

# Threshold Tensor Factor Model in CP Form

Stevenson Bolivar

Department of Statistics, Rutgers University  
and

Rong Chen \*

Department of Statistics, Rutgers University  
and

Yuefeng Han †

Department of Applied and Computational Mathematics and Statistics,  
University of Notre Dame

November 26, 2025

## Abstract

This paper proposes a new Threshold Tensor Factor Model in Canonical Polyadic (CP) form for tensor time series. By integrating a thresholding autoregressive structure for the latent factor process into the tensor factor model in CP form, the model captures regime-switching dynamics in the latent factor processes while retaining the parsimony and interpretability of low-rank tensor representations. We develop estimation procedures for the model and establish the theoretical properties of the resulting estimators. Numerical experiments and a real-data application illustrate the practical performance and usefulness of the proposed framework.

*JEL Classifications:* C13, C32, C55

*Keywords:* CP Decompositions, Factor models, Threshold autoregressive models, High-dimensional, Tensor data.

---

\*Rong Chen's research was supported in part by National Science Foundation grants DMS-2319260 and DMS-2413858.

†Yuefeng Han's research was supported in part by National Science Foundation grant DMS-2412578.

# 1 Introduction

Tensor time series data, denoted as  $\mathcal{X}_t \in \mathbb{R}^{d_1 \times d_2 \times \dots \times d_K}$  for  $t = 1, \dots, T$ , is a collection of multidimensional arrays observed sequentially over time. Such tensor time series arise in various scientific and financial applications, including multi-category product import-export volume among a group of countries over time (Chen et al. 2022), quarterly economic indices of different countries (Chen, Tsay & Chen 2020), returns of portfolio constructed by size, book to market ratio, and momentum (Li & Xiao 2021), multivariate spatial-temporal data (Chen, Yun, Chen & Yao 2020, Barigozzi et al. 2025), and many others.

Bolivar et al. (2025) provides a review of recent developments in autoregressive and factor modeling for such data. For example, Chen et al. (2021), Xiao et al. (2023) and Li & Xiao (2024) extend the autoregressive model to matrix-valued time series (MAR), while Li & Xiao (2021) and Wang et al. (2024) extend it further to tensor time series. Another widely used approach in high-dimensional data is dimension reduction through factor models. The two main decompositions for tensor data structures are Tucker decomposition and CP decomposition (Kolda & Bader 2009). Mimicking Tucker decomposition, Wang et al. (2019), Chen et al. (2022) and Chen & Fan (2023) introduced a tensor factor model in Tucker form, with factors in tensor form but of much smaller dimensions than the observed tensors. Other notable contributions in this area include Chen, Tsay & Chen (2020), Han, Chen, Yang & Zhang (2024), Han et al. (2022), Chen et al. (2025), and others.

In contrast, Han, Yang, Zhang & Chen (2024) propose a tensor factor model in CP form (TFM-cp) that uses a set of univariate factor processes that are uncorrelated, and rank-1 tensor factor loadings. This approach mimics the CP decomposition. This model structure facilitates the analysis of underlying dynamics in high-dimensional tensor time series through a set of univariate uncorrelated latent factor processes, making it significantly more convenient for studying the dynamics of the time-series. The TFM-cp is specified as:

$$\mathcal{X}_t = \sum_{j=1}^r \lambda_j f_{jt} \mathbf{u}_{j1} \circ \mathbf{u}_{j2} \circ \dots \circ \mathbf{u}_{jK} + \mathcal{E}_t, \quad (1)$$

where each  $f_{jt}$  is a one-dimensional latent factor,  $u_{jk} \in \mathbb{R}^{d_k}$  are the loading vectors normalized to  $\|u_{jk}\|_2 = 1$ ,  $E_t$  is a zero mean idiosyncratic noise tensor with potentially weak cross-correlations among its elements and assumed to be uncorrelated with the latent factors, and  $r$  is the rank of the decomposition, or equivalently the number of factors. Without loss of generality, it is assumed  $\mathbb{E} f_{jt}^2 \leq C < \infty$ , for all  $1 \leq j \leq r$ . Then, the signal strengths are contained in  $\lambda_j$ . Here,  $\{u_{jk}, 1 \leq i \leq r\}$  are not necessarily orthogonal. [Chang et al. \(2023\)](#), [Han, Yang, Zhang & Chen \(2024\)](#) showed that the model can be very useful in many applications.

We observe that the existing approaches mentioned above are all linear models. It has been widely acknowledged that in many applications linear time series models are not sufficient to capture various nonlinear phenomena encountered and there has been a large literature in study nonlinear time series models ([Tong 1990b](#), [Tsay & Chen 2018](#), [De Gooijer et al. 2017](#)). Modern machine learning algorithms such as deep neural networks are also large nonlinear systems and have shown to be extremely powerful in providing accurate predictions ([Zhou et al. 2025](#), [Luo et al. 2025](#)).

Motivated by observed regime-switching phenomenon in applications and encouraged by its simple form, regime-switching models, particularly the threshold models, are a class of powerful nonlinear time series models that have been studied extensively and have been shown to be a simple, elegant and useful parametric model ([Tong 1990b](#), [Chan & Tong 1986](#), [Sims & Zha 2006](#), [Tsay 1989](#), [Ang & Timmermann 2012](#), [Hansen 2011](#)) in many applications ranging from economics, biology and ecology, signal processing, and many others. As an attempt to move from the linear models to nonlinear ones for tensor time series, in this paper we propose a Threshold dynamic Tensor Factor Model in CP form (T-TFM-cp), which extends the standard TFM-cp model by incorporating a threshold mechanism into the latent factor dynamics. Specifically, we model the factors  $f_{jt}$  using a general Threshold Autoregressive (TAR) model:

$$f_{jt} = \text{TAR}(p_j, s_j, z_{jt}), \quad j = 1, \dots, r. \quad (2)$$

The TAR model allows the factor dynamics to switch between different regimes depending on the value of a threshold variable  $z_{jt}$ . This variable can be a function of observable variables (at time  $t$ ) or lagged latent factors. The parameters  $p_j$  and  $s_j$  determine the specific form of the TAR model for factor  $j$ .

There is a vast literature on TAR models ([Tong 1990b](#), [Chan & Tong 1986](#), [Sims & Zha 2006](#), [Tsay 1989](#), [Ang & Timmermann 2012](#)). For matrix autoregressive models, recent work such as [Bucci \(2024\)](#), [Wu & Chan \(2023\)](#), and [Yu, Li, Zhang & Tong \(2024\)](#) develops new methods for modeling regime-specific changes in the coefficient matrices. Related but conceptually distinct, [Liu & Chen \(2020\)](#) introduce a threshold vector factor model, and [Liu & Chen \(2022\)](#) study a threshold matrix factor model in Tucker form. In their framework, the loading matrices differ across regimes and regime switching is controlled by a threshold variable. The latent factors are assumed to be stationary but are not equipped with an explicit dynamic model, and therefore these approaches do not provide forecasting capability. In contrast, the proposed T-TFM-cp model assumes common loading structures across all regimes, while allowing the latent factor processes to switch between regimes according to one or more threshold variables. A distinctive feature of T-TFM-cp is its flexibility: all factors may share the same threshold variable, or each factor can be governed by its own thresholding rule. Moreover, unlike the threshold factor models mentioned above, T-TFM-cp incorporates an explicit dynamic specification for the factor processes, thereby enabling prediction, similar in spirit to the dynamic matrix factor model of [Yu, Chen, Xiao & Han \(2024\)](#).

In this paper, we extend the TFM-cp model to a nonlinear threshold framework—rather than extending the TFM-Tucker model—for several reasons. First, the TFM-cp model (1) is uniquely defined up to sign changes when the signal strengths are ordered, facilitating interpretation. Second, by decomposing the factor processes into uncorrelated univariate time series, TFM-cp enables a more flexible modeling of temporal dynamics. Compared with Tucker-based representations, CP models typically require fewer factors to capture the

essential signal structure, achieving a favorable balance between parsimony and explanatory power. Third, the uncorrelated univariate factor structure enables independent modeling of regime-switching dynamics for each factor, potentially with different threshold variables, threshold values, or numbers of regimes.

By combining CP decomposition with TAR dynamics, the proposed model offers a powerful and flexible framework for analyzing high-dimensional tensor time-series data exhibiting regime-switching behavior in the latent factors. Balancing parsimony, interpretability, and statistical efficiency, it provides a valuable tool for researchers working with complex multiway temporal data in fields such as neuroscience, finance, and signal processing.

The remainder of the paper is organized as follows. Section 2 presents the proposed model in detail, including the TAR dynamics of the latent factors and the estimation procedures. Section 3 establishes the theoretical properties of the estimators. Sections 4 and 5 demonstrate the effectiveness of the model through simulation studies and an empirical application, respectively. Section 6 concludes with a discussion of future research directions.

## 2 Model Setting and Estimation

In this section, we focus on a more specific specification of the T-TFM-cp model introduced in (1) and (2). In particular, we assume that each latent factor follows a threshold autoregressive (TAR) process of the form

$$f_{j,t} = \begin{cases} \phi_{j,1}^{(1)} f_{j,t-1} + \cdots + \phi_{j,p_j}^{(1)} f_{j,t-p_j} + \varsigma_j^{(1)} \xi_t & \text{if } z_{j,t} \in (-\infty, s_{j,1}], \\ \phi_{j,1}^{(2)} f_{j,t-1} + \cdots + \phi_{j,p_j}^{(2)} f_{j,t-p_j} + \varsigma_j^{(2)} \xi_t & \text{if } z_{j,t} \in (s_{j,1}, s_{j,2}], \\ \vdots & \\ \phi_{j,1}^{(L)} f_{j,t-1} + \cdots + \phi_{j,p_j}^{(L)} f_{j,t-p_j} + \varsigma_j^{(L)} \xi_t & \text{if } z_{j,t} \in (s_{j,L-1}, \infty), \end{cases} \quad (3)$$

where  $L$  is the number of regimes,  $s_{j,l}$  are the threshold values,  $p_j$  are the autoregressive orders,  $\phi_j^{(l)}$  are the autoregressive coefficients for regime  $l$ , and  $\xi_t$  are white noise processes

with variances 1 for  $l = 1, 2, \dots, L$ . In the Self-Exciting TAR (SETAR) case, the exogenous threshold variable  $z_t$  is assumed to be past values of  $f_{jt}$ , e.g.  $f_{j,t-\tau_j}$ , where  $\tau_j$  is the delay parameter. In the following we use the notations  $s_j = (s_{j,1}, \dots, s_{j,L-1})^\top \in \mathbb{R}^{L-1}$ ,  $\phi_j^{(l)} = (\phi_{j,1}^{(l)}, \dots, \phi_{j,p_j}^{(l)})^\top \in \mathbb{R}^{p_j}$ , and  $\phi_j = (\phi_j^{(1)\top}, \dots, \phi_j^{(L)\top})^\top \in \mathbb{R}^{Lp_j}$ .

The TAR model is a powerful and elegant nonlinear extension of the linear autoregressive (AR) model, capable of capturing regime switching driven by an observable threshold variable. It has been widely used to model economic and financial data exhibiting asymmetric dynamics and nonlinear patterns. By allowing different autoregressive behaviors depending on the regime of the threshold variable  $z_{j,t-d}$ , the model can represent phenomena such as asymmetric cycles, shifts in mean levels, and volatility clustering. This flexibility makes TAR models particularly suitable for characterizing distinct dynamics during economic expansions and recessions (Tong 1990b, Chan & Tong 1986, Tsay 1989).

We refer to the combination of the TFM-cp model in (1) and the dynamic TAR specification in (3) as the TAR Tensor Factor Model in CP form (T-TFM-cp). To estimate the parameters of the T-TFM-cp model, we adopt a two-step estimation strategy. In the first step, we estimate the TFM-cp model while ignoring the dynamic structure of the latent factors. This yields estimates of both the loading vectors  $\mathbf{u}_{jk}$  and the factor process  $\hat{f}_{jt}$ . The procedure follows Chen et al. (2024), which combines a composite PCA initialization with an iterative simultaneous orthogonalization (ISO) algorithm. Details and theoretical properties can be found in Chen et al. (2024). Denote the estimated loading vectors by  $\hat{u}_{jk}$  and the estimated factors by  $\hat{f}_{jt} = X_t \times_{k=1}^K \hat{v}_{jk}^\top$ , where  $\hat{v}_{jk}$  is the  $j$ -th column of  $\widehat{U}_k (\widehat{U}_k^\top \widehat{U}_k)^{-1}$ , with  $\widehat{U}_k = (\hat{u}_{1k}, \dots, \hat{u}_{rk}) \in \mathbb{R}^{d_k \times r}$ . Under this definition,  $\hat{f}_{jt}$  serves as an estimate of  $\lambda_j f_{jt}$  in (1), implicitly absorbing the signal strength  $\lambda_j$ . This scaling does not affect the results of the simulation study or the real data analysis; it only changes the magnitude of the estimated factors.

Second, once the factor process estimates  $\hat{f}_{jt}$  are obtained, we fit an individual TAR model to each estimated factor process using standard identification and estimation tech-

niques for TAR models (Tong 1990a, Tsay 1989, Tsay & Chen 2018). In particular, given the threshold variable, the number of regimes, and the AR orders of each regime, the AR coefficients and threshold values are estimated by least squares. Specifically, for each  $j = 1, \dots, r$ , the estimates  $\hat{\phi}_j$  and  $\hat{s}_j, j = 1, \dots, r$  minimizes

$$\sum_{t=p_j+1}^T \left( \hat{f}_{j,t} - \sum_{\ell}^L I(z_{j,t} \in (s_{j,\ell-1}, s_{j,\ell})) \sum_{k=1}^{p_j} \phi_{j,k}^{(\ell)} \hat{f}_{j,t-k} \right)^2. \quad (4)$$

Although the estimation of the threshold vector  $s_j$  is strictly a non-convex optimization problem, it is typically not computationally intensive since one only needs to check the observed values of the threshold variable; all values between two consecutive observations of the threshold variable are equivalent (Chan 1993, Hansen 2000). Sequential update of the least squares (when adding or dropping one observation) makes exhaustive search computationally efficient, though more advanced algorithms are available (Li & Tong 2016, Li et al. 2011). The number of regimes and the AR orders are typically determined by BIC or similar measures, as well as diagnostic model checking procedures involving residual analysis, testing for remaining nonlinearity, and evaluating forecasting performance (Tong 1990b, Hansen 2000, Li et al. 2011, Gonzalo & Pitarakis 2002, Chan & Tsay 1998).

The two-step estimation strategy is standard for models consisting of multiple components, such as the T-TFM-cp model. Similar procedures have been widely used in the literature on dynamic (vector) factor models; see, for example, Stock & Watson (2016), Gourioux & Jasiak (2001), Barigozzi & Hallin (2016), Otto & Salish (2022).

Threshold variable determination is essential for building a threshold model. When there is no prior knowledge on the threshold variable, a data-driven procedure is needed in order to search for a suitable one. In standard univariate threshold models, a typical candidate pool is the lag variables (Tong & Lim 1980, Tong 1990b, Chan 1993, Tong 2010) and identification is commonly performed by comparing a handful of plausible specifications. When multiple time series are involved, as in our setting, it may be desirable to identify one or a small number of common threshold variables, which can substantially improve

model interpretability. When one considers a large candidate pool consisting of exogenous variables, a trial-and-error approach can be extremely time consuming, complicated more by the multiple comparison problem at the end. In such cases, the reverse approach proposed in (Wu & Chen 2007, Liu & Chen 2016) provides an effective alternative. It is also worth noting that different factor processes may rely on different types of threshold variables: some may follow self-exciting dynamics using lagged values of the factor itself, while others may require observable exogenous threshold variables.

