \left\| \boldsymbol{A}^{(.,j)} \right\|_\infty \leq M_A.$$

Assumption A4. *There exists a positive constant* ν *such that*

$$\min_{1 \leq j \leq m_0} \left\| \boldsymbol{A}^{(.,j+1)} - \boldsymbol{A}^{(.,j)} \right\|_2 \geq \nu > 0.$$

Moreover, there exist constants l *and* u *such that* $r_j \in [l, u]$ *for* $1 \leq j \leq m_0$*. In addition, there exists a vanishing positive sequence* γ_n *such that as* $n \to \infty$, $\min_{1\leq j\leq m_0+1} |r_j - r_{j-1}| / \gamma_n \to +\infty$*. For hdTAR, we assume* $d_n^* \frac{\log(p^2 K)}{\sqrt{n\gamma_n}} \to 0$*, whereas for mvTAR we assume* $\sqrt{p^2 K} d_n^* \frac{\log(p^2 K)}{\sqrt{n\gamma_n}} \to 0$.

Assumption A5. $\{z_t\}$ *is a* β*-mixing stationary process with a geometric decaying mixing coefficient and positive density. In addition,* $\mathbb{E}|z_t|^{2+\iota} < \infty$ *for* $\iota > 0$.

The above assumptions are natural in high-dimensional settings and commonly used in the literature. Assumptions A1 and A2 are utilized to derive appropriate concentration inequalities needed to verify the asymptotic properties of the proposed methodology and have been used in the literature [29, 54]. The sub-Weibull distribution of error terms controls the tail effects, while the β-mixing condition ensures the dependence structure can be controlled appropriately. The latter is specifically needed due to the temporal correlation among observations. We can relax the β-mixing assumption to α-mixing if we restrict to Gaussian distributions, rather than sub-Weibull processes. However, to keep the distributional assumption more general, we consider here the β-mixing assumption. In Appendix 4, we also develop a sufficient condition for β-mixing processes by imposing constrains on the operator norm of transition matrices; this implies that the β-mixing condition is less restrictive. The assumption $\varkappa_0 \geq 2/3$ is to ensure a sharp consistency rate for estimating the thresholds and can be removed at the cost of worsening the consistency rate (see additional details in Remark 1). Assumption A3 ensures the sparsity of the model and is needed to quantify the effect of model misspecification, since exact recovery of threshold values is not possible. A similar assumption has been used in [41] in the context of change point detection. Further, Assumption A4 puts a minimum jump size on the transition matrices ensuring a detectable change occurred at threshold r_j; it also puts certain conditions on the detection rate, which is related to γ_n. Assumption A4 can be seen as an extension of Assumption H4 in [10] to high-dimensions. It can be seen that the assumption is more stringent for mvTAR, rendering this procedure not suitable for high dimensions. Finally, Assumption A5 is used to build the relationship between the length of each regime and the number of observations in that regime.

Our first theoretical result concerns the first step, i.e., the initial estimation of thresholds using group or regular fused lasso penalties. The penalized estimation (5) in this step does not guarantee parameter estimation consistency since the design matrix $\boldsymbol{Z}$ in Equation (4) may not satisfy the restricted eigenvalue condition [2], which is critical for establishing the parameter estimation consistency in high-dimensions [4]. However, with either penalty, the estimator over-estimates the true number of thresholds, as established next.

Let $\mathcal{A}_n = \{r_1, r_2, \dots, r_{m_0}\}$ be the set of the sorted true thresholds. Define the Hausdorff distance between two countable sets as:

$$d_H(A, B) = \max_{b \in B} \min_{a \in A} |b - a|.$$

Though not a distance, $d_H(A, B)$ proves useful in Theorem 1.

Theorem 1. *Under assumptions A1 to A5, there exist large constants* $C_1, C_2 > 0$ *such that* $\tilde{\lambda}_{1,n} = C_1 \frac{\log(p^2 K)}{\sqrt{n}}$*, and* $\tilde{\lambda}_{2,n} = \frac{C_2}{n} \frac{\log(p^2 K)}{\sqrt{n \gamma_n}}$*, where for hdTAR* $\lambda_{1,n} = \tilde{\lambda}_{1,n}$ *and* $\lambda_{2,n} = \tilde{\lambda}_{2,n}$*, whereas for mvTAR,* $\lambda_{1,n} = \sqrt{p^2 K}\tilde{\lambda}_{1,n}$ *and* $\lambda_{2,n} = \sqrt{p^2 K}\tilde{\lambda}_{2,n}$*. Then,*

$$\min\left\{\mathbb{P}\left(|\hat{\mathcal{A}}_n| \geq m_0\right), \mathbb{P}\left(d_H\left(\mathcal{A}_n, \hat{\mathcal{A}}_n\right) \leq \gamma_n\right)\right\} \to 1.$$

Theorem 1 shows that the number of estimated thresholds $\hat{m}$ in Step 1 is no less than the true number of thresholds m_0 with high probability. In addition, there exists at least one estimated threshold in the γ_n-radius neighborhood of the true thresholds. The rate of consistency for threshold detection, γ_n, depends on the number of time series p, the maximum considered lag K, and the minimum distance between consecutive true thresholds in the model. In addition, the convergence rate for using $\hat{r}_j$ to estimate r_j could be as low as $\log\log n \left(\log\left(p^2 K\right)\right)^2 / n$ when m_0 is finite.

The rate of consistency for thresholds detection, γ_n, for mvTAR also depends on the number of time series p, the maximum considered lag K, and the minimum distance between consecutive true thresholds in the model. However, the assumptions on γ_n for hdTAR and mvTAR are different, so the consistency rate for thresholds detection is different for these two methods. In addition, when using the ℓ_2 penalty, the convergence rate for using $\hat{r}_j$ to estimate r_j could be as low as $\log\log n\left(\log\left(p^2K\right)\right)^2 p^2K/n$ when m_0 is finite. Thus, convergence of mvTAR is only guaranteed in low to moderate dimensions and not in high dimensions. Finally, the minimum sample size requirement depends on the sub-Weibull parameter $\varkappa_1$ and β-mixing parameter $\varkappa_2$. For example, as mentioned in Lemma 5, we need $n \geq c_0\left(\log(p^2K)\right)^{2/\varkappa_0-1}$ where $\varkappa_0 := \left(\frac{2}{\varkappa_1}+\frac{1}{\varkappa_2}\right)^{-1}$. This indicates that if the sub-Weibull parameter $\varkappa_1$ increases (i.e., the tail probability decays faster), the minimum sample size will decrease; similarly, the minimum sample size decreases as the β-mixing parameter $\varkappa_2$ increases.

Next, we state Theorem 2 which shows the screening procedure (10) consistently estimates the number and values of thresholds. For that, we need two additional assumptions.

Assumption A6. *Let* $\Delta_n = \min_{1\leq j\leq m_0+1}|r_j - r_{j-1}|$. *Then,*

$$m_0\left(n\gamma_n\right)^{3/2} d_n^{*2}/\omega_n \to 0, \text{ and } n\Delta_n/(m_0\omega_n) \to +\infty.$$

Assumption A7. *There exist positive constants* c, c_1, c_2 *and* c_3 *such that for indexes* j' *and* $j'-1$ *and corresponding estimated thresholds* $s_{j'}$ *and* $s_{j'-1}$*,*

(a) if $|s_{j'} - s_{j'-1}| \leq \gamma_n$*, then* $\eta_{(s_{j'-1},s_{j'})} = c\sqrt{n\gamma_n}\log\left(p^2K\right)$*;*

(b) if there exist r_j *and* r_{j+1} *such that* $|s_{j'-1} - r_j| \leq \gamma_n$ *and* $|s_{j'} - r_{j+1}| \leq \gamma_n$*, then,*

$$\eta_{(s_{j'-1},s_{j'})} = \frac{2}{c_3}\left(c_1\frac{\log(p^2K)}{\sqrt{n(s_{j'}-s_{j'-1})}} + c_2M_Ad_n^*\frac{\gamma_n}{s_{j'}-s_{j'-1}}\right);$$

(c) otherwise $\eta_{(s_{j'-1},s_{j'})} = \frac{2}{c_3}\left(c_1\frac{\log(p^2K)}{\sqrt{n(s_{j'}-s_{j'-1})}} + c_2M_Ad_n^*\right)$.

Assumption A6 makes a unique connection between three important quantities: (1) minimum spacing between consecutive thresholds, Δ_n; (2) the consistency rate for estimating the threshold values, γ_n; (3) the penalty term in the definition of the information criterion, ω_n. This connection helps with quantifying the consistency rate for estimating the threshold values as discussed after Theorem 2.

Assumption A7 specifies three different tuning parameter rates for the screening step. Although this assumption may seem technical, but it is needed to get the sharpest consistency rate. It is possible to define a fixed tuning parameter for all cases in Assumption A7, but the consistency results will be worsened. Remark 5 in [41] shed some light into this issue.

Theorem 2. *Under Assumptions A1 to A7, if* $n \to +\infty$*, the minimizer*

$$(\tilde{m}, \tilde{r}_j, j = 1, 2, \ldots, \tilde{m})$$

of Equation (10) *satisfies:*

$$\mathbb{P}(\tilde{m} = m_0) \to 1. \tag{12}$$

In addition, there exists a constant $B > 0$ such that:

$$\mathbb{P}\left(\max_{1\le j\le m_0} |\tilde{r}_j - r_j| \le Bm_0 (\gamma_n)^{3/2} d_n^{*2}\sqrt{n}\right) \to 1. \tag{13}$$

When $p = cn^\kappa$, where $c > 0$ and $\kappa \in (0, 1)$, the proposed procedure for both hdTAR and mvTAR can also be applied to low-dimensional time series. The consistency results would be similar to those in Theorem 2. It is challenging to select ηs in practice, since the distance between estimated thresholds to the true thresholds is unknown. Instead, we set ηs to be the same and apply BIC/HBIC to select them.

Although the consistency rates for mvTAR and hdTAR are both functions of γ_n, the assumptions on γ_n for the two methods are different, leading to different rates of consistency. To illustrate this point, consider the case when m_0 is finite. Then, when using the ℓ_1 penalty in the first step, we can set $\gamma_n = (\log n)^\rho \left(\log\left(p^2K\right)\right)^{2+2\rho}/n$ for some $\rho > 0$. With this rate, the hdTAR model can have total sparsity $d_n^* = o\left(\left(\log n\left(\log\left(p^2K\right)\right)^2\right)^{\rho/2}\right)$. The consistency rate then becomes of order $\left((\log n)^{\frac{5}{2}\rho}\left(\log\left(p^2K\right)\right)^{3+5\rho}\right)/n$. In comparison, when using the ℓ_2 penalty, we can set $\gamma_n = (\log n)^{\rho'}\left(\log\left(p^2K\right)\right)^{2+2\rho'}\left(p^2K\right)^{1+\rho'}/n$ for some $0 < \rho' < 1$ to ensure that Assumption A3 is satisfied. With this rate, the mvTAR model can have total sparsity $d_n^* = o\left(\left(p^2K\log n\left(\log\left(p^2K\right)\right)^2\right)^{\rho'/2}\right)$. Using a similar calculation, the consistency rate for mvTAR becomes of order $\left((\log n)^{\frac{5}{2}\rho'}\left(\log\left(p^2K\right)\right)^{3+5\rho'}\left(p^2K\right)^{\frac{3}{2}+\frac{5}{2}\rho'}\right)/n$, further highlighting that mvTAR is not suitable in high dimensions, when $p = cn^\kappa$, where $c > 0$ and $\kappa \ge 1$.

Remark 1. *If we remove the assumption $\varkappa_0 \ge 2/3$ and only keep $\varkappa_0 < 1$, then, according to Lemma 7, the choice of γ_n would also depend on $\varkappa_0$. For the hdTAR model, we can set $\gamma_n = (\log n)^\rho\left(\log\left(p^2K\right)\right)^{2/\varkappa_0-1+2\rho}/n$ for some $\rho > 0$, and keep the total sparsity the same as above. The consistency rate then becomes of order $\left((\log n)^{\frac{5}{2}\rho}\left(\log\left(p^2K\right)\right)^{3/\varkappa_0-3/2+5\rho}\right)/n$. Similarly, for mvTAR model, we can set $\gamma_n = (\log n)^{\rho'}\left(\log\left(p^2K\right)\right)^{2/\varkappa_0-1+2\rho'}\left(p^2K\right)^{1+\rho'}/n$ for some $0 < \rho' < 1$, and keep the total sparsity the same as above. The consistency rate for mvTAR becomes of order $\left((\log n)^{\frac{5}{2}\rho'}\left(\log\left(p^2K\right)\right)^{3/\varkappa_0-3/2+5\rho'}\left(p^2K\right)^{\frac{3}{2}+\frac{5}{2}\rho'}\right)/n$.*

Our last theorem establishes the consistent estimation of regime-specific transition matrices in the third step.

Theorem 3. *Under Assumptions A1 to A7, and selecting*

$$\alpha_j = C\sqrt{\log(p^2K)/(n\gamma_n)}$$

for some large enough $C > 0$, with high probability approach to 1, there exists a positive constant C' such that we have for each fixed regime j:

$$\left\|\hat{\boldsymbol{\beta}}^{(.,j)} - \boldsymbol{A}^{(.,j)}\right\|_2 \le C'\sqrt{d_n^*\log(p^2K)/(n\gamma_n)}. \tag{14}$$

The consistency rate derived in Theorem 3 is similar to that of Wong et al. [54], Basu and Michailidis [2] for high-dimensional vector auto-regressive models.

5. Tuning parameter selection

We next provide guidance on selecting the tuning parameters for our three-step procedure.

$\lambda_{1,n}$ We choose $\lambda_{1,n}$ by cross-validation. We first randomly choose the order of switching variable z_t with equal space. Let $\mathbb{T}$ be a set of time points corresponding to selected switching variable. We use the rest of observations to estimate $\boldsymbol{\Theta}$ in the first step for a range of $\lambda_{1,n}$. To choose the optimal value of $\lambda_{1,n}$, we use the estimated $\boldsymbol{\Theta}$ to predict the series at time points in $\mathbb{T}$. The optimal $\lambda_{1,n}$ is selected as the value corresponding to the minimum mean squared prediction error over $\mathbb{T}$.

$\lambda_{2,n}$ The rate for $\lambda_{2,n}$ vanishes fast as n increases. Thus, to lower the computational cost, we set $\lambda_{2,n}$ to zero. It is possible to select $\lambda_{2,n}$ using cross-validation as well at the cost of increasing computation time. However, the sensitivity analysis reported in [41] indicates that setting $\lambda_{2,n} = 0$ is a reasonable choice.

$\boldsymbol{\eta}_n$ Selecting η_n is in general difficult. For $0 \leq m \leq \hat{m}$ (m̂ is the number of estimated thresholds in step 1), we choose different ηs for different regimes, and use HBIC and eBIC [52] across all regimes. For each time series l, $l \in 1, 2, \ldots, p$, and $j = 1, 2, \ldots, m+1$, set η_j^l as the tuning parameter for l-th time series at j-th regime. Then, the HBIC for interval $[s_{j-1}, s_j]$ is defined as

$$\text{HBIC}\left(j, \eta_j^l\right) = \log\left(\text{SSE}_{l,j} / \left|\mathcal{T}_{(s_j - s_{j-1})}\right|\right) + \frac{\gamma_1 \left\|\hat{\boldsymbol{\theta}}^l_{s_{j-1},s_j}\right\|_0}{\left|\mathcal{T}_{(s_j - s_{j-1})}\right|} \log\left(pK\right),$$

where $\gamma_1 = 2.8$ that is within the recommended range in [52]. Similarly, the eBIC for interval $[s_{j-1}, s_j]$ is defined as

$$\text{eBIC}\left(j, \eta_j^l\right) = \log\left(\text{SSE}_{l,j} / \left|\mathcal{T}_{(s_j - s_{j-1})}\right|\right) + \frac{\gamma_2 \left\|\hat{\boldsymbol{\theta}}^l_{s_{j-1},s_j}\right\|_0}{\left|\mathcal{T}_{(s_j - s_{j-1})}\right|} \left(\log\left(pK\right) + \log\left(\left|\mathcal{T}_{(s_j - s_{j-1})}\right|\right)\right),$$

where $\gamma_2 = 1.4$ that is within the recommended range in [52] as well. If $\left|\mathcal{T}_{(s_j - s_{j-1})}\right| \geq pK$, η_j^l is selected as:

$$\hat{\eta}_j^l = \arg\min_{\eta_j^l} \text{eBIC}\left(j, \eta_j^l\right). \tag{15}$$

If $\left|\mathcal{T}_{(s_j - s_{j-1})}\right| < pK$, η_j^l is selected as:

$$\hat{\eta}_j^l = \arg\min_{\eta_j^l} \text{HBIC}\left(j, \eta_j^l\right). \tag{16}$$

ω_n We first perform the backward elimination algorithm (BEA) until no break points are left. Then, we cluster the differences in the objective function L_n into two subgroups, small and large. If removing a threshold only leads to a small decrease in L_n, then the removed threshold is likely redundant. In contrast, true thresholds lead to larger decrease. We choose the smallest decrease in the second group as the optimal value of ω_n. To this end, we first calculate the minimum sum of squared error for removing all thresholds in $\hat{\mathcal{A}}_n$ one by one, denoted as $L'_0, L'_1, \ldots, L'_{\hat{m}}$. Then, ω_n is selected as the maximum values among $L'_{j+1} - L'_j$ for $j = 0, 1, \ldots, \hat{m} - 1$.

α_i For simplicity, we let all time series share the same α_i, denoted by α_n. For low to moderate dimensions, the tuning parameter α_n for parameter estimation is selected as the minimizer of the combined HBIC over all regimes. For $j = 1, 2, \ldots, \tilde{m} + 1$, we define the HBIC on interval $[\tilde{r}_{j-1}, \tilde{r}_j]$ as:

$$\text{HBIC}\left(j, \alpha_n\right) = \log\left(\det \hat{\boldsymbol{\Sigma}}_{\epsilon,j}\right) + \frac{\gamma \left\|\hat{\boldsymbol{\beta}}^{(\cdot,j)}\right\|_0}{\left|\mathcal{T}_{(\tilde{r}_{j-1},\tilde{r}_j)}\right|} \log\left(p^2 K\right),$$

where $\hat{\boldsymbol{\Sigma}}_{\epsilon,j}$ is the residual sample covariance matrix with $\hat{\boldsymbol{\beta}}$ estimated in Equation (11) and $\gamma = 2.8$. For high dimensions, we choose α_n by 10-fold cross validation.

6. Empirical evaluations

In this section, we present simulation results evaluating the performance of the proposed procedure in both moderate dimensions and high dimensions; the first four simulations scenarios presented are moderate-dimensional, while the last one is high-dimensional. Details of simulation settings are presented in Appendix 7. All results are averaged over 200 replicates.

We compare our method with Tsay [48], Li and Tong [28], and the threshold vectorized auto-regressive method [33]. These methods, which are denoted as Tsay (1998), Li (2016) and TVAR, respectively, assume a known number of thresholds or at least a known upper bound on the number of thresholds when establishing the asymptotic properties of their estimators. In practice, Tsay [48] proposes to perform a grid search to select the number of thresholds, when unknown, by minimizing AIC. They are also restricted to low dimensions. For instance, TVAR estimates the number of thresholds and the values of thresholds using two separate steps and assumes the number of thresholds is at most 2. This is in contrast to our developed mvTAR and hdTAR methods, which do not make any assumptions about the number of thresholds. Though Calderón V and Nieto [7] and Orjuela and Villanueva [37] do not require a known number of thresholds, we did not include a comparison with these methods, as they cannot handle larger dimensions, e.g., $p = 20$.

We compare the estimated thresholds and the percentage of simulations where thresholds are correctly estimated; this is defined as cases where the selected thresholds are close to the true thresholds. More specifically, a selected threshold is considered as close to the first true threshold, z_1, if it is in the interval $[-\infty, z_1 + 0.5(z_2 - z_1))$; similarly, a selected threshold is considered as close to the second true threshold, z_2, if it falls in the interval $[z_1 + 0.5(z_2 - z_1), \infty]$. Note that the number of thresholds is set to be known for Tsay (1998), Li (2016) and TVAR, since the first two require a known number of thresholds and TVAR does not perform well in selecting the number of threshold in its first step (Note that when the number of thresholds is not provided, TVAR's rates of correctly identifying the correct number of thresholds are 84%, 87%, 65%, and 16% (11.5% for $T = 300$) in the first four scenarios and the method is not applicable in the last scenario (scenario 5) due to high-dimensionality of model.)

6.1. Simulation results

We next compare the performances of the proposed hdTAR and mvTAR methods with Tsay (1998), Li (2016) and TVAR. Here, the selection rate of Tsay (1998), Li (2016) and TVAR is based on whether the estimated thresholds are within one standard deviation of true threshold.

Table 1 summarizes the results of threshold estimation. In all simulations, if any of the methods does not select a thresholds, we set the minimum value of the threshold variable as the selected threshold. The results indicate that Tsay (1998) and Li (2011) do not work well even for the first three scenarios, while hdTAR, mvTAR, and TVAR perform well in the first three scenarios; however, the estimation error and standard deviation of TVAR are larger than those of hdTAR and mvTAR. In Scenario 4, in which only a portion of time series components change at threshold values, the detection rate for both mvTAR and TVAR drops significantly, while hdTAR still achieves 100% threshold detection rate. This is expected since hdTAR is more flexible and mvTAR only works well for scenarios in which auto-regressive components change at the same threshold values. In Scenario 4, mvTAR tends to choose a large λ_1 which leads to selecting smaller number of threshold values in the first step than needed. Nonetheless, when the changes in the transition matrices are large enough, the threshold values can still be detected using the ℓ_2 penalty. Finally, hdTAR continues to offer excellent threshold detection in the high-dimensional setting of Scenario 5; in contrast, the other methods are not well suited for this scenario and are not included.

