Econometric Theory, 2014, Page 1 of 29.
doi:10.1017/S0266466614000863

ESTIMATION OF CHANGE-POINTS IN LINEAR AND NONLINEAR TIME SERIES MODELS

SHIQING LING
Hong Kong University of Science and Technology

This paper develops an asymptotic theory for estimated change-points in linear and nonlinear time series models. Based on a measurable 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, *Annals of Statistics* 15, 1321–1328). This provides a channel to connect our results with those in Picard (1985, *Advances in Applied Probability* 17, 841–867) and Bai, Lumsdaine, and Stock (1998, *Review of Economic Studies* 65, 395–432), where the magnitude of changed parameters depends on the sample size n and tends to zero as $n \to \infty$. The theory is applied for the self-weighted QMLE and the local QMLE of change-points in ARMA-GARCH/IGARCH models. A simulation study is carried out to evaluate the performance of these estimators in the finite sample.

1. INTRODUCTION

Structural change has been an important issue in econometrics, engineering, and statistics for a long time. As a recent comment by Hendry and Johansen (2014), the breaks and regime shifts are ubiquitous in economic time series and were widely recognized even by the time of Haavelmo (1944). More real examples are in Stock and Watson (1996) and Hansen (2001). The earliest test statistics go back to Chow (1960) and Quandt (1960). After that, 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 (1992), Andrews (1993), and Bai and Perron (1998); the exact likelihood ratio test in Horv*á*th (1993) and Davis, Huang, and Yao (1995); the empirical approach in Bai (1996) for regression models; and the sequential test in Lai (1995). Su and White (2010) proposed two tests for change-points in partially linear models. Breitung

I would like to thank two reviewers, Dr Gao Zhaoxing, the co-editor Pentti Saikkonen, and the editor Peter C.B. Phillips for their extremely helpful comments and Wang Kai for his simulation study. This project was supported by the Hong Kong Research Grants Council (Grants HKUST641912, 603413 and 19 FSGRF12SC12). Address correspondence to Shiqing Ling, Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong; e-mail : maling@ust.hk.

and Eickmeier (2011) and Han and Inoue (2014) studied some tests for structural breaks in dynamic factor models. Ling (2007a) developed an asymptotic theory of the Quandt-type tests for linear and nonlinear time series models. Aue, Hörmann, Horváth, and Reimherr (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 tests was further developed by Hidalgo and Seo (2013) under a larger 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 nonbinomial 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. Ritov (1990) studied the asymptotic efficient estimation of the change-point. Dümbgen (1991) investigated the nonparametric method for change-point estimators. 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. Hansen (2009) proposed an averaging estimator for regressions with a possible structural break. Perron and Yamamoto (2014) studied estimating and testing for multiple structural changes in models with endogenous regressors.

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_n which depends on the sample size n with $d_n \to 0$ 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 et al. (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, 2003 for ARMA-GARCH models). Davis, Lee, and Rodriguez-Yam (2006) proposed a minimum description length principle to locate the change points in the multiple structural change AR models. Saikkonen, Lükepohl, and Trenkler (2006) and Kejriwal and Perron (2008) used a similar method to estimate the change-point in VAR models and cointegrated regression models, respectively. 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 develops an asymptotic theory for estimating change-points in linear and nonlinear time series models. Based on a measurable 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 et al. (1998). The theory is applied for the self-weighted QMLE and the local QMLE of change-points in ARMA-GARCH/IGARCH models. A simulation study is carried out to evaluate the performance of these estimators in the finite sample.

This paper proceeds as follows. Section 2 presents our main results. Section 3 gives the approximating distribution of the estimated change-points. Section 4 presents the results for the structure-change ARMA-GARCH/IGARCH models. Section 5 reports simulation results. Sections 6 and 7 give the proofs of results in Sections 2 and 4, respectively. Section 8 gives a concluding remark. The consistency of the estimated change-point and its proof are given in Appendix.

2. MAIN RESULTS

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

$$y_t = g(\vartheta, Y_{t-1}, \eta_t), \tag{2.1}$$

where $\mathcal{F}_t$ is the σ-field generated by $\{\eta_t, \eta_{t-1}, \ldots\}$, $Y_t = (y_t, \ldots, y_{t-p+1})$, or $Y_t = (y_t, y_{t-1}, \ldots)$, ϑ is an $\tilde{m} \times 1$ unknown parameter vector, and $\{\eta_t\}$ is independently and identically distributed (i.i.d.). The structure of the time series $\{y_t\}$ is characterized by g and the parameter ϑ. We assume that the parameter space Θ is a bounded compact subset of R^m. We denote model (2.1) by $M(\vartheta_0)$ when the true parameter is $\vartheta = \vartheta_0$.

Let $\{y_1, \ldots, y_n\}$ be the random sample. We assume

$$\{y_1, \ldots, y_{k_0}\} \in M(\theta_{10}) \ \ \textit{and} \ \ \{y_{k_0+1}, \ldots, y_n\} \in M(\theta_{20}) \ \ \textit{with} \ \ \theta_{10} \neq \theta_{20},$$

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 θ_{10} and θ_{20}, based on the presample and the postsample, respectively:

$$L_{1n}(k, \theta_1) = \sum_{t=1}^{k} l_t(\theta_1) \ \ \textit{and} \ \ L_{2n}(k, \theta_2) = \sum_{t=k+1}^{n} l_t(\theta_2), \tag{2.2}$$

where $l_t(\vartheta) = l(\vartheta, y_t, y_{t-1}, \ldots)$ is a measurable function in terms of $\{y_t\}$ and is almost surely (a.s.) continuous with respect to ϑ. 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 $l_t(\vartheta)$ as that in LSE, MLE, quasi-MLE, LAD-type, or M-estimators, among others. Assume θ_{10} and θ_{20} are interior points in Θ. When $t > k_0$, $l_t(\vartheta) = l(\vartheta, y_t, \ldots, y_{k_0+1}, Y_{k_0})$ and when $t \le k_0$, $l_t(\vartheta) = l(\vartheta, y_t, \ldots, y_1, Y_0)$. We assume

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

That is, there are two processes $\{y_{1t}\} \in M(\theta_{10})$ and $\{y_{2t}\} \in M(\theta_{20})$ and we observe $y_t = y_{2t}$ when $t > k_0$ and $y_t = y_{1t}$ when $t \le k_0$. This assumption keeps the stationarity and ergodicity of y_t when $t > k_0$ and requires its initial values from $M(\theta_{20})$. Thus, the objective function (2.2) always involves these initial values and we need to replace Y_0 by some chosen constants in practice. Their effect needs to be addressed case by case. We will discuss them in Section 4 for ARMA-GARCH models.

Let $\hat{\theta}_{1n}(k)$ and $\hat{\theta}_{2n}(k)$ be the maximizers of $L_{1n}(k, \theta_1)$ and $L_{2n}(k, \theta_2)$ on Θ for each given k. k_0 is estimated by

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

In practice, $1 \le k \le n$ can be replaced by $\tilde{p} \le k \le n - \tilde{p}$ for some integer $\tilde{p}$. Other parameters are estimated by

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

In this procedure, one needs to run two sequential estimates 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 example, Bai, 1995 and Bai et al., 1998). We now introduce two assumptions as follows.

Assumption 2.1. When $\{y_s : s \le t\} \in M(\vartheta_0)$, $E \sup_{\vartheta \in \Theta} |l_t(\vartheta)| < \infty$, and $E[l_t(\vartheta)]$ has a unique maximizer at $\vartheta = \vartheta_0$.

Assumption 2.2. When $\{y_t : t = 0, \pm 1, \pm 2, \ldots\} \in M(\vartheta_0)$,

$$\frac{1}{u} \sup_{\Theta} \left| \sum_{t=-u}^{-1} [l_t(\vartheta) - E l_t(\vartheta)] \right| = o(1), \ a.s., \text{ as } u \to \infty.$$

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 $L^p(\nu)$ NED in terms of $\{\eta_t\}$ if $\sup_{-\infty < t < \infty} \|X_t\|_p < \infty$ and

$$\sup_{-\infty < t < \infty} \|X_t - E[X_t | \mathcal{F}_k(t)]\|_p = O(k^{-\nu}),$$

where $\|A\| = [tr(AA')]^{1/2}$ for a vector or matrix A, $\mathcal{F}_i(j)$ is the σ-field generated by $\{\eta_j, \eta_{j-1}, \ldots, \eta_{j-i+1}\}$ with $i \ge 1$, and $\mathcal{F}_0(j) = \{\emptyset, \Omega\}$, $p \ge 1$ and $\nu > 0$.

