

On buffered double autoregressive time series models

Zhao Liu

Department of Economics, Duke University*

Abstract

A buffered double autoregressive (BDAR) time series model is proposed in this paper to depict the buffering phenomenon of conditional mean and conditional variance in time series. To build this model, a novel flexible regime switching mechanism is introduced to modify the classical threshold time series model by capturing the stickiness of signal. Besides, considering the inadequacy of traditional models under the lack of information, a signal retrospection is run in this model to provide a more accurate judgment. Moreover, formal proofs suggest strict stationarity and geometric ergodicity of BDAR model under several sufficient conditions. A Gaussian quasi-maximum likelihood estimation (QMLE) is employed and the asymptotic distributions of its estimators are derived. It has been demonstrated that the estimated thresholds of the BDAR model are n -consistent, each of which converges weakly to a functional of a two-sided compound Poisson process. The remaining parameters are \sqrt{n} -consistent and asymptotically normal. Furthermore, a model selection criteria and its asymptotic property have been established. Simulation studies are constructed to evaluate the finite sample performance of QMLE and model selection criteria. Finally, an empirical analysis of Hang Seng Index (HSI) using BDAR model reveals the asymmetry of investors' preference over losses and gains as well as the asymmetry of volatility structure.

Some key words: Double autoregressive model; Buffered threshold model; Geometric ergodicity.

First version available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2829536

This version: May 2018

*E-mail: zhao.liu@duke.edu

1 Introduction

A buffering phenomenon refers to the variable of interest sticks in the previous state when the variable of signal stays in a certain uninformative region. Consider, for example, a decision making of treatment plan for patients with diabetes. The up-crossing of blood sugar level over a certain high level r_U will intrigue an inpatient treatment, while a blood sugar level lower than r_L will evoke outpatient treatment. A doctor may easily choose between two treatment plans when a patient's blood sugar level is too high ($> r_U$) or too low ($< r_L$). However, he/she may become very hesitant to make the decision when the sugar level dwells in $[r_L, r_U]$, an interval without any other explicit benchmarks or information. This chaotic interval $[r_L, r_U]$ is exactly a typical buffer zone, where a retrospection of the historical data is needed to judge the status of the variable of interest. Therefore, in this case, a wise doctor might read the patient's previous record of blood sugar level. If the blood sugar level swings around r_U within the last week, the patient will be hospitalized overnight. Otherwise, the doctor will only give a prescription and require the patient to have at-home treatment.

Such regime switching mechanism was first formulated by Li, Guan, Li, and Yu (2015), in an effort to extend the traditional threshold AR model. Only focusing on the conditional mean function, Li, Guan, Li, and Yu (2015) has not considered conditional heteroscedasticity, which is, however, an essential part in the models of financial time series. Therefore, Lo et al. (2016) applied the similar idea to the conditional volatility process and proposed the buffered pure GARCH process in the study of exchange rate. Zhu et al. (2017) then introduced the concept of the buffered autoregressive model with generalized autoregressive conditional heteroscedasticity (BAR-GARCH) and empirically compared its performance with AR-GARCH and Threshold-GARCH model by an analysis of exchange rate. However, the theoretical property of BAR-GARCH model remained unexplored, for which no statistical inference could be substantiated. Besides, the volatility could not be explicitly computed by observations in a direct way in BAR-GARCH model.

Statistical test for the existence of threshold and model selection criteria are also crucial in the construction of buffered models. A quasi-likelihood ratio for the thresholds of BAR model has been established by Zhu et al. (2014), yet similar tests for a more generalized model such as BAR-GARCH model seems not readily available. Moreover, the application of model selection criteria (AIC, AIC_C and BIC) to the order selection in BAR-GARCH model lacks necessary simulation tests as well as theoretical support.

In this paper, we proposed a buffered double AR (BDAR) model that captures the stickiness

through applying a novel regime switching mechanism established by Li, Guan, Li, and Yu (2015) to double AR (DAR) model (see Ling (2004, 2007), Zhu and Ling (2013)). The theoretical properties of the proposed model are thoroughly investigated. Unlike the BAR-GARCH model, the conditional variance of each regime in the BDAR model is governed directly by past observations which allows a visible dynamic behavior of conditional volatility.

The remainder of the paper is organized as follows. Section 2 proposes the buffered double autoregressive (BDAR) model. Formal proofs suggest strict stationarity and geometric ergodicity of BDAR model under several sufficient conditions. A Gaussian quasi-maximum likelihood estimation (QMLE) is employed to compute the estimators in this model. The strict proof reveals the asymptotic normality of the estimated coefficients governing the conditional mean and variance of each regime in the BDAR model. More importantly, it has been demonstrated that each of the estimated thresholds of the BDAR model is n -consistent and weakly converges to the smallest minimizer of a two-sided compound Poisson process. Our results include the buffered AR (BAR) model (Li, Guan, Li, and Yu (2015)), threshold double AR (TDAR) model (Li, Ling, and Zhang (2016)) and traditional threshold AR (TAR) model (Tong (1990)) as special cases. A Bayesian-type information criteria (BIC), as well as its asymptotic property, are formally derived in section 3 for selecting orders of DAR models embedded in the BDAR model. Section 4 conducts several Monte Carlo simulation experiments to evaluate the finite sample performance of the Gaussian QMLE and proposed BIC. Finally, the model is applied to Hang Seng Index (HSI) in Section 5 to show the asymmetry of investors' preference over losses and gains as well as the leverage effect of the stock market. All technical proofs are relegated to Appendix.

2 Buffered double autoregressive models

Consider the following buffered double autoregressive (BDAR) model:

$$y_t = \begin{cases} \phi_{10} + \sum_{i=1}^p \phi_{1i} y_{t-i} + \varepsilon_t \sqrt{\alpha_{10} + \sum_{j=1}^p \alpha_{1j} y_{t-j}^2} & \text{if } R_t = 1 \\ \phi_{20} + \sum_{i=1}^p \phi_{2i} y_{t-i} + \varepsilon_t \sqrt{\alpha_{20} + \sum_{j=1}^p \alpha_{2j} y_{t-j}^2} & \text{if } R_t = 0 \end{cases} \quad (2.1)$$

$$R_t = \begin{pmatrix} 1 & \text{if } y_{t-d} \leq r_L \\ 0 & \text{if } y_{t-d} > r_U \\ R_{t-1} & \text{otherwise} \end{pmatrix}$$

Where $\alpha_{k0} > 0$ with $\alpha_{kj} \geq 0$ for $k = 1, 2$ and $j = 1, \dots, p$. And $\{\varepsilon_t\}$ are identically and independently distributed (*i.i.d.*) random variables with mean zero and variance one. The integer $d > 0$ is the delay parameter and $r_L \leq r_U$ are the boundary parameters of buffer zone. We

denote the model by BDAR(p) for simplicity. Note (2.1) includes the BAR model proposed by Li, Guan, Li, and Yu (2015) as a special case. When $r_L = r_U$, (2.1), the model becomes the threshold double AR model proposed by Li et al. (2016).

Denote $Y_t = (y_t, y_{t-1}, \dots, y_{t-p+1}, R_t)'$, $m_{it} = \phi_{1i}I(A_t) + \phi_{2i}I(A_t^c)$, where A_t is the event $\{y_{t-d} \leq r_L\} \cup \{r_L < y_{t-d} \leq r_U, R_{t-1} = 1\}$ and A_t^c is its complement. After introducing notations M_{0t} , M_{1t} , $g_1(Y_{t-1})$ and $g_2(Y_{t-1})$ as follows, we can show that $Y_t = g_1(Y_{t-1}) + \epsilon_t g_2(Y_{t-1})$ forms a Markov Chain.

$$M_{0t} = \begin{bmatrix} \phi_{10}I(A_t) + \phi_{20}I(A_t^c) \\ 0 \\ \vdots \\ 0 \\ I(y_{t-d} \leq r_L) \end{bmatrix} \quad M_{1t} = \begin{bmatrix} m_{1t} & m_{2t} & \dots & m_{pt} & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & I(r_L < y_{t-d} \leq r_U) \end{bmatrix}$$

$$g_1(Y_{t-1}) = M_{0t} + M_{1t}Y_{t-1}$$

$$g_2(Y_{t-1}) = \begin{bmatrix} \sqrt{(\alpha_{10} + \sum_{j=1}^p \alpha_{1j}y_{t-j}^2)I(A_t) + (\alpha_{20} + \sum_{j=1}^p \alpha_{2j}y_{t-j}^2)I(A_t^c)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

By applying a method similar to Lee (2006), several sufficient conditions are obtained for the geometric ergodicity of $\{Y_t\}$.

Theorem 1. *Suppose the distribution of ε_t has a positive density f over \mathbb{R} and $E|\varepsilon_t|^s < \infty$ for some $s > 0$. Moreover, f is locally bounded away from 0 and satisfies $\sup_{x \in \mathbb{R}} \{(1+|x|)f(x)\} < \infty$, then the multivariate process $\{Y_t\}$ is geometrically ergodic and hence (2.1) admits a strictly stationary, geometrically ergodic solution if one of the following conditions holds:*

- (i). $\sum_{j=1}^p \left(\sup_{1 \leq i \leq 2} |\phi_{ij}|^r + \sup_{1 \leq i \leq 2} (\alpha_{ij}^{r/2}) E|\varepsilon_t|^r \right) < 1$, $r \in (0, 1]$.
- (ii). $\left(\sum_{j=1}^p \sup_{1 \leq i \leq 2} |\phi_{ij}| \right)^r + \sum_{j=1}^p \sup_{1 \leq i \leq 2} (\alpha_{ij}^{r/2}) E|\varepsilon_t|^r < 1$, $r \in (1, 2]$ and f symmetric.
- (iii). $(1 + 3E(\epsilon_t^2))(\sum_{j=1}^p \sup_{1 \leq i \leq 2} |\phi_{ij}|)^4 + (E(\epsilon_t^4) + 3E(\epsilon_t^2))(\sum_{j=1}^p \sup_{1 \leq i \leq 2} \alpha_{ij})^2 < 1$, $r = 4$

Such conditions are easy to check but may be restrictive. Obtaining more general sufficient and necessary stability conditions seems difficult and thus is left for future research.

3 The QML estimator of BDAR(p) model

3.1 Estimation procedure

This section considers the Gaussian quasi-maximum likelihood estimation of the buffered double AR model specified in (2.1).

