Theory and Applications of TAR Model with Two Threshold Variables

Chen, Haiqiang and Chong, Terence Tai Leung and Bai, Jushan

Xiamen University, The Chinese University of Hong Kong, Columbia University

1 January 2012

Theory and Applications of TAR Model with Two Threshold Variables

by

Haiqiang Chen, Terence Tai-Leung Chong and Jushan Bai*

1/1/2012

Abstract

A growing body of threshold models has been developed over the past two decades to capture the nonlinear movement of financial time series. Most of these models, however, contain a single threshold variable only. In many empirical applications, models with two or more threshold variables are needed. This paper develops a new threshold autoregressive model which contains two threshold variables. A likelihood ratio test is proposed to determine the number of regimes in the model. The finite-sample performance of the estimators is evaluated and an empirical application is provided.

JEL Classification: C22

Keywords: Threshold Autoregressive Model, Misspecification, Likelihood Ratio Test, Bootstrapping.

1 Introduction

A growing body of threshold models has been developed over the past two decades to capture the nonlinear movement of financial time series. Tong (1983) develops a

*We would like to thank Patrik Guggenberger, Guido Kuersteiner, Yongmiao Hong and the seminar participants of Tsinghua University and Nanyang Technological University for suggestions. All the remaining errors are ours. Haiqiang Chen, Department of Economics, Xiamen University. Jushan Bai, Department of Economics, Columbia University. Corresponding Author: Terence Chong, Department of Economics and Institute of Global Economics and Finance, The Chinese University of Hong Kong, Hong Kong. Email: chong2064@cuhk.edu.hk.

threshold autoregressive (TAR) model and uses it to predict stock price movements. A number of new models have been proposed since the seminal work of Tong (1983), including the smooth transition threshold autoregressive model (STAR) of Chan and Tong (1986) and the functional-coefficient autoregressive (FAR) model of Chen and Tsay (1993). Tsay (1998) develops a multivariate TAR model for the arbitrage activities in the security market. Dueker et al. (2007) develop a contemporaneous TAR model for the bond market.

Most of the aforementioned models, however, contain a single threshold variable only. In many empirical applications, a model with two or more threshold variables is more appropriate. For example, Leeper (1991) divides the policy parameter space into four disjoint regions according to whether monetary and fiscal policies are active or passive. Given these policy combinations, macroeconomic variables, such as real output, inflation and unemployment have different dynamics. Tiao and Tsay (1994) divide the U.S. quarterly real GNP growth rate into four regimes according to the level and sign of the past growth rate. Durlauf. and Johnson (1995) split that cross-country GDP growth rate into different regimes according to the level of per capita real GDP and literacy rate. In modelling currency crises, Sachs et al. (1996), Frankel and Rose (1996), Kaminsky (1998) and Edison (2000) argue that the occurrence of currency crises hints at the values of fiscal reserves, foreign reserves and interest rate differential between home countries and the U.S.. In these examples, TAR models with multiple threshold variables can be used to describe the dynamics of different regimes.[1]

As the distributional theory is rather involved, no asymptotic result has been developed for TAR models with multiple threshold variables.[2] This paper contributes to the literature by developing estimation and inference procedures for TAR models with two threshold variables. Our model is applied to identify the regimes of the Hong Kong stock market. The case of Hong Kong is of interest because of its rising role as a global financial center. In 2006, Hong Kong becomes the world's second most popular place for IPO after London. In 2007, the Hong Kong stock market ranks fifth in the world, while its warrant market ranks top worldwide in terms of turnover. Using the historical prices of the Hang Seng index and the market turnover as threshold variables, our estimation shows that the stock market of Hong Kong can be classified into a high-return stable regime, a low-return volatile regime and a neutral regime.

[1]Threshold model with two threshold variables can also be applied to the cross section of financial data. For example, in the Fama and French (1992) model, one may use firm size and book-to-market ratio as threshold variables to explain abnormal returns of a stock. Avramov et al. (2006) also sort stocks into different categories according to historical returns and liquidity level.

[2]A related empirical study is the nested threshold autoregressive (NeTAR) models of Astatkie et al. (1997).

This is different from the conventional bull-bear classification.

The remainder of the paper is organized as follows: Section 2 presents the model and discusses the estimation procedure. Section 3 derives the limiting distribution of the threshold estimators. Section 4 proposes a likelihood ratio test to determine the number of regimes. Monte Carlo simulations are conducted and the performance of the estimation procedure is evaluated in Section 5. An empirical application is provided in Section 6. Section 7 concludes the paper.

2 TAR Model with Two Threshold Variables

Consider the following TAR model with two threshold variables which classifies the observations y_t into four regimes:

$$y_t = \begin{cases} \beta_0^{(1)} + \beta_1^{(1)} y_{t-1} + \beta_2^{(1)} y_{t-2}, ..., + \beta_{p_1}^{(1)} y_{t-p_1} + u_t, \text{when } z_{1t} \leq \gamma_1^0, z_{2t} \leq \gamma_2^0 \\ \beta_0^{(2)} + \beta_1^{(2)} y_{t-1} + \beta_2^{(2)} y_{t-2}, ..., + \beta_{p_2}^{(2)} y_{t-p_2} + u_t, \text{ when } z_{1t} \leq \gamma_1^0, z_{2t} > \gamma_2^0 \\ \beta_0^{(3)} + \beta_1^{(3)} y_{t-1} + \beta_2^{(3)} y_{t-2}, ..., + \beta_{p_3}^{(3)} y_{t-p_3} + u_t, \text{ when } z_{1t} > \gamma_1^0, z_{2t} \leq \gamma_2^0 \\ \beta_0^{(4)} + \beta_1^{(4)} y_{t-1} + \beta_2^{(4)} y_{t-2}, ..., + \beta_{p_4}^{(4)} y_{t-p_4} + u_t, \text{ when } z_{1t} > \gamma_1^0, z_{2t} > \gamma_2^0 \end{cases}, \quad (1)$$

where

$z_t \overset{def}{=} (z_{1t}, z_{2t})$ are the threshold variables;

$\gamma^0 \overset{def}{=} (\gamma_{1,}^0 \gamma_2^0) \in \Omega$ where $\Omega = [\underline{\gamma_1}, \overline{\gamma_1}] \times [\underline{\gamma_2}, \overline{\gamma_2}]$ is a strict subset of the support of z_t. γ^0 is the threshold parameter vector pending to be estimated;

p_j $(j = 1, 2, 3, 4)$ is the order in each regime;

$\beta^{(j)} \overset{def}{=} (\beta_0^{(j)}, \beta_1^{(j)}, \beta_2^{(j)}, ..., \beta_{p_j}^{(j)})'$ are the structural parameters and $\beta^{(i)} \neq \beta^{(j)}$ for some $i \neq j$.[3]

The model is a linear AR model within each regime.[4] The threshold variables z_1 and z_2 can be exogenous variables or functions of the lags of y_t.[5] Given $\{y_t, z_t\}_{t=1}^T$, our

[3]Restrictions on the structural parameters can be imposed so that there are less than four regimes. For example, if $\beta^{(1)} = \beta^{(2)} = \beta^{(3)}$, the model will have two regimes only.

[4]An empirical example of the Model (1) is Tiao and Tsay (1994)'s four-regime TAR model for quarterly U.S. real GNP growth rates y_t:

$$y_t = \begin{cases} -0.015 - 1.076 y_{t-1} + \varepsilon_{1t}, \ y_{t-1} \leq y_{t-2} \leq 0 \\ 0.63 y_{t-1} - 0.76 y_{t-2} + \varepsilon_{2t}, \ y_{t-1} > y_{t-2}, y_{t-2} \leq 0 \\ 0.006 + 0.43 y_{t-1} + \varepsilon_{3t}, \ y_{t-1} \leq y_{t-2}, y_{t-2} > 0 \\ 0.433 y_{t-1} + \varepsilon_{4t}, \ y_{t-1} > y_{t-2} > 0 \end{cases}.$$

In their model, the process is divided into four regimes by $z_{1t} = y_{t-2}$ and $z_{2t} = y_{t-1} - y_{t-2}$, and the threshold values are set to zero. In practice, we need to estimate the threshold values.

[5]The model is a Self-Exciting Threshold Autoregressive (SETAR) model if the threshold variable is y_{t-d}.

objective is to estimate the threshold parameters γ^0 and the structural parameters $\beta^{(j)}$. Without loss of generality, we let $p = \max\{p_1, p_2, p_3, p_4\}$, and $\beta_q^{(j)} = 0$ when $q > p_j$, $j = 1, 2, 3, 4$. The model can be rewritten as

$$y_t = \sum_{j=1}^{4} \Psi_t^{(j)}\left(\gamma^0\right)(\beta_0^{(j)} + \sum_{i=1}^{p} \beta_i^{(j)} y_{t-i} + u_t), \tag{2}$$

where

$\Psi_t^{(j)}\left(\gamma^0\right)$ is an indicator function which equals one when the threshold condition is satisfied, and equals zero otherwise. Specifically,

$$\begin{aligned}
\Psi_t^{(1)}\left(\gamma^0\right) &= I\left(z_{1t} \leq \gamma_1^0, z_{2t} \leq \gamma_2^0\right); \\
\Psi_t^{(2)}\left(\gamma^0\right) &= I\left(z_{1t} \leq \gamma_1^0, z_{2t} > \gamma_2^0\right); \\
\Psi_t^{(3)}\left(\gamma^0\right) &= I\left(z_{1t} > \gamma_1^0, z_{2t} \leq \gamma_2^0\right); \\
\Psi_t^{(4)}\left(\gamma^0\right) &= I\left(z_{1t} > \gamma_1^0, z_{2t} > \gamma_2^0\right).
\end{aligned}$$

For analytical reasoning, it is convenient to rewrite the model (2) in the following matrix form:

$$Y = \sum_{j=1}^{4} I_j(\gamma^0) X \beta^{(j)} + U, \tag{3}$$

where

$$X = (x_T', x_{T-1}', ..., x_{p+1}')' = \begin{pmatrix} 1, y_{T-1}, & y_{T-2}, ..., y_{T-p} \\ 1, y_{T-2}, & y_{T-3}, ..., y_{T-p-1} \\ & .... \\ 1, \quad y_p, & y_{p-1}, ..., y_1 \end{pmatrix}_{(T-p)\times(p+1)},$$

$$x_t = (1, y_{t-1}, ..., y_{t-p})' \text{ for } t = p+1, ..., T.$$

$$I_j(\gamma^0) = diag\left\{\Psi_T^{(j)}\left(\gamma^0\right), \Psi_{T-1}^{(j)}\left(\gamma^0\right), ..., \Psi_{p+1}^{(j)}\left(\gamma^0\right)\right\},$$

$$Y = (y_T, y_{T-1}, ..., y_{p+1})',$$

$$U = (u_T, u_{T-1}, ..., u_{p+1})'.$$

We make the following assumptions:

$(A1)$ y_t is stationary ergodic and $E(y_t^4) < \infty$.

$(A2)$ $\{u_t\}$ is a sequence of i.i.d. normal errors with zero mean and finite variance σ^2.

$(A3)$ The threshold variables z_{1t} and z_{2t} are strictly stationary and have a continuous joint distribution $F(\gamma)$, which is differentiable with respect to both variables. Let $f(\gamma)$ denote the corresponding joint density function and $f_i(\gamma) = \dfrac{\partial F(\gamma)}{\partial \gamma_i}$. We assume that $0 < f(\gamma) \leq \overline{f} < \infty; 0 < f_i(\gamma) \leq \overline{f_i} < \infty$ for $i = 1, 2$.

$(A1)$ assumes that y_t is stationary ergodic, which allows us to apply the law of large number. A sufficient condition for $(A1)$ to hold is $\max_j \sum_i (|\beta_i^{(j)}|) < 1$.[6] $(A2)$ assumes that $\{u_t\}$ is a sequence of i.i.d. normal errors with finite second moment.[7] $(A3)$ requires the stationarity of the threshold variables. We also assume that the threshold variables are continuous with positive density everywhere, so that it is dense near γ^0 as the sample size increases. This assumption is needed for the consistent estimation of threshold values.

