Vector Threshold Autoregressive Models with Bivariate Threshold Variables

YUAN, Gan

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Philosophy
in
Risk Management Science

The Chinese University of Hong Kong
July 2019

Thesis Assessment Committee

Professor CHAN, Kin Wai (Chair)

Professor YAU, Chun Yip (Thesis Supervisor)

Professor CHAN, Ngai Hang (Committee Member)

The Chinese University of Hong Kong
Graduate School

The undersigned certify that we have read a thesis, entitled "Vector Threshold Autoregressive Models with Bivariate Threshold Variables" submitted to the Graduate School by YUAN, Gan ( ) in partial fulfillment of the requirements for the degree of Master of Philosophy in Risk Management Science. We recommend that it be accepted.

_____

Prof. C.Y. Yau

Supervisor

_____

Prof. K.W. Chan

_____

Prof. N.H. Chan

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

Acknowledgements

I wish to thank Prof. Chun Yip Yau for supervising my M.Phil. study. He has offered me with insightful suggestions and enormous encouragement to overcome the challenges in my research and the preparation of this thesis.

Besides, I would like to thank the other three professors in my thesis assessment committee, Prof. Kin Wai Chan, Prof. Ngai Hang Chan and Prof. Henry Lam, for their valuable comments. I am also grateful to my fellow graduate students and colleagues in the Department of Statistics. It is fantastic to work with and learn from you in the past two years.

Finally, I must express my very profound gratitude to my family and friends for their endless support. This accomplishment would not have been possible without them. Thank you!

Abstract

This paper proposes a multivariate extension of the well-known threshold autoregressive (TAR) model in nonlinear time series literature. We consider a k-dimensional multiple-regime vector threshold autoregressive (VAR) model, in which the regime switching is governed by another observable bivariate time series, known as threshold variables. Specifically, the regimes are induced by a partition of $\mathbb{R}^2$ plane by an unknown number of threshold lines. Within each regime, the process is governed by a specific VAR model. We formulate the tasks of model selection and parameter estimation into a minimization problem based on minimum description length (MDL) principle, and develop a genetic algorithm (GA) to estimate the number of threshold lines, parametric forms of threshold lines and VAR model parameters in each regime simultaneously. The consistency of these estimators is established theoretically with an illustration by numerical simulations. Finally, we demonstrate the practical use of the model through an real data analysis on US interest rates.

Keywords: genetic algorithm, minimum description length principle, nonlinear time series, model selection

TAR$k\mathbb{R}^2$VARMDLVARGA

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Univariate Threshold Autoregressive Model

Nonlinear time series models has been developed extensively in the past several decades. *Threshold autoregressive* (TAR) model, proposed by Tong (1978), is one particularly popular class that performs well in modelling many empirical phenomena in both physical and social science.

A typical TAR model for univariate time series $\{y_t\}_{t\geq 0}$ has the form

$$y_t = \sum_{j=1}^{m} \left(\phi_0^{(j)} + \sum_{\tau=1}^{p^{(j)}} \phi_\tau^{(j)} y_{t-\tau} + a_t^{(j)} \right) \mathbb{1}\left(z_t \in I_j\right), \tag{1.1}$$

where $\phi_0^{(j)}, \ldots, \phi_{p^{(j)}}^{(j)}$ are real-valued AR parameters within the j-th regime and $a_t^{(j)} \sim N(0, \sigma_j^2)$ for some $\sigma_j > 0$. The central characteristic for a TAR process is that the model governing the process $\{y_t\}_{t\geq 0}$ shifts based on another observable threshold variables $\{z_t\}_{t\geq 0}$ and a partition of real line $\mathbb{R} = \bigcup_{j=1}^{m+1} I_j$, where $I_j = (r_{j-1}, r_j]$ for $j = 1, \ldots, m+1$, with $-\infty = r_0 < r_1 < \cdots < r_m < r_{m+1} = \infty$. In particular, when $z_t = y_{t-d}$, i.e. $\{z_t\}_{t\geq 0}$ is the lag-d series of $\{y_t\}_{t\geq 0}$, (1.1) is called *self-excited threshold*

autoregressive (SETAR) model. TAR models are able to capture patterns such as limit cycles and periodic behaviors, and hence are particularly useful in some certain fields. For example, it is used to model annual sunspot numbers, see Tong and Lim (1980), Tong (1983). In particular, TAR models have been deeply influential in economics and finance literature, see Hansen (2011) for a detailed review.

On theoretical side, the estimation of thresholds $r_1, \ldots, r_m$ has been widely studied in statistcal literature. When m is known, the least squared estimators for thresholds are n-consistent, where n is the sample size, see Chan (1993) for $m = 1$ and Li and Ling (2012) for the general case. Similar results can be found for maximum likelihood estimators, see Qian (1998). For unknown m, Yau et al. (2015) proposed a procedure to estimate m and thresholds simultaneously using minimum description length (MDL) principle. They proved that the number of thresholds m can be correctly estimated for sufficiently large n and the MDL estimators are also n-consistent.

1.2 Multivariate Extensions

In practice, a univariate TAR model is often insufficient due to the multivariate nature of many real applications. In this section, we review several multivariate extensions of TAR model in the literature.

1.2.1 Vector Threshold AutoregressiveModel

In general, a *vector threshold autoregressive* (VTAR) model for $\{\boldsymbol{y}_t\}_{t\geq 0}$ with $\boldsymbol{y}_t \in \mathbb{R}^k$ has the form

$$\boldsymbol{y}_t = \sum_{j=1}^{m} \left(\Phi_0^{(j)} + \sum_{\tau=1}^{p^{(j)}} \Phi_\tau^{(j)} y_{t-\tau} + \boldsymbol{a}_t^{(j)} \right) \mathbb{I}\left(z_t \in I_j\right),$$

where $\Phi_0^{(j)}, \ldots, \Phi_{p^{(j)}}^{(j)}$ are matrices of VAR parameters within the j-th regime and $\boldsymbol{a}_t^{(j)} \sim N(\mathbf{0}, \Sigma^{(j)})$ for some $\Sigma^{(j)} > 0$. Notice that the threshold variable z_t remains to be univariate in VTAR models and I_j's have the same definition as in univariate TAR models. Tsay (1998) provided a detailed discussion on estimation and testing for VTAR model with exogenuous variables and illustrated the use of VTAR on modelling in security markets and interest rates.

The estimation of VTAR models is similar to that of TAR models due to the unchanged fashion of regime partition. The only difference is that the AR model within each regime is now replaced by a VAR model. The inferences on the thresholds for VTAR can be extended readily from that of TAR models.

1.2.2 Multivariate Threshold Variables

In time series literature, efforts are made to extend the univariate threshold model to allow multivariate threshold variables, i.e. $\boldsymbol{z}_t \in \mathbb{R}^s$ for $s > 1$. Correspondingly, a partition of the threshold space $\mathbb{R}^s = \bigcup_{j=1}^r R_j$ is also needed so that the process $\{\boldsymbol{y}_t\}_{t \geq 0}$ behaves differently among model regimes $\{\boldsymbol{z}_t \in R_j\}$ for $j = 1, \ldots, r$. Models with multivariate threshold variables are rather involved and difficult to estimate due to the complexity of this partition when $s > 1$. To deal with this issue, it is necessary to have some restrictions on the partition.

One possible way is to restrict each R_j to be a s-cell, i.e. $(a_1^{(j)}, b_1^{(j)}] \times \cdots \times (a_s^{(j)}, b_s^{(j)}]$. Examples include

- Adaptive spline autoregressive (ASTAR) proposed by Lewis and Stevens (1991);
- Multivariate self-exciting threshold autoregressive models with exogenous input (MSETARX) proposed by Addo (2014) and
- TAR model with two threshold variables by Chen et al. (2012).

As an alternative, one may restrict that R_j being induced by hyperlines passing through the origin. For example, Tiao and Tsay (1994) considered a four-regime model with bivariate threshold variable corresponding to $\mathbb{R}^2 = \bigcup_{j=1}^{4} R_j$, with R_j's induced by two lines passing through the origin.

It seems that all the aforementioned models are too restrictive in the ways of partition for threshold space. It seems natural to extend the model by allowing the threshold space to be partitioned by an arbitrary number of threshold hyperplanes, without additional assumptions such as passing through origin or parallel to any axes. Meanwhile,

1.3 Organization of the Thesis

The first contribution of this thesis is to propose a new multivariate model, *vector threshold autoregressive model with bivariate threshold variables* (VTARBT). Our model allow bivariate threshold variables $\boldsymbol{z}_t$ and unknown number of threshold lines for the partition of threshold space $\mathbb{R}^2$.

As the second contribution, this thesis suggests the use of Minimum Description Length (MDL) for the model selection and parameter estimation of VTARBT models and provide the theoretical property of the corresponding estimators. We prove that the number of threshold lines is correctly estimated and establish consistency results of parameter estimators for VTARBT models.

Finally, we develop an algorithm for minimizing the non-regular MDL function. The majority works in threshold model estimation used grid search method, including Tsay (1998), Li and Ling (2012), etc. Such method is particularly inefficient for multivariate setting, especially when the number of threshold lines also need to be estimated and each threshold line can be chosen arbitrarily. In this thesis, we develop a genetic algorithm (GA) utilizing the idea of evolutionary programming to provide a

computationally efficient solution to the optimization problem.

The rest of this thesis is organized as follows. In Chapter 2 we formulate VTARBT model with a polar coordinate representation of threshold lines. In Chapter 3 we construct an MDL objective function for VTARBT model and propose an estimation scheme based on genetic algorithm. Some theoretical results that support our methodology are presented in Chapter 4. The proposed method is applied to simulated and real dataset in Chapters 5 and 6 respectively. Finally, Chapter 7 concludes the thesis.

Chapter 2

VTARBT Model

A general r-regime vector threshold autoregressive model with bivariate threshold variable (VTARBT) for k-dimensional process $\{\boldsymbol{y}_t\}_{t\geq 1} \equiv \{(y_{1t}, y_{2t}, \cdots, y_{kt})^T\}_{t\geq 1}$ can be defined as

$$\boldsymbol{y}_t = \sum_{j=1}^{r} \left(\Phi_0^{(j)} + \sum_{\tau=1}^{p^{(j)}} \Phi_\tau^{(j)} \boldsymbol{y}_{t-\tau} + \boldsymbol{a}_t^{(j)} \right) \mathbb{1}\left(\boldsymbol{z}_t \in R_j\right), \tag{2.1}$$

where $\mathbb{1}(\cdot)$ is the indicator function; $\{\boldsymbol{z}_t\}_{t\geq 1}$ with $\boldsymbol{z}_t \in \mathbb{R}^2$ is a prespecified bivariate series, known as threshold variables; $\{R_1, \cdots, R_r\}$ is a partition of $\mathbb{R}^2$. Under such settings, Model (2.1) has r regimes corresponding to $\{\boldsymbol{z}_t\}_{t\geq 0}$ falling in R_j for $j = 1, \ldots, r$. Within the j-th regime, the process $\{\boldsymbol{y}_t\}_{t\geq 0}$ is governed by a k-dimensional VAR model, where $p^{(j)}$ is the AR order, $\Phi_0^{(j)}, \Phi_1^{(j)}, \ldots, \Phi_{p^{(j)}}^{(j)}$ are coefficient matrices and $\boldsymbol{a}_t^{(j)} \sim N(\boldsymbol{0}_k, \Sigma^{(j)})$ with $\Sigma^{(j)} > 0$. We define $\Psi^{(j)} \equiv (\Sigma^{(j)}, \Phi_0^{(j)}, \Phi_1^{(j)}, \cdots, \Phi_{p^{(j)}}^{(j)})$ as the matrix of model parameters.

For simplicity, we assume that all the regimes are induced by lines in $\mathbb{R}^2$ (*threshold lines*) in later development. Each threshold line can be fully described by two parameters θ and ρ. As shown in Figure 2.1, suppose A is a point on a threshold line l such that $\overrightarrow{OA} \perp l$. Then, we define ρ as the length of vector $\overrightarrow{OA}$, i.e. the distance

of origin and l, and θ as the angle between positive x_1-axis and $\overrightarrow{OA}$, measured counterclockwisely. Thus, every line in $\mathbb{R}^2$ plane can be uniquely represented by a pair (θ, ρ), where $\theta \in [0, 2\pi)$, $\rho \in [0, +\infty)$. Although extra efforts are needed for a more delicate parameterization and algorithm designing, the theoretical results in Chapter 4 can be readily generalized for regimes induced by more complex partitions, as long as the model complexity (as introduced in Chapter 3) is of order $O(\log n)$.

Once the threshold lines are determined, the partition $\{\tilde{R}_1, \ldots, \tilde{R}_r\}$ can be defined accordingly. When there is only one threshold line l: (θ, ρ) as in Figure 2.1, there are apparently two subregions

$$\begin{aligned} \tilde{R}_1(\theta, \rho) &= \{\boldsymbol{x} = (x_1, x_2)^T : d(\boldsymbol{x}, l) := x_1 \cos\theta + x_2 \sin\theta \geq \rho\} \quad \text{and} \\ \tilde{R}_2(\theta, \rho) &= \{\boldsymbol{x} = (x_1, x_2)^T : d(\boldsymbol{x}, l) := x_1 \cos\theta + x_2 \sin\theta < \rho\}. \end{aligned} \tag{2.2}$$

Similarly, if there are two threshold lines l_1: (θ_1, ρ_1) and l_2: (θ_2, ρ_2) as in Figure 2.2(a), there are at most four subregions

$$\begin{aligned} \tilde{R}_1(\boldsymbol{\theta}, \boldsymbol{\rho}) &= \{\boldsymbol{x} = (x_1, x_2)^T : d(\boldsymbol{x}, l_1) \geq \rho_1, d(\boldsymbol{x}, l_2) < \rho_2\}, \\ \tilde{R}_2(\boldsymbol{\theta}, \boldsymbol{\rho}) &= \{\boldsymbol{x} = (x_1, x_2)^T : d(\boldsymbol{x}, l_1) \geq \rho_1, d(\boldsymbol{x}, l_2) \geq \rho_2\}, \\ \tilde{R}_3(\boldsymbol{\theta}, \boldsymbol{\rho}) &= \{\boldsymbol{x} = (x_1, x_2)^T : d(\boldsymbol{x}, l_1) < \rho_1, d(\boldsymbol{x}, l_2) \geq \rho_2\}, \\ \tilde{R}_4(\boldsymbol{\theta}, \boldsymbol{\rho}) &= \{\boldsymbol{x} = (x_1, x_2)^T : d(\boldsymbol{x}, l_1) < \rho_1, d(\boldsymbol{x}, l_2) < \rho_2\}. \end{aligned} \tag{2.3}$$

Notice that if l_1 and l_2 are parallel, then one of $\tilde{R}_1, \tilde{R}_2, \tilde{R}_3, \tilde{R}_4$ is an empty set. The partition with more threshold lines can be defined similarly.

