Statistica Sinica **34** (2024), 817-836
doi:https://doi.org/10.5705/ss.202022.0096

FUNCTIONAL THRESHOLD AUTOREGRESSIVE MODEL

Yuanbo Li[1], Kun Chen[*2], Xunze Zheng[3] and Chun Yip Yau[3]

[1] *University of International Business and Economics*
[2] *Southwestern University of Finance and Economics*
and [3] *The Chinese University of Hong Kong*

Abstract: We propose a functional threshold autoregressive model for flexible functional time series modeling. In particular, the behavior of a function at a given time point can be described by different autoregressive mechanisms, depending on the values of a threshold variable at a past time point. Sufficient conditions for the strict stationarity and ergodicity of the functional threshold autoregressive process are investigated. We develop a novel criterion-based method simultaneously conducting dimension reduction and estimating the thresholds, autoregressive orders, and model parameters. We also establish the consistency and asymptotic distributions of the estimators of both thresholds and the underlying autoregressive models. Simulation studies and an application to U.S. Treasury zero-coupon yield rates are provided to illustrate the effectiveness and usefulness of the proposed methodology.

Key words and phrases: Functional time series, minimum description length principle, multiple thresholds.

1. Introduction

High-frequency data are becoming increasingly prevalent. A popular method of studying such data is to convert the raw data into a sequence of curves $Y_k = \{Y_k(t),\ t \in K\}$, for some measurable set $K \subset \mathbb{R}^n$, and then to use the functional data analysis (FDA) approach. A brief introduction to FDA is available in Ramsay and Silvermann (2005). Numerous studies have contributed to theoretical and practical developments in functional time series, in which some auto-dependence exists among the observed curves Y_k. Hörmann and Kokoszka (2010, 2012) developed covariance estimators to study the auto-dependence in functional time series. Similarly to classic real-valued time series analysis, the linear model most commonly used to describe this auto-dependency is the functional autoregressive (FAR) process. Bosq (2000) investigated the causality and stationarity of FAR(1) processes and constructed one-step-ahead predictors. Aue, Norinho, and Hörmann (2015) proposed a prediction algorithm for FAR(p) models that is accurate and easily implementable. Moreover, Liu, Xiao and Chen (2016) proposed the convolutional autoregressive model, which is a special type of

*Corresponding author.

FAR(p) model, and Kokoszka and Reimherr (2013) developed a multiple testing procedure for selecting the order of FAR(p) models. In addition, Hörmann, Horváth and Reeder (2013), Aue, Horváth and Pellatt (2017), and Cerovecki et al. (2019) have developed functional ARCH and GARCH models that describe the nonconstant volatility function.

In scientific studies, a regime-switching structure is often suitable for modeling phenomena. For example, in macroeconomics, the behaviour of quantities such as GDP, exchange rate, inflation, interest rate, equity returns, and volatility varies depending on the regime of the underlying financial and economic mechanism, which may depend on regular business cycle movements, monetary policy, or the health of the financial market; see Lange and Rahbek (2009). In particular, it is well known that the yield curve of bonds has three main shapes, namely, normal, inverted, and flat, and that these shapes are related to periods of the economic cycle. In classical time series analysis, the threshold autoregressive (TAR) model, originally proposed by Tong (1978), is often used to model the feature of regime switching. As a simple and intuitive approximation of a complicated dynamic function, the TAR model captures many nonlinear phenomena, such as asymmetric limit cycles and time irreversibility, that cannot be explained by linear time series models. Owing to these advantages, the TAR model is widely used in areas such as biological sciences, econometrics, and environmental sciences (see Tong (1990), Hansen (2011), and Chen, So and Liu (2011)). However, despite their popularity and importance in read- and vector-valued time series, no studies have examined threshold-type models in the context of functional time series. Hence, for example, existing methods cannot describe the mechanism through which a functional observation exhibits different autoregressive structures depending on whether the maximum of a previous observation is above a certain threshold.

In this paper, we propose two models, namely, the *functional threshold autoregressive* (fTAR) model and the *functional threshold autoregressive with exogenous variables* (fTARX) model for flexible modeling of functional time series. These models allow a functional time series to follow different autoregressive models based on the range to which a threshold variable belongs. The threshold variable can be a real-valued functional, such as the maximum or average, of the observed function, or an exogenous scalar variable, at a past time point. We establish the conditions necessary for the stationarity and ergodicity of the fTAR/fTARX process. Moreover, the proposed model requires new statistical estimation and inference theory. In particular, in real-valued time series, threshold-type models are commonly estimated using achieved by a likelihood-based approach. However, it is well known that the likelihood is undefined in functional data. Therefore, we develop a quasi-likelihood estimation method, based on a dimension reduction step, that approximates the fTAR or fTARX model using a vector threshold autoregressive model. We also provide a model selection procedure that simultaneously reduces the dimension and estimates the thresholds, autoregressive orders, and model parameters.

The rest of the paper is organized as follows. Section 2 introduces the fTAR and fTARX models. Section 3 establishes sufficient conditions for the strict stationarity and ergodicity of the models. Section 4 proposes estimation and model selection procedures for statistical inference on the models. Section 5 establishes the consistency and weak convergence of the estimators of both the thresholds and the underlying model parameters in each regime. In Section 6, we apply the proposed method to data on U.S. Treasury zero-coupon yield rates.

2. The fTAR Model

First we introduce some notation. Denote $H = L^2([0,1])$ as the Hilbert space of square integrable functions on $[0,1]$, equipped with the inner product $\langle x, y\rangle = \int_0^1 x(t)y(t)\,dt$ and the norm $\|x\| = (\int_0^1 x(t)^2\,dt)^{1/2}$. The interval $[0,1]$ is chosen for convenience, and does not restrict the generality of our results. If x and y are both functions of H, then xy denotes their point-wise or component-wise product. Furthermore, denote $\mathcal{L}(H)$ as the Banach space of bounded linear operators on H, equipped with the operator norm $\|A\|_{\mathcal{L}} \equiv \sup_{\|x\|\le 1} \|A(x)\|$. This norm is sub-multiplicative, that is, $\|AB\|_{\mathcal{L}} \le \|A\|_{\mathcal{L}}\|B\|_{\mathcal{L}}$. We use the standard convention for combining operators, that is, $AB \equiv A \circ B$ and $A^2 \equiv A \circ A$, for A, $B \in \mathcal{L}(H)$. Moreover, let $\mathcal{K}(H)$ be the class of integral or kernel operators in H; that is, if $A \in \mathcal{K}(H)$, then there is a kernel $a : [0,1] \times [0,1] \to \mathbb{R}$ such that $A(x)(t) = \int_0^1 a(t,s)x(s)\,ds$. The operator $A \in \mathcal{K}(H)$ is Hilbert–Schmidt if and only if $\iint_0^1 a^2(t,s)\,dtds < \infty$.

Let $\{Y_k\}_{k=1,\ldots,n}$ be a sequence of functional time series with a sample size n. Assume that Y_k is a random function in H defined on a common probability space $(\Omega, \mathcal{B}, P)$. That is, for each $\omega \in \Omega$, $Y_k(\cdot,\omega) \in H$. For notational simplicity, we suppress the dependence of ω in Y_k. Finally, denote $Y \in L_H^\delta = L_H^\delta(\Omega, \mathcal{B}, P)$ if $\mathrm{E}[\|Y\|^\delta] < \infty$, for some $\delta > 0$.

As an extension of the functional AR and classic TAR models, the r-regime functional TAR model is specified as

$$Y_k = \sum_{i=1}^{r} \left[a_i + \Psi_{i,1}(Y_{k-1}) + \Psi_{i,2}(Y_{k-2}) + \cdots + \Psi_{i,p_{Y,i}}(Y_{k-p_{Y,i}}) + \sigma_i \epsilon_k\right] I(z_{k-d} \in (\theta_{i-1}, \theta_i]), \tag{2.1}$$

where $\Psi_{i,j} \in \mathcal{L}(H) \cap \mathcal{K}(H)$, with $\Psi_{i,j}(x)(t) = \int_0^1 \psi_{i,j}(t,s)x(s)\,ds$, for a kernel function $\psi_{i,j}(s,t)$ satisfying $\iint_0^1 \psi_{i,j}^2(s,t)\,dsdt < \infty$, $\{\epsilon_k\}$ are independent and identically distributed (i.i.d.) innovations in L_H^4, and σ_i is a positive constant. Assume that ϵ_k is independent of past information $\{Y_{k-j} : j \ge 1\}$, has an almost everywhere continuous and positive density function, and satisfies

$\mathrm{E}(\epsilon_k(t)) = 0$ and $\mathrm{E}(\epsilon_k^2(t)) = 1$, for all $t \in [0,1]$. The *thresholds* are denoted as $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{r-1})$, where $-\infty = \theta_0 < \theta_1 < \cdots < \theta_{r-1} < \theta_r = \infty$. The thresholds divide the range of the *threshold variable* z_{k-d} into r regimes, where d is the *delay parameter*. For simplicity, assume that z_{k-d} is either a real-valued functional of a lagged observation, say $z_{k-d} = g(Y_{k-d})$, or a scalar exogenous variable. This setting includes many classical threshold models in univariate time series, including the self-excited threshold autoregressive model, the open-loop threshold models (Tong (1990)), where the threshold variable is an exogenous scalar variable, and the models in Wu and Chen (2007) and Lee and Huang (2002), where z_{k-d} is the square or weighted average, respectively, of several exogenous scalar and lagged observations. Moreover, the current setting allows us to model how the behavior of a function Y_t changes based on that of the lagged function Y_{t-d}, measured in terms of its average $z_{k-d} = \int_0^1 Y_{k-d}(t)\, dt$, supremum $z_{k-d} = \sup_t(Y_{k-d}(t))$, infimum $z_{k-d} = \inf_t(Y_{k-d}(t))$, or other functionals of Y_{k-d}. The autoregressive model order is denoted as $\boldsymbol{p}_Y = (p_{Y,1}, \ldots, p_{Y,r})$, where $p_{Y,i} \in \mathbb{Z}^+$ is the model order in the ith regime.