Denote by $\boldsymbol{\theta} = (\boldsymbol{\lambda}', r_L, r_U, d)'$ the parameter vector of model (2.1), $\boldsymbol{\lambda} = (\boldsymbol{\phi}_1', \boldsymbol{\alpha}_1', \boldsymbol{\phi}_2', \boldsymbol{\alpha}_2')$ with $\boldsymbol{\phi}_i' = (\phi_{i0}, \phi_{i1}, \dots, \phi_{ip})'$ and $\boldsymbol{\alpha}_i' = (\alpha_{i0}, \alpha_{i1}, \dots, \alpha_{ip})'$ and

$$\begin{aligned} u_t(\boldsymbol{\theta}) &= y_t - \mu_t(\boldsymbol{\theta}), \quad \mu_t(\boldsymbol{\theta}) = (\boldsymbol{\phi}_1' \mathbf{Y}_{1,t-1})R_t(r_L, r_U, d) + (\boldsymbol{\phi}_2' \mathbf{Y}_{2,t-1})(1 - R_t(r_L, r_U, d)) \\ h_t(\boldsymbol{\theta}) &= (\boldsymbol{\alpha}_1' \mathbf{X}_{1,t-1})R_t(r_L, r_U, d) + (\boldsymbol{\alpha}_2' \mathbf{X}_{2,t-1})(1 - R_t(r_L, r_U, d)) \end{aligned} \quad (3.1)$$

with $\mathbf{Y}_{1,t-1} = \mathbf{Y}_{2,t-1} = \mathbf{Y}_{t-1} = (1, y_{t-1}, \dots, y_{t-p})'$ and $\mathbf{X}_{1,t-1} = \mathbf{X}_{2,t-1} = \mathbf{X}_{t-1} = (1, y_{t-1}^2, \dots, y_{t-p}^2)'$. The conditional log-likelihood function (after multiplying -2 and omitting some constant) is defined as

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^n l_t(\boldsymbol{\theta}), \quad \text{where} \quad l_t(\boldsymbol{\theta}) = \log h_t(\boldsymbol{\theta}) + \frac{u_t^2(\boldsymbol{\theta})}{h_t(\boldsymbol{\theta})}$$

Let $\boldsymbol{\Lambda}$ be a compact set of \mathbb{R}^{4p+4} , $[a, b]$ be a predetermined interval, and d_{max} be a predetermined positive integer. Assume $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}, a \leq r_L \leq r_U \leq b$ and $d \in D = \{1, \dots, d_{max}\}$. The true parameter vector is $\boldsymbol{\theta}_0 = (\boldsymbol{\lambda}_0', r_{0L}, r_{0U}, d_0)'$ and the true regime indicator function is $R_{t0} = R_t(r_{0L}, r_{0U}, d_0)$

Let $n_0 = \max\{p, d_{max}\}$. For observed time series $\{y_t, -n_0 + 1 \leq t \leq n\}$ generated by (2.1), the likelihood functions defined in the above depend on past observations infinitely far away, due to the novel regime switching mechanism. Hence, initial values are needed to fit the model.

As discussed in Li, Guan, Li, and Yu (2015), for fixed (r_L, r_U, d) , the first few observations of the threshold variable y_{t-d} , say $1 \leq t \leq t_0$ may fall into the buffer zone $(r_L, r_U]$ such that the regime which y_1, \dots, y_{t_0} belong to cannot be identified. Note these t_0 observations come from the same regime since the threshold variables keep staying in the buffer zone. We may simply assign these t_0 observations to lower regime and denote the resulting regime indicator function by $\tilde{R}_t(r_L, r_U, d)$. Note the value of $R_{t_0+1}(r_L, r_U, d)$ is known since y_{t_0+1-d} is outside the buffer zone and it is clear $\tilde{R}_t(r_L, r_U, d) = R_t(r_L, r_U, d)$ when $t_0 < t \leq n$.

Denote by $\tilde{l}_t(\boldsymbol{\theta})$ and $\tilde{L}_n(\boldsymbol{\theta})$ the corresponding functions with $R_t(r_L, r_U, d)$ replaced by $\tilde{R}_t(r_L, r_U, d)$. We can then define the Gaussian QMLE of the true value $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ as $\hat{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \tilde{L}_n(\boldsymbol{\theta})$.

One can take two steps to search $\hat{\theta}_n$:

- First minimize $\hat{\theta}_n$ for each fixed (r_L, r_U, d) , i.e.

$$\hat{\lambda}_n(r_L, r_U, d) = \arg \min_{\lambda \in \Lambda} \tilde{L}_n(\lambda, r_L, r_U, d)$$

- Search for the estimators of (r_L, r_U, d) by

$$(\hat{r}_L, \hat{r}_U, \hat{d}) = \arg \min_{d \in D, a \leq r_L \leq r_U \leq b} \tilde{L}_n(\hat{\lambda}_n, (r_L, r_U, d), r_L, r_U, d)$$

Then we obtain $\hat{\theta}_n = (\hat{\lambda}'_n, \hat{r}_L, \hat{r}_U, \hat{d})'$.

3.2 Asymptotic Results

Assumption 1 $\text{pr}(y_t \in [a, b]) < 1$, ε_t is *i.i.d.* with zero mean and unit variance, and has a positive and continuous density $f(x)$ on \mathbb{R} .

Assumption 2 The parameter space

$\Lambda = \{\lambda \in \mathbb{R}^{4p+4} : \phi_1 \neq \phi_2 \text{ and } \alpha_1 \neq \alpha_2\}$ is compact. Moreover, each element in α_1 or α_2 is positive.

Theorem 2. Suppose Assumptions 1 and 2 hold and $\{y_t\}$ is strictly stationary and ergodic with $E y_t^2 < \infty$. Then $\hat{\theta}_n \rightarrow \theta_0$ a.s. as $n \rightarrow \infty$.

The proof of Theorem 2 follows the standard arguments in proving strong consistency. The delay parameter d only takes integer value. By theorem 2, when the sample size n is large, \hat{d} will be equal to d_0 . As in Li, Guan, Li, and Yu (2015), we assume the true delay parameter, d_0 , is known for the remainder of this subsection, and then it is deleted from the parameter vector θ and corresponding functions.

Assumption 3 $\kappa_4 \equiv E(\varepsilon_t^4) < \infty$ and $E(y_t^4) < \infty$.

Assumption 4 The conditional mean function $\mu_t(\theta)$ or volatility function $h_t(\theta)$ in (3.1) are discontinuous over the buffer zone $[r_{0L}, r_{0U}]$, i.e. there exist $p-1$ constants $z_{p-1}, \dots, z_{p-d+1}, z_{p-d-1}, \dots, z_0$ such that

$$\{(\phi_{10} - \phi_{20})'z\}^2 + \{(\alpha_{10} - \alpha_{20})'Z\}^2 > 0$$

for all $z_{p-d} \in [r_{0L}, r_{0U}]$ where $z = (1, z_{p-1}, \dots, z_0)'$ and $Z = (1, z_{p-1}^2, \dots, z_0^2)'$. Without loss of generality, here we assume that $d \leq p$.

Let $Y_t = (y_t, \dots, y_{t-p+1}, R_t)'$, then $\{Y_t\}$ is a Markov chain as shown in previous section. Denote its m-step transition probability function by $P^m(x, A)$, where $x \in \mathbb{R}^p \times \{0, 1\}$, $A \in \mathcal{B}_p \times \mathcal{U}$, where \mathcal{B}_p is the class of Borel sets of \mathbb{R}^p and $\mathcal{U} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$

Assumption 5 The time series $\{Y_t\}$ admits a unique invariant measure $\pi(\cdot)$, such that there exist $K > 0$ and $0 \leq \rho < 1$ for any $x \in \mathbb{R}^p \times \{0, 1\}$ and any m, $\|P^m(x, \cdot) - \pi(\cdot)\|_v \leq K(1 + \|x\|^2)\rho^m$, where $\|\cdot\|_v$ and $\|\cdot\|$ are respectively the total variation norm and Euclidean norm.

Under Assumption 5, $\{Y_t\}$ is said to be V-uniformly ergodic with $V(x) = K(1 + \|x\|^2)$, a condition stronger than geometric ergodicity.

Theorem 3. Suppose Assumptions 1 to 5 hold and θ_0 is an interior point of Θ . Then:

- (i). $n(\hat{r}_L - r_{0L}) = O_p(1)$ and $n(\hat{r}_U - r_{0U}) = O_p(1)$
- (ii). $\sqrt{n} \sup_{n(|r_U - r_{0U}| + |r_L - r_{0L}|) \leq B} \|\hat{\lambda}_n(r_L, r_U) - \hat{\lambda}_n(r_{0L}, r_{0U})\| = o_p(1)$ for any fixed constant $0 < B < \infty$, where $\hat{\lambda}_n(r_L, r_U)$ is the QMLE given r_L, r_U are known. Furthermore, it follows that:

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) = \sqrt{n}(\hat{\lambda}_n(r_{0L}, r_{0U}) - \lambda_0) + o_p(1) \Rightarrow N(\mathbf{0}, \Omega^{-1} \Sigma \Omega^{-1}), \text{ as } n \rightarrow \infty.$$

Where $\Omega = \text{diag}(\mathbf{A}_1, \mathbf{0.5B}_1, \mathbf{A}_2, \mathbf{0.5B}_2)$, $\Sigma = \text{diag}(\Sigma_1, \Sigma_2)$ with

$$\Sigma_i = \begin{pmatrix} \mathbf{A}_i & \frac{\kappa_3}{2} \mathbf{D}_i \\ \frac{\kappa_3}{2} \mathbf{D}_i' & \frac{\kappa_4 - 1}{4} \mathbf{B}_i \end{pmatrix},$$

where $\kappa_3 = E(\varepsilon_1^3)$

$$\mathbf{A}_i = E \left\{ \frac{Y_{i,t-1} Y_{i,t-1}'}{\alpha'_{i0} \mathbf{X}_{i,t-1}} g_i(r_{0L}, r_{0U}) \right\} \quad \mathbf{B}_i = E \left\{ \frac{\mathbf{X}_{i,t-1} \mathbf{X}_{i,t-1}'}{(\alpha'_{i0} \mathbf{X}_{i,t-1})^2} g_i(r_{0L}, r_{0U}) \right\}$$

$$\mathbf{D}_i = E \left\{ \frac{Y_{i,t-1} \mathbf{X}_{i,t-1}'}{(\alpha'_{i0} \mathbf{X}_{i,t-1})^{3/2}} g_i(r_{0L}, r_{0U}) \right\}$$

with $g_1(r_{0L}, r_{0U}) = R_t(r_{0L}, r_{0U})$ and $g_2(r_{0L}, r_{0U}) = 1 - R_t(r_{0L}, r_{0U})$

To study the limiting distribution of \hat{r}_L and \hat{r}_U , denote the $\xi_{1t} = \sum_{j=0}^{\infty} \zeta_{1,t+j} H_{t+j,j}$ and $\xi_{2t} =$

$\sum_{j=0}^{\infty} \zeta_{2,t+j} H_{t+j,j}$, where

$$\zeta_{1t} = \log \frac{\alpha'_{10} \mathbf{X}_{t-1}}{\alpha'_{20} \mathbf{X}_{t-1}} + \frac{\{(\phi_{10} - \phi_{20})' \mathbf{Y}_{t-1} - \epsilon_t \sqrt{\alpha'_{20} \mathbf{X}_{t-1}}\}^2}{\alpha'_{10} \mathbf{X}_{t-1}} - \epsilon_t^2$$

$$\zeta_{2t} = \log \frac{\alpha'_{20} \mathbf{X}_{t-1}}{\alpha'_{10} \mathbf{X}_{t-1}} + \frac{\{(\phi_{10} - \phi_{20})' \mathbf{Y}_{t-1} + \epsilon_t \sqrt{\alpha'_{10} \mathbf{X}_{t-1}}\}^2}{\alpha'_{20} \mathbf{X}_{t-1}} - \epsilon_t^2$$

$$H_{t,j} = \prod_{l=1}^j 1(r_{0L} < y_{t-d+1-l} \leq r_{0U}) \text{ with convention } \prod_{l=1}^0 = 1$$

Similar to Li, Guan, Li, and Yu (2015), for $i = 1, 2$, denote $F_{i,L}(\cdot|r)$ and $F_{i,U}(\cdot|r)$ be the conditional distribution functions of $\xi_{it}\{1-R_{t-1}(r_0)\}$ and $\xi_{it}R_{t-1}(r_0)$ given $y_{t-d} = r$ respectively, where $r_0 = (r_{0L}, r_{0U})'$. Denote $\pi(\cdot)$ the density function of y_t . Then define two independent one-dimensional two-sided compound Poisson process:

$$\begin{aligned}\varphi_L(t) &= 1(t \geq 0) \sum_{k=1}^{N_1^{(L)}(|t|)} \chi_k^{(1,L)} + 1(t < 0) \sum_{k=1}^{N_2^{(L)}(|t|)} \chi_k^{(2,L)} \\ \varphi_U(t) &= 1(t \geq 0) \sum_{k=1}^{N_1^{(U)}(|t|)} \chi_k^{(1,U)} + 1(t < 0) \sum_{k=1}^{N_2^{(U)}(|t|)} \chi_k^{(2,U)}\end{aligned}$$

where $\chi_k^{(i,L)}$ and $\chi_k^{(i,U)}$ have conditional distribution functions $F_{i,L}(\cdot|r_{0L})$ and $F_{i,U}(\cdot|r_{0U})$ respectively, $i = 1, 2$. $\{N_1^{(L)}(t), t \geq 0\}$, $\{N_2^{(L)}(t), t \geq 0\}$, $\{N_1^{(U)}(t), t \geq 0\}$, $\{N_2^{(U)}(t), t \geq 0\}$ are four independent Poisson Processes with $N_1^{(L)}(0) = N_2^{(L)}(0) = N_1^{(U)}(0) = N_2^{(U)}(0) = 0$. $N_1^{(L)}(\cdot)$, $N_2^{(L)}(\cdot)$ have jump rate $\pi(r_{0L})$, while $N_1^{(U)}(\cdot)$, $N_2^{(U)}(\cdot)$ have jump rate $\pi(r_{0U})$. Moreover, $N_1^{(L)}(\cdot)$ and $N_1^{(U)}(\cdot)$ are right continuous. $N_2^{(L)}(\cdot)$ and $N_2^{(U)}(\cdot)$ are left continuous.