This holds for many time series models. Theorem 2.1 of Ling (2007a) can be used to verify Assumption 2.2 if $l_t(\vartheta)$ is $L^p(\nu)$ NED with $p > 1$ and $\nu > 0$ (see the proof of Theorem 4.1 in Section 7).

THEOREM 2.1. *If Assumptions 2.1 and 2.2 hold, then*

(a) $\hat{\theta}_{in} = \theta_{i0} + o_p(1)$, $i = 1, 2$;

(b) $\hat{k}_n = k_0 + O_p(1)$.

We can write $\hat{k}_n = [n\hat{\tau}_n]$. Then $\hat{\tau}_n$ is an estimator of τ_0. This theorem implies that the rate of convergence of $\hat{\tau}_n$ is n which is faster than that in Picard (1985) and Bai et al. (1998) for AR models.

Assumption 2.3. When $\{y_t : t = 1, \ldots, n\} \in M(\vartheta_0)$, the following statements hold:

(i) for any $\vartheta_n \to_p \vartheta_0$,

$$\sum_{t=1}^{n}[l_t(\vartheta_n) - l_t(\vartheta_0)] = (\vartheta_n - \vartheta_0)' \sum_{t=1}^{n} D_t(\vartheta_0) - n(\vartheta_n - \vartheta_0)' \times \left[\frac{1}{2}\Sigma_{\vartheta_0} + o_p(1)\right](\vartheta_n - \vartheta_0),$$

(ii) $D_t(\vartheta_0)$ is a martingale difference in terms of $\mathcal{F}_t$ with the covariance Ω_{ϑ_0},

(iii) Ω_{ϑ_0} and Σ_{ϑ_0} are positive definite matrices.

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

$$W(k, \theta_{10}, \theta_{20}) = \begin{cases} \sum_{t=1}^{k}[l_t(\theta_{10}) - l_t(\theta_{20})], & k > 0, \\ 0, & k = 0, \\ \sum_{t=k}^{-1}[l_t(\theta_{20}) - l_t(\theta_{10})], & k < 0, \end{cases}$$

where $y_t \in M(\theta_{20})$ when $k > 0$ and $y_t \in M(\theta_{10})$ 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 $\vartheta_0 = \theta_{10}$ and θ_{20}, respectively, then $\hat{k}_n$, $\hat{\theta}_{1n}$, and $\hat{\theta}_{2n}$ are asymptotically independent and, when $n \to \infty$, it follows that*

(a) $\sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) \longrightarrow_{\mathcal{L}} N(0, \frac{1}{\tau_0}\Sigma_{\theta_{10}}^{-1}\Omega_{\theta_{10}}\Sigma_{\theta_{10}}^{-1})$

and $\sqrt{n}(\hat{\theta}_{2n} - \theta_{20}) \longrightarrow_{\mathcal{L}} N(0, \frac{1}{1-\tau_0}\Sigma_{\theta_{20}}^{-1}\Omega_{\theta_{20}}\Sigma_{\theta_{20}}^{-1})$,

(b) $\hat{k}_n - k_0 \longrightarrow_{\mathcal{L}} argmax_k W(k, \theta_{10}, \theta_{20})$,

where $\longrightarrow_{\mathcal{L}}$ denotes convergence in distribution.

Unlike the i.i.d. case in Hinkley (1970) and Bai (1995), the double-sided random walk $W(k, \theta_{10}, \theta_{20})$ is neither independent nor symmetric.

3. APPROXIMATING DISTRIBUTION OF ESTIMATED K_0

The distribution of $\operatorname{argmax}_k W(k, \theta_{10}, \theta_{20})$ does not have a closed form and is therefore difficult to be used directly for statistical inference. Denote

$$d = \theta_{10} - \theta_{20} \ \ and \ \ W_d(k) = W(k, \theta_{10}, \theta_{20}).$$

This section investigates the limiting distribution of $\operatorname{argmax}_k W_d(k)$ when $\|d\| \to 0$. Note that $y_t \in M(\theta_{10})$ is a function of θ_{10} and $\{\eta_t\}$ and similarly for $y_t \in M(\theta_{20})$. Thus, y_t changes when the value of d is changed. To make it simple, we fix θ_{20} and assume that $d = \theta_{10} - \theta_{20} \to 0$. In this case, when $k > 0$, $y_t \in M(\theta_{20})$ and is not changed when $d \to 0$. But when $k < 0$, $y_t \in M(\theta_{10})$ and is changed when $d \to 0$. To make it clear, when $y_t \in M(\theta_{i0})$, y_t is denoted by y_{it}, $l_t(\theta)$ by $l_{it}(\theta)$, and $D_t(\theta)$ by $D_{it}(\theta)$, $i = 1, 2$, etc. We make the following assumptions.

Assumption 3.1. Let $m = [(d'\Sigma_{20}d)^{-2}(d'\Omega_{20}d)]$. For each $z \in R$, we have

$$\sum_{t=-[mz]}^{-1} [l_{1t}(\theta_{20}) - l_{1t}(\theta_{10})] = -\sum_{t=-[mz]}^{-1} d' D_{2t}(\theta_{20}) - \frac{[mz]}{2} d' \left[\Sigma_{\theta_{20}} + o_p(1)\right] d,$$

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

In this assumption, $y_{1t} = g(\theta_{10}, y_{1t-1}, y_{1t-2}, \ldots, \eta_t)$ is a composite function of θ_{10} and $\{\eta_t\}$ and it is changing when θ_{10} changes, and so is $l_{1t}(\theta_{10})$. To check it, one needs to explore a function of θ_{10} and $\{\eta_t\}$ (see Lemma 7.3 in Section 7). It is usually more complicated than that for Assumption 2.3(i) in which one only needs to study a function of θ and $\{y_t\}$ since $\{y_t\}$ is generated by the same ϑ_0. This issue also appears in the change-point problem with assuming $d = d_n$, changing over sample size n.

Assumption 3.2. $D_{2t}(\theta_{20})$ is $L^{2+\iota}(\nu)$ NED in terms of $\{\eta_t\}$ with $\iota > 0$, where either $2\nu > 1$ or $2\nu = 1$ and there exist constants $\nu_1 > 0$ and $\iota_1 > 0$ with $2\nu_1 > 1$ such that

$$\sup_{-\infty < t < \infty} \|E[D_{2t}|\mathcal{F}_{k+1}(t)] - E[D_{2t}|\mathcal{F}_k(t)]\|_{2+\iota_1} = O(k^{-\nu_1}). \tag{3.1}$$

This assumption is to use the invariance principle in Ling (2007a) for the backward sum $\sum_{t=-k}^{-1} D_{2t}\ (\theta_{10})$. The usual invariance principle for the forward sum cannot be applied in this case. Our approximating distribution is as follows.

THEOREM 3.1. *Suppose that Assumptions 2.3(i) with* $\vartheta_n = \vartheta_0 + O_p(1/\sqrt{n})$, *Assumption 2.3(ii) and (iii), and Assumptions 3.1–3.2 hold. Then, for any fixed* M, *we have*

$$(d'\Sigma_{20}d)^2(d'\Omega_{20}d)^{-1} argmax_{z\in[-M,M]} W_d([mz]) \longrightarrow_{\mathcal{L}} argmax_{z\in[-M,M]}\Big[B(z) - \frac{1}{2}|z|\Big],$$

as $0 < \|d\| \to 0$, *where* $B(z)$ *is the standard Brownian motion in* R.