Remark 1. Because our primary goal is to model the regime-switching structure of functional time series, the variance of the noise σ_j^2 in each regime is assumed to be a constant, for simplicity. For the general case in which σ_j^2 is a function, we first apply the same estimation method in Section 4, and then use the residuals to estimate σ_j^2 using the method of Aue, Horváth and Pellatt (2017).

Next, we introduce a stochastic recursive equation (SRE) representation of the fTAR process (2.1), which we use to examine the stationarity and ergodicity of the process. Denote

$$Y_k^* = \begin{pmatrix} Y_k \\ Y_{k-1} \\ Y_{k-2} \\ \vdots \\ Y_{k-p+1} \end{pmatrix}, \ \Psi_i^* = \begin{pmatrix} \Psi_{i,1} & \Psi_{i,2} & \cdots & \Psi_{i,p-1} & \Psi_{i,p} \\ Id & 0 & \cdots & 0 & 0 \\ 0 & Id & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & Id & 0 \end{pmatrix},$$

$$\epsilon_k^* = \begin{pmatrix} \epsilon_k \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ a_i^* = \begin{pmatrix} a_i \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{2.2}$$

where $p = \max_i p_{Y,i}$, and $\Psi_{i,j} = 0$, for $j > p_{Y,i}$. The p-vector Y_k^* is a vector of functions taking values in the product space $H_p = H \times \cdots \times H = (L^2[0,1])^p$. The p-dimensional matrix Ψ_i^* is a matrix of operators, and the components Id and 0 denote the identity operator and the zero operator on H, respectively. For $x = (x_1, \ldots, x_p)^{\mathrm{T}}, y = (y_1, \ldots, y_p)^{\mathrm{T}} \in H_p$, we define the inner product $\langle x, y\rangle_p =$

$\sum_{l=1}^{p}\langle x_l, y_l\rangle$, and the norm $\|x\|_p = \sqrt{\langle x, x\rangle_p}$. Equipped with $\langle\cdot,\cdot\rangle_p$, the space H_p defines a separable Hilbert space. Here, $\mathcal{L}(H_p)$, $\|\cdot\|_{\mathcal{L}_p}$, and $L^{\delta}_{H_p}$ are defined analogously to $\mathcal{L}(H)$, $\|\cdot\|_{\mathcal{L}}$, and L^{δ}_{H}, respectively. Note that $Y_k^* \in L^2_{H_p}$ and $\epsilon_k^* \in L^4_{H_p}$, for $k = 1, \ldots, n$. In addition, $\Psi_i^* \in \mathcal{L}(H_p)$, for $i = 1, \ldots, r$.

For simplicity, assume that $\min_i p_{Y,i} \geq d$; otherwise, we may replace $p_{Y,i}$ in each regime with $p_{Y,i} \vee d$. Using (2.2), we can express the fTAR model (2.1) in a state-space form

$$Y_k^* = \sum_{i=1}^{r}[a_i^* + \Psi_i^*(Y_{k-1}^*) + \sigma_i\epsilon_k^*]I(z_{k-d} \in (\theta_{i-1}, \theta_i])\,, \tag{2.3}$$

which is called the SRE representation of Y_k.

To enjoy greater modeling flexibility by incorporating exogenous covariate effects, we extend the fTAR model to the fTARX model, as follows:

$$Y_k = \sum_{i=1}^{r}\left[a_i + \sum_{j=1}^{p_{Y,i}}\Psi_{i,j}(Y_{k-j}) + \sum_{m=1}^{p_X}\Phi_{i,m}(X_{k,m}) + \sigma_i\epsilon_k\right]I(z_{k-d} \in (\theta_{i-1}, \theta_i])\,, \tag{2.4}$$

where $X_{k,m}$ is an exogenous variable that may be scalar, vector, or functional. Correspondingly, the operator $\Phi_{i,m}$ may represent scalar or matrix multiplication. In particular, if $X_{k,m} \in L^2_H$, then we assume $\Phi_{i,m} \in \mathcal{L}(H)$.

3. Stationarity and Ergodicity

When studying the asymptotic properties of estimators of threshold-type models, the assumptions of strict stationarity and ergodicity of the process are often essential; see Li and Ling (2012) and Yau, Tang and Lee (2015). In this section, we develop some mild conditions on the functional process Y_k, exogenous variables $X_{k,m}$, and the coefficients of the fTAR/fTARX models. These conditions are easy to verify and sufficient to ensure stationarity and ergodicity.

The following theorem provides a general sufficient criterion for the strict stationarity and ergodicity of the fTAR process.

Theorem 1. *For any positive integer $u \leq r$,*

$$\max_{i_1,\ldots,i_u}\left\|\Psi_{i_u}^* \circ \Psi_{i_{u-1}}^* \circ \cdots \circ \Psi_{i_1}^*\right\|_{\mathcal{L}_p} < 1\,, \tag{3.1}$$

where Ψ_i^* is defined in (2.2) and the maximum is taken over $1 \leq i_1 < i_2 \cdots < i_u \leq r$. Then $\{Y_k\}$ in (2.1) is strictly stationary and ergodic.

To enhance the practicality of Theorem 1, in the following corollaries, we develop easily verifiable sufficient conditions for (3.1) to hold.

Corollary 1. *If $\max_i \sum_{j=1}^{p_{Y,i}} \|\Psi_{i,j}\|_{\mathcal{L}} < 1$, then (3.1) holds.*

Corollary 2. *Assume that the operator* $\Psi_{i,j} \in \mathcal{L}(H)$ *is Hilbert–Schmidt with norm* $\|\Psi_{i,j}\|_{\mathcal{S}} = \iint_0^1 \psi_{i,j}(t,s)^2\,dtds < \infty$. *If* $\max_i \sum_{j=1}^{p_{Y,i}} \|\Psi_{i,j}\|_{\mathcal{S}} < 1$, *then* $\max_i \sum_{j=1}^{p_{Y,i}} \|\Psi_{i,j}\|_{\mathcal{L}} < 1$, *and hence* (3.1) *holds.*

We also discuss the β-mixing property of the functional process $\{Y_k\}$ and the stationary and ergodicity of the fTARX models in Section A.1 of the Supplementary Material.

4. Estimation and Model Selection

In FDA, a major approach to statistical inference is to reduce the infinite-dimensional problem to a multivariate problem by approximating functionals using a finite number, say q, of basis functions or principle components. However, the choice of q is difficult, and often requires a pre-processing step (e.g., Aue, Norinho, and Hörmann (2015), Aue, Horváth and Pellatt (2017)). In this paper, we propose a novel criterion-based method that simultaneously reduces the dimension, selects the thresholds and model orders, and estimates model parameters.

4.1. Estimation procedure

In this section, we show how to estimate the model parameters for the fTAR/fTARX model, given a set of thresholds $\boldsymbol{\theta}$. We discuss estimating the thresholds in Section 4.2.

First, consider the fTAR model (2.1). The functional parameters of interest are the functional intercepts a_i and the operators $\Psi_{i,j}$, and the volatility functions σ_i are treated as nuisance parameters. First, we choose an orthonormal basis $\mathcal{U}_{q_{\text{total}}} = \{u_l\}_{l=1}^{q_{\text{total}}}$, where q_{total} can be finite or infinite, and project a_i and $\Psi_{i,j}$, for $i = 1, \ldots, r$ and $j = 1, \ldots, p_{Y,i}$, onto $\mathcal{U}_{q_{\text{total}}}$. See Remark 2 for the choice of basis. From Parseval's identity, any operator $\Psi \in \mathcal{L}(H)$ can be expressed as

$$\Psi(x) = \sum_{l',l=1}^{q_{\text{total}}} \langle x, u_l\rangle\langle\Psi(u_l), u_{l'}\rangle u_{l'}\,. \tag{4.1}$$

Thus, estimating a_i and $\Psi_{i,j}$ is equivalent to estimating the innerproducts $\langle a_i, u_{l'}\rangle$ and $\langle\Psi_{i,j}(u_l), u_{l'}\rangle$, respectively. Because q_{total} can be finite, but large or even infinite, in practice, we set a prespecified upper bound $q_{\max}$ for the number of basis functions in order to conduct a feasible estimation of the inner products. Theoretically, $q_{\max}$ can increase with n with an appropriate order to provide an accurate approximation to Ψ; see Assumption 3 of Section 5.1.