Given $\gamma = (\gamma_1, \gamma_2)$, the conditional least square (CLS) estimator for $\beta^{(j)}$ is defined as

$$\widehat{\beta}^{(j)}(\gamma) = (X' I_j(\gamma) X)^{-1} X' I_j(\gamma) Y, \qquad (j = 1, 2, 3, 4), \tag{4}$$

where

$$I_j(\gamma) = diag \left\{ \Psi_T^{(j)}(\gamma), \Psi_{T-1}^{(j)}(\gamma), ..., \Psi_{p+1}^{(j)}(\gamma) \right\}.$$

The residual sum of squares is

$$RSS_T(\gamma) = || \sum_{j=1}^{4} I_j(\gamma^0) X \beta^{(j)} + U - \sum_{j=1}^{4} I_j(\gamma) X \widehat{\beta}^{(j)}(\gamma) ||^2,$$

and we define the estimator of γ^0 as the value that minimizes $RSS_T(\gamma)$:

$$\widehat{\gamma} = \arg \min_{\gamma \in \Omega} RSS_T(\gamma). \tag{5}$$

---

[6]See Chan (1993) and Hansen (1997).

[7]In this paper, we generalize the TAR model to the one with two threshold variables. The error term u_t is assumed to be i.i.d. normal in order to derive the asymptotic distribution of the threshold estimators. We can relax this assumption and allow for heteroskedasticity of u_t. The estimators will still be consistent. See Hansen (1997) for more discussion on the heteroskedastic errors.

The structural estimators evaluated at the estimated threshold values are defined as:

$$\widehat{\beta}^{(j)}\left(\widehat{\gamma}\right) = (X' I_j(\widehat{\gamma}) X)^{-1} X' I_j(\widehat{\gamma}) Y, \tag{6}$$

Appendix 2 shows the consistency of the estimators $(\widehat{\gamma}, \widehat{\beta}^{(j)}\left(\widehat{\gamma}\right))$.

3 Limiting Distribution of $(\widehat{\gamma}_1, \widehat{\gamma}_2)$

In this section, the asymptotic joint distribution of the least-squares estimator $\widehat{\gamma}$ is derived under the assumption that the magnitude of change goes to zero at an appropriate rate. As pointed out by Hansen (2000), the assumption of decaying threshold effect is needed in order to obtain an asymptotic distribution of $\widehat{\gamma}$ free of nuisance parameters.[8] For notational simplicity, we rewrite Model (2) as:

$$Y = X\beta^{(1)} + \sum_{j=2}^{4} X_{\gamma^0}^{(j)} \delta^{(j)} + U, \tag{7}$$

where

$$X_{\gamma^0}^{(j)} = I_j\left(\gamma^0\right) X, \qquad (j = 2, 3, 4)$$

and

$$\delta^{(j)} = \beta^{(j)} - \beta^{(1)}, \qquad (j = 2, 3, 4)\,.$$

For any given γ, we define

$$X_{\gamma}^{(j)} = I_j\left(\gamma\right) X, \qquad (\, j = 1, 2, 3, 4)\,.$$

Observe that $X_{\gamma}^{(i)\prime} X_{\gamma}^{(j)} = 0$ if $i \neq j$, and $X' X_{\gamma}^{(j)} = X_{\gamma}^{(j)\prime} X_{\gamma}^{(j)}$.
Let $X_0^{(j)} = X_{\gamma^0}^{(j)}$, we have

$$X = \sum_{j=1}^{4} X_0^{(j)} = \sum_{j=1}^{4} X_{\gamma}^{(j)}.$$

We define the following conditional moment functionals:

$$D\left(\gamma\right) = E\left(x_t x_t' | z_t = \gamma\right), \tag{8}$$

$$V\left(\gamma\right) = E\left(x_t x_t' u_t^2 | z_t = \gamma\right). \tag{9}$$

Let $D = D(\gamma_0)$, $V = V(\gamma_0)$. Under the assumption $(A2)$, $V = \sigma^2 D$. We define block diagonal matrices $D^* = diag\{D, D\}$ and $V^* = diag\{V, V\}$.[9] We also need the

[8]This approach is first used in the literature of change points (Bai, 1997) and applied to threshold model by Hansen (2000).

[9]Note that D and V are $p \times p$ matrices and D^* and V^* are $2p \times 2p$ matrices.

following assumptions before the limiting distribution of $\widehat{\gamma}$ can be obtained. These assumptions mainly follow Hansen (1997, 2000).

$(A4)$ $M > M_j(\gamma) > 0$ for all $\gamma \in \Omega$, where $M = E\left(x_t x_t'\right), M_j\left(\gamma\right) = E\left(x_t x_t' \Psi_t^{(j)}\left(\gamma\right)\right)$, $j = 1, 2, 3, 4$.

$(A5)$ $\delta_T = (\delta^{(2)\prime}, \delta^{(3)\prime}, \delta^{(4)\prime})' = cT^{-\alpha} = (c_2', c_3', c_4')'T^{-\alpha}$, $0 < \alpha < \frac{1}{2}$, c is a $3p-$dimensional constant vector and c_i is a $p-$dimensional constant vector for $i = 2, 3, 4$.

$(A6)$ $D(\gamma)$ and $V(\gamma)$ are continuous at $\gamma = \gamma_0$.

$(A7)$ $d_1' D^* d_1 > 0, d_2' D^* d_2 > 0$, where $d_1 = (c_2' - c_4', c_3')', d_2 = (c_2', c_3' - c_4')'$.

$(A4)$ is the conventional full-rank condition which excludes perfect collinearity. Ω is restricted to be a proper subset of the support of z. $(A5)$ assumes that the parameter change is small and converges to zero at a slow rate when the sample size is large. Under this assumption, we are able to make the limiting distribution of $\widehat{\gamma}$ free of nuisance parameters (Chan, 1993). By letting δ_T go to zero, we reduce the rate of convergence of $\widehat{\gamma}$ from $O_p(T^{-1})$ to $O_p(T^{-1+2\alpha})$ and obtain a simpler limiting distribution of $\widehat{\gamma}$. $(A6)$ requires the moment functionals to be continuous so that one can obtain the Taylor expansion around γ_0. This condition excludes regime-dependent heteroskedasticity. $(A7)$ excludes the continuous threshold model.[10] Moreover, $d_1' D^* d_1 > 0$ and $d_2' D^* d_2 > 0$ impose the identification condition for γ_1^0 and γ_2^0 respectively.[11]

Theorem 1 *Under assumptions* $(A1)$ *to* $(A7)$*, we have*

$$
\begin{aligned}
& T^{1-2\alpha}\lambda_T\left((\widehat{\gamma}_1 - \gamma_1^0),\ (\widehat{\gamma}_2 - \gamma_2^0)\right) = (r_1, r_2) \\
& \xrightarrow{d} \underset{-\infty<r_1<\infty, -\infty<r_2<\infty}{\arg\max}\left[-\frac{1}{2}\left|r_1\right| + W_1\left(\left|r_1\right|\right) - \frac{1}{2}\left|r_2\right| + W_2(\left|r_2\right|)\right],
\end{aligned}
$$

where

$$
\lambda_T = \left(\frac{(d_1' D^* d_1) f_1^0}{\sigma^2}, \frac{(d_2' D^* d_2) f_2^0}{\sigma^2}\right),
$$

[10]This paper focuses on the discontinuous threshold effect. For continuous threshold models, one is referred to Chan and Tsay (1998).

[11]Note that $d_1 = (c_2' - c_4', c_3')'$ measures the size of the threshold effect for the first threshold variable z_1, while $d_2 = (c_2', c_3' - c_4')'$ measures the size of the threshold effect for the second threshold variable z_2. When $c_2 = c_4 \neq 0$ and $c_3 = 0$, we obtain a single threshold model with only two regimes separated by $z_2 = \gamma_2^0$. In this case, γ_1^0 is not identified. When $c_2 = 0$ and $c_3 = c_4 \neq 0$, we have a single threshold model with only two regimes separated by $z_1 = \gamma_1^0$ and γ_2^0 is not identified.

and $W_i(|r_i|)$ is a two-sided Brownian motion on the real line defined as:

$$W_i(|r_i|) = \begin{cases} \Lambda_i(-r_i) \text{ if } r_i < 0 \\ 0 \text{ if } r_i = 0 \\ \Lambda_i(r_i) \text{ if } r_i > 0 \end{cases},$$

$\Lambda_i(r_i)$, $i = 1, 2$, are two independent standard Brownian motions on $[0, \infty)$.

Proof. See Appendix 3. ■

The result of Hansen (1997) is a special case of Theorem 1 with $d_1 = 0$ or $d_2 = 0$. One can also use Theorem 1 to simulate the confidence interval of $(\widehat{\gamma}_1, \widehat{\gamma}_2)$. The parameter ratio λ_T can be estimated by a polynomial regression or kernel regression. See Hansen (1997, 2000).

4 Testing for and Estimation of the Threshold

To determine the number of regimes, we first consider the null hypothesis of no threshold effect:

$$H_0 : \beta^{(1)} = \beta^{(2)} = \beta^{(3)} = \beta^{(4)}.$$

Under the null hypothesis, there is only one regime. We define a likelihood ratio test statistic as:

$$J_T = \max_{\gamma \in \Omega} (T-p) \frac{\widetilde{\sigma}^2 - \widehat{\sigma}^2(\gamma)}{\widehat{\sigma}^2(\gamma)}. \tag{10}$$

$(T-p)\widetilde{\sigma}^2$ is the residual sum of squares under the null hypothesis, while $(T-p)\widehat{\sigma}^2(\gamma)$ is the residual sum of squares under the alternatives. If H_0 cannot be rejected, then the model is a simple AR model. Rejection of the null hypothesis suggests the existence of more than one regimes. The threshold estimator is defined as $\widehat{\gamma} = \arg\min \widehat{\sigma}^2(\gamma) = \arg\max J_T(\gamma)$. Since γ is not identified under the null hypothesis, the asymptotic distribution of $J_T(\widehat{\gamma})$ is not a standard χ^2. Hansen (1996) shows that the asymptotic distribution can be approximated by the following bootstrap procedure:

Let u_t^* $(t = 1, ...T)$ be i.i.d. $N(0, 1)$, and set $y_t^* = u_t^*$. Next, we regress y_t^* on $x_t = (1, y_{t-1}^*, y_{t-2}^*, ..y_{t-p}^*)$ to obtain the $J_T^*(\gamma) = (T-p)\dfrac{\widetilde{\sigma}^{*2} - \widehat{\sigma}^{*2}(\gamma)}{\widehat{\sigma}^{*2}(\gamma)}$ and $J_T^* = \max_{\gamma \in \Omega} J_T^*(\gamma)$.

The distribution of J_T^* converges weakly in probability to the distribution of J_T under the null hypothesis. Therefore, one can use the bootstrap value of J_T^* to approximate the asymptotic null distribution of J_T. The percentage of draws where the simulated statistic under H_0 exceeds the one obtained from the original sample is our bootstrapping p-value. The null hypothesis will be rejected if the p-value is small.

Rejection of the null hypothesis implies the presence of threshold effects. To determine the number of regimes, a general-to-specific approach is adopted. First, a three-regime model is tested against a four-regime model. Each of the following hypotheses

$(I)\quad H_0 : \beta^{(1)} = \beta^{(2)}$;
$(II)\quad H_0 : \beta^{(1)} = \beta^{(3)}$;
$(III)\ H_0 : \beta^{(1)} = \beta^{(4)}$;
$(IV)\ H_0 : \beta^{(2)} = \beta^{(3)}$;
$(V)\quad H_0 : \beta^{(2)} = \beta^{(4)}$;
$(VI)\ H_0 : \beta^{(3)} = \beta^{(4)}$.

is tested against the alternative hypothesis

H_1: there are four regimes.