Proof. Let $\gamma_d = (d'\Sigma_{\theta_{20}}d)(d'\Omega_{\theta_{20}}d)^{-1}$. Then $m\gamma_d^2 d'\Omega_{20}d \to 1$ as $d \to 0$. By Assumptions 2.3(ii) and (iii) and 3.2, and Theorem 2.2 of Ling (2007a), we can show that

$$\gamma_d W_m^+(z) \equiv (\sqrt{m}\gamma_d d')\frac{1}{\sqrt{m}}\sum_{t=1}^{[mz]} D_{2t}(\theta_{20}) \longrightarrow_{\mathcal{L}} B(z), \tag{3.2}$$

$$\gamma_d W_m^-(z) \equiv (\sqrt{m}\gamma_d d')\frac{1}{\sqrt{m}}\sum_{t=-[mz]}^{-1} D_{2t}(\theta_{20}) \longrightarrow_{\mathcal{L}} B(z), \tag{3.3}$$

on $D[-M, M]$ for any given M, as $m \to \infty$, where $D[-M, M]$ denotes the space of functions on [−M, M] which are right continuous and have left-hand limits, equipped with the Skorokhod topology as in Billingsley (1968). By Assumptions 2.3(i) and 3.1, we can show that $\gamma_d W_d([mz])$ has the uniform expansion on $z \in [-M, M]$,

$$\begin{aligned}\gamma_d W_d([mz]) &= -\frac{1}{2}|z| + \gamma_d W_m^+(z) I\{z > 0\} + \gamma_d W_m^-(z) I\{z \le 0\} + o_p(1)\\ &\longrightarrow_{\mathcal{L}} -\frac{1}{2}|z| + B(z) \text{ on } D[-M, M],\end{aligned}$$

as $0 < \|d\| \to 0$, where the last step holds by (3.2) and (3.3). The random element $\text{argmax}_{z\in[-M,M]} W_d([mz])$ has the same distribution as $\text{argmax}_{z\in[-M,M]}\{\gamma_d W_d([mz])\}$. By the previous equations and using the continuous mapping theorem for the argmax function, we can claim that the conclusion holds. ■

Let $F_d(x)$ be the distribution of $\text{argmax}_k W_d(k)$. Then, $F_d(x)$ and the distribution of $\text{argmax}_z\{\gamma_d W_d([z])\}$ are identical. In practice, d is fixed. Thus, there is a M such that

$$\big|F_d(x) - P(\text{argmax}_{z\in[-s,s]}\{\gamma_d W_d([z])\} \le x)\big| \le \varepsilon,$$

when $s > M$. Thus,

$$\begin{aligned}&\big|F_d(x) - P(m\,\text{argmax}_{z\in[-M,M]}\{\gamma_d W_d([mz])\}x)\big|\\ &\quad = \big|F_d(x) - P(\text{argmax}_{z\in[-mM,mM]}\{\gamma_d W_d([z])\} \le x)\big| \le \varepsilon.\end{aligned}$$

Since the probability of $-|z|/2+B(z)$ when $z \notin [-M, M]$ is small as M is large, we can see that

$$F_d(x) \approx P(\text{argmax}_{\gamma \in R}[B(\gamma) - |\gamma|/2] \leq x),$$

when d is small. Yao (1987) showed that the distribution $F(x)$ of $\text{argmax}_{\gamma \in R}$ $[B(\gamma) - |\gamma|/2]$ has the density function:

$$f(x) = \frac{3}{2}e^{|x|}\Phi\left(\frac{3}{2}\sqrt{|x|}\right) - \frac{1}{2}\Phi\left(\frac{\sqrt{|x|}}{2}\right), \quad \textit{where } \Phi(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du,$$

and $x \in R \equiv (-\infty, \infty)$. By Theorem 3.1, it is reasonable to approximate the distribution of $(d'\Sigma_{20}d)^2(d'\Omega_{20}d)^{-1}(\hat{k}_n - k_0)$ 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 $\theta_{10} = \theta_n$ and $\theta_{20} = \theta_{1n}$, $d = d_n = \theta_n - \theta_{1n} \to 0$ and $\|d_n\|\sqrt{n} \to \infty$ as $n \to \infty$. They estimate (θ_n, θ_{1n}) and show that the limiting distribution of the normalized $\hat{k}_n$ is $F(x)$. Their true parameter (θ_n, θ_{1n}) is changed with n, while the true parameter $(\theta_{10}, \theta_{20})$ in our model in Theorems 2.1 and 2.2 is fixed and hence d is fixed. Theorem 3.1 is only to give a reasonable approximating distribution to the limiting distribution of $\hat{k}_n$ in Theorem 3.2 when d is small. The confidence intervals of k_0 based on the two frameworks are identical when d or d_n is small since we use the same approximating distribution.

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:

$$\phi(B)y_t = \psi(B)\varepsilon_t, \tag{4.1}$$

$$\varepsilon_t = \eta_t\sqrt{h_t} \quad \textit{and} \quad h_t = \alpha_0 + \sum_{i=1}^{r}\alpha_i\varepsilon_{t-i}^2 + \sum_{i=1}^{s}\beta_i h_{t-i}, \tag{4.2}$$

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, $\alpha_0 > 0$, $\alpha_i \geq 0$ $(i = 1, \ldots, r)$, $\beta_j \geq 0$ $(j = 1, \ldots, s)$, and $\{\eta_t\}$ are a sequence of i.i.d. random variables with zero mean and variance 1. Denote $\gamma = (\phi_1, \ldots, \phi_p, \psi_1, \ldots, \psi_q)'$, $\delta = (\alpha_0, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s)'$, and $\vartheta = (\gamma', \delta')'$. The parameter space is $\Theta = \Theta_\gamma \times \Theta_\delta$, where $\Theta_\gamma \subset R^{p+q}$ and $\Theta_\delta \subset R_0^{r+s+1}$ are compact, where $R_0 = [0, \infty)$. We denote models (4.1) and (4.2) by $M(\vartheta_0)$ when the true value of ϑ is ϑ_0. We introduce the following conditions:

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

Assumption 4.2. $\alpha(z) \equiv \sum_{i=1}^{r} \alpha_i z^i$ and $\beta(z) \equiv 1 - \sum_{i=1}^{s} \beta_i z^i$ have no common root, $\alpha(1) \neq 0$, $\alpha_r + \beta_s \neq 0$, and $\sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j \leq 1$ for each $\vartheta \in \Theta$.

Assumption 4.3. η_t^2 has a nondegenerate distribution with $E\eta_t^2 = 1$.

Models (4.1) and (4.2) have a finite second moment when $\sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j < 1$. When $\sum_{i=1}^{r} \alpha_i + \sum_{j=1}^{s} \beta_j = 1$, model (4.2) is called the IGARCH model and in this case, $Ey_t^2 = \infty$ and $E|y_t|^{2\iota} < \infty$ for any $\iota \in (0, 1)$. We assume that

$$\{y_1, \ldots, y_{k_0}\} \in M(\theta_{10}) \text{ and } \{y_{k_0+1}, \ldots, y_n\} \in M(\theta_{20}), \tag{4.3}$$

and θ_{10} and θ_{20} are interior points in Θ. We first consider the self-weighted quasi-maximum likelihood estimator (SQMLE) of parameters $(k_0, \theta_{10}, \theta_{20})$. In this case,

$$L_{1n}(k, \theta_1) = \sum_{t=1}^{k} w_t l_t(\theta_1) \quad \textit{and} \quad L_{2n}(k, \theta_2) = \sum_{t=k+1}^{n} w_t l_t(\theta_2), \tag{4.4}$$

where

$$l_t(\vartheta) = -\frac{1}{2}\left[\log h_t(\vartheta) + \frac{\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}\right], \tag{4.5}$$

$\varepsilon_t(\vartheta) = y_t - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \psi_i \varepsilon_{t-i}(\vartheta)$, $h_t(\vartheta) = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2(\vartheta) + \sum_{i=1}^{s} \beta_i h_{t-i}(\vartheta)$, and

$$w_t = \left[1 + \sum_{i=1}^{\infty} i^{-2} |y_{t-i}|\right]^{-3}.$$

This particular weight w_t is just for simplicity. We refer to Ling (2007b) for other choices.