Table 2 summarizes the performance of the five methods in terms of auto-regressive parameter estimation. Since Tsay (1998) does not provide coefficients estimates, so we use the method in our Step 3 to estimate the parameters given the thresholds obtained by Tsay (1998). The results indicate that both hdTAR and mvTAR perform well in the first three scenarios, as measured by their high true positive rates and low false positive rates. Since TVAR does not perform variable selection, all estimated values of transition matrices using this method are non-zero. This leads to true positive and false positive rates that are both equal to 1, which are not meaningful and are hence excluded from the table.

The results also indicate that in Scenario 4 with $T = 600$ and Scenario 5 hdTAR performs satisfactorily, while in Scenario 4 with $T = 300$, its FPR

Table 1
Mean and standard deviation of estimated thresholds, the percentage of simulation runs where thresholds are correctly detected (selection rate) in different simulation scenarios. If the estimated thresholds is within one standard deviation of the true threshold, we consider the estimated thresholds as correctly detected.

	Threshold(s)	Methods	Mean	Std	Selection Rate
Scenario 1		hdTAR	4.05	0.05	1.00
		mvTAR	4.04	0.05	1.00
	4	TVAR	4.23	1.13	0.95
		Tsay (1998)	5.36	2.01	0.57
		Li (2016)	7.66	0.34	0.00
Scenario 2		hdTAR	4.05	0.06	1.00
		mvTAR	4.04	0.05	1.00
	4	TVAR	4.15	1.17	0.93
		Tsay (1998)	5.46	2.00	0.56
		Li (2016)	7.66	0.34	0.00
Scenario 3		hdTAR	4.04	0.06	1.00
		mvTAR	4.04	0.05	1.00
	4	TVAR	4.15	1.17	0.93
		Tsay (1998)	7.63	0.76	0.03
		Li (2016)	7.66	0.34	0.00
Scenario 4		hdTAR	4.00	0.15	1.00
(T = 600)	4	mvTAR	2.44	1.24	0.93
		TVAR	3.82	1.34	0.85
		hdTAR	6.02	0.09	1.00
	6	mvTAR	5.30	1.37	0.82
		TVAR	6.11	1.28	0.88
Scenario 4		hdTAR	4.03	0.47	1.00
(T = 300)	4	mvTAR	2.52	0.91	0.67
		TVAR	3.92	1.48	0.81
		hdTAR	6.00	0.31	1.00
	6	mvTAR	4.78	1.16	0.42
		TVAR	6.19	1.42	0.84
Scenario 5		hdTAR	5.06	0.29	1.00
	5	mvTAR	–	–	–
		TVAR	–	–	–

increases to around 20%. This is primarily due to the smaller sample size in this scenario for each of the three regimes. Recall from Table 1 that the selection rate of mvTAR in both of these scenarios was very low; as a result, in many simulation replicates there were fewer number estimated regimes than needed to obtain estimates of auto-regressive parameters. As a result, mvTAR is not included in the comparisons for Scenarios 4 and 5. These findings underscore the advantages of hdVAR in settings with complex patterns of changes in auto-regressive parameters as well as in high dimensions.

Box plots summarizing the results in Table 1 are presented in Figure 3.

7. Real data application

We demonstrate the utility of our penalized estimation framework in financial econometric applications by analyzing a bank balance sheet data. The data

TABLE 2
Results of parameter estimation for simulation scenarios. The table shows mean and standard deviation of relative estimation error (REE), true positive rate (TPR), and false positive rate (FPR) for estimated coefficients.

	Method	REE	SD(REE)	FPR	TPR
Scenario 1	hdTAR	0.31	0.04	0.03	0.95
	mvTAR	0.32	0.04	0.03	0.95
	TVAR	0.85	0.19	–	–
	Tsay (1998)	0.70	0.30	0.04	0.57
	Li (2016)	1.50	0.44	1.00	1.00
Scenario 2	hdTAR	0.31	0.04	0.03	0.95
	mvTAR	0.31	0.04	0.03	0.94
	TVAR	0.89	0.43	–	–
	Tsay (1998)	0.69	0.31	0.04	0.55
	Li (2016)	1.49	0.55	1.00	1.00
Scenario 3	hdTAR	0.34	0.04	0.04	0.89
	mvTAR	0.34	0.04	0.04	0.89
	TVAR	0.69	0.65	–	–
	Tsay (1998)	0.88	0.05	0.03	0.36
	Li (2016)	1.33	0.64	1.00	1.00
Scenario 4 (T = 600)	hdTAR	0.5	0.05	0.02	0.77
	mvTAR	–	–	–	–
	TVAR	0.67	0.07	–	–
Scenario 4 (T = 300)	hdTAR	0.77	0.09	0.19	0.71
	mvTAR	–	–	–	–
	TVAR	0.87	0.15	–	–
Scenario 5	hdTAR	0.80	0.04	0.51	0.86
	mvTAR	–	–	–	–
	TVAR	–	–	–	–

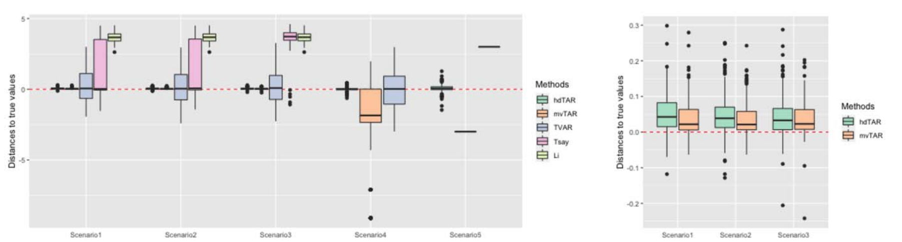


FIG 3. *Box plot of distances between the estimated final points and true values. The left panel shows the results for all the five scenarios with all the five models. The right panel zooms in the results in the first three scenarios using hdTAR and mvTAR.*

consists of total balances of the top 10 largest US banks over time, each measured in thousands of dollars (available from www.fdic.gov).

To assess the relationship between the state of the banking sector and the overall economic conditions, we fit a multivariate TAR model of the quarterly bank balance sheet data over the period of 1995 to 2018 with the growth rate of the US GDP as the switching variable. For the quarterly GDP data $y_t; t = 1, 2, \ldots, T$ over T observations, the growth rate is defined as

$$z_t = 100(\log y_t - \log y_{t-1}), \quad t = 2, 3, \ldots, T.$$

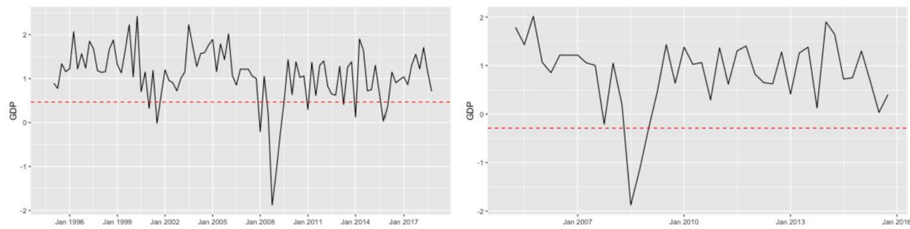


FIG 4. *The GDP growth rate and detected thresholds using data from ten top banks. The red dash line shows the estimated threshold. The left panel shows the GDP growth rate and detected thresholds based on data from 1995 to 2018, while the right panel shows the GDP growth rate and detected thresholds based on data from 2005 to 2015. In both cases, the proposed method divides economic patterns into only two conditions — recession and non-recession.*

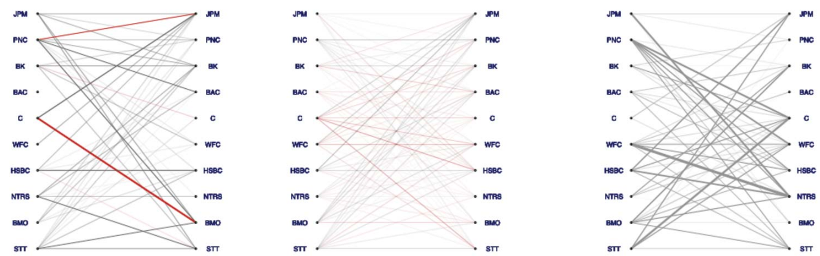
To reduce the non-stationarity, the bank balance sheet data $v_t; t = 1, 2, \dots, T$, is also transformed as

$$x_t = \log v_t - \log v_{t-1}, \quad t = 2, 3, \dots, T.$$

We applied the hdTAR on the entire time series consisting of $T = 98$ quarterly observations from 1995 to 2018. To examine how results change with smaller sample sizes, we also analyze the shorter time period of quarterly observations from 2005 to 2015. The detected threshold for both time periods are shown in Figure 4. Although hdTAR does not enforce the coefficients to change at the same threshold value, irrespective of the sample size it identifies a single threshold corresponding to the great recession of 2008. This further highlights the flexibility and adaptability of hdTAR for both moderate- and high-dimensional TAR models. As a comparison, we also applied the mvTAR to the same two data sets, but exclude the results due to the inconsistency in the estimated thresholds using mvTAR when applied to the same two data sets.

The Granger causal networks [3] of interactions among these ten banks in both recession and non-recession periods during 1995–2018 are shown in Figure 5. The red links in each panel represent the interactions that occur in that economic period only. The results show strong interactions between Citibank and Harris Bank and a comparable strong interaction between PNC and JP-Morgan Chase during the recession. The interactions become weaker during the non-recession period, but more interactions appear among banks. A similar observation was made in [31].

We only plot the estimated network structures during non-recession period from 2005–2015. This is because the detected threshold is very close to the lower boundary of the sorted values of the switching variable, resulting in very few observations in the recession regime. From Figure 5, the interactions among banks in non-recession period from 2005 to 2015 are similar to the structures detected using full data set. This further confirms the satisfactory performance of hdTAR in both larger data and smaller data sets.



(a) (b)

FIG 5. *The Granger causality graph for the top ten banks across time. Each vertex represents a bank, and the links display directed interactions between banks. Panel (a) corresponds to the longer time series (1995–2018) and panel (b) corresponds to the shorter time series (2005–2015). The left figure in panel (a) shows the interactions during the recession; the right figure shows the interactions in non-recession. The red links in each panel represent the interactions that occur in that economic period only. Panel (b) only show the interactions among banks identified in non-recession period from the shorter time series. Given the very small number of observations in the recession period in the shorter time series, the Granger causality graph for this period is not estimated.*

8. Discussion

We developed a three-step algorithm to estimate the number and values of thresholds, as well as the auto-regressive parameters in possibly high-dimensional TAR model. The proposed algorithm can utilize either an ℓ_2 or an ℓ_1 penalty, or more specifically, a grouped or regular fused lasso penalty. The ℓ_2 penalty corresponds to the natural extension of the original multivariate TAR model in which all coefficients are forced to change at the same thresholds. The ℓ_1 penalty, in contrast, is more flexible allowing each coefficient to potentially change at different thresholds. Although this flexibility potentially comes at the cost of a larger number of thresholds in the TAR model, our theoretical and empirical results indicate that mvTAR is not appropriate for high-dimensional settings and is better suited for moderate dimensions. In contrast, the more flexible hdTAR leads to consistent estimation and superior empirical performance in both moderate and high dimensions.

We established that both versions of our algorithm, termed mvTAR and hdTAR, consistently estimate the model parameters under natural conditions on the distribution and on the level of temporal correlations in the model. The consistency rates for both models depend explicitly on several model characteristics. Specifically, when the total number of thresholds, m_0, is finite, the rate of consistency for detecting the thresholds is based on: (1) the effective number of time points, n, (2) the number of time series components, p, (3) the number of lags, K, and (4) the total sparsity of the model, d_n^*. For mvTAR, if we set $d_n^* = o\left(\left(\log n \left(\log\left(p^2K\right)\right)^2 p^2K\right)^{\rho'/2}\right)$ for small $0 < \rho' < 1$, then the consistency rate becomes of order

$\left((\log n)^{\frac{5}{2}\rho'}\left(\log\left(p^2K\right)\right)^{3+5\rho'}\left(p^2K\right)^{\frac{3}{2}+\frac{5}{2}\rho'}\right)/n$. This confirms that mvTAR is suitable for moderate dimension but may not work in high dimensions. In contrast, for hdTAR, setting $d_n^* = o\left(\left(\log n\left(\log\left(p^2K\right)\right)^2\right)^{\rho/2}\right)$ for some small positive ρ, the consistency rate becomes of order $\left((\log n)^{\frac{5}{2}\rho}\left(\log\left(p^2K\right)\right)^{3+5\rho}\right)/n$. The first component of the rate, i.e. $(\log n)^{\frac{5}{2}\rho}$, is similar to some existing consistency rates for univariate TAR models [10] while the additional term $\left(\log\left(p^2K\right)\right)^{3+5\rho}$ quantifies the difficulty in estimating the thresholds in high-dimensions.

A limitation of the proposed procedure is that it requires several hyperparameters, especially in the second step. To lower the computational cost, we chose similar tuning parameters in the second step according to eBIC/ HBIC. However, regime-specific tuning parameters may improve the estimation performance in finite samples. Fast selection of regime-specific tuning parameters is an interesting future research direction. Identifying the switching variable in the TAR model is another challenge, specifically in applications. For example, in the bank data, we selected the GDP as the switching variable. However, it is not obvious whether this is an optimal choice; for example, the unemployment rate or the inflation rate could also serve as the switching variable. Selecting optimal (data-driven) switching variable is another fruitful future research direction.

Appendix

Notations. We first describe some notations which will be used across all proofs. For a symmetric matrix X, let $\lambda_{\max}(X)$ and $\lambda_{\max}(X)$ denote its maximum and minimum eigenvalues and $|||X|||$ denotes its operator norm $\sqrt{\lambda_{\max}(X'X)}$. For any matrix M, if $\{G_1, G_2, \ldots, G_{g_0}\}$ denote a partition of $\{1, 2, \ldots, |vec(M)|\}$ into g_0 non-overlapping groups, then we use $\|M\|_{2,\infty}$ to denote $\max_{g=1,2,\ldots,g_0} \|vec(M)_{G_g}\|_2$ and $\|M\|_{2,1}$ to denote $\sum_{g=1}^{g_0} \|vec(M)_{G_g}\|_2$, where $vec(M)_{G_g}$ represents all the elements of vectorized form M in G_g group. Let $\mathcal{S} = \{w_1, w_2, \ldots, w_{m_0}\}$, where w_j denotes the j-th order of true threshold. Set $m_0 = |\mathcal{S}|$. Let b_j denotes the order of the j-th estimated threshold in Step 2.

Appendix 1: Some definitions

Sub-Weibull random variable: A random variable U is sub-Weibull [39] if there exist constants $K_U > 0$ and $\varkappa' > 0$ such that

$$\|U\|_\psi := \sup_{c \geq 1 \wedge \varkappa'} c^{-\frac{1}{\varkappa'}} \left(\mathbb{E}\,|U|^c\right)^{1/c} \leq K_U; \tag{17}$$

Moreover, K_U is called the sub-Weibull constant while $\varkappa' > 0$ is called the sub-Weibull parameter.

Mixing conditions: We follow the definitions in [5]. Given the probability space $(\Omega, \mathcal{F}, P)$, for any σ-field $\mathcal{A} \subset \mathcal{F}$, define $L_2(\mathcal{A})$ to be the family of all square integrable $\mathcal{A}$-measurable random variables. For any two σ-fields $\mathcal{A}$ and $\mathcal{B} \subset \mathcal{F}$, we define:

$$\alpha'_n = \sup |P(A \cap B) - P(A) P(B)|, \quad A \in \mathcal{A}, B \in \mathcal{B}; \tag{18}$$

$$\beta'_n = \sup \frac{1}{2} \sum_{i_1=1}^{I} \sum_{i_2=1}^{J} |P(A_{i_1} \cap B_{i_2}) - P(A_{i_1}) P(B_{i_2})|, \tag{19}$$

where the supremum is taken over all pairs of (finite) partitions $\{A_1, A_2, \ldots, A_I\}$ and $\{B_1, B_2, \ldots, B_J\}$ of Ω such that $A_{i_1} \in \mathcal{A}$ for each i_1 and $B_{i_2} \in \mathcal{B}$ for each i_2. The stochastic process is said to be α-mixing (strongly mixing) if $\alpha'_n \to 0$, and β-mixing if $\beta'_n \to 0$. Note that β-mixing implies α-mixing.

Appendix 2: Technical lemmas

Lemma 4. *Under Assumptions A1, A2 and A5, for $x \in \mathbb{R}, 1 \le l, l' \le p, 1 \le k \le K$,*

$$x_{((t-k),l)} I(z_t \le x) \epsilon_{(t,l')}$$

is sub-Weibull with parameter $\frac{1}{1/\varkappa_1 + 1/\varkappa_c}$;

$$x_{((t-k),l)} I(z_t \le x) x_{(t,l')}$$

is sub-Weibull with parameter $\varkappa_1/2$.

Proof of Lemma 4: According to Assumptions A1 and A2, we know $x_{((t-k),l)}$ and $\epsilon_{(t,l')}$ are sub-Weibull with sub-Weibull parameter $\varkappa_1$ and $\varkappa_c$. From Proposition 3 in Vladimirova et al. [51], we have $x_{((t-k),l)}\epsilon_{(t,l')}$ is sub-Weibull with parameter $\frac{1}{1/\varkappa_1 + 1/\varkappa_c}$. Similarly, $x_{((t-k),l)} x_{(t,l')}$ is sub-Weibull with parameter $\varkappa_1/2$.

Combined with above statement and based on Theorem 1 in [51], there exists $K_C > 0$ such that for all $y_x \ge 0$, we have:

$$\begin{aligned} \mathbb{P}\left(\left|x_{((t-k),l)} I(z_t \le x) \epsilon_{(t,l')}\right| \ge y_x\right) &\le \mathbb{P}\left(\left|x_{((t-k),l)} \epsilon_{(t,l')}\right| \ge y_x\right) \\ &\le 2 \exp\left(-\left(y_x/K_C\right)^{\left(\frac{\varkappa_1 \varkappa_c}{\varkappa_1 + \varkappa_c}\right)}\right). \end{aligned} \tag{20}$$

By Theorem 1 in [51] again, $x_{((t-k),l)} I(z_t \le x) \epsilon_{(t,l')}$ is sub-Weibull with parameter $\frac{1}{1/\varkappa_1 + 1/\varkappa_c}$. By similar procedure, we can prove $x_{((t-k),l)} I(z_t \le x) x_{(t,l')}$ is sub-Weibull with sub-Weibull parameter $\varkappa_1/2$.

Lemma 5. *Under Assumptions A1 to A4, there exist positive constants C, c_0, c_1, c_2, c_3, such that for*

$$n \ge c_0 \left(\log(p^2 K)\right)^{2/\varkappa_0 - 1},$$

with probability at least $1 - c_3\eta_1 - \eta_2$, *we have:*

$$\frac{1}{n}\|\boldsymbol{Z}'\boldsymbol{E}\|_\infty \le C\frac{\log\left(p^2K\right)}{\sqrt{n}}, \tag{21}$$

where $\eta_1 = \exp\left(-c_1\log\left(p^2K\right)\right)$ *and*

$$\eta_2 = \exp\left(-c_2\frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}} + \log\left(np^2K\right)\right).$$

Proof of Lemma 5: First, we rewrite Equation (21) with respect to the switching variable z_t and t as:

$$\max_{1\le i\le n, 1\le l,l'\le p, 1\le k\le K}\frac{1}{n}\left|\sum_{t=1}^{n} x_{((t-k),l)} I\left(z_t \le z_{\pi(i)}\right)\epsilon_{(t,l')}\right|. \tag{22}$$

The main goal is to find a proper rate for Equation (22). The indicator term $I(z_t < z_{\pi(i)})$ makes the proof more complicated, since we need to maximize Equation (22) w.r.t. t and we have no control on $z_{\pi(i)}$. Hence, we rewrite Equation (22) in the following form:

$$\max_{x\in\mathbb{R}, 1\le l,l'\le p, 1\le k\le K}\frac{1}{n}\left|\sum_{t=1}^{n} x_{((t-k),l)} I\left(z_t \le x\right)\epsilon_{(t,l')}\right|. \tag{23}$$

Similar to [10], we use the bracketing technique to bound Equation (23). To simplify the notation, we denote $x_{t,l}^k = x_{((t-k),l)}$, and let $W_n^{(l',l,k)}(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^n x_{t,l}^k I\left(z_t \le x\right)\epsilon_{(t,l')}$. Define $\Gamma_{(x)}(a) = a_1 I_{(-\infty,x)}\left(a_2\right)$ for $a \in \mathbb{R}^2$ and $\mathcal{F} = \left\{\Gamma_{(x)} : x\in\mathbb{R}\right\}$. Write $\Gamma_{(x)}$ as Γ. Let $M_t^{(l',l,k)} = x_{t,l}^k\epsilon_{(t,l')}$ and $Y_{nt}^{(l',l,k)} = \left(M_t^{(l',l,k)}/\sqrt{n}, z_t\right)$ for $l, l' \in 1, 2, \ldots, p$ and $k \in 1, 2, \ldots, K$, then $W_n^{(l',l,k)}(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^n M_t^{(l',l,k)} I\left(z_t \le x\right) = \sum_{t=1}^n \Gamma_{(x)}\left(Y_{nt}^{(l',l,k)}\right)$.