### 3 Theoretical Properties of the Estimators

In this section, we present some theoretical properties of the proposed two-stage estimation procedures. We introduce some notations first. Let  $d = \prod_{k=1}^K d_k, d_{\max} = \max\{d_1, \dots, d_K\}$ . The matrix Frobenius norm is defined as  $\|A\|_F = (\sum_{ij} a_{ij}^2)^{1/2}$ . Define the spectral norm as

$$\|A\|_2 = \max_{\|x\|_2=1, \|y\|_2=1} \|x^\top A y\|_2.$$

Considering that the loading vector  $u_{jk}$  of TFM-cp can only be identified with a change in sign, we employ

$$\|\hat{u}_{jk} \hat{u}_{jk}^\top - u_{jk} u_{jk}^\top\|_2 = \sqrt{1 - (\hat{u}_{jk}^\top u_{jk})^2} = \sup_{z \perp u_{jk}} |z^\top \hat{u}_{jk}|$$

to quantify the discrepancy between  $\hat{u}_{jk}$  and  $u_{jk}$ .

To establish the asymptotic properties of the proposed procedures, we impose the following assumptions.

**Assumption 1.** The idiosyncratic noise process  $E_t$  are independent Gaussian tensors, conditioning on the factor process  $\{f_{jt}, 1 \leq j \leq r, t \in \mathbb{Z}\}$ . In addition, there exists some constant  $\sigma > 0$ , such that

$$\mathbb{E}(v^\top \text{vec}(E_t))^2 \leq \sigma^2 \|v\|_2^2, \quad \forall v \in \mathbb{R}^d.$$

**Assumption 2.** Let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$ . Suppose  $\lambda_1 \asymp \lambda_r \asymp \lambda$ .



**Assumption 3.** Assume the factor process  $f_{jt}, 1 \leq j \leq r$ , is stationary and strong  $\alpha$ -mixing in  $t$ , with  $\mathbb{E}f_{jt}^2 < \infty$ . Let  $F_t = (f_{1t}, \dots, f_{rt})^\top$ . For any  $v \in \mathbb{R}^r$  with  $\|v\|_2 = 1$ ,

$$\max_t \mathbb{P}(|v^\top F_t| \geq x) \leq c_1 \exp(-c_2 x^{\gamma_2}), \quad (5)$$

where  $c_1, c_2$  are some positive constants and  $0 < \gamma_2 \leq 2$ . In addition, the mixing coefficient satisfies

$$\alpha(m) \leq \exp(-c_0 m^{\gamma_1}) \quad (6)$$

for some constant  $c_0 > 0$  and  $\gamma_1 > 0$ , where

$$\alpha(m) = \sup_t \left\{ \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \right| : A \in \sigma(f_{js}, 1 \leq j \leq r, s \leq t), B \in \sigma(f_{js}, 1 \leq j \leq r, s \geq t + m) \right\}.$$

Assumption 1 closely resembles the noise conditions found in foundational studies such as Bai & Ng (2002), Bai (2003), Lam et al. (2011), Lam & Yao (2012), and other notable contributions in the factor model literature. For ease of exposition, we assume that the noise tensor is independent across time  $t$ , hence allowing for weak cross-sectional contemporaneous dependence. Although introducing a weak temporal correlation among noises, as proposed by Bai & Ng (2002), is feasible, it significantly complicates the theoretical framework. Additionally, we adopt the normality assumption for technical convenience, although it can be generalized to accommodate exponential-type tail conditions, as seen in Chen et al. (2024).

Assumption 2 is a standard condition in factor models, such as Chen et al. (2024) and Han, Yang, Zhang & Chen (2024). In the case of strong factor models, we have  $\lambda = \sqrt{d}$ . Assumption 3 requires the tail probability of  $f_{jt}$  to exhibit exponential decay, which is a standard assumption. In particular, when  $\gamma_2 = 2$ , this implies that  $f_{jt}$  follows a sub-Gaussian distribution. The mixing condition is a well-established assumption that encompasses a wide variety of time series models, including causal ARMA processes with continuous innovations, as discussed in Tong (1990b), Tsay (2005), Fan & Yao (2003), Rosenblatt (2012), Tsay & Chen (2018), among others.

Next, we impose assumptions for the TAR structure of the factor processes, depending on whether the threshold variable  $z_{jt}$  is self-exciting (i.e., the past values of  $f_{jt}$ ) or is an exogenous observable variable. If  $z_{jt}$  is self-exciting with  $z_{jt} = f_{j,t-\tau_j}$ , then  $Y_{jt} = (f_{j,t-1}, \dots, f_{j,t-(p_j \vee \tau_j)})$  is a Markov chain. Denote its  $m$ -step transition probability by  $\mathbf{P}^m(y, A)$ , where  $y \in \mathbb{R}^{p_j \vee \tau_j}$  and  $A$  is a Borel set. If  $z_{jt}$  is observable, let  $Y_{jt} = (f_{j,t-1}, \dots, f_{j,t-p_j}, z_{jt})$ , Denote its  $m$ -step transition probability by  $\mathbf{P}^m(y, A)$ , where  $y \in \mathbb{R}^{p_j+1}$  and  $A$  is a Borel set.

**Assumption 4.** The innovation of the latent TAR process (3),  $\xi_{jt}$  are i.i.d. with  $\mathbb{E}\xi_{jt} = 0$ ,  $\mathbb{E}\xi_{jt}^2 = 1$  and  $\mathbb{E}\xi_{jt}^4 < \infty$  for each  $1 \leq j \leq r$ , and  $\xi_{jt}$  has a bounded, continuous and positive density on  $\mathbb{R}$ .

**Assumption 5.** The Markov Chain  $Y_{jt}$  ( $1 \leq j \leq r$ ) admits a unique invariant measure  $\Pi(\cdot)$  such that there exist  $C_0 > 0$  and  $\rho \in [0, 1)$ , for any  $y$  and any  $m$ ,  $\|\mathbf{P}^m(y, \cdot) - \Pi(\cdot)\|_{\text{TV}} \leq C_0(1 + \|y\|)\rho^m$ , where  $\|\cdot\|_{\text{TV}}$  and  $\|\cdot\|$  denote the total variation norm and the Euclidean norm, respectively.

**Assumption 6.** Assume that

- (i) (when  $z_{jt}$  is self-exciting with  $z_{jt} = f_{j,t-\tau_j}$ ): There exist nonrandom vectors  $w_{j\ell} = (w_{j,\ell,1}, \dots, w_{j,\ell,p_j})^\top$  with  $w_{j,\ell,\tau_j} = s_{j,\ell}$  such that  $(\phi_j^{(\ell)} - \phi_j^{(\ell+1)})^\top w_{j\ell} \neq 0$  for  $\ell = 1, \dots, L-1$ .
- (ii) (when  $z_{jt}$  is an observable variable): Assume that  $z_{jt}$  has a bounded, continuous and positive density on  $\mathbb{R}$ , and is Markovian. There exist vector  $w_{j\ell} = \mathbb{E}[(f_{j,t-1}, \dots, f_{j,t-p_j})^\top | z_{jt} = s_{j,\ell}]$  such that  $(\phi_j^{(\ell)} - \phi_j^{(\ell+1)})^\top w_{j\ell} \neq 0$  for  $\ell = 1, \dots, L-1$ .

Assumptions 4, 5, 6 are commonly used in theoretical analysis of TAR models. Under Assumption 5,  $Y_{jt}$  is V-uniformly ergodic with  $V(\cdot) = K(1 + \|\cdot\|)$ , a condition that is stronger than geometric ergodicity. For a detailed explanation of V-uniform ergodicity, see

Chapter 16 in [Meyn & Tweedie \(2012\)](#). In the specific case where all threshold variables are lags of  $f_{jt}$  and errors are homoscedastic across all regimes, Assumption 4 combined with the condition  $\max_{\ell} \sum_{i=1}^{p_j} |\phi_{j,i}^{(\ell)}| < 1$  suffices for Assumption 5 to hold. More details can be found in [Chan & Tong \(1985\)](#) and [Chan \(1989\)](#). Assumption 6 implies that the autoregressive function is discontinuous at the thresholds  $s_{j,\ell}$ . In Assumption 6(i), if  $\tau_j > p_j$ , then  $w_{j,\ell,\tau_j}$  may not be a component of  $w_{j\ell}$ . In this scenario, Assumption 6(i) is equivalent to the conditions  $\phi_j^{(\ell)} \neq \phi_j^{(\ell+1)}$  for  $1 \leq \ell \leq L-1$ , which are both necessary and sufficient for the identification of all thresholds. Assumption 6 (ii) is similar to Assumption 3.6 in [Zhang et al. \(2024\)](#).

[Han, Yang, Zhang & Chen \(2024\)](#) proposed iterative procedure ISO to estimate the factor loading vectors of TFM-cp using auto-covariance tensor. When the number of factors  $r$  is fixed, the procedure achieves convergence rate  $\|\hat{u}_{jk}\hat{u}_{jk}^\top - u_{jk}u_{jk}^\top\|_2 = O_{\mathbb{P}}(\sigma d_{\max}^{1/2}\lambda^{-1}T^{-1/2} + \sigma^2\lambda^{-2})$ ,  $1 \leq k \leq K$ . Using their iterative procedure, we can further establish the theoretical properties of the estimated latent factor process  $\hat{f}_{jt}$  in (1).

**Proposition 3.1.** Suppose Assumptions 1, 2, 3 hold. Assume  $r$  is fixed. Let  $\hat{f}_{jt} = \lambda^{-1}X_t \times_{k=1}^K \hat{u}_{jk}^\top$  be the estimated factors using the iterative procedure in [Han, Yang, Zhang & Chen \(2024\)](#), assuming the initialization condition is satisfied and the signal strength  $\lambda$  is known. Then, with probability at least  $1 - T^{-c} - \sum_{k=1}^K e^{-d_k}$ ,

$$|\hat{f}_{jt} - f_{jt}| \leq C \left( \frac{\sigma\sqrt{d_{\max}}}{\lambda\sqrt{T}} + \frac{\sigma}{\lambda} \right), \quad (7)$$

for all  $1 \leq t \leq T$ , and

$$\left| \frac{1}{T-h} \sum_{t=h+1}^T \hat{f}_{i,t-h} \hat{f}_{jt} - \frac{1}{T-h} \sum_{t=h+1}^T f_{i,t-h} f_{jt} \right| \leq C \left( \frac{\sigma\sqrt{d_{\max}}}{\lambda\sqrt{T}} + \frac{\sigma^2}{\lambda^2} \right), \quad (8)$$

for all  $1 \leq i, j \leq r$  and  $0 \leq h \leq T/4$ , where  $c$  is a positive constant.

The proposition establishes a non-asymptotic bound for the estimated factors  $\hat{f}_{jt}$ . It demonstrates that consistent factor estimation requires the signal-to-noise ratio  $(\lambda/\sigma)$  to increase to infinity, ensuring sufficient information about the signal at each time point  $t$ .

As expected, a higher signal-to-noise ratio leads to faster convergence, as stronger factors provide more information in the observed data. Additionally, (8) shows that the error rate between the sample (auto-)covariance of the estimated factors and the true sample (auto-)cross-moment is  $o_{\mathbb{P}}(T^{-1/2})$  when  $\lambda/\sigma \gg \sqrt{d_{\max}} + T^{1/4}$ . In the case of a strong factor model setting (Bai & Ng 2002, Lam et al. 2011), where  $\lambda/\sigma \asymp \sqrt{d}$ , this condition is equivalent to  $T \ll d^2$ , which generally holds. This suggests that under such conditions, using the estimated factor processes for model building and inference of the TAR component is equivalent to using the underlying true factors, without any efficiency loss. The statistical rates in (7) and (8) provide a basis for further modeling of the estimated factors.

The following theorems present the convergence rate and asymptotic normality for the least squares estimators (4) in the second-stage TAR modeling of the latent factor process, using  $\hat{f}_{jt}$  as the true factor process. Theorems 3.1 and 3.2 separately examine the cases of self-exciting threshold variables and observable threshold variables.

**Theorem 3.1.** Consider  $z_{jt}$  in (3) as self-exciting with  $z_{jt} = f_{j,t-\tau_j}$ , for some  $1 \leq j \leq r$ . Suppose Assumptions 1, 2, 3, 4, 5, 6(i) hold. Assume  $r$  is fixed. Let  $\hat{f}_{jt} = \lambda^{-1} X_t \times_{k=1}^K \hat{v}_{jk}^{\top}$  be the estimated factors using the iterative procedure in Han, Yang, Zhang & Chen (2024), assuming the initialization condition is satisfied and the signal strength  $\lambda$  is known. If  $\sigma\sqrt{d_{\max}}/(\lambda\sqrt{T}) + \sigma^2/\lambda^2 \rightarrow 0$  as  $T \rightarrow \infty$ , then  $\hat{\tau}_j \rightarrow \tau_j$  in probability. Moreover,

$$\hat{s}_j - s_j = O_{\mathbb{P}} \left( \frac{1}{T} + \frac{\sigma\sqrt{d_{\max}}}{\lambda\sqrt{T}} + \frac{\sigma}{\lambda} \right), \quad (9)$$

$$\hat{\phi}_j - \phi_j = O_{\mathbb{P}} \left( \frac{1}{\sqrt{T}} + \frac{\sigma\sqrt{d_{\max}}}{\lambda\sqrt{T}} + \frac{\sigma}{\lambda} \right). \quad (10)$$

Furthermore, if the factor strength satisfies  $\lambda/\sigma \gg T^{1/2} + \sqrt{d_{\max}}$ , then

$$\sqrt{T} (\hat{\phi}_j - \phi_j) \Rightarrow N(0, \Sigma_j), \quad (11)$$

where  $\Sigma_j = \text{diag}\{(\varsigma_j^{(1)})^2 \Sigma_{j1}, \dots, (\varsigma_j^{(L)})^2 \Sigma_{jL}\}$  with  $\Sigma_{j,\ell}^{-1} = \mathbb{E}[\tilde{f}_{j,t-1} \tilde{f}_{j,t-1}^{\top} 1\{s_{j,\ell-1} < f_{j,t-\tau_j} \leq s_{j,\ell}\}]$ ,  $1 \leq \ell \leq L$ , and  $\tilde{f}_{j,t-1} = (f_{j,t-1}, \dots, f_{j,t-p_j})^{\top}$ ,  $s_{j,0} = -\infty$ ,  $s_{j,L} = +\infty$ , and  $\varsigma_j^{(\ell)}$  is defined in (3).

