Estimation of Change-points in ARMA-GARCH/IGARCH and General Time Series Models

Shiqing Ling∗

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

Abstract
This paper first develops a general theory for estimating change-points in a general class of linear and nonlinear time series models. Based on a general objective function, it is shown that the estimated change-point converges weakly to the location of the maxima of a double-sided random walk and other estimated parameters are asymptotically normal. When the magnitude \(d\) of changed parameters is small, it is shown that the limiting distribution can be approximated by the known distribution as in Yao (1987). This provides a channel to connect our results with those in Picard (1985) and Bai, Lumsdaine and Stock (1998), where the magnitude of changed parameters depends on the sample size \(n\) and tends to zero as \(n \to \infty\). We then focus on the self-weighted QMLE and the local QMLE of structure-change ARMA-GARCH/IGARCH models. The limiting distribution of the estimated change-point and its approximating distribution are obtained. Some simulation results are reported and one real example is given.

Keywords: ARMA, change-point, GARCH, near epoch dependence, time series models.
JEL classification: C13, C22.

1. Introduction
Structural change has been recognized as an important issue in econometrics, engineering, and statistics for a long time. Literature in this area is extensive. The earliest references go back to Chow (1960) and Quandt (1960). Many approaches have been developed to detect whether or not structural change exists in a statistical model. Examples are the weighted likelihood ratio test in Picard (1985), and Andrews and Ploberger (1994); Wald and Lagrange multiplier tests in Hansen (1993), Andrews (1993), and Bai and Perron (1998); the exact likelihood ratio test in Horváth (1993) and Davis, Huang and Yao (1995); the empirical approach in Bai (1996) for regression models; and the sequential test in Lai (1995). The structural change occurs in many time series data, see Stock and Watson (1996) and Hansen (2001). Ling (2007) developed a general theory on the Quandt-type tests for time series models. Aue, et al. (2009) studied the break detection in the covariance structure of multivariate time series models. Shao and Zhang (2010) studied Quandt-type test for the change of mean in time series. This type of test was further developed by Hidalgo and Seo (2011) under a more general framework. Empirically, we want to know not only that structural change exists, but also the location of change-point.

The first paper on the estimation of change-points is by Hinkley (1970), in which he investigated the maximum likelihood estimator (MLE) of the change-points in a sequence of i.i.d. random variables and proved that the estimated change-point converges in distribution to the location of the maxima of a double-sided random walk. Under the normality assumption, he showed that the limiting distribution can be tabulated by a numerical method. Hinkley and Hinkley (1970) used a similar method to investigate the binomial random variables and showed that the limiting distribution has a computable form. However, for the nonnormal or non-binomial cases, their results cannot be used as statistical inference for the change point. When the magnitude \(d\) of changed parameters is small, Yao (1987) showed that Hinkley’s (1970) limiting distribution can be approximated by a very nice distribution. Dmbgen (1991)

∗Corresponding author. Tel.: +852 23587459; fax: +852 23591016.
Email address: maling@ust.hk (Shiqing Ling)
investigated the nonparametric method for change-point estimators. Ritov (1990) studied the asymptotic efficient estimation of the change-point. Bai (1995) studied a structure-changed regression model with a fixed \( d \) and showed that the estimated change-point converges in distribution to the location of the maxima of a double-sided random walk. Qu and Perron (2007) investigated estimating and testing structural changes in multivariate regressions.

In the field of time series, Picard (1985) first studied the MLE of change-points in AR models. She assumed that the magnitude of changed parameters is \( d_0 \) which depends on the sample size \( n \) with \( d_0 \to 0 \) as \( n \to \infty \), and obtained the same limiting distribution as that in Yao (1987). Picard’s method was developed for the regression models by Bai (1994, 1995). Bai, Lumsdaine and Stock (1998) also used Picard’s method for the structure-changed multivariate AR model and cointegrating time series model, see also Chong (2001) for AR(1) models and Ling (2003b) for ARMA-GARCH models. Davis, et. al (2006) proposed a minimum description length principle to locate the change points in the multiple structural change AR models. Under Hinkley’s framework, as far as we know, no result has been obtained for the limiting distribution of the estimated change-points with a fixed \( d \) in time series models.

This paper first develops a general theory for estimating change-points in a general class of linear and nonlinear time series models. Based on a general objective function, it is shown that the estimated change-point converges weakly to the location of the maxima of a double-sided random walk and other estimated parameters are asymptotically normal. When the magnitude of changed parameters is small, it is shown that the limiting distribution can be approximated by the known distribution as in Yao (1987). This provides a channel to connect our results with those in Picard (1985) and Bai, Lumsdaine and Stock (1998). We then focus on the self-weighted QMLE and the local QMLE of structure-change ARMA-GARCH models. The limiting distribution of the estimated change-point and its approximating distribution are obtained. Some simulation results are reported and one real example is given.

This paper proceeds as follows. Section 2 presents general results. Section 3 gives the approximating distribution of the estimated change-points. Section 4 presents the results for the structure-change ARMA-GARCH models. Section 5 reports simulation results and gives one real example. Sections 6-7 give the proofs of results in Sections 2 and 4, respectively. Section 8 gives the proofs of three technical lemmas.

2. General Results

Assume that the real time series \( \{y_t : t = 0, \pm 1, \pm 2, \cdots \} \) is \( \mathcal{F}_t \)-measurable, strictly stationary and ergodic, and is generated by

\[
y_t = g(\theta, Y_{t-1}, e_t),
\]

where \( \mathcal{F}_t \) is the \( \sigma \)-field generated by \( \{e_t, e_{t-1}, \cdots \} \), \( Y_t = (y_0, \cdots, y_{t-p+1}) \) or \( Y_t = (y_t, y_{t-1}, \cdots) \), \( \theta \) is an \( m \times 1 \) unknown parameter vector, and \( \{e_t\} \) is independently and identically distributed (i.i.d.). The structure of the time series \( \{y_t\} \) is characterized by \( g \) and the parameter \( \theta \). We assume that the parameter space \( \Theta \) is a bounded compact subset of \( \mathbb{R}^m \).

We denote model (2.1) by \( M(\theta_0) \) when the true parameter is \( \theta = \theta_0 \). Let \( \{y_1, \cdots, y_n\} \) be the random sample. We assume

\[
\{y_1, \cdots, y_k\} \in M(\theta_{10}) \text{ and } \{y_{k+1}, \cdots, y_n\} \in M(\theta_{20}) \text{ with } \theta_{10} \neq \theta_{20}.
\]

When \( 1 \leq k < n \), we parameterize it as \( k = [n\tau] \) with \( \tau \in (0, 1) \), where \( [x] \) is the integer part of \( x \). \( k = [n\tau] \) is called the unknown change-point and \( k_0 = [n\tau_0] \) is its true change-point. For each \( k \), we use the following objective function to estimate \( \theta_{10} \) and \( \theta_{20} \), based on the pre-sample and the post-sample, respectively:

\[
L_{1n}(k, \theta_1) = \sum_{i=1}^{k} l_i(\theta_1) \text{ and } L_{2n}(k, \theta_2) = \sum_{i=k+1}^{n} l_i(\theta_2),
\]

where \( l_i(\theta) = l(\theta, y_t, y_{t-1}, \cdots) \) is a measurable function in terms of \( \{y_t\} \) and is almost surely (a.s.) continuous with respect to \( \theta \). The objective function based on the whole sample is

\[
L_n(k, \theta_1, \theta_2) = L_{1n}(k, \theta_1) + L_{2n}(k, \theta_2).
\]
We can take \( l_i(\theta) \) as that in LSE, MLE, quasi-MLE, LAD-type or M-estimators, among others. Assume \( \theta_{10} \) and \( \theta_{20} \) are interior points in \( \Theta \). When \( t > k_0 \), \( l_i(\theta) = l(\theta, y_t, \ldots, y_{t-1}, y_{t+k_0}) \) and when \( t \leq k_0 \), \( l_i(\theta) = l(\theta, y_t, \ldots, y_1, Y_0) \). We assume

\[
Y_{k_0} \in M(\theta_{20}) \quad \text{and} \quad Y_0 \in M(\theta_{10}). \tag{2.3}
\]

We will discuss the initial value issue in Section 4 for ARMA-GARCH models.

Let \( \hat{\theta}_{1n}(k) \) and \( \hat{\theta}_{2n}(k) \) be the maximizers of \( L_{1n}(k, \hat{\theta}) \) and \( L_{2n}(k, \hat{\theta}) \) on \( \Theta \) for each known \( k \). \( k_0 \) is estimated by

\[
\hat{k}_n = \text{argmax}_{1 \leq k \leq L_n} \left[ k, \hat{\theta}_{1n}(k), \hat{\theta}_{2n}(k) \right].
\]

In practice, the range \( 1 < k < n \) can be replaced by \( m \leq k \leq n - m \). Other parameters are estimated by

\[
(\hat{\theta}_{1n}, \hat{\theta}_{2n}) = \text{argmax}_{(\theta_1, \theta_2) \in \Theta} \left[ L_n(\hat{k}_n, \theta_1, \theta_2) \right].
\]

This procedure is to run two sequential estimation for the same model. Given the advanced computing technology today, it is not difficult to implement such a procedure. It has been used for AR models and the regression models, see for examples, Bai (1995) and Bai, Lumsdaine and Stock (1998). We now introduce two assumptions as follows.

**Assumption 2.1.** When \( \{y_t, s \leq t \} \in M(\theta_{10}) \), \( E_{\theta_{10}}[l_i(\theta)] < \infty \) and \( E[l_i(\theta)] \) has a unique maximizer at \( \theta = \theta_{10} \).

**Assumption 2.2.** When \( \{y_t, t = 1, \ldots, n \} \in M(\theta_{10}) \), it follows that

\[
\frac{1}{n} \sup_{\theta \in \Theta} \left| \sum_{t=0}^{n-1} [l_i(\theta) - E(l_i(\theta))] \right| = o(1), \text{ a.s., as } n \to \infty.
\]

We should mention that the ergodic theorem cannot be applied to Assumption 2.2. We need to check its near-epoch dependence (NED). A time series \( \{X_t\} \) is called to be \( L^2(\nu) \) NED in terms of \( \{\varepsilon_t\} \) if \( \sup_{-\infty < \nu < \infty} \|X_\nu\|_p < \infty \) and

\[
\sup_{-\infty < \nu < \infty} \|X_\nu - E[X_\nu|\mathcal{F}_0]\|_p = O(k^{-\nu}),
\]

where \( \|A\| = [\text{tr}(AA')]^{1/2} \) for a vector or matrix \( A \), \( \mathcal{F}(j) \) is the \( \sigma \)-field generated by \( \{\varepsilon_i, \varepsilon_{i-1}, \ldots, \varepsilon_{i-\nu+1}\} \) with \( i \geq 1 \), and \( \mathcal{F}_0(j) = \{\emptyset, \Omega\}, \rho \geq 1 \) and \( \nu > 0 \). This holds for many time series models. Theorem 2.1 of Ling (2007) can be used to verify Assumption 2.2 if \( l_i(\theta) \) is \( L^2(\nu) \) NED, see the proof of Theorem 4.1 in Section 7.