It is possible that some subregions from $\{\tilde{R}_1, \ldots, \tilde{R}_r\}$ are corresponding to one same regime. Under such situation, It is natural to consider to merge some of the subregions so that a one-to-one relationship is established between merged subregions and model regimes. For convenience, we define a dummy variable $q \in \mathbb{N}$ to indicate the merge

pattern. Table 2.1 shows a possible definition of q when there are two threshold lines. In general, there is an upper bound $\bar{q}(m)$ for number of different merge patterns, given that there are m threshold lines. Given m, q, $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$, we can then recover the original parameter r and $R_1, \ldots, R_r$ in Model (2.1). For example, if $m = 2, q = 4$, then $r = 3$, $R_1 = \tilde{R}_1(\boldsymbol{\theta}, \boldsymbol{\rho}) \cup \tilde{R}_4(\boldsymbol{\theta}, \boldsymbol{\rho})$ and $R_2 = \tilde{R}_2(\boldsymbol{\theta}, \boldsymbol{\rho})$ and $R_3 = \tilde{R}_3(\boldsymbol{\theta}, \boldsymbol{\rho})$. See Figure 2.2(b) for a graphical illustration.

Let $\mathcal{M}$ be the class of VTARBT models satisfying (2.1). A generic model from this class is denoted by $\mathcal{F} \equiv \mathcal{F}(\boldsymbol{\theta}, \boldsymbol{\rho}, q, \boldsymbol{p}, \boldsymbol{\Psi}) \in \mathcal{M}$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^T$, $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_m)^T$, $\boldsymbol{p} = (p^{(1)}, \ldots, p^{(r)})^T$, $\boldsymbol{\Psi} = (\Psi^{(1)}, \ldots, \Psi^{(r)})$, m is the number of threshold lines and $r \equiv r(\boldsymbol{\theta}, \boldsymbol{\rho}, q)$ is the number of threshold lines.

Value of q	Regions to Merge	Value of q	Regions to Merge
1	$\tilde{R}_1$ and $\tilde{R}_2$	8	$\tilde{R}_1$ and $\tilde{R}_3$, $\tilde{R}_2$ and $\tilde{R}_4$
2	$\tilde{R}_2$ and $\tilde{R}_3$	9	$\tilde{R}_1$ and $\tilde{R}_2$ and $\tilde{R}_3$
3	$\tilde{R}_3$ and $\tilde{R}_4$	10	$\tilde{R}_2$ and $\tilde{R}_3$ and $\tilde{R}_4$
4	$\tilde{R}_4$ and $\tilde{R}_1$	11	$\tilde{R}_3$ and $\tilde{R}_4$ and $\tilde{R}_1$
5	$\tilde{R}_1$ and $\tilde{R}_3$	12	$\tilde{R}_4$ and $\tilde{R}_1$ and $\tilde{R}_2$
6	$\tilde{R}_2$ and $\tilde{R}_4$	13	$\tilde{R}_1$ and $\tilde{R}_2$ and $\tilde{R}_3$ and $\tilde{R}_4$
7	$\tilde{R}_1$ and $\tilde{R}_2$, $\tilde{R}_3$ and $\tilde{R}_4$	14	No Merges

Table 2.1: *Value of q and Corresponding Merge Pattern for Two Threshold Lines*

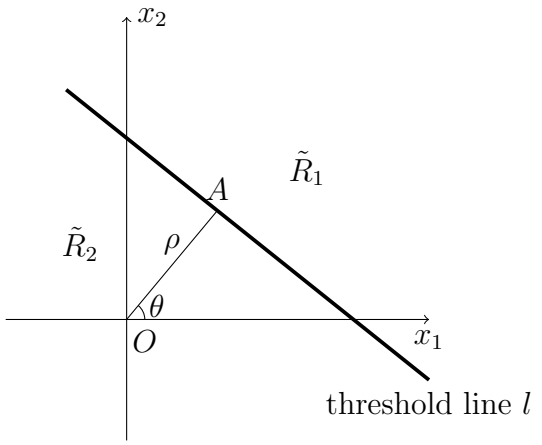
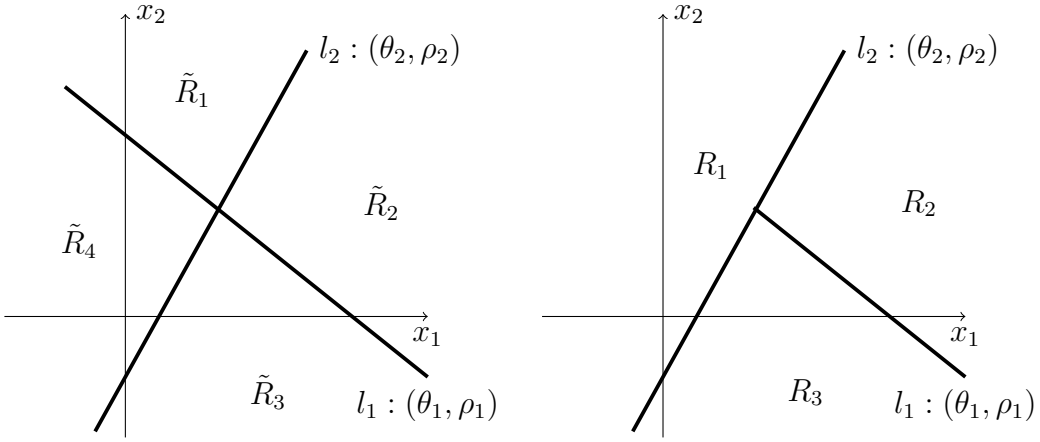


Figure 2.1: *Parameterization of Threshold Line and Subregions*



(a) Preliminary Partition (b) Merge Pattern $q = 4$

Figure 2.2: *Preliminary Partition and Merging with Two Threshold Lines*

Chapter 3

Model Selection and Parameter Estimation

In this chapter, we develop a criterion function (i.e. MDL function) using minimum description length, and adopt genetic algorithm to find the best fitting VTARBT model. Similar to Yau et al. (2015), we adopt "two-part MDL" introduced by Hansen and Yu (2001) and Lee (2001). In particular, a model $\mathcal{F} \in \mathcal{M}$ and corresponding residuals $\{\boldsymbol{a}_t(\mathcal{F})\}_{1\leq t\leq n}$ store all information in original data $\{\boldsymbol{y}_t\}_{1\leq t\leq n}$. This leads to the expression:

$$\boldsymbol{CL}_{\mathcal{F}}(\{\boldsymbol{y}_t\}_{1\leq t\leq n}) = \boldsymbol{CL}(\mathcal{F}) + \boldsymbol{CL}(\{\boldsymbol{a}_t(\mathcal{F})\}_{1\leq t\leq n})\,, \tag{3.1}$$

where $\boldsymbol{CL}(\mathcal{F})$ is the number of bits (codelength) needed for storing model $\mathcal{F}$ in computer storage and $\boldsymbol{CL}(\{\boldsymbol{a}_t(\mathcal{F})\}_{1\leq t\leq n})$ is the codelength needed for storing the residual under model $\mathcal{F}$. Our aim is to find the best fitted model $\hat{\mathcal{F}} \equiv \hat{\mathcal{F}}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\rho}}, \hat{q}, \hat{\boldsymbol{p}}, \hat{\boldsymbol{\Psi}})$, with which $\boldsymbol{CL}_{\hat{\mathcal{F}}}(\{\boldsymbol{y}_t\}_{1\leq t\leq n})$ is minimized. That is, we need to find the model with minimum description length:

$$\hat{\mathcal{F}} = \underset{\mathcal{F}\in\mathcal{M}}{\arg\min}\, \boldsymbol{CL}_{\mathcal{F}}(\{\boldsymbol{y}_t\}_{1\leq t\leq n})\,. \tag{3.2}$$

3.1 Detailed MDL for VTARBT Models

In this section, we discuss the calculation of $\boldsymbol{CL}_{\mathcal{F}}(\{\boldsymbol{y}_t\}_{1\leq t\leq n})$ in detail. First, we derive the expression for $\boldsymbol{CL}(\mathcal{F})$. Clearly, to store $\mathcal{F}$ is equivalent to store all the model parameters, i.e.

$$\boldsymbol{CL}(\mathcal{F}) = \boldsymbol{CL}(\boldsymbol{\theta}) + \boldsymbol{CL}(\boldsymbol{\rho}) + \boldsymbol{CL}(q) + \boldsymbol{CL}(\boldsymbol{p}) + \boldsymbol{CL}(\boldsymbol{\Psi})\,. \tag{3.3}$$

To encode integer-valued variable with a predetermined upper bound I, $\log_2 I$ bits are needed. Therefore, $\boldsymbol{CL}(q) = \log_2 \bar{q}(m)$ bits and encoding $p^{(j)}$ requires $\boldsymbol{CL}(p^{(j)}) = \log_2 \bar{p}$ bits, where $\bar{p}$ is a predetermined uniform upper bound for AR orders within all model regimes. Suppose n is the total sample size and n_j is the number of points falls in j-th regime. For each of the m-vectors $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$, $m \log n/2$ bits is needed for storage with an optimized truncation scheme introduced by Rissanen (1989). Similarly, $\boldsymbol{\Psi}^{(j)}$ can be regarded as $k + k^2p^{(j)} + k(k+1)/2 = k^2p^{(j)} + k(k+3)/2$ real-valued Gaussian maximum likelihood estimators, each requiring $\log(n_j)/2$ bits. Thus, $\boldsymbol{CL}(\boldsymbol{\Psi}^{(j)}) = (k^2p^{(j)} + k(k+3))\log n_j/4$. To summarize the above discussion,

$$\boldsymbol{CL}(\mathcal{F}) = \log_2 \bar{q}(m) + r\log_2 \bar{p} + \sum_{j=1}^{r}\left[\frac{(k^2p^{(j)} + k(k+3))\log n_j}{4}\right] + m\log n\,. \tag{3.4}$$

Next, we derive the expression of $\boldsymbol{CL}(\boldsymbol{a_t}(\mathcal{F}))$. Within the j-th regime, the model is:

$$\mathcal{F}_j : \boldsymbol{y}_t = \Phi_0^{(j)} + \sum_{\tau=1}^{p^{(j)}} \Phi_\tau^{(j)} \boldsymbol{y}_{t-\tau} + \boldsymbol{a}_t^{(j)}\,, \tag{3.5}$$

where $\boldsymbol{a}_t^{(j)} \sim N(0, \Sigma^{(j)})$. Following Rissanen (1989), the codelength of residual sequence can be well approximated by the negative log-likelihood of the model $\mathcal{F}_j$. Therefore, to minimize the codelength, maximum likelihood estimator (MLE) should be adopted.

Specifically, for fixed $\boldsymbol{\theta}, \boldsymbol{\rho}, q, p^{(j)}$, we extract all the observations that falls in j-th regime,

$$\left\{\boldsymbol{y}_{t_1}^{(j)}, \ldots, \boldsymbol{y}_{t_{n_j}}^{(j)}\right\} \equiv \left\{\boldsymbol{y}_t : \boldsymbol{z}_t \in R_j(\boldsymbol{\theta}, \boldsymbol{\rho}, q)\right\},$$

where $\bar{p}+1 \leq t_1 < \cdots < t_{n_j} \leq n$. Then, the MLE is

$$\hat{\Psi}_n^{(j)} \equiv \hat{\Psi}_n^{(j)}(\boldsymbol{\theta}, \boldsymbol{\rho}, q, p^{(j)}) = \arg\max_{\Psi} L(\boldsymbol{y}_{t_1}^{(j)}, \ldots, \boldsymbol{y}_{t_{n_j}}^{(j)} | \Psi), \tag{3.6}$$

where

$$L\left(\boldsymbol{y}_{t_1}^{(j)}, \ldots, \boldsymbol{y}_{t_{n_j}}^{(j)} | \Psi^{(j)}\right) = \prod_{i=1}^{n_j} \frac{1}{(2\pi)^{k/2}|\Sigma^{(j)}|^{\frac{1}{2}}} \exp\left(\frac{-\left(\boldsymbol{a}_{t_i}^{(j)}(\boldsymbol{\Phi}^{(j)})\right)^T \left(\Sigma^{(j)}\right)^{-1} \left(\boldsymbol{a}_{t_i}^{(j)}(\boldsymbol{\Phi}^{(j)})\right)}{2}\right), \tag{3.7}$$

and $\boldsymbol{a}_t^{(j)}(\boldsymbol{\Phi}^{(j)}) \equiv \boldsymbol{a}_t^{(j)}(\mathcal{F}) := \boldsymbol{y}_t - \left(\Phi_0^{(j)} + \Phi_1^{(j)}\boldsymbol{y}_{t-1} + \cdots + \Phi_{p^{(j)}}^{(j)}\boldsymbol{y}_{t-p^{(j)}}\right)$. Let $Y^{(j)} \in \mathbb{R}^{n_j \times k}$ with i-th row being $(\boldsymbol{y}_{t_i}^{(j)})^T$, $X^{(j)} \in \mathbb{R}^{n_j \times (kp^{(j)}+1)}$ with i-th row being $\left(1, (\boldsymbol{y}_{t_i-1}^{(j)})^T, \cdots, (\boldsymbol{y}_{t_i-p^{(j)}}^{(j)})^T\right)$, and $\boldsymbol{\Phi}^{(j)} = \left(\Phi_0^{(j)}, \cdots, \Phi_{p^{(j)}}^{(j)}\right)^T \in \mathbb{R}^{(kp^{(j)}+1)\times k}$. Then, by standard calculation, we have

$$\hat{\boldsymbol{\Phi}}_n^{(j)} = \left(\left(X^{(j)}\right)^T \left(X^{(j)}\right)\right)^{-1} \left(X^{(j)}\right)^T Y^{(j)}, \tag{3.8}$$

$$\hat{\Sigma}_n^{(j)} = \frac{1}{n_j}\left(Y^{(j)} - X^{(j)}\hat{\boldsymbol{\Phi}}_n^{(j)}\right)^T \left(Y^{(j)} - X^{(j)}\hat{\boldsymbol{\Phi}}_n^{(j)}\right). \tag{3.9}$$