**Theorem 3.2.** Consider  $z_{jt}$  in (3) as an observable variable, for some  $1 \leq j \leq r$ . Suppose Assumptions 1, 2, 3, 4, 5, 6(ii) hold. Assume  $r$  is fixed. Let  $\hat{f}_{jt} = \lambda^{-1} X_t \times_{k=1}^K \hat{v}_{jk}^\top$  be the estimated factors using the iterative procedure in Han, Yang, Zhang & Chen (2024), assuming the initialization condition is satisfied and the signal strength  $\lambda$  is known. Then,

$$\hat{s}_j - s_j = O_{\mathbb{P}}\left(\frac{1}{T}\right), \quad (12)$$

$$\hat{\phi}_j - \phi_j = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}} + \frac{\sigma\sqrt{d_{\max}}}{\lambda\sqrt{T}} + \frac{\sigma^2}{\lambda^2}\right). \quad (13)$$

Furthermore, if the factor strength satisfies  $\lambda/\sigma \gg T^{1/4} + \sqrt{d_{\max}}$ , then

$$\sqrt{T}(\hat{\phi}_j - \phi_j) \Rightarrow N(0, \Sigma_j), \quad (14)$$

where  $\Sigma_j = \text{diag}\{(\varsigma_j^{(1)})^2 \Sigma_{j1}, \dots, (\varsigma_j^{(L)})^2 \Sigma_{jL}\}$  with  $\Sigma_{j,\ell}^{-1} = \mathbb{E}[\tilde{f}_{j,t-1} \tilde{f}_{j,t-1}^\top \mathbf{1}\{s_{j,\ell-1} < z_{jt} \leq s_{j,\ell}\}]$ ,  $1 \leq \ell \leq L$ , and  $\tilde{f}_{j,t-1} = (f_{j,t-1}, \dots, f_{j,t-p_j})^\top$ ,  $s_{j,0} = -\infty$ ,  $s_{j,L} = +\infty$ , and  $\varsigma_j^{(\ell)}$  is defined in (3).

In the case of self-exciting threshold variables, the estimation error of the threshold level  $s_j$  in (9) comprises two components: the conventional  $1/T$  rate from Chan (1993), Li & Ling (2012) and Zhang et al. (2024) and the estimation error of the factor process from the first stage in (7). This additional error arises because the lagged estimated factor process serves as the threshold variable. Since the estimated factors determine the Markovian regime at each time  $t$ , we require a non-asymptotic convergence rate for all  $\hat{f}_{jt}$ , as established in Proposition 3.1. The convergence rate provided by Theorem 3 in Han, Yang, Zhang & Chen (2024) is insufficient for this purpose. In contrast, for observable threshold variables, the threshold level  $s_j$  in (12) achieves the standard  $1/T$  rate without additional factor estimation error.

The presence of factor estimation error in the self-exciting threshold variable case leads to larger AR coefficient estimation errors compared to the observable threshold variable case. For observable thresholds, the error consists of the parametric rate  $T^{-1/2}$  and the error from the sample covariance matrix using estimated factors as in (8). For self-exciting thresholds,

an additional term  $\sigma/\lambda$  appears, reflecting the signal-to-noise ratio. Consequently, the least squares estimator  $\hat{\phi}_j$  satisfies a central limit theorem only when the signal strength is sufficiently high: specifically,  $\lambda/\sigma \gg \sqrt{d_{\max}} + T^{1/4}$  for observable threshold variables and  $\lambda/\sigma \gg \sqrt{d_{\max}} + T^{1/2}$  for self-exciting threshold variables. When the signal is weaker, consistency remains achievable, but the factor loading and idiosyncratic noise  $E_t$  dominate the estimation error. In such regimes, the factor process error in (8) and/or (7) dominates the parametric rate, preventing the derivation of a tractable asymptotic distribution.

## 4 Simulations

We conducted a simulation study to assess the performance of the proposed T-TFM-cp model in matrix form across a range of settings. The scenarios systematically vary the dimensions, sample sizes, and signal-to-noise ratios (SNR,  $\lambda/\sigma$ ). Specifically, we generate data from weak factor model with  $\lambda = 1$ ,

$$X_t = M_t + E_t = \sum_{j=1}^3 u_{1,j} u'_{2,j} f_{j,t} + E_t, \quad (15)$$

where, for  $j = 1, 2, 3$ , the components  $u_{1,j}$  and  $u_{2,j}$  are generated by drawing i.i.d. Gaussian entries with zero mean and then normalizing the resulting vectors. All elements of  $E_t$  have variance  $\sigma^2$ . The latent factors follow the TAR models

$$\begin{aligned} f_{1,t} &= \begin{cases} 0.5f_{1,t-1} + 0.2f_{1,t-2} + a_{1,t}, & \text{if } f_{1,t-1} < 0, \\ 0.7f_{1,t-1} - 0.6f_{1,t-2} + a_{1,t}, & \text{if } f_{1,t-1} \geq 0, \end{cases} \\ f_{2,t} &= \begin{cases} 0.8f_{2,t-1} + 0.1f_{2,t-2} + a_{2,t}, & \text{if } f_{2,t-1} < 0, \\ -0.4f_{2,t-1} - 0.6f_{2,t-2} + a_{2,t}, & \text{if } f_{2,t-1} \geq 0, \end{cases} \\ f_{3,t} &= \begin{cases} 0.7f_{3,t-1} + a_{3,t}, & \text{if } f_{3,t-1} < 0, \\ -0.8f_{3,t-1} + a_{3,t}, & \text{if } f_{3,t-1} \geq 0, \end{cases} \end{aligned} \quad (16)$$

where  $a_{j,t}$  are independent Gaussian white noise with zero mean and unit variance ( $\varsigma_j^2 = 1$ ), for  $j = 1, 2, 3$ . All results are based on 100 simulation replications.

Given that the TFM-cp factors are identifiable only up to sign changes and permutations, we first aligned the estimated factors with the true simulated factors using optimal sign matching and permutation selection. We then evaluate the estimation accuracy of the loading vectors using the Mean Squared Error (MSE):

$$\text{MSE}_j = \frac{|\hat{u}_{1,j} - u_{1,j}|_2^2}{d_1} + \frac{|\hat{u}_{2,j} - u_{2,j}|_2^2}{d_2}, \quad j = 1, 2, 3.$$

We use the estimation procedure of [Han, Yang, Zhang & Chen \(2024\)](#) to estimate the TFM-cp model, using lag 1 ( $h = 1$ ) auto-moments in the estimation. Figure 1 presents boxplots of the logarithm of  $\text{MSE}_j$  for the three loading vectors, under three choices of matrix dimensions, three SNR levels, and three time-series lengths. As expected, the estimation error decreases as the sample size, dimensionality, and SNR increase.

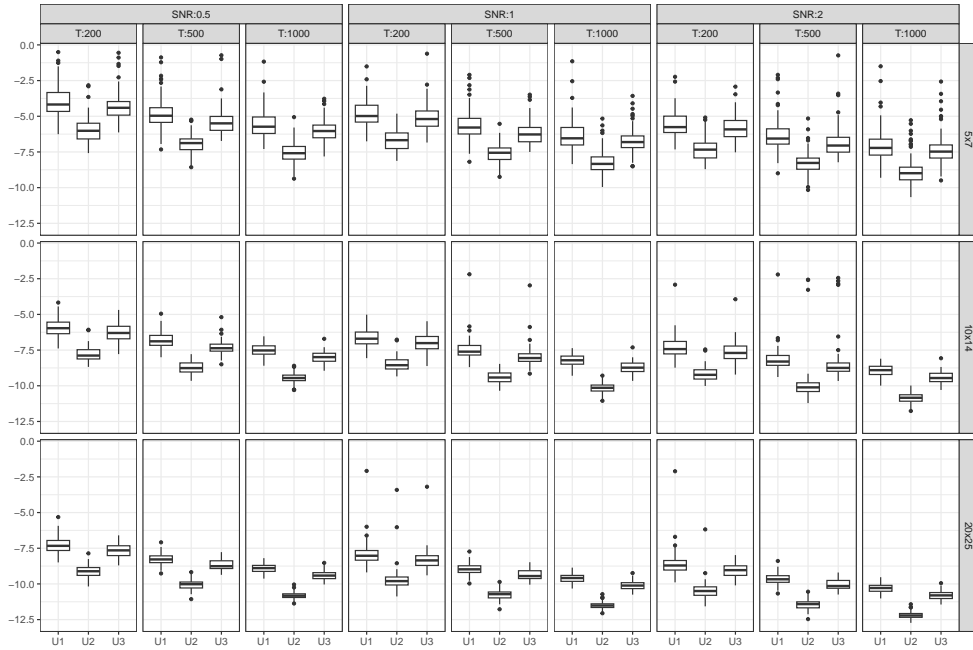


Figure 1: Boxplot of log of MSE for estimating the loading vectors ( $j = 1, 2, 3$ ), for combinations of matrix dimension ( $5 \times 7$ ,  $10 \times 14$ ,  $20 \times 25$ , SNR level (0.5, 1, 2) and time series length (200, 500, 1000).

Next we study the impact of using the estimated factor process as well as its lag variable

as the threshold variable, compared to using the underlying true factor process and its lag threshold variable. To isolate the impact, we fix the true TFM-cp rank  $r = 3$ , use the true AR order each factor as in (16), and set the threshold delay to its true value ( $\tau_j = 1$ ). Treating the estimated  $\hat{f}_{j,t}$  ( $j = 1, 2, 3$ ) as observed time series, we estimate the threshold model (the AR parameters and the threshold value) under the true configuration and the estimated threshold variable  $\hat{f}_{j,t-1}$ . Given the estimated threshold variable  $\hat{f}_{j,t-1}$  and the estimated threshold  $\hat{s}_{1,j}$ , each observation  $\hat{f}_{j,t}$  is classified into one of the two regimes (lower vs. upper). We then compute the proportion of observations for which the inferred regime matches the underlying true regime (i.e.  $f_{i,t-1} < 0$  for the lower regime and  $f_{i,t-1} \geq 0$  for the upper regime). For comparison, we repeat the procedure using the “true” factor process  $f_{i,t}$  and the “true” threshold variable  $f_{i,t-1}$  to estimate the AR coefficients and the threshold  $\tilde{s}_{1,j}$ . Using  $f_{j,t-1}$  and  $\tilde{s}_{1,j}$ , we compute the corresponding correct-classification proportion. Figure 2 shows the comparison of these two proportions under the same model settings as Figure 1. As expected, the results based on the true factor process do not vary across different dimensions or SNR levels, since they are independent of the first-stage TFM-cp estimation. In contrast, the results based on the estimated factor processes approach the performance of the true-factor case as the dimension or SNR increases, reflecting the improved accuracy of the first-stage estimation.

Prediction of  $X_{t+1}$  given  $X_1, \dots, X_t$  is carried out by first forecasting the factor process. To assess the prediction performance, we generated 200 additional observations for each simulation replicate beyond the original sample size  $T$ . Without updating the estimated model parameters, rolling one-step ahead forecasts are produced, and prediction accuracy is assessed using the sum of squared prediction errors

$$\sum_{t=T}^{T+199} \|\widehat{X}_t(1) - X_{t+1}\|_F^2.$$

Here, for  $t \geq T$ ,  $\widehat{X}_t(1) = \sum_{j=1}^3 \hat{u}_{1,j} \hat{u}'_{2,j} \hat{f}_{j,t}(1)$ , where  $\hat{f}_{j,t}(1)$  denotes the one-step-ahead forecast from the estimated TAR model. These forecasts are based on the TFM-cp factor estimates  $\hat{f}_{1,t}, \hat{f}_{2,t}, \hat{f}_{3,t}$ , obtained by projecting  $X_t$  onto the space spanned by the estimated



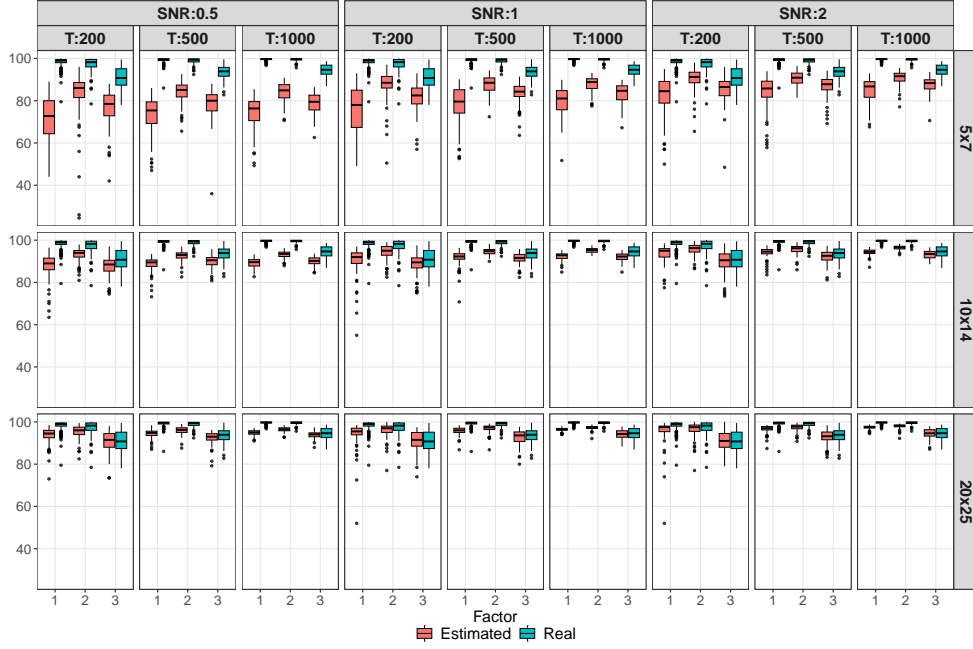


Figure 2: Proportion of correctly identified threshold regimes using the estimated factor process (marked as ‘Estimated’) versus the true factor process (marked as ‘Real’).

loading vectors  $\hat{u}_{1,j}, \hat{u}_{2,j}$ ,  $j = 1, 2, 3$ , which were estimated at  $T$ . For comparison, prediction errors were also computed using the underlying true factor processes  $f_{j,t}$  to build the TAR model and forecast  $f_{j,t+1}$ , which leads to predictions of  $X_{t+1}$ . Figures 3 presents the box plots of the prediction errors, comparing the prediction using estimated loading vectors and the estimated factor process (marked as “Estimated”), and that using the underlying true loading vectors and the underlying true factor processes (marked as “Real”). The results exhibit a pattern similar to that in Figure 2: when the dimension of  $X_t$  is large, the prediction performance based on estimated quantities approaches that obtained using the true underlying factors.

To eliminate the impact of the unpredictable noise in the observed  $X_{t+1}$ , Figure 4 shows the performance of

$$\sum_{t=T}^{T+199} \|\hat{X}_t(1) - M_{t+1}\|_F^2,$$

where  $M_t$  is the signal part in (15). The difference between the two approaches becomes slightly more pronounced, but the overall pattern remains consistent.

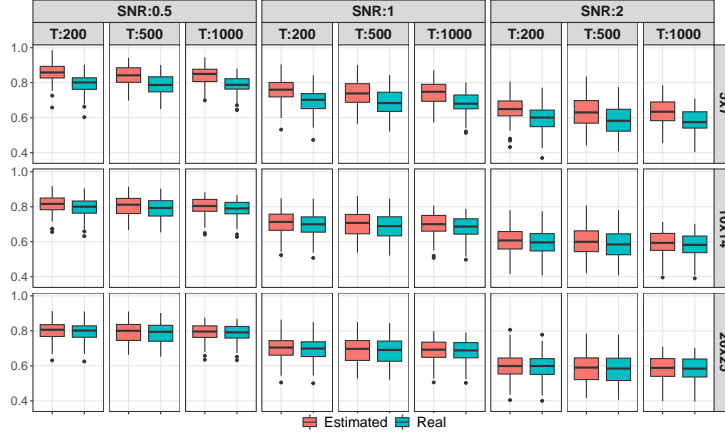


Figure 3: Boxplot of one-step ahead prediction error MSE for the prediction of the observations  $\mathbf{X}_t$  for  $t = T + 1, \dots, T + 200$ . The boxes marked as ‘Estimated’ is obtained using the estimated loading vectors and estimated threshold model parameters obtained at  $T$ . The boxes marked as ‘Real’ is obtained using the underlying true loading vectors and true factor processes.