**Theorem 2.1.** (a) If Assumption 2.1 holds, then \( \hat{\theta}_n(k) = \theta_{10} + o_p(1) \), \( i = 1, 2 \), uniformly in \( \tau \in (0, 1) \). (b) If Assumptions 2.1-2.2 hold, then \( \hat{k}_n = k_0 + O_p(1) \).

This theorem implies that the rate of convergence of \( \hat{\tau}_n \), the estimator of \( \tau_0 \), is \( n \) which is faster than that in Picard (1985) and Bai, Lumsdaine and Stock (1998) for AR models.

**Assumption 2.3.** When \( \{y_t, t = 1, \ldots, n \} \in M(\theta_{10}) \), the following statements hold:

(i) for any \( \theta_n \to \theta_{10} \), it follows that

\[
\sum_{i=1}^{n} [l_i(\theta_n) - l_i(\theta_{10})] = (\theta_n - \theta_{10})' \sum_{i=1}^{n} D_i(\theta_{10}) - n(\theta_n - \theta_{10})' \left[ \frac{1}{2} \sum_{i=1}^{n} \varepsilon_{i} + o_p(1) \right] (\theta_n - \theta_{10}),
\]

(ii) \( D_i(\theta_{10}) \) is a martingale difference in terms of \( \mathcal{F}_i \) with the covariance \( \Omega_{\varepsilon_{i}} \),

(iii) \( \Omega_{\varepsilon_{i}} \) and \( \Sigma_{\varepsilon_{i}} \) are positive definite matrices.

This assumption holds for the various estimation of time series models. The conditions for Assumption 2.3(i) is given in Ling and McAleer (2010) for a differentiable \( l_i(\theta) \). We now define a double-sided random walk:

\[
W_d(k) = \begin{cases} 
\sum_{i=1}^{k} [l_i(\theta_{10}) - l_i(\theta_{20})], & k > 0, \\
0, & k = 0, \\
\sum_{i=1}^{-k} [l_i(\theta_{20}) - l_i(\theta_{10})], & k < 0.
\end{cases}
\]

where \( d = \theta_{10} - \theta_{20} \). We should mention that \( y_i \in M(\theta_{20}) \) when \( k > 0 \) and \( y_i \in M(\theta_{10}) \) when \( k < 0 \). The limiting distribution of \( (\hat{k}_n, \hat{\theta}_{1n}, \hat{\theta}_{2n}) \) is as follows.
Theorem 2.2. If Assumptions 2.1-2.3 hold when \( \theta_0 = \theta_{10} \) and \( \theta_{20} \), respectively, then \( \hat{k}_n, \hat{\theta}_{1n} \) and \( \hat{\theta}_{2n} \) are asymptotically independent, and

\[
\sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) \to_d N(0, \frac{1}{\tau_0} \Sigma_{\theta_{10}}^{-1} \Omega_{\theta_{10}} \Sigma_{\theta_{10}}^{-1}) \quad \text{and} \quad \sqrt{n}(\hat{\theta}_{2n} - \theta_{20}) \to_d N(0, \frac{1}{1 - \tau_0} \Sigma_{\theta_{20}}^{-1} \Omega_{\theta_{20}} \Sigma_{\theta_{20}}^{-1}),
\]

(b) \( \hat{k}_n - k_0 \to_d \arg\max_k W_d(k) \).

Unlike the i.i.d. case in Hinkley (1970) and Bai (1995), the double-sided random walk \( W(d, k) \) is neither independent nor symmetric.

3. Approximating Distribution of Estimated \( k_0 \)

The distribution of \( \arg\max_k W_d(k) \) does not have a closed form. It is difficult to be used directly for statistical inference. This section investigates the limiting distribution of \( \arg\max_k W_d(k) \) when \( \|d\| \to 0 \). Note that \( y_i \in M(\theta_{10}) \) is a function of \( \theta_{10} \) and \( \{\epsilon_i\} \) and similarly for \( y_i \in M(\theta_{20}) \). Thus, \( y_i \) changes when the value of \( d \) is changed. To make it simple, we fix \( \theta_{20} \) and assume that \( d = \theta_{10} - \theta_{20} \to 0 \). In this case, when \( k > 0, y_i \in M(\theta_{20}) \) and is not changed when \( d \to 0 \). But when \( k < 0, y_i \in M(\theta_{10}) \) and is changed when \( d \to 0 \). To make it clear, when \( y_i \in M(\theta_{10}), j_i \) is denoted by \( y_i, \epsilon_i \) by \( \epsilon_i, h_i \) by \( h_i, D_0(\theta) \) by \( D_0(\theta), \) and \( P_i(\theta) \) by \( P_i(\theta), i = 1, 2, \ldots \). We make the following assumptions.

Assumption 3.1. Let \( m = [(d' \Sigma_{\theta_{20}}d)^{-1}(d' \Omega_{\theta_{20}}d)] \). For each \( z \in \mathbb{R} \), we have

\[
\sum_{i=[mz]}^{[mz]} D_{20}(\theta_{20}) = \sum_{i=[mz]}^{[mz]} D_{20}(\theta_{20}) - \frac{1}{2} d' \Sigma_{\theta_{20}} + o_p(1) \quad \text{as} \quad d \to 0.
\]

where \( o_p(1) \to 0 \) in probability as \( d \to 0 \).

Assumption 3.2. \( D_{20}(\theta_{20}) \) is \( L^{2+\epsilon}(\nu) \) NED in terms of \( \{\epsilon_i\} \) with \( \epsilon_i > 0 \), where either \( 2\nu > 1 \) or \( 2\nu = 1 \) and there exist constants \( \nu_1 > 0 \) and \( \nu_1 > 0 \) with \( 2\nu_1 > 1 \) such that

\[
\sup_{-\infty < z < \infty} \|E[D_{20}|F_{k+1}(t)] - E[D_{20}|F_k(t)]\|_{L^{2+\epsilon}} = O(k^{-\nu_1}).
\]

We note that the usual invariance principle for the forward sum can not be applied for the backward sum \( \sum_{i=-k}^{i=-1} D_{20}(\theta_{20}) \). This assumption is to use the invariance principle in Ling (2007). Our approximating distribution is as follows.

Theorem 3.1. Suppose that Assumptions 2.3 and 3.1-3.2 hold. Then

\[
(d' \Sigma_{\theta_{20}}d)^{-1}(d' \Omega_{\theta_{20}}d)^{-1} \arg\max_k W_d(k) \to_d \arg\max_{\gamma \in \mathbb{R}} \left[B(\gamma) - \frac{1}{2} |\gamma|\right],
\]

as \( 0 < \|d\| \to 0 \), where \( B(\gamma) \) is the standard Brownian motion in \( \mathbb{R} \).

Proof. First, by Assumption 2.3(ii)-(iii) and Assumption 3.2, and Theorem 2.3 of Ling (2007), we can show that

\[
W'_n(z) \equiv \frac{1}{m} \sum_{i=1}^{[mz]} D_i(\theta_{20}) \to \Omega_{\theta_{20}}^{1/2} B(z),
\]

\[
W''_n(z) \equiv \frac{1}{m} \sum_{i=1}^{[mz]} D_i(\theta_{20}) \to \Omega_{\theta_{20}}^{1/2} B(z),
\]

on \( C[-M, M] \) for any given \( M \), as \( m \to \infty \), where \( B(z) \) is a standard Brownian motion on \( C[-\infty, \infty] \). Let \( F_d(x) \) be the distribution of \( \arg\max_k W_d(k) \). It is the same as that of \( \arg\max_{\gamma \in \mathbb{R}} \left[\gamma_d W_d([\gamma])\right] \), where \( \gamma_d = (d' \Sigma_{\theta_{20}}d)(d' \Omega_{\theta_{20}}d)^{-1} \). Thus, there is a \( M \) such that

\[
\left| F_d(x) - P(\arg\max_{\gamma \in [-x, x]} \left[\gamma_d W_d([\gamma])\right] \leq x) \right| \leq \epsilon,
\]

as \( 0 < \|d\| \to 0 \).
when $s > M$. Thus,

$$
|F_d(x) - P(m \arg\max_{z \in [-M, M]} [\gamma_d W_d(|mz|)] \leq x) | = |F_d(x) - P(\arg\max_{z \in [-M, M]} [\gamma_d W_d(|z|)] \leq x) | \leq \varepsilon.
$$

By Assumptions 2.3(i) and 3.1, we can show that $\gamma_d W_d(|mz|)$ has the uniform expansion on $z \in [-M, M]$,

$$
\gamma_d W_d(|mz|) = -\frac{1}{2} |z| + \gamma_d' W_m(z) I[z > 0] + \gamma_d W_m(z) I[z \leq 0] + o_p(1),
$$

$$
\rightarrow_L -\frac{1}{2} |z| + B(z) \text{ on } C[-M, M],
$$

where the last step holds by (3.2)-(3.3). Note that the probability of $-|z|/2 + B(z)$ when $z \notin [-M, M]$ is small as $M$ is large. By the previous equations and using the continuous mapping theorem for the argmax function, we can claim that the conclusion holds. □

Yao (1987) showed that the distribution $F(x)$ of $\arg\max_{\gamma} [B(\gamma) - |\gamma|/2]$ has the density function:

$$
f(x) = \frac{3}{2} e^{i|g|} \frac{3}{2} \left( \frac{|x|}{2} \right) - \frac{1}{2} \Phi\left( \frac{|x|}{\sqrt{2}} \right) \text{ and } \Phi(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{u^2}{2} \right) du,
$$

where $x \in R$. By Theorem 3.1, it is reasonable to approximate the distribution of $(d' \Sigma d)^{1/2} (d' \Omega d)^{-1}(\hat{k}_n - k_0)$ by $F(x)$ when $d$ is small. $F(x)$ can be used to construct the confidence interval of $k_0$ and its percentiles can be found in Yao (1987). The simulation results in Yao (1987) for i.i.d. data show that $F(x)$ approximates the empirical distribution of $\hat{k}_n$ very well in finite samples. For time series models, some simulation results can be found in Bai, et al (1998) and Ling (2003). We note that our framework is different from that in Picard (1985) and Bai, et al (1998), where they assume that $\theta_0 = \theta_1$ and $\theta_0 = \theta_1$, $d = d_o = \theta_1 - \theta_2 \to 0$ and $|d_o| \sqrt{n} \to \infty$ as the sample size $n \to \infty$. They estimate $(\theta, \theta_0)$ and show that the limiting distribution of the normalized $\hat{k}_n$ is $F(x)$. Their true parameter $(\theta, \theta_0)$ is changed with $n$, while the true parameter $(\theta, \theta_0)$ in our model is fixed and hence $d$ is fixed. By Theorem 3.1, the statistical inference on $k_0$ based on the two frameworks are nearly identical when $d$ or $d_o$ is small.