A likelihood ratio test

$$J_T(\widehat{\gamma}) = (T-p)\frac{\widehat{\sigma}_0^2(\widehat{\gamma}) - \widehat{\sigma}_1^2(\widehat{\gamma})}{\widehat{\sigma}_1^2(\widehat{\gamma})} \tag{11}$$

is used to test these pairs of hypotheses, where $(T-p)\widehat{\sigma}_0^2(\widehat{\gamma})$ is the residual sum of squares under H_0, and $(T-p)\widehat{\sigma}_1^2(\widehat{\gamma})$ is the residual sum of squares under H_1. A parametric bootstrap method is applied to obtain the critical value. $\widehat{\gamma}$ is the estimated value from the unrestricted model. Let $y_t^* = \sum_{j=1}^{4}(\widehat{\beta}_0^{(j)} + \sum_{i=1}^{p} \widehat{\beta}_i^{(j)} y_{t-i}^*)\Psi_t^{(j)}(\widehat{\gamma}) + u_t^*$, where u_t^* are i.i.d. $N(0,1)$ and $\widehat{\beta}_i^{(j)\prime}s$ are estimated under the restricted model. We regress y_t^* on $x_t = (1, y_{t-1}^*, y_{t-2}^*, ..y_{t-p}^*)$ to obtain $J_T^*(\widehat{\gamma}) = (T-p)\frac{\widehat{\sigma}_0^{*2}(\widehat{\gamma}) - \widehat{\sigma}_1^{*2}(\widehat{\gamma})}{\widehat{\sigma}_1^{*2}(\widehat{\gamma})}$, and repeat this procedure a large number of times to calculate the percentage of draws for which the simulated statistic exceeds the actual value. The null is rejected if this p-value is too small.

Rejection of all the null hypotheses (I)-(VI) implies the existence of four regimes. If any one of them is accepted, then there are less than four regimes and we proceed to test a two-regime model against a three-regime model. For instance, if $(I)\ H_0 : \beta^{(1)} = \beta^{(2)}$ is accepted, then there are at most three regimes, and we proceed to test the two-regime model against the three-regime model. The following three hypotheses are tested using $J_T(\widehat{\gamma})$:

$H_0 : \beta^{(1)} = \beta^{(2)} = \beta^{(3)}$;
$H_0 : \beta^{(1)} = \beta^{(2)} = \beta^{(4)}$;
$H_0 : \beta^{(1)} = \beta^{(2)}$, $\beta^{(3)} = \beta^{(4)}$.

The alternative hypothesis is :
H_1 : There are three regimes with $\beta^{(1)} = \beta^{(2)}$.

If all the above null hypotheses are rejected, we conclude that there are three regimes. Otherwise, we conclude that the model has two regimes. In empirical studies, one can estimate the autoregressive order, the threshold value and the coefficients of the TAR model via the following procedure:

Step 1: First, a first-order TAR model is estimated:

$$y_t = \sum_{j=1}^{4}(\widehat{\beta}_0^{(j)} + \widehat{\beta}_1^{(j)} y_{t-1})\Psi_t^{(j)}(\gamma) + \widehat{u}_t,$$

and the initial threshold estimate $\widehat{\gamma}_T$ is obtained.

The first-order model is estimated for simplicity purposes (Chong, 2001). The initial threshold estimate will still be consistent even the true model is not of the first-order (Chong, 2003; Bai et al. 2008).[12]

Step 2: Given the threshold values obtained from step 1, we use the AIC (Tsay, 1998) to select the autoregressive order in each regime. In our case,

$$AIC_j(p_j) = n_j \ln[RSS_j(\widehat{\gamma}_T)/n_j] + 2(p_j + 1), \tag{12}$$

where

n_j is the number of observations in the j^{th} regime;
p_j is the order of autoregression in the j^{th} regime;
$RSS_j(\widehat{\gamma}_T)$ is the residual sum of squares for the j^{th} regime.

Define

$$\widehat{p}_j = \underset{p_j \in \{1,2,\ldots,P_{\max}\}}{\arg\min} AIC_j(p_j), \tag{13}$$

[12]The proof is available upon request.

where P_{max} is the maximum order considered in the model. The AIC for the whole model can be written as

$$NAIC = \sum_{j=1}^{s} AIC_j(\widehat{p}_j), \tag{14}$$

where s is the number of regimes.

Step 3: Perform the sequential likelihood ratio test to determine the number of regimes.

Step 4: Use the result obtained from step 3 to refine the threshold values, and repeat steps 2 and 3 until all the estimates converge.

5 Simulations

In the previous section, it is argued the threshold value can be consistently estimated even we start with a misspecified model in step 1. This result is obtained by Chong (2003) and Bai et al. (2008). The following experiments examine the consistency of the threshold estimator under model misspecifications.

The experiment is set up as follows:
Sample size: $T = 200$;
Number of replications: $N = 500$;
$u_t \sim N(0,1)$, $\varepsilon_t \sim N(0,1)$, $z_{1t} \sim N(0,1)$;
$P_{\max} = 10$.
We consider two cases for z_{2t} : (i) $z_{2t} \sim N(0,1)$, and (ii) $z_{2t} = z_{1t} + \varepsilon_t$.

The following data generating processes are examined:

DGP 1 : $y_t = (0.3y_{t-1} + 0.3y_{t-2})I(z_{1t} \leq 0 \text{ or } z_{2t} \leq 0) + (-0.3y_{t-1} - 0.3y_{t-2})I(z_{1t} > 0 \text{ and } z_{2t} > 0) + u_t$;

DGP 2 : $y_t = 0.3y_{t-1}I(z_{1t} \leq 0 \text{ or } z_{2t} \leq 0) - 0.3y_{t-1}I(z_{1t} > 0 \text{ and } z_{2t} > 0) + u_t$.

Three misspecified models are estimated:

Model A: $y_t = \sum_{j=1}^{4} \widehat{\beta}_1^{(j)} y_{t-1} \Psi_t^{(j)}(\gamma) + \widehat{u}_t$;

$$\text{Model B:} \quad y_t = \sum_{j=1}^{4} (\widehat{\beta}_1^{(j)} y_{t-1} + \widehat{\beta}_2^{(j)} y_{t-2}) \Psi_t^{(j)}(\gamma) + \widehat{u}_t;$$

$$\text{Model C:} \quad y_t = \widehat{\beta}_1^{(1)} y_{t-1} 1(z_{1t} \leq \gamma_1) + \widehat{\beta}_1^{(2)} y_{t-1} 1(z_{1t} > \gamma_1) + \widehat{u}_t.$$

Model A underestimates the autoregressive order p, while Model B overestimates the autoregressive order p. Both of them overestimate the number of regimes. The estimation results are reported in Table 1. For all misspecified estimated models, $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ converge to the true threshold value 0. The results for models A and B suggest that the consistency of the threshold estimators is unaffected by the misspecification of regressors. Therefore, if the number of threshold variables is known, one can obtain a preliminary and consistent threshold estimate using the simplest model possible. The preliminary estimate of the threshold value can be used to obtain the estimates of other parameters of interest. In the context of our model, such a preliminary threshold value allows us to determine the number of regimes, as well as the order and parameters of the autoregressive model within each regime.

Table 1: The Simulation Results

DGP	Estimated Model	z_{2t}	$\overline{\widehat{\gamma}_1}$	$Var(\widehat{\gamma}_1)$	$\overline{\widehat{\gamma}_2}$	$Var(\widehat{\gamma}_2)$
1	A	$N(0,1)$	0.007	0.030	−0.002	0.025
1	A	$z_{1t} + \varepsilon_t$	−0.002	0.036	−0.003	0.029
2	B	$N(0,1)$	−0.001	0.025	−0.005	0.034
2	B	$z_{1t} + \varepsilon_t$	−0.006	0.025	0.002	0.027
2	C	$N(0,1)$	−0.015	0.68		
2	C	$z_{1t} + \varepsilon_t$	−0.19	0.28		

The results of Chong (2003) and Bai et al. (2008) apply to cases where the threshold variables are correctly specified. Model C underspecifies the number of threshold variables. The results for Model C show that the estimators of the single threshold-variable model may not be consistent in the presence of two dependent threshold variables.

6 Empirical Application

Our model is applied to the daily return series of the Hang Seng Index. The Hong Kong stock market is studied because of its rising role as a global financial center. In 2006, Hong Kong becomes the world's second most popular place for IPO after London. In 2007, the Hong Kong stock market ranks fifth in the world, and its warrant market ranks first globally in terms of turnover. Most of the previous studies in the literature use the first lagged return as the threshold variable to identify the market regimes. Such a classification method does not take investors' sentiment into account and does not consider the information of market turnover. In this paper, we use the past information of price and market turnover to construct our threshold variables. Our sample period runs from January 3rd 1995 to January 13th 2005. The return series is defined as the log-difference of the Hang Seng Index (HSI). There are over 2500 observations in our sample. Figure 1 and Figure 2 show the time series data for daily return and market turnover.

Figure 1 about here

Figure 2 about here

The two threshold variables are analogous to those of Granville (1963) and Lee and Swaminathan (2000). We define the first threshold variable as

$$Rap_t = \frac{PMA20_t}{PMA250_t}, \tag{15}$$

where

$$PMA250_t = \frac{\sum_{j=1}^{250} p_{t-j}}{250}, \quad PMA20_t = \frac{\sum_{j=1}^{20} p_{t-j}}{20}.$$

$PMA250_t$ is the average price for the past 250 trading days;
$PMA20_t$ is the average price for the past 20 trading days.

The variable is a ratio of two moving averages, which is similar to that of Hong and Lee (2003). In particular, the 250-day moving average, which is widely used by investors to define the market state, is employed. If the price rises above (falls below) the 250-day moving average, an average investor who has taken a long position in the previous year (about 250 trading days) has made a profit (loss), suggesting that the market sentiment should be good (bad). To reduce noise, we use the crossing of the 20-day and 250-day moving averages to help identify the market regimes.

The second threshold variable contains the information of the market turnover, which has been widely used to measure the liquidity of the market, see Amihud and Mendelson (1986), Brennan et al. (1998) and Amihud (2002) among others. Several studies have shown that the autocorrelation in stock returns is related to turnover or trading volume. For example, Campbell et al. (1993) find that the first-order daily return autocorrelation tends to decline with turnover, and the returns accompanied by high volume tend to be reversed more strongly. Llorente et al. (2002) point out that intensive trading volume can help to identify the periods in which shocks occur. Therefore, we define the second threshold variable as:

$$Rav_t = \log\left(turnover_{t-1}\right) - MAV_{t-1}, \tag{16}$$

where

$$MAV_t = \frac{\sum_{j=1}^{250} \log\left(turnover_{t-j}\right)}{250}.$$

Figure 3 shows the two threshold variables Rap_t and Rav_t.

Figure 3 about here

Our four-regime threshold model on the return series is

$$y_t = \sum_{j=1}^{4} \Psi_t^{(j)}\left(\gamma^0\right)\left(\beta_0^{(j)} + \beta_1^{(j)} y_{t-1} + \beta_2^{(j)} y_{t-2} + ... + \beta_p^{(j)} y_{t-p_j}\right) + u_t, \tag{17}$$

where

y_t is the return series defined as the log-difference of the HSI;
$\Psi_t^{(1)}\left(\gamma^0\right) = I\left(Rap_t \le \gamma_1^0,\ Rav_t \le \gamma_2^0\right);$
$\Psi_t^{(2)}\left(\gamma^0\right) = I\left(Rap_t \le \gamma_1^0,\ Rav_t > \gamma_2^0\right);$
$\Psi_t^{(3)}\left(\gamma^0\right) = I\left(Rap_t > \gamma_1^0,\ Rav_t \le \gamma_2^0\right);$
$\Psi_t^{(4)}\left(\gamma^0\right) = I\left(Rap_t > \gamma_1^0,\ Rav_t > \gamma_2^0\right).$

The estimated threshold values from step 1 in Section 3 are: $\widehat{\gamma}_{rap} = 1.02$ and $\widehat{\gamma}_{rav} = 0.57$. The results of the sequential likelihood ratio test are shown in Tables 2a and 2b.

Table 2a: Results of the LR Test for 4 Regimes vs 3 Regimes

$H_1 : s = 4$	$\beta^{(i)} \neq \beta^{(j)}$ when $i \neq j$					
$H_0 : s = 3$	$\beta^{(2)} = \beta^{(1)}$	$\beta^{(3)} = \beta^{(2)}$	$\beta^{(4)} = \beta^{(1)}$	$\beta^{(3)} = \beta^{(1)}$	$\beta^{(2)} = \beta^{(4)}$	$\beta^{(3)} = \beta^{(4)}$
$J_T(\widehat{\gamma})$	126.7	136.3	5.45	16.59	93.05	18.79
p-value	< 0.01	< 0.01	> 0.05	< 0.05	< 0.01	< 0.05

Note from Table 2a that the null hypothesis $\beta^{(4)} = \beta^{(1)}$ cannot be rejected since $J_T(\widehat{\gamma})$ has a p-value larger than 0.05.[13] Next, we proceed to test the 3-regime model against the 2-regime model. The results from Table 2b suggest that the movement of the return series can be approximated by a three-regime model.