We assume that the initial condition (2.3) is satisfied. Denote $U_t(\vartheta) = \left([\partial\varepsilon_t(\vartheta)/\partial\vartheta]/\sqrt{h_t(\vartheta)}, \partial h_t(\vartheta)/\partial\vartheta/[\sqrt{2}h_t(\vartheta)]\right)$ and $\xi_t(\vartheta) = [\varepsilon_t(\vartheta)/\sqrt{h_t(\vartheta)}, (1 - \varepsilon_t^2(\vartheta)/h_t(\vartheta))/\sqrt{2}]'$. Then

$$D_t(\vartheta) = \partial l_t(\vartheta)/\partial\vartheta = U_t(\vartheta)\xi_t(\vartheta), \tag{4.6}$$

$$\begin{aligned} P_t(\vartheta) &= -\frac{\partial^2 l_t(\vartheta)}{\partial\vartheta\partial\vartheta'} \\ &= U_t(\vartheta)U_t(\vartheta)' + \left[\frac{\varepsilon_t^2(\vartheta)}{h_t(\vartheta)} - 1\right]R_{1t}(\vartheta) + \frac{\varepsilon_t(\vartheta)}{\sqrt{h_t(\vartheta)}}R_{2t}(\vartheta), \end{aligned} \tag{4.7}$$

where $R_{1t}(\vartheta) = [\partial h_t(\vartheta)/\partial\vartheta][\partial h_t(\vartheta)/\partial\vartheta']/2h_t^2(\vartheta) - [\partial^2 h_t(\vartheta)/\partial\vartheta\partial\vartheta']/2h_t(\vartheta)$ and $R_{2t}(\vartheta) = 2[\partial\varepsilon_t(\vartheta)/\partial\vartheta]\ [\partial h_t(\vartheta)/\partial\vartheta']/\sqrt{h_t(\vartheta)} + [\partial^2\varepsilon_t(\vartheta)/\partial\vartheta\partial\vartheta']/x$

$\sqrt{h_t(\vartheta)}$. When $\{y_t\} \in M(\vartheta_0)$, $\varepsilon_t(\vartheta_0) = \varepsilon_t$ and $h_t(\vartheta_0) = h_t$, and $\xi_t(\vartheta_0) = [\eta_t, (1-\eta_t^2)/\sqrt{2}]'$. Let $J = E[\xi_t(\vartheta_0)\xi_t'(\vartheta_0)]$. We have the following result:

THEOREM 4.1. *Suppose that Assumptions 4.1–4.3 hold, $E\eta_t^4 < \infty$ and $J > 0$. Then, $\hat{k}_n$, $\hat{\theta}_{1n}$, and $\hat{\theta}_{2n}$ are asymptotically independent and, when $n \to \infty$, it follows that*

(a) $\sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) \longrightarrow_{\mathcal{L}} N(0, \frac{1}{\tau_0}\Sigma_{s1}^{-1}\Omega_{s1}\Sigma_{s1}^{-1})$,

$$\sqrt{n}(\hat{\theta}_{2n} - \theta_{20}) \longrightarrow_{\mathcal{L}} N(0, \frac{1}{1-\tau_0}\Sigma_{s2}^{-1}\Omega_{s2}\Sigma_{s2}^{-1}),$$

(b) $\hat{k}_n - k_0 \longrightarrow_{\mathcal{L}} \arg\max_k W_s(k, \theta_{10}, \theta_{20})$,

where $\Sigma_{si} = E[w_t U_t(\theta_{i0})U_t'(\theta_{i0})]$ and $\Omega_{si} = E[w_t^2 U_t(\theta_{i0}) J U_t'(\theta_{i0})]$ as $i = 1, 2$ and $W_s(k, \theta_{10}, \theta_{20})$ is defined as $W(k, \theta_{10}, \theta_{20})$ in Theorem 2.2 with $l_t(\vartheta)$ replaced by $w_t l_t(\vartheta)$.

The SQMLE of $(\theta_{10}, \theta_{20})$ may not be as efficient as its QMLE (see a discussion in Ling, 2007b). This may affect the estimator of k_0. When $Ey_t^4 < \infty$, we can take $w_t = 1$ such that the SQMLE reduces to the QMLE. We refer to Francq and Zakoan (2004) for the QMLE of models (4.1) and (4.2) when $Ey_t^4 < \infty$. However, we cannot show that Theorem 4.1 holds when $Ey_t^4 = \infty$ with $w_t = 1$. We now consider the local QMLE without a weighted function w_t. Specifically, using $\hat{\theta}_{in}$ in Theorem 4.1 as an initial estimator of θ_{i0}, $i = 1, 2$, the local QMLE is obtained via the following one-step iteration:

$$\tilde{\theta}_{1n} = \hat{\theta}_{1n} - \left[\sum_{t=1}^{\hat{k}_n} P_t(\hat{\theta}_{1n})\right]^{-1} \sum_{t=1}^{\hat{k}_n} D_t(\hat{\theta}_{1n}), \tag{4.8}$$

$$\tilde{\theta}_{2n} = \hat{\theta}_{2n} - \left[\sum_{t=\hat{k}_n+1}^{n} P_t(\hat{\theta}_{2n})\right]^{-1} \sum_{t=\hat{k}_n+1}^{n} D_t(\hat{\theta}_{2n}), \tag{4.9}$$

$$\tilde{k}_n = \operatorname{argmax}_{k_L \le k \le n-k_L} \left[\sum_{t=1}^{k} l_t(\tilde{\theta}_{1n}) + \sum_{t=k+1}^{n} l_t(\tilde{\theta}_{2n})\right]. \tag{4.10}$$

For this local QMLE, we have the following result:

THEOREM 4.2. *Suppose that Assumptions 4.1–4.3 hold, $E\eta_t^4 < \infty$ and $J > 0$. If $(\tilde{\theta}_{1n}, \tilde{\theta}_{2n}, \tilde{k}_n)$ is obtained through (4.8)–(4.10), then, when $n \to \infty$, it follows that*

(a) $\sqrt{n}(\tilde{\theta}_{1n} - \theta_{10}) \longrightarrow_{\mathcal{L}} N(0, \frac{1}{\tau_0}\Sigma_1^{-1}\Omega_1\Sigma_1^{-1})$,

$$\sqrt{n}(\tilde{\theta}_{2n} - \theta_{20}) \longrightarrow_{\mathcal{L}} N(0, \frac{1}{1-\tau_0}\Sigma_2^{-1}\Omega_2\Sigma_2^{-1}),$$

(b) $\tilde{k}_n - k_0 \longrightarrow_{\mathcal{L}} arg\max_k W(k, \theta_{10}, \theta_{20})$,

where $\Sigma_i = E[U_t(\theta_{i0})U_t'(\theta_{i0})]$ *and* $\Omega_i = E[U_t(\theta_{i0})JU_t'(\theta_{i0})]$, $i = 1, 2$.

The approximating distribution in Theorem 3.1 can be used for both $\hat{k}_n$ and $\tilde{k}_n$. We only state one for $\tilde{k}_n$ here.

THEOREM 4.3. *If the assumptions of Theorem 4.2 hold, then for any fixed M, we have*

$$(d'\Sigma_2 d)^2 (d'\Omega_2 d)^{-1} argmax_{z\in[-M,M]} W([mz], \theta_{10}, \theta_{20})$$
$$\longrightarrow_{\mathcal{L}} argmax_{z\in[-M,M]}\Big[B(z) - \frac{1}{2}|z|\Big],$$

as $0 < \|d\| \to 0$, *where* $d = \theta_{10} - \theta_{20}$ *and* $m = [(d'\Sigma_2 d)^{-2}(d'\Omega_2 d)]$.