It is implied by Assumption 4 that $E[\chi_k^{(1,j)}]$ and $E[\chi_k^{(2,j)}] > 0$ for $j = L$ and U . Then $\varphi_j(t) \rightarrow \infty$ a.s. as $|t| \rightarrow \infty$. Then define $\varphi(z) = \varphi_L(z_L) + \varphi_U(z_U)$, where $z = (z_L, z_U) \in R^2$. As in Li, Guan, Li, and Yu (2015) and Li, Ling, and Zakoian (2015), there exists a unique random square $[M_-^{(L)}, M_+^{(L)}] \times [M_-^{(U)}, M_+^{(U)}]$ on which $\varphi(z)$ attains the global minimum, where $[M_-^{(j)}, M_+^{(j)}] = \operatorname{argmin}_{t \in R} \varphi_j(t)$.

The following theorem states that $n(\hat{r}_L - r_{0L})$ and $n(\hat{r}_U - r_{0U})$ converges weakly to a functional of the compound Poisson Process $\varphi(z)$.

Theorem 4. *If Assumption 1 to 5 hold, then $n(\hat{r}_L - r_{0L}) \rightarrow M_-^{(L)}$ and $n(\hat{r}_U - r_{0U}) \rightarrow M_-^{(U)}$. Furthermore, $n(\hat{r}_L - r_{0L})$, $n(\hat{r}_U - r_{0U})$ and $\sqrt{n}(\hat{\lambda}_n - \lambda_0)$ are asymptotically independent.*

3.3 Model Selection

We consider the Bayesian Information Criteria (BIC) for model selection, since BIC outperforms AIC for buffered type models, see Li, Guan, Li, and Yu (2015).

Write for simplicity \tilde{R}_t as the QMLE estimator of $\tilde{R}_t(\hat{r}_L, \hat{r}_U, \hat{d})$. $n_1 = \sum_{t=1}^n \tilde{R}_t$ and $n_2 = n - n_1$. Then define

$$\text{BIC}(p) = \sum_{t=1}^n \mathbf{l}_t(\hat{\theta}) + (2p + 2) \log n_1 + (2p + 2) \log n_2$$

where $l_t(\theta)$ is defined in Section 3.1. Let p_0 be the true order of the model and p_{\max} be a predetermined large order. $\hat{p}_n = \operatorname{argmin}_{0 \leq p \leq p_{\max}} \text{BIC}(p)$ Then,

Theorem 5. *If Assumption 1 to 5 hold, then $\Pr(\hat{p}_n = p_0) \rightarrow 1$ as $n \rightarrow \infty$.*

4 Simulation studies

We first examine the finite sample performance of QMLE. We use sample size $n = 400, 800$, each with replications 500 for the following model

$$y_t = \begin{cases} -0.1 + 0.2y_{t-1} + 0.1y_{t-2} + \varepsilon_t \sqrt{0.1 + 0.3y_{t-1}^2 + 0.05y_{t-2}^2} & R_t = 1 \\ 0.1 - 0.2y_{t-1} + 0.3y_{t-2} + \varepsilon_t \sqrt{0.05 + 0.2y_{t-1}^2 + 0.1y_{t-2}^2} & R_t = 0 \end{cases} \quad (4.1)$$

with the regime indicator

$$R_t = \begin{pmatrix} 1 & \text{if } y_{t-4} \leq -0.1 \\ 0 & \text{if } y_{t-4} > 0.15 \\ R_{t-1} & \text{otherwise} \end{pmatrix}$$

where $\{\varepsilon_t\}$ are independently and identically distributed with standard normality. The estimation procedure discussed in Section 3.1 is employed. The range of the boundary parameters r_L and r_U is set from 10th percentile to 90th percentile of each sample. And the maximum delay parameter d_{\max} is set to be 6. For each sample size $n = 400, 800$, d can be correctly identified for all 500 replications respectively. The bias, the empirical standard deviation (ESD) and the asymptotic standard deviation (ASD) are listed in Table 1. We can observe the bias and the empirical standard deviation decrease as the sample size increases and all the ESDs are close to corresponding ASDs, which are consistent with the asymptotic results in Theorem 2 and Theorem 3. Note the ASD of boundary parameters r_L and r_U cannot be estimated and only the ESD are included in Table 1. We observe the ESD of r_L and r_U are approximately reduced by one half when the sample size doubles, which is evidence supporting the super-consistency of boundary parameters. For each sample size, the histograms of $n(\hat{r}_L - r_{0L})$ and $n(\hat{r}_U - r_{0U})$ are displayed in Figure 1, whose shapes are close to the ones reported in Li et al. (2013) and Li, Guan, Li, and Yu (2015)

We now evaluate the finite sample performance of model selection criteria (BIC) in Section 3.3. The data generating process is the same as (4.1). The sample size is 800, and there are 150 replications. The maximum order p_{\max} for each regime's conditional mean and conditional variance is set to be 6. And the maximum delay parameter d_{\max} is still set to be 6. As a result, the BIC correctly identifies all true orders at a rate of 100%.

5 Empirical Analysis

This section considers the weekly closing prices of Hang Seng Index over the period January 2000 to December 2007. We focus on the log return in percentage $y_t = 100(\log P_t - \log P_{t-1})$, where P_t is the weekly closing price at time t . There are 417 observations of $\{y_t\}$ in total. The plot of $\{P_t\}$ and $\{y_t\}$ are displayed in Figure 2. Li et al. (2016) have studied this series and they conclude the existence of nonlinearity, ARCH effect and threshold effect through Tsay's test (Tsay (1986)), Mcleod-Li's test (Mcleod and Li (1983)) and their score-based test respectively. The two-regime threshold double autoregressive model (TDAR) is employed to fit the data in Li et al. (2016). Since TDAR model can be considered as a special case of BDAR model, it motivates us to consider the proposed BDAR model to fit the percentage log return series.

The range of boundary parameters r_L and r_U is from the 10th to the 90th empirical percentiles of observations. The maximum delay parameter d_{max} is set to be 6 and the maximum order p_{max} for each regime's conditional mean and conditional variance is set to be 5. Based on BIC, we have the following fitted model:

$$y_t = \begin{cases} 0.3937_{0.2621} + 0.0385_{0.1318}y_{t-1} + 0.2093_{0.0894}y_{t-2} + \varepsilon_t\sigma_t & R_t = 1 \\ -0.5992_{0.2665} + 0.2354_{0.0970}y_{t-1} + \varepsilon_t\sigma_t & R_t = 0 \end{cases} \quad (5.1)$$

where

$$\sigma_t^2 = \begin{cases} 5.3991_{0.9225} + 0.5432_{0.1583}y_{t-1}^2 + 0.1787_{0.1161}y_{t-2}^2 & R_t = 1 \\ 3.4473_{0.7086} + 0.0263_{0.0599}y_{t-1}^2 + 0.0678_{0.0578}y_{t-2}^2 + 0.1416_{0.0631}y_{t-3}^2 & R_t = 0 \end{cases}$$

with the regime indicator

$$R_t = \begin{cases} 1 & \text{if } y_{t-1} \leq -0.2048 \\ 0 & \text{if } y_{t-1} > 0.8770 \\ R_{t-1} & \text{otherwise} \end{cases}$$

where the subscripts of parameter estimates are their associated standard errors. Figure 3 displays respectively ACFs of residuals and squared residuals, and they slightly go beyond the 95% confidence bands only at a few lags, which partially suggests the adequacy of the fitted model. To further investigate the fit-adequacy, the Ljung-Box test statistic Q_m and the McLeod-Li test statistic \tilde{Q}_m are employed. Generally, m takes value 6 and 12, see Tse (2002) for a discussion on the choice of m . The p-values for of Q_6 , Q_{12} , \tilde{Q}_6 and \tilde{Q}_{12} are 0.6590, 0.3666, 0.8200 and 0.4013 respectively, which further suggests the fit-adequacy at 5% significance level.

The asymmetry of the buffered zone around 0 might be interpreted as market participants are more sensitive to losses and gains. Based on the fitted model, the investors require the return of

previous week to go above 0.8770% to confirm a good time, approximately three times larger in absolute value than the signal of a bad time. When the previous week's return lies in the buffer zone $[-0.2048\% \ 0.8770\%]$, investors are not sure about the market condition, so they trace back to learn R_{t-1} , which may depend on returns several periods ago. Therefore, the buffered double autoregressive model will further rely on the information contained in the dynamic momentum structure, when the past week's return alone is not quite informative on the market conditions. The BDAR model provides a more realistic characterization on the regime switching mechanism than the traditional two-regime threshold type model. The sudden switch of the probabilistic structure as in the traditional threshold time series models can be delayed by the incorporation of the buffer zone.

Another feature of the fitted BDAR model is the asymmetry of volatility. The coefficients in the lower regime's conditional variance is much larger than their counterparts in the upper regime. This asymmetry can be interpreted as the stock market is more volatile during bad times.

6 Appendix

6.1 Proof of Theorem 1

As in the main part of the paper, define

$$Y_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \\ R_t \end{bmatrix}$$

$$M_{0t} = \begin{bmatrix} \phi_{10}I(A_t) + \phi_{20}I(A_t^c) \\ 0 \\ \vdots \\ 0 \\ I(y_{t-d} \leq r_L) \end{bmatrix} \quad M_{1t} = \begin{bmatrix} m_{1t} & m_{2t} & \dots & m_{pt} & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & I(r_L < y_{t-d} \leq r_U) \end{bmatrix}$$

$$g_1(Y_{t-1}) = M_{0t} + M_{1t}Y_{t-1}$$

$$g_2(Y_{t-1}) = \begin{bmatrix} \sqrt{(\alpha_{10} + \sum_{j=1}^p \alpha_{1j}y_{t-j}^2)I(A_t) + (\alpha_{20} + \sum_{j=1}^p \alpha_{2j}y_{t-j}^2)I(A_t^c)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$Y_t = g_1(Y_{t-1}) + \epsilon_t g_2(Y_{t-1}) \quad \text{forms a Markov Chain}$$

Let \mathcal{B}^p be the class of Borel sets of \mathcal{R}^p , μ_{p+1} be the Lebesgue measure on \mathcal{R}^{p+1} and $\mathcal{U} = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$

We denote its state space by $(\mathcal{R}^p \times \{0, 1\}, \mathcal{B}^p \times \mathcal{U}, \mu_{p+1})$ and sets its transition probability function as:

$$P(x, A) = \int_{A_1} f_\epsilon(y) dy, \quad x \in \mathcal{R}^p \times \{0, 1\} \text{ and } A \in \mathcal{B}^p \times \mathcal{U}$$

Where $A_1 = \{\epsilon : \epsilon g_2(x) + g_1(x) \in A\}$ and $f_\epsilon(\cdot)$ is the density of ϵ_t .