Specifically, in view of (4.1), applying $\langle\cdot, u_l\rangle$ to (2.1) for $l = 1, \ldots, q_{\max}$ yields the following $q_{\max}$-dimensional *threshold vector autoregressive* (TVAR) model:

$$\boldsymbol{Y}_{k,q_{\max}} = \sum_{i=1}^{r}\left[\boldsymbol{a}_i^{q_{\max}} + \sum_{j=1}^{p_{Y,i}} \boldsymbol{\Psi}_{i,j}^{q_{max}} \boldsymbol{Y}_{k-j,q_{\max}} + \boldsymbol{W}_{k,i}^{q_{\max}}\right] I(z_{k-d} \in (\theta_{i-1}, \theta_i]), \quad (4.2)$$

where $\boldsymbol{a}_i^{q_{\max}} = (a_1^{(i)}, \ldots, a_{q_{\max}}^{(i)})^{\mathrm{T}}$ with $a_l^{(i)} = \langle a_i, u_l\rangle$, $\boldsymbol{Y}_{k,q_{\max}} = (y_{k,1}, \ldots, y_{k,q_{\max}})^{\mathrm{T}}$ with $y_{k,l} = \langle Y_k, u_l\rangle$, $\boldsymbol{W}_{k,i}^{q_{\max}} = (w_{k,1}^{(i)}, \ldots, w_{k,q_{\max}}^{(i)})^{\mathrm{T}}$ with $w_{k,l}^{(i)} = \langle \sigma_i \epsilon_k, u_l\rangle$, and $\boldsymbol{\Psi}_{i,j}^{q_{\max}} = (a_{l,l'}^{(i,j)})_{1\le l,l'\le q_{\max}}$, with $a_{l,l'}^{(i,j)} = \langle \Psi_{i,j}(u_l), u_{l'}\rangle$. It can be shown that $\boldsymbol{W}_{k,i}^{q_{\max}}$ is an error term with mean $\mathbf{0}$ and covariance matrix $\Sigma_{W,i}^{q_{\max}}$, with off-diagonal entries equal to zero.

The TVAR model (4.2) captures most necessary information in the fTAR model (2.1) if the parameters a_i and $\Psi_{i,j}$ are well represented by the $q_{\max}$ basis functions. With this multivariate formulation, estimating the fTAR model is reduced to estimating the vectors $\boldsymbol{a}_i^{q_{\max}}$ and matrices $\boldsymbol{\Psi}_{i,j}^{q_{\max}}$.

Choosing $q_{\max}$ is a difficult task, because different $q_{\max}$ result in TVAR models with different dimensions, and the corresponding likelihood functions are not directly comparable. To tackle this problem, instead of choosing $q_{\max}$, we set a large $q_{\max}$ to capture most information of the functional data, and assume that the true number of basis functions that generate a_i and $\Psi_{i,j}$ is finite $q(\ll q_{\max})$. Under this assumption, all entries of $\boldsymbol{a}_i^{q_{\max}}$ and $\boldsymbol{\Psi}_{i,j}^{q_{\max}}$ are zero, except $\boldsymbol{a}_i^{q}$, that is, the first q entries of $\boldsymbol{a}_i^{q_{\max}}$, and $\boldsymbol{\Psi}_{i,j}^{q}$, that is, the $q \times q$ submatrix at the top-left corner of $\boldsymbol{\Psi}_{i,j}^{q_{max}}$. This allows a trade-off between a lack of fit and model complexity in the choice of q, and the $q_{\max}$-dimensional TVAR model in (4.2) provides a benchmark for comparing different q. Consequently, the number of basis functions q can be regarded as a model order, and can be chosen using an information criterion; see Section 4.2.

For a fixed q, the log-likelihood of $\{\boldsymbol{Y}_{k,q_{\max}}\}$ is

$$L_n(\boldsymbol{\Psi}^q, r, d, \boldsymbol{\theta}, q, \boldsymbol{p}_Y) = \sum_{i=1}^{r}\sum_{k=1}^{n} l(\boldsymbol{\Psi}_i^q; \boldsymbol{Y}_{k,q_{\max}}, \ldots, \boldsymbol{Y}_{k-p_{Y,i},q_{\max}}) I(z_{k-d} \in (\theta_{i-1}, \theta_i]), \quad (4.3)$$

where $\boldsymbol{\Psi}_i^q = (\Psi_{i,1}^q, \ldots, \Psi_{i,p_{Y,i}}^q)$,

$$l(\boldsymbol{\Psi}_i^q; \boldsymbol{Y}_{k,q_{\max}}, \ldots, \boldsymbol{Y}_{k-p_{Y,i},q_{\max}}) = -\frac{1}{2}\Bigg[q_{\max}\log 2\pi + \log|\Sigma_{W,i}^{q_{\max}}| + \tilde{\mathbf{Y}}_{k,q}^{\mathrm{T}} {\Sigma_{W,i}^{q}}^{-1} \tilde{\mathbf{Y}}_{k,q}. + \sum_{m=q+1}^{q_{\max}} \frac{y_{k,m}^2}{\sigma_{m,i}^2}\Bigg],$$

$\boldsymbol{Y}_{k,q}$ are the first q components of $\mathbf{Y}_{k,q_{\max}}$, $\tilde{\mathbf{Y}}_{k,q} = \mathbf{Y}_{k,q} - \boldsymbol{a}_i^q - \sum_{j=1}^{p_{Y,i}} \boldsymbol{\Psi}_{i,j}^q \mathbf{Y}_{k-j,q}$, $\Sigma_{W,i}^q$ is a $q\times q$ submatrix of $\Sigma_{W,i}^{q_{\max}}$, and $\sigma_{m,i}^2$ is the mth diagonal entry of $\Sigma_{W,i}^{q_{\max}}$. The estimators $\widehat{\boldsymbol{a}}_i^q$, $\widehat{\boldsymbol{\Psi}}_{i,j}^q$, and $\widehat{\Sigma}_{W,i}^q$ of the TVAR model can be obtained by maximizing (4.3). Then, the estimator of the operator $\Psi_{i,j}$ of the fTAR model (2.1) can be computed using (4.1) for $\widehat{\Psi}_{i,j}(x) = (\widehat{\boldsymbol{\Psi}}_{i,j}^q \mathbf{c})^{\mathrm{T}}\mathbf{u}$, where $\mathbf{c} = (\langle x, u_1\rangle, \ldots, \langle x, u_q\rangle)^{\mathrm{T}}$

and $\mathbf{u} = (u_1, \ldots, u_q)^{\mathrm{T}}$.

The likelihood function (4.3) corresponds to the TVAR model (4.2), which serves as an approximation to the fTAR model. Although the likelihood function of functional data is not well defined, in general, this approximation is justified by Delaigle and Hall (2010). Specifically, they show that, for a set of functional data $\{Y_k\}$ and $y \in H$, the logarithm of a small ball probability, $p(y|h) = pr(\|Y_k - y\| \le h)$, can be approximated by the logarithms of the densities of its *functional principle component scores*, that is,

$$\log p(y|h) = C_1 + \sum_{l=1}^{q_{\max}} \log f_l(y_{l,q_{\max}}) + o(q_{\max}) \,, \tag{4.4}$$

where C_1 is a constant independent of y, $y_{l,q_{\max}} = \langle y, \hat{\nu}_l \rangle$, $h > 0$ is a small constant, and $f_l(\cdot)$ is the probability density function of $\langle Y_1, \hat{\nu}_l \rangle$. The term $\sum_{l=1}^{q_{\max}} \log f_l(y_{l,q_{\max}})$ in (4.4) diverges as $q_{\max} \to \infty$, implying that the log-density $l(y|q_{\max}) = \sum_{l=1}^{q_{\max}} \log f_l(y_{l,q_{\max}})$ captures the main variation of $\log p(y|h)$. Therefore, the proposed estimation procedure can be regarded as an approximate likelihood-based method, and is expected to achieve high efficiency.