These simulation results demonstrate the effectiveness of our proposed approach in accurately estimating the loading vectors, identifying regime switching in latent factors, and predicting future observations in tensor time series data. Across a wide range of model settings, including varying dimensionality, sample size, and signal-to-noise ratios, the method consistently delivers reliable estimation and forecasting performance.

## 5 Real data example: Multi-country economic indices

In this section, we demonstrate the proposed T-TFM-cp model by analyzing the matrix time series containing several economic indices from several countries.

### 5.1 Data

We use the dataset from [Chen, Tsay & Chen \(2020\)](#), which contains quarterly observations from 1990Q4 to 2016Q4 from the 14 countries including: United States of America (USA),

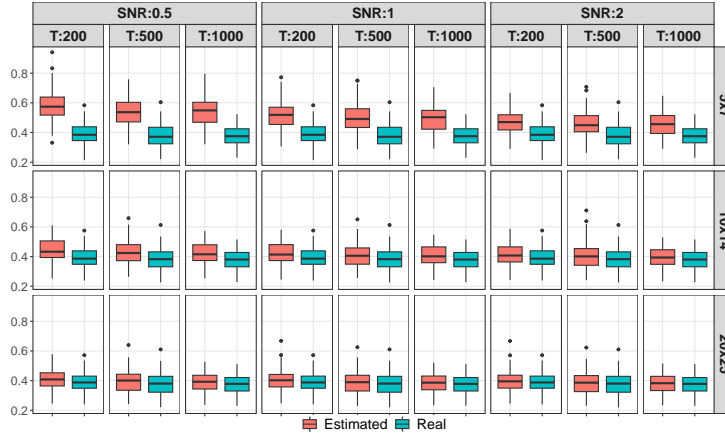


Figure 4: Boxplot of one-step ahead prediction error MSE for the prediction of the signal  $\mathbf{M}_t$  for  $t = T + 1, \dots, T + 200$ . The boxes marked as ‘Estimated’ is obtained using the estimated loading vectors and estimated threshold model parameters obtained at  $T$ . The boxes marked as ‘Real’ is obtained using the underlying true loading vectors and true factor processes.

Canada (CAN), New Zealand (NZL), Australia (AUS), Norway (NOR), Ireland (IRL), Denmark (DNK), United Kingdom (GBR), Finland (FIN), Sweden (SWE), France (FRA), Netherlands (NLD), Austria (AUT), Germany (DEU). Our analysis focuses on ten macroeconomic indicators: CPGDFD.d2lnsa, CPGREN.d2lnsa, CPALTT01.d2lnsa, IRLT.dlv, IR3TIB.dlv, PRINTO01.dln, PRMNT001.dln, LORSGPOR.dln, XTEXVA01.GP, and XTIMVA01.GP.

Table 1 summarizes each series, including its short name, its mnemonic (the series label used in the OECD database), the transformation applied to the series, and a brief description of the data. All data are obtained from the OECD Database. In the transformation column,  $\Delta$  denotes the first difference, and  $\Delta \ln$  denotes the first difference of the logarithm. GP denotes the growth rate measure from the last period.

Figure 5 shows the standardized time series data; the rows and the indices by the columns represent countries.

| Short name | Mnemonic | Tran           | Description  |
|------------|----------|----------------|--|
| CPI: Food  | CPGDFD   | $\Delta^2 \ln$ | Consumer Price Index: Food, seasonally adjusted        |
| CPI: Ener  | CPGREN   | $\Delta^2 \ln$ | Consumer Price Index: Energy, seasonally adjusted      |
| CPI: Tot   | CPALTT01 | $\Delta^2 \ln$ | Consumer Price Index: Total, seasonally adjusted       |
| IR: Long   | IRLT     | $\Delta$       | Interest Rates: Long-term government bond yields       |
| IR: 3-Mon  | IR3TIB   | $\Delta$       | Interest Rates: 3-month Interbank rates and yields     |
| P: TIEC    | PRINTO01 | $\Delta \ln$   | Production: Total industry excluding construction      |
| P: TM      | PRMNT001 | $\Delta \ln$   | Production: Total Manufacturing                        |
| GDP        | LQRSGPOR | $\Delta \ln$   | GDP: Original (Index 2010 = 1.00, seasonally adjusted) |
| IT: Ex     | XTEXVA01 | $\Delta \ln$   | International Trade: Total Export Value (Goods)        |
| IT: IM     | XTIMVA01 | $\Delta \ln$   | International Trade: Total import value (goods)        |

Table 1: Data transformations and variable definitions

## 5.2 Exploratory Analysis

As an exploratory analysis, univariate ARMA model is fitted to each individual time series, using corrected Akaike Information Criterion (AICc) approach ([Hurvich & Tsai 1989](#), [Hyndman & Khandakar 2008](#)) for model determination.

Among the 130 selected univariate ARMA models, 28 are MA(1), 23 are AR(1), 16 are AR of order 4 or 5, 15 are white noise, 10 are ARMA(1,1), and some other less frequently used models. Figure 6 displays the orders of the ARMA models for each time series selected by AICc. Notably, the first three indices (CPI related) show a different pattern from the remaining seven indices.

We stack each observed matrix into a 130 dimensional vector, and estimate a vector factor model (VFM) with two factors using the estimation procedure of [Lam & Yao \(2012\)](#), with  $h = 1$ . The number of factors is determined using the eigen-ratio criterion in [Lam & Yao \(2012\)](#). A TFM-cp model is estimated with two factors, again using the estimation procedure of [Han, Yang, Zhang & Chen \(2024\)](#) with lag 1 ( $h = 1$ ) auto-moments. Two factors are used so to be comparable with the VFM model. Note that the loading matrix in the VFM model uses  $130 \times 2 - 2$  parameters, while that in TFM-cp uses  $(13 - 1 + 10 - 1) \times 2$  parameters for the two (standardized) rank-one  $13 \times 10$  matrices.

Figure 7 shows the estimated factors of VFM (left) and TFM-cp (right). The two

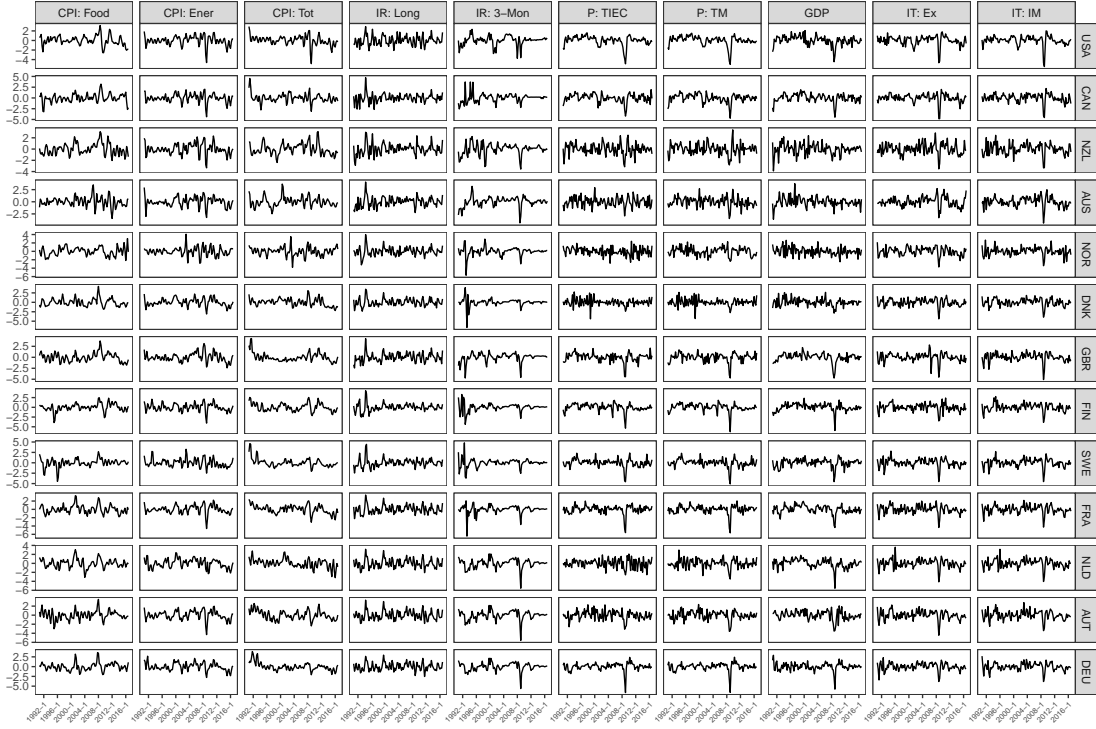


Figure 5: Scaled time series, mean zero and  $s^2 = 1$ .

figures are relatively similar, except for the financial crisis period around the beginning of 2009. During the period, VFM tries to capture the extreme event with one factor, while TFM-cp uses both factors with less magnitude.

To further compare the estimated VFM and TFM-cp, each column of the estimated loading matrix of the VFM model is rearranged into a matrix according to the column and row classifications of the corresponding observed element in the stacked vector. Figure 8 shows the rearranged matrix of the VFM loading vectors (left) and the estimated rank-one loading matrices of TFM-cp model (right). There are some similarity in both models. For example, the first three indices (CPI related) are loaded heavily on the second factor in both models. Index 4 (long interest rate) is very weakly related to Factor 1, but (negatively) loads on the second Factor in VFM, but not for TFM-cp. Most of the indices have negative loadings on Factor 2 for VFM, while the CPI related indices have negative loading on Factor 1 for TFM-cp. The TFM-cp model show more country-wide differences in the loading. For example, in TFM-cp, DNK, NOR and AUS are very weakly loaded on Factor 1 in TFM-cp,

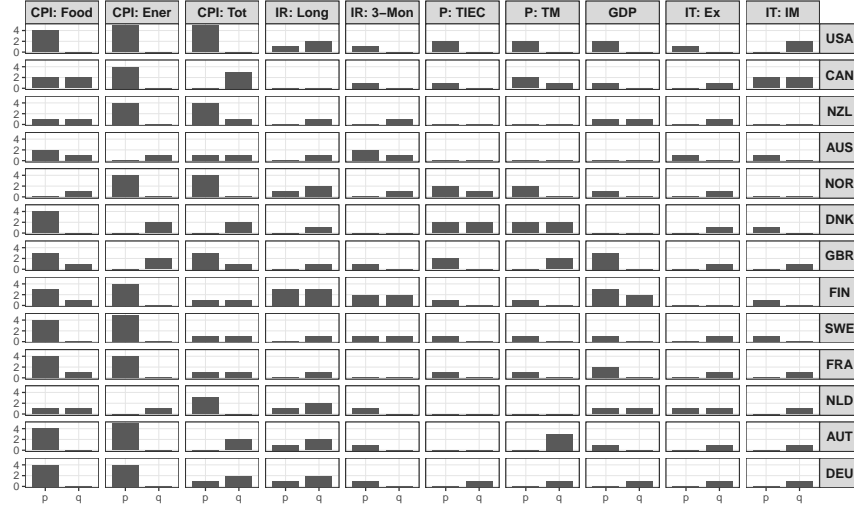


Figure 6: Order of  $ARMA(p, q)$  model for each individual time series, selected by corrected Akaike Information Criterion (AICc)

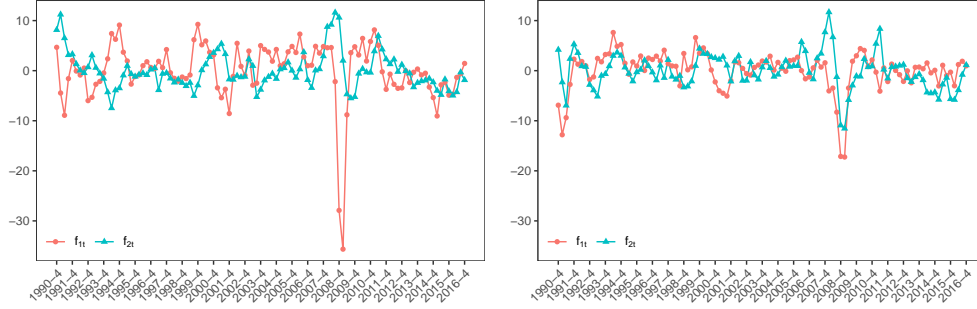


Figure 7: Estimated factors using VFM (left) and TFM-cp (right)

and NOR is weakly loaded on Factor 2.

For each model, we compute the in-sample fitted values and obtain residuals for each time series. Figure 9 reports the in-sample coefficient of determination,  $R^2 = 1 - \sum(y_t - \hat{y}_t)^2 / \sum y_t^2$  for each series under the best ARMA model selected by AICc shown in Figure 6, as well as the VFM and TFM-cp specifications. The overall in-sample  $R^2$  values for the ARMA, VFM, and TFM-cp models are 0.2993, 0.3839, and 0.2937, respectively, using 258,  $257 = 130 \times 2 - 3$ , and  $42 = 2(10 - 1) + (13 - 1)$  parameters.

Figure 9 indicates that the ARMA models provide markedly better fits for the first three CPI related series—albeit with relatively high ARMA orders (see Figure 6)—whereas the

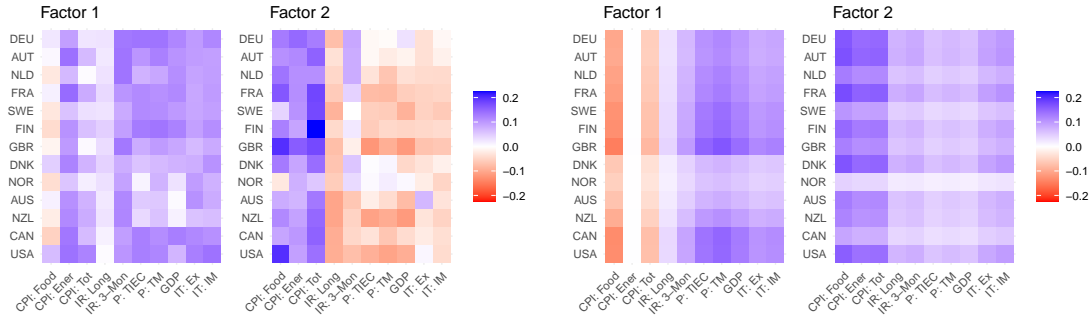


Figure 8: Estimated loading matrices. Left: VFM. Right: TFM-cp.

factor models yield better fits for indices 5–10. None of the models captures the dynamics of the fourth index (IR: Long) satisfactorily.

More specifically, for the first three series, the overall  $R^2$  values are 0.5941 (ARMA), 0.3331 (VFM), and 0.3437 (TFM-cp); for the remaining indices, the corresponding values are 0.1633, 0.3470, and 0.3521. It is worth noting that VFM employs substantially more parameters than TFM-cp.