For models (4.1) and (4.2), the initial condition (2.3) is not satisfied in practice. Since we have only one data set $\{y_n, \ldots, y_1\}$, we use this and replace Y_0 by some constant $\tilde{Y}_0$ to calculate $l_t(\theta)$. Although we do not know k_0, this calculation has implied that we replace Y_{k_0} by $\tilde{Y}_{k_0} = \{y_{k_0}, \ldots, y_1, \tilde{Y}_0\}$ when $t > k_0$. With these initial values, the expansion in Assumption 2.3(i) still holds and hence they do not affect the asymptotic results of $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ (see Zhu, 2010 for models (4.1) and (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

$$\tilde{l}_t(\theta_1) = l(\theta_1, y_t, \ldots, y_1, \tilde{Y}_0) \text{ when } t \le k_0$$
$$\text{and } \tilde{l}_t(\theta_2) = l(\theta_2, y_t, \ldots, y_{k_0+1}.\tilde{Y}_{k_0}) \text{ when } t > k_0.$$

From the proof of Theorem 4.2, we can see that

$$\tilde{k}_n - k_0 = argmin_k \tilde{W}_t(k, \theta_{10}, \theta_{20}) + o_p(1),$$

where

$$\tilde{W}_t(k, \theta_{10}, \theta_{20}) = \begin{cases} \sum_{t=k_0+1}^{k}[\tilde{l}_t(\theta_{20}) - \tilde{l}_t(\theta_{10})], & k > k_0, \\ 0, & k = k_0, \\ \sum_{t=k+1}^{k_0}[\tilde{l}_t(\theta_{10}) - \tilde{l}_t(\theta_{20})], & k < k_0. \end{cases}$$

Since the distributions of $W(k, \theta_{10}, \theta_{20})$ and $\tilde{W}_t(k, \theta_{10}, \theta_{20})$ are different, the initial values always affect the asymptotic distribution of the estimated k_0. However, by Taylor's expansion, we have

$$\sum_{t=k+1}^{k_0} \{[l_t(\theta_{10}) - \tilde{l}_t(\theta_{10})] - [l_t(\theta_{20}) - \tilde{l}_t(\theta_{20})]\}$$
$$= d \sum_{t=k+1}^{k_0} [D_t(\xi_1^*) - \tilde{D}_t(\xi_1^*)] = d\, O_p(\rho^{k_0-k}) = o(1),$$

when $d \to 0$, where ξ_1^* is between θ_{10} and θ_{20}, $\rho \in (0, 1)$ and the second equation is from Zhu (2010). Similarly, we have

$$\sum_{t=k_0+1}^{k} \{[l_t(\theta_{10}) - \tilde{l}_t(\theta_{10})] - [l_t(\theta_{20}) - \tilde{l}_t(\theta_{20})]\} = o(1),$$

when $d \to 0$. Thus, we can see that the approximation distribution in Theorem 4.3 is still valid in this case.

Models (4.1) and (4.2) include the ARMA model (i.e., the case with $h_t = \alpha_0$) and the GARCH model (i.e., the case with $y_t = \varepsilon_t$) 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 a finite variance. 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. SIMULATION STUDY

This section examines the performance of our asymptotic results in the finite samples via some Monte Carlo experiments. The data are generated by the following AR(1)-GARCH(1,1) model:

$$\begin{aligned} &y_t = \phi_1 y_{t-1} I_{\{t \le k_0\}} + \phi_2 y_{t-1} I_{\{t > k_0\}} + \varepsilon_t, \\ &\varepsilon_t = \eta_t \sqrt{h_t}, \\ &\text{and } h_t = (\alpha_{10} + \alpha_{11}\varepsilon_{t-1}^2 + \beta_{11} h_{t-1}) I_{\{t \le k_0\}} + (\alpha_{20} + \alpha_{21}\varepsilon_{t-1}^2 + \beta_{22} h_{t-1}) I_{\{t > k_0\}}, \end{aligned}$$

where $\eta_t \sim$ i.i.d. $N(0, 1)$. The true parameters are $\theta_{10} = (0.6, 0.1, 0.1, 0.45)'$ and $\theta_{20} = \theta_{10} + d(1, 1, 1, 1)'$ with $d = 0.05$, 0.1, and 0.2, respectively. We use 4,000 replications in all the experiments. The simulations are carried out by MATLAB and the optimization algorithm in the package Fmincon. Under the two sets of parameters, the model has a finite fourth moment. We can take the weight $w_t = 1$ in which case self-weighted MLE (SMLE) is the MLE since $\eta_t \sim N(0, 1)$. The SMLE here is the SQMLE defined in Section 4. We compare the performances of MLEs, SMLE, and local MLEs (LMLE).

The empirical means, standard deviations (SD), and asymptotic standard deviations (AD) of these estimators for θ_{10} and θ_{20} are summarized in Tables 1 and 2 when the sample sizes are $n = 400$ and $n = 600$ with $d = 0.1$, respectively. The results are similar for other cases and hence they are not reported here. From the two tables, we can see that the SD and AD of the MLE and LMLE are almost identical, but they are smaller than those of the SMLE, respectively. Furthermore, we see that the SDs and ADs of all estimators become smaller and SDs and ADs become closer as the sample size n is increased from 400 to 600. This is the same as the usual results in the AR-GARCH model.

TABLE 1. Mean and standard deviation of MLE, SMLE, and LMLE of θ_{10} and θ_{20}, $n = 400$ and $k_0 = 200$

		$\theta_{10} = (0.6, 0.1, 0.1, 0.45)$				$\theta_{20} = (0.7, 0.2, 0.2, 0.55)$			
		$\hat{\phi}_{1n}$	$\hat{\alpha}_{10n}$	$\hat{\alpha}_{11n}$	$\hat{\beta}_{11n}$	$\hat{\phi}_{2n}$	$\hat{\alpha}_{20n}$	$\hat{\alpha}_{21n}$	$\hat{\beta}_{21n}$
MLE	Mean	.5911	.0986	.1166	.4340	.6942	.2718	.2088	.4575
	SD	.0648	.0475	.0798	.2271	.0564	.1384	.0947	.2014
	AD	.0586	.0590	.0799	.2810	.0533	.1474	.0964	.2275
SMLE	Mean	.5899	.0915	.1427	.4448	.7015	.3096	.2520	.3790
	SD	.0828	.0560	.1331	.2791	.0842	.1899	.1489	.2691
	AD	.0720	.0616	.1062	.3018	.0754	.2185	.1453	.3334
LMLE	Mean	.5861	.0957	.1350	.4300	.6898	.2446	.2207	.4880
	SD	.0682	.0534	.1004	.2564	.0574	.1531	.1112	.2318
	AD	.0598	.0540	.0838	.2580	.0529	.1281	.0943	.1992

TABLE 2. Mean and standard deviation of MLE, SMLE, and LMLE of θ_{10} and θ_{20}, $n = 600$ and $k_0 = 300$

		$\theta_{10} = (0.6, 0.1, 0.1, 0.45)$				$\theta_{20} = (0.7, 0.2, 0.2, 0.55)$			
		$\hat{\phi}_{1n}$	$\hat{\alpha}_{10n}$	$\hat{\alpha}_{11n}$	$\hat{\beta}_{11n}$	$\hat{\phi}_{2n}$	$\hat{\alpha}_{20n}$	$\hat{\alpha}_{21n}$	$\hat{\beta}_{21n}$
MLE	Mean	.5952	.0992	.1092	.4396	.6958	.2562	.2059	.4749
	SD	.0519	.0462	.0659	.2241	.0458	.1216	.0795	.1841
	AD	.0479	.0551	.0655	.2642	.0434	.1179	.0788	.1852
SMLE	Mean	.5944	.0988	.1227	.4294	.7009	.2891	.2380	.4085
	SD	.0636	.0568	.0966	.2834	.0674	.1666	.1169	.2501
	AD	.0581	.0611	.0847	.2977	.0619	.1713	.1181	.2675
LMLE	Mean	.5918	.1004	.1161	.4285	.6927	.2370	.2128	.4982
	SD	.0527	.0526	.0742	.2544	.0461	.1345	.0878	.2075
	AD	.0483	.0527	.0664	.2512	.0433	.1060	.0774	.1675