By substituting (3.8) and (3.9) into (3.7) and taking logarithm,

$$\begin{aligned} l\left(\boldsymbol{y}_{t_1}^{(j)}, \ldots, \boldsymbol{y}_{t_{n_j}}^{(j)} | \hat{\Psi}_n^{(j)}\right) &= -\frac{n_j}{2}(k\log 2\pi + \log|\hat{\Sigma}_n^{(j)}|) - \frac{1}{2}tr(n_j I) \\ &= -\frac{n_j}{2}(k\log 2\pi + \log|\hat{\Sigma}_n^{(j)}| + k), \end{aligned} \tag{3.10}$$

where I is a $k \times k$ identity matrix. It follows that

$$\begin{aligned} \boldsymbol{CL}(\{\boldsymbol{a_t}(\mathcal{F})\}_{1\le t\le n}) &= \sum_{j=1}^{r} \frac{n_j}{2}(k\log 2\pi + \log|\hat{\Sigma}_n^{(j)}| + k) \\ &= \frac{nk}{2}(1+\log 2\pi) + \sum_{j=1}^{r}\frac{n_j}{2}\log|\hat{\Sigma}_n^{(j)}|\,. \end{aligned} \tag{3.11}$$

Finally, by combining (3.1), (3.4) and (3.11), we obtain the following expression for MDL function,

$$\begin{aligned} \mathrm{MDL}(\boldsymbol{\theta}, \boldsymbol{\rho}, q, \boldsymbol{p}) &= \boldsymbol{CL}_{\mathcal{F}\left(\boldsymbol{\theta},\boldsymbol{\rho},q,\boldsymbol{p},\hat{\Psi}(\boldsymbol{\theta},\boldsymbol{\rho},q,\boldsymbol{p})\right)}(\{\boldsymbol{y}_t\}_{1\le t\le n}) \\ &= \log_2 \bar{q}(m) + r\log_2 \bar{p} + \frac{nk}{2}(1+\log 2\pi) + m\log_2 n \\ &\quad + \sum_{j=1}^{r}\left[\frac{k^2p^{(j)} + k(k+3)}{4}\log_2 n_j + \frac{n_j}{2}\log|\hat{\Sigma}_n^{(j)}|\right]. \end{aligned} \tag{3.12}$$

By minimizing (3.12) with respect to $\boldsymbol{\theta}, \boldsymbol{\rho}, q, \boldsymbol{p}$, we can find the $\hat{\mathcal{F}}$, as defined in (3.2).

3.2 Minimizing MDL Using Genetic Algorithm

3.2.1 General Description and Settings

Our aim is to find $\boldsymbol{\theta}, \boldsymbol{\rho}, q, \boldsymbol{p}$ such that MDL function in (3.12) is minimized. As shown in Section 3.1, the MDL function is not continuous with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$. It means that regular optimization via differentiation does not work. In this thesis, we follow the successes in Lee (2001); Davis et al. (2006); Lu et al. (2010) and Yau et al. (2015), and apply genetic algorithm (GA) to achieve our goal of minimizing (3.12). In essence, GA incorporates the idea of *natural selection* to reduce MDL generation by generation. A brief summary of some related terminologies is as follows:

Chromosomes are vectors that store candidate solutions. The *population* is a set

of chromosomes that serve as the original generation. There are four major operations involved in genetic algorithm. *Crossover* is a procedure where two (parent) chromosomes are randomly picked from current generation to produce a (child) chromosome that constitutes the next generation. The odd of a parent being chosen is of probability inversely proportional to the rank of its MDL function value. Mimicing genetic heredity in sexual reproduction, the offsprings have equal chance to inherit its traits from either of the parents. Crossover is a critical step where good features can get combined and passed on from parents to their offsprings. *Mutation* prevents the algorithm from being trapped in local optima. Some appropriate random changes will be happened with predetermined probabilities so that more potential solutions will be possible to be explored. *Elite step* guarantees the monotonicity of GA by replacing the worst chromosome of the next generation with the best one in current generation. This guarantees that the best performed chromosome in one generation will be preserved no matter what happens in crossovers and mutations. Finally, We adopt *island model*, similar to the one raised in Alba and Troya (1999), to divide the original optimization problem into several sub-optimization problems. Island model can not only avoid the algorithm being trapped in local optima, but also more compatible for parallel computing. Communication among islands are realized through *migration* operations.

3.2.2 Implementation of GA

In this section, we describe the implementation of major steps of GA in details.

(1) *Chromosome Representation*

A chromosome is a vector that carries complete information for a VARTMT model. In our situation, we need to specify the number of threshold lines m, threshold line parameters $\boldsymbol{\theta}, \boldsymbol{\rho}$ and merge pattern q. Notice that the number of regimes $r \equiv r(\boldsymbol{\theta}, \boldsymbol{\rho}, q)$ is automatically determined. Besides, we need to specify the AR

order in each regime $p^{(1)}, \ldots, p^{(r)}$ for each regime. As long as these quantities are provided, a unique maximum likelihood estimators of other model parameters are automatically determined. Consider the difficulty of including merge pattern q in GA, we delay the fine-tuning step of choosing merge pattern after determining the threshold lines. By ignoring potential mergings, the estimated regimes given by GA can be finer than the true model. In particular, a typical chromosome has the form:

$$[m, \boldsymbol{\theta}, \boldsymbol{\rho}, \boldsymbol{p}] .$$

(2) *Initialization of Generation Zero*

The algorithm starts with initialization of Generation Zero, a population of N_C chromosomes. A typical chromosome is generated through the following scheme. The number of threshold lines m is randomly selected from $\{1, \ldots, \bar{m}\}$ with equal probabilities. In our application, we choose $\bar{m} = 4$. For each $j = 1, \ldots, m$, we generate θ_j from uniform distribution over $(0, 2\pi)$. Similarly, ρ_j can be generated from uniform distribution over $(0, \bar{\rho})$ where $\bar{\rho} := \sup\{\, |\boldsymbol{z}_t| \,|t = 1, \ldots, n\}$. To ensure that maximum likelihood estimation can be smoothly conducted within each model regime, we set a minimum requirement for number of observations $N_{min} = 2\bar{p}$ fulling in each regime. We should check whether this condition holds whenever a new choromosome is initialized or produced. Those chromosome violating this condition should be discarded. Finally, the AR order $p^{(j)}$ is randomly selected from $\{1, 2, \cdots, \bar{p}\}$ with equal probabilities. In our implementation, we set $\bar{p} = 4$.

(3) *Crossover and Mutation*

From the randomly initiated Generation Zero, we can perform either crossover or mutation to reproduce offspring generations. The optimization will be realized through generations. The fundamental idea of crossover is to let *good features* of

parents be inherited by their offspring. In our implementation, we use the following rules to select parents and reproduce offspring. First, two parent chromosomes are randomly picked from the incumbent generation, with the probability inversely proportional to their ranks sorted by MDL values. Such selection method ensures that chromosomes with smaller MDLs have higher probabilities to be picked. The rules for reproducing offspring are as follow. Suppose we have two parents

$$\textit{Parent 1} : [m_1, \boldsymbol{\theta}_1, \boldsymbol{\rho}_1, \boldsymbol{p}_1] \quad \text{and}$$
$$\textit{Parent 2} : [m_2, \boldsymbol{\theta}_2, \boldsymbol{\rho}_2, \boldsymbol{p}_2] \,.$$

Then a child with form

$$\textit{Child} : [m_c, \boldsymbol{\theta}_c, \boldsymbol{\rho}_c, \boldsymbol{p}_c]$$

is generated by the following rules. First, m_c is randomly picked from $\{m_1, m_2\}$ with equal probabilities. That is, the child chromosome has the same number of threshold lines with either one of its parents. Then m_c threshold lines are drawn randomly, again with equal probability from the set containing all threshold lines in both *Parent 1* and *Parent 2*. Up to now, all threshold lines and regimes are specified in the child chromosome. Finally, the AR order $p_j^{(c)}$ for $R_j^{(c)}$, is inherited from the most similar parent regime among all $R_j^{(i)}$ for $i = 1, 2$ and $j = 1, \ldots, r_i$. Specifically, we use Jaccard Index $J(A, B) = |A \cap B|/|A \cup B| \in (0, 1)$ to quantify the notion "similarity" between two sets A and B. For each $R_j^{(i)}$, we define $\Gamma_j^{(i)} := \{\boldsymbol{z}_t \in R_j^{(i)} : t = 1, \ldots, n\}$. Then, we choose $p_j^{(c)} = p_{k_j^*}^{(i_j^*)}$ with

$$(i_j^*, k_j^*) := \underset{i=1,2;\, k=1,\ldots,r_i}{\arg\max} \; J(\Gamma_j^{(c)}, \Gamma_k^{(i)}) \,.$$

For illustration, suppose we have selected two parent chromosomes from the pop-

ulation, i.e.

$$\textit{Parent 1} : [1, (0), (4), (3, 2)] \quad \text{and}$$
$$\textit{Parent 2} : [2, (3\pi/4, \pi/4), (2, 3), (4, 3, 2, 1)] .$$

Then a possible child is:

$$\textit{Child} : [1, (3\pi/4), (2), (2, 1)] .$$

To see how the child chromosome is generated, we may refer to Figure 3.1. First, the number of lines m_c is randomly selected from the set $\{m_1, m_2\} = \{1, 2\}$. In our example, we assume that $m_1 = 1$ is selected. Then, the threshold line l_1^2 with $\theta = \pi/4, \rho = 3$ among $\{l_1^1, l_1^2, l_2^2\}$ is selected. This yields two regimes $R_1^{(c)}$ and $R_2^{(c)}$ in the child chromosome as indicated in Figure 3.1(c). Finally, we need to decide AR orders within both regimes. Suppose we have $n = 10$ observations A to J as in Figure 3.1(d). Then, it can be easily seen that

$$\Gamma_2^{(c)} = \{B, C, D, E, F, G, H\} ,$$
$$\Gamma_1^{(1)} = \{A, B, C, D, F, G, J\} , \ \Gamma_2^{(1)} = \{E, H, I\} ,$$
$$\Gamma_1^{(2)} = \{F, G, H\} , \ \Gamma_2^{(2)} = \{A, J, I\} , \ \Gamma_3^{(2)} = \varnothing \quad \text{and} \quad \Gamma_4^{(2)} = \{B, C, D, E\} .$$

Among $\Gamma_1^{(1)}$, $\Gamma_2^{(1)}$, $\Gamma_1^{(2)}$, $\Gamma_2^{(2)}$, $\Gamma_3^{(2)}$ and $\Gamma_4^{(2)}$,

$$J(\Gamma_2^{(c)}, \Gamma_4^{(2)}) = 4/7$$

is the largest. Therefore, we choose $p_2^{(c)} = p_4^{(2)} = 1$. Similarly, we can do the same analysis and determine that $p_1^{(c)} = p_2^{(2)} = 2$.

(4) *Mutation and Elite Step*

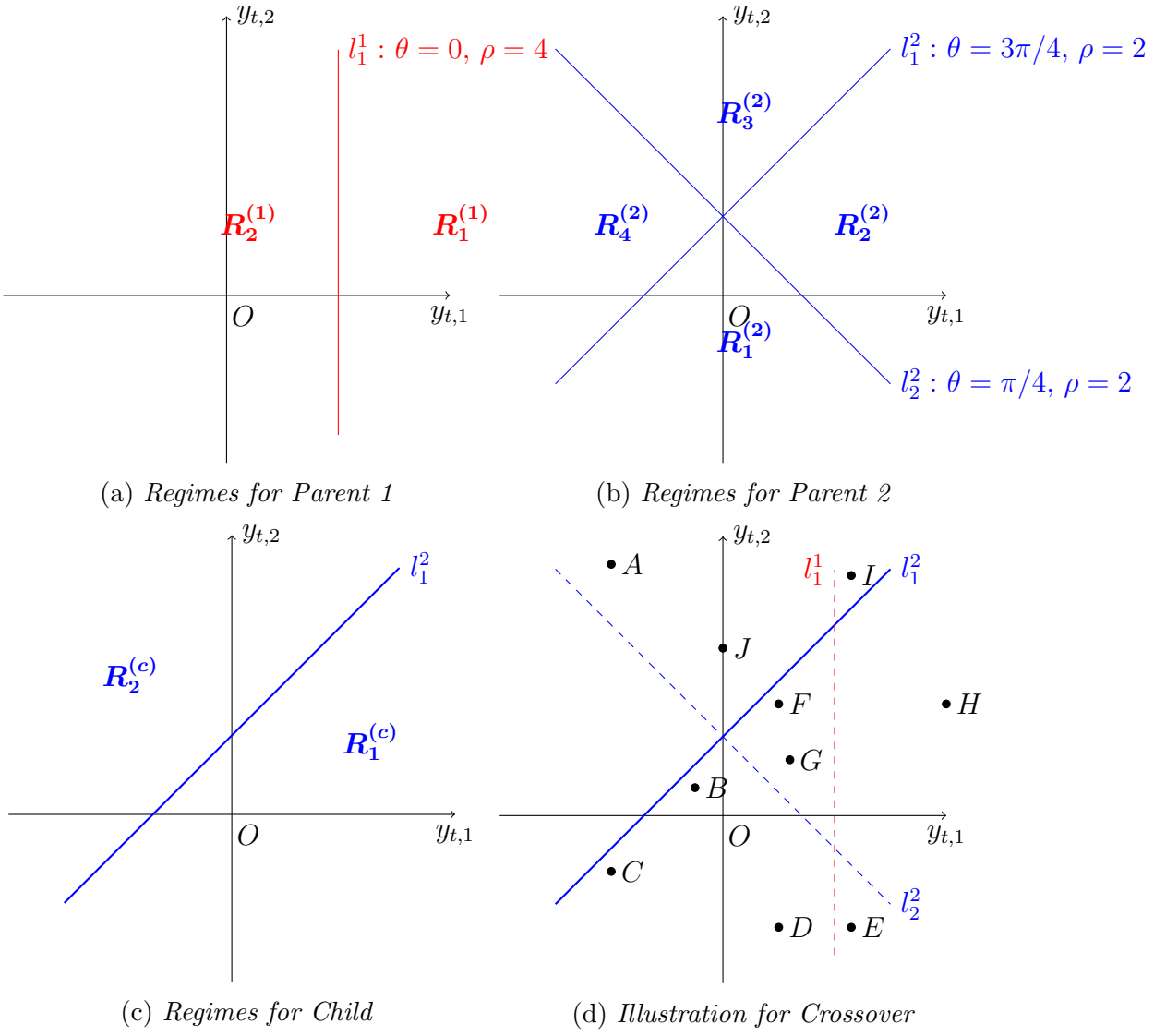
To avoid being trapped in local optima, we introduce mutation operations in the algorithm. Mutations happen during crossovers with probability $\pi_m = 0.1$. In such situation, one of the parents is randomly generated using the method in *Initialization of Population Generation* rather than being drawn from the current generation. Then, crossover will take place to generate a child chromosome from one mutated parent and one parent drawn from incumbent generation. Another critical procedure to accelerate the convergence of the algorithm is elite step. After conducting crossover for each generation, we replace some " worst performing" chromosomes (say, ten chromosome with largest MDL values) of the offspring generation by the " best performing" *elite chromosomes* of parent generation. Meanwhile, all but one of these elite chromosomes are slightly perturbed by adding a small random quantity following Uniform $(-\delta, \delta)$. In our practice, we choose $\delta = 0.02$. This enables us to find the local optima more efficiently.