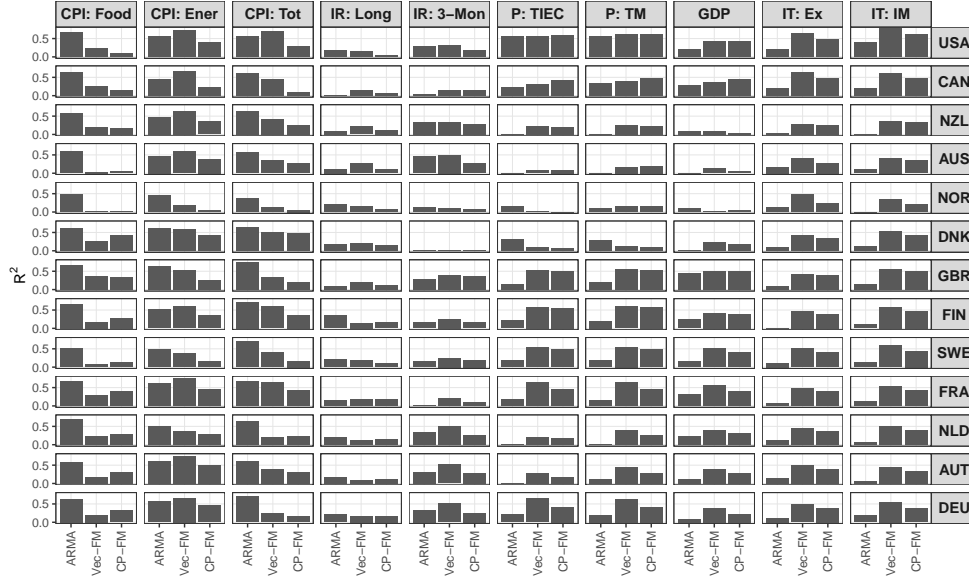


Figure 9: In-sample  $R^2$  for each time series under the AICc selected ARMA model, VFM and TFM-cp models.

### 5.3 Analysis using T-TFM-cp

The VFM and TFM-cp do not specify any temporal structure of the factors. In the following we present results of fitting ARMA and TAR models for the estimated factors under each setting. The results show that there is indeed a threshold phenomenon among the factors.

Table 2 reports the models estimated using the first 65 observations; the remaining 40 observations are reserved for evaluating out-of-sample predictive performance. Based on preliminary exploration, we use the log growth of U.S. GDP,  $z_t = \ln(US.GDP_t) - \ln(US.GDP_{t-1})$  as the threshold variable for all factors. U.S. GDP growth is often used as the threshold variable in economic studies as it is a good representative for the status of the economy (e.g., [Enders et al. 2007](#), [Osińska et al. 2020](#), [Hansen 2011](#), [Tiao & Tsay 1994](#)). Over the estimation period, the resulting threshold variable has a mean of 0.0130 and a standard deviation of 0.0050

The T-TFM-cp estimation results reveal markedly different dynamics across regimes for both factors. Factor 1 exhibits a clear threshold effect at 0.0113 (with delay of  $d = 4$  quarters), splitting the sample into a relatively small regime with 17 observations and a large regime with 44 observations. The upper regime (Regime 1) shows pronounced dynamics, with a strong  $AR(1)$  effect (0.90) and a significant negative  $AR(2)$  term ( $-0.43$ ), indicating substantial oscillatory behavior. In contrast, Regime 2 is well-described by a simple  $AR(1)$  structure with a more moderate coefficient (0.65), suggesting a smoother dynamics. The threshold variable used for factor 2 is the same as factor 1, with a smaller threshold value, resulting that the upper regime has more observations. Note that the overall residual variances and AIC values of both TAR model for factor 1 and factor 2 are much smaller than that of fitting the linear ARMA models to the (same) factors under TFM-cp. The overall variances and AIC values under T-TFM-cp is also smaller than that under VFM using both ARMA and TAR model for the factor processes.

The estimated models using the full sample size is shown in Table 4 in Appendix Appendix B. The T-TFM-cp model under full sample size is very similar to that of using



the first 65 observations. Even though the threshold values are slightly different, the proportions of observations in each regime are similar, showing the relative stability of the estimated models.

Figure 10 illustrates the regimes identified by the T-TFM-cp model. Red dots indicate Regime 1, and blue triangles indicate Regime 2. The top panel shows the threshold variable for reference, while the middle and bottom panels display Factors 1 and 2, respectively. The two factors exhibit slightly different thresholds, with values of 0.0113 for Factor 1 and 0.0121 for Factor 2, highlighting subtle differences in their regime separation.

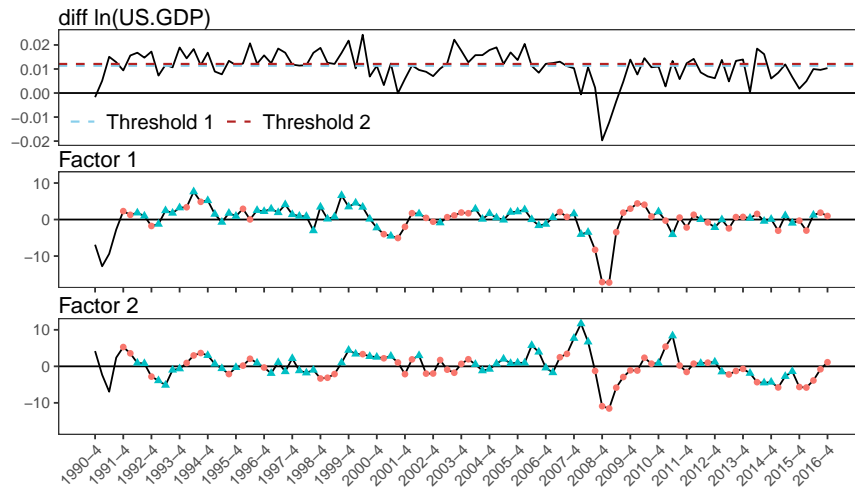


Figure 10: The threshold variable and the estimated factors with regime labels of the T-TFM-cp model.

To assess the performance of each model listed in Table 2, out-sample rolling 1-step prediction errors are obtained for the last 40 observations in the data set, from first quarter of 2007 to forth quarter of 2016. In the rolling prediction, the structure of the models (e.g. number of factors, the estimated loading vectors, the ARMA orders or the Threshold AR orders, and threshold variable, the number of regimes and the threshold values) are fixed as listed in Table 2. Other model parameters, including the factors in the prediction period, the parameters in ARMA or the TAR, are updated using observations 1 to  $t$  for the prediction of  $f_{j,t+1}$  and  $X_{t+1}$ .

Table 3 shows the prediction MSEs over the prediction period under five different models

| Model                        | Series  | Regime                   | Coefficients   | Threshold   | $\sigma^2$ | AIC    |
|------------------------------|---------|--------------------------|--|-------------|------------|--------|
| <b>TFM-cp</b><br><b>ARMA</b> | factor1 | –                        | ar1: 0.72 (0.09)   | –           | 5.77       | 302.13 |
|                              | factor2 | –                        | ma1: 0.68 (0.15)<br>ma2: 0.16 (0.15)   | –           | 4.74       | 290.06 |
| <b>VFM</b><br><b>ARMA</b>    | factor1 | –                        | ma1: 0.71 (0.12)<br>ma2: 0.18 (0.11)   | –           | 10.58      | 342.25 |
|                              | factor2 | –                        | ar1: 0.56 (0.16)<br>ma1: 0.42 (0.20)   | –           | 4.37       | 285.23 |
| <b>TFM-cp</b><br><b>TAR</b>  | factor1 | <b>Regime 1</b> (17 obs) | ar1: 0.90 (0.17)<br>ar2: -0.43 (0.16)  | 0.0113      | 4.18       | 88.07  |
|                              |         | <b>Regime 2</b> (44 obs) | ar1: 0.65 (0.13)   | ( $d = 4$ ) |            |        |
|                              | factor2 | <b>Regime 1</b> (25 obs) | ar1: 1.1 (0.21)<br>ar2: -0.67 (0.19)<br>ar3: 0.50 (0.20)<br>ar4: -0.58 (0.18)  | 0.0121      | 3.25       | 76.49  |
|                              |         | <b>Regime 2</b> (36 obs) | ar1: 0.57 (0.12)   | ( $d = 4$ ) |            |        |
| <b>VFM</b><br><b>TAR</b>     | factor1 | <b>Regime 1</b> (38 obs) | ar1: 0.61 (0.15)<br>ar2: -0.30 (0.18)<br>ar3: 0.33 (0.17)<br>ar4: -0.56 (0.16) | 0.0147      | 7.62       | 128.21 |
|                              |         | <b>Regime 2</b> (23 obs) | ar1: 0.88 (0.14)   | ( $d = 4$ ) |            |        |
|                              | factor2 | <b>Regime 1</b> (53 obs) | ar1: 0.63 (0.13)<br>ar2: -0.10 (0.13)  | 0.0179      | 3.41       | 80.15  |
|                              |         | <b>Regime 2</b> (10 obs) | ar1: 1.09 (0.22)   | ( $d = 2$ ) |            |        |

Table 2: Estimated parameters of ARMA and TAR models fitted to the factor process  $\hat{f}_{1t}$ , based on the first 65 observations (indices 1–10).

(individual ARMA, ARMA-VFM, ARMA-TFM-cp, T-VFM and T-TFM-cp). In addition, the table also reports the overall MS of the original data in the prediction period, and the in-sample MSE of TFM-cp in the same period, computed using the loading vectors estimated from the full dataset, as references. In the last two rows, we report the prediction performance of the first three indices and the others separately.

From the table, it is seen that T-TFM-cp performed the best considering all indices and indices 4-10 only. The individual ARMA models perform the best for the three CPI indices, similar to that has been revealed in Figure 9. It is also evident that there is indeed a threshold phenomenon and using the threshold model is useful, as seen from the comparison between ARMA-TFM-cp and T-TFM-cp, and the comparison between ARMA-VFM and

T-VFM.

| Indices | $MS(X_t)$ | TFM-cp<br>(MSE) | ARMA   | VFM<br>ARMA | VFM<br>TAR | TFM-cp<br>ARMA | TFM-cp<br>TAR |
|---------|-----------|-----------------|--------|-------------|------------|----------------|---------------|
| All     | 1.1937    | 0.6541          | 0.8494 | 1.0974      | 0.9301     | 1.0222         | 0.8218        |
| 4-10    | 1.1606    | 0.6183          | 0.9907 | 1.0588      | 0.9264     | 1.0583         | 0.7735        |
| 1-3     | 1.2710    | 0.7376          | 0.5196 | 1.1873      | 0.9387     | 0.9382         | 0.9344        |

Table 3: Prediction performance comparison: one-step rolling prediction mean squared error of the last 40 observations of five different models. The sample mean squares of observations in the prediction period and the in-sample mean square errors of TFM-cp in the prediction period are included as references.

Figure 11 shows the prediction MSE of each of the 40 prediction times. It seems that for the economic indices 4 to 10, the five models are comparable except during the financial crisis period (mid 2008 to mid 2009), in which T-TFM-cp out-performs ARMA and ARMA-TFM-cp in 2008, and ARMA-VFM in 2009. For the CPI related indices, individual ARMA models perform the best except for the second half of year 2008,

## 6 Conclusions

In this paper we proposed a threshold Tensor factor model in CP form, in which the latent factor processes in a TFM-cp model are assumed to follow threshold AR models. It effectively provides a threshold type dynamics for high dimensional tensor time series that allows regime change according to the threshold variable. The model also allows prediction capability while enjoys significant dimension reduction using the factor model structure.

While we have shown that the proposed model is useful in empirical studies, there are challenging issues that are worth further investigation. For example, the two-step estimation procedure works only when the signal to noise ratio in the factor model part is sufficiently high. An effective joint estimation procedure may be needed when the signal to noise ratio is relatively smaller. Hunting for an effective threshold variable is always a

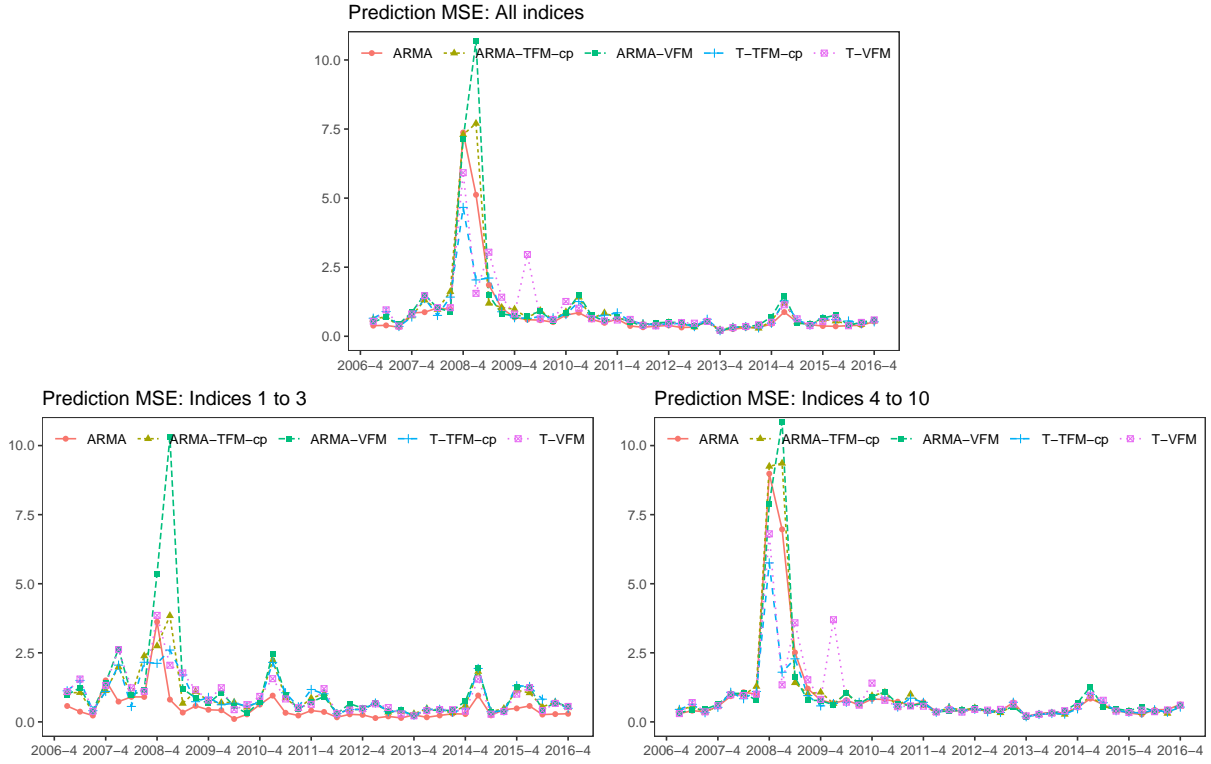


Figure 11: Prediction MSE at each  $t$  under five different methods.

challenge in threshold modeling.

## References

- Ang, A. & Timmermann, A. (2012), ‘Regime changes and financial markets’, *Annual Review of Financial Economics* **4**(1), 313–337.
- Bai, J. (2003), ‘Inferential theory for factor models of large dimensions’, *Econometrica* **71**(1), 135–171.
- Bai, J. & Ng, S. (2002), ‘Determining the number of factors in approximate factor models’, *Econometrica* **70**(1), 191–221.
- Barigozzi, M. & Hallin, M. (2016), ‘Generalized dynamic factor models and volatilities: recovering the market volatility shocks’, *The Econometrics Journal* **19**(1), C33–C60.