For the fTARX model, we adjust the estimation procedure to incorporate covariates. Specify an orthonormal basis $\mathcal{U}_{q_{X,m}} = \{u_{X_m,1}, \ldots, u_{X_m,q_{X,m}}\}$ for the mth exogenous variable. Define the $q_{X,m}$-dimensional vector $\boldsymbol{X}_{k,m} = (x_{k,m,1}, \ldots, x_{k,m,q_{X,m}})^{\mathrm{T}}$, where $x_{k,m,l} = \langle X_{k,m}, u_{X_m,l} \rangle$. In practice, we may choose $\{u_{X_m,l}\}$ as a standard basis or the empirical eigenfunctions of $X_{k,m}$ derived using a *functional principle component analysis* (FPCA). For each m, since the dimension $q_{X,m}$ is exactly the number of variables included in the TVAR model, choosing $q_{X,m}$ is a classical variable selection task, which can be done easily using the model selection criterion proposed in the next section. Extending (4.2) by incorporating $\boldsymbol{X}_{k,m}$, we have

$$\begin{aligned}\boldsymbol{Y}_{k,q_{\max}} = \sum_{i=1}^{r} &\left[\boldsymbol{a}_i^{q_{\max}} + \sum_{j=1}^{p_{Y,i}} \boldsymbol{\Psi}_{i,j}^{q_{\max}} \boldsymbol{Y}_{k-j,q_{\max}} + \sum_{m=1}^{p_X} \boldsymbol{\Phi}_{i,m}^{q_{\max},q_{X,m}} \boldsymbol{X}_{k,m} + \boldsymbol{W}_{k,i}^{q_{\max}} \right] \\ &\times I(z_{k-d} \in (\theta_{i-1}, \theta_i]) \,,\end{aligned} \tag{4.5}$$

where $\boldsymbol{\Phi}_{i,m}^{q_{\max},q_{X,m}}$ is a $q_{\max} \times q_{X,m}$-dimensional matrix, the (l, l')th element of which is $\langle u_l, \Phi_{i,m}(u_{X_m,l'}) \rangle$. Similarly to the estimation of the fTAR model, estimating the functional parameters a_i, $\Psi_{i,j}$, and $\Phi_{i,m}$ of the fTARX model (2.4) reduces to estimating the parameter matrices, $\boldsymbol{a}_i^{q_{\max}}$, $\boldsymbol{\Psi}_{i,j}^{q_{\max}}$, and $\boldsymbol{\Phi}_{i,m}^{q_{\max},q_{X,m}}$, respectively, for all i, j and m, of model (4.5). Here, we can use the same method as that used to estimate (4.2).

Remark 2. The aforementioned estimation procedure requires the specification of a basis. Typical choices include the Fourier, B-spline, wavelet, or Hermite polynomial bases. Alternatively, the basis functions can be obtained using a data-driven method, such as the FPCA (Aue, Norinho, and Hörmann (2015), Aue, Horváth and Pellatt (2017) and Liu, Xiao and Chen (2016)).

4.2. Threshold and model-order estimation

A threshold model with more regimes clearly provides a finer approximation to the data, and thus achieves a better fit, at the expense of greater model complexity. Therefore, in traditional threshold models, it is common to estimate the threshold using an information criterion, that consists of a likelihood function and a penalty for model complexity. The optimizer of the information criterion then provides an estimate for the number and locations of the thresholds and the model orders; see, for example, Li and Ling (2012). In this section, using the vectorization idea in Section 4.1, we derive a *functional minimum description length* (fMDL) criterion for threshold estimation for the fTAR model.

The fMDL proposed in this paper is motivated by the minimum description length (MDL). The MDL is the minimum length of computer code required to record the observed data $\boldsymbol{Y} = \{Y_1, \ldots, Y_k\}$. In general, given a parametric model, say $\mathcal{M}$, Lee (2000) defines the MDL as a sum of two parts, given by $\mathrm{MDL}(\mathcal{M}) = \mathrm{CL}(\mathcal{M}) + \log_2(e)\mathrm{CL}(\boldsymbol{Y} \mid \mathcal{M})$, where $\mathrm{CL}(\mathcal{M})$ and $\mathrm{CL}(\boldsymbol{Y} \mid \mathcal{M})$ are the lengths of code lengths required to record the model $\mathcal{M}$ and the observations given $\mathcal{M}$, respectively.

First, we derive $\mathrm{CL}(\mathcal{M})$ for the fTAR model. To record an fTAR model $\mathcal{M}$, we need to record the number of thresholds $r-1$, delay parameter d, values of the thresholds $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{r-1})^{\mathrm{T}}$, dimension q, autoregressive order $\boldsymbol{p}_Y = (p_{Y,1}, \ldots, p_{Y,r})^{\mathrm{T}}$, and values of the model parameters $\boldsymbol{\Psi}_i^q = (\boldsymbol{a}_i^q, \boldsymbol{\Psi}_{i,1}^q, \ldots, \boldsymbol{\Psi}_{i,p_{Y,i}}^q)$ in each of the r regimes. Because the model parameters can be estimated as in Section 5.1, given the aforementioned quantities, $\mathcal{M}$ can be represented by the vector $(r, d, \boldsymbol{\theta}, q, \boldsymbol{p}_Y)$. From Rissanen (1989, 2007), recording an integer m and a parameter estimator computed from a sample of size n requires code lengths of approximately $\log_2 m$ and $(\log_2 n)/2$ digits, respectively. To avoid the problem of $\log 0 = -\infty$, $\log_2 m$ is understood as $\max(\log_2 m, 0)$. Furthermore, the thresholds can be located by the ranks of the values of the threshold variable z_t, and thus can be encoded using integers. Hence, we use $\log_2(r)$ digits to record the number of regimes, $\log_2 d$ digits to record the delay parameter, $\sum_{i=1}^{r-1}(\log_2 n_i)$ digits to record the threshold estimators, $\sum_{i=1}^{r}\log_2(p_{Y,i}q+1)q$ digits to record the total number of parameters in a regime, and $\sum_{i=1}^{r}((p_{Y,i}q+1)q\log_2 n_i)/2$ digits to record the estimators of $\boldsymbol{\Psi}^q = (\boldsymbol{\Psi}_1^q, \ldots, \boldsymbol{\Psi}_r^q)$, where n_i is the number of observations in the ith regime. Therefore, the code length for encoding $\mathcal{M}$ is expressed as

$$\mathrm{CL}(\mathcal{M}) = \log_2(r) + \log_2 d + \sum_{i=1}^{r-1}\log_2 n_i + \sum_{i=1}^{r}\log_2(p_{Y,i}q+1)q$$

$$+\sum_{i=1}^{r}\frac{(p_{Y,i}q+1)q}{2}\log_2 n_i\,. \tag{4.6}$$

Next, we derive $\mathrm{CL}(\boldsymbol{Y}\mid\mathcal{M})$ for the fTAR model. From Rissanen (1989, 2007), this term can be approximated using the likelihood function specified by $\mathcal{M}$. Therefore, we can use the approximate likelihood in (4.3) to define

$$\mathrm{CL}(\boldsymbol{Y}\mid\mathcal{M}) = -L_n(\hat{\boldsymbol{\Psi}}^q, r, d, \boldsymbol{\theta}, q, \boldsymbol{p}_Y)\log_2 e\,, \tag{4.7}$$

where the estimator $\hat{\boldsymbol{\Psi}}^q$ is the maximizer of $L_n(\boldsymbol{\Psi}^q, r, d, \boldsymbol{\theta}, q, \boldsymbol{p}_Y)$, given a set of $\{r, d, \boldsymbol{\theta}, q, \boldsymbol{p}_Y\}$. Combining (4.6) and (4.7), the MDL for a fTAR model is defined as

$$\begin{aligned}
&\mathrm{fMDL}(r, d, \boldsymbol{\theta}, q, \boldsymbol{p}_Y) = \mathrm{CL}(\mathcal{M}) + \mathrm{CL}(\boldsymbol{Y}\mid\mathcal{M}) \\
&= \log_2(rd) + \sum_{i=1}^{r-1}\log_2 n_i + \sum_{i=1}^{r}\log_2(p_{Y,i}q+1)q + \sum_{i=1}^{r}\frac{(p_{Y,i}q+1)q}{2}\log_2 n_i \\
&\quad -L_n(\hat{\boldsymbol{\Psi}}^q, r, d, \boldsymbol{\theta}, q, \boldsymbol{p}_Y)\log_2 e\,.
\end{aligned} \tag{4.8}$$

The optimal fTAR model is obtained from $\{\hat{r}_n, \hat{d}_n, \hat{\boldsymbol{\theta}}_n, \hat{q}_n, \hat{\boldsymbol{p}}_Y\}$, which minimizes (4.8). The fMDL criterion for the fTARX model can be defined similarly; see Section C of the Supplementary Material. Note that the optimization of the fMDL in (4.8) is nonstandard, because the objective function fMDL is not differentiable with respective to the parameters $\{r, d, \boldsymbol{\theta}, q, \boldsymbol{q}_X, \boldsymbol{p}_X\}$. To tackle this, we develop a genetic algorithm to efficiently optimize the fMDL in Section D of the Supplementary Material.

Remark 3. As an alternative to the fMDL approach in this paper, the Bayesian approach is widely used for estimating threshold time series models; see Chen (1995), Chen and Lee (1995), Wu and Chen (2007), and Pan, Xia and Liu (2017). Compared with the Bayesian approach, the fMDL procedure avoids having to choose prior distributions and the corresponding hyperparameters. On the other hand, the Bayesian approach allows a simpler inference procedure, because it avoids having to estimate the asymptotic distribution of the estimators, and is worth considering for future research.