(5) *Island Model and Migration*

As stated earlier, we implement island model in practice. Instead of initializing and operating in a giant population, we simultaneously run $N_I = 50$ islands containing $N_C = 100$ chromosomes in each of them. To connect each of the sub-optimization processes, migration operation allows chromosomes to go through islands periodically, say every $M_i = 5$ generations. In detail, the worst two chromosomes from the j-th island are replaced by the best two chromosomes from the $(j-1)$-th island, where $j = 2, \ldots, N_I$. For $j = 1$, the worst two chromosomes are replaced from the best chromosomes from the N_I-th island.

3.2.3 Declaration of Convergence

Similar to Davis et al. (2006) and Yau et al. (2015), the overall best chromosome among all the island with smallest MDL is noted as the temporary solution for the optimization problem. If a temporary solution does not change for 10 consecutive generations, or if 30 generations of crossovers and mutations have been conducted, it will be taken as the final solution.



(a) *Regimes for Parent 1*

(b) *Regimes for Parent 2*

(c) *Regimes for Child*

(d) *Illustration for Crossover*

Figure 3.1: *Graph Illustration for Crossover*

Chapter 4

Theoretical Properties

In this chapter, we establish the consistency of the MDL procedure and find the rate of convergence for model parameter estimators.

4.1 Notations and Assumptions

In general, VTARBT model (2.1) is determined by a set of parameters $(\boldsymbol{\theta}, \boldsymbol{\rho}, q, \boldsymbol{\Psi}, \boldsymbol{p})$. For simplicity, we denote $\boldsymbol{\Xi} \equiv (\boldsymbol{\theta}, \boldsymbol{\rho}, q)$ as the parameters needed for identifying regimes. In particular, we can arrange the order $R^{(1)}, R^{(2)}, \ldots, R^{(r)}$ using some pre-specified rules, where $r = r(\boldsymbol{\Xi})$ is the total number of regimes. For example, one can number them by their shortest distance to the origin. Recall from (3.12) that

$$\mathrm{MDL}(\boldsymbol{\Xi}, \boldsymbol{p}) = \mathrm{Pen}(\boldsymbol{\Xi}, \boldsymbol{p}) + \frac{1}{2} \sum_{j=1}^{r} L_n(\hat{\Psi}^{(j)}, R^{(j)}(\boldsymbol{\Xi}), p^{(j)}) \,, \tag{4.1}$$

with

$$\mathrm{Pen}(\boldsymbol{\Xi}, \boldsymbol{p}) \equiv \sum_{j=1}^{r} \log_2 p^{(j)} + \sum_{j=1}^{r} \frac{(k^2 p^{(j)} + k(k+3)) \log_2 n^{(j)}}{4} + m(4 \log_2 n + 2) \ ,$$

$$L_n(\Psi^{(j)}, R^{(j)}(\boldsymbol{\Xi}), p^{(j)}) \equiv \sum_{t=1}^{n} \left(k \log 2\pi + \log |\Sigma^{(j)}| + \boldsymbol{a}_t'(\boldsymbol{\Phi}^{(j)})(\Sigma^{(j)})^{-1} \boldsymbol{a}_t(\boldsymbol{\Phi}^{(j)}) \right) I(\boldsymbol{z}_t \in R^{(j)}) \ ,$$

for $i = 1, 2, \ldots, n^{(j)}$; $p^{(j)}$ and $\Psi^{(j)} = (\Sigma^{(j)}, \Phi_0^{(j)}, \ldots \Phi_{p^{(j)}}^{(j)})$ are the AR order and AR parameters in $R^{(j)}$, and

$$\hat{\Psi}_n^{(j)} \equiv \hat{\Psi}_n^{(j)}(\boldsymbol{\Xi}, p^{(j)}) := \arg\min_{\Psi^{(j)}} L_n(\Psi^{(j)}, \boldsymbol{\Xi}, p^{(j)}) \, .$$

We assume that $\{\boldsymbol{y}_t\}_{t=1,\ldots,n}$ is a realization of Model (2.1) with true parameter $(\boldsymbol{\Xi}_0, \boldsymbol{\Psi}_0, \boldsymbol{p}_0)$ and denote the MDL estimators as $(\hat{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{\Psi}}_n, \hat{\boldsymbol{p}}_n)$. Correspondingly, the true and estimated regimes are $R_0^{(1)}, R_0^{(2)}, \ldots, R_0^{(r_0)}$ and $\hat{R}_n^{(1)}, \hat{R}_n^{(2)}, \ldots, \hat{R}_n^{(\hat{r}_n)}$. By definition, we can write:

$$(\hat{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) \equiv \left(\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nm}, \hat{\rho}_{n1}, \ldots, \hat{\rho}_{nm}, \hat{q}, \hat{p}_n^{(1)}, \ldots, \hat{p}_n^{(r)} \right) := \arg\min_{\boldsymbol{\Xi}, \boldsymbol{p}} \mathrm{MDL}(\boldsymbol{\Xi}, \boldsymbol{p}) \, ,$$

$$\hat{\boldsymbol{\Psi}}_n \equiv \left(\hat{\Psi}_n^{(1)}, \ldots, \hat{\Psi}_n^{(r)} \right) := \left(\hat{\Psi}^{(1)}(\hat{\boldsymbol{\Xi}}_n, \hat{p}_n^{(1)}), \ldots, \hat{\Psi}^{(r)}(\hat{\boldsymbol{\Xi}}_n, \hat{p}_n^{(r)}) \right) \, .$$

Further, we define

$$L(\Psi, R, p) := \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left\{ L_n(\Psi, R, p) \right\} \equiv \mathbb{E} \left\{ \left(k \log 2\pi + \log |\Sigma| + \hat{\boldsymbol{a}}_1'(\boldsymbol{\Phi}) \Sigma^{-1} \hat{\boldsymbol{a}}_1(\boldsymbol{\Phi}) \right) I(\boldsymbol{z}_1 \in R) \right\} \ ,$$

which can be regarded as the population version of log-likelihood function. For the development of the theoretical results, we make the following assumptions on the model.

Assumption 4.1 *For Model (2.1), we assume the following conditions for the model parameters and their corresponding parameter spaces:*

(a) There are no redundant threshold lines. That is, there does not exist an $m' < m$, such that some $\mathcal{F}(\boldsymbol{\theta}', \boldsymbol{\rho}', q', \boldsymbol{p}', \boldsymbol{\Psi}') \equiv \mathcal{F}(\boldsymbol{\theta}, \boldsymbol{\rho}, q, \boldsymbol{p}, \boldsymbol{\Psi})$ with $\boldsymbol{\theta}, \boldsymbol{\rho} \in \mathbb{R}^{m'}$;

(b) The number of threshold line m is bounded by some integer $\bar{m} > 0$. Consequently, the number of regimes is bounded above by some integer $\bar{r} = \bar{r}(\bar{m}) > 0$. Also, the AR order within j-th regime $p^{(j)}$ is bounded above by $\bar{p}$;

(c) For any m, the parameter space for $\boldsymbol{\theta}$ is $\Theta = [0, 2\pi)^m$ equipped with product metric $d_\Theta(\boldsymbol{\theta}, \boldsymbol{\vartheta}) = \sqrt{\sum_{l=1}^m |e^{i\theta_l} - e^{i\vartheta_l}|^2}$; and the parameter space for $\boldsymbol{\rho}$ is $P = [0, \infty]^m$ equipped with product metric $d_P(\boldsymbol{\rho}, \boldsymbol{\varrho}) = \sqrt{\sum_{l=1}^m |\arctan \rho_l - \arctan \varrho_l|^2}$. Within j-th regime with $p^{(j)}$ given, the parameter space $\Omega^{(j)}$ for model parameters $\Psi^{(j)}$ is convex and compact;

(d) The VTARBT model is discountinous, in the sense that for any $R^{(j)}$ and $R^{(j')}$ shares a border (i.e. the intersection of their closure $\overline{R^{(j)}}$ and $\overline{R^{(j')}}$ is nonempty), there exists some $(\boldsymbol{y}_{t-1}^T, \ldots, \boldsymbol{y}_{t-\max(p^{(j)}, p^{(j')})}^T)^T$ such that

$$\Phi_0^{(j)} + \sum_{\tau=1}^{p^{(j)}} \Phi_\tau^{(j)} \boldsymbol{y}_{t-\tau} \neq \Phi_0^{(j')} + \sum_{\tau=1}^{p^{(j')}} \Phi_\tau^{(j')} \boldsymbol{y}_{t-\tau}\,.$$

Remark By allowing merging, it is possible that a VTARBT model can be represented by different sets of parameters. Condition 4.1(a) makes sure that the representation in Model (2.1) is the most succinct one, hence unique. Condition 4.1(b) and 4.1(c) define the parameter space for estimation. In particular, metrics are chosen to make the underlying spaces compact. Condition 4.1(d) excludes the family of continous threshold models, see Chan and Tsay (1998) for estimations of univariate models in this family.

Assumption 4.2 *We assume the following conditions for the time series described by Model (2.1):*

(a) For each $j = 1, \ldots, r$*, the error term* $\boldsymbol{a}_t^{(j)}$ *is absolutely continuous with respect to Lebesgue measure and its density is uniformly continuous, positive and bounded above on* $\mathbb{R}^k$*. Furthermore,* $\mathbb{E}|\boldsymbol{a}_t^{(j)}|^4 < \infty$*;*

(b) There exists a $\tilde{p} \in \mathbb{N}$ *such that* $\{Z_t\}_{t \geq \tilde{p}}$*, with* $Z_t = (\boldsymbol{z}_t^T, \boldsymbol{z}_{t-1}^T, \ldots, \boldsymbol{z}_{t-\tilde{p}+1}^T)^T$*, is a Markov chain. Furthermore, define* $\pi(R) := \mathbb{P}(\boldsymbol{z}_1 \in R)$ *for any* $R \subset \mathbb{R}^2$*;*

(c) With $p^* := \max(\bar{p}, \tilde{p})$*, the time series* $\{\boldsymbol{y}_t\}_{t \geq p^*}$ *is stationary and ergodic with* $\mathbb{E}|\boldsymbol{y}_t|^4 < \infty$*;*

(d) The Markov chain $\{X_t\}_{t \geq p^*}$*, with* $X_t = (\boldsymbol{y}_t^T, \boldsymbol{y}_{t-1}^T, \ldots, \boldsymbol{y}_{t-p^*+1}^T, \boldsymbol{z}_t^T, \boldsymbol{z}_{t-1}^T, \ldots, \boldsymbol{z}_{t-p^*+1}^T)^T$ *for* $t \geq p^*$*, admits a unique invariant measure* $\Pi(\cdot)$ *such that* $\exists\, K > 0$*,* $\rho < 1$*,* $\forall\, \boldsymbol{x} \in \mathbb{R}^{(2+k)p^*}$*,* $\forall\, d \in \mathbb{N}$*,*

$$\|\mathbb{P}^d(\boldsymbol{x}, \cdot) - \Pi(\cdot)\| \leq K(1 + |\boldsymbol{x}|)\rho^d$$

where $\mathbb{P}^d(\boldsymbol{x}, A)$ *is the d-step transition probability for any Borel set* $A \subset \mathbb{R}^{(2+k)p^*}$*,* $\|\cdot\|$ *and* $|\cdot|$ *are total variation norm and Euclidean norm, respectively.*

Remark Conditions 4.2(a) and 4.2(c) are distributional conditions commonly used in the threshold model literature, see Chan (1993), Li and Ling (2012) and Yau et al. (2015) for their univariate versions. Condition 4.2(b) holds when the threshold variable $\{\boldsymbol{z}_t\}$ is an autoregressive process. Under Condition 4.2(d), $\{X_t\}$ is V-uniform ergodic with $V(\boldsymbol{x}) = K(1 + |\boldsymbol{x}|)$.