For any $x_1 < x_2$, we have:

$$\begin{aligned}
&\mathbb{E}[W_n^{(l',l,k)}(x_1) - W_n^{(l',l,k)}(x_2)]^2 \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{t=1}^n M_t^{(l',l,k)}\left[I\left(z_t \le x_1\right) - I\left(z_t \le x_2\right)\right]\right]^2 \\
&= \mathbb{E}\left[\left(M_t^{(l',l,k)}\right)^2 I\left(x_1 < z_t \le x_2\right)\right] \\
&= \left(\mathbb{E}\left(M_t^{(l',l,k)}\right)^2\right)\left(G^{(l',l,k)}\left(x_2\right) - G^{(l',l,k)}\left(x_1\right)\right),
\end{aligned} \tag{24}$$

where $G^{(l',l,k)}\left(x\right) = \mathbb{E}\left[\left(M_t^{(l',l,k)}\right)^2 I\left(z_t \le x\right)\right]/\mathbb{E}\left(M_t^{(l',l,k)}\right)^2$. Then for fixed l', l, k, we can construct a psudo-metric

$$d\left(x_1, x_2\right) = \sqrt{\left(\mathbb{E}\left(M_t^{(l',l,k)}\right)^2\right)\left|G^{(l',l,k)}\left(x_2\right) - G^{(l',l,k)}\left(x_1\right)\right|}.$$

For any $0 < \delta < 1$, the integral of the bracketing entropy satisfies the following

$$\int_0^\delta \sqrt{\log N(\epsilon, \mathcal{F}, L_2)}d\epsilon \le C \int_0^\delta \sqrt{-\log \epsilon}d\epsilon < \infty, \tag{25}$$

where $N(\epsilon, \mathcal{F}, L_2)$ denotes the brackets number, that is, the minimum number of ϵ-brackets needed to cover $\mathcal{F}$. Choose a fixed integer q_0 such that $4\delta \le 2^{-q_0} \le 8\delta$. Then, choose a nested sequence of partitions $\mathcal{F}_{q_{u'}}$ of $\mathcal{F}$ indexed by the integer $q \ge q_0$. By the Chaining Lemma [49], set $P_q = \{\Gamma_{(x)} : x \in B_{q_{u'}}, 1 \le u' \le N_q\}$ such that:

$$\sum_{q=q_0}^{\infty} 2^{-q}\sqrt{\log N_q} < \int_0^\delta \sqrt{\log N(\epsilon, \mathcal{F}, L_2)}d\epsilon,$$
$$\mathbb{E}\Lambda^2\left(B_{q_{u'}}\right) := \frac{1}{n}\mathbb{E}\sum_{t=1}^{n} \sup_{(x_1,x_2)\in B_{q_{u'}}} \left(M_t^{(l',l,k)}\right)^2 |I(x_1 < z_t \le x_2)| \le 2^{-2q}. \tag{26}$$

Fix q for each level $q > q_0$ and each partition $\mathcal{F}_{q_{u'}}$. For $x \in B_{q_{u'}}$, select a fixed $x_{q_{u'}} \in B_{q_{u'}}$ and define:

$$\pi_q x = x_{q_{u'}};$$
$$B_q x = B_{q_{u'}}.$$

Note that

$$\begin{aligned}
&\sum_{t=1}^{n} \Gamma\left(Y_{nt}^{(l',l,k)}\right)\\
&= \sum_{t=1}^{n} \Gamma_{(\pi_{q_0}x)}\left(Y_{nt}^{(l',l,k)}\right) + \sum_{t=1}^{n}\left(\Gamma\left(Y_{nt}^{(l',l,k)}\right) - \Gamma_{(\pi_{q_0}x)}\left(Y_{nt}^{(l',l,k)}\right)\right)\\
&= H_1^{(l',l,k)} + H_2^{(l',l,k)},
\end{aligned} \tag{27}$$

where

$$H_1^{(l',l,k)} = \sum_{t=1}^{n} \Gamma_{(\pi_{q_0}x)}\left(Y_{nt}^{(l',l,k)}\right)$$

and

$$H_2^{(l',l,k)} = \sum_{t=1}^{n}\left(\Gamma\left(Y_{nt}^{(l',l,k)}\right) - \Gamma_{(\pi_{q_0}x)}\left(Y_{nt}^{(l',l,k)}\right)\right).$$

To bound $H_1^{(l',l,k)}$, we apply Proposition 7 in [54], and take the union over N_{q_0} balls (which is a finite number). Let $\mathcal{K}$, c_0, c_1, and c_2 be positive constants. Then, for

$$n \ge c_0\left(\log(p^2K)\right)^{2/\varkappa_0 - 1},$$

we can get:

$$\mathbb{P}\left(\max_{1\le l,l'\le p,1\le k\le K}\sup_{\Gamma(x)\in\mathcal{F}}\sum_{t=1}^{n}\Gamma_{(\pi_{q_0}x)}\left(Y_{nt}^{(l',l,k)}\right)>c_1\mathcal{K}\sqrt{n}\sqrt{\frac{\log(p^2K)}{n}}\right)\\ \le N_{q_0}\left(2\exp\left(-c_2\log\left(p^2K\right)\right)\right). \tag{28}$$

To bound $H_2^{(l',l,k)}$, define

$$\begin{aligned} a_q &= 2^{-q}/\left[(\log n)^2\sqrt{\log N_{q+1}}\right],\\ \Omega_t^{(l',l,k)}(B) &= \sup_{(x_1,x_2)\in B}\left|\Gamma_{(x_1)}\left(Y_{nt}^{(l',l,k)}\right)-\Gamma_{(x_2)}\left(Y_{nt}^{(l',l,k)}\right)\right|,\\ A_{t,q_0}^{(l',l,k)} &= I\left(\Omega_t^{(l',l,k)}(B_{q_0}x)>a_{q_0}\right),\\ C_{t,q-1}^{(l',l,k)} &= I\left(\Omega_t^{(l',l,k)}(B_{q_0}x)\le a_{q_0},\ldots,\Omega_t^{(l',l,k)}(B_{q-1}x)\le a_{q-1}^{(l',l,k)}\right),\\ D_{t,q}^{(l',l,k)} &= I\Big(\Omega_t^{(l',l,k)}(B_{q_0}x)\le a_{q_0},\ldots,\\ &\qquad \Omega_t^{(l',l,k)}(B_{q-1}x)\le a_{q-1},\Omega_t^{(l',l,k)}(B_qx)>a_q\Big). \end{aligned} \tag{29}$$

Since $\mathbb{E}H_2^{(l',l,k)}=0$, $H_2^{(l',l,k)}$ can be decomposed into three parts

$$\begin{aligned} H_2^{(l',l,k)} &= \sum_{t=1}^{n}\Big\{\left[\Gamma_{(x)}(Y_{nt}^{(l',l,k)})-\Gamma_{(\pi_{q_0}x)}(Y_{nt}^{(l',l,k)})\right]A_{t,q_0}^{(l',l,k)}\\ &\qquad -\mathbb{E}\left[\left[\Gamma_{(x)}(Y_{nt}^{(l',l,k)})-\Gamma_{(\pi_{q_0}x)}(Y_{nt}^{(l',l,k)})\right]A_{t,q_0}^{(l',l,k)}\right]\Big\}\\ &\quad +\sum_{t=1}^{n}\sum_{q=q_0+1}^{\infty}\Big\{\left[\Gamma_{(\pi_qx)}(Y_{nt}^{(l',l,k)})-\Gamma_{(\pi_{q-1}x)}(Y_{nt}^{(l',l,k)})\right]C_{t,q-1}^{(l',l,k)}\\ &\qquad -\mathbb{E}\left[\left[\Gamma_{(\pi_qx)}(Y_{nt}^{(l',l,k)})-\Gamma_{(\pi_{q-1}x)}(Y_{nt}^{(l',l,k)})\right]C_{t,q-1}^{(l',l,k)}\right]\Big\}\\ &\quad +\sum_{t=1}^{n}\sum_{q=q_0+1}^{\infty}\Big\{\left[\Gamma_{(x)}(Y_{nt}^{(l',l,k)})-\Gamma_{(\pi_qx)}(Y_{nt}^{(l',l,k)})\right]D_{t,q}^{(l',l,k)}\\ &\qquad -\mathbb{E}\left[\left[\Gamma_{(x)}(Y_{nt}^{(l',l,k)})-\Gamma_{(\pi_qx)}(Y_{nt}^{(l',l,k)})\right]D_{t,q}^{(l',l,k)}\right]\Big\}\\ &=: S_{n1}^{(l',l,k)}+S_{n2}^{(l',l,k)}+S_{n3}^{(l',l,k)}. \end{aligned} \tag{30}$$

By Lemma 4 and Proposition 3 from [51], $\Gamma_{(x)}(Y_{nt}^{(l',l,k)})-\Gamma_{(\pi_{q_0}x)}(Y_{nt}^{(l',l,k)})$ is sub-Weibull. So there exists a constant C_0 and $\varkappa_1>0$ such that

$$\left(\mathbb{E}\left|\Gamma_{(x)}(Y_{nt}^{(l',l,k)})-\Gamma_{(\pi_{q_0}x)}(Y_{nt}^{(l',l,k)})\right|^c\right)^{1/c}\le C_0c^{(1/\varkappa_1+1/\varkappa_c)}$$

for all $c \geq 1$. Then, using the sub-Weibull property (first part of Theorem 2.1 in [51]), for any $\varepsilon > 0$, we can get

$$\begin{aligned}\mathbb{P}\left(\sup_{\Gamma(x)\in\mathcal{F}}\left|S_{n1}^{(l',l,k)}\right| > \varepsilon\right) &\leq \sum_{t=1}^{n}\mathbb{P}\left(\left|M_t^{(l',l,k)}\right| > \sqrt{n}a_{q_0}\right)\\ &\leq n\exp\left(-c_3(\sqrt{n}a_{q_0})^{\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}\right), \qquad (31)\end{aligned}$$

where c_3 is a positive constant and $\varkappa_1, \varkappa_c > 0$. Now, take the union over p^2K. Then, for a positive constant c_4 such that:

$$\begin{aligned}&\mathbb{P}\left(\max_{1\leq l,l'\leq p,1\leq k\leq K}\sup_{\Gamma(x)\in\mathcal{F}}\left|S_{n1}^{(l',l,k)}\right| > \varepsilon\right)\\ &\leq np^2K\exp\left(-c_3\left(\sqrt{n}2^{-q_0}/\left[(\log n)^2\sqrt{\log N_{q_0+1}}\right]\right)^{\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}\right) \qquad (32)\\ &\leq \exp\left(-c_4\frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}} + \log\left(np^2K\right)\right).\end{aligned}$$

The second inequality satisfies since q_0 is fixed and its value will be determined later. To bound $S_{n2}^{(l',l,k)}$ and $S_{n3}^{(l',l,k)}$, we apply a similar procedure as in Lemma A.1 in [10]. Specifically, let

$$\begin{aligned}\zeta_{nt}^{(l',l,k)}\left(q,\Gamma_{(x)}\right) = &\Big\{\left(\Gamma_{\pi_q x}\left(Y_{nt}^{(l',l,k)}\right) - \Gamma_{(\pi_{q-1}x)}\left(Y_{nt}^{(l',l,k)}\right)\right)C_{t,q-1}^{(l',l,k)}\\ &- \mathbb{E}\left[\left(\Gamma_{\pi_q x}\left(Y_{nt}^{(l',l,k)}\right) - \Gamma_{(\pi_{q-1}x)}\left(Y_{nt}^{(l',l,k)}\right)\right)C_{t,q-1}^{(l',l,k)}\right]\Big\}.\end{aligned}$$

For any $q \geq q_0$, since $\Gamma_{(x_q)}(Y_{nt}^{(l',l,k)})$ and $\Gamma_{(x_{q-1})}(Y_{nt}^{(l',l,k)})$ are points lying in one of the balls $B_{q-1,u'}$, $u' \leq N_{q-1}$, $\{\zeta_{nt}^{(l',l,k)}(q,\Gamma_{(x)})\}$ is a centered β-mixing process. Since β-mixing implies α-mixing, we can use the results in Lemma A.1 in [10]. For any $y \geq 0$, there exists a positive constant c_5 such that

$$\mathbb{P}\left(\sup_{\Gamma_{(x)}\in\mathcal{F}}\left|S_{n2}^{(l',l,k)}\right| \geq h_{q_0}y\right) \leq \sum_{q=q_0+1}^{\infty}N_q\exp\left\{-\frac{c_5y^2\log N_q}{2+y}\right\}, \qquad (33)$$

where $h_{q_0} = \sum_{q=q_0}^{\infty}2^{-q/2}\sqrt{\log N_q}$, and $q_0, n \geq 3$. Since i and j are fixed, we take the union over p^2K again and get the bound of $S_{n2}^{(l',l,k)}$ as

$$\begin{aligned}&\mathbb{P}\left(\max_{1\leq l,l'\leq p,1\leq k\leq K}\sup_{\Gamma_{(x)}\in\mathcal{F}}\left|S_{n2}^{(l',l,k)}\right| \geq h_{q_0}y\right)\\ &\leq \sum_{q=q_0+1}^{\infty}N_q\exp\left\{-\frac{c_5y^2\log N_q}{2+y} + \log\left(p^2K\right)\right\}.\end{aligned} \qquad (34)$$

Use the same argument of $S_{n2}^{(l',l,k)}$, we have:

$$\begin{aligned} &\mathbb{P}\left(\max_{1\le l,l'\le p,1\le k\le K}\sup_{\Gamma_{(x)}\in\mathcal{F}}\left|S_{n3}^{(l',l,k)}\right|\ge h_{q_0}y\right)\\ &\le\sum_{q=q_0+1}^{\infty}N_q\exp\left\{-\frac{c_5y^2\log N_q}{2+y}+\log\left(p^2K\right)\right\}. \end{aligned} \tag{35}$$

When n is large enough, we can combine $S_{n1}^{(l',l,k)}$, $S_{n2}^{(l',l,k)}$, and $S_{n3}^{(l',l,k)}$. Thus, we have:

$$\begin{aligned} &\mathbb{P}\left\{\max_{1\le l,l'\le p,1\le k\le K}\sup_{\Gamma_{(x)}\in\mathcal{F}}H_2^{(l',l,k)}\ge 2h_{q_0}\left(y+1\right)\right\}\\ &\le 2\sum_{q=q_0+1}^{\infty}N_q\exp\left\{-\frac{c_5y^2\log N_q}{2+y}+\log\left(p^2K\right)\right\}+\\ &\quad\exp\left(-c_4\frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}+\log\left(np^2K\right)}\right)\\ &\le 2p^2K\sum_{q=q_0+1}^{\infty}N_q\exp\left(-\frac{c_5y^2\log N_q}{2+y}\right)+\\ &\quad np^2K\exp\left(-c_4\frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}}\right)\\ &\le 2p^2K\sum_{q=q_0+1}^{\infty}N_q^{1-\frac{c_5y^2}{2+y}}+np^2K\exp\left(-c_4\frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}}\right) \end{aligned} \tag{36}$$

Now, take $y = C_0'\frac{\log\left(p^2K\right)}{\sqrt{n}}\sqrt{n}$ and q_0 to be a smallest integer such that $q_0\ge 3$ and $h_{q_0}\le 1$. Note that N_{q_0} is a constant since q_0 is selected to be fixed. Equation (28) and Equation (36) give Lemma 5 as desired. Specifically, if we choose C_0' large enough, there exist positive constants C, c_6, c_7, c_8, $c_9>1$, c_{10} and $\frac{c_5y^2}{2+y}>3$ such that

$$\begin{aligned} &\mathbb{P}\left(\sup_{x\in\mathbb{R},1\le l,l'\le p,1\le k\le K}\frac{1}{n}\left|\sum_{t=1}^{n}x_{((t-k),l)}I\left(z_t\le x\right)\epsilon_{(t,l')}\right|\ge C\frac{\log\left(p^2K\right)}{\sqrt{n}}\right)\\ &\le N_{q_0}\left(2\exp\left(-c_2\log\left(p^2K\right)\right)\right)+2p^2K\sum_{q=q_0+1}^{\infty}N_q^{1-\frac{c_5y^2}{2+y}}\\ &\quad+np^2K\exp\left(-c_4\frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}}\right)\\ &\le N_{q_0}\left(2\exp\left(-c_2\log\left(p^2K\right)\right)\right)+c_7p^2K\exp\left(-c_8y\right) \end{aligned}$$

$$
\begin{aligned}
&+ np^2K \exp\left(-c_4 \frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}}\right) \\
&\le N_{q_0}\left(2\exp\left(-c_2\log\left(p^2K\right)\right)\right) + c_7\exp\left(-c_8C_0'\log\left(p^2K\right)+\log\left(p^2K\right)\right) \\
&\quad + \exp\left(-c_4 \frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}} + \log\left(np^2K\right)\right) \\
&\le c_6\exp\left(-c_9\log\left(p^2K\right)+\log\left(p^2K\right)\right) \\
&\quad + \exp\left(-c_4 \frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}} + \log\left(np^2K\right)\right) \\
&\le c_6\exp\left(-c_{10}\log\left(p^2K\right)\right) + \exp\left(-c_4 \frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}} + \log\left(np^2K\right)\right).
\end{aligned} \tag{37}
$$

Lemma 6. *Under Assumptions A1 to A4, there exist positive constants C, c_0, c_1, c_2, and c_3, such that for*

$$
n \ge c_0\left(\log(p^2K)\right)^{2/\varkappa_0-1},
$$

with probability at least $1-c_3\eta_1-\eta_2$, we have:

$$
\frac{1}{n}\|\boldsymbol{Z}'\boldsymbol{E}\|_{2,\infty} \le C\frac{\sqrt{p^2K}\log\left(p^2K\right)}{\sqrt{n}}, \tag{38}
$$

where $\eta_1 = \exp\left(-c_1\log\left(p^2K\right)\right)$ and

$$
\eta_2 = \exp\left(-c_2 \frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}} + \log\left(np^2K\right)\right).
$$

Proof of Lemma 6: By combining Equation (28) and Equation (36) in Lemma 5, we have:

$$
\begin{aligned}
&\mathbb{P}\left(\frac{1}{n}\|\boldsymbol{Z}'\boldsymbol{E}\|_\infty \ge C\frac{\log\left(p^2K\right)}{\sqrt{n}}\right) \\
&\le c_3\exp\left(-c_1\log\left(p^2K\right)\right) + \exp\left(-c_2 \frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}} + \log\left(np^2K\right)\right),
\end{aligned} \tag{39}
$$

where C, c_1, c_2, c_3 are positive constants.

Let G_g represents the group in group lasso for $g = 1, 2, \ldots, n$, where $G_1 = \{1, 2, \ldots, p^2K\}, G_2 = \{p^2K+1, p^2K+2, \ldots, 2p^2K\}, \ldots, G_n = \{(n-1)p^2K + 1, (n-1)p^2K+2, \ldots, np^2K\}$. Note that for none overlapping group G_g with size p^2K, we have:

$$\frac{1}{\sqrt{p^2K}}\|vec(\boldsymbol{Z}_l'\boldsymbol{E}), l \in G_g\|_2 = \sqrt{\frac{1}{p^2K}\sum_{l\in G_g}|vec(\boldsymbol{Z}_l'\boldsymbol{E})|^2} \leq \max_{l\in G_g}|vec(\boldsymbol{Z}_l'\boldsymbol{E})|, \tag{40}$$

where $\boldsymbol{Z}_l'$ represents l-th row of $\boldsymbol{Z}'$. Thus,

$$\begin{aligned}\frac{1}{\sqrt{p^2K}}\|vec(\boldsymbol{Z}'\boldsymbol{E})\|_{2,\infty} &= \max_{g=1,\dots,n}\frac{1}{\sqrt{p^2K}}\|vec(\boldsymbol{Z}_l'\boldsymbol{E}), l\in G_g\|_2 \\ &\leq \max_{g=1,\dots,n}\max_{l\in G_g}|vec(\boldsymbol{Z}_l'\boldsymbol{E})| \\ &= \|\boldsymbol{Z}'\boldsymbol{E}\|_\infty.\end{aligned} \tag{41}$$

Combining the Lemma 5 and Equation (41), we have:

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{n}\|vec\left(\boldsymbol{Z}'\boldsymbol{E}\right)\|_{2,\infty} \geq C\sqrt{p^2K}\frac{\log\left(p^2K\right)}{\sqrt{n}}\right) \\ &\leq \mathbb{P}\left(\frac{1}{n}\|\boldsymbol{Z}'\boldsymbol{E}\|_\infty \geq C\frac{\log\left(p^2K\right)}{\sqrt{n}}\right) \\ &\leq c_3\exp\left(-c_1\log\left(p^2K\right)\right) + \exp\left(-c_2\frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}} + \log\left(np^2K\right)\right).\end{aligned} \tag{42}$$

Lemma 7. *Set* $\sigma^2(s) = \mathbb{E}\left(x_{(t-k,g')}I\left(r_j - s < z_t \leq r_j\right)\right)^2$ *for any given* $1 \leq g' \leq p$. *Under Assumptions A1 to A4, there exist positive constants* c_i, C, C', C_i', C_i'' *for* $i = 1, 2, \dots$, *such that for any given* j*-th threshold, with probability at least* $1 - \delta_1$,

$$\begin{aligned}\sup_{\substack{1\leq k\leq K,\\ |s|\geq\gamma_n}}&\left(n\sigma^2(s)\right)^{-1}\left\|\sum_{t=1}^n \boldsymbol{x}_{(t-k)}\boldsymbol{x}_{(t-k)}'I\left(r_j - s < z_t \leq r_j\right)\right. \\ &\left.- \mathbb{E}\left(\boldsymbol{x}_{(t-k)}\boldsymbol{x}_{(t-k)}'I\left(r_j - s < z_t \leq r_j\right)\right)\right\|_\infty \\ &\leq C\left(\frac{\left(\log(p^2K)\right)^{1/\varkappa_0 - 1/2}}{\sqrt{n\gamma_n}}\right),\end{aligned} \tag{43}$$

where

$$\begin{aligned}\delta_1 = \max\Bigg\{&\exp\left(-C_1'\left(\frac{n}{\gamma_n}\right)^{\varkappa_0/2}\left(\log\left(p^2K\right)\right)^{1-\varkappa_0/2} + \log\left(p^2K\right) + \log\left(n\right)\right) \\ &+ \exp\left(-C_2'\frac{1}{\gamma_n}\log\left(p^2K\right)^{2/\varkappa_0 - 1} + \log\left(p^2K\right)\right),\end{aligned}$$

$$\exp\left(-C_3'\left(n\gamma_n\right)^{\varkappa_0/2}\left(\log\left(p^2K\right)\right)^{1-\varkappa_0/2}+\log\left(p^2K\right)+\log\left(n\gamma_n\right)\right)$$
$$\left.+\exp\left(-C_4'\log\left(p^2K\right)^{2/\varkappa_0-1}+\log\left(p^2K\right)\right)\right\}.$$