Table 3 reports the mean, SD, 90% range, and estimated asymptotic confidence interval (EACI) of k_0 when $n = 400$, 600, and 900. The empirical mean and SD are the average and SD of the estimated k_0 from the 4,000 replications. The 90% range is the range between the 5% and 95% quantiles of the distribution of the estimated k_0. For the case with LMLE, the EACI is computed by the following formula:

$$\left[\tilde{k}_n - [\Delta F_{\omega/2}] - 1, \tilde{k}_n - [\Delta F_{\omega/2}] + 1\right],$$

where $F_{\omega/2}$ is the ωth quantile of the distribution F, $\Delta = (\tilde{d}'\tilde{\Omega}_2 d)(\tilde{d}'\tilde{\Sigma}_2\tilde{d})^{-2}$, $\tilde{d} = \tilde{\theta}_{2n} - \tilde{\theta}_{1n}$, and $\tilde{\Sigma}_2$ and $\tilde{\Omega}_2$ are the corresponding sample estimators of Σ_2

TABLE 3. MLE, SMLE, LMLE, and confidence intervals of the change point k_0

n	k_0		Mean	SD	90% Range	EACI
				d = 0.05		
400	200	MLE	203.2733	36.2125	[136,273]	[175,231]
		SMLE	191.3768	42.4147	[110,269]	[163,221]
		LMLE	194.9743	41.4121	[113,270]	[164,225]
600	300	MLE	303.9853	35.4312	[239,372]	[271,336]
		SMLE	295.0708	40.3188	[217,369]	[260,333]
		LMLE	299.5288	38.0858	[225,369]	[267,331]
900	450	MLE	453.4170	33.5231	[391,515]	[417,488]
		SMLE	445.9773	39.6844	[370,519]	[405,487]
		LMLE	451.7503	35.9959	[383,519]	[416,486]
				d = 0.1		
400	200	MLE	202.7985	15.8064	[183,227]	[193,212]
		SMLE	195.0858	26.0558	[140,228]	[182,206]
		LMLE	198.3490	24.3610	[150,229]	[188,208]
600	300	MLE	302.5913	15.6842	[282,328]	[292,312]
		SMLE	296.9988	22.7459	[254,327]	[283,310]
		LMLE	300.3470	19.0976	[270,327]	[290,310]
900	450	MLE	452.6083	14.6081	[433,475]	[441,463]
		SMLE	447.9118	20.9562	[409,477]	[432,462]
		LMLE	451.0370	16.9072	[426,474]	[440,462]
				d = 0.2		
400	200	MLE	201.3980	5.6372	[195,210]	[198,204]
		SMLE	195.0540	17.7501	[157,207]	[187,200]
		LMLE	197.7855	14.4514	[176,208]	[194,201]
600	300	MLE	301.4453	5.5491	[295,310]	[298,304]
		SMLE	296.4353	15.4377	[271,307]	[289,302]
		LMLE	298.6930	11.8694	[286,308]	[295,302]
900	450	MLE	451.4530	5.2890	[445,460]	[448,454]
		SMLE	447.8860	11.7595	[430,458]	[440,454]
		LMLE	449.9695	8.3877	[441,458]	[446,453]

and $\mathbf{\Omega}_2$, respectively, based on the LMLE $\tilde{\theta}_{2n}$ of θ_{20} and the data set $\{y_{\tilde{k}}, \ldots, y_n\}$. Using the density function $f(x)$ in Section 3, we obtain $F_{0.05} = 7.792$. The EACI is computed similarly for other cases. From Table 3, we see that the means are almost unbiased in all cases. The SD and the length of EACI change just a little

when n increases. This is because the estimated k_0 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. However, the SD and the length of EACI decrease a lot when d increases, which implies that the estimators are more accurate when d is larger. The 90% range is slightly wider than EACI in all cases. The EACIs based on MLE and SMLE are almost identical, but they are generally narrower than those based on the SMLE. This simulation study indicates that our results should be useful in practice.

6. PROOFS OF THEOREMS 2.1 AND 2.2

Proof of Theorem 2.1. First, when $y_t \in M(\vartheta)$, using the ergodic theorem and Assumption 2.1 and partitioning Θ into finite balls with radius δ small enough, we can show that

$$\lim_{n\to\infty} \max_{\Theta} \left| \frac{1}{[n\tau_2]-[n\tau_1]} \sum_{t=[n\tau_1]+1}^{[n\tau_2]} [l_t(\vartheta) - El_t(\vartheta)] \right| = 0 \ \ a.s., \tag{6.1}$$

for any fixed τ_1 and τ_2 with $\tau_2 > \tau_1 \geq 0$, as $n \to \infty$.

(a) We prove only for the case when $\hat{k}_n \leq k_0$, while other case is similar. Denote

$$\Delta_n(k,\theta_1,\theta_2) = L_n(k,\theta_1,\theta_2) - L_n(k_0,\theta_{10},\theta_{20}).$$

We use the convention: $\sum_{t=k_0+1}^{k_0} X_t = 0$ for any series X_t. When $k \leq k_0$, we have

$$\begin{aligned}\Delta_n(k,\theta_1,\theta_2) = &\sum_{t=1}^{k}\{l_t(\theta_1) - l_t(\theta_{10})\} + \sum_{t=k_0+1}^{n}\{l_t(\theta_2) - l_t(\theta_{20})\} \\ &+ \sum_{t=k+1}^{k_0}\{l_t(\theta_2) - l_t(\theta_{10})\}.\end{aligned} \tag{6.2}$$

By Lemma 9.1 in Appendix, we can assume that $\hat{k}_n, k \in [k_L, n-k_L]$, where $k_L = [n\tilde{\tau}]$, $\tilde{\tau} \in (0,1/2)$ and $\tau_0 \in (\tilde{\tau}, 1-\tilde{\tau})$. Let $\Theta_\delta = \{\theta_1 : \|\theta_1 - \theta_{10}\| \geq \delta\}$. By Assumption 2.1, $C = \max_{\theta_1\in\Theta_\delta}[El_t(\theta_1) - El_t(\theta_{10})] < 0$ when $t \leq k_0$. Thus, by (6.1) and Lemma 1 in Chow and Teicher (1968, p. 31), we have

$$\begin{aligned}&\frac{1}{n}\max_{k_L\leq k\leq k_0}\max_{\theta_1\in\Theta_\delta}\left[\sum_{t=1}^{k}\{l_t(\theta_1) - l_t(\theta_{10})\}\right] \\ &\quad\leq \frac{2}{n}\max_{k_L\leq k\leq k_0}\max_{\theta_1\in\Theta_\delta}\left|\sum_{t=1}^{k}[l_t(\theta_1) - El_t(\theta_1)]\right| + \tilde{\tau}\max_{\theta_1\in\Theta_\delta}[El_t(\theta_1) - El_t(\theta_{10})] \\ &\quad= \tilde{\tau}C + o_p(1).\end{aligned} \tag{6.3}$$

Since $\max_{\theta_2\in\Theta}[El_t(\theta_2)-El_t(\theta_{20})]=0$ when $t>k_0$, by (6.1), it follows that

$$\begin{aligned}
&\frac{1}{n}\max_{\theta_2\in\Theta}\left[\sum_{t=k_0+1}^{n}\{l_t(\theta_2)-l_t(\theta_{20})\}\right]\\
&\quad\le\frac{2}{n}\max_{\theta_2\in\Theta}\left|\sum_{t=k_0+1}^{n}[l_t(\theta_2)-El_t(\theta_2)]\right|+(1-\tau_0)\max_{\theta_2\in\Theta}[El_t(\theta_2)-El_t(\theta_{20})]\\
&\quad=o_p(1).
\end{aligned}\tag{6.4}$$