4.2 Main Results

Proposition 4.1 *Under Assumption 4.1 and 4.2,*

(a) The number of estimated regimes will not be underestimated almost surely. That is, $\mathbb{P}(\hat{r}_n \geq r_0) = 1$ *for sufficiently large* n*. Furthermore,*

$$\max_{j=1,\ldots,r_0} \min_{\ell=1,\ldots,\hat{r}_n} \pi(R^{(\ell)}(\hat{\boldsymbol{\Xi}}_n) \backslash R_0^{(j)}) \xrightarrow{a.s} 0\,. \tag{4.2}$$

(b) The number of threshold lines will not be underestimated almost surely. That is, the dimension $\hat{m}_n$ *of* $\hat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ *(or* $\hat{\boldsymbol{\rho}}_{\boldsymbol{n}}$*) , is greater than or equal to the dimension* m_0 *of* $\boldsymbol{\theta_0}$ *(or* $\boldsymbol{\rho_0}$*) with probability 1 for sufficiently large* n*. Furthermore,*

$$\max_{i=1,\ldots,m_0} \min_{i'=1,\ldots,\hat{m}_n} (\hat{\theta}_{ni'} - \theta_{0i}, \hat{\rho}_{ni'} - \rho_{0i}) \xrightarrow{a.s} (0,0)\,. \tag{4.3}$$

(c) If the ℓ*-th estimated regime is nested in the* j*-th true regime, then* $\hat{p}_n^{(\ell)} \geq p_0^{(j)}$ *almost surely for suffciently large* n*.*

Proposition 4.2 *Under Assumption 4.1 and 4.2,*

$$\max_{j=1,\ldots,r_0} \min_{\ell=1,\ldots,\hat{r}_n} \pi(R^{(\ell)}(\hat{\boldsymbol{\Xi}}_n) \backslash R_0^{(j)}) = O_p(n^{-1})\,, \tag{4.4}$$

$$\max_{i=1,\ldots,m_0} \min_{i'=1,\ldots,\hat{m}_n} |\hat{\theta}_{ni'} - \theta_{0i}| = O_p(n^{-1})\,, \tag{4.5}$$

$$\max_{i=1,\ldots,m_0} \min_{i'=1,\ldots,\hat{m}_n} |\hat{\rho}_{ni'} - \rho_{0i}| = O_p(n^{-1})\,. \tag{4.6}$$

Remark Proposition 4.1 ensures that each true regime can be well approximated by one or more estimated regimes. Proposition 4.2 is useful in establishing the n-consistency of the threshold line estimators.

Theorem 4.1 *Let $\{\boldsymbol{y}_t\}_{t\geq 1}$ be a k-dimensional time series following Model (2.1) with true parameters $(\boldsymbol{\Psi}_0, \boldsymbol{\Xi}_0, \boldsymbol{p}_0)$. Under Assumption 4.1 and 4.2,*

(a) The number of total regimes will be corrected estimated with probability approaching 1, for sufficiently large n. Furthermore, for each $j = 1, \ldots, r_0$, there exists an $\ell = 1, \ldots, \hat{r}_n$ such that

$$\pi(R^{(\ell)}(\hat{\boldsymbol{\Xi}}_n) \backslash R_0^{(j)}) \xrightarrow{p} 0 \quad \textit{and} \quad \pi(R_0^{(j)} \backslash R^{(\ell)}(\hat{\boldsymbol{\Xi}}_n)) \xrightarrow{p} 0\,. \tag{4.7}$$

(b) The MDL estimators $(\hat{\boldsymbol{\Psi}}_n, \hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\rho}}_n, \hat{\boldsymbol{p}}_n)$ satisfy

$$(\hat{\boldsymbol{\Psi}}_n, \hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\rho}}_n, \hat{\boldsymbol{p}}_n) \xrightarrow{p} (\boldsymbol{\Psi}_0, \boldsymbol{\theta}_0, \boldsymbol{\rho}_0, \boldsymbol{p}_0)\,. \tag{4.8}$$

(c) The following results regarding convergence rates hold:

$$\pi(R^{(\ell)}(\hat{\boldsymbol{\Xi}}_n) \backslash R_0^{(j)}) = O_p(n^{-1})\,, \tag{4.9}$$

$$\pi(R_0^{(j)} \backslash R^{(\ell)}(\hat{\boldsymbol{\Xi}}_n)) = O_p(n^{-1})\,, \tag{4.10}$$

$$d_\Theta(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_p(n^{-1})\,, \tag{4.11}$$

$$d_P(\hat{\boldsymbol{\rho}}_n, \boldsymbol{\rho}_0) = O_p(n^{-1})\,, \tag{4.12}$$

$$\|\hat{\boldsymbol{\Psi}}_n - \boldsymbol{\Psi}_0\| = O_p(n^{-1/2})\,. \tag{4.13}$$

4.3 Proofs and Technical Lemmas

Lemma 4.1 (Information Inequality) *For $1 \leq j \leq r(\boldsymbol{\Xi})$ and $R \subset R_0^{(j)} \equiv R^{(j)}(\boldsymbol{\Xi}_0) \subset \mathbb{R}^2$, then*

$$L(\Psi, R, p) \geq L(\Psi_0^{(j)}, R_0^{(j)}, p_0^{(j)})\,, \tag{4.14}$$

for any $\Psi \in \Omega^{(j)}$ *and* $p = 1, \ldots, \bar{p}$. *Furthermore, the equatlity holds only when* $p = p_{0j} + s$ *for some non-negative integer* s *and* $\Psi^{(j)} = (\Psi_0^{(j)}, \mathbf{0}_{k\times(ks)})$.

Lemma 4.1 is a standard result in information theory, hence we omit the proof. See Cover and Thomas (2006) for details. ∎

Lemma 4.2 *Under Assumption 4.1, for fixed* $j = 1, \ldots, r_0$, $p^{(j)} = 1, \ldots, P$:

$$\sup_{\{\boldsymbol{\Xi}_n\}_{n\geq 1}\in\mathscr{N}_{j,\ell}(\boldsymbol{\Xi}_0)} \sup_{\Psi^{(j)}\in\Omega^{(j)}} \left|\frac{1}{n}L_n(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p^{(j)}) - L(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p^{(j)})\right| \xrightarrow{a.s.} 0\,, \tag{4.15}$$

$$\sup_{\{\boldsymbol{\Xi}_n\}_{n\geq 1}\in\mathscr{N}_{j,\ell}(\boldsymbol{\Xi}_0)} \sup_{\Psi^{(j)}\in\Omega^{(j)}} \left|\frac{1}{n}L'_n(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p^{(j)}) - L'(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p^{(j)})\right| \xrightarrow{a.s.} 0\,, \tag{4.16}$$

$$\sup_{\{\boldsymbol{\Xi}_n\}_{n\geq 1}\in\mathscr{N}_{j,\ell}(\boldsymbol{\Xi}_0)} \sup_{\Psi^{(j)}\in\Omega^{(j)}} \left|\frac{1}{n}L''_n(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p^{(j)}) - L''(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p^{(j)})\right| \xrightarrow{a.s.} 0\,, \tag{4.17}$$

where $\mathscr{N}_{j,\ell}(\boldsymbol{\Xi}_0) := \left\{\{\boldsymbol{\Xi}_n\}_{n\geq 1} | \limsup_{n\to\infty} R^{(\ell)}(\boldsymbol{\Xi}_n) \in R^{(j)}(\boldsymbol{\Xi}_0)\right\}$.

Proof For notational simplicity, we omit the argument $p^{(j)}$ as they are fixed. We will show that (4.15) holds, and (4.16) and (4.17) follow similarly. Define

$$\mathring{\mathscr{N}}_{j,\ell}(\boldsymbol{\Xi}_0) := \left\{\{\boldsymbol{\Xi}_n\}_{n\geq 1} | R^{(\ell)}(\boldsymbol{\Xi}_n) \subset R^{(j)}(\boldsymbol{\Xi}_0) \text{ for suffciently large } n\right\}\,,$$
$$\mathring{\mathscr{Q}}_{j,\ell}(\boldsymbol{\Xi}_0) := \left\{\{\boldsymbol{\Xi}_n\}_{n\geq 1} \in \mathring{\mathscr{N}}_{j,\ell}(\boldsymbol{\Xi}_0) | \boldsymbol{\theta}_n, \boldsymbol{\rho}_n \in \mathbb{Q}^m, 1 \leq m \leq M\right\}\,.$$

For any fixed $\{\boldsymbol{\Xi}_n\}_{n\geq 1}$ in $\mathring{\mathscr{N}}(\boldsymbol{\Xi}_0)$ and $\Psi^{(j)}$ in $\Omega^{(j)}$, by ergodic thoerem we have

$$\left|\frac{1}{n}L_n(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n)) - L(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n))\right| \xrightarrow{a.s.} 0\,. \tag{4.18}$$

By convexity and compactness of $\Omega^{(j)}$ and the equicontinuity of $\{\frac{1}{n}L_n(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n))\}_{n\geq 1}$

with respect to $\Psi^{(j)}$, it is routine to show that

$$\sup_{\Psi^{(j)}\in\Omega^{(j)}}\left|\frac{1}{n}L_n(\Psi^{(j)},R^{(\ell)}(\boldsymbol{\Xi}_n))-L(\Psi^{(j)},R^{(\ell)}(\boldsymbol{\Xi}_n))\right|\xrightarrow{a.s.}0\,. \tag{4.19}$$

In particular, for any $\{\boldsymbol{\Xi}_n\}_{n\geq 1}\in\mathring{\mathscr{Q}}_{j,\ell}(\boldsymbol{\Xi}_0)$, (4.19) holds on a probability one set. By the countability of $\mathring{\mathscr{Q}}_{j,\ell}(\boldsymbol{\Xi}_0)$, we have

$$\sup_{\{\boldsymbol{\Xi}_n\}_{n\geq 1}\in\mathring{\mathscr{Q}}_{j,\ell}(\boldsymbol{\Xi}_0)}\sup_{\Psi^{(j)}\in\Omega^{(j)}}\left|\frac{1}{n}L_n(\Psi^{(j)},R^{(\ell)}(\boldsymbol{\Xi}_n))-L(\Psi^{(j)},R^{(\ell)}(\boldsymbol{\Xi}_n))\right|\xrightarrow{a.s.}0\,. \tag{4.20}$$

For any fixed $\{\boldsymbol{\Xi}_n\}_{n\geq 1}\in\mathring{\mathscr{N}}_{j,\ell}(\boldsymbol{\Xi}_0)$, there exists $\{\widetilde{\boldsymbol{\Xi}}_n\}_{n\geq 1}\in\mathring{\mathscr{Q}}_{j,\ell}(\boldsymbol{\Xi}_0)$ such that,

$$\begin{aligned}
&\sup_{\Psi^{(j)}\in\Omega^{(j)}}\left|\frac{1}{n}L_n(\Psi^{(j)},R^{(\ell)}(\boldsymbol{\Xi}_n))-\frac{1}{n}L_n(\Psi^{(j)},R^{(\ell)}(\widetilde{\boldsymbol{\Xi}}_n))\right|\\
\leq\;&\frac{1}{n}\sum_t\sup_{\Psi^{(j)}\in\Omega^{(j)}}\left|k\log 2\pi+\log\left|\Sigma^{(j)^{-1}}\right|+\hat{\boldsymbol{a}}_t^T\Sigma^{(j)^{-1}}\hat{\boldsymbol{a}}_t\right|\\
\xrightarrow{a.s.}\;&0\,,
\end{aligned}$$

where the summation is over $\{t|\boldsymbol{z}_t\in R^{(\ell)}(\boldsymbol{\Xi}_n)\backslash R^{(\ell)}(\widetilde{\boldsymbol{\Xi}}_n)\}\cup\{t|\boldsymbol{z}_t\in R^{(\ell)}(\widetilde{\boldsymbol{\Xi}}_n)\backslash R^{(\ell)}(\boldsymbol{\Xi}_n)\}$. Therefore, it is clear that

$$\sup_{\{\boldsymbol{\Xi}_n\}_{n\geq 1}\in\mathring{\mathscr{N}}_{j,\ell}(\boldsymbol{\Xi}_0)}\sup_{\Psi^{(j)}\in\Omega^{(j)}}\left|\frac{1}{n}L_n(\Psi^{(j)},R^{(\ell)}(\boldsymbol{\Xi}_n))-L(\Psi^{(j)},R^{(\ell)}(\boldsymbol{\Xi}_n))\right|\xrightarrow{a.s.}0\,. \tag{4.21}$$

Finally, for any $\{\boldsymbol{\Xi}_n\}_{n\geq 1}\in\mathscr{N}_{j,\ell}(\boldsymbol{\Xi}_0)$, $R^{(\ell)}(\boldsymbol{\Xi}_n)$ can be partitioned as the union of $R^{(\ell)}(\widetilde{\boldsymbol{\Xi}}_n)$ and $R^{(\ell)}(\boldsymbol{\Xi}_n)\backslash R^{(j)}(\boldsymbol{\Xi}_0)$ for some $\{\widetilde{\boldsymbol{\Xi}}_n\}_{n\geq 1}\in\mathring{\mathscr{N}}_{j,\ell}(\boldsymbol{\Xi}_0)$ with $R^{(\ell)}(\widetilde{\boldsymbol{\Xi}}_n)=R^{(\ell)}(\boldsymbol{\Xi}_n)\cap$

$R^{(j)}(\boldsymbol{\Xi}_0)$. Hence,

$$
\begin{aligned}
& \sup_{\{\boldsymbol{\Xi}_n\}_{n\geq 1}\in\mathscr{N}_{j,\ell}(\boldsymbol{\Xi}_0)} \sup_{\Psi^{(j)}\in\Omega^{(j)}} \left|\frac{1}{n}L_n(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n)) - L(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n))\right| \\
\leq & \sup_{\{\boldsymbol{\Xi}_n\}_{n\geq 1}\in\mathscr{N}^{\circ}_{j,\ell}(\boldsymbol{\Xi}_0)} \sup_{\Psi^{(j)}\in\Omega^{(j)}} \left|\frac{1}{n}L_n(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n)) - L(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n))\right| \\
& \quad + \sup_{\{\boldsymbol{\Xi}_n\}_{n\geq 1}\in\mathscr{N}_{j,\ell}(\boldsymbol{\Xi}_0)} \sup_{\Psi^{(j)}\in\Omega^{(j)}} \left|\frac{1}{n}L_n(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n)\backslash R^{(\ell)}(\boldsymbol{\Xi}_0)) - L(\Psi^{(j)}, R^{(\ell)}(\boldsymbol{\Xi}_n)\backslash R^{(\ell)}(\boldsymbol{\Xi}_0))\right|.
\end{aligned}
$$