In addition, with probability at least $1-\delta_2$,

$$\sup_{\substack{1\le l,l'\le p,\\ 1\le k\le K,\\ |s|\ge\gamma_n}}\left(n\sigma^2\left(s\right)\right)^{-1}\left|\sum_{t=1}^{n}x_{(t-k,l)}I\left(r_j-s<z_t\le r_j\right)\epsilon_{(t,l')}\right|\le C'\frac{\log(p^2K)}{\sqrt{n\gamma_n}},\quad (44)$$

where

$$\delta_2=c_3\exp\left(-c_4\log\left(p^2K\right)\right)+\exp\left(-c_5\frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}}+\log\left(np^2K\right)\right).\quad (45)$$

Proof of Lemma 7: The proof for this lemma is along the lines of the proof of Lemma A.2 in [10]. Assume a fixed small number $D>0$. Since $\sigma^2(s)$ is non-decreasing in given distance s, and $\boldsymbol{\epsilon}_t$ and z_t both have bounded positive density based on Assumptions A1 and A5,

$$\sigma^2(s)\ge\sigma^2(D)\ge CD\quad\text{if}\quad s\ge D,\quad (46)$$

where C is a constant greater than 0. Similar to Lemma 5, for $s\ge D\ge\gamma_n$, according to Equation (28) and Equation (36), for a given j-th threshold, there exist large enough positive constant C_0, and positive constants C', $c_{h'}$, $C'_{h'}$ for $h'=1,2,\ldots,12$ such that

$$\mathbb{P}\left(\sup_{\substack{1\le l,l'\le p,\\ 1\le k\le K,\\ |s|\ge D}}\left(n\sigma^2\left(s\right)\right)^{-1}\left|\sum_{t=1}^{n}x_{(t-k,l)}I\left(r_j-s<z_t\le r_j\right)\epsilon_{(t,l')}\right|\ge\left(\frac{C_0}{CD}\right)\frac{\log\left(p^2K\right)}{\sqrt{n}}\right)$$
$$\le\mathbb{P}\left(\sup_{\substack{1\le l,l'\le p,\\ 1\le k\le K,\\ |s|\ge D}}\left(n\sigma^2\left(s\right)\right)^{-1}\left|\sum_{t=1}^{n}x_{(t-k,l)}I\left(r_j-s<z_t\le r_j\right)\epsilon_{(t,l')}\right|\ge C_0\frac{\log\left(p^2K\right)}{\sqrt{n}}/\sigma^2\left(s\right)\right)$$
$$\le c_3\exp\left(-c_1\log\left(p^2K\right)\right)+\exp\left(-c_2\frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}}+\log\left(np^2K\right)\right).\quad (47)$$

Thus, with high probability, we obtain Equation (44) when $s\ge D$.

When $s \in [\gamma_n, D]$, we want to partition the interval into small pieces, and prove the consistency in each piece. Let $M = [\log(D/\gamma_n)/\log b]$, where $b > 1$ is a constant. Now,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\substack{1\le l,l'\le p,\\ 1\le k\le K,\\ s\in[\gamma_n,D]}} \left(n\sigma^2(s)\right)^{-1}\left|\sum_{t=1}^{n} x_{(t-k,l)} I\left(r_{j-1}-s<z_t\le r_{j-1}\right)\epsilon_{(t,l')}\right| \ge y_1\right)\\
&\le \sum_{g=0}^{M}\mathbb{P}\left(\sup_{\substack{1\le l,l'\le p,\\ 1\le k\le K,\\ s\in[b^g\gamma_n,b^{g+1}\gamma_n]}} \left(n\sigma^2(b^g\gamma_n)\right)^{-1}\right.\\
&\qquad\left.\left|\sum_{t=1}^{n} x_{(t-k,l)} I\left(r_j-s<z_t\le r_j-b^g\gamma_n\right)\epsilon_{(t,l')}\right| \ge y_1/2\right)\\
&\quad+\sum_{g=0}^{M}\mathbb{P}\left(\sup_{\substack{1\le l,l'\le p,\\ 1\le k\le K}} \left(n\sigma^2(b^g\gamma_n)\right)^{-1}\right.\\
&\qquad\left.\left|\sum_{t=1}^{n} x_{(t-k,l)} I\left(r_j-b^g\gamma_n<z_t\le r_j\right)\epsilon_{(t,j)}\right| \ge y_1/2\right)\\
&\le \sum_{g=0}^{M}\mathbb{P}\left(\sup_{\substack{1\le l,l'\le p,\\ 1\le k\le K}} \left(n\sigma^2(b^g\gamma_n)\right)^{-1}\right.\\
&\qquad\left.\sum_{t=1}^{n} \left|x_{(t-k,l)} I\left(r_j-b^{g+1}\gamma_n<z_t\le r_j-b^g\gamma_n\right)\epsilon_{(t,l')}\right| \ge y_1/2\right)\\
&\quad+\sum_{g=0}^{M}\mathbb{P}\left(\sup_{\substack{1\le l,l'\le p,\\ 1\le k\le K}} \left(n\sigma^2(b^g\gamma_n)\right)^{-1}\right.\\
&\qquad\left.\left|\sum_{t=1}^{n} x_{(t-k,l)} I\left(r_j-b^g\gamma_n<z_t\le r_j\right)\epsilon_{(t,l')}\right| \ge y_1/2\right)\\
&\le \sum_{g=0}^{M}\mathbb{P}\left(\sup_{\substack{1\le l,l'\le p,\\ 1\le k\le K}} \left(C_1' n\gamma_n b^g\right)^{-1}\right.\\
&\qquad\left.\sum_{t=1}^{n} \left|x_{(t-k,l)} I\left(r_j-b^{g+1}\gamma_n<z_t\le r_j-b^g\gamma_n\right)\epsilon_{(t,l')}\right| \ge y_1/2\right)\\
&\quad+\sum_{g=0}^{M}\mathbb{P}\left(\sup_{\substack{1\le l,l'\le p,\\ 1\le k\le K}} \left(C_1' n\gamma_n b^g\right)^{-1}\right.\\
&\qquad\left.\left|\sum_{t=1}^{n} x_{(t-k,l)} I\left(r_j-b^g\gamma_n<z_t\le r_j\right)\epsilon_{(t,l')}\right| \ge y_1/2\right)
\end{aligned}
$$

$$=: \sum_{g=0}^{M} H_{ng} + \sum_{g=0}^{M} I_{ng}. \tag{48}$$

Recall that $1 > \varkappa_0 > 0$, and the fact that the function of a β-mixing process is also a β-mixing. Since $\boldsymbol{x}_t$ and z_t are β-mixing,

$$x_{(t-k,l)} I\left(r_j - b^g \gamma_n < z_t \le r_j\right) \epsilon_{(t,l')}$$

and

$$x_{(t-k,l)} I\left(r_j - b^{g+1} \gamma_n < z_t \le r_j - b^g \gamma_n\right) \epsilon_{(t,l')}$$

are β-mixing. To bound H_{ng} and I_{ng}, we can apply Proposition 7 from [54]. Set

$$y_1/2 = C_2' \frac{\log(p^2 K)}{\sqrt{n\gamma_n}}.$$

For

$$C_1' b^g n \gamma_n \ge C_3' \left(\log\left(p^2 K\right)\right)^{2/\varkappa_0 - 1},$$

we can get

$$\begin{aligned} I_{ng} < \mathbb{P}\Bigg(& \left(C_1' n \gamma_n b^g\right)^{-1} \left| \sum_{t=1}^{n} x_{(t-k,l)} I\left(r_j - b^g \gamma_n < z_t \le r_j\right) \epsilon_{(t,l')} \right| \\ & \ge C_2' \sqrt{\frac{\log(p^2 K)}{n \gamma_n}} \Bigg) \\ \le & \, 2 \exp\left(-C_4' \log\left(p^2 K\right)\right). \end{aligned} \tag{49}$$

Then, we can get

$$\sum_{g=0}^{M} I_{ng} \le C_5' \exp\left(-C_4' \log\left(p^2 K\right)\right). \tag{50}$$

Similarly, we can get

$$\sum_{g=0}^{M} H_{ng} \le C_6' \exp\left(-C_4' \log\left(p^2 K\right)\right). \tag{51}$$

By Equation (51), Equation (50), and Equation (47), we then can get:

$$\sup_{\substack{1 \le l, l' \le p, \\ 1 \le k \le K, \\ |s| \ge \gamma_n}} \left(n \sigma^2(s)\right)^{-1} \left| \sum_{t=1}^{n} x_{(t-k,l)} I\left(r_j - s < z_t \le r_j\right) \epsilon_{(t,l')} \right| \le C' \frac{\log(p^2 K)}{\sqrt{n\gamma_n}}, \tag{52}$$

with probability $1 - \delta_3$, where

$$\delta_3 = c_4 \exp\left(-c_5 \log\left(p^2 K\right)\right) + \exp\left(-c_2 \frac{n^{\varkappa_1 \varkappa_c / 2(\varkappa_1 + \varkappa_c)}}{(\log n)^{2\varkappa_1 \varkappa_c / (\varkappa_1 + \varkappa_c)}} + \log\left(np^2 K\right)\right). \tag{53}$$

Thus, this proves Equation (44). Similarly, we can prove Equation (43). Note that

$$\mathbb{E}\left(\boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)}I\left(r_j - s < z_t \leq r_j\right)\right)$$

is non-decreasing in s and $\mathbb{E}|\boldsymbol{x}_t|^2$ is positive and bounded from Assumption A2. For $s \geq D \geq \gamma_n$, we first fix l, l' in $1, 2, \ldots, p$ and k in $1, 2, \ldots, K$. Recall that

$$\sigma^2(s) \geq \sigma^2(D) \geq CD \quad \text{if} \quad s \geq D, \tag{54}$$

where C is a constant greater than 0. Note that $\varkappa_0 = \left(1/(\varkappa_1/2) + 1/\varkappa_2\right)^{-1} < 1$. By Assumption A2, Lemma 4, and Fact 1 and Lemma 13 in Wong et al. [54], for $n > 4$,

$$\begin{aligned}
&\mathbb{P}\Bigg(\sup_{|s|\geq D}\left(n\sigma^2(s)\right)^{-1}\Bigg|\sum_{t=1}^{n} x_{(t-k,l)}x_{(t-k,l')}I\left(r_j - s < z_t \leq r_j\right) - \\
&\quad \mathbb{E}\left(x_{(t-k,l)}x_{(t-k,l')}I\left(r_j - s < z_t \leq r_j\right)\right)\Bigg| \geq y_2\Bigg) \\
&\leq \mathbb{P}\Bigg(\sup_{|s|\geq D}(CDn)^{-1}\Bigg|\Bigg(\sum_{t=1}^{n} x_{(t-k,l)}x_{(t-k,l')}I\left(r_j - s < z_t \leq r_j\right) - \\
&\quad \mathbb{E}\left(x_{(t-k,l)}x_{(t-k,l')}I\left(r_j - s < z_t \leq r_j\right)\right)\Bigg)\Bigg| \geq y_2\Bigg) \\
&\leq n\exp\left(-c_6(ny_2)^{\varkappa_0}\right) + \exp\left(-c_7 n y_2^2\right).
\end{aligned} \tag{55}$$

Then, we take the union over p^2K:

$$\begin{aligned}
&\mathbb{P}\Bigg(\sup_{\substack{1\leq k\leq K,\\ |s|\geq D}}\left(n\sigma^2(s)\right)^{-1}\Bigg\|\sum_{t=1}^{n}\boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)}I\left(r_j - s < z_t \leq r_j\right) - \\
&\quad \mathbb{E}\left(\boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)}I\left(r_j - s < z_t \leq r_j\right)\right)\Bigg\|_{\infty} \geq y_2\Bigg) \\
&= \mathbb{P}\Bigg(\sup_{\substack{1\leq l,l'\leq p,\\ 1\leq k\leq K,\\ |s|\geq D}}\left(n\sigma^2(s)\right)^{-1}\Bigg|\sum_{t=1}^{n} x_{(t-k,l)}x_{(t-k,l')}I\left(r_j - s < z_t \leq r_j\right) - \\
&\quad \mathbb{E}\left(x_{(t-k,l)}x_{(t-k,l')}I\left(r_j - s < z_t \leq r_j\right)\right)\Bigg| \geq y_2\Bigg) \\
&\leq n\exp\left(-c_6(ny_2)^{\varkappa_0} + \log\left(p^2K\right)\right) + \exp\left(-c_7 n y_2^2 + \log\left(p^2K\right)\right).
\end{aligned} \tag{56}$$

For $s \in [\gamma_n, D]$, we have:

$$
\begin{aligned}
&\mathbb{P}\Bigg(\sup_{\substack{1\le k\le K,\\ D\ge|s|\ge\gamma_n}} \left(n\sigma^2(s)\right)^{-1} \Bigg\| \sum_{t=1}^{n} \boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)} I\left(r_j - s < z_t \le r_j\right) - \\
&\qquad \mathbb{E}\left(\boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)} I\left(r_j - s < z_t \le r_j\right)\right) \Bigg\|_\infty \ge y_2 \Bigg) \\
&\le \sum_{g=0}^{M} \mathbb{P}\Bigg(\sup_{\substack{1\le k\le K,\\ s\in[b^g\gamma_n, b^{g+1}\gamma_n]}} (n\sigma^2(b^g\gamma_n))^{-1} \\
&\quad \Bigg\| \sum_{t=1}^{n} \boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)} I\left(r_j - s < z_t \le r_j - b^g\gamma_n\right) - \\
&\qquad \mathbb{E}\left(\boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)} I\left(r_j - s < z_t \le r_j - b^g\gamma_n\right)\right) \Bigg\|_\infty \ge y_2/2 \Bigg) \\
&\quad + \sum_{g=0}^{M} \mathbb{P}\Bigg(\sup_{\substack{1\le k\le K,\\ s\in[b^g\gamma_n, b^{g+1}\gamma_n]}} \left(n\sigma^2(b^g\gamma_n)\right)^{-1} \\
&\quad \Bigg\| \sum_{t=1}^{n} \boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)} I\left(r_j - b^g\gamma_n < z_t \le r_j\right) - \\
&\qquad \mathbb{E}\left(\boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)} I\left(r_j - b^g\gamma_n < z_t \le r_j\right)\right) \Bigg\|_\infty \ge y_2/2 \Bigg) \\
&\le \sum_{g=0}^{M} \mathbb{P}\Bigg(\left(n\sigma^2(b^g\gamma_n)\right)^{-1} \Bigg\| \sum_{t=1}^{n} \boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)} I\left(r_j - b^{g+1}\gamma_n < z_t \le r_j - b^g\gamma_n\right) - \\
&\qquad \mathbb{E}\left(\boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)} I\left(r_j - b^{g+1}\gamma_n < z_t \le r_j - b^g\gamma_n\right)\right) \Bigg\|_\infty \ge y_2/2 \Bigg) \\
&\quad + \sum_{g=0}^{M} \mathbb{P}\Bigg(\sup_{\substack{1\le k\le K,\\ s\in[b^g\gamma_n, b^{g+1}\gamma_n]}} \left(n\sigma^2(b^g\gamma_n)\right)^{-1} \\
&\quad \Bigg\| \sum_{t=1}^{n} \boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)} I\left(r_j - b^g\gamma_n < z_t \le r_j\right) - \\
&\qquad \mathbb{E}\left(\boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)} I\left(r_j - b^g\gamma_n < z_t \le r_j\right)\right) \Bigg\|_\infty \ge y_2/2 \Bigg) \\
&=: \sum_{g=0}^{M} I_{1ng} + \sum_{g=0}^{M} I_{2ng}.
\end{aligned}
\tag{57}
$$

Note that $\mathbb{E}(\boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)}I(r_j - s < z_t \le r_j))$ is non-decreasing in s and $\mathbb{E}|\boldsymbol{x}_t|^2$ is positive and bounded from Assumption A2. By Lemma 4, Lemma 13 in Wong et al. [54] and taking union over p^2K, for $n\gamma_n > 4$, we can obtain

$$\begin{aligned}\sum_{g=0}^{M} I_{1ng} \le n\gamma_n \exp\left(-c_8(n\gamma_n y_2)^{\varkappa_0} + \log\left(p^2K\right)\right) \\ + \exp\left(-c_9 n\gamma_n y_2^2 + \log\left(p^2K\right)\right).\end{aligned} \tag{58}$$

Similarly, we have

$$\begin{aligned}\sum_{g=0}^{M} I_{2ng} \le n\gamma_n \exp\left(-c_{10}(n\gamma_n y_2)^{\varkappa_0} + \log\left(p^2K\right)\right) \\ + \exp\left(-c_{11} n\gamma_n y_2^2 + \log\left(p^2K\right)\right).\end{aligned} \tag{59}$$

Combining Equation (58), Equation (59) and Equation (56) and setting

$$y_2/2 = c_{12}\left(\frac{\left(\log(p^2K)\right)^{1/\varkappa_0 - 1/2}}{\sqrt{n\gamma_n}}\right)$$

with large enough c_{12}, we have

$$\begin{aligned}&\sup_{\substack{1\le k\le K,\\ |s|\ge\gamma_n}} (n\sigma^2(s))^{-1}\Bigg\| \sum_{t=1}^{n} \boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)} I\left(r_j - s < z_t \le r_j\right) \\ &\qquad - \mathbb{E}\left(\boldsymbol{x}_{(t-k)}\boldsymbol{x}'_{(t-k)} I\left(r_j - s < z_t \le r_j\right)\right)\Bigg\|_\infty \\ &\le c_{12}\left(\frac{\left(\log(p^2K)\right)^{1/\varkappa_0 - 1/2}}{\sqrt{n\gamma_n}}\right),\end{aligned} \tag{60}$$

with probability $1 - \delta_4$, where

$$\begin{aligned}\delta_4 = \max\Bigg\{&n\exp\left(-c_6(ny_2)^{\varkappa_0} + \log\left(p^2K\right)\right) + \exp\left(-c_7 n y_2^2 + \log\left(p^2K\right)\right), \\ &n\gamma_n \exp\left(-c_8(n\gamma_n y_2)^{\varkappa_0} + \log\left(p^2K\right)\right) + \exp\left(-c_9 n\gamma_n y_2^2 + \log\left(p^2K\right)\right)\Bigg\} \\ = \max\Bigg\{&n\exp\left(-c_6\left(nc_{12}\left(\frac{\left(\log(p^2K)\right)^{1/\varkappa_0-1/2}}{\sqrt{n\gamma_n}}\right)\right)^{\varkappa_0} + \log\left(p^2K\right)\right) \\ &+ \exp\left(-c_7 n\left(c_{12}\left(\frac{\left(\log(p^2K)\right)^{1/\varkappa_0-1/2}}{\sqrt{n\gamma_n}}\right)\right)^2 + \log\left(p^2K\right)\right),\end{aligned}$$

$$
\begin{aligned}
& n\gamma_n \exp\left(-c_8\left(n\gamma_n c_{12}\left(\frac{\left(\log(p^2K)\right)^{1/\varkappa_0-1/2}}{\sqrt{n\gamma_n}}\right)\right)^{\varkappa_0}+\log\left(p^2K\right)\right) \\
& + \exp\left(-c_9 n\gamma_n\left(c_{12}\left(\frac{\left(\log(p^2K)\right)^{1/\varkappa_0-1/2}}{\sqrt{n\gamma_n}}\right)\right)^2+\log\left(p^2K\right)\right)\Bigg\} \\
= \max & \Bigg\{\exp\left(-C_7'\left(\frac{n}{\gamma_n}\right)^{\varkappa_0/2}\left(\log\left(p^2K\right)\right)^{1-\varkappa_0/2}+\log\left(p^2K\right)+\log\left(n\right)\right) \\
& + \exp\left(-C_8'\frac{1}{\gamma_n}\log\left(p^2K\right)^{2/\varkappa_0-1}+\log\left(p^2K\right)\right), \\
& \exp\left(-C_9'\left(n\gamma_n\right)^{\varkappa_0/2}\left(\log\left(p^2K\right)\right)^{1-\varkappa_0/2}+\log\left(p^2K\right)+\log\left(n\gamma_n\right)\right) \\
& + \exp\left(-C_{10}'\log\left(p^2K\right)^{2/\varkappa_0-1}+\log\left(p^2K\right)\right)\Bigg\}. \qquad (61)
\end{aligned}
$$

Note that $C'\left(n\gamma_n\right)^{\varkappa_0/2} > \left(\log\left(p^2K\right)\right)^{\varkappa_0}$ by Assumption A4. So

$$
C' n^{\varkappa_0/2}\left(\log\left(p^2K\right)\right)^{1-\varkappa_0/2} \geq \log(p^2K).
$$

Thus, both

$$
\exp\left(-C_7'\left(\frac{n}{\gamma_n}\right)^{\varkappa_0/2}\left(\log\left(p^2K\right)\right)^{1-\varkappa_0/2}+\log\left(p^2K\right)+\log\left(n\right)\right)
$$

and

$$
\exp\left(-C_9'\left(n\gamma_n\right)^{\varkappa_0/2}\left(\log\left(p^2K\right)\right)^{1-\varkappa_0/2}+\log\left(p^2K\right)+\log\left(n\gamma_n\right)\right)
$$

will converge to zero as sample size n tends to infinity. According to Assumption A2, $\varkappa_0 < 1$, so $2/\varkappa_0 - 1 > 1$. As a result, δ_4 will converge to 0.