4. Estimation of Change-point in ARMA GARCH/IGARCH Model

This section considers the following autoregressive moving-average (ARMA) model with the generalized autoregressive conditional heteroscedasticity (GARCH) errors:

$$
\phi(B)y_t = \psi(B)\varepsilon_t, \quad (4.1)
$$

$$
\varepsilon_t = \eta_t \sqrt{h_t} \text{ and } h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j h_{t-j}, \quad (4.2)
$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$, $\psi(z) = 1 + \psi_1 z + \cdots + \psi_q z^q$, $p$, $q$, $r$ and $s$ are known, $\alpha_0 > 0$, $\alpha_i > 0$ ($i = 1, \ldots, r$), $\beta_j \geq 0$ ($j = 1, \ldots, s$), and $\eta_t$ is a sequence of i.i.d. random variables with zero mean and variance 1. Denote $\gamma = (\mu, \phi_1, \ldots, \phi_p, \psi_1, \ldots, \psi_q)'$, $\delta = (\alpha_0, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s)'$, and $\theta = (\gamma, \delta)'$. The parameter subspaces is $\Theta = \Theta_\gamma \times \Theta_\delta$, where $\Theta_\gamma \subset R^{p+q+1}$ and $\Theta_\delta \subset R^{r+s+1}$ are compact, where $R = (-\infty, \infty)$ and $R_0 = [0, \infty)$. We denote mode (4.1)-(4.2) by $M(\theta_0)$ when the true value of $\theta$ is $\theta_0$. We introduce the following conditions:

**Assumption 4.1.** For each $\vartheta \in \Theta$, $\phi(\vartheta) \neq 0$ and $\psi(\vartheta) \neq 0$ when $|\vartheta| \leq 1$, and $\phi(z)$ and $\psi(z)$ have no common root with $\phi_p \neq 0$ or $\psi_q \neq 0$.

**Assumption 4.2.** $\alpha(z)$ and $\beta(z)$ have no common root, $\alpha(1) \neq 0$, $\alpha_r + \beta_s \neq 0$, and $\sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j \leq 1$.

**Assumption 4.3.** $\eta_t^2$ has a non-degenerate distribution with $E\eta_t^2 = 1$. 

5
We assume that
\[ \{y_1, \cdots, y_n\} \in M(\theta_0) \text{ and } \{y_{k+1}, \cdots, y_n\} \in M(\theta_2), \]  
and \( \theta_0 \) and \( \theta_2 \) are interior points in \( \Theta \). We first consider the self-weighted quasi-maximum likelihood estimator (SQMLE) of parameters \((k_0, \theta_0, \theta_2)\). In this case,
\[ L_{1n}(k, \theta_1) = \sum_{i=1}^{k} w_i l_i(\theta_1) \text{ and } L_{2n}(k, \theta_2) = \sum_{i=k+1}^{n} w_i l_i(\theta_2), \]
where
\[ l_i(\theta) = -\frac{1}{2} \left[ \log h_i(\theta) + \frac{\varepsilon_i^2(\theta)}{h_i(\theta)} \right]. \]

\[ \varepsilon_i(\theta) = y_i - \mu - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \psi_i \varepsilon_{t-i}(\theta), \]
where \( h_i(\theta) = \alpha_0 + \sum_{i}^{p} \alpha_i \phi_i + \sum_{i}^{q} \beta_i \varepsilon_{t-i}(\theta) \) and \( w_i = [1 + \sum_{i=1}^{p} i^{-2} |y_{t-i}|]^{-1}. \)

The weighted function here is just for simplicity. We refer to Ling (2007) for other choices.

We assume the initial condition (2.3) is satisfied. Denote \( U_i(\theta) = \left\{ (\partial \varepsilon_i(\theta)/\partial \theta)/\sqrt{h_i(\theta)}, \partial h_i(\theta)/\partial \theta / [\sqrt{h_i(\theta)}] \right\} \) and \( \xi_i(\theta) = [\varepsilon_i(\theta)/\sqrt{h_i(\theta)}], (1 - \varepsilon_i^2(\theta)/h_i(\theta))/\sqrt{2}]. \) Then
\[ D_i(\theta) = \partial h_i(\theta)/\partial \theta = U_i(\theta) \xi_i(\theta), \]
\[ P_i(\theta) = -\frac{\partial^2 \varepsilon_i(\theta)}{\partial \theta^2} = U_i(\theta) U_i(\theta)' + \left[ \frac{\varepsilon_i^2(\theta)}{h_i(\theta)} - 1 \right] R_{1i}(\theta) + \frac{\varepsilon_i(\theta)}{\sqrt{h_i(\theta)}} R_{2i}(\theta), \]
where
\[ R_{1i}(\theta) = [\partial h_i(\theta)/\partial \theta][\partial h_i(\theta)/\partial \theta]^{-1} \frac{1}{2} h_i^2(\theta) - [\partial^2 h_i(\theta)/\partial \theta^2][\partial h_i(\theta)/\partial \theta]^{-1} \frac{1}{2} h_i^2(\theta) \] and
\[ R_{2i}(\theta) = 2 (\partial \varepsilon_i(\theta)/\partial \theta)[\partial h_i(\theta)/\partial \theta]/\sqrt{h_i(\theta)} \].

Let \( J = E[\xi_i(\theta) \xi_i(\theta)'] \). We have the result as follows.

**Theorem 4.1.** Suppose that Assumptions 4.1-4.3 hold, \( E\eta^4 < \infty \) and \( J > 0 \). Then, \( \hat{k}_n, \hat{\theta}_{1n}, \hat{\theta}_{2n} \) are asymptotically independent, and
\[ (a) \quad \sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) \longrightarrow_L N(0, \frac{1}{\tau_0} \sum_{i=1}^{i} \Omega_1 \Sigma_i^{-1}), \quad \sqrt{n}(\hat{\theta}_{2n} - \theta_{20}) \longrightarrow_L N(0, \frac{1}{1 - \tau_0} \sum_{i=1}^{i} \Omega_2 \Sigma_i^{-1}), \]
\[ (b) \quad \hat{k}_n - k_0 \longrightarrow_L \arg \max_k W_{ad}(k), \]
where \( \Sigma_0 = E[w_i U_i(\theta_0) U_i(\theta_0)'] \) and \( \Omega_0 = E[w_i^2 U_i(\theta_0) U_i(\theta_0)] \) as \( i = 1, 2 \) and \( W_{ad}(k) \) is defined as \( W_{ad}(k) \) in Theorem 2.3 with \( l_i(k) \) replaced by \( w_i l_i(k) \).

The SQMLE of \((\theta_{10}, \theta_{20})\) may not be as efficient as its QMLE, see a discussion in Ling (2007). This may affect the estimator of \( k_0 \). We now consider the local QMLE without a weighted function \( w_i \). Specifically, using \( \theta_{ad} \) in Theorem 4.1 as an initial estimator of \( \theta_{ad}, i = 1, 2 \), the local QMLE is obtained via the following one-step iteration:
\[ \hat{\theta}_{1n} = \hat{\theta}_{1n} - \frac{k_n}{\sum_{i=1}^{k_n} P_i(\hat{\theta}_{1n})} \sum_{i=1}^{k_n} D_i(\hat{\theta}_{1n}) \text{ and } \hat{\theta}_{2n} = \hat{\theta}_{2n} - \frac{n}{\sum_{i=k_n+1}^{n} P_i(\hat{\theta}_{2n})} \sum_{i=k_n+1}^{n} D_i(\hat{\theta}_{2n}). \]
\[ \hat{k}_n = \arg \max_k \left[ \sum_{i=1}^{k_n} l_i(\hat{\theta}_{1n}) + \sum_{i=k_n+1}^{n} l_i(\hat{\theta}_{2n}) \right]. \]

For this local QMLE, we have the following result:

**Theorem 4.2.** Suppose that Assumptions 4.1-4.3 hold, \( E\eta^4 < \infty \) and \( J > 0 \). If \((\hat{\theta}_{1n}, \hat{\theta}_{1n}, \hat{k}_n)\) is obtained through (4.8)-(4.9), then
\[ (a) \quad \sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) \longrightarrow_L N(0, \frac{1}{\tau_0} \sum_{i=1}^{i} \Omega_1 \Sigma_i^{-1}), \quad \sqrt{n}(\hat{\theta}_{2n} - \theta_{20}) \longrightarrow_L N(0, \frac{1}{1 - \tau_0} \sum_{i=1}^{i} \Omega_2 \Sigma_i^{-1}), \]
\[ (b) \quad \hat{k}_n - k_0 \longrightarrow_L \arg \max_k W_{ad}(k), \]
where \( \Sigma_i = E[U_i(\theta_0) U_i(\theta_0)'] \) and \( \Omega_i = E[U_i(\theta_0) U_i(\theta_0)], i = 1, 2. \)
The approximating distribution in Theorem 3.1 can be used for both $\hat{k}_n$ and $\tilde{k}_n$. For simplicity, we only state it for $\hat{k}_n$ as follows.

**Theorem 4.3.** Suppose that Assumptions 4.1-4.3 hold, $E\gamma^4 < \infty$ and $J > 0$. Then, as $0 < ||d|| \to 0$, it follows:

$$(d'\Sigma_{20}d)^2(d'\Omega_{20}d)^{-1}\text{argmax}_\gamma W_d(k) \to _{\text{L}} \text{argmax}_{\gamma \in \mathcal{B}} [B(\gamma) - \frac{1}{2}|\gamma|].$$

For model (4.1)-(4.2), the initial condition (2.3) is not satisfied in practice. We have to replace $Y_0$ by some constant $\tilde{Y}_0$ because it is not available. With these initial values, the expansions in Assumption 2.3(i) and Assumption 3.1 still hold and hence they do not affect the asymptotic results of $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$, see Zhu (2011) for model (4.1)-(4.2). Ling and McAleer (2010) gave a set of initial conditions for a class of time series models. To see their effect on the estimated change-point $k_0$, we denote

$$\tilde{l}(\theta_t) = l(\theta_1, y_t, \cdots, y_T, \tilde{Y}_0)$$

when $t \leq k_0$ and

$$\tilde{l}(\theta_t) = l(\theta_2, y_t, \cdots, y_{T+n}, \tilde{Y}_0)$$

when $t > k_0$.

From the proof of Theorem 2.3, we can show that

$$\hat{k}_n - k_0 = \text{argmin}_k W_d(k) + o_p(1),$$

where

$$\tilde{W}_d(k) = \begin{cases} 
\sum_{i=k+1}^{T} \tilde{l}(\theta_i) - l(\theta_i), & k > k_0, \\
0, & k = k_0, \\
\sum_{i=k}^{T} \tilde{l}(\theta_i) - l(\theta_i), & k < k_0.
\end{cases}$$

It can be seen that $W_d(k)$ and $\tilde{W}_{d-r-k}(k)$ are always different and hence the initial values always affect the asymptotic distribution of the estimated $k_0$. When $||d|| \to 0$, by Lemma 8.2, we can show that

$$\sum_{i=1}^{k_0} E[|l(\theta_{10}) - \tilde{l}(\theta_{10})| + |l(\theta_{20}) - \tilde{l}(\theta_{20})|] + \sum_{k_0+1}^{\infty} E[|l(\theta_{10}) - \tilde{l}(\theta_{10})| + |l(\theta_{20}) - \tilde{l}(\theta_{20})|] = o(1).$$

(4.10)

Thus, we can see that the approximation distribution in Theorem 4.3 is still valid in this case.