- Barigozzi, M., La Vecchia, D. & Liu, H. (2025), ‘General spatio-temporal factor models for high-dimensional random fields on a lattice’, *The Annals of Statistics* **53**(1), 268–294.
- Bolivar, S., Huang, S.-C. & Chen, R. (2025), ‘Analysis of tensor time series’, *Annual Review of Statistics and Its Application* **13**.
- Bucci, A. (2024), ‘A smooth transition autoregressive model for matrix-variate time series’, *Computational Economics* pp. 1–30.
- Chan, K. (1989), ‘A note on the geometric ergodicity of a markov chain’, *Advances in Applied Probability* **21**(3), 702–704.
- Chan, K.-S. (1993), ‘Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model’, *The Annals of Statistics* pp. 520–533.
- Chan, K. S. & Tong, H. (1985), ‘On the use of the deterministic lyapunov function for the ergodicity of stochastic difference equations’, *Advances in Applied Probability* **17**(3), 666–678.
- Chan, K.-S. & Tsay, R. S. (1998), ‘Limiting properties of the least squares estimator of a continuous threshold autoregressive model’, *Biometrika* **85**(2), 413–426.
- Chan, K. & Tong, H. (1986), ‘On estimating thresholds in autoregressive models’, *Journal of Time Series Analysis* **7**, 179–190.
- Chang, J., He, J., Yang, L. & Yao, Q. (2023), ‘Modelling matrix time series via a tensor cp-decomposition’, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **85**(1), 127–148.
- Chen, B., Han, Y. & Yu, Q. (2024), ‘Estimation and inference for CP tensor factor models’, *arXiv preprint arXiv:2406.17278* .
- Chen, B., Han, Y. & Yu, Q. (2025), ‘Diffusion index forecast with tensor data’, *arXiv preprint arXiv:2511.02235* .

- Chen, E. Y. & Fan, J. (2023), ‘Statistical inference for high-dimensional matrix-variate factor models’, *Journal of the American Statistical Association* **118**(542), 1038–1055.
- Chen, E. Y., Tsay, R. S. & Chen, R. (2020), ‘Constrained factor models for high-dimensional matrix-variate time series’, *Journal of the American Statistical Association* **115**(530), 775–793.
- Chen, E. Y., Yun, X., Chen, R. & Yao, Q. (2020), ‘Modeling multivariate spatial-temporal data with latent low-dimensional dynamics’, *arXiv preprint arXiv:2002.01305* .
- Chen, R., Xiao, H. & Yang, D. (2021), ‘Autoregressive models for matrix-valued time series’, *Journal of Econometrics* **222**(1, Part B), 539–560.
- Chen, R., Yang, D. & Zhang, C.-H. (2022), ‘Factor models for high-dimensional tensor time series (with discussion)’, *Journal of the American Statistical Association* **117**(537), 94–116.
- De Gooijer, J. G. et al. (2017), *Elements of nonlinear time series analysis and forecasting*, Vol. 37, Springer.
- Enders, W., Falk, B. & Siklos, P. L. (2007), ‘A threshold model of real US GDP and the problem of constructing confidence intervals in tar models’, *Studies in Nonlinear Dynamics and Econometrics*, *Forthcoming* .
- Fan, J. & Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer Series in Statistics, Springer-Verlag, New York.
- Gonzalo, J. & Pitarakis, J.-Y. (2002), ‘Estimation and model selection based inference in single and multiple threshold models’, *Journal of Econometrics* **110**(2), 319–352.
- Gourieroux, C. & Jasiak, J. (2001), ‘Dynamic factor models’, *Econometric Reviews* **20**(4), 385–424.

- Han, Y., Chen, R., Yang, D. & Zhang, C.-H. (2024), ‘Tensor factor model estimation by iterative projection’, *The Annals of Statistics* **52**(6), 2641–2667.
- Han, Y., Chen, R. & Zhang, C.-H. (2022), ‘Rank determination in tensor factor model’, *Electronic Journal of Statistics* **16**(1), 1726–1803.
- Han, Y., Yang, D., Zhang, C.-H. & Chen, R. (2024), ‘CP factor model for dynamic tensors’, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **86**(5), 1383–1413.
- Hansen, B. E. (2000), ‘Sample splitting and threshold estimation’, *Econometrica* **68**(3), 575–603.
- Hansen, B. E. (2011), ‘Threshold autoregression in economics’, *Statistics and its Interface* **4**(2), 123–127.
- Hurvich, C. M. & Tsai, C.-L. (1989), ‘Regression and time series model selection in small samples’, *Biometrika* **76**(2), 297–307.
- Hyndman, R. J. & Khandakar, Y. (2008), ‘Automatic time series forecasting: the forecast package for R’, *Journal of Statistical Software* **27**, 1–22.
- Kolda, T. G. & Bader, B. W. (2009), ‘Tensor decompositions and applications’, *SIAM review* **51**(3), 455–500.
- Lam, C. & Yao, Q. (2012), ‘Factor modeling for high-dimensional time series: inference for the number of factors’, *The Annals of Statistics* **40**(2), 694–726.
- Lam, C., Yao, Q. & Bathia, N. (2011), ‘Estimation of latent factors for high-dimensional time series’, *Biometrika* **98**(4), 901–918.
- Li, D., Li, W. K. & Ling, S. (2011), ‘On the least squares estimation of threshold autoregressive moving-average models’, *Statistics and its Interface* .

- Li, D. & Ling, S. (2012), ‘On the least squares estimation of multiple-regime threshold autoregressive models’, *Journal of Econometrics* **167**(1), 240–253.
- Li, D. & Tong, H. (2016), ‘Nested sub-sample search algorithm for estimation of threshold models’, *Statistica Sinica* pp. 1543–1554.
- Li, Z. & Xiao, H. (2021), ‘Multi-linear tensor autoregressive models’, *arXiv preprint arXiv:2110.00928* .
- Li, Z. & Xiao, H. (2024), ‘Cointegrated matrix autoregression models’, *arXiv preprint arXiv:2409.10860* .
- Liu, X. & Chen, E. Y. (2022), ‘Identification and estimation of threshold matrix-variate factor models’, *Scandinavian Journal of Statistics* **49**(3), 1383–1417.
- Liu, X. & Chen, R. (2016), ‘Regime-switching factor models for high-dimensional time series’, *Statistica Sinica* pp. 1427–1451.
- Liu, X. & Chen, R. (2020), ‘Threshold factor models for high-dimensional time series’, *Journal of Econometrics* **216**(1), 53–70.
- Luo, Z., Han, Y. & Yu, X. (2025), ‘Supervised dynamic dimension reduction with deep neural network’, *arXiv preprint arXiv:2508.03546* .
- Meyn, S. P. & Tweedie, R. L. (2012), *Markov chains and stochastic stability*, Springer Science & Business Media.
- Osińska, M., Kufel, T., Błażejowski, M. & Kufel, P. (2020), ‘Modeling mechanism of economic growth using threshold autoregression models’, *Empirical Economics* **58**(3), 1381–1430.
- Otto, S. & Salish, N. (2022), ‘Approximate factor models for functional time series’, *arXiv preprint arXiv:2201.02532* .



- Rosenblatt, M. (2012), *Markov Processes, Structure and Asymptotic Behavior: Structure and Asymptotic Behavior*, Vol. 184, Springer Science & Business Media.
- Sims, C. & Zha, T. (2006), ‘Were there regime switches in US monetary policy?’, *The American Economic Review* pp. 54–81.
- Stock, J. & Watson, M. (2016), Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics, in J. B. Taylor & H. Uhlig, eds, ‘Handbook of Macroeconomics’, Vol. 2 of *Handbook of Macroeconomics*, Elsevier, chapter 0, pp. 415–525.
- Tiao, G. C. & Tsay, R. S. (1994), ‘Some advances in non-linear and adaptive modelling in time-series’, *Journal of Forecasting* **13**(2), 109–131.
- Tong, H. (1990a), *Non-linear Time Series: a Dynamical System Approach*, Oxford University Press.
- Tong, H. (1990b), *Nonlinear Time Series Analysis: A Dynamical System Approach*, London: Oxford University Press.
- Tong, H. (2010), ‘Threshold models in time series analysis 30-years on’, *Reserch Report, University of Hong Kong* **471**(1-38).
- Tong, H. & Lim, K. S. (1980), ‘Threshold autoregression, limit cycles and cyclical data (with discussion)’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **42**, 245–292.
- Tsay, R. S. (1989), ‘Testing and modeling threshold autoregressive processes’, *Journal of the American Statistical Association* **84**, 231–240.
- Tsay, R. S. (2005), *Analysis of Financial Time Series*, Vol. 543, John Wiley & Sons.
- Tsay, R. S. & Chen, R. (2018), *Nonlinear Time Series Analysis*, John Wiley & Sons.

- Wang, D., Liu, X. & Chen, R. (2019), ‘Factor models for matrix-valued high-dimensional time series’, *Journal of Econometrics* **208**(1), 231–248.
- Wang, D., Zheng, Y. & Li, G. (2024), ‘High-dimensional low-rank tensor autoregressive time series modeling’, *Journal of Econometrics* **238**(1), 105544.
- Wu, F. & Chan, K.-S. (2023), ‘Mixture matrix-valued autoregressive model’, *arXiv preprint arXiv:2312.06098* .
- Wu, S. & Chen, R. (2007), ‘Threshold variable selection and threshold variable driven switching autoregressive models’, *Statistica Sinica* **17**, 241–264.
- Xiao, H., Han, Y., Chen, R. & Liu, C. (2023), ‘Reduced rank autoregressive models for matrix time series’, *Working Paper* .
- Yu, C., Li, D., Zhang, X. & Tong, H. (2024), ‘Two-way threshold matrix autoregression’, *arXiv preprint arXiv:2407.10272* .
- Yu, R., Chen, R., Xiao, H. & Han, Y. (2024), ‘Dynamic matrix factor models for high dimensional time series’, *arXiv preprint arXiv:2407.05624* .
- Zhang, X., Li, D. & Tong, H. (2024), ‘On the least squares estimation of multiple-threshold-variable autoregressive models’, *Journal of Business & Economic Statistics* **42**(1), 215–228.
- Zhou, G., Han, Y. & Yu, X. (2025), Factor augmented tensor-on-tensor neural networks, in ‘Proceedings of the AAAI Conference on Artificial Intelligence’, Vol. 39, pp. 22928–22936.

# Supplementary Material to “Threshold Tensor Factor Model in CP Form”

In this supplementary material, Appendix A provides the complete proofs of the main theorems presented in the article, while Appendix B presents the estimated models using the full dataset for the real data application.

## A Proofs of Main Theorems

**Proof of Proposition 3.1.** The proof of Proposition 3.1 can be derived using similar arguments to those in Theorem 4.3 of [Chen et al. \(2024\)](#) and Theorem 3 of [Han, Yang, Zhang & Chen \(2024\)](#).  $\square$

We first provide a theorem for the consistency of TAR parameters. Given that the least squares estimation of TAR involves a non-convex optimization problem, this result forms the basis for further analysis of the statistical convergence rate and asymptotic normality.

**Theorem A.1.** Suppose Assumptions 1, 2, 3, 4 hold. Assume that  $\phi_j^{(\ell)} \neq \phi_j^{(\ell+1)}$  for  $1 \leq j \leq r, 1 \leq \ell \leq L - 1$ . Let

$$\frac{\sigma \sqrt{d_{\max}}}{\lambda \sqrt{T}} + \frac{\sigma^2}{\lambda^2} \rightarrow 0,$$

as  $T \rightarrow \infty$ .

- (i) Suppose  $z_{jt}$  is self-exciting with  $z_{jt} = f_{j,t-\tau_j}$ , for some  $1 \leq j \leq r$ . Let  $\theta_j = (\phi_j^\top, s_j^\top, \tau_j)^\top$ . Then, in an event with probability at least  $1 - T^{-c} - \sum_{k=1}^K e^{-d_k}$ ,  $\hat{\theta}_j \rightarrow \theta_j$  as  $T \rightarrow \infty$ .
- (ii) Suppose  $z_{jt}$  is an observable variable with a bounded, continuous and positive density on  $\mathbb{R}$ , and is Markovian, for some  $1 \leq j \leq r$ . Let  $\theta_j = (\phi_j^\top, s_j^\top)^\top$ . Then, in an event with probability at least  $1 - T^{-c} - \sum_{k=1}^K e^{-d_k}$ ,  $\hat{\theta}_j \rightarrow \theta_j$  as  $T \rightarrow \infty$ .

**Proof of Theorem A.1.** We only prove part (i). The proof of part (ii) is similar. Define the squared loss functions with estimated and true factors,

$$\widehat{L}_T(\theta_j) = \sum_{t=p_j+1}^T \left( \widehat{f}_{j,t} - \sum_{\ell}^L I(z_{j,t} \in (s_{j,\ell-1}, s_{j,\ell})) \sum_{k=1}^{p_j} \phi_{j,k}^{(\ell)} \widehat{f}_{j,t-k} \right)^2, \quad (17)$$

$$L_T(\theta_j) = \sum_{t=p_j+1}^T \left( f_{j,t} - \sum_{\ell}^L I(z_{j,t} \in (s_{j,\ell-1}, s_{j,\ell})) \sum_{k=1}^{p_j} \phi_{j,k}^{(\ell)} f_{j,t-k} \right)^2, \quad (18)$$

and define the residuals

$$\widehat{\varepsilon}_t(\theta_j) = \widehat{f}_{j,t} - \sum_{\ell}^L I(z_{j,t} \in (s_{j,\ell-1}, s_{j,\ell})) \sum_{k=1}^{p_j} \phi_{j,k}^{(\ell)} \widehat{f}_{j,t-k}, \quad (19)$$

$$\varepsilon_t(\theta_j) = f_{j,t} - \sum_{\ell}^L I(z_{j,t} \in (s_{j,\ell-1}, s_{j,\ell})) \sum_{k=1}^{p_j} \phi_{j,k}^{(\ell)} f_{j,t-k}. \quad (20)$$

The following two Lemmas can be obtained from Zhang et al. (2024), Li & Ling (2012).

**Lemma A.2.** If the conditions in Theorem A.1 hold, then  $\mathbb{E}\varepsilon_t^2(\tilde{\theta}_j) \geq \mathbb{E}\varepsilon_t^2(\theta_j)$  for all  $\tilde{\theta}_j \in \Theta$ , and the equality holds if and only if  $\tilde{\theta}_j = \theta_j$ .

**Lemma A.3.** If the conditions of Theorem A.1 hold, then, for any  $\tilde{\theta}_j \in \Theta$  and its open neighborhood  $U_{\tilde{\theta}_j}$ ,

$$\mathbb{E} \sup_{\theta_j^* \in U_{\tilde{\theta}_j}} |\varepsilon_t^2(\theta_j^*) - \varepsilon_t^2(\tilde{\theta}_j)| \rightarrow 0 \quad \text{as } U_{\tilde{\theta}_j} \text{ shrinks to } \tilde{\theta}_j.$$

According to Proposition 3.1, we have in an event  $\Omega$  with probability at least  $1 - T^{-c} - \sum_{k=1}^K e^{-d_k}$ ,

$$\left| \frac{1}{T-h} \sum_{t=h+1}^T \widehat{f}_{i,t-h} \widehat{f}_{jt} - \frac{1}{T-h} \sum_{t=h+1}^T f_{i,t-h} f_{jt} \right| \leq c \left( \frac{\sigma \sqrt{d_{\max}}}{\lambda \sqrt{T}} + \frac{\sigma^2}{\lambda^2} \right) =: \delta. \quad (21)$$

In the following proof, we restrict to the event  $\Omega$ .