By definition of $\mathscr{N}_{j,\ell}(\boldsymbol{\Xi}_0)$, the second term is neglible. Thus, the desired convergence (4.15) follows. ∎

Lemma 4.3 *Under Assumption 4.1, for any $j = 1, \ldots, r_0$, and $p = 1, \ldots, \bar{p}$,*

$$
\sup_{\{\boldsymbol{\Xi}_n\}_{n\geq 1}\in\mathscr{N}_{j,\ell}(\boldsymbol{\Xi}_0)} \left|\hat{\Psi}^{(\ell)}_n(\boldsymbol{\Xi}_{\boldsymbol{n}}, p) - \tilde{\Psi}^{(\ell)}(\boldsymbol{\Xi}_{\boldsymbol{n}}, p)\right| \xrightarrow{a.s.} 0\,, \tag{4.22}
$$

where $\hat{\Psi}^{(\ell)}_n$ and $\tilde{\Psi}^{(\ell)}$ are defined respectively by

$$
\hat{\Psi}^{(\ell)}_n \equiv \hat{\Psi}^{(\ell)}_n(\boldsymbol{\Xi}, p) = \operatorname*{arg\,min}_{\Psi^{(\ell)}\in\Omega^{(\ell)}} L_n(\Psi^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}), p)\,, \quad \textit{and} \tag{4.23}
$$

$$
\tilde{\Psi}^{(\ell)} \equiv \tilde{\Psi}^{(\ell)}(\boldsymbol{\Xi}, p) = \operatorname*{arg\,min}_{\Psi^{(\ell)}\in\Omega^{(\ell)}} L(\Psi^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}), p)\,. \tag{4.24}
$$

In particular, given $p = p^{(j)}_0 + s$ for some non-negative integer s, we have

$$
\tilde{\Psi}^{(\ell)} = (\Psi^{(j)}_0, \mathbf{0}_{k\times(ks)})\,.
$$

Proof By definition of $\hat{\Psi}_n^{(\ell)}$ and $\tilde{\Psi}^{(\ell)}$, we have

$$\begin{aligned}
0 \;&\le\; L(\hat{\Psi}_n^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p) - L(\tilde{\Psi}^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p) \\
&\le\; L(\hat{\Psi}_n^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p) - \frac{1}{n} L_n(\hat{\Psi}_n^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p) + \frac{1}{n} L_n(\tilde{\Psi}^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p) - L(\tilde{\Psi}^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p) \\
&\le\; \left| L(\hat{\Psi}_n^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p) - \frac{1}{n} L_n(\hat{\Psi}_n^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p) \right| + \left| \frac{1}{n} L_n(\tilde{\Psi}^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p) - L(\tilde{\Psi}^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p) \right| \\
&\xrightarrow{a.s.} 0 \,.
\end{aligned}$$

By Lemma 4.2, the convergence is uniform in $\{\boldsymbol{\Xi}_n\}_{n\ge 1} \in \mathscr{N}_j(\boldsymbol{\Xi}_0)$. Since L is convex function in Ψ, its minimizer is unique and thus (4.22) holds. The final assertion follows from Lemma 4.1. ∎

Lemma 4.4 *Let $m_0 = 1$, $\boldsymbol{\Xi}_0 = (\theta_0, \rho_0, 0)$. For any $\varepsilon > 0, \eta > 0$, there exist some $c > 0, \Delta \in (0,1)$ such that, for any n,*

(a) If $\boldsymbol{\Xi}_n = (\theta_n, \rho_0, 0)$, then

$$\mathbb{P}\left(\sup_{\theta_n \in [cn^{-1}, \Delta)} \left| \frac{\sum_t x_t^{(2)} \mathbb{1}\left(\boldsymbol{z}_t \in R^{(1)}(\boldsymbol{\Xi}_n) \backslash R^{(1)}(\boldsymbol{\Xi}_0)\right)}{n\pi\left(\boldsymbol{z}_t \in R^{(1)}(\boldsymbol{\Xi}_n) \backslash R^{(1)}(\boldsymbol{\Xi}_0)\right)} - \bar{x}^{(2)} \right| > \eta \right) < \varepsilon \,, \qquad (4.25)$$

$$\mathbb{P}\left(\sup_{\theta_n \in [cn^{-1}, \Delta)} \left| \frac{\sum_t x_t^{(1)} \mathbb{1}\left(\boldsymbol{z}_t \in R^{(2)}(\boldsymbol{\Xi}_n) \backslash R^{(2)}(\boldsymbol{\Xi}_0)\right)}{n\pi\left(\boldsymbol{z}_t \in R^{(2)}(\boldsymbol{\Xi}_n) \backslash R^{(2)}(\boldsymbol{\Xi}_0)\right)} - \bar{x}^{(1)} \right| > \eta \right) < \varepsilon \,, \qquad (4.26)$$

where $(x_t^{(j)}, \bar{x}^{(j)})$ can take values from any of $(1,1), (\boldsymbol{a}_t^{(j)}, \mathbf{0}), \left(\boldsymbol{a}_t^{(j)^T} M \boldsymbol{y}_{t-\tau}, 0\right)$ and $\left(\boldsymbol{a}_t^{(j)^T} M \boldsymbol{a}_t^{(j)} \Big/ tr(M\Sigma^{(j)}), 1\right)$, $\tau = 1, \ldots, p^{(j)}$ and M is any $k \times k$ matrix with finite 2-norm.

(b) If $\boldsymbol{\Xi}_n = (\theta_0, \rho_n, 0)$*, then*

$$\mathbb{P}\left(\sup_{\rho_n \in [cn^{-1}, \Delta)} \left| \frac{\sum_t x_t^{(2)} \mathbb{1}\left(\boldsymbol{z}_t \in R^{(1)}(\boldsymbol{\Xi}_n) \backslash R^{(1)}(\boldsymbol{\Xi}_0) \right)}{n\pi\left(\boldsymbol{z}_t \in R^{(1)}(\boldsymbol{\Xi}_n) \backslash R^{(1)}(\boldsymbol{\Xi}_0) \right)} - \bar{x}^{(2)} \right| > \eta \right) < \varepsilon\,, \tag{4.27}$$

$$\mathbb{P}\left(\sup_{\rho_n \in [cn^{-1}, \Delta)} \left| \frac{\sum_t x_t^{(1)} \mathbb{1}\left(\boldsymbol{z}_t \in R^{(2)}(\boldsymbol{\Xi}_n) \backslash R^{(2)}(\boldsymbol{\Xi}_0) \right)}{n\pi\left(\boldsymbol{z}_t \in R^{(2)}(\boldsymbol{\Xi}_n) \backslash R^{(2)}(\boldsymbol{\Xi}_0) \right)} - \bar{x}^{(1)} \right| > \eta \right) < \varepsilon\,, \tag{4.28}$$

where $(x_t^{(j)}, \bar{x}^{(j)})$ *can take values from any of* $(1,1), (\boldsymbol{a}_t^{(j)}, \mathbf{0}), \left({\boldsymbol{a}_t^{(j)}}^T M \boldsymbol{y}_{t-\tau}, 0 \right)$ *and* $\left({\boldsymbol{a}_t^{(j)}}^T M \boldsymbol{a}_t^{(j)} \Big/ tr(M\Sigma^{(j)}), 1 \right)$*,* $\tau = 1, \ldots, p^{(j)}$ *and* M *is any* $k \times k$ *matrix with finite* 2*-norm.*

Proof The detailed proof are similar to Claim 2 in Chan (1993) and Proposition 1 in Qian (1998), and hence are omitted.

4.3.1 Proof of Proposition 4.1

Proof We focus the discussion on the probability one set within which Lemmas 4.2 and 4.3 hold. By compactness of parameter spaces, there exists a subsequence $\{n_s\}$, along which $\hat{\boldsymbol{\Psi}}_{n_s} \to \boldsymbol{\Psi}_*, \hat{\boldsymbol{\Xi}}_{n_s} \to \boldsymbol{\Xi}_*, \hat{\boldsymbol{p}}_{n_s} \to \boldsymbol{p}_*$ for some $\boldsymbol{\Psi}_*, \boldsymbol{\Xi}_*, \boldsymbol{p}_*$. For convenience, we write n instead of n_s throughout the proof for this particular subsequence. Under any of the following situation:

- The number of estimated regimes is underestimated;
- The number of threshold lines is underestimated;
- Either (4.2) or (4.3) does not hold;

there must be some estimated regime(s) intercepting with at least two of the true regimes. Specifically, we define for each $l = 1, \ldots, r_*$:

$$\mathcal{P}_l := \left\{ R_*^{(l)} \cap R_0^{(j)} : j = 1, \ldots, r_0 \right\} \equiv \{ P_{lw} : w = 1, \ldots, |\mathcal{P}_l| \}\,, \tag{4.29}$$

then $|\mathcal{P}_l| \geq 1$ for each l and at least one of them is strictly larger than 1. By Lemmas 4.2 and 4.3,

$$\frac{1}{n}\sum_{\ell=1}^{\hat{r}_n} L_n\left(\hat{\Psi}_n^{(\ell)}, R^{(\ell)}(\hat{\boldsymbol{\Xi}}_n), \hat{p}_{n\ell}\right) \to \sum_{l=1}^{r_*} L\left(\Psi_*^{(l)}, R^{(l)}(\boldsymbol{\Xi}_*), p_{*l}\right) \tag{4.30}$$

$$= \sum_{l=1}^{r_*}\sum_{w=1}^{|\mathcal{P}_l|} L\left(\Psi_*^{(l)}, P_{lw}, p_{*l}\right) \tag{4.31}$$

$$> \sum_{j=1}^{r_0} L\left(\Psi_0^{(j)}, R^{(j)}(\boldsymbol{\Xi}_0), p_{0j}\right) . \tag{4.32}$$

The last inequality is due to Lemma 4.1 and the fact $\Psi_*^{(l)}$ cannot equal to the true model parameters for all $|\mathcal{P}_l|$ regimes simultaneuously. Consequently,

$$\begin{aligned}
\frac{2}{n}\mathrm{MDL}(\hat{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) &= \frac{2}{n}\mathrm{Pen}(\hat{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) + \frac{1}{n}\sum_{\ell=1}^{\hat{r}_n} L_n\left(\hat{\Psi}_n^{(\ell)}, R_\ell(\hat{\boldsymbol{\Xi}}_n), \hat{p}_n^{(\ell)}\right) \\
&> \frac{2}{n}\mathrm{Pen}(\hat{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) + \sum_{j=1}^{r_0} L\left(\Psi_0^{(j)}, R_j(\boldsymbol{\Xi}_0), p_0^{(j)}\right) \\
&= \frac{2}{n}\mathrm{MDL}(\boldsymbol{\theta_0}, \boldsymbol{\rho_0}, \boldsymbol{p_0}) + o_p(1) .
\end{aligned} \tag{4.33}$$

This leads to an obvious contradiction to the definition of $\hat{\boldsymbol{\Xi}}_n$ and $\hat{\boldsymbol{p}}_n$. Therefore, part (a) and (b) of the theorem must hold. By assuming $p_* < p_0^{(j)}$ and applying Lemma 4.1 to estimated regime $\hat{R}^{(l)}(\boldsymbol{\Xi}_n)$ which is nested in $R^{(j)}(\boldsymbol{\Xi}_0)$, one can obtain contradiction and prove (c) in a similar manner. ∎

4.3.2 Proof of Proposition 4.2

Proof Equation (4.4) follow directly from (4.5) and (4.6), thus it suffices to establish (4.5) and (4.6). For any $i = 1, \ldots, m_0$, by Proposition 4.1, there exists an $i' \in \{1, \ldots, \hat{r}_n\}$ such that both $|\hat{\theta}_{ni'} - \theta_{0i}|$ and $|\hat{\rho}_{ni'} - \rho_{0i}|$ converge to 0 almost surely. For simple illustration, we assume that $m_0 = 1$. The generalization of the proof to un-

known m_0 is similar to that from Chan (1993) to Li and Ling (2012). We first assume that the MDL procedure specifies one threshold line $(\hat{\theta}_n, \hat{\rho}_n)$. It suffices to show that for some $c > 0$,

$$\mathbb{P}\left(|\hat{\theta}_n - \theta_0| > cn^{-1}\right) \to 0\,, \tag{4.34}$$

$$\mathbb{P}\left(|\hat{\rho}_n - \rho_0| > cn^{-1}\right) \to 0\,, \tag{4.35}$$

as $n \to \infty$.