Note that $\max_{\theta\in\Theta}[El_t(\theta)-El_t(\theta_{10})]=0$ when $t\le k_0$. When $|k_0-k|\ge M$,

$$\begin{aligned}
&\frac{1}{n}\max_{k-k_0\le -M}\sup_{\theta_2\in\Theta}\left[\sum_{t=k+1}^{k_0}\{l_t(\theta_2)-l_t(\theta_{10})\}\right]\\
&\quad\le\frac{2}{n}\max_{k-k_0\le -M}\sup_{\theta\in\Theta}\left|\sum_{t=k+1}^{k_0}[l_t(\theta)-El_t(\theta)]\right|+\frac{k_0-k}{n}\max_{\theta\in\Theta}[El_t(\theta)-El_t(\theta_{10})]\\
&\quad\le 2\max_{k-k_0\le -M}\sup_{\theta\in\Theta}\frac{1}{k_0-k}\left|\sum_{t=k+1}^{k_0}[l_t(\theta)-El_t(\theta)]\right|\\
&\quad=_d 2\max_{u\ge M}\sup_{\theta\in\Theta}\frac{1}{u}\left|\sum_{t=-u+1}^{0}[l_t(\theta)-El_t(\theta)]\right|\ \text{(by stationarity of }\{y_t\})\\
&\quad=o_{pM}(1),
\end{aligned}\tag{6.5}$$

where $o_{pM}(1)\to 0$ in probability when $M\to\infty$ and it holds uniformly in n, where the last step holds by Assumption 2.2 and Lemma 1 in Chow and Teicher (1968, p. 31), and "$=_d$" denotes "$=$" in distribution. On the event $\{\|\hat\theta_{1n}-\theta_{10}\|\ge\delta, |\hat k_n-k_0|\ge M\}$,

$$\max_{(\theta_1,\theta_2)\in\Theta_\delta\times\Theta}\ \max_{|k_0-k|\ge M}\Delta_n(k,\theta_1,\theta_2)\ge 0.$$

Furthermore, by (6.2)–(6.5), we have

$$\begin{aligned}
P(\|\hat\theta_{1n}-\theta_{10}\|\ge\delta, |\hat k_n-k_0|\ge M)&\le P\left(\frac{1}{n}\max_{(\theta_1,\theta_2)\in\Theta_\delta\times\Theta}\ \max_{|k_0-k|\ge M}\Delta_n(k,\theta_1,\theta_2)\ge 0\right)\\
&\le P(\tilde\tau C+o_{pM}(1)+o_p(1)\ge 0)\to 0,
\end{aligned}\tag{6.6}$$

as $M,n\to\infty$. When $|k_0-k|\le M$, the third term of (6.2) is less than

$$2\sum_{t=k+1}^{k_0}\max_{\theta_1\in\Theta}|l_t(\theta_1)|=o_p(n).\tag{6.7}$$

On the event $\{\|\hat{\theta}_{1n} - \theta_{10}\| \geq \delta, |\hat{k}_n - k_0| \leq M\}$,

$$\max_{(\theta_1,\theta_2)\in\Theta_\delta\times\Theta} \max_{|k_0-k|\leq M} \Delta_n(k,\theta_1,\theta_2) \geq 0.$$

Furthermore, by (6.2)–(6.4) and (6.7), we have

$$\begin{aligned} P(\|\hat{\theta}_{1n} - \theta_{10}\| \geq \delta, |\hat{k}_n - k_0| \leq M) &\leq P\left(\frac{1}{n}\max_{\Theta_\delta} \max_{|k_0-k|\leq M} \Delta_n(k,\theta_1,\theta_2) \geq 0\right) \\ &\leq P(\tilde{\tau}C + o_p(1) \geq 0) \to 0, \end{aligned} \tag{6.8}$$

as $n \to \infty$ for any given M. By (6.6) and (6.8), we can see that $\hat{\theta}_{1n} - \theta_{10} = o_p(1)$. Similarly, we can show that $\hat{\theta}_{2n} - \theta_{20} = o_p(1)$. Thus, (a) holds.

(b) We note that

$$\sum_{t=1}^{\hat{k}_n} l_t(\hat{\theta}_{1n}) + \sum_{\hat{k}_n+1}^{n} l_t(\hat{\theta}_{2n}) \geq L_n[k_0, \hat{\theta}_{1n}(k_0), \hat{\theta}_{2n}(k_0)] \geq \sum_{t=1}^{k_0} l_t(\hat{\theta}_{1n}) + \sum_{k_0+1}^{n} l_t(\hat{\theta}_{2n}).$$

Thus,

$$-\sum_{t=\hat{k}_n+1}^{k_0} l_t(\hat{\theta}_{1n}) + \sum_{t=\hat{k}_n+1}^{k_0} l_t(\hat{\theta}_{2n}) \geq 0.$$

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

$$\begin{aligned} &\lim_{n\to\infty} |El_t(\hat{\theta}_{in}) - El_t(\theta_{i0})| \\ &\quad \leq \lim_{\delta\to 0}\left[E \sup_{\|\theta_i-\theta_{i0}\|<\delta} |l_t(\theta_i) - l_t(\theta_{i0})| + \lim_{n\to\infty} E\xi_t I\{\|\hat{\theta}_{in} - \theta_{i0}\| \geq \delta\}\right] = 0, \end{aligned}$$

$i = 1, 2$, uniformly in t, where $\xi_t = 2\max_{\theta_i\in\Theta} |l_t(\theta_i)|$.

Denote $C_0 = El_t(\theta_{10}) - El_t(\theta_{20})$. Then $C_0 > 0$ when $t \leq k_0$ by Assumption 2.1. Thus, by the previous two inequalities, we have

$$\begin{aligned} &\frac{2}{k_0 - \hat{k}_n} \sup_{\theta\in\Theta} \left| \sum_{t=\hat{k}_n+1}^{k_0} [l_t(\theta) - El_t(\theta)] \right| \\ &\quad \geq \frac{1}{k_0 - \hat{k}_n} \left\{ -\sum_{t=\hat{k}_n+1}^{k_0} [l_t(\hat{\theta}_{1n}) - El_t(\hat{\theta}_{1n})] + \sum_{t=\hat{k}_n+1}^{k_0} [l_t(\hat{\theta}_{2n}) - El_t(\hat{\theta}_{2n})] \right\} \\ &\quad \geq \frac{1}{k_0 - \hat{k}_n} \sum_{t=\hat{k}_n+1}^{k_0} \left[El_t(\hat{\theta}_{1n}) - El_t(\hat{\theta}_{2n})\right] = C_0 + o(1), \end{aligned}$$

as $n \to \infty$. By the previous inequality and Assumption 2.2, for any $\varepsilon > 0$, we have

$$\begin{aligned}
P(k_0 - \hat{k}_n > M) &= P\left(k_0 - \hat{k}_n > M, \frac{2}{k_0 - \hat{k}_n} \sup_{\theta \in \Theta} \left| \sum_{t=\hat{k}_n+1}^{k_0} [l_t(\theta) - El_t(\theta)] \right| \geq C_0 + o(1)\right) \\
&\leq P\left(\max_{k_0 - k > M} \frac{2}{k_0 - k} \sup_{\theta \in \Theta} \left| \sum_{t=k+1}^{k_0} [l_t(\theta) - El_t(\theta)] \right| \geq C_0 + o(1)\right) \\
&\leq P\left(\max_{k_0 - k > M} \frac{2}{k_0 - k} \sup_{\theta \in \Theta} \left| \sum_{t=k+1}^{k_0} [l_t(\theta) - El_t(\theta)] \right| \geq \frac{C_0}{2}\right) + \frac{\varepsilon}{2} \\
&= P\left(\max_{u > M} \frac{2}{u} \sup_{\theta \in \Theta} \left| \sum_{t=-u}^{-1} [l_t(\theta) - El_t(\theta)] \right| \geq \frac{C_0}{2}\right) + \frac{\varepsilon}{2} < \varepsilon,
\end{aligned}$$

as $M > 0$ is large enough, where the last equation holds by the stationarity of $\{y_t\}$. Thus, $k_0 - \hat{k}_n = O_p(1)$. This completes the proof. ■

Proof of Theorem 2.2. Denote

$$\begin{aligned}
\hat{u}_1 &= \sqrt{k_0}(\hat{\theta}_{1n} - \theta_{10}), \\
\hat{u}_2 &= \sqrt{n - k_0}(\hat{\theta}_{2n} - \theta_{20}), \\
u_1^* &= \frac{\Sigma_{\theta_{10}}^{-1}}{\sqrt{k_0}} \sum_{t=1}^{k_0} D_t(\theta_{10}), \\
u_2^* &= \frac{\Sigma_{\theta_{20}}^{-1}}{\sqrt{n - k_0}} \sum_{t=k_0+1}^{n} D_t(\theta_{10}).
\end{aligned}$$