By a method similar to Lemma 1 and 2 of Lu (1998), it can be shown that the chain is μ_{p+1} -irreducible and aperiodic, and non-null compact sets are small sets.

Before we proceed, we need to state one lemma and one theorem. The lemma is proved in Lee (2006) and the theorem comes from Tweedie (1983):

Lemma 6.1. (Lemma 2.2 in Lee (2006)) *Let $V(z) = \sum_{i=1}^n \gamma_i |z_i|^r, z = (z_1, z_2, \dots, z_n), n \in \mathbb{Z}^+, r > 0$. If $\sum_{i=1}^n \xi_i < 1$ with $\xi_i \geq 0$, we may choose $\gamma_i > 0, i = 1, \dots, n$ so that for some positive constant $\rho < 1, \gamma_1(\sum_{i=1}^n \xi_i |z_i|^r) + \sum_{i=2}^n \gamma_i |z_{i-1}|^r \leq \rho V(z)$.*

Theorem 6.2. (Theorem 4 in Tweedie (1983)) *Suppose that the Markov process $\{Y_t : t \geq 0\}$ is aperiodic μ -irreducible and \mathcal{B} is a small set. Suppose there are constants $\rho < 1, \epsilon > 0$ and a measurable function $V \geq 1$ such that*

$$E[V(Y_t)|Y_{t-1} = z] \leq \rho V(z) - \epsilon, z \in \mathcal{B}^c$$

$$\sup_{z \in \mathcal{B}} E[V(Y_t)|Y_{t-1} = z] < \infty$$

then the Markov process Y_t is geometrically ergodic.

We continue the proof. Let $b_j = \max_{1 \leq i \leq 2} |\phi_{ij}|, 0 \leq j \leq p, \quad a_j = \max_{1 \leq i \leq 2} \alpha_{ij}, 0 \leq j \leq p$. Define a test function V by:

$$V(Y_t) = \sum_{i=1}^p \gamma_i |y_{t+1-i}|^r + 1$$

$$|y_t| \leq b_0 + \sum_{j=1}^p b_j |y_{t-j}| + |\epsilon_t| (a_0 + \sum_{j=1}^p a_j y_{t-j}^2)^{\frac{1}{2}}$$

Define:

$$s_1(y_{t-1}, \dots, y_{t-p}) = b_0 + \sum_{j=1}^p |y_{t-j}| b_j$$

$$s_2(y_{t-1}, \dots, y_{t-p}) = [a_0 + \sum_{j=1}^p y_{t-j}^2 a_j]^{\frac{1}{2}}$$

Similar to Lee (2006), we have:

$$s_1^r \leq \begin{cases} b_0^r + \sum_{j=1}^p b_j^r |y_{t-j}|^r, & 0 < r \leq 1 \\ (1 + \epsilon)^r \left(\sum_{j=1}^p b_j \right)^{r-1} \left(\sum_{j=1}^p b_j |y_{t-j}|^r \right), & 1 < r \leq 2, \forall \epsilon > 0, \exists M(\epsilon) \text{ s.t. } \|Y_{t-1}\| > M(\epsilon) \end{cases}$$

$$s_2^r \leq a_o^{\frac{r}{2}} + \sum_{j=1}^p a_j^{\frac{r}{2}} |y_{t-j}|^r, \quad 0 < r \leq 2$$

Moreover, when $0 < r \leq 1$, by basic inequality, we have $E[|y_t|^r | Y_{t-1}] \leq s_1^r + s_2^r E|\epsilon_t|^r$

When $1 < r \leq 2$, note ϵ_t is symmetric by assumption, by the following inequality:

$$(1+x)^r + (1-x)^r \leq 2(|x|^r + 1), \quad -1 \leq x \leq 1$$

we can show that

$$E[|s_1 + s_2 \epsilon_t|^r | Y_{t-1}] = \frac{1}{2} E[|s_1 + s_2 \epsilon_t|^r + |s_1 - s_2 \epsilon_t|^r | Y_{t-1}] \leq s_1^r + s_2^r E|\epsilon_t|^r$$

Therefore, $E[V(Y_t) | Y_{t-1}]$

$$\begin{aligned} &\leq \gamma_1 (s_1^r + s_2^r E|\epsilon_t|^r) + \sum_{i=2}^p \gamma_i |y_{t+1-i}|^r + 1 \\ &\leq \gamma_1 \sum_{i=1}^p \xi_i |y_{t-i}|^r + \sum_{i=2}^p \gamma_i |y_{t+1-i}|^r + 1 + C_1, \end{aligned}$$

where C_j is a generic notation for positive constants

$$\sum_{i=1}^p \xi_i = \begin{cases} \sum_{j=1}^p b_j^r + \sum_{j=1}^p a_j^{\frac{r}{2}} E|\epsilon_t|^r, & 0 < r \leq 1 \\ (1 + \epsilon)^r \left(\sum_{j=1}^p b_j \right)^r + \sum_{j=1}^p a_j^{\frac{r}{2}} E|\epsilon_t|^r, & 1 < r \leq 2, \forall \epsilon > 0, \exists M(\epsilon) \text{ s.t. } \|Y_{t-1}\| > M(\epsilon) \end{cases}$$

By Lemma 6.1, we can set $\sum_{i=1}^p \xi_i < 1$ and find $\gamma_i > 0, i = 1, \dots, p$ such that for $\|z\| > M, \exists$ some constants $\rho < 1, E[V(Y_t) | Y_{t-1} = z] \leq \rho V(z) + C_2$. Thus, inequalities in Theorem 6.2 hold with some $\epsilon > 0$ and compact set $\mathcal{B} = \{\|z\| \leq M\}$ for sufficiently large $M < \infty$, since $V(z)$ increases as $\|z\|$ increases. Hence, we obtain the geometric ergodicity and hence the strict stationarity.

To prove the third part of the theorem, define $V(z) = \sum_{i=1}^p \gamma_i z_i^4 + 1$ from $E(\epsilon_t) = E(\epsilon_t^3) = 0$ for $\|z\| > M$.

$$E[(s_1 + s_2 \epsilon_t)^4 | Y_{t-1}] \leq [(1 + 3E(\epsilon_t^2))s_1^4 + (E\epsilon_t^4 + 3E\epsilon_t^2)s_2^4]$$

$$s_1^4 \leq (1 + \epsilon)^4 \left(\sum_{j=1}^p b_j y_{t-j}^4 \right) \left(\sum_{j=1}^p b_j \right)^3$$

$$s_2^4 \leq (1 + \epsilon)^2 \left(\sum_{j=1}^p a_j y_{t-j}^4 \right) \left(\sum_{j=1}^p a_j \right)$$

Therefore for $\|z\| > M$

$$E[V(Y_t) | Y_{t-1} = z] \leq \gamma_1 \{ [1 + 3E(\epsilon_t^2)]s_1^4 + [E(\epsilon_t^4) + 3E(\epsilon_t^2)]s_2^4 \} + \sum_{i=2}^p \gamma_i |y_{t+1-i}|^4 + 1$$

$$\leq \gamma_1 \sum_{i=1}^p \xi_i |y_{t-i}|^4 + \sum_{i=2}^p \gamma_i |y_{t+1-i}|^4 + 1$$

$$\text{where } \sum_{i=1}^p \xi_i = [1 + 3E(\epsilon_t^2)](1 + \epsilon)^4 \left(\sum_{j=1}^p b_j \right)^4 + [E(\epsilon_t^4) + 3E(\epsilon_t^2)](1 + \epsilon)^2 \left(\sum_{j=1}^p a_j \right)^2$$

By Lemma 6.1, we can set $\sum_{i=1}^p \xi_i < 1$ and obtain the geometric ergodicity as we do in proving the first two parts of the theorem.

6.2 Proof of Theorem 2

Before we prove Theorem 2, we introduce some notations and a preliminary result.

Denote the parameter space as $\Theta = \Lambda \times [a, b] \times [a, b] \times D$, where $\Lambda = \Lambda_1 \times \Lambda_2$.

Parametre vector $\theta = (\phi'_1, \alpha'_1, \phi'_2, \alpha'_2, r_L, r_U, d)' = (\lambda'_1, \lambda'_2, r_L, r_U, d)'$, where

$$\lambda_1 = \begin{bmatrix} \phi_1 \\ \alpha_1 \end{bmatrix} \quad \lambda_2 = \begin{bmatrix} \phi_2 \\ \alpha_2 \end{bmatrix}$$

Recall in the main body of the paper, we define

$$\begin{aligned} \mathbf{l}_t(\theta) &= \log \mathbf{h}_t(\theta) + \frac{\mathbf{u}_t^2(\theta)}{\mathbf{h}_t(\theta)} \\ &= R_t \left\{ \log \mathbf{h}_{1t}(\theta) + \frac{[y_t - \mu_{1t}(\theta)]^2}{\mathbf{h}_{1t}(\theta)} \right\} + (1 - R_t) \left\{ \log \mathbf{h}_{2t}(\theta) + \frac{[y_t - \mu_{2t}(\theta)]^2}{\mathbf{h}_{2t}(\theta)} \right\} \end{aligned}$$

where for $i=1,2$

$$\mu_{it}(\theta) = \phi_{i0} + \phi_{i1}y_{t-1} + \cdots + \phi_{ip}y_{t-p}$$

$$\mathbf{h}_{it}(\theta) = \alpha_{i0} + \alpha_{i1}y_{t-1}^2 + \cdots + \alpha_{ip}y_{t-p}^2$$

We further denote it as $\mathbf{l}_t(\theta) = R_t \mathbf{l}_{1t}(\theta) + (1 - R_t) \mathbf{l}_{2t}(\theta)$, with

$$l_{1t}(\theta) = \log h_{1t}(\theta) + \frac{[y_t - \mu_{1t}(\theta)]^2}{h_{1t}(\theta)}$$

$$l_{2t}(\theta) = \log h_{2t}(\theta) + \frac{[y_t - \mu_{2t}(\theta)]^2}{h_{2t}(\theta)}$$

If Assumption 2 and conditions in Theorem 2 hold, Lemma B.2 in Ling (2007) implies that

$$E[\sup_{\theta \in \Theta} |l_{it}(\theta)|] < \infty \quad i=1,2 \text{ and } E[\sup_{\theta \in \Theta} |l_t(\theta)|] < \infty \quad (6.3)$$

This preliminary result will be used to invoke dominance convergence theorem in the following.

We are ready to prove Theorem 2 now.