5. Asymptotic Theory

In this section, we develop asymptotic theory for inferences on the fTARX model, which includes the fTAR model as a special case. Let $(a_1^0, \ldots, a_{r^0}^0, \Psi_{1,1}^0, \ldots, \Psi_{r^0, p_{Y,r^0}^0}^0, \Phi_{1,1}^0, \ldots, \Phi_{r^0, p_X^0}, \theta_1^0, \ldots, \theta_{r^0-1}^0, d^0)$ be the true parameter values of model (2.4). For the delay, number of thresholds, thresholds, dimension, autoregressive orders, and covariate indicators, the estimators are denoted as $\hat{d}_n$, $\hat{r}_n$, $\hat{\boldsymbol{\theta}}_n = \{\hat{\theta}_1, \ldots, \hat{\theta}_{\hat{r}-1}\}$, $\hat{q}_n$, $\hat{q}_{X,m}$, $\hat{\boldsymbol{p}}_Y$, and $\hat{\boldsymbol{p}}_X$, respectively. In addition, for the vector autoregressive model in each regime, denote the true values as $\mathbf{\Psi}_i^0 = \{a_i^0, \Psi_{i,1}^0, \ldots, \Psi_{i,p_{Y,i}^0}^0\}$, $\mathbf{\Phi}_i^0 = \{\Phi_{i,1}^0, \ldots, \Phi_{i,p_X^0}^0\}$, $\mathbf{\Psi}^0 = \{\mathbf{\Psi}_1^0, \ldots, \mathbf{\Psi}_{r^0}^0\}$, and $\mathbf{\Phi}^0 = \{\mathbf{\Phi}_1^0, \ldots, \mathbf{\Phi}_{r^0}^0\}$. To simplify the notation, we suppress the superscript 0 for the true values in the following, unless specified otherwise. First, we impose the following assumptions.

5.1. Assumptions

We assume the conditions in Theorem 1 or Corollaries 1–2 hold, such that the fTARX series $\{Y_k\}$ is stationary and ergodic. In addition, the following assumption ensures that $\{Y_k\}$ is in L_H^4.

Assumption 1. *Denote* $\tilde{\Psi}_k^* = \sum_{i=1}^r \Psi_i^* I(z_{k-d} \in (\theta_{i-1}, \theta_i])$. *Assume that there exists some positive integer* $u > 0$ *such that* $\max_i \mathrm{E}(\|\tilde{\Psi}_{k+u-1}^* \circ \cdots \circ \tilde{\Psi}_k^*\|_{\mathcal{L}_p}^4 \mid z_{k-d-1} \in (\theta_{i-1}, \theta_i]) < 1$.

Next, Assumptions 2–6 are imposed in order to establish the consistency of the estimates.

Assumption 2. *Let* $\Omega \times \{1, \ldots, D_0\}$ *be the parameter space, where* $\Omega = \Omega_{\mathbf{\Psi}} \times \Omega_{\mathbf{\Phi}} \times \Omega_{\boldsymbol{\theta}}$ *is a compact subset of* $\mathbb{R}^{qr+\sum_{i=1}^r (p_{Y,i}q+\sum_{m=1}^{p_X} p_{X,i,m}q_{X,m})q} \times \mathcal{R}^{r-1}$, D_0 *is the maximum delay, and* $\mathcal{R}^{r-1} = \{(\theta_1, \ldots, \theta_{r-1}) : -\infty < \theta_1 < \cdots < \theta_{r-1} < \infty\}$. *Moreover, the true parameters* $(\mathbf{\Psi}, \mathbf{\Phi})$ *are in the interior of* $\Omega_{\mathbf{\Psi}} \times \Omega_{\mathbf{\Phi}}$.

Assumption 3. *There exists a smallest positive integer* q *such that, for any* $q^* > q$, $a_i^{(q^*)}$, $\langle \Psi_{i,j}(u_l), u_{q^*}\rangle$, *and* $\langle \Psi_{i,j}(u_{q^*}), u_m\rangle$ *are equal to zero, for all* $1 \le l, m \le q^*$ *and* $i = 1, \ldots, r$. *In addition, assume that* $q_{\max} = O(\sqrt{n/\log n})$.

Assumption 4. *The autoregressive function varies across regimes. That is, there exists some* $(Y_0, \ldots, Y_{p-1}, X_{p,1}, \ldots, X_{p,p_X})$ *such that*

$$a_i + \sum_{j=1}^{p_{Y,i}} \Psi_{i,j}(Y_{p-j}) + \sum_{m=1}^{p_X} \Phi_{i,m}(X_{p,m}) \neq a_{i+1} + \sum_{j=1}^{p_{Y,i+1}} \Psi_{i+1,j}(Y_{p-j}) + \sum_{m=1}^{p_X} \Phi_{i+1,m}(X_{p,m}),$$

for $i = 1, \ldots, r-1$.

Assumption 5. *For each* $i = 1, \ldots, r$, *the joint conditional distribution function of* $\{\boldsymbol{Y}_{k-1,q}, \ldots, \boldsymbol{Y}_{k-p_{Y,i},q}, \boldsymbol{X}_{k,1}, \ldots, \boldsymbol{X}_{k,p_X}\}$ *given* $z_{k-d} = \theta_i$ *is continuous.*

Assumption 6. *The threshold variable* z_k *has a marginal probability density* $\pi_z(\cdot)$ *that is continuous and positive at* $\theta_1, \ldots, \theta_{r-1}$. *Furthermore, the joint density of* $\{z_{k-d_1}, z_{k-d_2}\}$, *denoted as* $\pi_{z,|d_1-d_2|}(\cdot,\cdot)$ *for* $d_1, d_2 \in \{1, \ldots, D_0\}$, *is uniformly bounded and positive everywhere. In addition, for any* $X_k^* = (X_{k,1}, \ldots, X_{k,p_X})^{\mathrm{T}}$ *and* $\Phi^* = (\Phi_1, \ldots, \Phi_{p_X})^{\mathrm{T}}$ *such that* $\sum_{j=1}^{p_X} \|\Phi_j\|_{\mathcal{L}} = 1$, *there exists an* $\epsilon > 0$ *such that* $\mathrm{pr}(\|\Phi^{*\mathrm{T}} X_k^*\| > \epsilon | z_{t-d_1}, z_{t-d_2}) > 0$, *almost surely, with respect to the joint distribution of* (z_{t-d_1}, z_{t-d_2}).

Assumption 2 is a standard regularity condition for asymptotic properties of parameter estimators. Assumption 3 requires that we can express the coefficients in the fTAR model using a finite number of bases, which essentially restricts the number of unknown parameters in the associated TVAR model (4.2). A similar requirement is adopted in Aue, Horváth and Pellatt (2017). The rate of $q_{\max}$ is used in the proof for Theorem 2. Assumption 4 assumes that consecutive regimes behave such that the thresholds are identifiable. Assumption 5 is a mild regularity condition. Assumption 6 removes linear dependence and, hence, redundancy from the exogenous covariates.

Assumption 7. *For each* $i = 1, \ldots, r-1$, *there exists some* $\Delta > 0$ *such that the process*

$$\{Y_k I(z_{k-d} \in [\theta_i - \Delta, \theta_i + \Delta]), z_{k-d} I(z_{k-d} \in [\theta_i - \Delta, \theta_i + \Delta])\}_{k=1,2,\ldots}$$

is ρ*-mixing, with summable mixing coefficients* $\{\rho(m)\}_{m=1,2,\ldots}$.

Assumption 7 is required in order to prove the convergence rates and the asymptotic distributions of the estimators of the thresholds and the parameters. Note that if $\boldsymbol{W}_{k,i}^{q}$ follows a joint normal distribution, then the ρ-mixing property can be deduced from the α-mixing of Gaussian processes (Kolmogorov and Rozanov (1960)). Let $\mathcal{A}$ and $\mathcal{A}^*$ be the σ-algebras generated by the stochastic processes $\{w_t\}_{t \leq j}$ and $\{w_t\}_{t \geq j+k}$, respectively, for any integer j. For any ρ-mixing process $\{w_t\}$, there exists a sequence $\{\rho(m)\}_{m=1,2,\ldots}$ with $\lim_{m\to\infty} \rho(m) \to 0$ such that, for all square-integrable random variables g and h that are $\mathcal{A}$ and $\mathcal{A}^*$ measurable, respectively, we have $|\mathrm{corr}(g,h)| \leq \rho(m)$; see Doukhan (1994).

5.2. Main results

Theorem 2. *If Assumptions* 1–6 *hold, then*

$$\{\hat{\boldsymbol{\Psi}}_n^{\hat{q}_n}, \hat{\boldsymbol{\Phi}}_n^{\hat{q}_n}, \hat{r}_n, \hat{d}_n, \hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{p}}_Y, \hat{\boldsymbol{p}}_X, \hat{q}_n, \hat{q}_{X,m}\} \xrightarrow{a.s.} \{\boldsymbol{\Psi}, \boldsymbol{\Phi}, r, d, \boldsymbol{\theta}, \boldsymbol{p}_Y, \boldsymbol{p}_X, q, q_{X,m}\}.$$

From Theorem 2, the model orders r_n, d_n, $\boldsymbol{p}_Y$, $\boldsymbol{p}_X$, q_n, and $q_{X,m}$ can be estimated consistently. Hence, without loss of generality, we can assume that they are known, and thus suppress the superscripts q and $\hat{q}_n$ for notational simplicity. Next, we derive the convergence rate of $\hat{\boldsymbol{\theta}}_n$.