Lemma 8. *Set $\sigma^2(s) = \mathbb{E}\left(x_{(t-k,g')}I\left(r_j - s < z_t < r_j\right)\right)^2$ for any given $1 \leq g' \leq p$. Let $\mathbb{I}_s \in \mathbb{R}^{np\times np}$ be the diagonal matrix with diagonal $I_1(s), \ldots, I_n(s)$, where $I_t(s)$ is a $p\times p$ diagonal matrix with all diagonal elements equal to $I\left(r_j - s < z_t \leq r_j\right)$ for $t = 1, 2, \ldots, n$. Under Assumptions A1 to A4, there exist positive constants c_3, c_4, c_5, C', C_1'', such that for any given j-th threshold, with probability at least $1-\delta_2'$,*

$$
\sup_{|s|\geq\gamma_n}\left(n\sigma^2\left(s\right)\right)^{-1}\left\|\boldsymbol{Z}'\mathbb{I}_{r_j}(s)\boldsymbol{E}\right\|_{2,\infty} \leq C'\frac{\sqrt{p^2K}\log(p^2K)}{\sqrt{n\gamma_n}}, \qquad (62)
$$

where

$$
\delta_2' = c_3\exp\left(-c_4\log\left(p^2K\right)\right) + \exp\left(-c_5\frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}}+\log\left(np^2K\right)\right).
$$

Proof of Lemma 8: By Lemma 7, we have:

$$\mathbb{P}\left(\sup_{\substack{1\le l,l'\le p,\\ 1\le k\le K,\\ |s|\ge\gamma_n}} \left(n\sigma^2(s)\right)^{-1}\left|\sum_{t=1}^{n} x_{(t-k,l)} I\left(r_j - s < z_t \le r_j\right)\epsilon_{(t,l')}\right| \le C'\frac{\log(p^2K)}{\sqrt{n\gamma_n}}\right)$$
$$\le c_3\exp\left(-c_4\log\left(p^2K\right)\right) + \exp\left(-c_5\frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}} + \log\left(np^2K\right)\right) \tag{63}$$

where C', c_3, c_4, c_5 are positive constants.

Note $\sup_{\substack{1\le l,l'\le p,\\ 1\le k\le K,\\ |s|\ge\gamma_n}} \left(n\sigma^2(s)\right)^{-1}\left|\sum_{t=1}^{n} x_{(t-k,l)} I\left(r_j - s < z_t \le r_j\right)\epsilon_{(t,l')}\right|$ can be written as $\sup_{|s|\ge\gamma_n}\left(n\sigma^2(s)\right)^{-1}\left\|\boldsymbol{Z}'\mathbb{I}_{r_j}(s)\boldsymbol{E}\right\|_\infty$ for a given r_j.

Recall that G_g represents the group in group lasso for $g = 1, 2, \dots, n$, where $G_1 = \{1, 2, \dots, p^2K\}, G_2 = \{p^2K + 1, p^2K + 2, \dots, 2p^2K\}, \dots, G_n = \{(n-1)p^2K + 1, (n-1)p^2K + 2, \dots, np^2K\}$. For none overlapping group G_g with size p^2K, we have:

$$\begin{aligned}\frac{1}{\sqrt{p^2K}}\|vec(\boldsymbol{Z}_l'\mathbb{I}_{r_j}(s)\boldsymbol{E}), l\in G_g\|_2 &= \sqrt{\frac{1}{p^2K}\sum_{l\in G_g}|vec(\boldsymbol{Z}_l'\mathbb{I}_{r_j}(s)\boldsymbol{E})|^2} \\ &\le \max_{l\in G_g}|vec(\boldsymbol{Z}_l'\mathbb{I}_{r_j}(s)\boldsymbol{E})|,\end{aligned} \tag{64}$$

where $\boldsymbol{Z}_l'$ represents the l-th row in $\boldsymbol{Z}'$. Thus,

$$\begin{aligned}\frac{1}{\sqrt{p^2K}}\|vec(\boldsymbol{Z}'\mathbb{I}_{r_j}(s)\boldsymbol{E})\|_{2,\infty} &= \max_{g=1,\dots,n}\frac{1}{\sqrt{p^2K}}\|vec(\boldsymbol{Z}_l'\mathbb{I}_{r_j}(s)\boldsymbol{E}), l\in G_g\|_2 \\ &\le \max_{g=1,\dots,n}\max_{l\in G_g}|vec(\boldsymbol{Z}_l'\mathbb{I}_{r_j}(s)\boldsymbol{E})| \\ &= \|\boldsymbol{Z}'\mathbb{I}_{r_j}(s)\boldsymbol{E}\|_\infty.\end{aligned} \tag{65}$$

Combining the Lemma 7 and Equation (65), we have:

$$\mathbb{P}\left(\sup_{|s|\ge\gamma_n}\left(n\sigma^2(s)\right)^{-1}\|vec\left(\boldsymbol{Z}'\mathbb{I}_{r_j}(s)\boldsymbol{E}\right)\|_{2,\infty} \ge C\sqrt{p^2K}\frac{\log\left(p^2K\right)}{\sqrt{n}}\right)$$
$$\le \mathbb{P}\left(\sup_{|s|\ge\gamma_n}\left(n\sigma^2(s)\right)^{-1}\|\boldsymbol{Z}'\mathbb{I}_{r_j}(s)\boldsymbol{E}\|_\infty \ge C\frac{\log\left(p^2K\right)}{\sqrt{n}}\right)$$
$$\le c_3\exp\left(-c_1\log\left(p^2K\right)\right) + \exp\left(-c_2\frac{n^{\varkappa_1\varkappa_c/2(\varkappa_1+\varkappa_c)}}{(\log n)^{2\varkappa_1\varkappa_c/(\varkappa_1+\varkappa_c)}} + \log\left(np^2K\right)\right). \tag{66}$$

Lemma 9. *Under the Assumptions A1 to A5, for $m < m_0$, there exist constants $c_1, c_2 > 0$ such that*

$$\mathbb{P}\Bigg(\min_{\{s_1,s_2,\ldots,s_m\}\subset\{1,2,\ldots,n\}} L_n(s_1, s_2, \ldots, s_m\ \eta_n) \\ > \sum_{i=1}^{n} \left\|\epsilon_{\pi(i)}\right\|_2^2 + c_1 n\Delta_n - c_2 m d_n^{*2} \left(n\gamma_n\right)^{3/2} \Bigg) \to 1. \tag{67}$$

where $\Delta_n = \min_{1\leq j\leq m_0+1} |r_j - r_{j-1}|$.

Proof of Lemma 9: The road-map for the proof of Lemma 9 is similar to that of Lemma 4 in [41], once adapted to the TAR modeling framework.

Denote $b_{j'}$ as the order of the given j'-th estimated threshold $s_{j'}$. Since $m < m_0$, there exists a threshold r_j such that $|s_{j'} - r_j| > \Delta_n/4$. In order to find a lower bound on the sum of the least squares, without loss of generality, we try to find a lower bound for the sum of squared errors plus penalty term in the following three cases: (a) $|s_{j'} - s_{j'-1}| \leq \gamma_n$; (b) there exist two true thresholds r_j, r_{j+1} such that $|s_{j'-1} - r_j| \leq \gamma_n$ and $|s_{j'} - r_{j+1}| \leq \gamma_n$; and (c) otherwise.

Based on the Assumption A5, $\{z_t\}$ is a β-mixing process, then $I(u_0 < z_t \leq u_1)$ is an β-mixing process for fixed u_0 and u_1. By the second inequality of Theorem 1 in [35], there exists a certain positive constant c_B such that:

$$\left|\sum_{t=1}^{n} I\left(u_0 < z_t \leq u_1\right)\right| < c_B n\mathbb{E}\left|I\left(u_0 < z_t \leq u_1\right)\right| \tag{68}$$

with high probability. Since $n\mathbb{E}\left|I\left(u_0 < z_t \leq u_1\right)\right| \leq n\left(u_1 - u_0\right)$,

$$\left|\sum_{t=1}^{n} I\left(u_0 < z_t \leq u_1\right)\right| < nc_B\mathbb{E}\left|I\left(u_0 < z_t \leq u_1\right)\right| \leq nc_B\left|u_1 - u_0\right| \tag{69}$$

with high probability. Recall that according to Assumptions A1 and A5, the density of $\{\boldsymbol{\epsilon}_t\}$ and z_t are positive, so

$$\sigma^2\left(s_{j'} - s_{j'-1}\right) \geq c_0\left|s_{j'} - s_{j'-1}\right|, \tag{70}$$

where c_0 is certain positive constant.

Use $\hat{\boldsymbol{\theta}}_{s_{j'}-1,s_{j'}}$ to denote the estimated parameter in the estimated regime $(s_{j'} - 1, s_{j'}]$. Recall that $b_{j'}$ represents the order of the given j'-th estimated threshold $s_{j'}$. To simplify the notation, set $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{s_{j'}-1,s_{j'}}$. For case (a), consider the case where $r_j < s_{j'-1} < s_{j'} < r_{j+1}$. Then,

$$\sum_{i=b_{j'-1}+1}^{b_{j'}} \left\|\boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}}\boldsymbol{Y}_{\pi(i)}\right\|_2^2 \\ = \sum_{i=b_{j'-1}+1}^{b_{j'}} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + \sum_{i=b_{j'-1}}^{b_{j'}-1} \left\|\left(\boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}}\right)\boldsymbol{Y}_{\pi(i)}\right\|_2^2$$

$$
\begin{aligned}
&+2\sum_{i=b_{j'-1}+1}^{b_{j'}} \boldsymbol{Y}'_{\pi(i)}\left(\boldsymbol{A}^{(.,j+1)}-\hat{\boldsymbol{\theta}}\right)'\boldsymbol{\epsilon}_{\pi(i)}\\
=&\sum_{i=b_{j'-1}+1}^{b_{j'}}\left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2+\sum_{t=1}^{n}\left\|\left(\boldsymbol{A}^{(.,j+1)}-\hat{\boldsymbol{\theta}}\right)\boldsymbol{Y}_t\right\|_2^2 I\left(s_{j'-1}<z_t\le s_{j'}-1\right)\\
&+2\sum_{i=b_{j'-1}+1}^{b_{j'}} \boldsymbol{Y}'_{\pi(i)}\left(\boldsymbol{A}^{(.,j+1)}-\hat{\boldsymbol{\theta}}\right)'\boldsymbol{\epsilon}_{\pi(i)}\\
\ge&\sum_{i=b_{j'-1}+1}^{b_{j'}}\left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2-2\left|\sum_{i=b_{j'-1}+1}^{b_{j'}} \boldsymbol{Y}'_{\pi(i)}\left(\boldsymbol{A}^{(.,j+1)}-\hat{\boldsymbol{\theta}}\right)'\boldsymbol{\epsilon}_{\pi(i)}\right|\\
\ge&\sum_{i=b_{j'-1}+1}^{b_{j'}}\left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2-c_2\left|\sum_{i=b_{j'-1}+1}^{b_{j'}} \boldsymbol{Y}'_{\pi(i)}\boldsymbol{\epsilon}_{\pi(i)}\right|_\infty\left\|\boldsymbol{A}^{(.,j+1)}-\hat{\boldsymbol{\theta}}\right\|_1. \qquad (71)
\end{aligned}
$$

In case (a), $|s_{j'}-s_{j'-1}|\le\gamma_n$. Based on Lemma 5 and Equation (69), we can get

$$
\begin{aligned}
&\sum_{i=b_{j'-1}+1}^{b_{j'}}\left\|\boldsymbol{x}_{\pi(i)}-\hat{\boldsymbol{\theta}}\boldsymbol{Y}_{\pi(i)}\right\|_2^2\\
\ge&\sum_{i=b_{j'-1}+1}^{b_{j'}}\left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2-c_2\sqrt{n\gamma_n}\log\left(p^2K\right)\left\|\boldsymbol{A}^{(.,j+1)}-\hat{\boldsymbol{\theta}}\right\|_1.
\end{aligned} \qquad (72)
$$

According to Assumption A7, we obtain

$$
\begin{aligned}
&\sum_{i=b_{j'-1}+1}^{b_{j'}}\left\|\boldsymbol{x}_{\pi(i)}-\hat{\boldsymbol{\theta}}\boldsymbol{Y}_{\pi(i)}\right\|_2^2+\eta_{(s_{j'-1},s_{j'})}\left\|\hat{\boldsymbol{\theta}}\right\|_1\\
\ge&\sum_{i=b_{j'-1}+1}^{b_{j'}}\left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2-c_2\sqrt{n\gamma_n}\log\left(p^2K\right)\left\|\boldsymbol{A}^{(.,j+1)}\right\|_1.
\end{aligned} \qquad (73)
$$

For case (b), consider the case where $s_{j'-1}<r_j$ and $s_{j'}<r_{j+1}$.

$$
\begin{aligned}
&\frac{1}{b_{j'}-b_{j'-1}}\sum_{i=b_{j'-1}+1}^{b_{j'}}\left\|\boldsymbol{x}_{\pi(i)}-\hat{\boldsymbol{\theta}}\boldsymbol{Y}_{\pi(i)}\right\|_2^2+\eta_{(s_{j'-1},s_{j'})}\left\|\hat{\boldsymbol{\theta}}\right\|_1\\
\le&\frac{1}{b_{j'}-b_{j'-1}}\sum_{i=b_{j'-1}+1}^{b_{j'}}\left\|\boldsymbol{x}_{\pi(i)}-\boldsymbol{A}^{(.,j+1)}\boldsymbol{Y}_{\pi(i)}\right\|_2^2+\eta_{(s_{j'-1},s_{j'})}\left\|\boldsymbol{A}^{(.,j+1)}\right\|_1.
\end{aligned} \qquad (74)
$$

By rearrangement, there exist constants $c'>0$, $c_1>0$, $c_2>0$, $c_3>0$, and

$c_4 > 0$ that satisfy

$$\begin{aligned}
0 \leq & c' \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_2^2 \\
\leq & \frac{1}{n\sigma^2(s_{j'} - s_{j'-1})} \sum_{i=b_{j'-1}+1}^{b_{j'}} \boldsymbol{Y}'_{\pi(i)} \left(\boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right)' \left(\boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right) \boldsymbol{Y}_{\pi(i)} \\
\leq & \frac{2}{n\sigma^2(s_{j'} - s_{j'-1})} \sum_{i=b_{j'-1}+1}^{b_{j'}} \boldsymbol{Y}'_{\pi(i)} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{A}^{(.,j+1)} \right)' \left(\boldsymbol{x}_{\pi(i)} - \boldsymbol{A}^{(.,j+1)} \boldsymbol{Y}'_{\pi(i)} \right) \\
+ & \frac{|b_{j'} - b_{j'-1}|}{n\sigma^2(s_{j'} - s_{j'-1})} \eta_{(s_{j'-1}, s_{j'})} \left(\left\| \boldsymbol{A}^{(.,j+1)} \right\|_1 - \left\| \hat{\boldsymbol{\theta}} \right\|_1 \right) \\
\leq & \left(c_1 \frac{\log(p^2 K)}{\sqrt{n\left(s_{j'} - s_{j'-1}\right)}} + c_2 M_A d_n^* \frac{|b_{j'} - b_{j'-1}|}{n\sigma^2\left(s_{j'} - s_{j'-1}\right)} \right) \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_1 \\
& + c_3 \eta_{(s_{j'-1}, s_{j'})} \left(\left\| \boldsymbol{A}^{(.,j+1)} \right\|_1 - \left\| \hat{\boldsymbol{\theta}} \right\|_1 \right) \\
\leq & \left(c_1 \frac{\log(p^2 K)}{\sqrt{n\left(s_{j'} - s_{j'-1}\right)}} + c_2 M_A d_n^* \frac{n\gamma_n}{n\left(s_{j'} - s_{j'-1}\right)} \right) \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_1 \\
& + c_3 \eta_{(s_{j'-1}, s_{j'})} \left(\left\| \boldsymbol{A}^{(.,j+1)} \right\|_1 - \left\| \hat{\boldsymbol{\theta}} \right\|_1 \right) \\
\leq & \frac{c_3 \eta_{(s_{j'-1}, s_{j'})}}{2} \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_1 + c_3 \eta_{(s_{j'-1}, s_{j'})} \left(\left\| \boldsymbol{A}^{(.,j+1)} \right\|_1 - \left\| \hat{\boldsymbol{\theta}} \right\|_1 \right) \\
\leq & \frac{3 c_3 \eta_{(s_{j'-1}, s_{j'})}}{2} \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_{1,\mathcal{I}} - \frac{c_3 \eta_{(s_{j'-1}, s_{j'})}}{2} \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_{1,\mathcal{I}^c} \\
\leq & 2 c_3 \eta_{(s_{j'-1}, s_{j'})} \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_1 .
\end{aligned} \tag{75}$$

This implies $3 \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_{1,\mathcal{I}} \geq \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_{1,\mathcal{I}^c}$, thus, $4 \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_{1,\mathcal{I}} \geq \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_1$. By Cauchy–Schwarz inequality, we can get $4 \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_{1,\mathcal{I}} \leq 4\sqrt{d_n^*} \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_2$. In addition, we can get $\left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_2 \leq c_4 \sqrt{d_n^*} \eta_{(b_{j'-1}, b_{j'})}$ from Equation (75).

Recall that w_j denotes the j-th order of the true threshold. By Equation (75), we can use the same procedure as in the case (a). For some constants $c_{h'} > 0$ for $h' = 5, 6, \ldots, 11$, we have:

$$\begin{aligned}
\sum_{i=w_j+1}^{b_{j'}} \left\| \boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}} \boldsymbol{Y}_{\pi(i)} \right\|_2^2 \geq & \sum_{i=w_j+1}^{b_{j'}} \left\| \boldsymbol{\epsilon}_{\pi(i)} \right\|_2^2 + c_5 \left| b_{j'} - w_j \right| \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_2^2 \\
& - c_6 \sqrt{\left| n(s_{j'} - r_j) \right|} \log(p^2 K) \| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \|_1 \\
\geq & \sum_{i=w_j+1}^{b_{j'}} \left\| \boldsymbol{\epsilon}_{\pi(i)} \right\|_2^2 + c_5 \left| n(s_{j'} - r_j) \right| \left\| \boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}} \right\|_2
\end{aligned}$$

$$
\begin{aligned}
&\left(\left\|\boldsymbol{A}^{(.,j+1)}-\hat{\boldsymbol{\theta}}\right\|_2-\frac{c_6}{c_5}\frac{\log\left(p^2K\right)\sqrt{d_n^*}}{\sqrt{|n(s_{j'}-r_j)|}}\right)\\
\geq&\sum_{i=w_j+1}^{b_{j'}}\left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2-c_7\left|n(s_{j'}-r_j)\right|\left\|\boldsymbol{A}^{(.,j+1)}-\hat{\boldsymbol{\theta}}\right\|_2\\
&\left(c_8\sqrt{d_n^*}\eta_{(s_{j'-1},s_{j'})}+\frac{\log\left(p^2K\right)\sqrt{d_n^*}}{\sqrt{|n(s_{j'}-r_j)|}}\right)\\
\geq&\sum_{i=w_j+1}^{b_{j'}}\left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2-c_7\left|n(s_{j'}-r_j)\right|c_9\sqrt{d_n^*}\eta_{(s_{j'-1},s_{j'})}\\
&\left(c_8\sqrt{d_n^*}\eta_{(s_{j'-1},s_{j'})}+\frac{\log\left(p^2K\right)\sqrt{d_n^*}}{\sqrt{|n(s_{j'}-r_j)|}}\right)\\
\geq&\sum_{i=w_j+1}^{b_{j'}}\|\boldsymbol{\epsilon}_{\pi(i)}\|_2^2-c_{10}d_n^*\left(\log(p^2K)\right)^2. \qquad (76)
\end{aligned}
$$

For the threshold interval $(s_{j'-1}, r_j)$, there exist positive constants $C_{h'}$ for $h' = 1, 2, \ldots, 9$ such that

$$
\begin{aligned}
&\sum_{i=b_{j'-1}+1}^{w_j}\left\|\boldsymbol{x}_{\pi(i)}-\hat{\boldsymbol{\theta}}\boldsymbol{Y}_{\pi(i)}\right\|_2^2\\
\geq&\sum_{i=b_{j'-1}+1}^{w_j}\left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2-C_2\sqrt{n\gamma_n}\log\left(p^2K\right)\left\|\boldsymbol{A}^{(.,j)}-\hat{\boldsymbol{\theta}}\right\|_1\\
\geq&\sum_{i=b_{j'-1}+1}^{w_j}\left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2-C_2\sqrt{n\gamma_n}\log(p^2K)\left(\left\|\boldsymbol{A}^{(.,j+1)}-\hat{\boldsymbol{\theta}}\right\|_1\right.\\
&\left.+\left\|\boldsymbol{A}^{(.,j+1)}-\boldsymbol{A}^{(.,j)}\right\|_1\right)\\
\geq&\sum_{i=b_{i-1}+1}^{w_i}\left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2-C_2\sqrt{n\gamma_n}\log(p^2K)\left(d_n^*\eta_{(s_{j'-1},s_{j'})}+\left\|\boldsymbol{A}^{(.,j+1)}-\boldsymbol{A}^{(.,j)}\right\|_1\right)\\
\geq&\sum_{i=b_{j'-1}+1}^{w_j}\left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2-C_4d_n^*\left(C_1\frac{\log(p^2K)}{\sqrt{n\left|s_{j'}-s_{j'-1}\right|}}+C_3M_A\frac{\gamma_n}{|s_{j'}-s_{j'-1}|}\right)\\
&\sqrt{n\gamma_n}\log(p^2K)\\
\geq&\sum_{i=b_{j'-1}+1}^{w_j}\left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2-C_5d_n^*\sqrt{n\gamma_n}\left(\log(p^2K)\right)^2. \qquad (77)
\end{aligned}
$$