Proof of Theorem 2

Following the method in Huber(1967), it is sufficient for us to verify the following three claims:

- S1: $\sup_{\theta \in \Theta} \frac{1}{n} |\tilde{L}_n(\theta) - L_n(\theta)| \xrightarrow{a.s.} 0$, where $\tilde{L}_n(\theta)$ is the modified likelihood function defined in the paper.
- S2: $E[l_t(\theta)]$ is uniquely minimized at $\theta = \theta_0 = (\phi'_{10}, \alpha'_{10}, \phi'_{20}, \alpha'_{20}, r_{0L}, r_{0U}, d_0)'$
- S3: $E[\sup_{\theta^* \in U_\eta(\theta)} |l_t(\theta) - l_t(\theta^*)|] \xrightarrow{a.s.} 0$ as $\eta \rightarrow 0$, where $U_\eta(\theta) = \{\theta^* \in \Theta, ||\theta^* - \theta|| < \eta\}$

Let us first show S1:

$$\begin{aligned} \frac{1}{n} |\tilde{L}_n(\theta) - L_n(\theta)| &= \frac{1}{n} \left| \sum_{t=1}^{k_n} (\tilde{l}_t(\theta) - l_t(\theta)) \right| \\ &= \{(1 - R_1) \prod_{t=1}^{k_n} 1(r_L < y_{t-d} \leq r_U)\} \left\{ \frac{1}{n} \sum_{t=1}^{k_n} [l_{1t}(\theta) - l_{2t}(\theta)] \right\} \\ &\leq \prod_{t=1}^{k_n} 1(a \leq y_{t-d} \leq b) \left\{ \frac{1}{n} \sum_{t=1}^{k_n} [l_{1t}(\theta) - l_{2t}(\theta)] \right\} \end{aligned}$$

If $\frac{k_n}{n} \rightarrow 0$, then clearly $\sup_{\theta \in \Theta} \frac{1}{n} |\tilde{L}_n(\theta) - L_n(\theta)| \rightarrow 0$

If $\frac{k_n}{n} \rightarrow$ a positive number (constant), then by ergodic theorem, Assumption 1 in the paper and (6.3), we have:

$$\sup_{\theta \in \Theta} \frac{1}{k_n} \sum_{t=1}^{k_n} [\mathbf{l}_{1t}(\theta) - \mathbf{l}_{2t}(\theta)] \xrightarrow{a.s.} E[\sup_{\theta \in \Theta} [\mathbf{l}_{1t}(\theta) - \mathbf{l}_{2t}(\theta)]] \leq E[\sup_{\theta \in \Theta} [\mathbf{l}_{1t}(\theta)]] + E[\sup_{\theta \in \Theta} [\mathbf{l}_{2t}(\theta)]] < \infty$$

Moreover, by Assumption 1,

$$\frac{1}{k_n} \sum_{t=1}^{k_n} 1(y_{t-d} < a \text{ or } y_{t-d} > b) \rightarrow Pr(y_{t-d} < a \text{ or } y_{t-d} > b) > 0$$

$$Pr(\lim_{k_n \rightarrow \infty} \prod_{t=1}^{k_n} 1(a \leq y_{t-d} \leq b) = 1) = Pr(\lim_{k_n \rightarrow \infty} \frac{1}{k_n} \sum_{t=1}^{k_n} 1(y_{t-d} < a \text{ or } y_{t-d} > b) = 0) = 0$$

$$i.e., \lim_{k_n \rightarrow \infty} \prod_{t=1}^{k_n} 1(a \leq y_{t-d} \leq b) \rightarrow 0 \text{ as } k_n \rightarrow \infty.$$

Therefore, we still have $\sup_{\theta \in \Theta} \frac{1}{n} |\tilde{\mathbf{L}}_n(\theta) - \mathbf{L}_n(\theta)| \rightarrow 0$. We finish the proof of S1.

Let us then show S2:

$$\begin{aligned} & E_{\theta_0} \{\mathbf{l}_t(\theta)\} - E_{\theta_0} \{\mathbf{l}_t(\theta_0)\} \\ &= E_{\theta_0} \{[\mathbf{l}_{1t}(\theta) - \mathbf{l}_{1t}(\theta_0)]R_{0t}R_t\} + E_{\theta_0} \{[\mathbf{l}_{2t}(\theta) - \mathbf{l}_{2t}(\theta_0)](1 - R_{0t})(1 - R_t)\} \\ &+ E_{\theta_0} \{[\mathbf{l}_{1t}(\theta) - \mathbf{l}_{2t}(\theta_0)](1 - R_{0t})R_t\} + E_{\theta_0} \{[\mathbf{l}_{2t}(\theta) - \mathbf{l}_{1t}(\theta_0)]R_{0t}(1 - R_t)\} \end{aligned}$$

Consider $E_{\theta_0} [(\mathbf{l}_{1t}(\theta) - \mathbf{l}_{2t}(\theta_0))R_{0t}R_t]$

$$= E_{\theta_0} \left\{ \left(\log \frac{\alpha'_1 \mathbf{X}_{t-1}}{\alpha'_{10} \mathbf{X}_{t-1}} + \frac{\alpha'_{10} \mathbf{X}_{t-1}}{\alpha'_1 \mathbf{X}_{t-1}} - 1 + \frac{[(\phi_{10} - \phi_1)' \mathbf{Y}_{t-1}]^2}{\alpha'_1 \mathbf{X}_{t-1}} \right) R_{0t}R_t \right\}$$

By $\log \frac{1}{x} + x - 1 \geq 0$, we know:

$$\begin{aligned} E_{\theta_0} \{[\mathbf{l}_{1t}(\theta) - \mathbf{l}_{2t}(\theta)]R_{0t}R_t\} &\geq 0 \text{ with equality iff } \frac{\alpha'_1 \mathbf{X}_{t-1}}{\alpha'_{10} \mathbf{X}_{t-1}} = 1 \text{ and} \\ &(\phi_{10} - \phi_1)' \mathbf{Y}_{t-1} = 0, i.e., \alpha_1 = \alpha_{10}, \phi_1 = \phi_{10} \end{aligned}$$

Similarly, we have:

$$\begin{aligned} E_{\theta_0} \{[\mathbf{l}_{2t}(\theta) - \mathbf{l}_{2t}(\theta_0)](1 - R_{0t})(1 - R_t)\} &\geq 0 \text{ with equality iff } \frac{\alpha'_2 \mathbf{X}_{t-1}}{\alpha'_{20} \mathbf{X}_{t-1}} = 1 \text{ and} \\ &(\phi_{20} - \phi_2)' \mathbf{Y}_{t-1} = 0, i.e., \alpha_2 = \alpha_{20}, \phi_2 = \phi_{20} \end{aligned}$$

For the third term, we have $E_{\theta_0} \{[\mathbf{l}_{1t}(\theta) - \mathbf{l}_{2t}(\theta_0)](1 - R_{0t})R_t\} \geq 0$. Since we assume $\phi_{10} \neq \phi_{20}$ and $\alpha_{10} \neq \alpha_{20}$, $E_{\theta_0} \{[\mathbf{l}_{1t}(\theta) - \mathbf{l}_{2t}(\theta_0)](1 - R_{0t})R_t\} = 0$ implies $E_{\theta_0} \{(1 - R_{0t})R_t\} = 0$, from which it can be deduced, following the same argument in proving Theorem 2 in Li, Guan, Li, and Yu (2015), that $d = d_0, r_L \geq r_{0L}, r_U \geq r_{0U}$. Similarly, from the last term $E_{\theta_0} \{[\mathbf{l}_{2t}(\theta) - \mathbf{l}_{1t}(\theta_0)]R_{0t}(1 - R_t)\} = 0$, we have $d = d_0, r_L \leq r_{0L}, r_U \leq r_{0U}$. We thus finish the proof of S2.

Let us further show S3:

$$\theta^* \in U_\eta(\theta)$$

$$\begin{aligned} l_t(\theta^*) - l_t(\theta) &= \{l_{1t}(\theta^*) - l_{1t}(\theta)\}R_t(r_L^*, r_U^*, d) \\ &\quad + l_{1t}(\theta)[R_t(r_L^*, r_U^*, d) - R_t(r_L, r_U, d)] \\ &\quad + [l_{2t}(\theta^*) - l_{2t}(\theta)][1 - R_t(r_L^*, r_U^*, d)] \\ &\quad + l_{2t}(\theta)[R_t(r_L, r_U, d) - R_t(r_L^*, r_U^*, d)] \end{aligned} \tag{6.4}$$

$$\begin{aligned} l_{1t}(\theta^*) - l_{1t}(\theta) &= \log \alpha_1^{*'} X_{t-1} - \log \alpha_1' X_{t-1} \\ &\quad + \frac{(y_t - \phi_1^{*'} Y_{t-1})^2 - (y_t - \phi_1' Y_{t-1})^2}{\alpha_1^{*'} X_{t-1}} \\ &\quad + (y_t - \phi_1' Y_{t-1})^2 \left[\frac{1}{\alpha_1^{*'} X_{t-1}} - \frac{1}{\alpha_1' X_{t-1}} \right] \end{aligned}$$

By Taylor's expansion and a method similar to Zhu and Ling (2013), we have:

$$\begin{aligned} E\left[\sup_{\theta^* \in U_\eta(\theta)} |\log \alpha_1^{*'} X_{t-1} - \log \alpha_1' X_{t-1}| \right] &\rightarrow 0 \quad \text{as } \eta \rightarrow 0 \\ E\left[\sup_{\theta^* \in U_\eta(\theta)} \left| \frac{(y_t - \phi_1^{*'} Y_{t-1})^2 - (y_t - \phi_1' Y_{t-1})^2}{\alpha_1^{*'} X_{t-1}} \right| \right] &\rightarrow 0 \quad \text{as } \eta \rightarrow 0 \\ E\left[\sup_{\theta^* \in U_\eta(\theta)} \left| (y_t - \phi_1' Y_{t-1})^2 \left(\frac{1}{\alpha_1^{*'} X_{t-1}} - \frac{1}{\alpha_1' X_{t-1}} \right) \right| \right] &\rightarrow 0 \quad \text{as } \eta \rightarrow 0 \end{aligned}$$

Therefore, the first and third terms in expression (6.4) can be dealt with.

We now look at the second term in (6.4).

Since $l_{1t}(\theta)$ and $R_t(r_L^*, r_U^*, d) - R_t(r_L, r_U, d)$ have non-overlapping parameters,

$$E\left\{ \sup_{\theta^* \in U_\eta(\theta)} l_{1t}(\theta) [R_t(r_L^*, r_U^*, d) - R_t(r_L, r_U, d)] \right\} = E\left\{ \sup_{\theta^* \in U_\eta(\theta)} l_{1t}(\theta) \sup_{\theta^* \in U_\eta(\theta)} [R_t(r_L^*, r_U^*, d) - R_t(r_L, r_U, d)] \right\}$$

Li, Guan, Li, and Yu (2015) shows that $E\left\{ \sup_{\theta^* \in U_\eta(\theta)} |R_t(r_L^*, r_U^*, d) - R_t(r_L, r_U, d)| \right\} \rightarrow 0$ as $\eta \rightarrow 0$.

Combining this with $E\left[\sup_{\theta^* \in U_\eta(\theta)} |l_{1t}(\theta)| \right] < \infty$ and using dominated convergence theorem, we have:

$$E\left\{ \sup_{\theta^* \in U_\eta(\theta)} l_{1t}(\theta) [R_t(r_L^*, r_U^*, d) - R_t(r_L, r_U, d)] \right\} \rightarrow 0 \text{ as } \eta \rightarrow 0$$

The fourth term in (6.4) can be dealt with in the same way. We thus finish the proof of $S3$.

Based on $S1$ - $S3$ and Huber (1967), Theorem 2 is proved.

6.3 Proof of Theorem 3

We first prove the super-consistency of the estimated boundary parameters at Claim (i). Note in Theorem 2, we know $\hat{\theta}_n$ is strong consistent. Therefore, without loss of generality, we restrict the parameter space to a neighborhood of θ_0 .