Table 2b: Results of the LR Test for 3 Regimes vs 2 Regimes

$H_1 : s = 3$	$\beta^{(4)} = \beta^{(1)}$		
$H_0 : s = 2$	$\beta^{(2)} = \beta^{(3)}, \beta^{(4)} = \beta^{(1)}$	$\beta^{(1)} = \beta^{(2)} = \beta^{(4)}$	$\beta^{(3)} = \beta^{(4)} = \beta^{(1)}$
$J_T(\widehat{\gamma})$	107.9	159.5	18.77
p-value	< 0.01	< 0.01	< 0.05

Table 3 shows the final estimation results. The threshold estimates are revised to $\widehat{\gamma} = (1.02, 0.53)$.

Table 3: The Estimated TAR Model

Regime	Estimation Results
I	$y_t = 0.0003 + 0.065y_{t-1}$, if $Rap_t > 1.02$ and $Rav_t \leq 0.53$
II	$y_t = 0.0067 - 0.3y_{t-1} - 0.4y_{t-2} + 0.18y_{t-3} + 0.09y_{t-4} - 0.12y_{t-5} + 0.54y_{t-6}$
	$-0.5y_{t-7} - 0.18y_{t-8}$, if $Rap_t \leq 1.02$ and $Rav_t > 0.53$
III	$y_t = 0.00014 + 0.096y_{t-1}$ Otherwise.

Figure 4 plots the estimated residuals of the model.[14]

Figure 4 about here

Using the Markov-switching model, Maheu and McCurdy (2000) divide the stock market into a high-return stable regime and a low-return volatile regime. From Table 3, we are able to classify the stock market of Hong Kong into three regimes. Since high turnover is usually associated with volatile returns (Karpoff, 1987; Foster and Viswanathan, 1995), Regime I generated by our model corresponds to the high-return

[13]In some cases, if two or more hypotheses cannot be rejected, we choose the one with the largest p-value as the candidate model in the subsequent step.

[14]A Ljung-Box test has been conducted and the results suggest that the residuals are white noise. The details of the test can be obtained from the authors upon request.

stable regime, while Regime II is the low-return volatile regime.[15] Regime III is a neutral regime. Table 4 shows a chronology of major events affecting the Hong Kong stock market between 1996 and 2005.[16]

Table 4: A Chronology of the Hong Kong Stock Market and the Corresponding Regimes

Date	Event	Regime
1997.7	The establishment of the Hong Kong Special Administration Region	I
1997.10.23	Asian currency turmoil triggered by the floating of Thai Baht	II
1998.1-1999.3	The burst of the property market	III, II
1999.9-2000.3	Global technology stock boom and the admission of China into the WTO	I, III
2000.4-2000.6	The burst of the high-tech bubble	II, III
2001.9.11	The 911 incident	I
2001.11.13	The accession of China to the WTO	I
2003.2-2003.6	The outbreak of SARS	I, II
2003.6.29	The launch of Closer Economic Partnership Arrangement with China	I

7 Conclusion

Conventional threshold models only allow for a single threshold variable. In many applications, the use of multiple threshold variables is needed. In this paper, a new

[15]Note that the first-order coefficient for Regime I is 0.065, which is positive as compared to that of -0.3 for Regime II. This agrees with Campbell et al. (1993) that the first-order daily return autocorrelation tends to decline when turnover increases.

[16]We associate the estimated regimes with these major events. For example, the establishment of the Hong Kong Special Administration Region in July 1997 falls into Regime I. During the Asian Financial Crisis, the crash of the stock market of Hong Kong and the burst of the property market fall into Regime II. The market experiences a volatile year in the millennium. Driven by the technology bubble and the accession of China to the World Trade Organization, the Hang Seng Index reaches a record high of 18301 in March 2000. However, the burst of the bubble in 2000 brings the stock market back into Regime II again. The market enters Regime I at the end of 2001. Note that the Hong Kong stock market is not seriously affected by the 911 incident. The accession of China to the World Trade Organization in 2001 is a good news for Hong Kong. In the beginning of 2003, the outbreak of the Severe Acute Respiratory Syndrome (SARS) threatens the economy. The market switches from Regime I to Regime II during the SARS period, but it rebounds sharpy in the second half of the year. In June, China and Hong Kong sign the Closer Economic Partnership Arrangement (CEPA), a free trade agreement between Hong Kong and China which gives Hong Kong a preferential access to the Chinese market.

TAR model with two threshold variables is developed. In addition, the consistency and limiting distribution of the estimators are established. A likelihood ratio test is also constructed to detect the threshold effect. Our model is applied to identify the regimes of the Hong Kong stock market. The two threshold variables used in this paper are analogous to those of Granville (1963) and Lee and Swaminathan (2000). Unlike the conventional bull-bear classification, it is shown that the Hong Kong stock market can be classified into three regimes, namely, a high-return stable regime, a low-return volatile regime and a neutral regime. It should be mentioned that our model assumes a single threshold for each threshold variable. It can be extended to allow for the existence of multiple thresholds (Gonzalo and Pitarakis, 2002). For example, if there are two threshold variables and each threshold variable has two threshold values, then the model can have at most nine regimes. One may also define the threshold condition as a nonlinear function of the two threshold variables. Finally, one may relax the i.i.d. assumption of the error term to allow for serial dependence and regime-dependent heteroskedasticity. Such extensions, however, are beyond the scope of this paper and are left for future research.

References

1. Amihud, Y. and H. Mendelson (1986). Asset Pricing and the Bid-Ask Spread. *Journal of Financial Economics* 17, 223-249.

2. Amihud, Y. (2002). Illiquidity and Stock Returns: Cross-Section and Time Series Effects. *Journal of Financial Markets* 5, 31-56.

3. Astatkie, T., D.G. Watts and W.E. Watt (1997). Nested Threshold Autoregressive (NeTAR) Models. *International Journal of Forecasting* 13, 105-116.

4. Avramov, D., T. Chorida and A. Goyal (2006). Liquidity and Autocorrelation in Individual Stock Returns. *Journal of Finance* LXI, 2365-2390.

5. Bai, J., H. Chen, T.T.L. Chong and X. Wang (2008). Generic Consistency of the Break-Point Estimator under Specification Errors in a Multiple-Break Model. *Econometrics Journal* 11, 287-307.

6. Bai, J. (1997). Estimation of a Change Point in Multiple Regressions. *Review of Economics and Statistics*, 79, 551-563.

7. Brennan, M.J., T. Cordia and A. Subrahmanyam (1998). Alternative Factor Specifications, Security Characteristics, and the Cross-Section of Expected

Stock Returns. *Journal of Financial Economics* 49, 345-73.

8. Campbell, J.Y., S.J. Grossman and J. Wang (1993). Trading Volume and Serial Correlation in Stock Returns. *Quarterly Journal of Economics* 108, 905-939.

9. Caner M. and B.E. Hansen (2001). Threshold Autoregression with a Unit Root. *Econometrica* 69, 1555-1596.

10. Chan, K.S. and H. Tong (1986). On Estimating Thresholds in Autoregressive Models. *Journal of Time Series Analysis* 7, 179-190.

11. Chan, K.S. (1993). Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model. *Annals of Statistics* 21, 520-533.

12. Chen, R. and R.S. Tsay (1993). Functional-Coefficient Autoregressive Models. *Journal of the American Statistical Association* 88, 298-308.

13. Chong, T.T.L. (2003). Generic Consistency of the Break-Point Estimator under Specification Errors. *Econometrics Journal* 6, 167-192.

14. Chong, T.T.L. (2001). Structural Change in AR(1) Models. *Econometric Theory* 17, 87-155.

15. Dueker, M., S. Martin and F. Spagnolo (2007). Contemporaneous Threshold Autoregressive Models: Estimation, Testing and Forecasting. *Journal of Econometrics* 141, 517-547.

16. Durlauf, S. N. and P. A. Johnson (1995). Multiple Regimes and Cross-country Growth Behavior. *Journal of Applied Econometrics* 10, 365-384.

17. Edison, H.J. (2000). Do Indicators of Financial Crises Work? An Evaluation of an Early Warning System. Board of Governors of Federal Reserve System, International Finance Discussion Papers Number 675.

18. Foster, F.D. and S. Viswanathan (1995). Can Speculative Trading Explain the Volume-Volatility Relation? *Journal of Business and Economic Statistics* 13, 379-396.

19. Frankel, J.A. and A. Rose (1996). Currency Crashes in Emerging Markets: Empirical Indicators. *NBER* WP5437.

20. Gonzalo, J. and J. Pitarakis (2002). Estimation and Model Selection Based Inference in Single and Multiple Threshold Models. *Journal of Econometrics* 110, 319-352.

21. Granville, J. (1963). *Granville's New Key to Stock Market Profits*, Prentice Hall.

22. Hansen, B.E. (2000). Sample Splitting and Threshold Estimation. *Econometrica* 68, 575-603.

23. Hansen, B.E. (1997). Inference in TAR models. *Studies in Nonlinear Dynamics and Economics* 2, 1-14.

24. Hansen, B.E. (1996). Inference when a Nuisance Parameter is not Identified under the Null Hypothesis. *Econometrica* 64, 413-430.

25. Hong, Y. and T.H. Lee (2003). Inference on Predictability of Foreign Exchange Rate via Generalized Spectrum and Nonlinear Time Series Models. *Review of Economics and Statistics* 85, 1048-1062.

26. Kaminsky, G.L. (1998). Currency and Banking Crises: The Early Warnings of Distress. Board of Governors of the Federal Reserve System. International Finance Discussion Papers Number 629.

27. Karpoff, J.M. (1987). The Relation between Price Changes and Trading Volume: A Survey. *Journal of Financial and Quantitative Analysis* 22,109-125.

28. Lee, M.C. and B. Swaminathan (2000). Price Momentum and Trading Volume. *Journal of Finance* IV, 2017-2069.

29. Leeper, E.M. (1991). Equilibria under "Active" and "Passive" Monetary and Fiscal Policies. *Journal of Monetary Economics* 27 129-147.

30. Llorente, G., R. Michaely, G. Saar and J. Wang (2002). Dynamic Volume-Return relation of Individual Stocks. *Review of Financial Studies* 15, 1005-1047.

31. Maheu, J.M. and T.H. McCurdy (2000). Identifying Bull and Bear Markets in Stock Returns. *Journal of Business and Economic Statistics* 18, 100-112.

32. Sachs, J., A. Tornell and A. Velasco (1996). Financial Crises in Emerging Markets: The Lessons from 1995. *Brookings Papers on Economic Activity No.1,* 147-215.

33. Tsay, R.S. (1998). Testing and Modeling Multivariate Threshold Models. *Journal of the American Statistical Association* 93, 231-240.

34. Tiao, G.C. and R.S. Tsay (1994). Some Advances in Nonlinear and Adaptive Modeling in Time-series. *Journal of Forecasting* 13, 109-131.

35. Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis* (Lecture Notes in Statistics No. 21), New York: Springer-Verlag.

Appendix 1: Lemmas

Throughout the Appendix, let $||A|| = (tr(A'A))^{1/2}$ denote the Euclidean norm of a matrix A. Let $||A||_r = (E|A|^r)^{1/r}$ denote the L^r-norm of a random matrix and $\Rightarrow$ denote weak convergence with respect to the uniform metric.

Let $x_t = (1, y_{t-1}, y_{t-2}, ..., y_{t-p})'$ for $t = p+1,\, p+2, ..., T$;

$X = (x'_T, x'_{T-1}, ..., x'_{p+1})_{(T-p)\times(p+1)}$;

$Y = (y_T, ...y_{p+1})'$;

$U = (u_T, u_{T-1}, ..., u_{p+1})'$;

$I_j(\gamma) = diag\left\{\Psi_T^{(j)}(\gamma), \Psi_{T-1}^{(j)}(\gamma), ..., \Psi_{p+1}^{(j)}(\gamma)\right\}$,

where $\Psi_t^{(j)}(\gamma)$ is defined in Section 2.

Let M and $M_j(\gamma)$ be moment functionals defined as:

$M = E(x_t x'_t)$, $M_j(\gamma) = E\left(x_t x'_t \Psi_t^{(j)}(\gamma)\right)$, $j = 1, 2, 3, 4$.

Lemma 1: Under assumptions $(A1) - (A2)$, it can be shown that

(a) $\dfrac{1}{T}X'X \xrightarrow{p} M$;

(b) $\dfrac{1}{T}X'U \xrightarrow{p} 0$.