Theorem 3. *If Assumptions* 1–7 *hold and* $E\|\epsilon_k\|^{4+\delta} < \infty$, *for some* $\delta > 0$, *then* $\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|_2 = O_p(n^{-1})$.

Next, we discuss the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$. Let the difference between the log-likelihood under the parameters $\{\boldsymbol{\Psi}_i, \boldsymbol{\Phi}_i\}$ and $\{\boldsymbol{\Psi}_j, \boldsymbol{\Phi}_j\}$ be

$$\begin{aligned}
&\xi_k^{(i,j)}(\boldsymbol{Y}_k, \ldots, \boldsymbol{Y}_{k-p_{Y,i}}, \boldsymbol{X}_{k,1}, \ldots, \boldsymbol{X}_{k,p_X}) \\
&= l(\boldsymbol{\Psi}_i, \boldsymbol{\Phi}_i; \boldsymbol{Y}_k, \ldots, \boldsymbol{Y}_{k-p_{Y,i}}, \boldsymbol{X}_{k,1}, \ldots, \boldsymbol{X}_{k,p_X}) \\
&\quad - l(\boldsymbol{\Psi}_j, \boldsymbol{\Phi}_j; \boldsymbol{Y}_k, \ldots, \boldsymbol{Y}_{k-p_{Y,j}}, \boldsymbol{X}_{k,1}, \ldots, \boldsymbol{X}_{k,p_X}).
\end{aligned}$$

Let $F_{(i,j)}(\cdot|\theta)$ be the conditional distribution function of $\xi_{d+1}^{(i,j)}$, given $z_1 = \theta$. Then, we can define the $(r-1)$ independent one-dimensional two-sided compound Poisson process $\{\mathcal{P}_i(\kappa), \kappa \in \mathbb{R}\}$ as

$$\mathcal{P}_i(\kappa) = \mathcal{P}_{i,1}(\kappa)I(\kappa < 0) + \mathcal{P}_{i,2}(\kappa)I(\kappa > 0)\,, \tag{5.1}$$

for $i = 1, \ldots, r-1$, where $\mathcal{P}_{i,1}(\kappa) = \sum_{k=1}^{N_{i,1}(-\kappa)} \xi_k^{(i+1,i)}(\tilde{\boldsymbol{Y}}_k^{(1)}, \ldots, \tilde{\boldsymbol{Y}}_{k-p_{Y,i}}^{(1)}, \tilde{\boldsymbol{X}}_{k,1}^{(1)}, \ldots, \tilde{\boldsymbol{X}}_{k,p_X}^{(1)})$ and $\mathcal{P}_{i,2}(\kappa) = \sum_{k=1}^{N_{i,2}(\kappa)} \xi_k^{(i,i+1)}(\tilde{\boldsymbol{Y}}_k^{(2)}, \ldots, \tilde{\boldsymbol{Y}}_{k-p_{Y,i}}^{(2)}, \tilde{\boldsymbol{X}}_{k,1}^{(2)}, \ldots, \tilde{\boldsymbol{X}}_{k,p_X}^{(2)})$, where $(\tilde{\boldsymbol{Y}}_k^{(j)}, \ldots, \tilde{\boldsymbol{Y}}_{k-p_{Y,i}}^{(j)}, \tilde{\boldsymbol{X}}_{k,1}^{(j)}, \ldots, \tilde{\boldsymbol{X}}_{k,p_X}^{(j)})$, for $j = 1, 2$, are independent copies of $(\boldsymbol{Y}_k, \ldots, \boldsymbol{Y}_{k-p_{Y,i}}, \boldsymbol{X}_{k,1} \ldots, \boldsymbol{X}_{k,p_X})$. The processes $\{N_{i,1}(\kappa), \kappa \geq 0\}$ and $\{N_{i,2}(\kappa), \kappa \geq 0\}$ are two independent Poisson processes with $N_{i,1}(0) = N_{i,2}(0) = 0$ almost surely, and with the same jump rate $\pi_z(\theta_i)$. In addition, $\{\xi_k^{(i,i+1)} : k \geq 1\}$ are i.i.d. random variables with the distribution $F_{(i,i+1)}(\cdot|\theta_i)$, and $\{\xi_k^{(i+1,i)} : k \geq 1\}$ are i.i.d. random variables with the distribution $F_{(i+1,i)}(\cdot|\theta_i)$. Furthermore, $\{\xi_k^{(i,i+1)} : k \geq 1\}$ and $\{\xi_k^{(i+1,i)} : k \geq 1\}$ are mutually independent. Combining $\mathcal{P}_{i,1}$ and $\mathcal{P}_{i,2}$, we can define a double-sided compound Poisson process $\mathcal{P}_i(\kappa) = \mathcal{P}_{i,1}(\kappa) + \mathcal{P}_{i,2}(\kappa)$, and further define a spatial compound Poisson process

$$\mathcal{P}(\boldsymbol{\kappa}) = \sum_{i=1}^{r-1} \mathcal{P}_i(\kappa_i), \quad \boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_{r-1})^{\mathrm{T}} \in \mathbb{R}^{r-1}\,.$$

By Assumption 4, for $i = 1, \ldots, r-1$, $\mathcal{P}(\boldsymbol{\kappa}) \to \infty$ as $\|\boldsymbol{\kappa}\|_2 \to \infty$. Hence, there exists a unique random $(r-1)$-dimensional cube $[\mathbf{M}_-, \mathbf{M}_+) = [M_-^{(1)}, M_+^{(1)}) \times \cdots \times [M_-^{(r-1)}, M_+^{(r-1)})$ on which the process $\mathcal{P}(\boldsymbol{\kappa})$ attains its global minimum almost surely; that is, $[\mathbf{M}_-, \mathbf{M}_+) = \operatorname{argmin}_{\boldsymbol{\kappa} \in \mathbb{R}^{r-1}} \mathcal{P}(\boldsymbol{\kappa})$. This immediately implies that $[M_-^{(i)}, M_+^{(i)}) = \operatorname{argmin}_{\kappa_i \in \mathbb{R}} \mathcal{P}_i(\kappa_i)$. In addition, note that the processes $\{\mathcal{P}_i(\kappa_i) : i = 1, \ldots, r-1\}$, and thus $\{M_-^{(i)} : i = 1, \ldots, r-1\}$, are mutually independent. The following theorem gives the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$.

Theorem 4. *If Assumptions* 1–7 *hold, then* $n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ *weakly converges to* $\mathbf{M}_-$ *as* $n \to \infty$*, and its components are asymptotically independent.*

Finally, the following theorem gives the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\Psi}}_i - \boldsymbol{\Psi}_i)$ and $\sqrt{n}(\hat{\boldsymbol{\Phi}}_i - \boldsymbol{\Phi}_i)$. For matrices $A_1, \ldots, A_k$, let $\operatorname{vec}(A_1, \ldots, A_k) = (\operatorname{vec}(A_1)^{\mathrm{T}}, \ldots, \operatorname{vec}(A_k)^{\mathrm{T}})^{\mathrm{T}}$, where $\operatorname{vec}(A_i)$ is the vectorization of the matrix A_i, for example, $\operatorname{vec}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = (a, c, b, d)^{\mathrm{T}}$. In addition, denote $A \otimes B$ as the Kronecker product for matrices A and B.

Theorem 5. *For* $i = 1, \ldots, r$*, define* $\boldsymbol{\beta}_i = \boldsymbol{\beta}_i(\boldsymbol{\theta}) = \operatorname{vec}(\boldsymbol{a}_i, \boldsymbol{\Psi}_{i,1}, \ldots, \boldsymbol{\Psi}_{i,p_{Y,i}}, \boldsymbol{\Phi}_{i,1}, \ldots, \boldsymbol{\Phi}_{i,p_X})$*. Let the estimator be* $\hat{\boldsymbol{\beta}}_i \equiv \hat{\boldsymbol{\beta}}_i(\hat{\boldsymbol{\theta}}_n)$*. If Assumptions* 1–7 *hold, then*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) \xrightarrow{d.} N(\mathbf{0}, \Gamma_i^{-1} \otimes \Sigma_{W,i})\,,$$

where $\mathbf{0}$ *is a zero vector of dimension* $(p_{Y,i}q + \boldsymbol{q}_X^{\mathrm{T}}\boldsymbol{p}_X + 1)q$ *and*

$$\Gamma_i = \mathrm{E}[\mathrm{vec}(1, \boldsymbol{Y}_{k-1}, \ldots, \boldsymbol{Y}_{k-p_{Y,i}}, \boldsymbol{X}_{k,1}, \ldots, \boldsymbol{X}_{k,p_X}) \\ (\mathrm{vec}(1, \boldsymbol{Y}_{k-1}, \ldots, \boldsymbol{Y}_{k-p_{Y,i}}, \boldsymbol{X}_{k,1}, \ldots, \boldsymbol{X}_{k,p_X}))^{\mathrm{T}}] .$$