Define $w(\Delta) = \{\lambda \in \Lambda, a < r_L < r_U < b : \|\lambda - \lambda_0\| < \Delta, |r_L - r_{0L}| < \Delta, |r_U - r_{0U}| < \Delta\}$, where $0 < \Delta < \min\{1, \frac{r_{0U} - r_{0L}}{2}\}$ will be determined later. As in Chan (1993), it is sufficient to show for any $\epsilon > 0$, \exists a positive K such that:

$$\Pr(\tilde{L}_n(\lambda, r_{0L} + z_L, r_{0U} + z_U) - \tilde{L}_n(\lambda, r_{0L}, r_{0U}) > 0) > 1 - \epsilon \quad (6.5)$$

where $\lambda \in w(\Delta), |z_L| > \frac{K}{n}, |z_U| > \frac{K}{n}$.

We consider the case $p = d = 1, z_L < 0$ and $z_U > 0$. Denote the disjoint events:

$$A_0 = \{r_{0U} < y_{t-1} \leq r_{0U} + z_U, R_{t-1} = 1\}$$

$$B_0 = \{r_{0L} + z_L < y_{t-1} \leq r_{0L}, R_{t-1} = 0\}$$

$$A_i = \{y_{t-1} \in (r_{0L}, r_{0U}], \dots, y_{t-i} \in (r_{0L}, r_{0U}], r_{0U} < y_{t-i-1} \leq r_{0U} + z_U, R_{t-i-1} = 1\}, i \geq 1$$

$$B_i = \{y_{t-1} \in (r_{0L}, r_{0U}], \dots, y_{t-i} \in (r_{0L}, r_{0U}], r_{0L} + z_L < y_{t-i-1} \leq r_{0L}, R_{t-i-1} = 0\}, i \geq 1$$

where $R_t \equiv R_t(r_{0L} + z_L, r_{0U} + z_U)$. Denote $A_t(z_L, z_U) = \bigcup_{j=0}^{\infty} A_j, B_t(z_L, z_U) = \bigcup_{j=0}^{\infty} B_j$

As in Li, Guan, Li, and Yu (2015), we have:

$$R_t(r_{0L} + z_L, r_{0U} + z_U) - R_t(r_{0L}, r_{0U}) = 1\{A_t(z_L, z_U)\} - 1\{B_t(z_L, z_U)\}$$

Moreover, $B_t(z_L, z_U) \subset \{R_t(r_{0L}, r_{0U}) = 1\}, A_t(z_L, z_U) \subset \{R_t(r_{0L}, r_{0U}) = 0\}$. As a result,

$$\begin{aligned} & \mathbf{L}_n(\lambda, r_{0L} + z_L, r_{0U} + z_U) - \mathbf{L}_n(\lambda, r_{0L}, r_{0U}) \\ &= \sum_{t=1}^n (l_{1t}(\theta) - l_{2t}(\theta)) [1\{A_t(z_L, z_U)\} - 1\{B_t(z_L, z_U)\}] \\ &= \sum_{t=1}^n (l_{1t}(\theta) - l_{2t}(\theta)) 1\{A_t(z_L, z_U)\} + \sum_{t=1}^n (l_{2t}(\theta) - l_{1t}(\theta)) 1\{B_t(z_L, z_U)\} \\ &= \mathbf{L}_{1n}(z_L, z_U) + \mathbf{L}_{2n}(z_L, z_U) \end{aligned}$$

Our next step is to show that

$$\Pr(\mathbf{L}_{1n}(z_L, z_U) > 0) > 1 - \epsilon \quad (6.6)$$

when $\lambda \in w(\Delta)$, $-\Delta < z_L < 0$, $\frac{K}{n} < z_U < \Delta$.

Define $Q_{z_L}(z_U) = E[1\{A_t(z_L, z_U)\}]$

$$Q_{n,z_L}(z_U) = \frac{1}{n} \sum_{t=1}^n 1\{A_t(z_L, z_U)\}$$

By a calculation, we can obtain

$$\begin{aligned} \frac{\mathbf{L}_{1n}(z_L, z_U)}{nQ_{z_L}(z_U)} &\equiv \frac{\sum_{t=1}^n [l_{1t}(\theta) - l_{2t}(\theta)][1\{A_t(z_L, z_U)\}]}{nQ_{z_L}(z_U)} \\ &= \frac{\sum_{t=1}^n [\log \frac{\alpha'_{10}\mathbf{X}_{t-1}}{\alpha'_{20}\mathbf{X}_{t-1}} + \frac{\alpha'_{20}\mathbf{X}_{t-1}}{\alpha'_{10}\mathbf{X}_{t-1}} - 1 + \frac{\{(\phi_{20} - \phi_{10})'\mathbf{Y}_{t-1}\}^2}{\alpha'_{10}\mathbf{X}_{t-1}}] 1\{A_t(z_L, z_U)\}]}{nQ_{n,z_L}(z_U)} \\ &\quad + \frac{\sum_{t=1}^n \frac{2(\phi_{20} - \phi_{10})'\mathbf{Y}_{t-1}\sqrt{\alpha'_{20}\mathbf{X}_{t-1}}}{\alpha'_{10}\mathbf{X}_{t-1}} \epsilon_t 1\{A_t(z_L, z_U)\}}{nQ_{n,z_L}(z_U)} \\ &\quad + \frac{\sum_{t=1}^n \frac{(\alpha_{20} - \alpha_{10})'\mathbf{X}_{t-1}}{\alpha'_{10}\mathbf{X}_{t-1}} (\epsilon_t^2 - 1) 1\{A_t(z_L, z_U)\}}{nQ_{n,z_L}(z_U)} + Op(\Delta) \end{aligned}$$

Note in the first term,

$$\log \frac{\alpha'_{10}\mathbf{X}_{t-1}}{\alpha'_{20}\mathbf{X}_{t-1}} + \frac{\alpha'_{20}\mathbf{X}_{t-1}}{\alpha'_{10}\mathbf{X}_{t-1}} - 1 + \frac{\{(\phi_{20} - \phi_{10})'\mathbf{Y}_{t-1}\}^2}{\alpha'_{10}\mathbf{X}_{t-1}} > 0$$

where $\mathbf{X}_{t-1} = (1, y_{t-1}^2)$, $\mathbf{Y}_{t-1} = (1, y_{t-1})$

In the second term, $\sum_{t=1}^n \frac{2(\phi_{20} - \phi_{10})'\mathbf{Y}_{t-1}\sqrt{\alpha'_{20}\mathbf{X}_{t-1}}}{\alpha'_{10}\mathbf{X}_{t-1}} \epsilon_t 1\{A_t(z_L, z_U)\}$ is bounded in absolute value by $K_1 \left| \sum_{t=1}^n \epsilon_t 1(A_t(z_L, z_U)) \right|$

For the third term, the numerator is bounded in absolute value by $K_2 \left| \sum_{t=1}^n (\epsilon_t^2 - 1) 1(A_t(z_L, z_U)) \right|$, where K_1 and K_2 are some constants independent of n .

Following the methond similar to Theorem 3 in Li, Guan, Li, and Yu (2015), we can verify the following three conditions: $\forall \epsilon > 0, \eta > 0, \exists$ a positive constant K such that as n is large

enough,

$$\begin{aligned}
& Pr\left[\sup_{\frac{k}{n} < z_U < \Delta, -\Delta < z_L \leq 0} \left| \frac{1}{nQ_{z_L}(z_U)} \sum_{t=1}^n 1\{A_t(z_L, z_U)\} - 1 \right| < \eta\right] > 1 - \epsilon \\
& Pr\left[\sup_{\frac{k}{n} < z_U < \Delta, -\Delta < z_L \leq 0} \left| \frac{1}{nQ_{z_L}(z_U)} \sum_{t=1}^n \epsilon_t 1\{A_t(z_L, z_U)\} \right| < \eta\right] > 1 - \epsilon \\
& Pr\left[\sup_{\frac{k}{n} < z_U < \Delta, -\Delta < z_L \leq 0} \left| \frac{1}{nQ_{z_L}(z_U)} \sum_{t=1}^n (\epsilon_t^2 - 1) 1\{A_t(z_L, z_U)\} \right| < \eta\right] > 1 - \epsilon
\end{aligned}$$

Therefore, (6.6) is implied. We can further show a similar result for $\mathbf{L}_{2n}(z_L, z_U)$, and then that for $L_n(\boldsymbol{\lambda}, r_{0L} + z_L, r_{0U} + z_U) - L_n(\boldsymbol{\lambda}, r_{0L}, r_{0U})$. (6.5) is thus proved for the case that $p = d = 1$, $z_L < 0$ and $z_U > 0$.

The proof for other cases is similar and hence is omitted. The proof of Claim (ii) in this theorem is similar to that of Theorem 2.2 in Li, Ling, and Li (2013) and that of Theorem 3.2 in Li, Ling, and Zakoian (2015), and is also omitted.

6.4 Proof of Theorem 4

From Theorem 3, we know $n(\hat{r}_L - r_{0L}) = O_p(1)$ and $n(\hat{r}_U - r_{0U}) = O_p(1)$. To further characterize their limiting distributions, we consider the limiting behavior of a sequence of normalized profile objective function defined by:

$$\tilde{\mathbf{L}}_n(z) = \mathbf{L}_n(\tilde{\boldsymbol{\lambda}}_n(r_0 + \frac{z}{n}), r_0 + \frac{z}{n}) - \mathbf{L}_n(\tilde{\boldsymbol{\lambda}}_n(r_0), r_0)$$

where $r_0 = (r_{0L}, r_{0U})^T$ and $z = (z_L, z_U)^T$

$$\begin{aligned}
\text{Denote } \varphi_n(z) = & \sum_{t=1} \xi_{1t} \{1 - R_{t-1}(r_0)\} 1\{r_{0L} < y_{t-d} \leq r_{0L} + \frac{z_L}{n}\} 1(z_L \geq 0) \\
& + \sum_{t=1} \xi_{2t} \{1 - R_{t-1}(r_0)\} 1\{r_{0L} + \frac{z_L}{n} < y_{t-d} \leq r_{0L}\} 1(z_L < 0) \\
& + \sum_{t=1} \xi_{1t} R_{t-1}(r_0) 1\{r_{0U} < y_{t-d} \leq r_{0U} + \frac{z_U}{n}\} 1(z_U \geq 0) \\
& + \sum_{t=1} \xi_{2t} R_{t-1}(r_0) 1\{r_{0U} + \frac{z_U}{n} < y_{t-d} \leq r_{0U}\} 1(z_U < 0)
\end{aligned}$$

$$\begin{aligned}
\text{where } \xi_{1t} &= \sum_{j=0}^{\infty} \zeta_{1,t+j} H_{t+j,j} \\
\xi_{2t} &= \sum_{j=0}^{\infty} \zeta_{2,t+j} H_{t+j,j} \\
\zeta_{1t} &= \log \frac{\alpha'_{10} \mathbf{X}_{t-1}}{\alpha'_{20} \mathbf{X}_{t-1}} + \frac{\{(\phi_{10} - \phi_{20})' \mathbf{Y}_{t-1} - \epsilon_t \sqrt{\alpha'_{20} \mathbf{X}_{t-1}}\}^2}{\alpha'_{10} \mathbf{X}_{t-1}} - \epsilon_t^2 \\
\zeta_{2t} &= \log \frac{\alpha'_{20} \mathbf{X}_{t-1}}{\alpha'_{10} \mathbf{X}_{t-1}} + \frac{\{(\phi_{10} - \phi_{20})' \mathbf{Y}_{t-1} + \epsilon_t \sqrt{\alpha'_{10} \mathbf{X}_{t-1}}\}^2}{\alpha'_{20} \mathbf{X}_{t-1}} - \epsilon_t^2 \\
H_{t,j} &= \prod_{l=1}^j 1(r_{0L} < y_{t-d+1-l} \leq r_{0U})
\end{aligned}$$

We first prove that, for any $B > 0$,

$$\sup_{\|z\| \leq B} |\tilde{\mathbf{L}}_n(z) - \varphi_n(z)| = o_p(1) \quad (6.7)$$

It is sufficient to verify the following two conditions:

$$\sup_{\|z\| \leq B} |\tilde{\mathbf{L}}_n(z) - \{\mathbf{L}_n(\boldsymbol{\lambda}_0, r_0 + \frac{z}{n}) - \mathbf{L}_n(\boldsymbol{\lambda}_0, r_0)\}| = o_p(1) \quad (6.8)$$

$$\sup_{\|z\| \leq B} |\varphi_n(z) - \{\mathbf{L}_n(\boldsymbol{\theta}_0, r_0 + \frac{z}{n}) - \mathbf{L}_n(\boldsymbol{\theta}_0, r_0)\}| = o_p(1) \quad (6.9)$$

By a method similar to similar to the proof of Theorem 4 in Li, Guan, Li, and Yu (2015), we can use Taylor's expansion and results of Theorem 3 in current paper to show that (6.8) holds. Below we will verify (6.9).