By Equation (76) and Equation (77), for certain constant $C_1' > 0$, we have:

$$\sum_{i=b_{j'-1}+1}^{b_{j'}} \left\| \boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}} \boldsymbol{Y}_{\pi(i)} \right\|_2^2 \geq \sum_{i=b_{j'-1}+1}^{b_{j'}} \left\| \boldsymbol{\epsilon}_{\pi(i)} \right\|_2^2 - C_1' d_n^* \sqrt{n\gamma_n} \left(\log \left(p^2 K \right) \right)^2 . \tag{78}$$

In case (c), $s_{j'-1} < r_j < s_{j'}$ with $|s_{j'-1} - r_j| > \Delta_n/4$ and $|s_{j'} - r_j| > \Delta_n/4$. In this case, the restricted eigenvalue condition does not hold, since the distance between two consecutive true thresholds is very large, which leads to a large distance to the intersection of the estimated thresholds. However, if the tuning parameters are chosen properly, then the deterministic part of the deviation bound argument holds. Consider threshold intervals $(s_{j'-1}, r_j)$ and $(r_j, s_{j'})$

$$\begin{aligned} &\sum_{i=b_{j'-1}+1}^{w_j} \left\| \boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}} \boldsymbol{Y}_{\pi(i)} \right\|_2^2 \\ \geq &\sum_{i=b_{j'-1}+1}^{w_j} \left\| \boldsymbol{\epsilon}_{\pi(i)} \right\|_2^2 + \sum_{i=b_{j'-1}+1}^{w_j} \left\| \boldsymbol{A}^{(.,j)} - \hat{\boldsymbol{\theta}} \right\|_2^2 \left\| \boldsymbol{Y}_{\pi(i)} \right\|_2^2 \\ &- 2 \sum_{i=b_{j'-1}+1}^{w_j} \left\| \boldsymbol{Y}_{\pi(i)} \boldsymbol{\epsilon}_{\pi(i)} \left(\boldsymbol{A}^{(.,j)} - \hat{\boldsymbol{\theta}} \right) \right\|_1 . \end{aligned} \tag{79}$$

According to the results from Lemma 7 and Equation (79), we have

$$\begin{aligned} &\sum_{i=b_{j'-1}+1}^{w_j} \left\| \boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}} \boldsymbol{Y}_{\pi(i)} \right\|_2^2 \\ \geq &\sum_{i=b_{j'-1}+1}^{w_j} \left\| \boldsymbol{\epsilon}_{\pi(i)} \right\|_2^2 + \left\| \boldsymbol{A}^{(.,j)} - \hat{\boldsymbol{\theta}} \right\|_2 \left(\sum_{i=b_{j'-1}+1}^{w_j} \left\| \boldsymbol{Y}_{\pi(i)} \right\|_2^2 \left\| \boldsymbol{A}^{(.,j)} - \hat{\boldsymbol{\theta}} \right\|_2 \right. \\ &\left. - C_6 n \sigma^2 (r_j - s_{j'-1}) \frac{\log(p^2K)\sqrt{d_n^*}}{\sqrt{n\gamma_n}} \right) \\ \geq &\sum_{i=b_{j'-1}+1}^{w_j} \left\| \boldsymbol{\epsilon}_{\pi(i)} \right\|_2^2 + C_7' \left\| \boldsymbol{A}^{(.,j)} - \hat{\boldsymbol{\theta}} \right\|_2 n \mathbb{E} \left(x_{(t-k,l)} x'_{(t-k,l)} I \left(s_{j'-1} < z_t \leq r_j \right) \right) \\ &\left(\left\| \boldsymbol{A}^{(.,j)} - \hat{\boldsymbol{\theta}} \right\|_2 - C_7 \frac{\log(p^2K)\sqrt{d_n^*}}{\sqrt{n\gamma_n}} \right) \\ \geq &\sum_{i=b_{j'-1}+1}^{w_j} \left\| \boldsymbol{\epsilon}_{\pi(i)} \right\|_2^2 + C_7' \left\| \boldsymbol{A}^{(.,j)} - \hat{\boldsymbol{\theta}} \right\|_2^2 n \mathbb{E} \left(x_{(t-k,l)} x'_{(t-k,l)} I \left(s_{j'-1} < z_t \leq r_j \right) \right) \\ \geq &\sum_{i=b_{j'-1}+1}^{w_j} \left\| \boldsymbol{\epsilon}_{\pi(i)} \right\|_2^2 + C_8 n \left(r_j - s_{j'-1} \right) \left\| \boldsymbol{A}^{(.,j)} - \hat{\boldsymbol{\theta}} \right\|_2^2 \\ \geq &\sum_{i=b_{j'-1}+1}^{w_j} \left\| \boldsymbol{\epsilon}_{\pi(i)} \right\|_2^2 + C_9 n \Delta_n, \end{aligned} \tag{80}$$

Similarly, we have

$$\sum_{i=w_j+1}^{b_{j'}} \left\|\boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}}\boldsymbol{Y}_{\pi(i)}\right\|_2^2 \geq \sum_{i=w_j+1}^{b_{j'}} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c_1' \left|b_{j'} - w_j\right| \left\|\boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}}\right\|_2^2 - c_2\sqrt{\left|b_{j'} - w_j\right|}\log(p^2K)\left\|\boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}}\right\|_1$$
$$\geq \sum_{i=w_j+1}^{b_{j'}} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c_1' \left|b_{j'} - w_j\right| \left\|\boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}}\right\|_2 \left(\left\|\boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}}\right\|_2 - \frac{c_2'}{c_1'}\frac{\log(p^2K)\sqrt{d_n^*}}{\sqrt{\left|b_{j'} - w_j\right|}}\right), \tag{81}$$

where c_1', c_2' are some positive constants.

Based on the Assumption A4, $\left\|\boldsymbol{A}^{(.,j+1)} - \boldsymbol{A}^{(.,j)}\right\|_2 \geq v > 0$, then either $\left\|\boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}}\right\|_2 \geq v/4$ or $\left\|\boldsymbol{A}^{(.,j)} - \hat{\boldsymbol{\theta}}\right\|_2 \geq v/4$. Assume that $\left\|\boldsymbol{A}^{(.,j)} - \hat{\boldsymbol{\theta}}\right\|_2 \geq v/4$. Based on Equation (79) and Equation (81), when $|s_{j'} - r_j| \leq \gamma_n$, there exist positive constants $c_{h'}'$ for $h' = 3, 4, \ldots, 8$ such that:

$$\sum_{i=b_{j'-1}}^{w_j-1} \left\|\boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}}\boldsymbol{Y}_{\pi(i)}\right\|_2^2 \geq \sum_{i=b_{j'-1}+1}^{w_j} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c_3' n\Delta_n, \tag{82}$$

and

$$\sum_{i=w_j+1}^{b_{j'}} \left\|\boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}}\boldsymbol{Y}_{\pi(i)}\right\|_2^2 \geq \sum_{i=w_j+1}^{b_{j'}} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 - c_4' d_n^* \left(\log(p^2K)\right)^2. \tag{83}$$

According to Equation (82) and Equation (83), we can get:

$$\sum_{i=b_{j'-1}+1}^{b_{j'}} \left\|\boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}}\boldsymbol{Y}_{\pi(i)}\right\|_2^2 \geq \sum_{i=b_{j'-1}+1}^{b_{j'}} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c_5' n\Delta_n - c_6' d_n^* \left(\log(p^2K)\right)^2. \tag{84}$$

When both $|s_{j'-1} - r_j| > \gamma_n$ and $|s_{j'} - r_j| > \gamma_n$, we can follow the similar steps that obtain Equation (84) and obtain:

$$\sum_{i=b_{j'-1}+1}^{b_{j'}} \left\|\boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}}\boldsymbol{Y}_{\pi(i)}\right\|_2^2 \geq \sum_{i=b_{j'-1}+1}^{b_{j'}} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c_7' n\Delta_n - c_8' \left(n\gamma_n\right)^{3/2} d_n^{*2}. \tag{85}$$

Combining above three cases (a),(b), and (c), we can prove the results.

Appendix 3: proof of main results

A.1. Proof of Theorem 1

The idea is similar to Theorem 2 in [41]. However, in our case, we do not have the assumptions related to spectral density matrices. Instead, we assume the random variables are β-mixing, sub-Weibull and stationary. For a matrix $\boldsymbol{A} \in \mathbb{R}^{pK\times p}$, let $\|\boldsymbol{A}\|_{1,\mathcal{I}} = \sum_{(i',l')\in\mathcal{I}} |a_{i'l'}|$, where $\mathcal{I}$ is the set of non-zero indices of $\boldsymbol{A}$ and $a_{i'l'}$ is the element at i'-th row and l'-th column. First, we prove the second part. For some $j = 1, 2, \ldots, m_0$, given the estimated threshold $\hat{r}_j$, $|r_j - \hat{r}_j| > \gamma_n$, there exists a true threshold point r_{j_0} which is isolated from all the estimated thresholds, i.e., $\min_{1\le j\le m_0} |\hat{r}_j - r_{j_0}| > \gamma_n$. In other words, there exists an estimated threshold $\hat{r}_j$ such that, $r_{j_0} - r_{j_0-1} \vee \hat{r}_j \ge \gamma_n$ and $r_{j_0+1} \vee \hat{r}_{j+1} - r_{j_0} \ge \gamma_n$. The idea of the proof is to show the estimated parameters in the interval $[r_{j_0-1} \vee \hat{r}_j, r_{j_0+1} \wedge \hat{r}_{j+1}]$ converges in ℓ_2 to both $\boldsymbol{A}^{(.,j_0)}$ and $\boldsymbol{A}^{(.,j_0+1)}$ which contradicts with the Assumption A4. The length of the interval is large enough to verify restricted eigenvalue and deviation bound inequalities needed to show parameter estimation consistency.

Define a new parameter sequence $\boldsymbol{\varphi}_q$, $q = 1, 2, \ldots, n$ with $\boldsymbol{\varphi}_q = \hat{\boldsymbol{\theta}}_q$ except for two thresholds $q = \hat{r}_j$ and $q = r_{j_0}$. For these two points, set $\boldsymbol{\varphi}_{\hat{r}_j} = \boldsymbol{A}^{(.,j_0)} - \hat{\boldsymbol{A}}_j$ and $\boldsymbol{\varphi}_{r_{j_0}} = \hat{\boldsymbol{A}}_{j+1} - \boldsymbol{A}^{(.,j_0)}$, where $\hat{\boldsymbol{A}}_j = \sum_{q=1}^{w_{j_0-1}\vee \hat{w}_j - 1} \hat{\boldsymbol{\theta}}_q$ and $\hat{\boldsymbol{A}}_{j+1} = \sum_{q=1}^{w_{j_0}\vee \hat{w}_j} \hat{\boldsymbol{\theta}}_q$, i.e. $\hat{\boldsymbol{\theta}}_{w_{j_0}\vee \hat{w}_j} = \hat{\boldsymbol{A}}_{j+1} - \hat{\boldsymbol{A}}_j$. Denote $\Psi = \text{vector}(\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_n) \in \mathbb{R}^{np^2K\times 1}$. By Equation (5) and focusing on the case of lasso (i.e. hdTAR), we have

$$\begin{aligned} &\frac{1}{n}\left\|\boldsymbol{Y} - \boldsymbol{Z}\hat{\Theta}\right\|_2^2 + \lambda_{1,n}\left\|\hat{\Theta}\right\|_1 + \lambda_{2,n}\sum_{i=1}^{n}\left\|\sum_{i'=1}^{i}\hat{\boldsymbol{\theta}}_{i'}\right\|_1 \\ &\le \frac{1}{n}\left\|\boldsymbol{Y} - \boldsymbol{Z}\Psi\right\|_2^2 + \lambda_{1,n}\left\|\Psi\right\|_1 + \lambda_{2,n}\sum_{i=1}^{n}\left\|\sum_{i'=1}^{i}\boldsymbol{\varphi}_{i'}\right\|_1 . \end{aligned} \tag{86}$$

By rearrangement, for a constant c, we can get

$$\begin{aligned} 0 &\le c\left\|\boldsymbol{A}^{(.,j_0)} - \hat{\boldsymbol{A}}_{j+1}\right\|_2^2 \\ &\le \frac{1}{n\sigma^2(r_{j_0} - r_{j_0-1}\vee\hat{r}_j)}\left\|\sum_{i=(w_{j_0-1}\vee\hat{w}_j)+1}^{w_{j_0}} \boldsymbol{Y}'_{\pi(i)}\left(\boldsymbol{A}^{(.,j_0)} - \hat{\boldsymbol{A}}_{j+1}\right)\right\|_2^2 \\ &\le \frac{2}{n\sigma^2(r_{j_0} - r_{j_0-1}\vee\hat{r}_j)}\sum_{i=(w_{j_0-1}\vee\hat{w}_j)+1}^{w_{j_0}} \boldsymbol{Y}'_{\pi(i)}\left(\boldsymbol{A}^{(.,j_0)} - \hat{\boldsymbol{A}}_{j+1}\right)\boldsymbol{\epsilon}_{\pi(i)} \\ &\quad + \frac{n\lambda_{1,n}}{n\sigma^2(r_{j_0} - r_{j_0-1}\vee\hat{r}_j)}\Bigg(\left\|\boldsymbol{A}^{(.,j_0)} - \hat{\boldsymbol{A}}_{j+1}\right\|_1 + \left\|\boldsymbol{A}^{(.,j_0)} - \hat{\boldsymbol{A}}_j\right\|_1 \\ &\qquad - \left\|\hat{\boldsymbol{A}}_{j+1} - \hat{\boldsymbol{A}}_j\right\|_1\Bigg) \end{aligned}$$

$$
\begin{aligned}
&+ \frac{n\lambda_{2,n}}{n\sigma^2\left(r_{j_0}-r_{j_0-1}\vee\hat{r}_j\right)} n\sigma^2\left(r_{j_0}-r_{j_0-1}\vee\hat{r}_j\right)\left(\left\|\boldsymbol{A}^{(.,j_0)}\right\|_1-\left\|\hat{\boldsymbol{A}}_{j+1}\right\|_1\right)\\
&\le \left(\frac{2n\lambda_{1,n}}{n\sigma^2\left(r_{j_0}-r_{j_0-1}\vee\hat{r}_j\right)}+C\frac{\log(p^2K)}{\sqrt{n\gamma_n}}\right)\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_1\\
&\quad + n\lambda_{2,n}\left(\left\|\boldsymbol{A}^{(.,j_0)}\right\|_1-\left\|\hat{\boldsymbol{A}}_{j+1}\right\|_1\right)\\
&\le \frac{n\lambda_{2,n}}{2}\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_1+n\lambda_{2,n}\left(\left\|\boldsymbol{A}^{(.,j_0)}\right\|_1-\left\|\hat{\boldsymbol{A}}_{j+1}\right\|_1\right)\\
&\le \frac{3n\lambda_{2,n}}{2}\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_{1,\mathcal{I}}-\frac{n\lambda_{2,n}}{2}\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_{1,\mathcal{I}^c}.
\end{aligned}
\tag{87}
$$

According to Lemma 7 and the fact that $r_{j_0}-r_{j_0-1}\vee\hat{r}_j\ge\gamma_n$, the second inequality holds with high probability converging to 1 in Equation (87). The third inequality holds because $w_{j_0}-w_{j_0-1}\vee\hat{w}_j\le c_1 n\sigma^2(s)$ for a certain positive constant c_1. The fourth inequality holds with high probability converging to 1 according to second part of Lemma 7 and triangular inequality. The fifth inequality is based on Assumption A4 and the selection for $\lambda_{2,n}$ in the statement of the theorem. The last inequality holds by Assumption A3. Thus,

$$
\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_2 = o_p\left(d_n^*\frac{\log\left(p^2K\right)}{\sqrt{n\gamma_n}}\right), \tag{88}
$$

which means that it converges to zero in probability based on Assumption A4. The convergence of $\left\|\boldsymbol{A}^{(.,j_0+1)}-\hat{\boldsymbol{A}}_{j+1}\right\|_2$ can be proved in the same procedure in the interval $[r_{j_0}, r_{j_0+1}\wedge\hat{r}_{j+1}]$, which contradicts Assumption A4. Thus, the results are as desired.

Similarly, we can prove the first part. Assume $\left|\hat{\mathcal{A}}_n\right| < m_0$. There exist an isolated true threshold r_{j_0} such that $r_{j_0}-r_{j_0-1}\vee\hat{r}_j\ge\gamma_n/3$ and $r_{j_0+1}\wedge\hat{r}_{j+1}-r_{j_0}\ge\gamma_n/3$. Similar procedure in the second part is applied to the interval $[r_{j_0-1}\vee\hat{r}_j, r_{j_0}]$ and $[r_{j_0}, r_{j_0-1}\wedge\hat{r}_{j+1}]$, then the proof is completed for the hd-TAR case.

Similar procedure can be applied to mvTAR which is briefly described next. We obtain:

$$
\begin{aligned}
0 &\le c\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_2^2\\
&\le \frac{1}{n\sigma^2(r_{j_0}-r_{j_0-1}\vee\hat{r}_j)}\left\|\sum_{i=(w_{j_0-1}\vee\hat{w}_j)+1}^{w_{j_0}}\boldsymbol{Y}'_{\pi(i)}\left(\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right)\right\|_2^2\\
&\le \frac{2}{n\sigma^2(r_{j_0}-r_{j_0-1}\vee\hat{r}_j)}\sum_{i=(w_{j_0-1}\vee\hat{w}_j)+1}^{w_{j_0}}\boldsymbol{Y}'_{\pi(i)}\left(\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right)\boldsymbol{\epsilon}_{\pi(i)}\\
&\quad + \frac{n\lambda_{1,n}}{n\sigma^2(r_{j_0}-r_{j_0-1}\vee\hat{r}_j)}\left(\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_2+\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_j\right\|_2\right.\\
&\qquad\left.-\left\|\hat{\boldsymbol{A}}_{j+1}-\hat{\boldsymbol{A}}_j\right\|_2\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{n\lambda_{2,n}}{n\sigma^2\left(r_{j_0}-r_{j_0-1}\vee\hat{r}_j\right)}n\sigma^2\left(r_{j_0}-r_{j_0-1}\vee\hat{r}_j\right)\left(\left\|\boldsymbol{A}^{(.,j_0)}\right\|_1-\left\|\hat{\boldsymbol{A}}_{j+1}\right\|_1\right)\\
\le\ &\frac{2}{n\sigma^2(r_{j_0}-r_{j_0-1}\vee\hat{r}_j)}\left\|\boldsymbol{Z}'\mathbb{I}_{r_j}(r_{j_0}-r_{j_0-1}\vee\hat{r}_j))\boldsymbol{E}\right\|_{2,\infty}\left\|\left(\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right)\right\|_{2,1}\\
&+\frac{n\lambda_{1,n}}{n\sigma^2(r_{j_0}-r_{j_0-1}\vee\hat{r}_j)}\Bigg(\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_2+\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_j\right\|_2\\
&\qquad\qquad\qquad\qquad\qquad-\left\|\hat{\boldsymbol{A}}_{j+1}-\hat{\boldsymbol{A}}_j\right\|_2\Bigg)\\
&+\frac{n\lambda_{2,n}}{n\sigma^2\left(r_{j_0}-r_{j_0-1}\vee\hat{r}_j\right)}n\sigma^2\left(r_{j_0}-r_{j_0-1}\vee\hat{r}_j\right)\left(\left\|\boldsymbol{A}^{(.,j_0)}\right\|_1-\left\|\hat{\boldsymbol{A}}_{j+1}\right\|_1\right)\\
\le\ &\left(\frac{2n\lambda_{1,n}}{n\sigma^2\left(r_{j_0}-r_{j_0-1}\vee\hat{r}_j\right)}+C\frac{\sqrt{p^2K}\log(p^2K)}{\sqrt{n\gamma_n}}\right)\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_2\\
&+n\lambda_{2,n}\left(\left\|\boldsymbol{A}^{(.,j_0)}\right\|_1-\left\|\hat{\boldsymbol{A}}_{j+1}\right\|_1\right)\\
\le\ &\left(\frac{2n\lambda_{1,n}}{n\sigma^2\left(r_{j_0}-r_{j_0-1}\vee\hat{r}_j\right)}+C\frac{\sqrt{p^2K}\log(p^2K)}{\sqrt{n\gamma_n}}\right)\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_1\\
&+n\lambda_{2,n}\left(\left\|\boldsymbol{A}^{(.,j_0)}\right\|_1-\left\|\hat{\boldsymbol{A}}_{j+1}\right\|_1\right)\\
\le\ &\frac{n\lambda_{2,n}}{2}\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_1+n\lambda_{2,n}\left(\left\|\boldsymbol{A}^{(.,j_0)}\right\|_1-\left\|\hat{\boldsymbol{A}}_{j+1}\right\|_1\right)\\
\le\ &\frac{3n\lambda_{2,n}}{2}\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_{1,\mathcal{I}}-\frac{n\lambda_{2,n}}{2}\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_{1,\mathcal{I}^c}.
\end{aligned}
\tag{89}
$$

Note that there is only one group in $\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_{2,1}$, so it is $\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_2$. According to the first part in Lemma 7 and the fact that $r_{j_0}-r_{j_0-1}\vee\hat{r}_j\ge\gamma_n$, the second inequality holds with high probability converging to 1 in Equation (89). The third inequality holds because $w_{j_0}-w_{j_0-1}\vee\hat{w}_j\le c_1n\sigma^2(s)$ for a certain positive constant c_1. The fifth inequality holds with high probability converging to 1 according to Lemma 8 and triangular inequality. The sixth inequality holds due to Minkowski inequality. The seventh inequality is based on Assumption A4 and the selection for $\lambda_{2,n}$ in the statement of the theorem. The last inequality holds by Assumption A3. Thus,

$$
\left\|\boldsymbol{A}^{(.,j_0)}-\hat{\boldsymbol{A}}_{j+1}\right\|_2=o_p\left(d_n^*\frac{\sqrt{p^2K}\log\left(p^2K\right)}{\sqrt{n\gamma_n}}\right),
\tag{90}
$$

The remaining parts are similar to hdTAR case, hence details are omitted to avoid duplication. This completes the proof for both cases.