Hence, $\|\hat{\boldsymbol{\Psi}}_{i,j} - \boldsymbol{\Psi}_{i,j}\|_2 = O_p(n^{-1/2})$ *and* $\|\hat{\boldsymbol{\Phi}}_{i,m} - \boldsymbol{\Phi}_{i,m}\|_2 = O_p(n^{-1/2})$*, for all* $j = 1, \ldots, p_{Y,i}$*,* $m = 1, \ldots, p_X$*, and* $i = 1, \ldots, r$*. In addition,* $\sqrt{n}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)$ *and* $n(\hat{\theta}_i - \theta_i)$*, for* $i = 1, \ldots, r-1$*, are asymptotically independent. Finally, for the functional operators, we have* $\|\hat{\Psi}_{i,j} - \Psi_{i,j}\|_{\mathcal{L}} = O_p(n^{-1/2})$ *and* $\|\hat{\Phi}_{i,m} - \Phi_{i,m}\|_{\mathcal{L}} = O_p(n^{-1/2})$*, for all* $j = 1, \ldots, p_{Y,i}$*,* $m = 1, \ldots, p_X$*, and* $i = 1, \ldots, r$*.*

6. Real Application

In this section, we use the fTARX model to fit a yield curve, which is a plot of the yields (interest rates) of bonds with equal credit quality against maturity dates. In finance, the modeling of yield rates plays a critical role in the pricing and risk management of fixed-income products. A qualified model for the yield curve should consider the cross-sectional and serial dependence of interest rates for all maturities at any give time. Here, we model the entire yield curve as a functional time series, and study the dynamics of the yield curve over time. In particular, we consider the fTAR model, for two reasons. First, the classical TAR model is often used to model interest rates; see Tsay (1998). Second, and more importantly, it is well known that the yield curve has three main shapes: normal (upward-sloping curve), inverted (downward-sloping curve), and flat. The normal yield curve refers to periods of economic expansion, and the inverted yield curve corresponds to periods of economic recession. Thus, the economic cycle is related to the regime classification of the yield curve. Therefore, the proposed fTARX model is particularly suitable for analyzing the serial dependence and regime classifications of yield curves.

The data set consists of daily off-the-run zero-coupon treasury yield rates from September 4, 2007, to January 3, 2011, with values shown on a percentage scale with bond maturity from 1 to 30 years. The raw data are collected from the U.S. Federal Reserve Data Release (downloaded from `https://www.federalreserve.gov/data/nominal-yield-curve.htm`). We transform 30 daily observations with different maturities into a functional observation by using a B-spline basis with order five and dimension nine. In addition, we extract the yield rate of the U.S. Generic Government 10-year treasury note on the previous trading day, as a proxy for the risk-free rate for long-term U.S. investment. Our goal is to model the excess yield curves, denoted as $\{Y_k(\cdot)\}_{k=1,\ldots}$. The curves $\{Y_k(\cdot)\}_{k=1,\ldots}$ are depicted in Figure 1, where the mean curve and ± 0.5 times various eigenfunctions are plotted to demonstrate the mean effect and the effect of major principal components (PCs), respectively. The first PC represents the

Table 1. Estimation results of fTARX and functional ARX models for the treasury yield rates from September 4, 2007, to January 3, 2011.

Model	Regime	$\hat{q}_n$	$\hat{p}_{Y,i}$	Exogenous covariates
fTARX	$(-\infty, 5.174]$	3	2	X_{csfb}, X_{gsli}
	$(5.174, 10.026]$	3	1	X_{csfb}, X_{vix}
	$(10.026, \infty)$	3	1	X_{csfb}
functional ARX	-	3	3	$X_{csfb}, X_{ip}, X_{gsli}$

overall level, the second PC represents the slope, and the third PC represents the curvature of the yield curves.

From Ang and Piazzesi (2003) and Diebold, Rudebusch and Aruoba (2006), macroeconomic factors such as inflation and real economic activities are considered when modeling yield rates. Analogously, we include the percentage change in the monthly U.S. consumer pricing index (CPI) and industrial production index (IPI) as candidate exogenous covariates for macroeconomic conditions, denoted as X_{cpi}, and X_{ipi}, respectively. In addition, we include the Credit Suisse Fear Barometer level (X_{csfb}), CBOE/CBOT 10-year U.S. Treasury Note Volatility Index (X_{vix}), and U.S. government securities liquidity index (X_{gsli}) on the previous trading day as potential exogenous covariates. These three indices measure market conditions, namely, the fear level in the stock market, and the volatility and the liquidity level in the treasury bond market, respectively. Increases in X_{csfb} and X_{vix} represent increases in market fear and the volatility level, respectively, whereas an increase in X_{gsli} represents a decrease in the market liquidity level. Furthermore, we adopt the historical value of X_{gsli} as the threshold variable, with the delay parameter d tested from 1 to 10. All exogenous variable data are obtained from the Bloomberg Terminal.

The optimal model is selected with $\hat{d} = 1$; that is, the U.S. government securities liquidity index on the previous trading day is selected as the threshold variable. To demonstrate the significance of the threshold effect, we compare the fittings of the optimal fTARX model with those of a functional ARX model; see Table 1.

From Table 1, the fTARX model suggests a three-regime classification with respect to high, medium, and low market liquidity, which are specified as normal, intermediate, and financial crisis periods, respectively. Compared with the single-regime functional ARX model, the functional ARX model corresponding to each regime of the fTARX model is simpler. For the model diagnostics, we plot the residuals of the intermediate-step vector TARX model and the vector ARX model in Figures 3 and 4, respectively, in the Supplementary Material. Significant heteroskedasticity and frequent outliers are observed in Figure 4, which indicating a poor fit from the functional ARX model. On the other hand, Figure 3 suggests that the fTARX model provides an adequate fit. Therefore, we conclude that

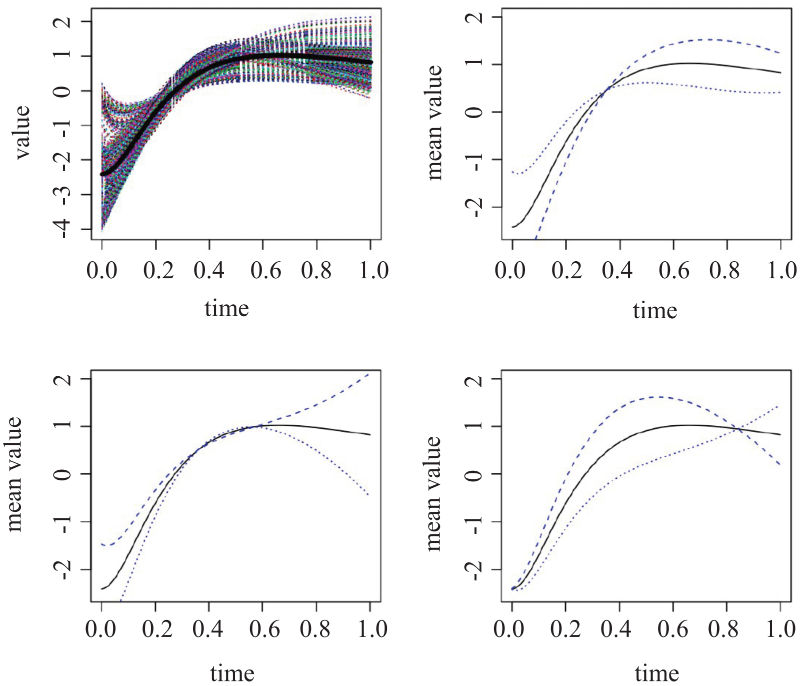

Figure 1. Plot of excess yield curve data, with the sample mean curve shown in bold (top-left). Plot of the mean curve, after adding and subtracting 0.5 times the first (top-right), second (bottom-left), and third (bottom-right) eigenfunctions, shown as dashed and dotted curve, respectively.

the fTARX model is more suitable for estimating the yield curve than is the functional ARX model. The estimates of the intercept vectors and coefficient matrices of the fTARX model are listed in (S.45) in the Supplementary Material. To quantify the performance of the models, we define the squared fitting error at time k as

$$SE_k = \int_0^1 [Y_k(t) - \hat{Y}_k(t)]^2 \, dt, \tag{6.1}$$

where $\hat{Y}_k$ denotes for the fitted value of Y_k. The mean squared fitting error (MSE) is computed as the average of the squared fitting errors from September 4, 2007, to January 3, 2011. The ratio of the MSEs of the fTARX and functional ARX models is 0.923, indicating that the fitting errors of the functional TARX model are smaller than those of the functional ARX model.