Proof: The proof is straightforward by applying the law of large number for stationary ergodic processes.■

Lemma 2: For any $\gamma \in \Omega$, under assumptions $(A1) - (A3)$, we have, for $j = 1, 2, 3, 4$,

(a) $\dfrac{1}{T}X'I_j(\gamma)X \xrightarrow{p} M_j(\gamma)$;

(b) $\dfrac{1}{T}X'I_j(\gamma)U \xrightarrow{p} 0$;

(c) $\dfrac{1}{T}(X'I_j(\gamma)U)'(X'I_j(\gamma)U) \xrightarrow{p} E(x_t x'_t u_t^2 \Psi_t^{(j)}(\gamma)) = \sigma^2 M_j(\gamma)$.

Proof: The proof of part (a) for $j = 1$ is similar to the proof of Lemma A1 in Hansen (1996) by replacing $\{w_t \leq \gamma\}$ with $\{z_{1t} \leq \gamma_1, z_{2t} \leq \gamma_2\}$. For $j = 2$, we have $\dfrac{1}{T}X'I_2(\gamma)X = \dfrac{1}{T}\sum x_t x'_t\{z_{1t} \leq \gamma_1\} - \dfrac{1}{T}\sum x_t x'_t\{z_{1t} \leq \gamma_1, z_{2t} \leq \gamma_2\} \xrightarrow{p}$ $E\left(x_t x'_t\{z_{1t} \leq \gamma_1\}\right) - E\left(x_t x'_t\{z_{1t} \leq \gamma_1, z_{2t} \leq \gamma_2\}\right) = M_2(\gamma)$. Similar proof can be applied to the cases where $j = 3$ and 4. The proofs for (b) and (c) are analogous and are therefore skipped. ■

Appendix 2: Consistency of Estimators

To prove the consistency of the estimator $\widehat{\gamma} = \arg\min_{\gamma\in\Omega} RSS_T(\gamma)$, it suffices to show that $RSS_T(\gamma)$ converges uniformly to a function $b(\gamma)$ which is minimized at γ_0. For simplicity, denote $\widehat{\beta}^{(j)} = \widehat{\beta}^{(j)}(\gamma)$ for $j = 1, 2, 3, 4$. Let $\widehat{Y}(\gamma) = \sum_{j=1}^{4} I_j(\gamma)X\widehat{\beta}^{(j)}$. The residual sum of squares can be written as:

$$RSS_T(\gamma) = ||Y - \widehat{Y}(\gamma)||^2 = Y'Y - \widehat{Y}(\gamma)'\widehat{Y}(\gamma)$$
$$= \sum_{j=1}^{4}\left(\beta^{(j)\prime}X'I_j(\gamma^0)X\beta^{(j)} - \widehat{\beta}^{(j)\prime}X'I_j(\gamma)X\widehat{\beta}^{(j)}\right) + 2\sum_{j=1}^{4}U'I_j(\gamma^0)X\beta^{(j)} + U'U.$$

Next, we prove that $RSS_T(\gamma)$ has a unique minimum at $\gamma = \gamma^0$. We partition the threshold space into four regions.

Case 1: $\gamma_1 \leq \gamma_1^0$ and $\gamma_2 \leq \gamma_2^0$

Using Lemmas 1 and 2, and the facts that

$I_1(\gamma)I_1(\gamma^0) = I_1(\gamma), I_1(\gamma)I_j(\gamma^0) = 0$ for $j = 2, 3, 4$;

$I_2(\gamma)I_1(\gamma^0) = I_2(\gamma) - I_2(\gamma_1, \gamma_2^0)$, $I_2(\gamma)I_2(\gamma^0) = I_2(\gamma_1, \gamma_2^0)$;
$I_2(\gamma)I_j(\gamma^0) = 0$, for $j = 3, 4$;

$I_3(\gamma)I_1(\gamma^0) = I_3(\gamma) - I_3(\gamma_1^0, \gamma_2)$, $I_3(\gamma)I_2(\gamma^0) = 0$;
$I_3(\gamma)I_3(\gamma^0) = I_3(\gamma_1^0, \gamma_2)$, $I_3(\gamma)I_4(\gamma^0) = 0$;

$I_4(\gamma)I_1(\gamma^0) = I_1(\gamma^0) + I_1(\gamma) - I_1(\gamma_1^0, \gamma_2) - I_1(\gamma_1, \gamma_2^0)$;
$I_4(\gamma)I_2(\gamma^0) = 0$, $I_4(\gamma)I_3(\gamma^0) = I_4(\gamma_1^0, \gamma_2) - I_4(\gamma^0)$, $I_4(\gamma)I_4(\gamma^0) = I_4(\gamma^0)$;

it can be shown that

$$\widehat{\beta}^{(1)} = (X'I_1(\gamma)X)^{-1}X'I_1(\gamma)Y = (X'I_1(\gamma)X)^{-1}X'I_1(\gamma)[\sum_{j=1}^{4} I_j(\gamma^0)X\beta^{(j)} + U]$$
$$= \beta^{(1)} + \frac{1}{\sqrt{T}}(\frac{X'I_1(\gamma)X}{T})^{-1}(\frac{X'I_1(\gamma)U}{\sqrt{T}}) \xrightarrow{p} \beta^{(1)};$$

$$\widehat{\beta}^{(2)} = (X'I_2(\gamma)X)^{-1}X'I_2(\gamma)Y$$

$\xrightarrow{p} M_2^{-1}(\gamma)(M_2(\gamma)-M_2(\gamma_1,\gamma_2^0))(\beta^{(1)}-\beta^{(2)})+\beta^{(2)};$

$$\widehat{\beta}^{(3)}=(X'I_3(\gamma)X)^{-1}X'I_3(\gamma)Y$$
$$\xrightarrow{p} M_3^{-1}(\gamma)(M_3(\gamma)-M_3^0(\gamma_1^0,\gamma_2))(\beta^{(1)}-\beta^{(2)})+M_3^{-1}(\gamma)M_3(\gamma_1^0,\gamma_2)(\beta^{(3)}-\beta^{(2)})+\beta^{(2)};$$

$$\widehat{\beta}^{(4)}=(X'I_4(\gamma)X)^{-1}X'I_4(\gamma)Y$$
$$\xrightarrow{p} M_4^{-1}(\gamma)\left[M_4(\gamma^0)-M_4(\gamma_1^0,\gamma_2)-M_4(\gamma_1,\gamma_2^0)+M_4(\gamma)\right](\beta^{(1)}-\beta^{(2)})$$
$$+M_4^{-1}(\gamma)M_4(\gamma^0)(\beta^{(4)}-\beta^{(2)})+M_4^{-1}(\gamma)(M_4(\gamma_1^0,\gamma_2)-M_4(\gamma^0))(\beta^{(3)}-\beta^{(2)})+\beta^{(2)}.$$

Therefore,

$$\begin{aligned}
&\tfrac{1}{T}\left(RSS_T(\gamma)-U'U\right)\\
&=\tfrac{1}{T}\sum_{j=1}^{4}\left(\beta^{(j)\prime}X'I_j(\gamma^0)X\beta^{(j)}-\widehat{\beta}^{(j)\prime}X'I_j(\gamma)X\widehat{\beta}^{(j)}\right)+\tfrac{2}{T}\sum_{j=1}^{4}U'I_j(\gamma^0)X\beta^{(j)}\\
&=\sum\nolimits_{j=1}^{4}\beta^{(j)\prime}M_j(\gamma^0)(\beta^{(j)}-\beta^{(2)})\\
&\quad-\left[\widehat{\beta}^{(1)\prime}M_1(\gamma)+\widehat{\beta}^{(2)\prime}(M_2(\gamma)-M_2(\gamma_1,\gamma_2^0))\right](\beta^{(1)}-\beta^{(2)})\\
&\quad+\widehat{\beta}^{(3)\prime}[M_3(\gamma)-M_3(\gamma_1^0,\gamma_2)](\beta^{(1)}-\beta^{(2)})\\
&\quad+\widehat{\beta}^{(4)\prime}[M_4(\gamma^0)-M_4(\gamma_1^0,\gamma_2)-M_4(\gamma_1,\gamma_2^0)+M_4(\gamma)](\beta^{(1)}-\beta^{(2)})\\
&\quad-\left[\widehat{\beta}^{(3)\prime}M_3(\gamma_1^0,\gamma_2)+\widehat{\beta}^{(4)\prime}(M_4(\gamma_1^0,\gamma_2)-M_4(\gamma^0))\right](\beta^{(3)}-\beta^{(2)})\\
&\quad-\widehat{\beta}^{(4)\prime}M_4(\gamma^0)(\beta^{(4)}-\beta^{(2)})+o_p(1)\\
&=(\beta^{(1)}-\beta^{(2)})'[M_1(\gamma^0)-M_1(\gamma)-M_2^{-1}(\gamma)(M_2(\gamma)-M_2(\gamma_1,\gamma_2^0))^2\\
&\quad-M_3^{-1}(\gamma)(M_3(\gamma)-M_3(\gamma_1^0,\gamma_2))^2\\
&\quad-M_4^{-1}(\gamma)(M_4(\gamma^0)-M_4(\gamma_1^0,\gamma_2)-M_4(\gamma_1,\gamma_2^0)+M_4(\gamma))^2]\times(\beta^{(1)}-\beta^{(2)})\\
&\quad+(\beta^{(3)}-\beta^{(2)})'[M_3(\gamma^0)-M_4^{-1}(\gamma)(M_4(\gamma_1^0,\gamma_2)-M_4(\gamma^0))^2\\
&\quad-M_3^{-1}(\gamma)(M_3(\gamma_1^0,\gamma_2))^2](\beta^{(3)}-\beta^{(2)})\\
&\quad+(\beta^{(4)}-\beta^{(2)})'\left[M_4(\gamma^0)-M_4^{-1}(\gamma)(M_4(\gamma^0))^2\right](\beta^{(4)}-\beta^{(2)})+o_p(1)\\
&=(\beta^{(1)}-\beta^{(2)})'Q_1(\beta^{(1)}-\beta^{(2)})+(\beta^{(3)}-\beta^{(2)})'Q_2(\beta^{(3)}-\beta^{(2)})\\
&\quad+(\beta^{(4)}-\beta^{(2)})'Q_3(\beta^{(4)}-\beta^{(2)})+o_p(1)\\
&=b_1(\gamma)+o_p(1).
\end{aligned}$$

For any $\gamma_1\le\gamma_1^0$ and $\gamma_2\le\gamma_2^0$, it is obvious that Q_3 is semi-positive definite since $M_4(\gamma)>M_4(\gamma^0)$. Meanwhile, using the following results:

$$\begin{aligned}
&M_1(\gamma^0)-M_1(\gamma)=E\left(x_tx_t'\Psi_t^{(1)}(\gamma)\right)\\
&=E(x_tx_t'[\Psi_t^{(2)}(\gamma)-\Psi_t^{(2)}(\gamma_1,\gamma_2^0)+\Psi_t^{(3)}(\gamma)-\Psi_t^{(3)}(\gamma_1^0,\gamma_2)+\Psi_t^{(4)}(\gamma^0)-\Psi_t^{(4)}(\gamma_1^0,\gamma_2)\\
&\qquad-\Psi_t^{(4)}(\gamma_1,\gamma_2^0)+\Psi_t^{(4)}(\gamma)])\\
&=M_2(\gamma)-M_2(\gamma_1,\gamma_2^0)+M_3(\gamma)-M_3(\gamma_1^0,\gamma_2)\\
&\quad+M_4(\gamma^0)-M_4(\gamma_1^0,\gamma_2)-M_4(\gamma_1,\gamma_2^0)+M_4(\gamma);
\end{aligned}$$

$M_3(\gamma^0) = M_4(\gamma_1^0, \gamma_2) - M_4(\gamma^0) + M_3(\gamma_1^0, \gamma_2)$,

it can be shown that Q_1 and Q_2 are semi-positive definite. Thus, $b_1(\gamma) \geq b_1(\gamma^0) = 0$, and the equation holds if and only if $\gamma = \gamma^0$.

By analogy, for the remaining three cases,

$\frac{1}{T}\left(RSS_T(\gamma) - U'U\right) = b_j(\gamma) + o_p(1)$ and $b_j(\gamma) \geq b_j(\gamma^0) = 0$ for $j = 2, 3, 4$.