A.2. Proof of Theorem 2

For the first part, we want to prove that $\mathbb{P}\left(\tilde{m}<m_0\right)\to0$, and $\mathbb{P}\left(\tilde{m}>m_0\right)\to0$. From Theorem 1, we know that there exist estimated thresholds $\hat{r}_j\in\hat{\mathcal{A}}_n$ such that $\max_{1\le j\le m_0}|\hat{r}_j - r_j| \le \gamma_n$, where $r_j \in \mathcal{A}_n$. Recall that w_j denotes the j-th order of the thresholds, and $b_{j'}$ denotes the j'-th order of the estimated threshold.

Without loss of generality, we only show one of the estimated regimes. For $s_{j'-1} < r_j < s_i$ with $|r_j - s_{j'-1}| \le \gamma_n$, the estimated coefficient is denoted as $\hat{\boldsymbol{\theta}}$ in $(s_{j'-1}, s_{j'})$. Similar to case (b) in the proof of Lemma 9, recall that $|b_{j'} - w_j| \le nc_B |s_{j'} - r_j|$ according to Equation (69). For the threshold interval $(r_j, s_{j'})$, we have

$$\begin{aligned}
&\sum_{i=w_j+1}^{b_{j'}} \left\|\boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}}\boldsymbol{Y}_{\pi(i)}\right\|_2^2 \\
\le &\sum_{i=w_j+1}^{b_{j'}} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c_3 |b_{j'} - w_j| \left\|\boldsymbol{A}^{(\cdot,j+1)} - \hat{\boldsymbol{\theta}}\right\|_2^2 \\
&+ c_4\sqrt{|b_{j'} - w_j|}\log(p^2K)\left\|\boldsymbol{A}^{(\cdot,j+1)} - \hat{\boldsymbol{\theta}}\right\|_1 \\
\le &\sum_{i=w_j+1}^{b_{j'}} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 \\
&+ c_5 n |s_{j'} - r_j| d_n^* \left(c_1 \frac{\log(p^2K)}{\sqrt{n(s_{j'} - s_{j'-1})}} + c_2 M_A d_n^* \frac{\gamma_n}{s_{j'} - s_{j'-1}}\right)^2 \\
&+ c_6\sqrt{|b_i - w_j|}\log\left(p^2K\right) d_n^* \left(c_1 \frac{\log(p^2K)}{\sqrt{n(s_{j'} - s_{j'-1})}} + c_2 M_A d_n^* \frac{\gamma_n}{s_{j'} - s_{j'-1}}\right) \\
\le &\sum_{i=w_j+1}^{b_{j'}} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c_7\sqrt{n\gamma_n}\left(\log\left(p^2K\right) d_n^*\right)^2 \\
\le &\sum_{i=w_j+1}^{b_{j'}} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c\left(n\gamma_n\right)^{3/2} d_n^{*2},
\end{aligned} \tag{91}$$

where c, $c_{h'}$ are positive constants for $h' = 1, 2, \dots, 7$.

Let $c'_{h'}$ be positive constants for $h' = 1, 2, \dots, 6$. For regime (s_{i-1}, r_j), we can get

$$\begin{aligned}
&\sum_{i=b_{j'-1}+1}^{w_j} \left\|\boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}}\boldsymbol{Y}_{\pi(i)}\right\|_2^2 \\
\le &\sum_{i=b_{j'-1}+1}^{w_j} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c'_1 |w_j - b_{j'-1}| \left\|\boldsymbol{A}^{(\cdot,j)} - \hat{\boldsymbol{\theta}}\right\|_2^2 \\
&+ c'_2\sqrt{|w_j - b_{j'-1}|}\log\left(p^2K\right)\left\|\boldsymbol{A}^{(\cdot,j)} - \hat{\boldsymbol{\theta}}\right\|_1
\end{aligned}$$

$$\leq \sum_{i=b_{j'-1}+1}^{w_j} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c_1' \left|w_j - b_{j'-1}\right| \left(\left\|\boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}}\right\|_2^2 + \left\|\boldsymbol{A}^{(.,j+1)} - \boldsymbol{A}^{(.,j)}\right\|_2^2 \right)$$
$$+ c_2' \sqrt{\left|w_j - b_{j'-1}\right|} \log\left(p^2 K\right) \left(\left\|\boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}}\right\|_1 + \left\|\boldsymbol{A}^{(.,j+1)} - \boldsymbol{A}^{(.,j)}\right\|_1 \right)$$
$$\leq \sum_{i=b_{j'-1}+1}^{w_j} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c_3' d_n^{*2} \sqrt{n\gamma_n} \left(\log\left(p^2 K\right)\right)^2$$
$$\leq \sum_{i=w_j+1}^{b_{j'}} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c_4' \left(n\gamma_n\right)^{3/2} d_n^{*2}. \tag{92}$$

Since

$$\eta_{\left(s_{j'-1}, s_{j'}\right)} \left\|\hat{\boldsymbol{\theta}}\right\|_1 \leq \eta_{\left(s_{j'-1}, s_{j'}\right)} \left(\left\|\boldsymbol{A}^{(.,j+1)} - \hat{\boldsymbol{\theta}}\right\|_1 + \left\|\boldsymbol{A}^{(.,j+1)}\right\|_1 \right) \leq c_5' d_n^*, \tag{93}$$

we combine Equation (91) to Equation (93) and get

$$\sum_{i=b_{j'-1}}^{b_{j'}-1} \left\|\boldsymbol{x}_{\pi(i)} - \hat{\boldsymbol{\theta}} \boldsymbol{Y}_{\pi(i)}\right\|_2^2 + \eta_{(b_{j'-1}, b_{j'})} \left\|\hat{\boldsymbol{\theta}}\right\|_1 \leq \sum_{i=b_{j'-1}}^{b_{j'}-1} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c_6' \left(n\gamma_n\right)^{3/2} d_n^{*2}. \tag{94}$$

Since there are $m_0 + 1$ regimes, we can get:

$$L_n(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_{m_0}; \eta_n) \leq \sum_{i=1}^{n} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c_7' m_0 \left(n\gamma_n\right)^{3/2} d_n^{*2}. \tag{95}$$

Given subset from the candidate thresholds found in Step 1. Let $C_{h'}$ be positive constants for $h' = 1, 2, \ldots, 6$. Assume $\tilde{m} < m_0$. By Lemma 9, we can get

$$\begin{aligned}
&\mathrm{IC}\left(\tilde{r}_1, \ldots, \tilde{r}_{\tilde{m}}\right) \\
&= L_n\left(\tilde{r}_1, \ldots, \tilde{r}_{\tilde{m}}; \eta_n\right) + \tilde{m}\omega_n \\
&> \sum_{i=1}^{n} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + C_1 n\Delta_n - C_2 \tilde{m} d_n^{*2} \left(n\gamma_n\right)^{3/2} + \tilde{m}\omega_n \\
&\geq L_n(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_{m_0}; \eta_n) + m_0\omega_n + C_1 n\Delta_n - C_2 \tilde{m} d_n^{*2} \left(n\gamma_n\right)^{3/2} \\
&\quad - C_3 m_0 \left(n\gamma_n\right)^{3/2} d_n^{*2} - \left(m_0 - \tilde{m}\right)\omega_n \\
&\geq L_n(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_{m_0}; \eta_n) + m_0\omega_n + C_1 n\Delta_n - C_4 m_0 \left(n\gamma_n\right)^{3/2} d_n^{*2} - (m_0 - \tilde{m})\omega_n.
\end{aligned} \tag{96}$$

According to Assumption A6, we have

$$m_0 \left(n\gamma_n\right)^{3/2} d_n^{*2} / \omega_n \to 0 \text{ and } m_0 \omega_n / n\Delta_n \to 0.$$

Then,

$$C_1 n\Delta_n - C_4 m_0 \left(n\gamma_n\right)^{3/2} d_n^{*2} - \left(m_0 - \tilde{m}\right)\omega_n \geq 0. \tag{97}$$

Thus, $\mathrm{IC}(\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_{\tilde{m}}) \geq L_n\left(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_{m_0}\,; \eta_n\right) + m_0\omega_n$, which proves

$$\mathbb{P}\left(\tilde{m} < m_0\right) \to 0.$$

Next, we want to prove $\mathbb{P}(\tilde{m} > m_0) \to 0$. Similar procedure in Lemma 9 can be used to get:

$$L_n\left(\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_{\tilde{m}}\,; \eta_n\right) \geq \sum_{i=1}^{n} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 - C_5 \tilde{m} d_n^{*2} \left(n\gamma_n\right)^{3/2}. \tag{98}$$

Then,

$$\begin{aligned}
\sum_{i=1}^{n} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 - C_5 \tilde{m} d_n^{*2} \left(n\gamma_n\right)^{3/2} + \tilde{m}\omega_n &\leq \mathrm{IC}(\tilde{r}_1, \tilde{r}_2, \cdots, \tilde{r}_{\tilde{m}}) \\
&\leq \mathrm{IC}(\hat{r}_1, \hat{r}_2, \cdots, \hat{r}_{m_0}) \\
&\leq \sum_{i=1}^{n} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + C_6 m_0 \left(n\gamma_n\right)^{3/2} d_n^{*2} \\
&\quad + m_0\omega_n.
\end{aligned} \tag{99}$$

Thus,

$$\left(\tilde{m} - m_0\right)\omega_n \leq C_5 \left(n\gamma_n\right)^{3/2} \tilde{m} d_n^{*2} + C_6 m_0 \left(n\gamma_n\right)^{3/2} d_n^{*2}. \tag{100}$$

If $\tilde{m} > m_0$, it contradicts assumption that $m_0 \left(n\gamma_n\right)^{3/2} d_n^{*2}/\omega_n \to 0$. Then, we can get $\mathbb{P}\left(\tilde{m} - m_0\right) \to 1$.

Next, we prove $\mathbb{P}\left(\max_{1\leq j\leq m_0} |\tilde{r}_j - r_j| \leq B m_0 \left(\gamma_n\right)^{3/2} d_n^{*2}\sqrt{n}\right)$. Given certain two constants $C_7 > 0$ and $c' > 0$, let $B = 2C_7/c'$. Suppose there exists a threshold r_j such that $\min_{1\leq j\leq m_0} |\tilde{r}_j - r_j| \geq B m_0 \left(\gamma_n\right)^{3/2} d_n^{*2}\sqrt{n}$. Applying similar procedure to Lemma 9, we can get:

$$\begin{aligned}
\sum_{i=1}^{n} \left\|\boldsymbol{\epsilon}_{\pi(i)}\right\|_2^2 + c' B m_0 \left(n\gamma_n\right)^{3/2} d_n^{*2} &\leq L_n(\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_{m_0}\,; \eta_n) \\
&\leq L_n(\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_{m_0}\,; \eta_n) \\
&\leq \sum_{i=1}^{n} \|\boldsymbol{\epsilon}_{\pi(i)}\|_2^2 + C_7 m_0 \left(n\gamma_n\right)^{3/2} d_n^{*2},
\end{aligned} \tag{101}$$

which contradicts the value of B.

A.3. Proof of Theorem 3

Theorem 3 can be proved according to the modification of Corollary 9 in [54]. In Corollary 9 [54], we know that for regime j, we have

$$\|\mathrm{vec}(\hat{\boldsymbol{\beta}}^{(\cdot,j)}) - \mathrm{vec}(\boldsymbol{A}^{(\cdot,j)})\|_2 \leq c_1 \alpha_j \sqrt{d_n^*} \tag{102}$$

with high probability, where α_j represent a tuning parameter determined by p, K, and the number of observations in each regime j. Let $\tilde{w}_j$ be the order of estimated threshold $\tilde{r}_j$. What is left is to find a lower bound on the number of observations. Due to Assumption A5, z_t has positive density. Combining Assumption A5 and Corollary 9 of Wong et al. [54], we have:

$$\begin{aligned}\tilde{w}_{j-1} - \tilde{w}_j &= n\mathbb{P}\left(\tilde{r}_{j-1} < z_t \le \tilde{r}_j\right) \\ &\ge c_2 n \left|\tilde{r}_j - \tilde{r}_{j-1}\right|,\end{aligned} \tag{103}$$

where $c_2 > 0$ is a constant. Now, by plugging in the optimal value of α_j, we get

$$\begin{aligned}\left\|\hat{\boldsymbol{\beta}}^{(.,j)} - \boldsymbol{A}^{(.,j)}\right\|_2 &\le c_3 \sqrt{\frac{d_n^* \log\left(p^2 K\right)}{\left(\tilde{w}_{j-1} - \tilde{w}_j\right)}} \\ &\le c_4 \sqrt{\frac{d_n^* \log\left(p^2 K\right)}{n\gamma_n}},\end{aligned} \tag{104}$$

where $c_3, c_4 > 0$ are constants.

Appendix 4: A sufficient condition for β-mixing

In this section, we provide a sufficient condition for the TAR process $\boldsymbol{x}_t$ to be β-mixing by imposing a restriction on the operator norm of transition matrices. To that end, note that the TAR process,

$$\boldsymbol{x}_t = \sum_{k=1}^{K} \boldsymbol{A}^{(k,j)} \boldsymbol{x}_{t-k} + \boldsymbol{\epsilon}_t, \tag{105}$$

can be rewritten as a Kp-dimensional TAR process with lag 1; that is,

$$\tilde{\boldsymbol{X}}_t = \tilde{\boldsymbol{B}}^{(j)} \tilde{\boldsymbol{X}}_{t-1} + \tilde{\boldsymbol{U}}_t, \tag{106}$$

where $\tilde{\boldsymbol{X}}_t = \begin{pmatrix}\boldsymbol{x}_t' & \boldsymbol{x}_{t-1}' & \dots & \boldsymbol{x}_{t-K+1}'\end{pmatrix}' \in \mathbb{R}^{Kp\times 1}$,

$$\tilde{\boldsymbol{B}}^{(j)} = \begin{pmatrix} \boldsymbol{A}^{(1,j)} & \boldsymbol{A}^{(k,j)} & \cdots & \boldsymbol{A}^{(K-1,j)} & \boldsymbol{A}^{(K,j)} \\ \boldsymbol{I}_p & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_p & & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & & \ddots & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_p & \boldsymbol{0} \end{pmatrix} \in \mathbb{R}^{Kp\times Kp}$$

for $\boldsymbol{I}_p$ is a $p \times p$ identity matrix, and $\tilde{\boldsymbol{U}}_t = \begin{pmatrix}\boldsymbol{\epsilon}_t' & \boldsymbol{0} & \dots & \boldsymbol{0}\end{pmatrix}' \in \mathbb{R}^{KP\times 1}$. Let $\tilde{\boldsymbol{B}}_{\max} = \arg\max_{j=1,\dots,m_0+1} \left|\left|\left|\tilde{\boldsymbol{B}}^{(j)}\right|\right|\right|$ where $\left|\left|\left|\tilde{\boldsymbol{B}}\right|\right|\right|$ denotes the operator norm of matrix $\tilde{\boldsymbol{B}}$; that is $\left|\left|\left|\tilde{\boldsymbol{B}}\right|\right|\right| = \sqrt{\lambda_{\max}\left(\tilde{\boldsymbol{B}}'\tilde{\boldsymbol{B}}\right)}$.

Lemma 10. *For the TAR model in Equation (105), if* $\left|\left|\left|\tilde{\boldsymbol{B}}_{\max}\right|\right|\right| < 1$*, then* $\boldsymbol{x}_t$ *is* β*-mixing with a geometrically decaying mixing coefficient. If, in addition,* $\boldsymbol{\epsilon}_t$ *follows a sub-Weibull distribution, then* $\boldsymbol{x}_t$ *is also sub-Weibull. In other words, Assumption A2 holds.*

Remark 2. *The condition based on the operator norm of transition matrices may not be the optimal for* $\boldsymbol{x}_t$ *to be* β*-mixing and sub-Weibull, and a condition based on the spectral norm could be less restrictive. However, a condition based on the spectral norm does not seem achievable as the argument used for VAR models does not hold in this case. Specifically, in VAR models, we have a sufficient condition based on the spectral norm according to Lemma 8.2 in Fan and Yao [16] stating that the geometric Ergodicity of any subsequence with deterministic index entails the geometric Ergodicity of the original series. But this result does not hold for the TAR models, as the index of the sub-sequence in the TAR model is not deterministic.*

Proof of Lemma 10: The proof of Lemma 10 is similar to that in Appendix E.1 of [54]. We mainly need to apply Proposition 1 and Proposition 2 in [30] and the fact that any measurable function of a stationary process is β-mixing if the original stationary process is β-mixing. Proposition 1 in [30] gives the result that the sequence is geometrically Ergodic based on certain conditions, and we can show that the sequence will be β-mixing with geometrically decaying mixing coefficients, by using Proposition 2 in [30]. Finally, we verify the sub-Weibull assumption by using the definition of sub-Weibull distributions.

To apply Proposition 1 in [30], we check the three conditions, where we set the corresponding parameters $E = \mathbb{R}^p$, and μ as the Lebesgue measure. Condition (i) is satisfied if we set the parameter m in the Proposition 1 of [30] to 1. (Note that here m is not the number of thresholds.) For condition (ii), we set $\bar{m} = \lceil \inf_{u \in C, v \in A} \|u - v\|_2 \rceil$ the minimum "distance" between the sets C and set A in [30], where A is any set that $A \in \mathcal{B}$ where $\mathcal{B}$ is the σ-algebra of Borel sets of E, and C is any compact set that $C \subset E$. Since C is bounded and A is Borel, $\bar{m}$ is finite. For condition (iii), the function $Q(\cdot) = \|\cdot\|$ and set $K_c = \{x \in \mathbb{R}^p : \|x\| \le \frac{4C_{ac}}{c}\}$ where $c = 1 - \left|\left|\left|\tilde{\boldsymbol{B}}_{\max}\right|\right|\right|$ and $C_{ac} := \mathbb{E}\|\boldsymbol{\epsilon}_t\|$. Since $\max_{j=1,2,\ldots,m_0+1} \left|\left|\left|\tilde{\boldsymbol{B}}^{(j)}\right|\right|\right| < 1$,

- For all $\tilde{y} \in E \setminus K_c$; i.e. $\tilde{y}$ such that $\|\tilde{y}\| > \frac{4C_{ac}}{c}$,

$$\begin{aligned}
\mathbb{E}\left[\|\tilde{\boldsymbol{X}}_{t+1}\| | \tilde{\boldsymbol{X}}_t = \tilde{y}\right] &= \mathbb{E}_{z_t}\left[\mathbb{E}\left[\|\tilde{\boldsymbol{X}}_{t+1}\| | \tilde{\boldsymbol{X}}_t = \tilde{y}, z_t\right]\right] \\
&\le \mathbb{E}_{z_t}\left[\left|\left|\left|\tilde{\boldsymbol{B}}_{\max}\right|\right|\right| \|\tilde{y}\| + \mathbb{E}\|\boldsymbol{\epsilon}_t\|\right] \\
&= \left|\left|\left|\tilde{\boldsymbol{B}}_{\max}\right|\right|\right| \|\tilde{y}\| + \mathbb{E}\|\boldsymbol{\epsilon}_t\|
\end{aligned}$$

$$\equiv (1-c)\|\tilde{y}\| + C_{ac}$$
$$< (1-\frac{c}{2})\|\tilde{y}\| - C_{ac}.$$

- For all $\tilde{y} \in K_c$,

$$\begin{aligned}
\mathbb{E}\left[\|\tilde{\boldsymbol{X}}_{t+1}\| | \tilde{\boldsymbol{X}}_t = \tilde{y}\right] &= \mathbb{E}_{z_t}\left[\mathbb{E}\left[\|\tilde{\boldsymbol{X}}_{t+1}\| | \tilde{\boldsymbol{X}}_t = \tilde{y}, z_t\right]\right] \\
&< \mathbb{E}_{z_t}\left[\left|\!\left|\!\left|\tilde{\boldsymbol{B}}_{\max}\right|\!\right|\!\right| \|\tilde{y}\| + C_{ac}\right] \\
&\le \frac{4C_{ac}(1-c)}{c} + C_{ac}.
\end{aligned}$$

- For all $\tilde{y} \in K_C$,

$$0 \le \|\tilde{y}\| \le \frac{4C_{ac}}{c}.$$

By Proposition 1 in [30], $\tilde{\boldsymbol{X}}_t$ is geometrically Ergodic and stationary. By Proposition 2 in [30], the sequence will be β-mixing with geometrically decaying mixing coefficients.

Next, we verify the sub-Weibull distribution. Let $\varkappa$ be the sub-Weibull parameter associated with $\tilde{\boldsymbol{U}}_t$ in (106). Since

$$\|\tilde{\boldsymbol{X}}_t\|_\psi \le \left|\!\left|\!\left|\tilde{\boldsymbol{B}}_{\max}\right|\!\right|\!\right| \|\tilde{\boldsymbol{X}}_{t-1}\|_\psi + \|\tilde{\boldsymbol{U}}_{t-1}\|_\psi,$$

and $\left|\!\left|\!\left|\tilde{\boldsymbol{B}}_{\max}\right|\!\right|\!\right| < 1$,

$$\|\tilde{\boldsymbol{X}}_t\|_\psi \le \frac{\|\boldsymbol{\epsilon}_t\|_\psi}{1 - \left|\!\left|\!\left|\tilde{\boldsymbol{B}}_{\max}\right|\!\right|\!\right|} < \infty.$$

Now, given that $\tilde{\boldsymbol{X}}_t$ is an (equivalent) representation for $\boldsymbol{x}_t$, it follows that $\boldsymbol{x}_t$ is also sub-Weibull. Therefore, Assumption A2 holds.