For any given open neighborhood  $V$  of  $\theta_j \in \Theta$  and any  $\vartheta \in V^c = \Theta \setminus V$ ,  $\mathbb{E}\varepsilon_t^2(\vartheta) > \mathbb{E}\varepsilon_t^2(\theta_j)$  by Lemma A.2. Lemma A.3 implies that  $\mathbb{E}\varepsilon_t^2(\vartheta)$  is continuous in  $\vartheta$ . Applying the compactness of  $V^c$ , there exists a  $\vartheta_0 \in V^c$  such that

$$\inf_{\vartheta \in V^c} \mathbb{E}\varepsilon_t^2(\vartheta) = \mathbb{E}\varepsilon_t^2(\vartheta_0) > \mathbb{E}\varepsilon_t^2(\theta_j).$$

Set  $\kappa_0 = (\mathbb{E}\varepsilon_t^2(\vartheta_0) - \mathbb{E}\varepsilon_t^2(\theta_j))/3 > 0$ . For any  $\vartheta \in V^c$ , by Lemma A.3 again, there exists  $\eta > 0$  such that

$$\mathbb{E} \inf_{\vartheta^* \in U_{\vartheta}(\eta)} \varepsilon_t^2(\vartheta^*) \geq \mathbb{E}\varepsilon_t^2(\vartheta) - \kappa_0 \geq \mathbb{E}\varepsilon_t^2(\vartheta_0) - \kappa_0 = 2\kappa_0 + \mathbb{E}\varepsilon_t^2(\theta_j). \quad (22)$$

Since  $V^c$  is compact, there exists a finite covering of  $V^c$ :  $\{U_{\vartheta_j}(\eta), \vartheta_j \in V^c, j = 1, 2, \dots, J\}$  such that  $\bigcup_{j=1}^J U_{\vartheta_j}(\eta) = V^c$ . By the ergodic theorem and (22), we have a.s. for sufficiently large  $T$  and  $1 \leq j \leq J$

$$\inf_{\vartheta^* \in U_{\vartheta_j}(\eta)} \frac{1}{T} \widehat{L}_T(\vartheta^*) \geq \frac{1}{T} \sum_{t=1}^T \inf_{\vartheta^* \in U_{\vartheta_j}(\eta)} \widehat{\varepsilon}_t^2(\vartheta^*) \geq \mathbb{E} \inf_{\vartheta^* \in U_{\vartheta_j}(\eta)} \varepsilon_t^2(\vartheta^*) - \kappa_0 - c_1 \delta \geq \kappa_0 + \mathbb{E}\varepsilon_t^2(\theta_j) - c_1 \delta$$

and

$$\inf_{\vartheta \in V} \frac{1}{T} \widehat{L}_T(\vartheta) \leq \frac{1}{T} \widehat{L}_T(\theta_j) = \frac{1}{T} \sum_{t=1}^T \widehat{\varepsilon}_t^2(\theta_j) \leq \mathbb{E}\varepsilon_t^2(\theta_j) + \frac{\kappa_0}{2} + c_1 \delta.$$

Therefore, as  $\delta = o(1)$ , for any neighborhood  $V$  of  $\theta_j \in \Theta$ , it follows that for sufficiently large  $T$

$$\inf_{\vartheta^* \in V^c} \frac{1}{T} \widehat{L}_T(\vartheta^*) \geq \min_{1 \leq j \leq J} \inf_{\vartheta^* \in U_{\vartheta_j}(\eta)} \frac{1}{T} \widehat{L}_T(\vartheta^*) \geq \kappa_0 + \mathbb{E}\varepsilon_t^2(\theta_j) - c_1 \delta > \inf_{\vartheta \in V} \frac{1}{T} \widehat{L}_T(\vartheta),$$

which implies that  $\widehat{\theta}_j \in V$  a.s. By the arbitrariness of  $V$ , we can get  $\widehat{\theta}_j \rightarrow \theta_j$  in the high probability event  $\Omega$ . The proof is complete.  $\square$

**Proof of Theorem 3.1.** According to Proposition 3.1, we have in an event  $\Omega$  with probability at least  $1 - T^{-c} - \sum_{k=1}^K e^{-d_k}$ ,

$$|\widehat{f}_{jt} - f_{jt}| \leq c \left( \frac{\sigma \sqrt{d_{\max}}}{\lambda \sqrt{T}} + \frac{\sigma}{\lambda} \right) =: \vartheta_1, \quad (23)$$

for all  $1 \leq t \leq T$ , and

$$\left| \frac{1}{T-h} \sum_{t=h+1}^T \widehat{f}_{i,t-h} \widehat{f}_{jt} - \frac{1}{T-h} \sum_{t=h+1}^T f_{i,t-h} f_{jt} \right| \leq c \left( \frac{\sigma \sqrt{d_{\max}}}{\lambda \sqrt{T}} + \frac{\sigma^2}{\lambda^2} \right) =: \vartheta_2, \quad (24)$$

for all  $1 \leq i, j \leq r$  and  $0 \leq h \leq T/4$ . In the following proof, we restrict to the event  $\Omega$ . For simplicity of presentation, we omit the sub-subscripts and use  $\theta, \widehat{\theta}, \phi, s, \tau, f_t$

to represent  $\theta_j, \hat{\theta}_j, \phi_j, s_j, \tau_j, f_{jt}$ , etc. Without loss of generality, consider two regime TAR models and let  $L = 2$ . The extension to general  $L$  follows similarly.

By Theorem A.1, in the event  $\Omega, \hat{\theta} \rightarrow \theta$ . Thus, in  $\Omega, \hat{\tau}_j = \tau_j$ , and we restrict to the parameter space of an open neighborhood of  $\theta$ . Define  $V_\delta = \{\tilde{\theta} \in \Theta : \|\tilde{\phi} - \phi\| < \delta, |\tilde{s} - s| < \delta\}$  for some  $0 < \delta < 1$  to be determined later.

Define the residuals

$$\hat{\varepsilon}_t(\phi_0, s_0) = \hat{f}_t - \sum_{\ell=1}^2 I(\hat{f}_{t-\tau} \in (s_{0,\ell-1}, s_{0,\ell})) \sum_{k=1}^{p_j} \phi_{0k}^{(\ell)} \hat{f}_{t-k}, \quad (25)$$

$$\varepsilon_t(\phi_0, s_0) = f_t - \sum_{\ell=1}^2 I(f_{t-\tau} \in (s_{0,\ell-1}, s_{0,\ell})) \sum_{k=1}^{p_j} \phi_{0k}^{(\ell)} f_{t-k}, \quad (26)$$

with  $s_{0,0} = -\infty, s_{0,2} = +\infty$ , and the squared loss functions with estimated and true factors,

$$\hat{L}_T(\phi_0, s_0) = \sum_{t=p_j+1}^T \hat{\varepsilon}_t^2(\phi_0, s_0), \quad (27)$$

$$L_T(\phi_0, s_0) = \sum_{t=p_j+1}^T \varepsilon_t^2(\phi_0, s_0). \quad (28)$$

Note that the difference of  $\hat{\varepsilon}_t(\phi_0, s_0)$  and  $\hat{\varepsilon}_t(\phi_0, s)$  is nonzero only in the region  $\mathbf{1}\{s_0 \wedge s < \hat{f}_{t-\tau} \leq s_0 \vee s\}$ . Here we only treat the case  $s_0 > s$ . Proofs of the other case is similar. Write  $s_0 = s + u$  for some  $0 < u < 1$ , and let  $\hat{\alpha}_t^{(\ell)} = (\hat{f}_t - \phi^{(\ell)\top} \hat{f}_{t-1})^2$ ,  $\hat{\alpha}_t^{(\ell 0)} = (\hat{f}_t - \phi_0^{(\ell)\top} \hat{f}_{t-1})^2$  and  $\alpha_t^{(\ell)} = (f_t - \phi^{(\ell)\top} f_{t-1})^2$ ,  $\alpha_t^{(\ell 0)} = (f_t - \phi_0^{(\ell)\top} f_{t-1})^2$ , where  $\hat{f}_{t-1} = (\hat{f}_{t-1}, \dots, \hat{f}_{t-p_j})^\top$ .

Then

$$\begin{aligned} R_T(\phi_0, u) &= \hat{L}_T(\phi_0, s_0) - \hat{L}_T(\phi_0, s) = \sum_{t=p_j+1}^T (\hat{\alpha}_t^{(20)} - \hat{\alpha}_t^{(10)}) \mathbf{1}\{s < \hat{f}_{t-\tau} \leq s + u\} \\ &= [\hat{L}_T(\phi, s_0) - \hat{L}_T(\phi, s)] + \{[\hat{L}_T(\phi_0, s_0) - \hat{L}_T(\phi, s_0)] - [\hat{L}_T(\phi_0, s) - \hat{L}_T(\phi, s)]\} \\ &:= R_{T1}(u) + R_{T2}(\phi_0, u), \end{aligned}$$

where

$$R_{T1}(u) = \sum_{t=p_j+1}^T (\hat{\alpha}_t^{(2)} - \hat{\alpha}_t^{(1)}) \mathbf{1}\{s < \hat{f}_{t-\tau} \leq s+u\},$$

$$R_{T2}(\phi_0, u) = \sum_{t=p_j+1}^T [(\hat{\alpha}_t^{(20)} - \hat{\alpha}_t^{(2)}) - (\hat{\alpha}_t^{(10)} - \hat{\alpha}_t^{(1)})] \mathbf{1}\{s < \hat{f}_{t-\tau} \leq s+u\}.$$

For  $R_{T1}(u)$ , by Assumption 6, for all  $i \neq j$ , there exist some positive constants  $c_0$  and  $\rho$  such that  $|(\phi^{(i)} - \phi^{(j)})^\top f_{t-1}| \geq c_0 > 0$  for all  $\|f_{t-1} - w_1\| \leq \rho$ . Then

$$[(\phi^{(i)} - \phi^{(j)})^\top f_{t-1}]^2 + 2\xi_t(\phi^{(i)} - \phi^{(j)})^\top f_{t-1} \geq c_0^2 \mathbf{1}\{\|f_{t-1} - w_1\| \leq \rho\} + 2\xi_t(\phi^{(i)} - \phi^{(j)})^\top f_{t-1}. \quad (29)$$

Let  $\varpi := \max_{i \neq j} \|\phi^{(i)} - \phi^{(j)}\|$ . By (29) and Lemma A.4, for  $R_{T1}(u)$ , we have

$$R_{T1}(u) \geq c_0^2 G_T^*(u - 2\vartheta_1) - 2\varpi \left( \left| \sum_{t=p_j+1}^T A_t(u + 2\vartheta_1) \right| + \left| \sum_{t=p_j+1}^T D_t(u + 2\vartheta_1) \right| \right) - 3(\varpi^2 + 2\varpi)\vartheta_1 G_T(u + 2\vartheta_1),$$

where  $G_T^*(u - 2\vartheta_1)$ ,  $A_t(u + 2\vartheta_1)$ ,  $D_t(u + 2\vartheta_1)$ , and other notations used below are defined in Lemma A.4. Thus

$$\begin{aligned} & \inf_{\frac{B}{T} < u < \delta} \frac{R_{T1}(u)}{TG(u - 2\vartheta_1)} \\ & \geq c_0^2 \inf_{\frac{B}{T} < u < \delta} \frac{G_T^*(u - 2\vartheta_1)}{TG^*(u - 2\vartheta_1)} \frac{G^*(u - 2\vartheta_1)}{G(u - 2\vartheta_1)} \\ & \quad - 2\varpi \sup_{\frac{B}{T} < u < \delta} \frac{\left| \sum_{t=p_j+1}^T A_t(u + 2\vartheta_1) \right| + \left| \sum_{t=p_j+1}^T D_t(u + 2\vartheta_1) \right|}{TG(u + 2\vartheta_1)} \frac{G(u + 2\vartheta_1)}{G(u - 2\vartheta_1)} \\ & \quad - \inf_{\frac{B}{T} < u < \delta} \frac{3(\varpi^2 + 2\varpi)\vartheta_1 G_T(u + 2\vartheta_1)}{TG(u - 2\vartheta_1)} \\ & \geq c_0^2 \left( 1 - \sup_{\frac{B}{T} < u < \delta} \left| \frac{G_T^*(u - 2\vartheta_1)}{TG^*(u - 2\vartheta_1)} - 1 \right| \right) \frac{G^*(u - 2\vartheta_1)}{G(u - 2\vartheta_1)} - \inf_{\frac{B}{T} < u < \delta} \frac{3(\varpi^2 + 2\varpi)\vartheta_1 G_T(u + 2\vartheta_1)}{TG(u - 2\vartheta_1)} \\ & \quad - 2\varpi \sup_{\frac{B}{T} < u < \delta} \frac{\left| \sum_{t=p_j+1}^T A_t(u + 2\vartheta_1) \right| + \left| \sum_{t=p_j+1}^T D_t(u + 2\vartheta_1) \right|}{TG(u + 2\vartheta_1)} \frac{G(u + 2\vartheta_1)}{G(u - 2\vartheta_1)}. \end{aligned}$$

Note that as  $\vartheta_1 = o(1)$ ,

$$\begin{aligned} \frac{G^*(u - 2\vartheta_1)}{G(u - 2\vartheta_1)} &= \mathbb{P}(\|f_{t-1} - w_1\| \leq \rho | s < f_{t-\tau} \leq s + u - 2\vartheta_1) \\ &\longrightarrow \mathbb{P}(\|f_{t-1} - w_1\| \leq \rho | f_{t-\tau} = s) > 0 \quad \text{as } u - 2\vartheta_1 \downarrow 0, \end{aligned}$$

which implies that the infimum below exists for sufficiently small  $\delta > 2\vartheta_1 > 0$ , namely,

$$c_1 := \inf_{0 < u < \delta} \frac{G^*(u - 2\vartheta_1)}{G(u - 2\vartheta_1)} > 0.$$

As  $\vartheta_1, \vartheta_2 \rightarrow 0$ , choose  $\eta_0 > 0$  such that  $2\gamma = c_0^2 c_1 - (c_0^2 + 8\varpi)\eta_0 - 6(\varpi^2 + 2\varpi)\vartheta_1\eta_0$ . By Lemma A.4, it follows that

$$\begin{aligned} &\mathbb{P} \left( \inf_{\frac{B}{T} < u < \delta} \frac{R_{T1}(u)}{TG(u - 2\vartheta_1)} > 2\gamma \right) \\ &\geq \mathbb{P} \left( \sup_{\frac{B}{T} < u < \delta} \left| \frac{G_T^*(u - 2\vartheta_1)}{TG^*(u - 2\vartheta_1)} - 1 \right| < \eta_0, \sup_{\frac{B}{T} < u < \delta} \frac{\left| \sum_{t=p_j+1}^T A_t(u + 2\vartheta_1) \right|}{TG(u + 2\vartheta_1)} < \eta_0, \right. \\ &\quad \left. \sup_{\frac{B}{T} < u < \delta} \left| \frac{G_T(u - 2\vartheta_1)}{TG(u - 2\vartheta_1)} - 1 \right| < \eta_0, \sup_{\frac{B}{T} < u < \delta} \frac{\left| \sum_{t=p_j+1}^T D_t(u + 2\vartheta_1) \right|}{TG(u + 2\vartheta_1)} < \eta_0 \right) \\ &\geq 1 - \varepsilon. \end{aligned} \tag{30}$$