Define a non-stochastic function $b(\gamma)$ as $b_j(\gamma)$ for the j^{th} case, we have

$$\sup_{\gamma \in \Omega} \left| \frac{1}{T}\left(RSS_T(\gamma) - U'U\right) - b(\gamma) \right| = o_p(1). \tag{18}$$

Thus, $b(\gamma)$ is minimized if and only if $\gamma = \gamma^0$. This implies that the limit of $\frac{1}{T} RSS_T(\gamma)$ is minimized at γ^0. By the superconsistency of $\widehat{\gamma}$, $\widehat{\beta}^{(j)}$ will also be consistent.

Appendix 3: The Limiting Distribution of $\widehat{\gamma}$

To derive the limiting distribution of $\widehat{\gamma}$ for shrinking break, we let $\delta = (\delta^{(2)\prime}, \delta^{(3)\prime}, \delta^{(4)\prime})' = cT^{-\alpha}, 0 < \alpha < \frac{1}{2}$, $c = (c_2', c_3', c_4')'$ is a $3p-$dimensional vector of constants. Define

$$\widehat{\gamma} = \arg\min_{\gamma \in \Omega} RSS_T(\gamma) = \arg\min_{\gamma \in \Omega} \left[RSS_T(\gamma) - RSS_T(\gamma^0)\right].$$

To obtain the limiting distribution of $\widehat{\gamma}$, we first examine the asymptotic behavior of $RSS_T(\gamma) - RSS_T(\gamma^0)$ in the neighborhood of the true thresholds.

Recall from Equation (7) that the true model can be written as:

$$Y = X\beta^{(1)} + \sum_{j=2}^{4} X_{\gamma^0}^{(j)} \delta^{(j)} + U = X\beta^{(1)} + X_0\delta + U,$$

where $X_0 = (X_{\gamma^0}^{(2)}, X_{\gamma^0}^{(3)}, X_{\gamma^0}^{(4)})$. Let

$\widehat{\delta}^{(j)} = \widehat{\delta}^{(j)}(\gamma)$, $\widehat{\delta}_0^{(j)} = \widehat{\delta}^{(j)}(\gamma^0)$, for $j = 2, 3, 4$;
$\widehat{\delta}(\gamma) = (\widehat{\delta}^{(2)\prime}, \widehat{\delta}^{(3)\prime}, \widehat{\delta}^{(4)\prime})'$ and $\widehat{\delta}(\gamma^0) = (\widehat{\delta}_0^{(2)\prime}, \widehat{\delta}_0^{(3)\prime}, \widehat{\delta}_0^{(4)\prime})'$.

For any γ, define $X_\gamma = (X_\gamma^{(2)}, X_\gamma^{(3)}, X_\gamma^{(4)})$. We have

$$\widehat{\beta}^{(1)}(\gamma) = (X_\gamma^{(1)\prime} X_\gamma^{(1)})^{-1} X_\gamma^{(1)\prime} Y = \beta^{(1)} + (X_\gamma^{(1)\prime} X_\gamma^{(1)}) X_\gamma^{(1)\prime} X_0 \delta + (X_\gamma^{(1)\prime} X_\gamma^{(1)})^{-1} X_\gamma^{(1)\prime} U.$$

$$\widehat{\beta}^{(1)}(\gamma^0) = (X_{\gamma_0}^{(1)\prime} X_{\gamma_0}^{(1)})^{-1} X_{\gamma_0}^{(1)\prime} Y = \beta^{(1)} + (X_{\gamma_0}^{(1)\prime} X_{\gamma_0}^{(1)})^{-1} X_{\gamma_0}^{(1)\prime} U.$$

Since $\widehat{\gamma}$ is a consistent estimator, we study its asymptotic behavior in the neighborhood of the true thresholds. Let $\gamma_1 = \gamma_1^0 + \dfrac{\upsilon}{T^{1-2\alpha}}$, $\gamma_2 = \gamma_2^0 + \dfrac{\omega}{T^{1-2\alpha}}$. By Lemmas 1 and 2,

$$\begin{aligned}
&\widehat{\beta}^{(1)}(\gamma) - \widehat{\beta}^{(1)}(\gamma^0) = (X_\gamma^{(1)\prime} X_\gamma^{(1)})^{-1} X_\gamma^{(1)\prime} X_0 \delta + (X_\gamma^{(1)\prime} X_\gamma^{(1)})^{-1} X_v^{(1)\prime} U - (X_{\gamma_0}^{(1)\prime} X_{\gamma_0}^{(1)})^{-1} X_{\gamma_0}^{(1)\prime} U \\
&= \sum_{j=2}^{4} (X_\gamma^{(1)\prime} X_\gamma^{(1)})^{-1} X_\gamma^{(1)\prime} (X_0^{(j)} - X_\gamma^{(j)}) \delta + (X_\gamma^{(1)\prime} X_\gamma^{(1)})^{-1} (X_\gamma^{(1)\prime} U - X_{\gamma_0}^{(1)\prime} U) \\
&\quad + \left((X_\gamma^{(1)\prime} X_\gamma^{(1)})^{-1} - (X_{\gamma_0}^{(1)\prime} X_{\gamma_0}^{(1)})^{-1} \right) X_{\gamma_0}^{(1)\prime} U \\
&= O_p\left(\frac{1}{T^{1-2\alpha}} \frac{1}{T^\alpha}\right) + O_p\left(\frac{1}{T^{1/2-\alpha}} \frac{1}{T^{1/2}}\right) + O_p\left(\frac{1}{T^{1-2\alpha}} \frac{1}{T^{1/2}}\right) \\
&= O_p\left(\frac{1}{T^{1-\alpha}}\right).
\end{aligned}$$

By the $\sqrt{T}$ consistency of the OLS estimator, we have

$$\widehat{\beta}^{(1)}(\gamma^0) - \beta^{(1)} = O_p\left(\frac{1}{T^{1/2}}\right)$$

and

$$\begin{aligned}
\widehat{\beta}^{(1)}(\gamma) - \beta^{(1)} &= \left(\widehat{\beta}^{(1)}(\gamma) - \widehat{\beta}^{(1)}(\gamma^0)\right) + \left(\widehat{\beta}^{(1)}(\gamma^0) - \beta^{(1)}\right) \\
&= O_p\left(\frac{1}{T^{1-\alpha}}\right) + O_p\left(\frac{1}{T^{1/2}}\right) = O_p\left(\frac{1}{T^{1/2}}\right). \qquad (19)
\end{aligned}$$

Moreover, since

$$\widehat{\delta}(\gamma) = \left(X_\gamma' X_\gamma\right)^{-1} X_\gamma'(X_0\delta + U) = \left(X_\gamma' X_\gamma\right)^{-1} X_\gamma' X_0 \delta + \left(X_\gamma' X_\gamma\right)^{-1} X_\gamma' U$$

and

$$\widehat{\delta}(\gamma^0) = \left(X_0' X_0\right)^{-1} X_0'(X_0\delta + U) = \delta + \left(X_0' X_0\right)^{-1} X_0' U,$$

we have

$$\begin{aligned}
\widehat{\delta}(\gamma) - \widehat{\delta}(\gamma^0) &= \left(X_\gamma' X_\gamma\right)^{-1} X_\gamma'(X_0 - X_\gamma)\delta + \left(X_\gamma' X_\gamma\right)^{-1} X_\gamma' U - \left(X_0' X_0\right)^{-1} X_0' U \\
&= O_p\left(\frac{1}{T^{1-2\alpha}} T^{-\alpha}\right) + O_p\left(\frac{1}{T^{1/2-\alpha}} \frac{1}{T^{1/2}}\right) + O_p\left(\frac{1}{T^{1-2\alpha}} T^{-1/2}\right) \\
&= O_p\left(\frac{1}{T^{1-\alpha}}\right).
\end{aligned}$$

We also have

$$\widehat{\delta}(\gamma^0) - \delta = \left(X_0'X_0\right)^{-1} X_0'U = O_p\left(\frac{1}{T^{1/2}}\right).$$

Therefore,

$$\begin{aligned}
\widehat{\delta}(\gamma) - \delta &= \left(\widehat{\delta}(\gamma) - \widehat{\delta}(\gamma^0)\right) + \left(\widehat{\delta}(\gamma^0) - \delta\right) \\
&= O_p\left(\frac{1}{T^{1-\alpha}}\right) + O_p\left(\frac{1}{T^{1/2}}\right) = O_p\left(\frac{1}{T^{1/2}}\right). \qquad (20)
\end{aligned}$$

By (19) and (20), we have,

$$\begin{aligned}
&RSS_T(\gamma) - RSS_T\left(\gamma^0\right) \\
&= \left(Y - X\widehat{\beta}^{(1)}(\gamma) - X_\gamma\widehat{\delta}(\gamma)\right)' \left(Y - X\widehat{\beta}^{(1)}(\gamma) - X_\gamma\widehat{\delta}(\gamma)\right) \\
&\quad - \left(Y - X\widehat{\beta}^{(1)}(\gamma^0) - X_0\widehat{\delta}(\gamma^0)\right)' \left(Y - X\widehat{\beta}^{(1)}(\gamma^0) - X_0\widehat{\delta}(\gamma^0)\right) \\
&= \left(Y - X\widehat{\beta}^{(1)}(\gamma) - X_\gamma\widehat{\delta}(\gamma)\right)' \left(Y - X\widehat{\beta}^{(1)}(\gamma) - X_\gamma\widehat{\delta}(\gamma)\right) \\
&\quad - \left(Y - X\widehat{\beta}^{(1)}(\gamma) - X_0\widehat{\delta}(\gamma)\right)' \left(Y - X\widehat{\beta}^{(1)}(\gamma) - X_0\widehat{\delta}(\gamma)\right) + o_p(1) \\
&= -2\widehat{\delta}(\gamma)'(X_\gamma - X_0)U + \widehat{\delta}(\gamma)'(X_\gamma - X_0)'(X_\gamma - X_0)\widehat{\delta}(\gamma) \\
&\quad +2\widehat{\delta}(\gamma)'(X_\gamma - X_0)'(X_\gamma - X_0)(\widehat{\beta}^{(1)}(\gamma) - \beta^{(1)}) + o_p(1) \\
&= \delta'(X_\gamma - X_0)'(X_\gamma - X_0)\delta + 2\widehat{\delta}(\gamma)'(X_\gamma - X_0)'(X_\gamma - X_0)(\widehat{\beta}^{(1)}(\gamma) - \beta^{(1)}) \\
&\quad -2\widehat{\delta}(\gamma)'(X_\gamma - X_0)U + (\delta + \widehat{\delta}(\gamma))'(X_\gamma - X_0)'(X_\gamma - X_0)(\widehat{\delta}(\gamma) - \delta) + o_p(1) \\
&= T^{-2\alpha}c'\left(X_\gamma - X_0\right)'\left(X_\gamma - X_0\right)c - 2T^{-\alpha}U'(X_\gamma - X_0)c + O_p(T^{-1/2+\alpha}) + o_p(1) \\
&= R_1 + R_2 + o_p(1)
\end{aligned}$$

where

$$\begin{aligned}
R_1 &= T^{1-2\alpha}\frac{\left[\left(X_\gamma - X_0\right)c\right]'\left(X_\gamma - X_0\right)c}{T} \\
&= T^{1-2\alpha}\frac{1}{T}\sum_{t=p+1}^{T} \left|\left|\sum_{j=2}^{4} c_j' x_t \left(\Psi_t^{(j)}(\gamma) - \Psi_t^{(j)}\left(\gamma^0\right)\right)\right|\right|^2, \qquad (21)
\end{aligned}$$

and

$$\begin{aligned}
R_2 &= -2T^{-\alpha}U'(X_\gamma - X_0)c \qquad (22) \\
&= -2T^{-\alpha}\sum_{j=2}^{4}\sum_{t=p+1}^{T} x_t' u_t \left(\Psi_t^{(j)}(\gamma) - \Psi_t^{(j)}\left(\gamma^0\right)\right)c_j.
\end{aligned}$$

To examine the asymptotic behavior of $RSS_T(\gamma) - RSS_T\left(\gamma^0\right)$, we study the asymptotics of R_1 and R_2. We consider four different cases and provide the proof for the case where $v > 0$ and $\omega > 0$. The proofs for the other 3 cases are analogous.