Appendix 5: Algorithms

In this section, we present two algorithms for solving the optimization Equation (5). In high dimension, we use Algorithm 1, while in moderate dimension, we use Algorithm 2. Let $S(\cdot;\lambda)$ be the element-wise soft thresholding operator. Recall that throughout the paper, for a $m \times n$ matrix A, $\|A\|_\infty = \max_{1\le i\le m, 1\le j\le n} |a_{ij}|$. The algorithms are as follows:

Initialize $\boldsymbol{\theta}_i = 0$, for $i = 1, \cdots, n$.;
while $h <$ *maximum iteration* **do**
  for $i = 1, \cdots, n$ **do**
    Calculate the $(h+1)$th iteration of $\boldsymbol{\theta}_i^{h+1}$ by KKT condition:

$$\boldsymbol{\theta}_i^{'(h+1)} = \left(\sum_{l=i}^{n} \boldsymbol{Y}_{\pi(l)} \boldsymbol{Y}'_{\pi(l)}\right)^{-1} S\left(\sum_{l=i}^{n} \boldsymbol{Y}_{\pi(l)} \boldsymbol{x}'_{\pi(l)} - \sum_{j\neq i}\left(\sum_{l=\max(i,j)} \boldsymbol{Y}_{\pi(l)} \boldsymbol{Y}'_{\pi(l)}\right) \boldsymbol{\theta}_j^{'(h)} ; \lambda_1\right),$$

    where $\boldsymbol{Y}'_{\pi(l)} = \left(\boldsymbol{x}_{\pi(l)}, \cdots, \boldsymbol{x}_{\pi(l)-K+1}\right)_{1\times pK}$ and

$$S(y\,;\lambda) = \begin{cases} y-\lambda & \text{if } y > \lambda \\ y+\lambda & \text{if } y < -\lambda \\ 0 & \text{otherwise} \end{cases}.$$

  end
  if $\max_{1\le i\le n} \|\boldsymbol{\theta}_i^{(h+1)} - \boldsymbol{\theta}_i^{(h)}\|_\infty < \delta$, *where* δ *is the tolerance set to* $2e^{-4}$ *in the paper* **then**
    Stop and denote the final estimate by $\boldsymbol{\Theta}^{(\text{intermediate})}$.
  end
end
Apply soft-thresholding to the partial sums of $\boldsymbol{\Theta}^{(\text{intermediate})}$, i.e. $\sum_{i=1}^{k} \boldsymbol{\theta}_i^{(\text{intermediate})}$ to find the optimizer in Equation (5). In other words, $\widehat{\boldsymbol{\theta}}_1 = S\left(\boldsymbol{\theta}_1^{(\text{intermediate})}; \lambda_2\right)$ and
$\widehat{\boldsymbol{\theta}}_k = S\left(\sum_{i=1}^{k} \boldsymbol{\theta}_i^{(\text{intermediate})}; \lambda_2\right) - S\left(\sum_{i=1}^{k-1} \boldsymbol{\theta}_i^{(\text{intermediate})}; \lambda_2\right)$ for $k = 2, 3, \ldots, n$. Finally $\widehat{\boldsymbol{\Theta}} = \left(\widehat{\boldsymbol{\theta}}_1, \ldots, \widehat{\boldsymbol{\theta}}_n\right)$.

Algorithm 1: The fused lasso algorithms

Initialize $\boldsymbol{\theta}_i = 0$, for $i = 1, \cdots, n.$;
while $h <$ *maximum iteration* **do**
  for $i = 1, \cdots, n$ **do**
    Calculate the $(h+1)$th iteration of $\boldsymbol{\theta}_i^{h+1}$: Let

$$\Omega = \boldsymbol{\theta}_i^{'(h)} + \gamma \left(\sum_{l=i}^{n} \boldsymbol{Y}_{\pi(l)} \boldsymbol{x}'_{\pi(l)} - \sum_{j \neq i} \left(\sum_{l=\max(i,j)} \boldsymbol{Y}_{\pi(l)} \boldsymbol{Y}'_{\pi(l)} \right) \boldsymbol{\theta}_j^{'(h)} - \sum_{l=i}^{n} \left(\boldsymbol{Y}_{\pi(l)} \boldsymbol{Y}'_{\pi(l)} \right) \boldsymbol{\theta}_i^{'(h)} \right)$$

$$\begin{aligned} \boldsymbol{\theta}_i^{'(h+1)} &= \frac{1}{2\gamma} \arg\min_U \|U - \Omega\|_2^2 + \|\Omega\|_2 \\ &= \left(1 - \frac{\gamma \lambda_1}{\|U\|_2}\right)_+ \Omega \end{aligned} \tag{107}$$

    .
  end
  if $\max_{1 \le i \le n} \|\boldsymbol{\theta}_i^{(h+1)} - \boldsymbol{\theta}_i^{(h)}\|_\infty < \delta$, *where* δ *is the tolerance set to* $2e^{-4}$ *in the paper* **then**
    Stop and denote the final estimate by $\boldsymbol{\Theta}^{(\text{final})}$.
  end
end

Algorithm 2: The group lasso algorithms

Appendix 6: extended literature review

In this section, we summarise the existing methods for estimating multivariate TARs, along with their treatment of the number of thresholds m_0 and dimension of the TAR model.

Table 3 highlights the limitations of existing approaches and the fact that our methods are the only available approach that can handle both low- and high-dimensional settings, while allowing for an *unknown* number of thresholds that could diverge with the number of observations T. Allowing for an unknown number of thresholds amounts significantly complicates the problem as previous approaches for multivariate TAR models need to first estimate the number of thresholds and then proceed with estimating the location of thresholds. An incorrect estimation of number of thresholds in the first step may result in biased estimation of thresholds due to having misspecified components in the estimation procedure.

As seen in Table 3, many methods assume that the number of thresholds is known, even though this information is often not available in practice. Thus,

Table 3
Comparison of existing methods for estimating multivariate TAR models. Here m_0 represents the number of thresholds and T the length of the time series.

Paper	m_0	m_0 assumed known?	Dimension
Tsay [48]	finite (at most three)	No	low
Lo and Zivot [33] (TVAR)	at most 2	Yes	low
Hansen and Seo [21]	1	Yes	low (bi-variate)
Nieto [36]	finite	No	low (bi-variate)
Dueker et al. [14]	finite	Yes	low
Li and Tong [28]	1	Yes	low
Calderón V and Nieto [7]	at most 3	No	low
Orjuela and Villanueva [37]	finite	No	low
Our method	diverging with T	No	moderate & high

in the remainder of this section we discuss how existing approaches treat the number of thresholds.

Utilizing this assumption, Tsay [48] performs a grid search, estimating the coefficient using simple linear model for each interval, and selecting the threshold based on the Akaike information criterion (AIC). Lo and Zivot [33] instead assume the model as at most 2 thresholds. While a relaxation compared to a known number of thresholds, this assumption still considerably simplifies the problem. Using this assumption, Lo and Zivot [33] use nested hypothesis tests (testing whether the data can be modeled by the linear model versus a TAR model) to detect the thresholds, and apply the grid search method to estimate the values of the thresholds based on the results of the hypothesis testing. As an alternative, Hansen and Seo [21] couples the grid search with a maximum likelihood estimation (MLE) of the model parameters. However, the algorithm is difficult to implement in higher dimensions, and the consistency and/or distribution of the MLE estimator is not investigated. Dueker et al. [14] restricts the switching variable to be constructed based on the lags of the original time series that is being modeled and performs a grid search with respected to certain log likelihood function. The key advantage of this method is that it allows for multiple switching variables, but with only one threshold for each switching variable. Li and Tong [28] provides a nested sub-sample search algorithm to reduce the time complexity of the grid search.

A few methods have recently tried to estimate multivariate TAR models under less restrictive assumptions on the number of thresholds. However, these methods can only handle finite number of thresholds or only work in low-dimensional settings. To our knowledge, Nieto [36], Calderón V and Nieto [7] and Calderón V and Nieto [7] are the only methods that do not require a known number of thresholds or a bound on the number of thresholds. This is achieved by utilizing a Bayesian estimation framework. However, the consistency of the number of estimated thresholds is not investigated for these Bayesian methods, which could be a challenging problem. Our proposed methods and the corresponding theory thus bridge a gap in the existing literature, as the only methods that allow for an unknown and diverging number of thresholds, m_0, while also facilitating estimation of moderate and high-dimensional time series.

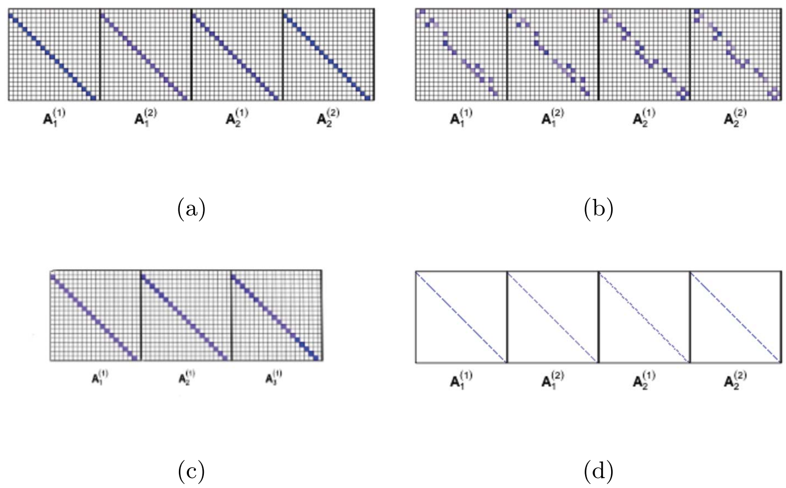


Fig 6. *Images of true auto-regressive coefficients in different simulation scenarios considered. (a): The two regimes in Simulation Scenario 1 and 2. (b): The two regimes in Simulation Scenario 3. (c): The three regimes in Simulation Scenario 4. (d): The two regimes in Simulation Scenario 5.*

Appendix 7: simulation settings

In all simulation scenarios, the switching variable is generated from an AR(1) process with coefficient 0.6. The error term follows normal distribution with mean 0 and standard deviation 2.

Simulation Scenario 1 (Simple $\boldsymbol{A}$ with uncorrelated error) In this scenario, $T = 300$, $p = 20$, and $K = 2$. There is only one threshold value $r_1 = 4$, which is not close to the boundary. The auto-regressive coefficients are chosen to have the same structure but different values (see Figure 6).

Simulation Scenario 2 (Simple $\boldsymbol{A}$ with correlated error) This is the same settings as in Scenario 1, but the covariance matrix of the error term is changed. Specifically, we set $\boldsymbol{\Sigma}_\epsilon = 0.02(\sigma_{ij})_{n\times n}$ with $\sigma_{ij} = \rho^{|i-j|}$, where $\rho = 0.5$.

Simulation Scenario 3 (Random $\boldsymbol{A}$ with uncorrelated error) This setting is also similar to Scenario 1. However, the auto-regressive coefficients are chosen at random (see Figure 6).

Simulation Scenario 4 (Simple $\boldsymbol{A}$ with correlated error allowing changes in different regimes) In this scenario, $T = 600$, $p = 20$, and $K = 1$. There are two threshold values $r_1 = 4$ and $r_2 = 6$. The auto-regressive coefficients are chosen to have the same structure as in Scenario 1 but the values change at different thresholds (see Figure 6). We also include an additional simulation setting with $T = 300$ and threshold points $r_1 = 4$ and $r_2 = 6$ for this scenario.

Simulation Scenario 5 (Simple high-dimensional *A* with uncorrelated error) In this scenario, $T = 80$, $p = 100$, and $K = 2$. There is only one threshold value $r_1 = 5$. The auto-regressive coefficients are chosen to have the same structure as in Scenario 1 but with different values (see Figure 6).

The auto-regressive coefficients for the above simulation scenarios are visualized in Figure 6, where different coefficient values are represented by different colors. For Scenarios 1, 2 and 5, the 1-off diagonal values for the two lags in the two regimes are 0.49, −0.3, −0.4, and 0.49, respectively. For Scenario 4, the auto-regressive coefficients are allowed to change in different regimes. The 1-off diagonal values for one lag in the first regime are 0.25. In the second regime, the first $p/3$ values are decreased to −0.2. In the third regime, the last $p/4$ values are increased to 0.49. For Scenario 3, the auto-regressive coefficients are chosen at random.

Acknowledgments

We thank two anonymous reviewers and the AE for their helpful comments.

References

[1] Sipan Aslan, Ceylan Yozgatligil, and Cem Iyigun. Temporal clustering of time series via threshold autoregressive models: Application to commodity prices. *Annals of Operations Research*, in press, 01 2018. MR3741553
[2] Sumanta Basu and George Michailidis. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567, 2015. MR3357870
[3] Sumanta Basu, Ali Shojaie, and George Michailidis. Network granger causality with inherent grouping structure. *Journal of Machine Learning Research*, 16(13):417–453, 2015. MR3335801
[4] Peter J Bickel, Ya'acov Ritov, and Alexandre B Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of statistics*, 37(4):1705–1732, 2009. MR2533469
[5] Richard C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probab. Surveys*, 2:107–144, 2005. MR2178042
[6] Hansen Bruce. Inference in TAR Models. *Studies in Nonlinear Dynamics & Econometrics*, 2(1):1–16, April 1997. MR1467458
[7] Sergio A Calderón V and Fabio H Nieto. Bayesian analysis of multivariate threshold autoregressive models with missing data. *Communications in Statistics-Theory and Methods*, 46(1):296–318, 2017. MR3553033
[8] K. S. Chan. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, 21(1):520–533, 1993. MR1212191
[9] Ngai Hang Chan and Yury A. Kutoyants. On parameter estimation of threshold autoregressive models. *Statistical Inference for Stochastic Processes*, 15(1):81–104, Apr 2012. MR2892589

[10] Ngai Hang Chan, Chun Yip Yau, and Rong-Mao Zhang. LASSO estimation of threshold autoregressive models. *Journal of Econometrics*, 189(2):285–296, 2015. MR3414900

[11] Cathy WS Chen, Feng-Chi Liu, and Mike KP So. A review of threshold time series models in finance. *Statistics and its Interface*, 4(2):167–181, 2011. MR2812813

[12] Rong Chen. Threshold variable selection in open-loop threshold autoregressive models. *Journal of Time Series Analysis*, 16(5):461–481, 1995. MR1365642

[13] Shizhe Chen, Ali Shojaie, and Daniela M Witten. Network reconstruction from high-dimensional ordinary differential equations. *Journal of the American Statistical Association*, 112(520):1697–1707, 2017. MR3750892

[14] Michael J Dueker, Zacharias Psaradakis, Martin Sola, and Fabio Spagnolo. Multivariate contemporaneous-threshold autoregressive models. *Journal of Econometrics*, 160(2):311–325, 2011. MR2748555

[15] Walter Enders, Barry L Falk, and Pierre Siklos. A threshold model of real u.s. gdp and the problem of constructing confidence intervals in tar models. *Studies in Nonlinear Dynamics & Econometrics*, 11:1322–1322, 02 2007.

[16] Jianqing Fan and Qiwei Yao. Nonlinear time series. nonparametric and parametric methods. page 384, 01 2005. MR1964455

[17] Jerome Friedman, Trevor Hastie, Holger Höfling, and Robert Tibshirani. Pathwise coordinate optimization. Technical report, Annals of Applied Statistics, 2007. MR2415737

[18] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010. URL `http://www.jstatsoft.org/v33/i01/`.

[19] Fang Han and Han Liu. Transition matrix estimation in high dimensional time series. In *International Conference on Machine Learning*, pages 172–180, 2013.

[20] Bruce E Hansen. Threshold autoregression in economics. *Statistics and its Interface*, 4(2):123–127, 2011. MR2812805

[21] Bruce E Hansen and Byeongseon Seo. Testing for two-regime threshold cointegration in vector error-correction models. *Journal of econometrics*, 110(2):293–318, 2002. MR1928307

[22] Z. Harchaoui and C. Lévy-Leduc. Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association*, 105(492):1480–1493, 2010. MR2796565

[23] Junzhou Huang and Tong Zhang. The benefit of group sparsity. *The Annals of Statistics*, 38(4):1978–2004, 2010. MR2676881

[24] Zhenyu Jiang, Chengan Du, Assen Jablensky, Hua Liang, Zudi Lu, Yang Ma, and Kok Lay Teo. Analysis of schizophrenia data using a nonlinear threshold index logistic model. *PloS one*, 9(10):e109454, 2014.

[25] Luciana Juvenal and Mark P. Taylor. Threshold adjustment of deviations from the law of one price. *Studies in Nonlinear Dynamics and Econometrics (Online)*, Vol.12(No.3):Article 8, 2008. MR2443327

[26] Clifford Lam and Qiwei Yao. Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 40(2):694–726, 2012. MR2933663

[27] Chien-Hui Lee, Bwo-Nung Huang, et al. The relationship between exports and economic growth in east asian countries: A multivariate threshold autoregressive approach. *Journal of Economic Development*, 27(2):45–68, 2002.

[28] Dong Li and Howell Tong. Nested sub-sample search algorithm for estimation of threshold models. LSE Research Online Documents on Economics 68880, London School of Economics and Political Science, LSE Library, October 2016. MR3586227

[29] Dong Li, Shiqing Ling, and Howell Tong. On moving-average models with feedback. *Bernoulli*, 18(2):735–745, 05 2012. URL `https://doi.org/10.3150/11-BEJ352`. MR2922468

[30] Eckhard Liebscher. Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *Journal of Time Series Analysis*, 26(5):669–689, 2005. MR2188304

[31] Jiahe Lin and George Michailidis. Regularized estimation and testing for high-dimensional multi-block vector-autoregressive models. *Journal of Machine Learning Research*, 18(117):1–49, 2017. URL `http://jmlr.org/papers/v18/17-055.html`. MR3725456

[32] Xialu Liu and Rong Chen. Threshold factor models for high-dimensional time series. *Journal of Econometrics*, 216(1):53–70, 2020. MR4077381

[33] Ming Chien Lo and Eric Zivot. Threshold cointegration and nonlinear adjustment to the law of one price. *Macroeconomic Dynamics*, 5(4):533–576, 2001.

[34] Po-Ling Loh and Martin J Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. In *Advances in Neural Information Processing Systems*, pages 2726–2734, 2011. MR3015038

[35] Florence Merlevède, Magda Peligrad, and Emmanuel Rio. *Bernstein inequality and moderate deviations under strong mixing conditions*, volume Volume 5 of *Collections*, pages 273–292. Institute of Mathematical Statistics, 2009. MR2797953

[36] Fabio H. Nieto. Modeling bivariate threshold autoregressive processes in the presence of missing data. *Communications in Statistics – Theory and Methods*, 34(4):905–930, 2005. MR2163094

[37] Lizet Viviana Romero Orjuela and Sergio Alejandro Calderón Villanueva. Bayesian estimation of a multivariate tar model when the noise process follows a student-t distribution. *Communications in Statistics-Theory and Methods*, 50(11):2508–2530, 2021. MR4252744

[38] Alessandro Rinaldo. Properties and refinements of the fused lasso. *The Annals of Statistics*, 37(5B):2922–2952, 2009. MR2541451

[39] Mark Rudelson and Roman Vershynin. Hanson-wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18:9 pp., 2013. MR3125258

[40] Takumi Saegusa and Ali Shojaie. Joint estimation of precision matrices in heterogeneous populations. *Electronic journal of statistics*, 10(1):1341, 2016. MR3507368

[41] Abolfazl Safikhani and Ali Shojaie. Joint Structural Break Detection and Parameter Estimation in High-Dimensional Non-Stationary VAR Models. *Journal of American Statistical Association (Theory and Methods)*, page To Appear, 2020.

[42] Alex Tank, Emily B Fox, and Ali Shojaie. Granger causality networks for categorical time series. *arXiv preprint arXiv:1706.02781*, 2017. MR4205091

[43] Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2005. MR2136641

[44] Howell Tong. On a threshold model. *In Chen, C, (ed.) Pattern Recognition and Signal Processing. NATO ASI Series E: Applied Sc.*, 29:575–586, 1978. MR0552641

[45] Howell Tong. Threshold models in time series analysis—30 years on. *Statistics and its Interface*, 4(2):107–118, 2011. MR2812802

[46] Howell Tong and Keng S Lim. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42:245–268, 07 1980.

[47] Ruey S. Tsay. Testing and modeling multivariate threshold models. *Journal of the American Statistical Association*, 93(443):1188–1202, 1998. MR1649212

[48] Ruey S Tsay. Testing and modeling multivariate threshold models. *journal of the american statistical association*, 93(443):1188–1202, 1998. MR1649212

[49] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998. MR1652247

[50] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000. MR1652247

[51] Mariia Vladimirova, Stéphane Girard, Hien Nguyen, and Julyan Arbel. Sub-weibull distributions: Generalizing sub-gaussian and sub-exponential properties to heavier tailed distributions. *Stat*, 9(1), Jan 2020. MR4193421

[52] Tao Wang and Lixing Zhu. Consistent tuning parameter selection in high dimensional sparse linear regression. *Journal of Multivariate Analysis*, 102(7):1141 – 1151, 2011. ISSN 0047-259X. MR2805654

[53] Laurence Watier and Sylvia Richardson. Modelling of an epidemiological time series by a threshold autoregressive model. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 44(3):353–364, 1995.

[54] Kam Chung Wong, Zifan Li, and Ambuj Tewari. Lasso guarantees for β-mixing heavy-tailed time series. *Annals of Statistics*, 48(2):1124–1142, 2020. MR4102690