Model (4.1)-(4.2) includes the ARMA model (i.e. the case with $h_t = \alpha_0$) and the GARCH model (i.e. the case with $y_t = \epsilon_t$) as two important special cases. By deleting the corresponding components in Theorem 4.1, we can obtain the asymptotic results of the self-weighted LSE of the structural change ARMA model with the finite or infinite variance errors. By deleting the corresponding components in Theorem 4.2, we can obtain the asymptotic results of the local QMLE of the structural change GARCH/IGARCH models. Similarly, the approximating distribution in Theorem 4.3 still can be used. Even for the two special cases, our results are the first time to be given in the literature.

5. **Empirical Study**

We first examine the performance of our asymptotic results in the finite samples via some Monte Carlo experiments. The data is generated by the following AR(1)-GARCH(1,1) model:

$$y_t = \phi_1 y_{t-1} I_{[0,\infty)}(\epsilon_t) + \phi_2 y_{t-1} I_{(\infty,\infty)}(\epsilon_t) + \epsilon_t,$$

$$\epsilon_t = \eta_t \sqrt{h_t}$$

where $\eta_t \sim \text{i.i.d.}N(0,1)$. The true parameters are $\theta_{10} = (0.2, 1.0, 0.2, 0.7)'$ and $\theta_{20} = (0.6, 0.5, 0.05, 0.95)'$. We use 4000 replications in all the experiments. These experiments are carried out by Fortran 77 and the optimization algorithm from Fortran subroutine DBCOAH in the IMSL library is used.

In Tables 1 and 2, we summarize the empirical means and standard deviations (SD) of the SMLE and the local MLEs of $\theta_{10}$ and $\theta_{20}$. From the two tables, we see that the SDs of $\hat{\theta}_{1n}$ and $\hat{\theta}_{1n}$ are decreased, while those of $\hat{\theta}_{2n}$ and
As \( k_0 \) is increased, the sample size \( n \) is increased from 250 to 400, both the corresponding biases and SDs become smaller. This is the same as the usual results in the AR-GARCH model.

### Table 1

<table>
<thead>
<tr>
<th>( k_0 )</th>
<th>( \phi_{1n} )</th>
<th>( \alpha_{10n} )</th>
<th>( \alpha_{1n} )</th>
<th>( \beta_{1n} )</th>
<th>( \phi_{2n} )</th>
<th>( \alpha_{20n} )</th>
<th>( \alpha_{2n} )</th>
<th>( \beta_{2n} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>124 Mean</td>
<td>.4954</td>
<td>.9566</td>
<td>.1707</td>
<td>.7251</td>
<td>-.5884</td>
<td>.4933</td>
<td>.5128</td>
<td>.0457</td>
</tr>
<tr>
<td>SD</td>
<td>.0847</td>
<td>.1748</td>
<td>.1105</td>
<td>.0985</td>
<td>.0726</td>
<td>.1271</td>
<td>.1949</td>
<td>.0971</td>
</tr>
<tr>
<td>125 Mean</td>
<td>.4952</td>
<td>.9580</td>
<td>.1706</td>
<td>.7251</td>
<td>-.5885</td>
<td>.4934</td>
<td>.5131</td>
<td>.0449</td>
</tr>
<tr>
<td>SD</td>
<td>.0845</td>
<td>.1692</td>
<td>.1104</td>
<td>.0980</td>
<td>.0728</td>
<td>.1277</td>
<td>.1960</td>
<td>.0988</td>
</tr>
<tr>
<td>126 Mean</td>
<td>.4952</td>
<td>.9579</td>
<td>.1712</td>
<td>.7246</td>
<td>-.5887</td>
<td>.4937</td>
<td>.5119</td>
<td>.0452</td>
</tr>
<tr>
<td>SD</td>
<td>.0841</td>
<td>.1686</td>
<td>.1098</td>
<td>.0975</td>
<td>.0730</td>
<td>.1291</td>
<td>.1973</td>
<td>.0996</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>( k_0 )</th>
<th>( \phi_{1n} )</th>
<th>( \alpha_{10n} )</th>
<th>( \alpha_{1n} )</th>
<th>( \beta_{1n} )</th>
<th>( \phi_{2n} )</th>
<th>( \alpha_{20n} )</th>
<th>( \alpha_{2n} )</th>
<th>( \beta_{2n} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>198 Mean</td>
<td>.4966</td>
<td>.9884</td>
<td>.1823</td>
<td>.7144</td>
<td>-.5968</td>
<td>.4851</td>
<td>.5408</td>
<td>.0495</td>
</tr>
<tr>
<td>SD</td>
<td>.0655</td>
<td>.1696</td>
<td>.0858</td>
<td>.0843</td>
<td>.0556</td>
<td>.0994</td>
<td>.1524</td>
<td>.0741</td>
</tr>
<tr>
<td>200 Mean</td>
<td>.4977</td>
<td>.9690</td>
<td>.1815</td>
<td>.7162</td>
<td>-.5931</td>
<td>.4943</td>
<td>.5288</td>
<td>.0399</td>
</tr>
<tr>
<td>SD</td>
<td>.0650</td>
<td>.1331</td>
<td>.0852</td>
<td>.0744</td>
<td>.0562</td>
<td>.0996</td>
<td>.1534</td>
<td>.0755</td>
</tr>
<tr>
<td>202 Mean</td>
<td>.4977</td>
<td>.9692</td>
<td>.1816</td>
<td>.7161</td>
<td>-.5932</td>
<td>.4944</td>
<td>.5288</td>
<td>.0399</td>
</tr>
<tr>
<td>SD</td>
<td>.0646</td>
<td>.1317</td>
<td>.0848</td>
<td>.0738</td>
<td>.0566</td>
<td>.0999</td>
<td>.1543</td>
<td>.0762</td>
</tr>
</tbody>
</table>

Table 3 reports the median (Med), mean, 90% range, estimated asymptotic confidence interval (EACI), and asymptotic confidence interval (ACI) of \( k_0 \) when \( n = 250 \) and 400. The empirical mean is the average of \( \hat{k}_n \) from the 4000 replications. The empirical median and the 90% range are respectively the 50%-quantile and the range between the 5% and 95% quantiles of the distribution of \( \hat{k}_n \). The EACI and ACI are computed by the following formulas:

\[
\text{EACI} = \left[ \hat{k}_n - [\Delta F_{\omega/2}] - 1, \hat{k}_n - [\Delta F_{\omega/2}] + 1 \right]
\]

and

\[
\text{ACI} = \left[ \hat{k}_0 - [\Delta F_{\omega/2}] - 1, \hat{k}_0 - [\Delta F_{\omega/2}] + 1 \right],
\]

respectively, where \( F_{\omega/2} \) is the \( \omega \)th quantile of the distribution \( F \) and \( \Delta = (d' \Omega_{\omega/2} d)(d' \Sigma_{\omega/2} d)^{-2} \) for \( \hat{k}_n \), and similarly for \( \hat{k}_{1n} \) and \( \hat{k}_{2n} \). Using the density function \( f(x) \) in Section 2, we obtain \( F_{0.05} = 7.792 \). \( \Sigma_{\omega/2} \) is estimated by \( n^{-1} \sum_{t=1}^{n} \partial^2 l(\hat{\theta}_n)/\partial \theta \partial \theta' \), and similarly for \( \Omega_{\omega/2} \). From Table 3, we see that the median and mean are unbiased in all cases, and the ACI is exactly the same as EACI. The 90% range is slightly wider than EACI and ACI in all cases. As \( n \) is increased from 250 to 400, the EACIs and ACIs of \( \hat{k}_n - k_0 \) have not been improved. This is because \( \hat{k}_n \) is not a consistent estimator of \( k_0 \). This finding is similar to those in Bai (1995) for the structure-change regression model and in Bai et al. (1998) for the structure-change multivariate AR models and cointegrating time series models.
We now use our theory to study the Hang Seng Index (HSI). It is one of the most important indices in the Asian financial market and has been extensively investigated in the literature. It is well-known that South-East Asia’s economic crisis actually started from Thailand and then spread into the Philippines, Malaysia before the economic and financial crisis occurred in Hong Kong around October in 1997. This finding is reasonable since South-East Asia’s economic crisis actually started from Thailand and then spread into the Philippines, Malaysia.

The mean and standard deviation of $y_i$ are $-2.87 \times 10^{-4}$ and 0.02169, respectively. The $t$-test statistic shows that the mean of $y_i$ is equal to zero (at the 0.05 significant level). The autocorrelation check also shows that $y_i$ is not white noise. Thus, we use the first-order AR-GARCH model without intercept to fit the data $y_i$. The model is as follows:

$$y_i = 0.0745 y_{i-1} + \epsilon_i,$$

$$\epsilon_i = \eta_i \sqrt{h_i} \text{ and } h_i = 4.9 \times 10^{-6} + 0.1346 \epsilon_{i-1}^2 + 0.8594 h_{i-1}. \tag{5.1}$$

The log-likelihood is $-476.61$. To check whether or not model (5.1)-(5.2) is adequate for the data $y_i$, we use the portmanteau test statistic $Q(K)$ in Li and Mak (1994) for the square-residual sequence $\{x_i^2(\hat{\theta}_n)/h_i(\hat{\theta}_n) - 1, i = 2, \cdots, 521\}$. It is shown that the $p$-values corresponding to $Q(6), Q(12)$ and $Q(18)$ are 0.7281, 0.8366 and 0.8099, respectively, and hence model (5.1)-(5.2) is adequate for data $y_i$. We use the Sup-Lagrange multiplier statistic in Andrews (1993), $\sup_{\tau \in [0.05, 0.95]} LM(\tau)$, to test whether or not there exists a structural change in the model. It is obtained that $\sup_{\tau \in [0.05, 0.95]} LM(\tau) = 22.3112$ which exceeds the critical value 17.56 (at the 0.05 significance level), and hence it is believed that a structural change exists in the model.