The selection of exogenous variables from the fTARX model fitting is interpreted as follows. First, the macroeconomic factors (i.e., the CPI and the IPI) are not selected in the fTARX model. This is in line with the findings of Ang and Piazzesi (2003) and Diebold, Rudebusch and Aruoba (2006) that although one-month lagged macroeconomic factors can contribute to error reduction and an improved R^2 statistic, the coefficients are not statistically significant and may cause an over-fitting problem.

Second, we observe that X_{csfb} is incorporated in all three regimes, which implies that the market fear factor plays an important role in modeling of a yield curve. In regime 1, the negative estimate $\hat{\mathbf{\Phi}}_{1,X_{csfb}}$ suggests that an increase in the fear of stock investment will lead to a decrease in the yield. Note that bond yields are inversely related to the price changes of bonds. Thus, the bond price will decrease in regime 1. However, as the market deteriorates from regime 1 to 2 and to 3, the coefficients of X_{csfb} increase and eventually become positive, which implies a reverse of the effect of stock market fear. Thus, the bond price decreases in regime 2 and increases in regime 3 as the stock market fear increases. These phenomena coincide with observations in the real financial market. In regime 1 of normal periods, investors transfer their risky stock investments to safer investments, such as treasury bonds, when the stock market fear rises, which increases the demand, and hence the price of treasury bonds. This transfer effect is called “flight to safety” in financial terminology. However, in regime 3 of financial crisis periods, the cross-asset contagion effect between the bond and the stock markets means that the fear in the bond market increases with that in the stock market. Owing to the possible panic, investors may leave the bond market, and bond prices decrease. Furthermore, in regime 2 of an intermediate state, stock market fear exhibits both the transfer effect and the contagion effect. The negative estimate $\hat{\mathbf{\Phi}}_{2,X_{csfb}}$ seems to indicate that the transfer effect dominates in this period. The sign of the coefficients of X_{csfb} under certain market conditions can indicate which of the two effects dominates at that time, and the corresponding absolute values of the coefficients measure the magnitude of this effect.

Furthermore, market volatility and liquidity are selected in some regimes. In particular, the larger the values of X_{gsli}, the less liquid the market becomes. In regime 1, the positive estimate $\hat{\mathbf{\Phi}}_{1,X_{gsli}}$ suggests a liquidity premium for a bond yield during the normal period. In addition, this liquidity premium is higher for bonds with medium or long maturity than it is for those with short maturity. However, X_{gsli} is excluded from regimes 2 and 3. This phenomenon can be explained as follows. When the market becomes less liquid in the intermediate period, financial traders are more likely to sell risky investments and buy treasury bonds to ensure safety. Thus, the bond price increases and the yields decrease, offsetting the effect of the liquidity premium.

Finally, X_{vix} is included in regime 2. According to Kalimipalli and Warga (2002), bond market volatility is positively related to the bid-ask spread, and negatively related to trading volume. Hence, the volatility reveals the degree of traders’ participation and confidence in the treasury market, and implies an upcoming recession or expansion during the intermediate period. The negative estimate of $\hat{\mathbf{\Phi}}_{2,X_{vix}}$ in regime 2 shows that the market volatility is negatively correlated with bond yields, because an increase in volatility changes the shape of the yield curve, and thus reduces the yields.

Lastly, we compute rolling real-time forecasts for each day from January 4, 2011, to October 19, 2011. The squared fitting error is modified to quantify the performance of the forecasting, as follows. The data until time $T-1$ are used to estimate the unknown parameters and model orders, as well as to predict the curve at time T. Then, we compute the forecast error at time T similarly to the computation of SE_T in (6.1), but replacing the fitted value $\hat{Y}_k$ with the prediction. Next, we obtain the mean squared prediction error (MSPE) by averaging the forecast errors of these 200 days. The ratio of the MSPEs of the fTARX and functional ARX models is 0.922, showing that the fTARX model exhibits better forecasting performance than that of the functional ARX model.

Supplementary Material

The supplementary materials contains proofs of the theorems, further probabilistic properties of the fMDL of fTARX models, and the optimization algorithms, simulation studies, and additional estimation and diagnostic results. It also describes how to construct confidence intervals for the parameters, and we extend the proposed method in Section 4 to the case where the functional parameters are parametrized by infinite-dimensional parameters.

Acknowledgments

We would like to thank the editor, associate editor and two anonymous referees for their critical comments and thoughtful suggestions. Chen's research was supported by the National Natural Science Foundation of China under Contract No. 12001444, the MOE (Ministry of Education in China) Project of Humanities and Social Sciences (20YJC910001), and the Fundamental Research Funds for the Central Universities. Yau's research was supported by HKSAR-RGC-GRF Nos. 14305517, 14302719, and 14304221.

References

Ang, A. and Piazzesi, M. (2003). A no arbitrage-autoregression of term structure dynamics with macroeconomic and latent factors. *Journal of Monetary Economics* **50**, 745–787.

Aue, A., Norinho, D. D. and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the Amercian Statistical Association* **110**, 378–392.

Aue, A., Horváth, L. and Pellatt, D. F. (2017). Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis* **38**, 3–21.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer-Verlag, New York.

Cerovecki, C., Francq, C., Hörmann, S. and Zakoian, J. M. (2019). Functional GARCH models: The quasi-likelihood approach and its applications. *Journal of Econometrics* **209**, 353–375.

Chen, C. W. S. and Lee, J. C. (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis* **16**, 483–492.

Chen, C. W. S., So, M. K. P. and Liu, F. C. (2011). A review of threshold time series models in finance. *Statistics and Its Interface* **4**, 167–181.

Chen, R. (1995). Threshold variable selection in open-loop threshold autoregressive models. *Journal of Time Series Analysis* **16**, 461–481.

Delaigle, A. and Hall, P. (2010). Defining probability density for a distribution of random functions. *The Annals of Statistics* **38**, 1171–1193.

Diebold, F. X., Rudebusch, G. D. and Aruoba, S. B. (2006). The macroeconomy and the yield curve: A dynamic latent factor approach. *Journal of Econometrics* **131**, 309–338.

Doukhan, P. (1994). *Mixing: Properties and Examples*. Springer-Verlag, New York.

Hansen, B. E. (2011). Threshold autoregression in economics. *Statistics and Its Interface* **4**, 123–127.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics* **38**, 1845–1884.

Hörmann, S. and Kokoszka, P. (2012). Functional time series. *Handbook of Statistics: Time Series Analysis–Methods and Applications* **110**, 157–186.

Hörmann, S., Horváth, L. and Reeder, R. (2013). A functional version of the ARCH model. *Econometric Theory* **29**, 267–288.

Kalimipalli, M. and Warga, A. (2002). Bid-ask spread, volatility and volume in the corporate bond market. *The Journal of Fixed Income* **11**, 31–42.

Kokoszka, P. and Reimherr, M. (2013). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis* **34**, 116–129.

Kolmogorov, A. N. and Rozanov, Y. A. (1960). On strong mixing conditions for stationary Gaussian proceses. *Theory Probability and Its Applications* **5**, 204–208.

Lange, T. and Rahbek, A. (2009). An introduction to regime switching time series models. In *Handbook of Financial Time Series*. Springer, Berlin, Heidelberg.

Lee, C. H. and Huang, B. N. (2002). The relationship between exports and economic growth in East Asian countries: A multivariate threshold autoregressive approach. *Journal of Economic Development* **27**, 45–68.

Lee, T. C. M. (2000). An introduction to coding theory and the two-part minimum description length principle. *International Statistical Review* **69**, 169–183.

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

Liu, X. L., Xiao, H. and Chen, R. (2016). Convolutional autoregressive models for functional time series. *Journal of Econometrics* **194**, 263–282.

Pan, J., Xia, Q. and Liu, J. (2017). Bayesian analysis of multiple thresholds autoregressive model. *Computational Statistics* **32**, 219–237.

Ramsay, J. O. and Silvermann, B. W. (2005). *Functional Data Analysis*. 2nd Edition. Springer-Verlag, New York.

Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry*. World Scientific, Singapore.

Rissanen, J. (2007). *Information and Complexity in Statistical Modeling*. Springer, New York.

Tong, H. (1978). On a threshold model. *Pattern Recognition and Signal Processing* **29**, 575–586.

Tong, H. (1990). *Threshold Models in Non-linear Time Series*. Springer, New York.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* **93**, 1188–1202.

Wu, S. L. and Chen, R. (2007). Threshold variable determination and threshold variable driven switching autoregressive models. *Statistica Sinica* **17**, 241–264.

Yau, C. Y., Tang, C. M. and Lee, T. C. M. (2015). Estimation of multiple-regime threshold autoregressive models with structural breaks. *Journal of the American Statistical Association* **110**, 1175–1186.

Yuanbo Li

School of Statistics, University of International Business and Economics, Beijing 100029, China.

E-mail: liyuanbo@uibe.edu.cn

Kun Chen

School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China.

E-mail: chenkun@swufe.edu.cn

Xunze Zheng

Department of Statistics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong.

E-mail: s1155002000@sta.cuhk.edu.hk

Chun Yip Yau

Department of Statistics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong.

E-mail: cyyau@sta.cuhk.edu.hk

(Received March 2022; accepted September 2022)