Denote $R_t(z) = R_t(r_0 + \frac{z}{n})$ for simplicity. As in Li, Guan, Li, and Yu (2015), we have:

$$\begin{aligned}
R_t(z) - R_t(0) &= \{R_{t-1}(z) - R_{t-1}(0)\}1(r_{0L} < y_{t-d} \leq r_{0U}) \\
&+ \{1 - R_{t-1}(z)\}1(r_{0L} < y_{t-d} \leq r_{0L} + \frac{z_L}{n})1(z_L \geq 0) \\
&- \{1 - R_{t-1}(z)\}1(r_{0L} + \frac{z_L}{n} < y_{t-d} \leq r_{0L})1(z_L < 0) \\
&+ R_{t-1}(z)1(r_{0U} < y_{t-d} \leq r_{0U} + \frac{z_U}{n})1(z_U \geq 0) \\
&- R_{t-1}(z)1(r_{0U} + \frac{z_U}{n} < y_{t-d} \leq r_{0U})1(z_U < 0)
\end{aligned}$$

$$\begin{aligned}
\text{and then } R_t(z) - R_t(0) &= \sum_{j=0}^{\infty} H_{t,j} \{1 - R_{t-1-j}(z)\}1(r_{0L} < y_{t-d-j} \leq r_{0L} + \frac{z_L}{n})1(z_L \geq 0) \\
&- \sum_{j=0}^{\infty} H_{t,j} \{1 - R_{t-1-j}(z)\}1(r_{0L} + \frac{z_L}{n} < y_{t-d-j} \leq r_{0L})1(z_L < 0) \\
&+ \sum_{j=0}^{\infty} H_{t,j} R_{t-1-j}(z)1(r_{0U} < y_{t-d-j} \leq r_{0U} + \frac{z_U}{n})1(z_U \geq 0) \\
&- \sum_{j=0}^{\infty} H_{t,j} R_{t-1-j}(z)1(r_{0U} + \frac{z_U}{n} < y_{t-d-j} \leq r_{0U})1(z_U < 0)
\end{aligned} \tag{6.10}$$

where $H_{t,j} = \prod_{l=1}^j 1(r_{0L} < y_{t-d+1-l} \leq r_{0U})$

Denote $\zeta_t = l_{1t}(\theta_0) - l_{2t}(\theta_0)$

$$= \log \alpha'_{10} X_{t-1} + \frac{(y_t - \phi'_{10} Y_{t-1})^2}{\alpha'_{10} X_{t-1}} - \log \alpha'_{20} X_{t-1} - \frac{(y_t - \phi'_{20} Y_{t-1})^2}{\alpha'_{20} X_{t-1}}$$

and it holds that $L_n(\lambda_0, r_0 + \frac{z}{n}) - L_n(\lambda_0, r_0) = \sum_{t=1}^n \zeta_t \{R_t(z) - R_t(0)\}$.

For the first term of (6.10), i.e., the case with $z_L \geq 0$, by a method similar to Li, Guan, Li, and Yu (2015), it can be shown that:

$$\sup_{||z|| \leq B} \left\| \sum_{t=1}^n \zeta_t \sum_{j=0}^{\infty} H_{t,j} \{R_{t-1-j}(z) - R_{t-1-j}(0)\}1(r_{0L} < y_{t-d-j} \leq r_{0L} + \frac{z_L}{n}) \right\| = o_p(1)$$

Note $\zeta_t = \zeta_{1t}$ when $R_t(0) = 0$ and $\zeta_t = -\zeta_{2t}$ when $R_t(0) = 1$. Thus, we have

$$\begin{aligned}
& \sum_{t=1}^n \zeta_t \sum_{j=0}^{\infty} H_{t,j} \{1 - R_{t-1-j}(0)\} 1(r_{0L} < y_{t-d-j} \leq r_{0L} + \frac{z_L}{n}) \\
&= \sum_{j=0}^{\infty} \sum_{t=1-j}^{n-j} \zeta_{t+j} H_{t+j,j} \{1 - R_{t-1}(0)\} 1(r_{0L} < y_{t-d} \leq r_{0L} + \frac{z_L}{n}) \\
&= \sum_{j=0}^{\infty} (\sum_{t=1}^n + \sum_{t=1-j}^0 - \sum_{t=n-j+1}^n) \zeta_{t+j} H_{t+j,j} \{1 - R_{t-1}(0)\} 1(r_{0L} < y_{t-d} \leq r_{0L} + \frac{z_L}{n}) \\
&= \sum_{j=0}^{\infty} \sum_{t=1-j}^n \zeta_{t+j} H_{t+j,j} \{1 - R_{t-1}(0)\} 1(r_{0L} < y_{t-d} \leq r_{0L} + \frac{z_L}{n}) + o_p(1) \\
&= \sum_{t=1-j}^n \xi_{1t} \{1 - R_{t-1}(0)\} 1(r_{0L} < y_{t-d} \leq r_{0L} + \frac{z_L}{n}) + o_p(1)
\end{aligned}$$

where the $o_p(1)$ is uniformly for $\|z\| \leq B$. See also the proof of Theorem 4 in Li, Guan, Li, and Yu (2015) and Lemma 8.2 in Li, Ling, and Li (2013).

Similarly, we can handle the remaining terms in (6.10). Therefore, (6.9) is implied. With (6.8) and (6.9), it is clear that (6.7) holds.

For the space $\mathbb{D}(\mathbb{R}^2)$, we define the skorohod metric as $d(f, g) = \sum_{k=1}^{\infty} 2^{-k} \min\{1, d_k(f, g)\}$ for $f, g \in \mathbb{D}(\mathbb{R}^2)$, where $d_k(f, g)$ is the skorohod metric on $\mathbb{D}([-k, k] \times [-k, k])$, see also Li, Guan, Li, and Yu (2015), Li and Ling (2012) and (16.4) in Billingsley (1999).

By a technique similar that used in the proof of Theorem 3.3 in Li and Ling (2012) and in the proof of Theorem 5 in Li, Ling, and Zhang (2016), together with Theorem 5.5 in Straf (1972), we are able to conclude that $\{\varphi_n(z), z \in \mathbb{R}^2\}$ converges weakly to $\{\varphi(z), z \in \mathbb{R}^2\}$ as $n \rightarrow \infty$.

Moreover, (6.7) implies that $d(\tilde{L}_n(z), \varphi_n(z)) \rightarrow 0$ in probability as $n \rightarrow \infty$. By Theorem 3.1 in Seijo and Sen (2011), it is readily seen that $n(\hat{r}_L - r_{0L}) \rightarrow M_-^{(L)}$ and $n(\hat{r}_U - r_{0U}) \rightarrow M_-^{(U)}$ in distribution as $n \rightarrow \infty$ respectively. The remainder of the proof is similar to that of Theorem 2 in Chan (1993).

6.5 Proof of Theorem 5

From Theorem 2 and 3, \hat{r}_L and \hat{r}_U are super-consistent, and \hat{d} is consistent with integer values. Therefore, it can be assumed that the true values of (r_L, r_U, d) are known, indicating the true Regime indicators R_t are known.

By definition,

$$\begin{aligned}
\text{BIC}(p) &= \sum_{t=1}^n l_t(\hat{\theta}) + (2p+2) \log n_1 + (2p+2) \log n_2 \\
&= \sum_{t=1}^n l_{1t}(\hat{\theta}) R_t + \sum_{t=1}^n l_{2t}(\hat{\theta})(1-R_t) + (2p+2) \log n_1 + (2p+2) \log n_2 \\
&= \sum_{t=1}^n l_{1t}(\hat{\lambda}_1) R_t + \sum_{t=1}^n l_{2t}(\hat{\lambda}_2)(1-R_t) + (2p+2) \log n_1 + (2p+2) \log n_2
\end{aligned}$$

where $l_{1t}(\theta) = l_{1t}(\lambda_1)$ and $l_{2t}(\theta) = l_{2t}(\lambda_2)$ are defined at the beginning of the proof of Theorem 2.

We first consider the case $p > p_0$. Below we introduce notations λ^p, λ_0^p and $\hat{\lambda}^p$ to emphasize their dependence on the order p . Under the situation $p > p_0$, the model with order p corresponds to a bigger model, and we have

$$l_{1t}(\lambda_{10}^p) = l_{1t}(\lambda_{10}^{p_0}), l_{2t}(\lambda_{20}^p) = l_{2t}(\lambda_{20}^{p_0})$$

By an intermediate result in the proof of Theorem 4 in Li et al. (2017), we know

$$\begin{aligned}
&\sum_{t=1}^n l_{1t}(\hat{\lambda}_1^{p_0}) R_t - \sum_{t=1}^n l_{1t}(\lambda_{10}^{p_0}) R_t = O_p(1) \\
&\sum_{t=1}^n l_{1t}(\hat{\lambda}_1^p) R_t - \sum_{t=1}^n l_{1t}(\lambda_{10}^p) R_t = O_p(1) \\
&\sum_{t=1}^n l_{2t}(\hat{\lambda}_2^{p_0})(1-R_t) - \sum_{t=1}^n l_{2t}(\lambda_{20}^{p_0})(1-R_t) = O_p(1) \\
&\sum_{t=1}^n l_{2t}(\hat{\lambda}_2^p)(1-R_t) - \sum_{t=1}^n l_{2t}(\lambda_{20}^p)(1-R_t) = O_p(1)
\end{aligned}$$

As a result,

$$\begin{aligned}
\sum_{t=1}^n l_{1t}(\hat{\lambda}_1^p) R_t - \sum_{t=1}^n l_{1t}(\hat{\lambda}_1^{p_0}) R_t &= \left[\sum_{t=1}^n l_{1t}(\hat{\lambda}_1^p) R_t - \sum_{t=1}^n l_{1t}(\lambda_{10}^p) R_t \right] \\
&\quad - \left[\sum_{t=1}^n l_{1t}(\hat{\lambda}_1^{p_0}) R_t - \sum_{t=1}^n l_{1t}(\lambda_{10}^{p_0}) R_t \right] \\
&\quad + \left[\sum_{t=1}^n l_{1t}(\lambda_{10}^p) R_t \right] - \sum_{t=1}^n l_{1t}(\lambda_{10}^{p_0}) R_t = O_p(1) \\
\sum_{t=1}^n l_{2t}(\hat{\lambda}_2^p)(1-R_t) - \sum_{t=1}^n l_{2t}(\hat{\lambda}_2^{p_0})(1-R_t) &= \left[\sum_{t=1}^n l_{2t}(\hat{\lambda}_2^p)(1-R_t) - \sum_{t=1}^n l_{2t}(\lambda_{20}^p)(1-R_t) \right] \\
&\quad - \left[\sum_{t=1}^n l_{2t}(\hat{\lambda}_2^{p_0})(1-R_t) - \sum_{t=1}^n l_{2t}(\lambda_{20}^{p_0})(1-R_t) \right] \\
&\quad + \left[\sum_{t=1}^n l_{2t}(\lambda_{20}^p)(1-R_t) - \sum_{t=1}^n l_{2t}(\lambda_{20}^{p_0})(1-R_t) \right] \\
&= O_p(1)
\end{aligned}$$

Therefore, $\text{BIC}(p) - \text{BIC}(p_0) = O_p(1) + 2(p - p_0) \log n_1 n_2 \rightarrow \infty$ as $n \rightarrow \infty$, since $\frac{n_1}{n} \rightarrow \Pr(R_t = 1) > 0$ and $\frac{n_2}{n} \rightarrow \Pr(R_t = 0) > 0$ as $n \rightarrow \infty$.