The structure-change AR-GARCH model is used to fit the data and the following model is obtained:

$$y_i = 0.0673 y_{i-1} I_{[5 \leq 324]} + 0.0616 y_{i-1} I_{[5 > 324]} + \epsilon_i,$$

$$\epsilon_i = \eta_i \sqrt{h_i} \text{ and } h_i = (3.4 \times 10^{-7} + 0.0125 \epsilon_{i-1}^2 + 0.9862 h_{i-1}) I_{[5 \leq 324]} + (4.8 \times 10^{-5} + 0.1789 \epsilon_{i-1}^2 + 0.7849 h_{i-1}) I_{[5 > 324]}. \tag{5.3}$$

The log-likelihood is $-324.24$. The test statistic $Q(K)$ is used for the square-residual sequence $\{x_i^2(\hat{\theta}_n)/h_i(\hat{\theta}_n) - 1, [x_i^2(\hat{\theta}_n)/h_j(\hat{\theta}_n)] - 1, i = 1, \cdots, 324, j = 325, 521\}$, and the $p$-values of $Q(6), Q(12)$ and $Q(18)$ are 0.9157, 0.9804 and 0.9565, respectively. Thus, model (5.3)-(5.4) is also adequate for the data $y_i$. From the values of the log-likelihood and the related $p$-values of $Q(K)$, model (5.3)-(5.4) seems more suitable than model (5.1)-(5.2) for the data $y_i$.

The MLE of change-point is $\hat{k}_n = 324$ and its 90% confidence interval is $[321, 327]$. This tight interval indicates the MLE of $k_0$ is very accurate. After checking with the data, it is found that the estimated change-point $\hat{k}_n$ is on August 17, 1997. From this analysis, our finding is that the structure of HSI had been changed around August 17 in 1997 before the economic and financial crisis occurred in Hong Kong around October in 1997. This finding is reasonable since South-East Asia’s economic crisis actually started from Thailand and then spread into the Philippines, Malaysia.
Figure 1: HSI

Figure 2: HSI
and Indonesia around July and August in 1997. In this period, the financial and stock markets in these countries had been quite unstable. For example, the Thai central bank depleted its reserves of foreign currency and the currency of Thailand was devalued by more than 10 percent in July. The stock mark in Hong Kong was sensitive to such unfavorable information and seems as a leading indicator of the coming crisis.

6. Proofs of Theorems 2.1-2.2

Proof of Theorem 2.1. We prove only for the case when \( k \leq k_0 \), while other case is similar.

(a). First, when \( y_t \in \mathcal{M}(\theta) \), by the ergodic theorem and Assumption 2.1, we can show that

\[
\lim_{n \to \infty} \max_{\theta} \left| \frac{1}{\tau n} \sum_{t=\lfloor \tau n \rfloor + 1}^{\lfloor \tau n \rfloor} [l_t(\theta) - EL_t(\theta)] \right| = 0 \text{ a.s.,} \tag{6.1}
\]

for any fixed \( \tau_1 \) and \( \tau_2 \), as \( n \to \infty \). Note that

\[
0 \leq L_n[k, \hat{\theta}_{1a}(k), \hat{\theta}_{2a}(k)] - L_n(k_0, \theta_{10}, \theta_{20})
= \sum_{k_1=1}^{k_1} [l_t(\hat{\theta}_{1a}(k_1)] - L_n(\theta_{10}) + \sum_{k_1=k_0+1}^{n} [l_t(\hat{\theta}_{2a}(k)] - L_n(\theta_{20})]
- \sum_{k_1=k_0}^{n} [l_t(\hat{\theta}_{2a}(k)] - L_n(\theta_{10})]. \tag{6.2}
\]

The third term is bounded by

\[
2 \max_{1 \leq k \leq k_0} \left| \sum_{i=1}^{k} [l_t(\theta) - EL_t(\theta)] \right| + (k_0 - k) \max_{\theta} [EL_t(\theta_1) - EL_t(\theta_{10})] = o_p(n),
\]

by (6.1). Let \( \Theta_\delta = \{ \theta_t : \|\theta_t - \theta_{10}\| \geq \delta \} \). By Assumption 2.1, \( C = \max_{\theta \in \Theta_\delta} [EL_t(\theta_1) - EL_t(\theta_{10})] < 0 \) when \( t \leq k_0 \) and \( \max_{\theta} [EL_t(\theta_2) - EL_t(\theta_{20})] = 0 \) when \( t > k_0 \). On the event \( \|\theta_{1a}(k) - \theta_{10}\| \geq \delta \), by (6.1) and (6.2),

\[
0 \leq \frac{2}{n} \max_{1 \leq k \leq k_0} \left| \sum_{i=1}^{k} [l_t(\theta_{1a}(k)] - EL_t(\theta_{1a}(k)) \right| + \frac{2}{n} \max_{\theta} \left| \sum_{i=1}^{n} [l_t(\theta_{2a}(k)] - EL_t(\theta_{2a}(k)) \right|
+ \tau_0 \max_{\theta} [EL_t(\theta_1) - EL_t(\theta_{10})] + (1 - \tau_0) \max_{\theta} [EL_t(\theta_2) - EL_t(\theta_{20})] + o_p(1) = \tau_0 C + o_p(1),
\]

uniformly in \( \tau \in (0, 1) \). Thus,

\[
P(\max_{1 \leq k \leq k_0} \|\theta_{1a}(k) - \theta_{10}\| \geq \delta) \leq P \left( \bigcup_{k=1}^{k_0} \{ |\tau_0 C + o_p(1) \geq 0 \} \right) \leq P(\tau_0 C + o_p(1) \geq 0) \to 0,
\]

as \( n \to \infty \), i.e., \( \hat{\theta}_{1a}(k) = \theta_{10} + o_p(1) \) uniformly in \( \tau \in (0, 1) \). Similarly, we can show that \( \hat{\theta}_{2a}(k) = \theta_{20} + o_p(1) \) uniformly in \( \tau \in (0, 1) \). Thus, (a) holds.

(b). We note that

\[
\sum_{i=1}^{k} l_t(\hat{\theta}_{1a}(k)] + \sum_{i=k_0+1}^{n} l_t(\hat{\theta}_{2a}(k)] \geq L_n[k_0, \hat{\theta}_{1a}(k_0), \hat{\theta}_{2a}(k_0)] \geq \sum_{i=1}^{k_0} l_t(\hat{\theta}_{1a}(k_0)] + \sum_{i=k_0+1}^{n} l_t(\hat{\theta}_{2a}(k_0)]
\]

Thus,

\[
- \sum_{i=k_0}^{k} l_t(\hat{\theta}_{1a}(k_0)] + \sum_{i=k_0+1}^{n} l_t(\hat{\theta}_{2a}(k_0)] \geq 0.
\]

By (a) of this Theorem, Assumption 2.1 and the dominated convergence theorem, we have

\[
\lim_{n \to \infty} [EL_t(\hat{\theta}_{2a}(k_n))] - EL_t(\hat{\theta}_{1a}(k_n)]) \leq \lim_{n \to \infty} \max_{1 \leq k \leq n} [EL_t(\hat{\theta}_{2a}(k)] - EL_t(\hat{\theta}_{1a}(k)]) = EL_t(\theta_{20}) - EL_t(\theta_{10}) \equiv -C < 0.
\]
By (6.3) and (6.4), we have
\[
\frac{1}{k_0 - \hat{k}_n} \left\{ - \sum_{k=1}^{k_0} (l_i[I_1(\theta_{1n}(k))] - EL_I[\theta_{1n}(\hat{k}_n)]) + \sum_{k=1}^{k_0} (l_i[I_2(\theta_{2n}(\hat{k}_n)]) - l_i[I_2(\theta_{2n}(\hat{k}_n))]) \right\} \geq EL[I[\theta_{1n}(\hat{k}_n)] - EL[I[\theta_{2n}(\hat{k}_n)]] = C + o(1).
\]

From this, it follows that
\[
\frac{2}{k_0 - \hat{k}_n} \sup_{\theta} \left| \sum_{k=1}^{k_0} [l_i(\theta) - EL_I(\theta)] \right| \geq C + o(1).
\]

By Assumption 2.2, for any \( \varepsilon > 0 \), we have
\[
P(k_0 - \hat{k}_n > M) = P(k_0 - \hat{k}_n > M, \frac{2}{k_0 - \hat{k}_n} \sup_{\theta} \left| \sum_{k=1}^{k_0} [l_i(\theta) - EL_I(\theta)] \right| \geq C + o(1)) \leq P(\max_{k_0 - \hat{k}_n > M} \frac{2}{k_0 - \hat{k}_n} \sup_{\theta} \left| \sum_{k=1}^{k_0} [l_i(\theta) - EL_I(\theta)] \right| \geq C + o(1)) < \varepsilon,
\]
as \( M > 0 \) is large enough, i.e., \( k_0 - \hat{k}_n = O_P(1) \). This completes the proof. \( \Box \)

**Proof of Theorem 2.2.** Denote
\[
\hat{u}_1 = \sqrt{k_0}(\hat{\theta}_{1n}(k) - \theta_{10}),
\hat{u}_2 = \sqrt{n - k_0}(\hat{\theta}_{2n}(k) - \theta_{20}),
\] 
\[
u_i^* = \frac{\Sigma^{-1}_{\theta_{10}}}{\sqrt{k_0} \Sigma_{\theta_{10}}} \sum_{t=1}^{k_0} D_t(\theta_{10}),
\] 
\[ u_i^* = \frac{\Sigma^{-1}_{\theta_{10}}}{\sqrt{n - k_0}} \sum_{t=k_0+1}^{n} D_t(\theta_{10}).
\] 

By Assumption 2.3, we have
\[
\sum_{i=1}^{k_0} [l_i[I_1(\theta_{10})] - l_i(\theta_{10})] = \hat{u}_1^* \Sigma_{\theta_{10}} u_1^* - \hat{u}_1^* \left[ \frac{1}{2} \Sigma_{\theta_{10}} + o_P(1) \right] \hat{u}_1 = \frac{1}{2} \left[ ||\Sigma_{\theta_{10}}^{1/2}(\hat{u}_1 - u_1^*)||^2 + ||u_1^*||^2 \right] (1 + o_P(1)), \tag{6.3}
\]
\[
\sum_{i=k_0+1}^{n} [l_i[I_2(\theta_{20})] - l_i(\theta_{20})] = \hat{u}_2^* \Sigma_{\theta_{20}} u_2^* - \hat{u}_2^* \left[ \frac{1}{2} \Sigma_{\theta_{20}} + o_P(1) \right] \hat{u}_2 = \frac{1}{2} \left[ ||\Sigma_{\theta_{20}}^{1/2}(\hat{u}_2 - u_2^*)||^2 + ||u_2^*||^2 \right] (1 + o_P(1)). \tag{6.4}
\]

Denote
\[
\Delta L_n(k, \theta_{1}, \theta_{2}) = L_n(k, \theta_{1}, \theta_{2}) - L_n(k_0, \theta_{10}, \theta_{20}).
\]