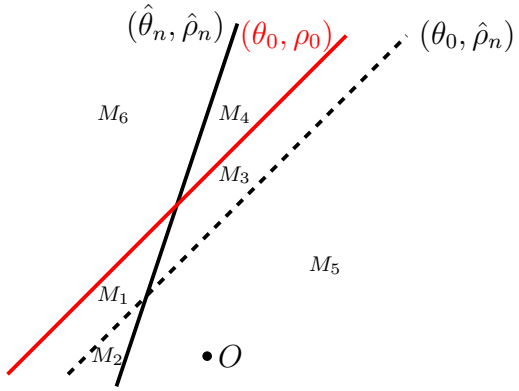


Figure 4.1: *Graphical Illustration for Misclassifications*

Figure 4.1 depicts the general relationship between the true threshold line (θ_0, ρ_0) and the estimated one $(\hat{\theta}_n, \hat{\rho}_n)$. Using the notation in Figure 4.1, we have two true regimes, namely $R_0^{(1)} = M_1 \cup M_2 \cup M_3 \cup M_5$ and $R_0^{(2)} = M_4 \cup M_6$. Denote $\hat{\boldsymbol{\Xi}}_n := (\hat{\theta}_n, \hat{\rho}_n, 0)$, $\tilde{\boldsymbol{\Xi}}_n := (\theta_0, \rho_0, 0) = \boldsymbol{\Xi}_0$. Apparently, once $\hat{\boldsymbol{\Xi}}_n$ is estimated, it is easy to get $\hat{\boldsymbol{\Psi}}_n = (\hat{\Psi}_n^{(1)}, \hat{\Psi}_n^{(2)})$. Notice that

$$
\begin{aligned}
& \mathrm{MDL}(\hat{\boldsymbol{\Psi}}_n, \tilde{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) - \mathrm{MDL}(\hat{\boldsymbol{\Psi}}_n, \hat{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) \\
= \quad & \sum_t \left(l_t(\hat{\Psi}_n^{(2)}) - l_t(\hat{\Psi}_n^{(1)}) \right) \left[\mathbb{1}\left(\boldsymbol{z}_t \in M_4\right) - \mathbb{1}\left(\boldsymbol{z}_t \in M_1 \cup M_2\right) \right] \\
= \quad & \left\{ \sum_t \left[\log \left| \left(\Sigma_0^{(1)}\right)^{-1} \Sigma_0^{(2)} \right| + \mathrm{tr} \left(\left(\left(\Sigma_0^{(2)}\right)^{-1} - \left(\Sigma_0^{(1)}\right)^{-1} \right) \Sigma_0^{(2)} \right) \right. \right. \\
& \left. \left. - \ \Delta_{1,2}^T \left(\Sigma_0^{(1)}\right)^{-1} \Delta_{1,2} \right] \cdot \mathbb{1}\left(\boldsymbol{z}_t \in M_4\right) \right\} (1 + o_p(1)) \\
& + \left\{ \sum_t \left[\log \left| \left(\Sigma_0^{(2)}\right)^{-1} \Sigma_0^{(1)} \right| + \mathrm{tr} \left(\left(\left(\Sigma_0^{(1)}\right)^{-1} - \left(\Sigma_0^{(2)}\right)^{-1} \right) \Sigma_0^{(1)} \right) \right. \right. \\
& \left. \left. - \ \Delta_{2,1}^T \left(\Sigma_0^{(2)}\right)^{-1} \Delta_{2,1} \right] \cdot \mathbb{1}\left(\boldsymbol{z}_t \in M_1 \cup M_2\right) \right\} (1 + o_p(1)) \ , \qquad (4.36)
\end{aligned}
$$

where

$$
\begin{aligned}
l_t(\Psi) &:= \log|\Sigma| + \boldsymbol{a}_t(\Phi)'\Sigma^{-1}\boldsymbol{a}_t(\Phi) \quad \text{and} \\
\Delta_{i,i'} &:= \left(\Phi_0^{(i')} - \Phi_0^{(i)}\right) + \sum_{\tau=1}^{p} \left(\Phi_\tau^{(i')} - \Phi_\tau^{(i)}\right) \boldsymbol{y}_{t-\tau} \, .
\end{aligned}
$$

The last equality in (4.36) is due to Lemmas 4.3, 4.4 and the fact

$$
\mathbb{1}\left(\boldsymbol{z}_t \in M_4\right) - \mathbb{1}\left(\boldsymbol{z}_t \in M_1 \cup M_2\right) = \mathbb{1}\left(\boldsymbol{z}_t \in M_3 \cup M_4\right) - \mathbb{1}\left(\boldsymbol{z}_t \in M_2\right) - \mathbb{1}\left(\boldsymbol{z}_t \in M_1 \cup M_3\right) \, ,
$$

given $|\hat{\theta}_n - \theta_0| > cn^{-1}$, $|\hat{\rho}_n - \rho_0| > cn^{-1}$. Notice that Lemma 4.4 applies here since all $M_3 \cup M_4$, M_2 and $M_1 \cup M_3$ are induced by the small perturbation in one of θ and ρ parameter, while keeping the other unchanged. It is clear that both quantities in (4.36) are negative with probability approaching 1. Therefore, we have

$$
\begin{aligned}
\mathbb{P}\left(\mathrm{MDL}(\hat{\boldsymbol{\Psi}}_n, \tilde{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) - \mathrm{MDL}(\hat{\boldsymbol{\Psi}}_n, \hat{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) > 0, |\hat{\theta}_n - \theta_0| > cn^{-1} \right) \to 0 \, , \\
\mathbb{P}\left(\mathrm{MDL}(\hat{\boldsymbol{\Psi}}_n, \tilde{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) - \mathrm{MDL}(\hat{\boldsymbol{\Psi}}_n, \hat{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) > 0, |\hat{\rho}_n - \rho_0| > cn^{-1} \right) \to 0 \, ,
\end{aligned}
$$

By the definition of MDL, $\text{MDL}(\hat{\boldsymbol{\Psi}}_n, \tilde{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) \geq \text{MDL}(\hat{\boldsymbol{\Psi}}_n, \hat{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n)$. Hence, (4.34) and (4.35) must hold.

Next, we consider the cases where MDL procedure specifies more than m_0 threshold line, $\hat{\boldsymbol{\Xi}}_n = (\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\rho}}_n, q)$ with $m_0 > 1$. Without loss of generality, we assume $|\hat{\theta}_{n1} - \theta_0| \xrightarrow{a.s.} 0$ and $|\hat{\rho}_{n1} - \rho_0| \xrightarrow{a.s.} 0$. Then, define $\tilde{\boldsymbol{\Xi}}_n = (\tilde{\boldsymbol{\theta}}_n, \tilde{\boldsymbol{\rho}}_n, q)$ with $\tilde{\boldsymbol{\theta}}_n = (\theta_{01}, \hat{\theta}_{n2}, \ldots, \hat{\theta}_{n,m_0})^T$ and $\tilde{\boldsymbol{\rho}}_n = (\rho_{01}, \hat{\rho}_{n2}, \ldots, \hat{\rho}_{n,m_0})^T$. Use similar arguments as in (4.36) will lead to the fact that $\text{MDL}(\hat{\boldsymbol{\Psi}}_n, \tilde{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) - \text{MDL}(\hat{\boldsymbol{\Psi}}_n, \hat{\boldsymbol{\Xi}}_n, \hat{\boldsymbol{p}}_n) < 0$ with probability approaching 1. The remaining is identical to the case $m_0 = 1$. ∎

4.3.3 Proof of Theorem 4.1

Proof The n-consistency of threshold variable estimators and the $n^{-1/2}$-consistency for common maximum likelihood estimators for VAR parameters ensure that (4.13) follows directly from (4.9) to (4.12). With Proposition 4.1 and 4.2, it suffices to show that the number of regimes will not be overestimated with probablity approaching to 1 as $n \to \infty$. Equivalently, we can establish the result by searching for a set of $(\boldsymbol{\Xi}_n^\dagger, \boldsymbol{p}_n^\dagger)$ with $r(\boldsymbol{\Xi}_n^\dagger) = r(\boldsymbol{\Xi}_0) \equiv r_0$ and

$$\mathbb{P}\left\{ \min_{\{\boldsymbol{\Xi}_n, \boldsymbol{p}_n\} \in A} \text{MDL}(\boldsymbol{\Xi}_n, \boldsymbol{p}_n) > \text{MDL}(\boldsymbol{\Xi}_n^\dagger, \boldsymbol{p}_n^\dagger) \right\} \to 1\,,$$

as $n \to \infty$, where A is the collection of $\{\boldsymbol{\Xi}_n, \boldsymbol{p}_n\}$ satisfying Proposition 4.1 and $r_n \equiv r(\boldsymbol{\Xi}_n) > r_0$ for all large n. In particular, let $\boldsymbol{\Xi}_n^\dagger$ to be any sequence that converge to

$\boldsymbol{\Xi}_0$ with rate n^{-1} in probability; $\boldsymbol{p}_n^\dagger \equiv \boldsymbol{p}_0$ and $\boldsymbol{\Psi}_n^\dagger \equiv \boldsymbol{\Psi}_0$. Then, for sufficiently large n,

$$
\begin{aligned}
& \frac{2}{n}\mathrm{MDL}(\boldsymbol{\Xi}_n, \boldsymbol{p}_n) - \mathrm{MDL}(\boldsymbol{\Xi}_n^\dagger, \boldsymbol{p}_n^\dagger) \\
\geq\ & \frac{2}{n}\left[\mathrm{Pen}(\boldsymbol{\Xi}_n, \boldsymbol{p}_n) - \mathrm{Pen}(\boldsymbol{\Xi}_n^\dagger, \boldsymbol{p}_0)\right] \\
& \quad + \frac{1}{n}\left[\sum_{\ell=1}^{r_n} L_n(\hat{\Psi}_n^{(\ell)}(\boldsymbol{\Xi}_n, \boldsymbol{p}_n), R^{(\ell)}(\boldsymbol{\Xi}_n), p_n^{(\ell)}) - \sum_{j=1}^{r_0} L_n(\Psi_0^{(j)}, R^{(j)}(\boldsymbol{\Xi}_n^\dagger), p_0^{(j)})\right] \\
=\ & \frac{2}{n}\left[\mathrm{Pen}(\boldsymbol{\Xi}_n, \boldsymbol{p}_n) - \mathrm{Pen}(\boldsymbol{\Xi}_n^\dagger, \boldsymbol{p}_0)\right] && (4.37) \\
& \quad + \frac{1}{n}\left[\sum_{\ell=1}^{r_n} L_n(\tilde{\Psi}^{(\ell)}(\boldsymbol{\Xi}_n, \boldsymbol{p}_n), R^{(\ell)}(\boldsymbol{\Xi}_n), p_n^{(\ell)}) - \sum_{j=1}^{r_0} L_n(\Psi_0^{(j)}, R^{(j)}(\boldsymbol{\Xi}_n^\dagger), p_0^{(j)})\right] && (4.38) \\
& \quad - \frac{1}{n}\sum_{\ell=1}^{r_n} (\tilde{\Psi}^{(\ell)} - \hat{\Psi}_n^{(\ell)})^T L_n''(\Psi_+^{(\ell)}, R^{(\ell)}(\boldsymbol{\Xi}_n), p_n^{(\ell)})(\tilde{\Psi}^{(\ell)} - \hat{\Psi}_n^{(\ell)})\,. && (4.39)
\end{aligned}
$$

It is clear that for any $\{\boldsymbol{\Xi}_n, \boldsymbol{p}_n\}$ in A, quantity (4.37) is strictly positive and of order $O_p(\log n/n)$; quantity (4.38) is exact 0 since $\tilde{\Psi}^{(\ell)}$ and $\Psi_0^{(j)}$ specify the same model in $R^{(\ell)}(\boldsymbol{\Xi}_n) \subset R^{(j)}(\boldsymbol{\Xi}_0)$ due to Lemma 4.3; quantity (4.39) is of order $O_p(1/n)$ due to Proposition 4.2. Hence, the quantity (4.37) dominates the whole expression and the proof is completed. ∎

Chapter 5

Simulation Study

In this chapter, we evaluate the finite sample performance of the proposed MDL estimator through simulations. We consider the following bivariate model for $\{\boldsymbol{y}_t\}_{t\geq 1}$ with $\boldsymbol{y}_t = (y_{t,1}, y_{t,2})'$,

$$\begin{aligned}\boldsymbol{y}_t = &\left(\Phi_0^{(1)} + \Phi_1^{(1)}\boldsymbol{y}_{t-1} + \Phi_2^{(1)}\boldsymbol{y}_{t-2} + \boldsymbol{a}_t^{(1)}\right) \mathbb{1}\left(y_{t-1,1} + \sqrt{3}y_{t-1,2} \geq -2\sqrt{3}\,, -y_{t-1,1} + y_{t-1,2} < 2\sqrt{2}\right) \\ &+ \left(\Phi_0^{(2)} + \Phi_1^{(2)}\boldsymbol{y}_{t-1} + \boldsymbol{a}_t^{(2)}\right) \mathbb{1}\left(y_{t-1,1} + \sqrt{3}y_{t-1,2} \geq -2\sqrt{3}\,, -y_{t-1,1} + y_{t-1,2} \geq 2\sqrt{2}\right) \\ &+ \left(\Phi_0^{(3)} + \Phi_1^{(3)}\boldsymbol{y}_{t-1} + \boldsymbol{a}_t^{(3)}\right) \mathbb{1}\left(y_{t-1,1} + \sqrt{3}y_{t-1,2} < -2\sqrt{3}\,, -y_{t-1,1} + y_{t-1,2} \geq 2\sqrt{2}\right) \\ &+ \left(\Phi_0^{(3)} + \Phi_1^{(3)}\boldsymbol{y}_{t-1} + \boldsymbol{a}_t^{(3)}\right) \mathbb{1}\left(y_{t-1,1} + \sqrt{3}y_{t-1,2} < -2\sqrt{3}\,, -y_{t-1,1} + y_{t-1,2} < 2\sqrt{2}\right)\,,\end{aligned} \tag{5.1}$$

where $\boldsymbol{a}_t^{(j)} \overset{iid}{\sim} N(0, \Sigma^{(j)})$ for $j = 1, 2, 3, 4$ and

$$\Phi_0^{(1)} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \Phi_1^{(1)} = \begin{pmatrix} 0.4 & -0.4 \\ -0.7 & -0.7 \end{pmatrix}, \quad \Phi_2^{(1)} = \begin{pmatrix} 0.2 & 0.5 \\ -0.3 & -0.1 \end{pmatrix};$$
$$\Phi_0^{(2)} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \Phi_1^{(2)} = \begin{pmatrix} -0.4 & -0.6 \\ 0.2 & 0.6 \end{pmatrix};$$

$$\Phi_0^{(3)} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad \Phi_1^{(3)} = \begin{pmatrix} 0.5 & 0.4 \\ -0.8 & 0.3 \end{pmatrix};$$
$$\Phi_0^{(4)} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}, \quad \Phi_1^{(4)} = \begin{pmatrix} -0.5 & 0.3 \\ -0.1 & -0.3 \end{pmatrix}.$$
$$\Sigma_1 = \Sigma_3 = \boldsymbol{I_2}\,, \quad \Sigma_2 = \Sigma_4 = 4\boldsymbol{I_2}\,.$$

Clearly, (5.1) is a self-excited VTARBT model with $d = 1$. The four regimes are induced by two non-parallel threshold lines

$$l_1 : (\theta_1, \rho_1) = (\pi/6, 1)\,,$$
$$l_2 : (\theta_2, \rho_2) = (3\pi/4, 2)\,.$$

See Figure 5.1 for a realization of Model (5.1) with sample size $n = 400$.