Note that

$$\begin{aligned} &[(\hat{\alpha}_t^{(20)} - \hat{\alpha}_t^{(2)}) - (\hat{\alpha}_t^{(10)} - \hat{\alpha}_t^{(1)})] \mathbf{1}\{s < \hat{f}_{t-\tau} \leq s + u\} \\ &= \{[(\phi^{(2)} - \phi^{(20)})^\top \hat{f}_{t-1}]^2 - [(\phi^{(1)} - \phi^{(10)})^\top \hat{f}_{t-1}]^2 + (\phi^{(2)} - \phi^{(20)})^\top \hat{f}_{t-1} (\phi^{(1)} - \phi^{(10)})^\top \hat{f}_{t-1} \\ &\quad + 2\hat{\xi}_t(\phi^{(10)} - \phi^{(1)} + \phi^{(2)} - \phi^{(20)})^\top \hat{f}_{t-1}\} \end{aligned}$$

For  $R_{T2}(\phi_0, u)$ , by the above identity and Lemma A.4, we have

$$\begin{aligned} |R_{T2}(\phi_0, u)| &\leq C\delta \left( (\mathbb{E}\|f_{t-1}\|^2 + \vartheta_2)G_T(u + 2\vartheta_1) + \left| \sum_{t=p_j+1}^T H_t(u + 2\vartheta_1) \right| + \left| \sum_{t=p_j+1}^T A_t(u + 2\vartheta_1) \right| \right. \\ &\quad \left. + \left| \sum_{t=p_j+1}^T D_t(u + 2\vartheta_1) \right| \right). \end{aligned}$$



It implied that

$$\begin{aligned}
& \sup_{\frac{B}{T} < u < \delta, \theta \in V_\delta} \frac{|R_{T2}(\phi_0, u)|}{TG(u - 2\vartheta_1)} \\
& \leq C\delta \left( (\mathbb{E}\|f_{t-1}\|^2 + \vartheta_2) \left[ \sup_{\frac{B}{T} < u < \delta} \left| \frac{G_T(u + 2\vartheta_1)}{TG(u - 2\vartheta_1)} - 1 \right| + 1 \right] + \right. \\
& \quad \left. + \sup_{\frac{B}{T} < u < \delta} \frac{\left| \sum_{t=p_j+1}^T H_t(u + 2\vartheta_1) \right| + \left| \sum_{t=p_j+1}^T A_t(u + 2\vartheta_1) \right| + \left| \sum_{t=p_j+1}^T D_t(u + 2\vartheta_1) \right|}{TG(u - 2\vartheta_1)} \right) \\
& = O_{\mathbb{P}}(\delta). \tag{31}
\end{aligned}$$

Then by (30) and (31), for sufficiently small  $\delta > 0$ , we have

$$\begin{aligned}
& \mathbb{P} \left( \inf_{\frac{B}{T} < u < \delta} \frac{R_T(u)}{TG(u - 2\vartheta_1)} > \gamma \right) \geq \mathbb{P} \left( \inf_{\frac{B}{T} < u < \delta} \frac{R_{T1}(u)}{TG(u - 2\vartheta_1)} - \sup_{\frac{B}{T} < u < \delta, \theta \in V_\delta} \frac{|R_{T2}(\phi_0, u)|}{TG(u - 2\vartheta_1)} > \gamma \right) \\
& \geq 1 - \mathbb{P} \left( \inf_{\frac{B}{T} < u < \delta} \frac{R_{T1}(u)}{TG(u - 2\vartheta_1)} \leq 2\gamma \right) - \mathbb{P} \left( \sup_{\frac{B}{T} < u < \delta, \theta \in V_\delta} \frac{|R_{T2}(\phi_0, u)|}{TG(u - 2\vartheta_1)} \geq \gamma \right) \\
& \geq 1 - 2\epsilon.
\end{aligned}$$

Thus, we finish the proof of (9).

Next, we move to the proof of (10) and (11).

Let  $\hat{\ell}_T(\phi_0, s_0) = \hat{L}_T(\phi_0, s_0)/T$  and  $\ell_T(\phi_0, s_0) = L_T(\phi_0, s_0)/T$ . By the Taylor expansion of  $\partial \hat{\ell}_T(\phi_0, s_0)/\partial \phi_0$ , we have

$$0 = \frac{\partial \hat{\ell}_T(\hat{\phi}(s_0), s_0)}{\partial \phi_0} = \frac{\partial \hat{\ell}_T(\phi, s_0)}{\partial \phi_0} + \frac{\partial^2 \hat{\ell}_T(\tilde{\phi}, s_0)}{\partial \phi_0 \partial \phi'_0} (\hat{\phi}(s_0) - \phi), \tag{32}$$

where  $\tilde{\phi}$  lies in between  $\hat{\phi}(s_0)$  and  $\phi$ , i.e.,  $\|\tilde{\phi} - \phi\| \leq \|\hat{\phi}(s_0) - \phi\|$ .

Let  $\tilde{\Sigma} = \text{diag}(\Sigma_1, \Sigma_2)$ . Since  $\mathbb{E}(\hat{f}_t^2) < \infty$  and  $\vartheta_1 = o(1)$ ,  $\vartheta_2 = o(1)$ , by the law of large numbers, it follows that

$$\frac{\partial^2 \hat{\ell}_T(\theta)}{\partial \phi_0 \partial \phi'_0} \rightarrow 2\tilde{\Sigma}, \quad \text{a.s. as } T \rightarrow \infty.$$

Then, by (32) and Lemma A.5, we have

$$\sup_{\|s - s_0\| \leq B(1/T + \vartheta_1)} \left\| [\hat{\phi}(s_0) - \phi] + (2\tilde{\Sigma})^{-1} \frac{\partial \ell_T(\phi, s)}{\partial \phi_0} \right\| = O_{\mathbb{P}} \left( \frac{1}{\sqrt{T}} + \vartheta_1 + \vartheta_2 \right).$$

This implies that (10).

Furthermore, if  $\lambda/\sigma \gg T^{1/2} + \sqrt{d_{\max}}$ , then  $\vartheta_1 + \vartheta_2 = o(T^{-1/2})$ . It follows that

$$\begin{aligned} \sup_{\|s-s_0\| \leq B(1/T+\vartheta_1)} \sqrt{T} \|\hat{\phi}(s_0) - \hat{\phi}(s)\| &\leq \sup_{\|s-s_0\| \leq B(1/T+\vartheta_1)} \left\| \sqrt{T} [\hat{\phi}(s_0) - \phi] + (2\tilde{\Sigma})^{-1} \sqrt{T} \frac{\partial \ell_T(\phi, s)}{\partial \phi_0} \right\| \\ &\quad + \left\| \sqrt{T} [\hat{\phi}(s) - \phi] + (2\tilde{\Sigma})^{-1} \sqrt{T} \frac{\partial \ell_T(\phi, s)}{\partial \phi_0} \right\| = o_{\mathbb{P}}(1). \end{aligned}$$

Note that

$$\sqrt{T} \frac{\partial \ell_T(\phi, s)}{\partial \phi_0} = \frac{2}{\sqrt{T}} \sum_{t=p_j+1}^T \frac{\partial \varepsilon_t(\phi, s)}{\partial \phi_0} \varepsilon_t(\phi, s),$$

and  $\{\varepsilon_t(\phi, s) \partial \varepsilon_t(\phi, s) / \partial \phi_0\}$  is a martingale difference sequence with respect to  $\{\mathcal{F}_t\}$ . By the martingale central limit theorem, it follows that

$$\sqrt{T} \frac{\partial \ell_T(\phi, s)}{\partial \phi_0} \xrightarrow{d} \mathcal{N}(0, 4\bar{\Sigma}),$$

where

$$\bar{\Sigma} = \text{diag}((\varsigma^{(1)})^2 \Sigma_1, (\varsigma^{(2)})^2 \Sigma_2).$$

Thus,

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} \mathcal{N}(0, \Sigma).$$

The proof of (11) is complete. □

**Proof of Theorem 3.2.** The proof of Theorem 3.2 closely parallels that of Theorem 3.1, with the key distinction that the threshold variable is observable. Consequently, we can eliminate the term  $\vartheta_1$  that appears in the original proof of Theorem 3.1, thereby obtaining improved convergence rates. The details are omitted due to their similarity. □

**Lemma A.4.** If Assumptions 3, 4, 5, 6 hold, then, for any  $\epsilon > 0$ ,  $\eta > 0$  and  $\delta \in (0, 1)$ , there exists a positive constant  $B$  such that for  $T$  large enough, we have

- (i).  $\mathbb{P} \left( \sup_{B/T < u < \delta} \left| \frac{G_T(u)}{TG(u)} - 1 \right| < \eta \right) > 1 - \epsilon,$
- (ii).  $\mathbb{P} \left( \sup_{B/T < u < \delta} \left| \frac{G_T^*(u)}{TG^*(u)} - 1 \right| < \eta \right) > 1 - \epsilon,$

$$(iii). \quad \mathbb{P} \left( \sup_{B/T < u < \delta} \frac{|\sum_{t=1}^T A_t(u)|}{TG(u)} < \eta \right) > 1 - \epsilon,$$

$$(iv). \quad \mathbb{P} \left( \sup_{B/T < u < \delta} \frac{|\sum_{t=1}^T D_t(u)|}{TG(u)} < \eta \right) > 1 - \epsilon,$$

$$(v). \quad \mathbb{P} \left( \sup_{B/T < u < \delta} \frac{|\sum_{t=1}^T H_t(u)|}{TG(u)} < \eta \right) > 1 - \epsilon,$$

where

$$G(u) = \mathbb{P}(s < z_t \leq s + u),$$

$$G_T(u) = \sum_{t=1}^T I(s < z_t \leq s + u),$$

$$G_T^*(u) = \sum_{t=1}^T I(s < z_t \leq s + u, \|f_{t-1} - w_1\| \leq \rho),$$

$$G^*(u) = \mathbb{P}(s < z_t \leq s + u, \|f_{t-1} - w_1\| \leq \rho),$$

$$A_t(u) = \xi_t I(s < z_t \leq s + u),$$

$$D_t(u) = f_{t-1} \xi_t I(s < z_t \leq s + u),$$

$$H_t(u) = (\|f_{t-1}\|^2 - \mathbb{E}\|f_{t-1}\|^2) I(s < z_t \leq s + u).$$

**Lemma A.5.** If the conditions in Theorem 3.1 hold, then, for any  $0 < B < \infty$ ,

$$\sup_{\|s-s_0\| \leq B(1/T+\vartheta_1)} \left\| \frac{\partial \hat{\ell}_T(\phi, s_0)}{\partial \phi_0} - \frac{\partial \hat{\ell}_T(\theta)}{\partial \phi_0} \right\| = O_{\mathbb{P}}(T^{-1} + \vartheta_1), \quad (33)$$

$$\sup_{\|s-s_0\| \leq B(1/T+\vartheta_1)} \left\| \frac{\partial^2 \hat{\ell}_T(\phi_0, s_0)}{\partial \phi_0 \partial \phi'_0} - \frac{\partial^2 \hat{\ell}_T(\theta)}{\partial \phi_0 \partial \phi'_0} \right\| = O_{\mathbb{P}}(T^{-1} + \vartheta_1), \quad (34)$$

where

$$\vartheta_1 = c \left( \frac{\sigma \sqrt{d_{\max}}}{\lambda \sqrt{T}} + \frac{\sigma}{\lambda} \right).$$

**Lemma A.6.** If the conditions in Theorem 3.2 hold, then, for any  $0 < B < \infty$ ,

$$\sup_{\|s-s_0\| \leq B/T} \left\| \frac{\partial \hat{\ell}_T(\phi, s_0)}{\partial \phi_0} - \frac{\partial \hat{\ell}_T(\theta)}{\partial \phi_0} \right\| = O_{\mathbb{P}}(T^{-1}), \quad (35)$$

$$\sup_{\|s-s_0\| \leq B/T} \left\| \frac{\partial^2 \hat{\ell}_T(\phi_0, s_0)}{\partial \phi_0 \partial \phi'_0} - \frac{\partial^2 \hat{\ell}_T(\theta)}{\partial \phi_0 \partial \phi'_0} \right\| = O_{\mathbb{P}}(T^{-1}). \quad (36)$$

Lemma A.4 is obtained directly from Zhang et al. (2024), while Lemmas A.5 and A.6 can be derived using similar arguments as in Zhang et al. (2024).

## B Additional Real Data Application Results

Table 4 shows the estimated models using the full data set. Note that the residual variance and AIC values are all significantly higher than that in Table 2. This is due to the fact that the full data include the period of financial crises, with large deviations from the majority of the data.

| Model                      | Series  | Regime                   | Coefficients   | Threshold   | $\sigma^2$ | AIC    |
|----------------------------|---------|--------------------------|--|-------------|------------|--------|
| <b>TFM-cp<br/>ARMA</b>     | factor1 | –                        | ar1: 0.70 (0.07)   | –           | 7.807      | 517.42 |
|                            | factor2 | –                        | ma1: 0.98 (0.09)<br>ma2: 0.40 (0.09)   | –           | 5.993      | 490.99 |
| <b>VFM<br/>ARMA</b>        | factor1 | –                        | ma1: 1.00 (0.10)<br>ma2: 0.40 (0.10)   | –           | 17.47      | 603.36 |
|                            | factor2 | –                        | ar1: 0.62 (0.09)<br>ma1: 0.42 (0.11)   | –           | 4.901      | 469.99 |
| <b>TFM-<br/>cp<br/>TAR</b> | factor1 | <b>Regime 1</b> (35 obs) | ar1: 1.21 (0.13)<br>ar2: -0.35 (0.12)  | 0.0103      | 5.7616     | 179.76 |
|                            |         | <b>Regime 2</b> (66 obs) | ar1: 0.44 (0.09)   | ( $d = 4$ ) |            |        |
|                            | factor2 | <b>Regime 1</b> (49 obs) | ar1: 1.03 (0.14)<br>ar2: -0.76 (0.18)<br>ar3: 0.45 (0.18)<br>ar4: -0.31 (0.13) | 0.0118      | 5.1425     | 169.24 |
|                            |         | <b>Regime 2</b> (52 obs) | ar1: 0.78 (0.09)   | ( $d = 4$ ) |            |        |
| <b>VFM<br/>TAR</b>         | factor1 | <b>Regime 1</b> (50 obs) | ar1: 1.08 (0.14)<br>ar2: -0.94 (0.21)<br>ar3: 0.54 (0.20)<br>ar4: -0.36 (0.14) | 0.0119      | 15.6982    | 261.11 |
|                            |         | <b>Regime 2</b> (51 obs) | ar1: 0.73 (0.09)   | ( $d = 4$ ) |            |        |
|                            | factor2 | <b>Regime 1</b> (42 obs) | ar1: 1.13 (0.13)<br>ar2: -0.49 (0.13)  | 0.0112      | 4.2950     | 149.37 |
|                            |         | <b>Regime 2</b> (61 obs) | ar1: 0.65 (0.09)   | ( $d = 2$ ) |            |        |

Table 4: Estimated parameters of ARMA and TAR models fitted to the factor process  $\hat{f}_{1t}$ , based on the entire sample (indices 1–10).