By (6.3) and (6.4), we have
\[
\Delta L_n \left[ k, \hat{\theta}_{1n}(k), \hat{\theta}_{2n}(k) \right] = - \frac{1}{2} \left[ ||\Sigma_{\theta_{10}}^{1/2}(\hat{u}_1 - u_1^*)||^2 + ||\Sigma_{\theta_{20}}^{1/2}(\hat{u}_2 - u_2^*)||^2 \right] (1 + o_P(1)) + \frac{1}{2} \left( ||u_1^*||^2 + ||u_2^*||^2 \right) (1 + o_P(1)) + \sum_{i=k_0+1}^{k_0} [l_i[I_2(\theta_{2n}(k))] - l_i[I_1(\theta_{1n}(k))]]. \tag{6.5}
\]
Let \( \theta_i' = \theta_0 + u_i', i = 1, 2 \). By Assumption 2.3, similar to (6.5), we have

\[
\Delta L_0(k, \theta_0', \theta_0'') = \frac{1}{2} (||u_1'||^2 + ||u_2'||^2) (1 + o_P(1)) + \sum_{i=k+1}^{k_0} \{l_i(\theta_0'') - l_i(\theta_0') \}.
\]

Note that \( l_i(\theta_0') \to_P l_i(\theta_0) \) and \( l_i(\hat{\theta}_m(k)) \to_P l_i(\theta_0), i = 1, 2 \). Given any \( M > 0 \), when \( k_0 - k < M \), we have

\[
0 \leq \Delta L_0 \left[ k, \hat{\theta}_m(k), \hat{\theta}_m(k) \right] - \Delta L_0(k, \theta_0', \theta_0'') = \frac{1}{2} \left( ||\Sigma_{\theta_0}^{-1/2}(\tilde{u}_1 - u_1')||^2 + ||\Sigma_{\theta_0}^{-1/2}(\tilde{u}_2 - u_2')||^2 \right) (1 + o_P(1)) + o_P(1)
\]

Thus, we can claim that, when \( k_0 - k < M \) for any given \( M > 0 \), we have

\[
\tilde{u}_i - u_i' = o_P(1), \quad i = 1, 2.
\]

Since \( \hat{k}_n - k_0 = O_P(1) \), we can claim that (a) holds by the central limiting theorem. Furthermore, we have

\[
\Delta L_0 \left[ k, \hat{\theta}_m(k), \hat{\theta}_m(k) \right] = \frac{1}{2} \left( ||u_1'||^2 + ||u_2'||^2 \right) + \sum_{i=k+1}^{k_0} \{l_i(\theta_10) - l_i(\theta_10) \} + o_P(1).
\]

Thus, by the strict stationarity of \( \{y_t\} \),

\[
\hat{k}_n - k_0 = \arg\max_k \left\{ \sum_{i=k+1}^{k_0} \{l_i(\theta_10) - l_i(\theta_10) \} + o_P(1) \right\} \to \arg\max_k W_d(k),
\]

as \( n \to \infty \). This completes the proof. \( \square \)

7. Proofs of Theorems 4.1-4.3

We first give one lemma which is used for the \( L^p(\nu) \)-NED of \( w_i \). Their proof are in Section 8.

**Lemma 7.1.** If Assumptions 4.1-4.2 hold, then for and, \( \epsilon \in (0, 1) \), it follows that

\[
(a) \quad E[h_1 - E[h_1|F_{i}(t)]] = O(\rho^\epsilon),
\]

\[
(b) \quad E[\epsilon_i - E[\epsilon_i|F_{i}(t)]] = O(\rho^\epsilon),
\]

\[
(c) \quad E[y_i - E[y_i|F_{i}(t)]] = O(\rho^\epsilon).
\]

**Proof of Theorem 4.1.** Assumption 2.1 was verified in Ling (2007). We only need to verify Assumption 2.2. First, by Assumptions 4.1-4.2, \( \epsilon_i(\theta) \) and \( h_i(\theta) \) have the following expansions:

\[
\epsilon_i(\theta) = \sum_{i=0}^{\infty} a_i(\theta) y_{t-i} \quad \text{and} \quad h_i(\theta) = \sum_{i=0}^{\infty} b_i(\theta) \epsilon_i^2(\theta), \quad (7.1)
\]

where \( \sup_{\theta} a_i(\theta) = O(\rho^\epsilon) \) and \( \sup_{\theta} b_i(\theta) = O(\rho^\epsilon) \) with \( \rho \in (0, 1) \). Thus, we have \( \sup_{\theta \in \Theta} |\epsilon_i(\theta)| \leq \xi_{it}^\epsilon \) and \( \sup_{\theta \in \Theta} |h_i(\theta)| \leq \xi_{it}^\epsilon \), where \( \xi_{it} = C + C \sum_{i=0}^{\infty} \rho^i |y_{t-i}| \) and \( C \) is a constant. Using this, we can show that \( E \sup_{\theta \in \Theta} |w_i l_i(\theta)|^{1+\epsilon} < \infty \).

We now show that

\[
\sup_{\theta} E[w_i l_i(\theta) - E[w_i l_i(\theta)|F_{i}(t)]]^{1+\epsilon} = O(k^{-\epsilon}), \quad (7.2)
\]

for some \( \nu > 0 \). By Lemma 7.1 and (7.1), it is straightforward to show that

\[
\sup_{\theta} E[\epsilon_i^2(\theta) - E[\epsilon_i^2(\theta)|F_{i}(t)]] = O(\rho^\epsilon) \quad \text{and} \quad \sup_{\theta} E[h_i(\theta) - E[h_i(\theta)|F_{i}(t)]] = O(\rho^\epsilon),
\]
for any $t \in (0, 1)$. Furthermore, we can show that, for small enough $t > 0$,

$$E[w_i - E[w_i|F_k(t)]] = O(k^{-\gamma})$$

and

$$\sup_{\Theta} E \left[ \frac{\epsilon_i^2(\theta)}{h_i(\theta)} - E \left[ \frac{\epsilon_i^2(\theta)}{h_i(\theta)} \right] | F_k(t) \right]^2 = o(\epsilon^2).$$

Note that $w_i$ and $h_i$ is bounded and $E[\epsilon_i(\theta)] = 0$. By the previous inequality and Lemma 7.1(d), we have

$$E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} - E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} \right] | F_k(t) \right]^2 \leq O(1) E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} - E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} \right] \right]^2 | F_k(t) |^{2i} \leq O(1) E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} - E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} \right] \right]^2 + O(1) \left( E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} \right] \right)^{2i} w_i - E[w_i|F_k(t)] \right)^{1/2} = O(k^{-\gamma}).$$

where $O(1)$ holds uniformly in $\Theta$. By Holder’s inequality and the previous inequality, we have

$$E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} - E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} \right] | F_k(t) \right]^{1+i} \leq \left[ E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} - E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} \right] \right]^2 | F_k(t) |^{2i} \right]^{1/2} \left[ E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} - E \left[ \frac{w_i \epsilon_i^2(\theta)}{h_i(\theta)} \right] \right]^2 \right]^{1/2} \leq \sqrt{E[w_i \epsilon_i^2(\theta)]^2} O(\epsilon^2) = O(k^{-\gamma}).$$

where $O(1)$ holds uniformly in $\Theta$. Similarly, we can show that $E[w_i \log h_i(\theta) - E[w_i \log h_i(\theta)|F_k(t)] |^{1+i} = O(k^{-\gamma})$ uniformly in $\Theta$. Thus, (7.2) holds.

By (7.2), $(w_i l_i(\theta))$ is an $L^{1+i}(\nu)$-NED sequence. By Theorem 2.1 of Ling (2007), for each $\theta \in \Theta$,

$$\frac{1}{n} \sum_{i=n}^{n-1} w_i l_i(\theta) = E[w_i l_i(\theta)] + o(1). \tag{7.3}$$

Denote $V_\delta = \{ \theta^* : ||\theta - \theta^*|| \leq \delta \} \subset \Theta$. Let

$$\xi_i = \sup_{\theta^* \in V_\delta} |w_i(\theta^*) - w_i(\theta)| \quad \text{and} \quad \tilde{\xi}_i = \sup_{\theta^* \in V_\delta} |E[w_i(\theta^*)|F_k(t)] - E[w_i(\theta)|F_k(t)]|. \tag{7.4}$$

Then

$$|\xi_i - \tilde{\xi}_i| \leq \sup_{\theta^* \in V_\delta} \left| w_i(\theta^*) - E[w_i(\theta)|F_k(t)] - w_i(\theta) + E[w_i(\theta)|F_k(t)] \right| \leq 2 \sup_{\theta^* \in V_\delta} |w_i(\theta^*) - E[w_i(\theta)|F_k(t)]|. \tag{7.5}$$

Furthermore, by (7.2), we have

$$E[|\xi_i - E[\xi_i|F_k(t)]|^{1+i}] \leq E[|\xi_i - \tilde{\xi}_i| + E[|\xi_i - E[\xi_i|F_k(t)]|^{1+i}] \leq CE[|\xi_i - \tilde{\xi}_i|^{1+i}] = O(k^{-\gamma}). \tag{7.6}$$

Thus, $(\xi_i)$ is an $L^{1+i}(\nu)$-NED sequence. By Theorem 2.1 of Ling (2007), we have

$$\frac{1}{n} \sum_{i=n}^{n-1} \xi_i = E\xi_i + o(1). \tag{7.7}$$

Using (7.3) and (7.4) and partitioning $\Theta$ into finite balls with a radian $\delta$ small enough, we can show that Assumption 2.2 holds. By Theorem 2.3, we completes the proof. \qed

**Proof of Theorem 4.2.** First, from the proof of Lemma A.6 in Ling (2007), we can see that

$$\frac{1}{n} \sum_{i=1}^{k_n} \|P_\delta(\hat{\theta}_{in}) - P_\delta(\theta_{10})\| = o_P(1). \tag{7.8}$$
Note that \( \hat{k} - k_0 \) is bounded in probability. When \( \hat{k}_n < k_0 \), we can show that

\[
\frac{1}{n} \sum_{i=r_{k+1}}^{k_0} \left\| P_i(\hat{\theta}_{1n}) \right\| \leq \frac{1}{n} \sum_{i=r_{k+1}}^{k_0} \left\| P_i(\hat{\theta}_{1n}) - P_i(\theta_{10}) \right\| + \frac{1}{n} \sum_{i=r_{k+1}}^{k_0} \left\| P_i(\theta_{10}) \right\| = o_p(1). \tag{7.6}
\]

Thus, by the previous two inequalities, we have

\[
\frac{1}{n} \sum_{i=1}^{k_0} P_i(\hat{\theta}_{1n}) = \tau_0 + o_p(1). \tag{7.7}
\]

By Taylor’s expansion, we have

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{k_0} D_i(\hat{\theta}_{1n}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{k_0} D_i(\theta_{10}) + \left[ \frac{1}{n} \sum_{i=1}^{k_0} P_i(\hat{\xi}_{1n}) \right] \left[ \sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) \right] \tag{7.8}
\]

where \( \hat{\xi}_{1n} \) lies between \( \hat{\theta}_{1n} \) and \( \theta_{10} \). By (7.7)-(7.8), we have

\[
\hat{\theta}_{1n} = \theta_{10} + \frac{\sum_{i=1}^{k_0} D_i(\theta_{10}) + o_p(1)}{\sqrt{n}}.
\]

A similar expansion holds for \( \hat{\theta}_{2n} \). Thus, by the central limiting theorem, we see that (a) holds. For (b), by Taylor’s expansion, we have

\[
L_n[k, \hat{\theta}_{1n}, \hat{\theta}_{2n}] - L_n(k_0, \theta_{10}, \theta_{20}) = \frac{1}{2} \left( \|u_1^*\|^2 + \|u_2^*\|^2 \right) + \sum_{i=r_{k+1}}^{k_0} \left[ l_i(\theta_{20}) - l_i(\theta_{10}) \right] + o_p(1),
\]

where \( u_i^* = \sqrt{n}(\tilde{\theta}_n - \theta_{10}) \), \( i = 1, 2 \). Thus, by the strict stationarity of \( \{y_i\} \),

\[
\hat{k}_n - k_0 = \text{argmax} \left\{ \sum_{i=r_{k+1}}^{k_0} \left[ l_i(\theta_{20}) - l_i(\theta_{10}) \right] + o_p(1) \right\} \rightarrow \chi \text{ argmax}_k W_d(k),
\]

as \( n \to \infty \). This completes the proof. □

We now further give two lemmas. The first one is for the NED properties of \( D_2(\theta_{20}) \) and \( P_2(\theta_{20}) \) which are used to verify both Assumptions 3.1-3.2. The second lemma is used only for Assumption 3.1. The proofs are in Section 8.