For each realization, we can obtain one set of MDL estimators. By repeatly obtaining MDL estimator via GA for different sample sizes (e.g. $n = 400$ and 800) for 400 replications, we summarize the estimation performance in Table 5.1. Observe that the threshold lines l_1 and l_2 can be estimated with high accuracy and precision, i.e. estimators in Table 5.1 have small biases and standard errors. However, the regime specification with $n = 400$ is not entirely satisfied, since 91 out of 400 the replications merge subregions $\hat{R}_2$ and $\hat{R}_4$ (i.e. $\hat{q} = 6$). This deviates from the true model, Model (5.1), and identifies a three-regime model. This misspecification problem is alleviated significantly with $n = 800$, with only 1 out of 400 replications indicates the same false merge pattern (i.e. $\hat{q} = 6$). Based on all the replications that correctly specified the number of regime, the AR orders within each regime is well identified for most replications, i.e. $(\hat{p}^{(1)}, \hat{p}^{(2)}, \hat{p}^{(3)}, \hat{p}^{(4)})' = (2, 1, 1, 1)'$, see Table 5.2 for the relative frequencies of AR order specification. Finally, with all the regimes and AR orders accurately identified, the AR parameters within each regime are effectively estimated.

$\hat{m}$		$\hat{\boldsymbol{\theta}}$			$\hat{\boldsymbol{\rho}}$	
Value	%		$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$
$n = 400$						
≤ 1	0.0%	MSE	0.0004	0.0012	0.0056	0.0206
2	100.0%	S.D.	0.0192	0.0347	0.0746	0.1433
≥ 3	0.0%	Bias	0.0037	0.0013	0.0062	−0.0087
$n = 800$						
≤ 1	0%	MSE	0.0001	0.0003	0.0012	0.0053
2	100.0%	S.D.	0.0080	0.0177	0.0343	0.0726
≥ 3	0.0%	Bias	0.0028	−0.0009	0.0053	−0.0035

Table 5.1: *Summary of Threshold Line Estimators with* $n = 400$ *and* 800

$n = 400$	(309 valid replications)				
order	0	1	2	3	≥ 4
$\hat{p}^{(1)}$	0	0	99.7	0.3	0
$\hat{p}^{(2)}$	0	100	0	0	0
$\hat{p}^{(3)}$	0	100	0	0	0
$\hat{p}^{(4)}$	0	100	0	0	0
$n = 800$	(399 valid replications)				
order	0	1	2	3	≥ 4
$\hat{p}^{(1)}$	0	0	100	0	0
$\hat{p}^{(2)}$	0	100	0	0	0
$\hat{p}^{(3)}$	0	100	0	0	0
$\hat{p}^{(4)}$	0	99.7	0.3	0	0

Table 5.2: *Relative frequencies for AR Order Specification*

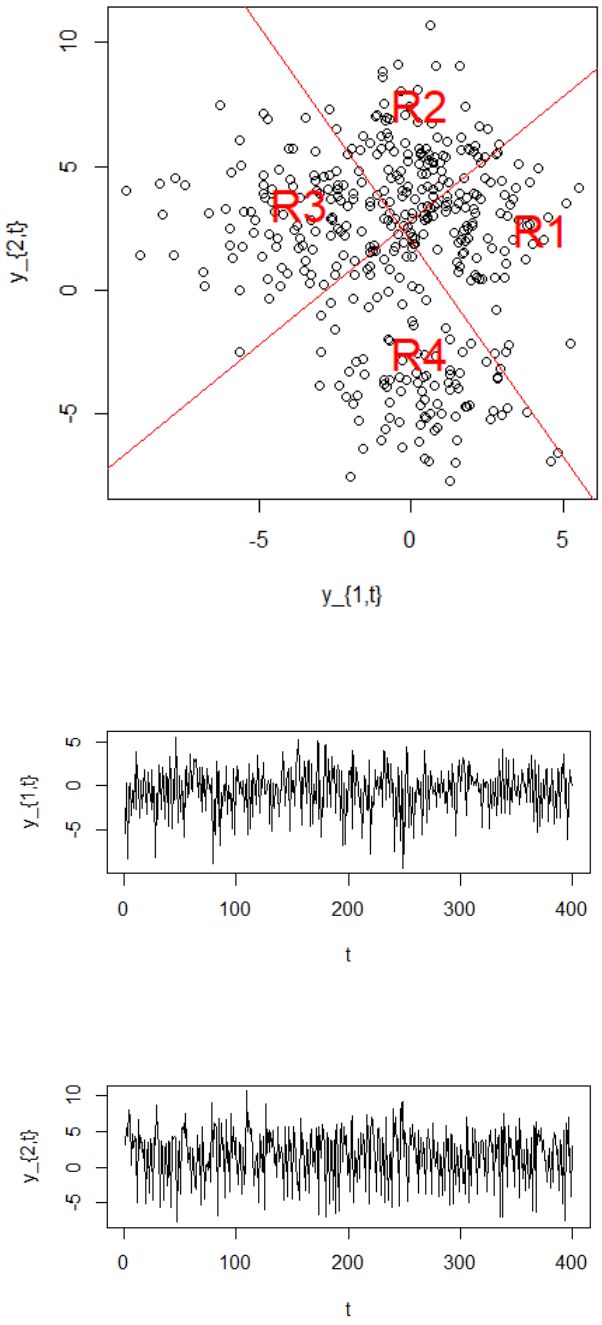


Figure 5.1: *A Realization Generated from the VTARBT Model (5.1)*

Chapter 6

Real Case Application

In Tsay (1998), an empirical study on US interest rates was conducted. The dataset consists of a bivarite time series $\{\boldsymbol{Y}_t\}_{t=1,\ldots,409}$ with $\boldsymbol{Y}_t = (Y_{1t}, Y_{2t})^T$, the monthly log yields of three-month treasury bill and three-year treasury note from 1959.1 to 1993.2. In particular, he modeled the growth series $\{\boldsymbol{y}_t\}_{t=2,\ldots,409}$ with $\boldsymbol{y}_t = \boldsymbol{Y}_t - \boldsymbol{Y}_{t-1}$ with TAR model with univariate threshold variable

$$z_t = \bar{Y}^{(3)}_{2,t-d} - \bar{Y}^{(3)}_{1,t-d} \ ,$$

for $t \geq 3$, where $d = 4$ and $\bar{Y}^{(3)}_{i,t} = (Y_{it} + Y_{i,t-1} + Y_{i,t-2})/3$. The threshold variable z_t is interpreted as a lag-d series of *interest rate spread*, the difference between intermediate term (three year) interest yield and short term (three month) yield. Interest rate spread is a useful index for predicting economic conditions. In general, a large positive z_t is an indicator for economic booming, a small positive one for normal cases, while a negative one for contraction. Correspondingly, a three-regime model is adopted with regimes $\{r_2 \leq z_t\}$, $\{r_1 \leq z_t < r_2\}$ and $\{z_t < r_1\}$. After roughly deciding the possible range for r_1 and r_2 by economic sense, he did grid search for optimal r_1 and r_2 to minimize AIC.

The aforementioned model can be improved in several ways. First, interest rate

spread is not the only option for the threshold variable. Although the interest rate spread has meaningful interpretation from economic judgment, the choice of threshold variable may also be rely on the data. Secondly, the number of thresholds as well as the approximated location are predetermined. Specifically, r_1 and r_2 are assumed to fall in the intervals $[-0.30, -0.20]$ and $[-0.15, 0.05]$, respectively. Our VTARBT model, on the other hand, provides more flexibility on threshold variable. We use bivariate threshold variable

$$\boldsymbol{z}_t = (\bar{Y}^{(3)}_{1,t-d}, \bar{Y}^{(3)}_{2,t-d})',$$

and do not prespecify the number of threshold lines. Under this setting, the model in Tsay (1998) is a special case with two parallel threshold lines, see Model (A) in Figure 6.1.

We consider three different models in the Section. Model (A) is the model used in Tsay (1998), with univariate threshold variable, obtained by minimizing AIC. Models (B) and (C) are VTARBT model with minimized AIC and MDL, respectively. The configurations of threshold lines, AR orders for each model and their AICs (MDLs) are summarized in Table 6.1. Apparently, Model (B) has the smallest AIC, while Model (C) has the smallest MDL among all three models. Meanwhile, Model (A) is dominated by both Models (B) and (C), which shows that superiority of VARTBT model over VTAR model with univariate threshold variables. By comparing Models (B) and (C), we find that using MDL as the model selection criterion yields a more parsimonious model, as it carries a larger penalty on model parameters.

Notably, the threshold lines in Model (A) are significantly different from those in the other two models. The superiority of Model (B) and (C) indicates that the interest rate spread is not the most suitable threshold variable. The overall level of interest rates plays a more significant role, according to the dataset provided. For example, the

three regimes in Model (B) are

$$R_1 = \{\bar{Y}^{(3)}_{1,t-d} + 0.9094\bar{Y}^{(3)}_{2,t-d} < 1.9433, \bar{Y}^{(3)}_{1,t-d} + 6.4338\bar{Y}^{(3)}_{2,t-d} < 17.0236\},$$
$$R_2 = \{\bar{Y}^{(3)}_{1,t-d} + 0.9094\bar{Y}^{(3)}_{2,t-d} \geq 1.9433, \bar{Y}^{(3)}_{1,t-d} + 6.4338\bar{Y}^{(3)}_{2,t-d} < 17.0236\} \text{ and}$$
$$R_3 = \{\bar{Y}^{(3)}_{1,t-d} + 0.9094\bar{Y}^{(3)}_{2,t-d} \geq 1.9433, \bar{Y}^{(3)}_{1,t-d} + 6.4338\bar{Y}^{(3)}_{2,t-d} \geq 17.0236\}.$$

We interpret R_1 as the low interest rate regime, R_2 as the normal or intermediate interest rate regime, and R_3 as the high interest rate regime. Indeed, the level of interest rates is closely related to the whole economy, since it has deep influence on the financial markets and other investments.

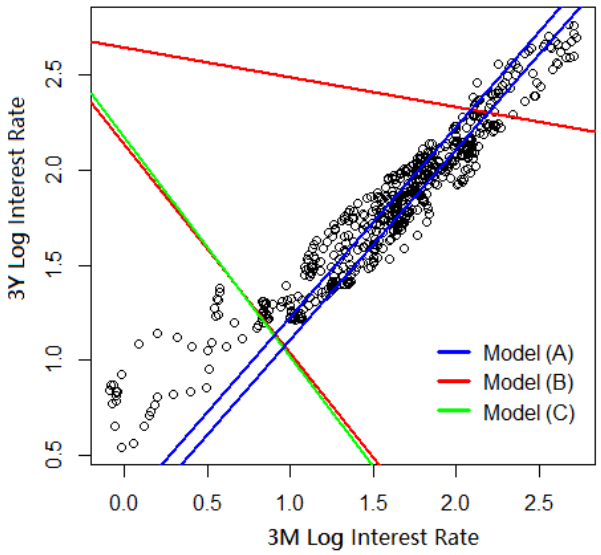


Figure 6.1: *Estimated Threshold Lines for Model (A), (B) and (C)*

	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	MDL	AIC
Model (A)*	2.3562	2.3562	0.1613	0.0735	7	6	2	-	-1556	-3651
Model (B)	0.7380	1.4166	1.4377	2.6146	7	3	1	-	-1681	-3813
Model (C)	0.7125	-	1.4212	-	1	2	-	-	-1740	-3701

Table 6.1: *Optimal Models (A), (B) and (C)*
*: Model (A) is directly from a reparametrization of the one in Tsay (1998).

Chapter 7

Conclusion

This thesis generalizes the TAR model by allowing multivariate time series with bivariate threshold variables. A method for model selection and parameter estimation through the MDL criterion is proposed. To minimize MDL function, GA with Island Model is adopted. Consistency of threshold line estimators is established with the n^{-1} convergence rate and effectiveness of the proposed method on finite samples has been studied through a stimulation study. Finally, the model is used to analyze US interest rates. For later research, possible directions include:

(1) The asymptotic marginal and joint distributions for threshold line estimators;

(2) Testing methods for VTARBT model against VTAR model with univarite threshold variables, or against ordinary VAR model;

(3) Robustness of the MDL estimators for slight misspecfication of the model;

(4) Extension of the model to allow nonlinear partition of the space for threshold variables.

Bibliography

Addo, P. M. (2014). Multivariate self-exciting threshold autoregressive models with exogenous input. *arXiv*, pages 1–11.

Alba, E. and Troya, J. M. (1999). A survey of parallel distributed genetic algorithms. *Complexity*, 4:31–52.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The Annals of Statistics*, 21:520–533.

Chan, K. S. and Tsay, R. S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, 85(2):413–426.

Chen, H. C., Chong, T. T.-L., and Bai, J. (2012). Theory and applications of tar model with two threshold variables. *Econometric Reviews*, 31:142–170.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. Wiley, New York.

Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101:223–239.

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

Hansen, M. H. and Yu, B. (2001). Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 96:746–774.

Lee, T. C. M. (2001). An introduction to coding theory and the two-part minimum description length principle. *International Statistical Review*, 69:169–183.

Lewis, P. A. W. and Stevens, J. G. (1991). Nonlinear modeling of time series using multivariate adaptive regression splines. *Journal of the American Statistical Association*, 86:864–877.

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

Lu, Q., Lund, R., and Lee, T. C. M. (2010). An MDL approach to the climate segmentation problem. *The Annals of Applied Statistics*, 4:299–319.

Qian, L. (1998). On maximum likelihood estimators for a threshold autoregression. *Journal of Statistical Planning and Inference*, 75:21–46.

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

Tiao, G. C. and Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time-series. *Journal of Forecasting*, 13:109–131.

Tong, H. (1978). On a threshold model. In Chen, C., editor, *Pattern Recognition and Signal Processing. NATO ASI Series E: Applied Sc. (29).*, pages 575–586. Oxford University Press.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. Springer, New York.

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

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

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.