Case 1: $v > 0$ and $\omega > 0$

$c_2' x_t \left(\Psi_t^{(2)}(\gamma) - \Psi_t^{(2)}(\gamma^0) \right)$
$= c_2' x_t \left(\Psi_t^{(2)}(\gamma_1, \gamma_2) - \Psi_t^{(2)}(\gamma_1^0, \gamma_2) + \Psi_t^{(2)}(\gamma_1^0, \gamma_2) - \Psi_t^{(2)}(\gamma_1^0, \gamma_2^0) \right)$
$= c_2' x_t \left(I(\gamma_1^0 \le z_{1t} < \gamma_1, z_{2t} > \gamma_2) - I(z_{1t} \le \gamma_1^0, \gamma_2^0 \le z_{2t} < \gamma_2) \right)$

$c_3' x_t \left(\Psi_t^{(3)}(\gamma) - \Psi_t^{(3)}(\gamma^0) \right)$
$= c_3' x_t \left(\Psi_t^{(3)}(\gamma_1, \gamma_2) - \Psi_t^{(3)}(\gamma_1^0, \gamma_2) + \Psi_t^{(3)}(\gamma_1^0, \gamma_2) - \Psi_t^{(3)}(\gamma_1^0, \gamma_2^0) \right)$
$= c_3' x_t \left(-I(\gamma_1^0 \le z_{1t} < \gamma_1, z_{2t} \le \gamma_2) + I(z_{1t} > \gamma_1^0, \gamma_2^0 \le z_{2t} < \gamma_2) \right)$

$c_4' x_t \left(\Psi_t^{(4)}(\gamma) - \Psi_t^{(4)}(\gamma^0) \right)$
$= c_4' x_t \left(\Psi_t^{(4)}(\gamma_1, \gamma_2) - \Psi_t^{(4)}(\gamma_1^0, \gamma_2) + \Psi_t^{(4)}(\gamma_1^0, \gamma_2) - \Psi_t^{(4)}(\gamma_1^0, \gamma_2^0) \right)$
$= c_4' x_t \left(-I(\gamma_1^0 \le z_1 < \gamma_1, z_2 > \gamma_2) - I(z_1 > \gamma_1^0, \gamma_2^0 \le z_2 < \gamma_2) \right)$

Summing up the three terms, we have
$\sum_{j=2}^4 c_j' x_t \left(\Psi_t^{(j)}(\gamma) - \Psi_t^{(j)}(\gamma^0) \right)$
$= (c_2 - c_4)' x_t I(\gamma_1^0 \le z_1 < \gamma_1, z_2 > \gamma_2) - c_2' x_t I(z_{1t} \le \gamma_1^0, \gamma_2^0 \le z_{2t} < \gamma_2)$
$\quad - c_3' x_t I(\gamma_1^0 \le z_{1t} < \gamma_1, z_{2t} \le \gamma_2) + (c_3 - c_4)' x_t I(z_1 > \gamma_1^0, \gamma_2^0 \le z_2 < \gamma_2).$

Since the four terms are orthogonal, by Lemma 2, we have

$\frac{1}{T} \sum_{t=p+1}^T || \sum_{j=2}^4 c_j' x_t \left(\Psi_t^{(j)}(\gamma) - \Psi_t^{(j)}(\gamma^0) \right) ||^2$
$\xrightarrow{p} (c_2 - c_4)'(M_2(\gamma_1, \gamma_2) - M_2(\gamma_1^0, \gamma_2))(c_2 - c_4) + c_2'((M_1(\gamma_1^0, \gamma_2) - M_1(\gamma_1^0, \gamma_2^0))c_2$
$\quad + c_3'(M_1(\gamma_1, \gamma_2) - M_1(\gamma_1^0, \gamma_2))c_3 + (c_3 - c_4)'(M_3(\gamma_1^0, \gamma_2) - M_3(\gamma_1^0, \gamma_2^0))(c_3 - c_4).$

By the continuity of $D(\gamma)$ around γ^0, we can apply the first-order Taylor approximation to the moment functionals and obtain the following results:

$$(c_2 - c_4)'(M_2(\gamma_1, \gamma_2) - M_2(\gamma_1^0, \gamma_2))(c_2 - c_4) = |\gamma_1 - \gamma_1^0|(c_2 - c_4)' D f_1^0 (c_2 - c_4) + o(1),$$

$$c_2'((M_1(\gamma_1^0, \gamma_2) - M_1(\gamma_1^0, \gamma_2^0))c_2 = |\gamma_2 - \gamma_2^0| c_2' D f_2^0 c_2 + o(1),$$

$$c_3'(M_1(\gamma_1, \gamma_2) - M_1(\gamma_1^0, \gamma_2))c_3 = |\gamma_1 - \gamma_1^0| c_3{}' D f_1^0 c_3 + o(1),$$

$$(c_3 - c_4)'(M_3(\gamma_1^0, \gamma_2) - M_3(\gamma_1^0, \gamma_2^0))(c_3 - c_4) = |\gamma_2 - \gamma_2^0|(c_3 - c_4)' D f_2^0 (c_3 - c_4)) + o(1),$$

where $f_i^0 = \frac{\partial F(\gamma)}{\partial \gamma_i}|_{\gamma=\gamma^0}$ for $i = 1, 2$ and $D = E\left(x_t x_t' | z_t = \gamma_0\right).$

Thus,

$$\begin{aligned}
R_1 &= T^{1-2\alpha}\frac{1}{T}\sum_{t=p+1}^{T} ||\sum_{j=2}^{4} c_j' x_t \left(\Psi_t^{(j)}(\gamma) - \Psi_t^{(j)}\left(\gamma^0\right)\right)||^2 \\
&= T^{1-2\alpha}(|\gamma_1 - \gamma_1^0|(c_2 - c_4)' D f_1^0 (c_2 - c_4) + |\gamma_2 - \gamma_2^0| c_2' D f_2^0 c_2 \\
&+ |\gamma_1 - \gamma_1^0| {c_3}' D f_1^0 c_3 + |\gamma_2 - \gamma_2^0|(c_3 - c_4)' D f_2^0 (c_3 - c_4)) + o_p(1) \\
&= |\upsilon| d_1' D^* f_1^0 d_1 + |\omega| {d_2}' D^* f_2^0 d_2 + o_p(1), \qquad (23)
\end{aligned}$$

where $D^* = diag\{D, D\}, d_1 = ((c_2 - c_4)', c_3')', d_2 = (c_2', (c_3 - c_4)')'$.

Next, we consider the asymptotic property of R_2 for $\upsilon > 0$ and $\omega > 0$:

$T^{-\alpha}\sum_{t=p+1}^{T} x_t' u_t \left(\Psi_t^{(2)}(\gamma) - \Psi_t^{(2)}\left(\gamma^0\right)\right) c_2$
$= -2T^{-\alpha}\sum_{t=p+1}^{T} x_t' u_t \left(\Psi_t^{(2)}(\gamma_1, \gamma_2) - \Psi_t^{(2)}(\gamma_1^0, \gamma_2) + \Psi_t^{(2)}(\gamma_1^0, \gamma_2) - \Psi_t^{(2)}(\gamma_1^0, \gamma_2^0)\right) c_2$
$= -2T^{-\alpha}\sum_{t=p+1}^{T} x_t' u_t ((\gamma_1^0 \le z_{1t} < \gamma_1, z_{2t} > \gamma_2) - I(z_{1t} \le \gamma_1^0, \gamma_2^0 \le z_{2t} < \gamma_2)) c_2$
$\Rightarrow -2(B_1(\upsilon) - B_2(\omega)) c_2;$

$T^{-\alpha}\sum_{t=p+1}^{T} x_t' u_t \left(\Psi_t^{(3)}(\gamma) - \Psi_t^{(3)}\left(\gamma^0\right)\right) c_3$
$= -2T^{-\alpha}\sum_{t=p+1}^{T} x_t' u_t \left(\Psi_t^{(3)}(\gamma_1, \gamma_2) - \Psi_t^{(3)}(\gamma_1^0, \gamma_2) + \Psi_t^{(3)}(\gamma_1^0, \gamma_2) - \Psi_t^{(3)}(\gamma_1^0, \gamma_2^0)\right) c_3$
$= -2T^{-\alpha}\sum_{t=p+1}^{T} x_t' u_t (-I(\gamma_1^0 \le z_{1t} < \gamma_1, z_{2t} \le \gamma_2) + I(z_{1t} > \gamma_1^0, \gamma_2^0 \le z_{2t} < \gamma_2)) c_3$
$\Rightarrow -2(-B_3(\upsilon) + B_4(\omega)) c_3;$

$T^{-\alpha}\sum_{t=p+1}^{T} x_t' u_t \left(\Psi_t^{(4)}(\gamma) - \Psi_t^{(4)}\left(\gamma^0\right)\right) c_4$
$= -2T^{-\alpha}\sum_{t=p+1}^{T} x_t' u_t \left(\Psi_t^{(4)}(\gamma_1, \gamma_2) - \Psi_t^{(4)}(\gamma_1^0, \gamma_2) + \Psi_t^{(4)}(\gamma_1^0, \gamma_2) - \Psi_t^{(4)}(\gamma_1^0, \gamma_2^0)\right) c_4$
$= -2T^{-\alpha}\sum_{t=p+1}^{T} x_t' u_t (-I(\gamma_1^0 \le z_{1t} < \gamma_1, z_{2t} > \gamma_2) - I(z_{1t} > \gamma_1^0, \gamma_2^0 \le z_{2t} < \gamma_2)) c_4$
$\Rightarrow -2(-B_1(\upsilon) - B_4(\omega)) c_4.$

Summing up the three terms, we have

$$\begin{aligned}
R_2 &= -2T^{-\alpha}\sum_{j=2}^{4}\sum_{t=p+1}^{T} x_t' u_t \left(\Psi_t^{(j)}(\gamma) - \Psi_t^{(j)}\left(\gamma^0\right)\right) c_j \\
&\Rightarrow -2[B_1(\upsilon)(c_2 - c_4) - B_2(\omega) c_2 - B_3(\upsilon) c_3 + B_4(\omega)(c_3 - c_4)], \qquad (24)
\end{aligned}$$

where $B_j(\cdot)$, $j=1,2,3,4$, are independent Brownian motion vectors corresponding to the four disjointed regions. The covariance matrix of $B_j(\cdot)$ is given by

$$\begin{aligned} E\left(B_j(1)\,B_j(1)'\right) &= Vf_1^0, \qquad \text{for } j=1,3, \\ E\left(B_j(1)\,B_j(1)'\right) &= Vf_2^0, \qquad \text{for } j=2,4, \end{aligned}$$

where $V=E\left(x_t x_t' u_t^2 | z_t=\gamma_0\right)=\sigma^2 D$ and $f_i^0=\dfrac{\partial F(\gamma)}{\partial \gamma_i}|_{\gamma=\gamma^0}$ for $i=1,2$.

Let $B_1^*(v)=(B_1(v),-B_3(v)), B_2^*(\omega)=(-B_2(\omega),B_4(\omega))$. $B_1^*(v)$ and $B_2^*(\omega)$ are two independent Brownian motion vectors with covariance matrix $E\left(B_1^*(1)\,B_1^*(1)'\right)=V^*f_1^0$, $E\left(B_2^*(1)\,B_2^*(1)'\right)=V^*f_2^0$ respectively, where $V^*=diag\{V,V\}$.

Thus, (24) can be rewritten as

$$\begin{aligned} R_2 &\Rightarrow -2[B_1^*(v)(c_2'-c_4',c_3')'+B_2^*(\omega)(c_2',c_3'-c_4')'] \\ &= -2[B_1^*(v)d_1+B_2^*(\omega)d_2] \\ &= -2(\sqrt{d_1'V^*d_1f_1^0}W_1(v)+\sqrt{d_2'V^*d_2f_2^0}W_2(\omega)), \end{aligned} \tag{25}$$

where $W_1(v)$ and $W_2(\omega)$ are independent standard Brownian motions.