**Lemma 7.2.** If \( \{y_i\} \in \mathcal{M}(\theta_0) \) and Assumptions 4.1-4.2 hold, then there exist a constant \( c \in (0,1) \) such that

\[
(a) \quad E \left\| \frac{1}{\sqrt{h_i}} \frac{\partial \xi_i(\theta_0)}{\partial m} - E \left[ \frac{1}{\sqrt{h_i}} \frac{\partial \xi_i(\theta_0)}{\partial m} \right] \xi_i(t) \right\|^2 = O_p(c),
\]

\[
(b) \quad E \left\| \frac{1}{h_i} \frac{\partial h_i(\theta_0)}{\partial \theta} - E \left[ \frac{1}{h_i} \frac{\partial h_i(\theta_0)}{\partial \theta} \right] \xi_i(t) \right\|^2 = O_p(c),
\]

where \( c \) is a constant with \( c \in (0,1) \).

**Lemma 7.3.** If Assumptions of Theorem 4.2 hold, then, as \( d = \theta_{10} - \theta_{20} \to 0 \),

\[
(a) \quad E \left\| \frac{1}{h_{11}} \frac{\partial \xi_{11}(\theta_0)}{\partial m} - \frac{1}{h_{12}} \frac{\partial \xi_{12}(\theta_{20})}{\partial m} \right\|^2 = o(1),
\]

\[
(b) \quad E \left\| \frac{1}{h_{11}} \frac{\partial h_{11}(\theta_0)}{\partial \theta} - \frac{1}{h_{22}} \frac{\partial h_{22}(\theta_{20})}{\partial \theta} \right\|^2 = o(1).
\]

15
Proof of Theorem 4.3. Taking Taylor expansion at \( \theta_{10} \), we have
\[
\sum_{r=-[m]}^{1} [l_1(\theta_{20}) - l_1(\theta_{10})] = - \sum_{r=-[m]}^{1} d' D_1(\theta_{10}) - \frac{1}{2} d'' \sum_{r=-[m]}^{1} P_{11}(\xi^d) d, \tag{7.9}
\]
where \( \xi^d \) lies between \( \theta_{10} \) and \( \theta_{20} \), and \( D_1(\theta_{10}) = U_1(\theta_{10}) \xi_t \), where \( \xi_t = [\eta_t, (1 - \eta_t^2)/\sqrt{2}]' \). Note that \( \xi^* = \theta_{10} + O(1/\sqrt{m}) \). From Lemma A.6 of Ling (2007), we have
\[
\frac{1}{m} \sum_{r=-[m]}^{1} P_{11}(\xi^d) = \frac{1}{m} \sum_{r=-[m]}^{1} P_{11}(\theta_{10}) + o_p(1), \tag{7.10}
\]
where \( P_{11}(\theta_{10}) = U_1(\theta_{10}) U_1'(\theta_{10}) + \eta_t R_{11u}(\theta_{10}) + (\eta_t^2 - 1) R_{12u}(\theta_{10}) \). Using Lemma 7.3, we can show that
\[
\lim_{[d] \to 0} E[||U_1(\theta_{10}) - U_2(\theta_{20})||^2] = 0 \quad \text{and} \quad \lim_{[d] \to 0} ||P_{11}(\theta_{10}) - P_{22}(\theta_{20})|| = 0, \tag{7.11}
\]
where \( P_{22}(\theta_{20}) = U_2(\theta_{20}) U_2'(\theta_{20}) + \eta_t R_{21u} + (\eta_t^2 - 1) R_{22u} \). By (7.9)-(7.11), we can show that
\[
\sum_{r=-[m]}^{1} [l_1(\theta_{20}) - l_1(\theta_{10})] = - \sum_{r=-[m]}^{1} d' U_2(\theta_{20}) \xi_t - \frac{1}{2} d'' \sum_{r=-[m]}^{1} P_{22}(\theta_{20}) d + o_p(1). \tag{7.12}
\]
Using Lemma 7.2, we can show that \( ||U_2(\theta_{20})||^2 \) is \( L^{1+\epsilon}(v) \)-NED. Similarly, we can show that \( ||R_{2i}||, i = 1, 2 \), are also \( L^{1+\epsilon}(v) \)-NED. By Theorem 2.1 of Ling (2007), we have
\[
\frac{1}{m} \sum_{r=-[m]}^{1} U_2(\theta_{20}) U_2(\theta_{20})' = \Sigma_{20} + o_p(1) \quad \text{and} \quad \frac{1}{m} \sum_{r=-[m]}^{1} \eta_t R_{12u} + (\eta_t^2 - 1) R_{22u} = o_p(1).
\]
By the previous three equations, we can see that Assumption 3.1 holds. By Lemma 7.2, \( U_2(\theta_{20}) \) is \( L^{2+\epsilon}(v) \)-NED and hence so is \( D_2(\theta_{20}) \), i.e., Assumption 3.2 holds. By Theorem 3.1, we can complete the proof.

8. Proofs of Lemmas 7.1-7.3

We first gives one lemma, which is from Ling and (1997), see also Bougerol and Picard (1992).

Lemma 8.1. Suppose \( \{y_t\} \in M(\theta) \). If Assumptions 4.1-4.2 hold, then \( h_t \) has the expansion:
\[
h_t = \alpha_0 [1 + \sum_{j=1}^{\infty} \prod_{i=0}^{j-1} A_{i-t-\xi_i}].
\]
where \( \xi_i = (\eta_i^2, 0, \cdots, 0, 1, \cdots, 0)_{(r+j)\times 1}^t \) with the first component \( \eta_i^2 \) and the \( (r+1) \)th component 1, \( u = (0, \cdots, 1, \cdots, 0)_{(r+j)\times 1}^t \) with the \( (r+1) \)th component 1 and
\[
A_t = \begin{pmatrix}
\alpha_1 \eta_1^2 & \cdots & \alpha_r \eta_r^2 & \beta_1 \eta_1^2 & \cdots & \beta_r \eta_r^2 \\
I_{r-1} & O \\
\alpha_1 & \cdots & \alpha_r & \beta_1 & \cdots & \beta_r \\
O & I_{r-1} & O
\end{pmatrix}.
\]

Proof of Lemma 7.1. By Theorem 2.1 of Ling (2007), Assumption 4.2 implies that, for any \( \epsilon \in (0, 1) \), it follows that
\[
E\left[ \prod_{i=0}^{h_i} A_{i-t-\xi_i} \right] < 1,
\]
16
Thus, (a) holds. Let \( \xi_i \in \mathbb{R} \) for some \( \xi \). By (8.1) and Lemma A.5(ii) of Ling (2007), there exists \( \tilde{\xi} \) is an i.i.d. series. Thus, (a) holds. (b) and (c) are directly from (a). This completes the proof.

**Proof of Lemma 7.2.** By Assumption 4.1, we have the expansion:

\[
\frac{\partial e_i(\theta_0)}{\partial \theta_j} = \sum_{j=0}^{\infty} u_j \xi_{t-j},
\]

where \( u_j = O(p^j) \). Denote \( A_i(k) = \sum_{j=0}^{k-1} u_j \xi_{t-j} \) and \( \tilde{A}_i(k) = \sum_{j=0}^{\infty} u_j \xi_{t-j} \). Then,

\[
E \left[ \frac{\tilde{A}_i(k)}{\sqrt{h_t}} - E \left[ \frac{\tilde{A}_i(k)}{\sqrt{h_t}} \mid F_t(t) \right] \right]^{2-2i} \leq C \left( E \left[ \frac{1}{\sqrt{h_t}} - E \left[ \frac{1}{\sqrt{h_t}} \mid F_t(t) \right] \right] \right)^{2-2i} = O(p^i)
\]

By Holder inequality with \( p = (2-i)/(2-2i) \) and \( q = (2-i)/i \), we have

\[
E \left( A_i(k) \right) \left( \frac{1}{\sqrt{h_t}} - E \left[ \frac{1}{\sqrt{h_t}} \mid F_t(t) \right] \right)^{2-2i} \leq \left( E \left[ A_i(k) \right] \right)^{2-1} \delta \left( E \left[ \frac{1}{\sqrt{h_t}} - E \left[ \frac{1}{\sqrt{h_t}} \mid F_t(t) \right] \right] \right)^{2-2i} = O(p^i)
\]

where the last step holds by Lemma 7.1(a). By (8.2)-(8.3), we can show that

\[
E \left[ \frac{1}{\sqrt{h_t}} \frac{\partial e_i(\theta_0)}{\partial \gamma} - E \left[ \frac{1}{\sqrt{h_t}} \frac{\partial e_i(\theta_0)}{\partial \gamma} \mid F_t(t) \right] \right]^{2-2i} = O(p^i)
\]

for some \( \epsilon > 0 \). By (8.1) and Lemma A.5(ii) of Ling (2007), there exists \( \tilde{\epsilon} \) and \( \epsilon \) such that

\[
\left\| \frac{1}{\sqrt{h_t}} \frac{\partial e_i(\theta_0)}{\partial \gamma} \right\| \leq \frac{\tilde{\epsilon} \epsilon^{-1}}{\sqrt{h_t}} \leq C \epsilon^{1-i}
\]

where \( \xi_x = \sum_{i=0}^{\infty} x^{i} |\xi_{t-i}| \) with \( x \in (0, 1) \). By (8.4)-(8.5) and taking \( \epsilon \) small enough such that \((1 + 3\epsilon)(1 - i) < 1\), we have

\[
E \left[ \frac{1}{\sqrt{h_t}} \frac{\partial e_i(\theta_0)}{\partial \gamma} - E \left[ \frac{1}{\sqrt{h_t}} \frac{\partial e_i(\theta_0)}{\partial \gamma} \mid F_t(t) \right] \right]^{2+2i} \leq \left( E \left[ \frac{1}{\sqrt{h_t}} \frac{\partial e_i(\theta_0)}{\partial \gamma} - E \left[ \frac{1}{\sqrt{h_t}} \frac{\partial e_i(\theta_0)}{\partial \gamma} \mid F_t(t) \right] \right]^{2+2i} \right)^{1/2} \left( E \left[ \frac{1}{\sqrt{h_t}} \frac{\partial e_i(\theta_0)}{\partial \gamma} - E \left[ \frac{1}{\sqrt{h_t}} \frac{\partial e_i(\theta_0)}{\partial \gamma} \mid F_t(t) \right] \right]^{2-2i} \right)^{1/2} \leq E \epsilon_{\tilde{\epsilon} \epsilon^{-1}} O(p^i) = O(p^i).
\]