We then consider the case $p < p_0$. Again by an intermediate result in the proof of Theorem 4 in Li et al. (2017), we know

$$\begin{aligned}\sum_{t=1}^n \mathbf{l}_{1t}(\hat{\boldsymbol{\lambda}}_1^p) R_t - \sum_{t=1}^n \mathbf{l}_{1t}(\hat{\boldsymbol{\lambda}}_1^{p_0}) R_t &= O_p(1) + o_p(n_1) + c_1 n_1 \\ \sum_{t=1}^n \mathbf{l}_{2t}(\hat{\boldsymbol{\lambda}}_2^p)(1 - R_t) - \sum_{t=1}^n \mathbf{l}_{2t}(\hat{\boldsymbol{\lambda}}_2^{p_0})(1 - R_t) &= O_p(1) + o_p(n_2) + c_2 n_2\end{aligned}$$

where c_1, c_2 are positive constants, defined in the same way as constant c in Li et al. (2017).

$$\begin{aligned}\text{Therefore, } \text{BIC}(p) - \text{BIC}(p_0) &= c_1 n_1 + c_2 n_2 + O_p(1) + o_p(n_1) + o_p(n_2) + 2(p - p_0) \ln n_1 n_2 \\ &= c_3 n + O_p(1) + o_p(n) + O(\ln n)\end{aligned}$$

$$\rightarrow \infty \text{ as } n \rightarrow \infty$$

since constant $c_3 > 0$, $\frac{n_1}{n} \rightarrow \Pr(R_t = 1) > 0$ and $\frac{n_2}{n} \rightarrow \Pr(R_t = 0) > 0$ as $n \rightarrow \infty$.

References

- Billingsley, P. (1999). *Convergence of Probability Measures*. New York: Wiley, 2nd ed.
- 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.
- Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, Berkeley, pp. 221–233. University of California Press.
- Lee, O. (2006). Stationarity and β -mixing property of a mixture ar-arch models. *Bull. Korean Math. Soc* 43, 813–820.
- Li, D. and S. Ling (2012). On the least squares estimation of multiple-regime threshold autoregressive models. *Journal of Econometrics* 167, 240–253.
- Li, D., S. Ling, and K. Li, W (2013). Asymptotic theory on the least squares estimation of threshold moving-average model. *Econometric Theory* 29, 482–516.
- Li, D., S. Ling, and J. Zakoian (2015). Asymptotic inference in multiple-threshold double autoregressive models. *Journal of Econometrics* 189, 415–427.
- Li, D., S. Ling, and R. Zhang (2016). On a threshold double autoregressive model. *Journal of Business and Economic Statistics* 34, 68–80.
- Li, G., B. Guan, W. K. Li, and P. L. H. Yu (2015). Hysteretic autoregressive time series models. *Biometrika* 102, 717–723.
- Li, G., Q. Zhu, Z. Liu, and W. K. Li (2017). On mixture double autoregressive time series models. *Journal of Business and Economic Statistics* 35, 306–317.
- Ling, S. (2004). Estimation and testing stationarity for double-autoregressive models. *Journal of the Royal Statistical Society, Series B* 66, 63–78.
- Ling, S. (2007). A double AR(p) model: structure and estimation. *Statistica Sinica* 17, 161–175.
- Lo, P., W. Li, and G. Li (2016). On buffered threshold garch model. *Statistica Sinica* 26, 1555–1567.
- Lu, Z. (1998). On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Statistica Sinica* 8, 1205–1217.
- McLeod, A. and W. K. Li (1983). Diagnostic checking arma time series models using squared residual autocorrelations. *Journal of Time Series Analysis* 4, 269–273.

- Seijo, E. and B. Sen (2011). A continuous mapping theorem for the smallest argmax functional. *Electronic Journal of Statistics* 5, 421–439.
- Straf, M. L. (1972). Weak convergence of stochastic processes with several parameters. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. II: Probability Theory, Berkeley, California*.
- Tong, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.
- Tsay, R. S. (1986). Nonlinearity test for time series. *Biometrika* 73, 461–466.
- Tse, Y. K. (2002). Residual-based diagnostics for conditional heteroscedasticity models. *The Econometrics Journal* 5, 358–373.
- Tweedie, R. L. (1983). Criteria for rates of convergence of markov chains, with application to queueing and storage theory. In J. F. C. Kingman and G. E. H. Reuter (Eds.), *Probability, Statistics and Analysis*. Cambridge: Cambridge University Press.
- Zhu, K. and S. Ling (2013). Quasi-maximum exponential likelihood estimators for a double AR(p) model. *Statistica Sinica* 23, 251–270.
- Zhu, K., P. L. H. Yu, and W. K. Li (2014). Testing for the buffered autoregressive processes. *Statistica Sinica* 24, 971–984.
- Zhu, K., P. L. H. Yu, and W. K. Li (2017). Buffered autorgressive models with conditional heteroscedasticity: An application to exchange ratess. *Journal of Businees and Economic Statistics* 35, 528–542.

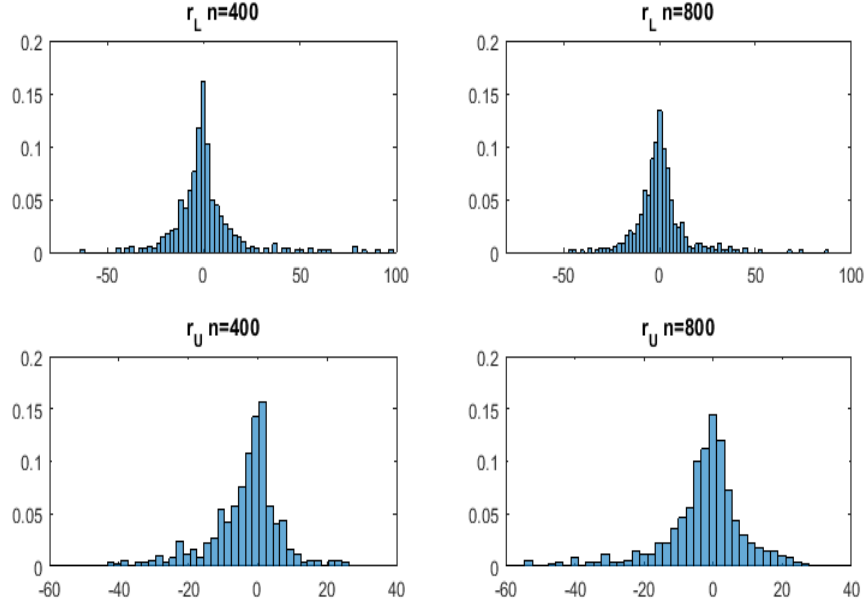


Figure 1: Histograms of random variables $n(\hat{r}_L - r_L)$ (upper panels) and $n(\hat{r}_U - r_U)$ (lower panels)



Figure 2: Time plots of weekly closing prices of Hang Seng Index (HSI) from January 2000 to December 2007(upper panel) and corresponding log returns(lower panel)

Table 1: Simulation results for buffered double autoregressive model

n		1st regime						2nd regime						r_U	
		ϕ_{10}	ϕ_{11}	ϕ_{12}	α_{10}	α_{11}	α_{12}	ϕ_{20}	ϕ_{21}	ϕ_{22}	α_{20}	α_{21}	α_{22}		r_L
400	Bias	0.0020	-0.0013	-0.0124	-0.0024	-0.0102	0.0054	-0.0039	-0.0008	-0.0041	-0.0010	0.0010	-0.0045	0.0013	-0.0102
	ESD	0.0288	0.0883	0.0765	0.0165	0.1161	0.0657	0.0225	0.0835	0.0680	0.0103	0.0959	0.0693	0.0421	0.0314
	ASD	0.0255	0.0798	0.0723	0.0152	0.1057	0.0719	0.0197	0.0681	0.0627	0.0086	0.0741	0.0571		
800	Bias	0.0028	-0.0025	-0.0030	-0.0005	-0.0037	-0.0030	-0.0004	-0.0055	-0.0020	-0.0005	0.0007	-0.0039	0.0005	-0.0039
	ESD	0.0188	0.0616	0.0529	0.0112	0.0855	0.0471	0.0155	0.0503	0.0467	0.0068	0.0631	0.0494	0.0173	0.0165
	ASD	0.0184	0.0573	0.0518	0.0113	0.0786	0.0506	0.0142	0.0486	0.0448	0.0063	0.0551	0.0414		

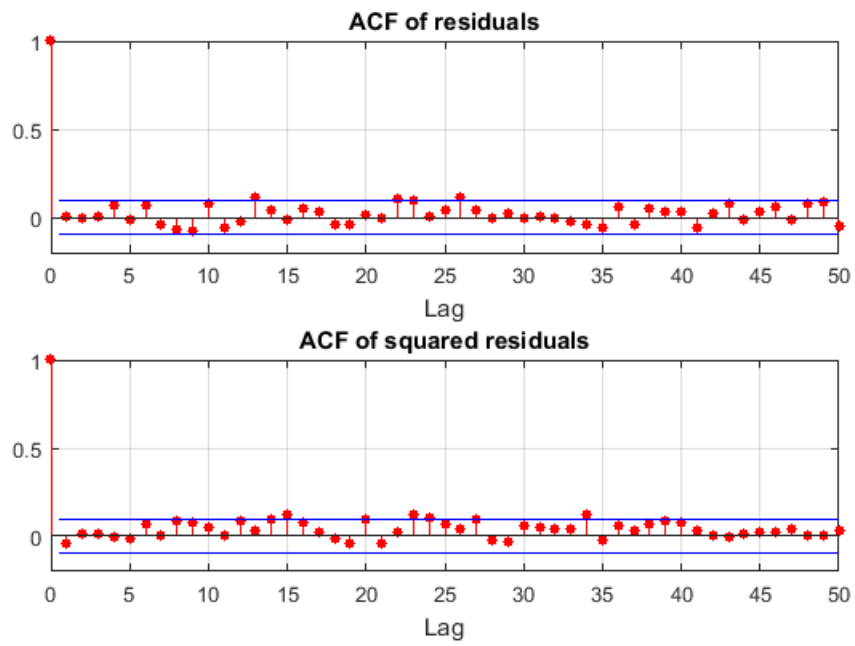


Figure 3: Sample ACFs of residuals and squared residuals of fitted buffered double autoregressive model in empirical analysis