Similarly, for the other three cases, we can show

$$R_1=|v|d_1'D^*f_1^0d_1+|\omega|d_2{}'D^*f_2^0d_2+o_p(1);$$

$$R_2\Rightarrow -2(\sqrt{d_1'V^*d_1f_1^0}W_1(v)+\sqrt{d_2'V^*d_2f_2^0}W_2(\omega)).$$

Making the change-of-variables

$$v=\frac{d_1'V^*d_1}{(d_1'D^*d_1)^2f_1^0}r_1,$$

$$\omega=\frac{d_2'V^*d_2}{(d_2'D^*d_2)^2f_2^0}r_2,$$

and noting $d_1'V^*d_1=\sigma^2d_1'D^*d_1$ and $d_2'V^*d_2=\sigma^2d_2'D^*d_2$, we have

$$\begin{aligned} & RSS_T(\gamma)-RSS_T\left(\gamma^0\right)\xrightarrow{p} R_1+R_2 \\ \Rightarrow\ & \sigma^2|r_1|+\sigma^2|r_2|-2\sigma^2W_1(r_1)-2\sigma^2W_2(r_2) \\ =\ & \sigma^2(|r_1|+|r_2|-2W_1(r_1)-2W_2(r_2)). \end{aligned}$$

Define

$$\lambda_T=(\frac{(d_1'V^*d_1)f_1^0}{\sigma^2},\frac{(d_2'V^*d_2)f_2^0}{\sigma^2}).$$

The asymptotic distribution can be expressed as

$$
\begin{aligned}
& T^{1-2\alpha}\lambda_T\left((\widehat{\gamma}_1-\gamma_1^0),\ (\widehat{\gamma}_2-\gamma_2^0)\right) \\
= \ & (r_1, r_2) \\
\Rightarrow \ & \underset{-\infty<v_1<\infty,-\infty<\omega_1<\infty}{\arg\min}\left[(\frac{1}{2}\left|r_1\right|-W_1\left(r_1\right))+(\frac{1}{2}\left|r_2\right|-W_2\left(r_2\right))\right] \\
= \ & \underset{-\infty<r_1<\infty,-\infty<r_2<\infty}{\arg\max}\left[(-\frac{1}{2}\left|r_1\right|+W_1\left(r_1\right))+(-\frac{1}{2}\left|r_2\right|+W_2\left(r_2\right))\right]. \qquad (26)
\end{aligned}
$$

■

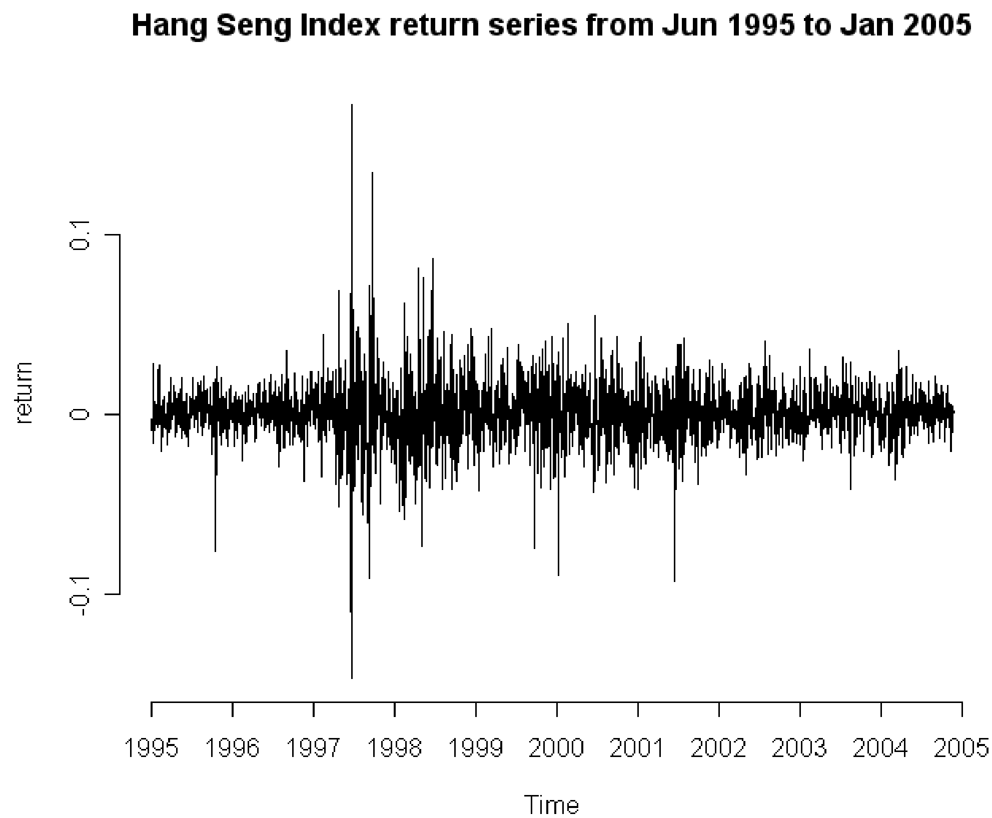

Figure 1: Hang Seng Index Return Series

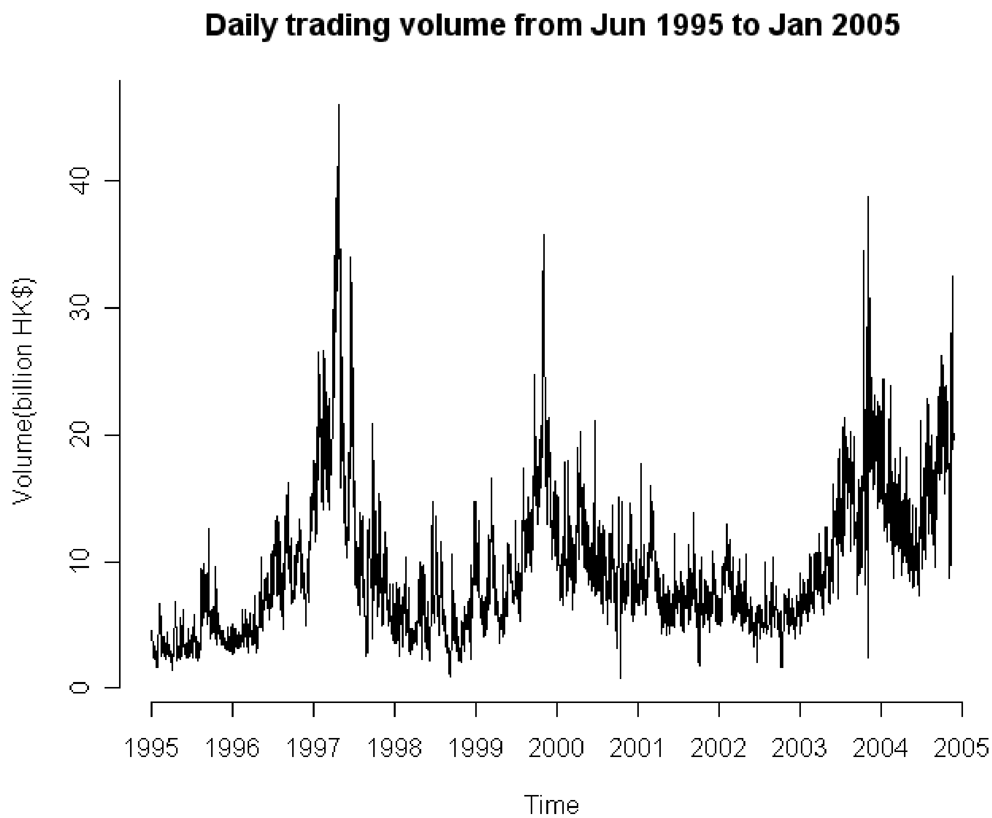

Figure 2: Hang Seng Index Daily Trading Volume

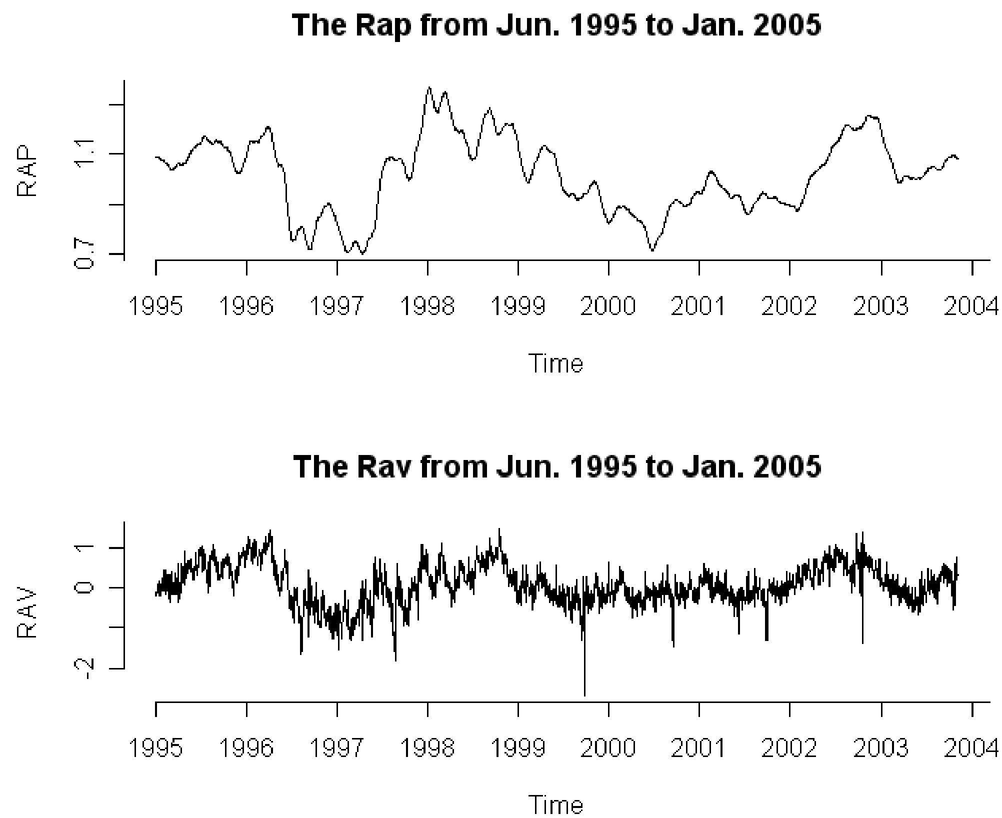

Figure 3: The Two Threshold Variables

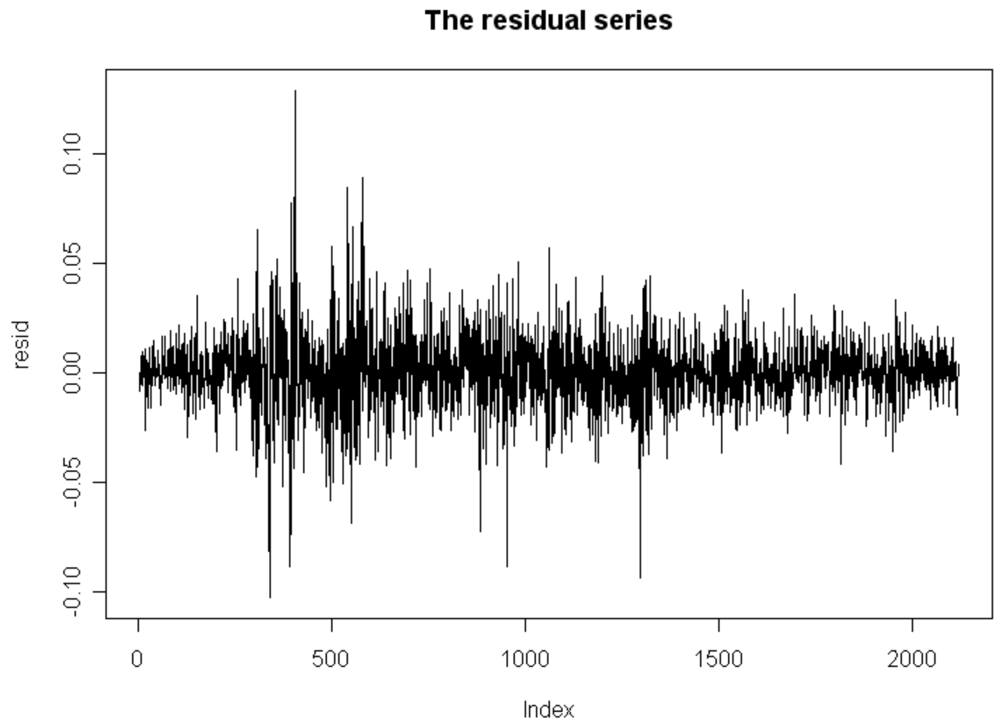

Figure 4: The Residual Series