Thus, (a) holds.
By (2.5) of Ling (2007) and Assumption 4.2, we can show that
\[ h_{i}(\theta_{0}) = \alpha_{0}\beta^{-1}(1) + \sum_{j=1}^{\infty} v_{h_{j}\varepsilon_{j-1}} \text{ and } \frac{\partial h_{i}(\theta_{0})}{\partial \gamma} = 2 \sum_{j=1}^{\infty} \frac{v_{h_{j}\varepsilon_{j-1}}}{\partial \gamma}, \tag{8.7} \]
where \( v_{h_{j}} = O(\rho_{1}^{\frac{1}{2}}) \). Thus, \( \varepsilon_{i-1}/\sqrt{h_{t}} = \varepsilon_{i-1}/\sqrt{h_{t}}^{\frac{1}{2}} \leq 1/\sqrt{h_{t}} \). For any \( t \in (0, 1/2) \), there exist an \( \rho_{1} \in (0, 1) \) such that
\[ E \left[ \sum_{j=1}^{\infty} \frac{v_{h_{j}\varepsilon_{j-1}}}{\partial \gamma} \right]^{2} = O(\rho_{1}). \tag{8.8} \]
Using Lemma 7.1, we can show that there exists a \( t > 0 \) such that
\[ E \left[ \frac{\varepsilon_{i-1}}{\sqrt{h_{t}}} \right]^{2} = O(\rho_{1}). \tag{8.9} \]
By (8.4) and (8.9), and using Hölder’s inequality, we have
\[ E \left[ \frac{\varepsilon_{i-1}}{\sqrt{h_{t}}} \right]^{2} \leq O(1)E \left[ \frac{\varepsilon_{i-1}}{\sqrt{h_{t}}} \right]^{2} = O(\rho_{1}). \tag{8.10} \]
Let \( J = [0.5k \min[1, \ln \rho/\ln(\rho_{1}/\beta_{10})]] \). Note that \( h_{i-1}/h_{i} \leq \beta_{10}^{\frac{1}{2}} \). By (8.9), (8.10) and Hölder’s inequality, we have
\[ E \left[ \frac{\varepsilon_{i-1}}{\sqrt{h_{t}}} \right]^{2} \leq O(1)E \left[ \frac{\varepsilon_{i-1}}{\sqrt{h_{t}}} \right]^{2} + O(\rho_{1}^{\frac{1}{2}}) = O(\rho_{1}). \tag{8.11} \]
where \( u_{0} \geq 1 \) is a constant. By Lemma A.5(ii) of Ling (2007), we have
\[ \left\| \sum_{j=0}^{\infty} \frac{v_{h_{j}\varepsilon_{j-1}}}{\partial \gamma} \right\| \leq O(1) \sum_{j=0}^{\infty} \frac{\varepsilon_{j-1}}{\sqrt{h_{t}}} \frac{\partial \varepsilon_{j-1}(\theta_{0})}{\partial \gamma} \leq C\xi_{\rho_{1}}^{\frac{1}{2}} \leq C\xi_{\rho_{1}}^{\frac{1}{2}}. \]
Taking \( \varepsilon \) small enough such that \( (1 + 3\varepsilon)(1 - \varepsilon) = 1 \) and using Cauchy-Schwarz inequality, we can show that
\[ E \left[ \frac{1}{h_{t}} \frac{\partial h_{t}(\theta_{0})}{\partial \gamma} \right]^{2} = O(\rho_{1}). \tag{8.12} \]
By (8.11)-(8.12), we can show that (b) holds. This completes the proof.
Lemma 8.2. If Assumptions 4.1-4.2 hold, then, for any $0 < \epsilon < 1$, as $\|d\| \to 0$, it follows that

(a) $E[h_{1\epsilon} - h_2\epsilon^2] = o(1)$,

(b) $E[y_{1\epsilon} - y_2\epsilon^2] = o(1)$,

(c) $E[y_{1\epsilon} - y_2\epsilon] = o(1)$.

Proof. Let $A_{\epsilon \tau}$ be defined as $A_{\epsilon}$ in Lemma 8.1 when $\theta = \theta_{1\epsilon} = (\alpha_{0\epsilon}, \alpha_{1\epsilon}, \ldots, \alpha_{\epsilon}, \beta_{0\epsilon}, \beta_{1\epsilon}, \ldots, \beta_{\epsilon})'$, $\nu = 1, 2$. By Theorem 2.1 of Ling (2007), Assumption 4.2 implies that, for any $0 < \epsilon < 1$, there exists an $i_0 > 0$ such that

$$E[\left| \sum_{\epsilon=0}^{i_0-1} A_{\epsilon \epsilon} \right|] < 1/C,$$

where $C > 1$ is any given constant. Let $B_{\epsilon \tau} = \xi + \sum_{i=1}^{i_0-1} A_{\epsilon \epsilon} \xi_{\epsilon \tau-i}$ and $A_{\epsilon \tau} = \sum_{i=0}^{i_0-1} A_{\epsilon \epsilon-i}$, $\nu = 1, 2$. Thus, (c) holds. This completes the proof.

Proof of Lemma 7.3. By Assumption 4.2, we have

$$h_{\epsilon} = \alpha_{0\epsilon} \beta_{0\epsilon} + \sum_{j=1}^{\infty} \nu_{\epsilon j} \varepsilon_{\epsilon j}$$

where $\nu_{\epsilon j} = O(\rho_j^\epsilon)$ and $\nu = 1, 2$. Similar to Lemma 8.2(c), we can show that, for any $1 < \epsilon < 2$,\n
$$E \left[ \left\| \frac{\partial \varepsilon_{1\epsilon} (\theta_{1\epsilon})}{\partial \theta_{1\epsilon}} - \frac{\partial \varepsilon_{2\epsilon} (\theta_{2\epsilon})}{\partial \theta_{2\epsilon}} \right\| \right] = o(1),$$

as $\|d\| \to 0$. By Lemma 8.2, we can show that

$$E \left[ \left( \frac{\nu_{1\epsilon j} E_{1\epsilon-j}}{h_{1\epsilon}} - \frac{\nu_{2\epsilon j} E_{2\epsilon-j}}{h_{2\epsilon}} \right) \right] \leq C \left[ \left| \nu_{1\epsilon j} - \nu_{2\epsilon j} \right| + \nu_{2\epsilon j} E_{1\epsilon-j} + \nu_{2\epsilon j} E_{2\epsilon-j} + \nu_{2\epsilon j} E \left[ \varepsilon_{2\epsilon-j} \left( \frac{1}{h_{1\epsilon}} - \frac{1}{h_{2\epsilon}} \right) \right] \right] = o(1),$$
uniformly $j$. Take $t_1 \in (t, 2)$ and $v$ such that $v(t_1 - t) < t$. By the previous equation, we have

$$E\left[\left(\frac{v_1 h_1}{h_2} - \frac{v_2 h_2}{h_2}\right)\left|\frac{\partial E_2(\theta_2|Y)}{\partial Y}\right|^t\right] \leq E\left[\left(\frac{v_1 h_1}{h_2} - \frac{v_2 h_2}{h_2}\right)\left|\frac{\partial E_2(\theta_2|Y)}{\partial Y}\right|^t\right] = o(1), \quad (8.15)$$

uniformly $j$. By (8.13), $v_{i-j}/\sqrt{v_{i-j}} \leq 1/\sqrt{v_{i-j}}$. Furthermore, by (8.13)-(8.15), we have

$$E\left[\left(\frac{v_1 h_1}{h_2} - \frac{v_2 h_2}{h_2}\right)\left|\frac{\partial E_1(\theta_1|Y)}{\partial Y}\right|^t\right] \leq C \sum_{j=1}^{\infty} \left\{E\left[\left(\frac{v_1 h_1}{h_2} - \frac{v_2 h_2}{h_2}\right)\left|\frac{\partial E_2(\theta_2|Y)}{\partial Y}\right|^t\right] + C \sum_{j=1}^{\infty} \left\{E\left[\left(\frac{v_1 h_1}{h_2} - \frac{v_2 h_2}{h_2}\right)\left|\frac{\partial E_2(\theta_2|Y)}{\partial Y}\right|^t\right]\right\}$$

$$= o(1) + C \sum_{j=1}^{\infty} \left\{E\left[\left(\frac{v_1 h_1}{h_2} - \frac{v_2 h_2}{h_2}\right)\left|\frac{\partial E_2(\theta_2|Y)}{\partial Y}\right|^t\right]\right\}$$

$$= o(1). \quad (8.16)$$

By (8.13) and Lemma 5.4(ii) of Ling (2007), there exists an $i \in (0, 1)$ such that

$$\left\|\frac{1}{h_i} \frac{\partial h_i(\theta_0)}{\partial Y}\right\| = \frac{C}{\sqrt{v_h}} \sum_{j=1}^{\infty} \left\{E\left[\left(\frac{v_1 h_1}{h_2} - \frac{v_2 h_2}{h_2}\right)\left|\frac{\partial E_2(\theta_2|Y)}{\partial Y}\right|^t\right]\right\} \leq \xi^{1+i}_j$$

$$\xi_{j+i} = \sum_{j=1}^{\infty} \left\{E\left[\left(\frac{v_1 h_1}{h_2} - \frac{v_2 h_2}{h_2}\right)\left|\frac{\partial E_2(\theta_2|Y)}{\partial Y}\right|^t\right]\right\} = o(1), \quad (8.17)$$

where $E\xi_{j+i}^{(1+i)(1-i)}(d) \to E\xi_{j+i}^{2(1+i)(1-i)} < \infty$. Thus, by (8.16)-(8.17), we have

$$E\left[\left(\frac{v_1 h_1}{h_2} - \frac{v_2 h_2}{h_2}\right)\left|\frac{\partial E_1(\theta_1|Y)}{\partial Y}\right|^t\right] \leq E\left(\xi_{j+i}^{1+i} + \xi_{j+i}^{2(1+i)}\right)^{1+i} \leq E\left(\xi_{j+i}^{1+i} + \xi_{j+i}^{2(1+i)}\right)^{2(1+i)} = o(1).$$

This completes the proof. □

References