By Assumption 2.3, we have

$$\begin{aligned}
\sum_{t=1}^{k_0} \{l_t(\hat{\theta}_{1n}) - l_t(\theta_{10})\} &= \hat{u}_1' \Sigma_{\theta_{10}} u_1^* - \hat{u}_1' \left[\frac{1}{2} \Sigma_{\theta_{10}} + o_p(1)\right]]\hat{u}_1 \\
&= \frac{1}{2}\left[-\| \Sigma_{\theta_{10}}^{1/2}(\hat{u}_1 - u_1^*)\|^2 + \| \Sigma_{\theta_{10}}^{1/2} u_1^* \|^2\right](1 + o_p(1)), \qquad \textbf{(6.9)} \\
\sum_{t=k_0+1}^{n} \{l_t(\hat{\theta}_{2n}) - l_t(\theta_{20})\} &= \hat{u}_2' \Sigma_{\theta_{20}} u_2^* - \hat{u}_2' \left[\frac{1}{2} \Sigma_{\theta_{20}} + o_p(1)\right] \hat{u}_2 \\
&= \frac{1}{2}\left[-\| \Sigma_{\theta_{20}}^{1/2}(\hat{u}_2 - u_2^*)\|^2 + \| \Sigma_{\theta_{20}}^{1/2} u_2^* \|^2\right](1 + o_p(1)). \qquad \textbf{(6.10)}
\end{aligned}$$

When $\hat{k}_n < k_0$, by (6.2), (6.9), and (6.10), we have

$$\begin{aligned}
&\Delta_n\left(\hat{k}_n, \hat{\theta}_{1n}, \hat{\theta}_{2n}\right) \\
&\quad = -\frac{1}{2}\left[\| \Sigma_{\theta_{10}}^{1/2}(\hat{u}_1 - u_1^*)\|^2 + \| \Sigma_{\theta_{20}}^{1/2}(\hat{u}_2 - u_2^*)\|^2\right](1 + o_p(1))
\end{aligned}$$

$$+\frac{1}{2}\left(\|\Sigma_{\theta_{10}}^{1/2}u_1^*\|+\|\Sigma_{\theta_{20}}^{1/2}u_2^*\|^2\right)(1+o_p(1))$$
$$+\sum_{t=\hat{k}_n+1}^{k_0}\{l_t(\hat{\theta}_{2n})-l_t(\hat{\theta}_{1n})\}. \tag{6.11}$$

Let $\theta_1^*=\theta_{10}+u_1^*/\sqrt{k_0}$ and $\theta_2^*=\theta_{20}+u_2^*/\sqrt{n-k_0}$. By Assumption 2.3, similar to (6.11), we have

$$\Delta_n(\hat{k}_n,\theta_1^*,\theta_2^*)=\frac{1}{2}(\|\Sigma_{\theta_{10}}^{1/2}u_1^*\|^2+\|\Sigma_{\theta_{20}}^{1/2}u_2^*\|^2)(1+o_p(1))$$
$$+\sum_{t=\hat{k}_n+1}^{k_0}\{l_t(\theta_2^*)-l_t(\theta_1^*)\}.$$

Note that $l_t(\theta_i^*)\to_p l_t(\theta_{i0})$ and $l_t[\hat{\theta}_{in}]\to_p l_t(\theta_{i0})$, $i=1,2$. Given any $M>0$, on the event $K_n=\{0<k_0-\hat{k}_n<M\}$,

$$0\le\Delta_n\left(\hat{k}_n,\hat{\theta}_{1n},\hat{\theta}_{2n}\right)-\Delta_n(\hat{k}_n,\theta_1^*,\theta_2^*)$$
$$=-\frac{1}{2}\left[\|\Sigma_{\theta_{10}}^{1/2}(\hat{u}_1-u_1^*)\|^2+\|\Sigma_{\theta_{20}}^{1/2}(\hat{u}_2-u_2^*)\|^2\right](1+o_p(1))+o_p(1).$$

Thus, we can claim that, on K_n,

$$\hat{u}_i-u_i^*=o_p(1),\quad i=1,2. \tag{6.12}$$

Similarly, we can show that on the event $K_n=\{-M<k_0-\hat{k}_n<0\}$, (6.12) holds. Since $\hat{k}_n-k_0=O_p(1)$, we can claim that (a) holds by the central limit theorem. Furthermore, we have

$$\Delta_n\left(\hat{k}_n,\hat{\theta}_{1n},\hat{\theta}_{2n}\right)=\frac{1}{2}\left(\|\Sigma_{\theta_{10}}^{1/2}u_1^*\|^2+\|\Sigma_{\theta_{20}}^{1/2}u_2^*\|^2\right)$$
$$+I\{\hat{k}_n\le k_0\}\sum_{t=\hat{k}_n+1}^{k_0}[l_t(\theta_{20})-l_t(\theta_{10})]$$
$$+I\{\hat{k}_n>k_0\}\sum_{t=k_0+1}^{\hat{k}_n}[l_t(\theta_{20})-l_t(\theta_{10})]+o_p(1).$$

Thus, by the strict stationarity of $\{y_t\}$,

$$\hat{k}_n-k_0=\text{argmax}_k\left\{I\{k\le k_0\}\sum_{t=k+1}^{k_0}[l_t(\theta_{20})-l_t(\theta_{10})]+I\{k>k_0\}\right.$$
$$\left.\times\sum_{t=k_0+1}^{k}[l_t(\theta_{10})-l_t(\theta_{20})]\right\}$$

$$-k_0 + o_p(1) \longrightarrow_{\mathcal{L}} \operatorname{argmax}_k W(k, \theta_{10}, \theta_{20}),$$

as $n \to \infty$. This completes the proof. ■

7. PROOFS OF THEOREMS 4.1–4.3

We first give one lemma which is used for the $L^p(\nu)$-NED of $w_t l_t(\vartheta)$. Its proof is given in Ling (2014).

LEMMA 7.1. *Suppose that* $\{y_t\} \in M(\vartheta)$. *If Assumptions 4.1–4.2 hold, then, for any* $\iota \in (0,1)$, *it follows that*

(a) $E|h_t - E[h_t|\mathcal{F}_k(t)]|^\iota = O(\rho^k)$,
(b) $E|\varepsilon_t - E[\varepsilon_t|\mathcal{F}_k(t)]|^{2\iota} = O(\rho^k)$,
(c) $E|y_t - E[y_t|\mathcal{F}_k(t)]|^{2\iota} = O(\rho^k)$.

Proof of Theorem 4.1. Assumption 2.1 was verified in Ling (2007b). We only need to verify Assumption 2.2. First, by Assumptions 4.1–4.2, $\varepsilon_t(\vartheta)$ and $h_t(\vartheta)$ have the following expansions:

$$\varepsilon_t(\vartheta) = \sum_{i=0}^{\infty} a_i(\vartheta) y_{t-i} \text{ and } h_t(\vartheta) = \sum_{i=0}^{\infty} b_i(\vartheta)\varepsilon_t^2(\vartheta), \tag{7.1}$$

where $\sup_\Theta a_i(\vartheta) = O(\rho^i)$ and $\sup_\Theta b_i(\vartheta) = O(\rho^i)$ with $\rho \in (0,1)$. Thus, we have $\sup_{\vartheta\in\Theta}|\varepsilon_t(\vartheta)| \le \xi_{\rho t}$ and $\sup_{\vartheta\in\Theta} h_t(\vartheta) \le \xi_{\rho t}^2$, where $\xi_{\rho t} = C + C \sum_{i=0}^{\infty} \rho^i |y_{t-i}|$ and C is a constant. Using this, we can show that $E\sup_{\vartheta\in\Theta} |w_t l_t(\vartheta)|^{1+\iota} < \infty$.

We now show that

$$E\sup_\Theta |wl_t(\vartheta) - E[w_t l_t(\vartheta)|\mathcal{F}_k(t)]|^{1+\iota} = O(k^{-\nu}), \tag{7.2}$$

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

$$E\sup_\Theta |\varepsilon_t^2(\vartheta) - E[\varepsilon_t^2(\vartheta)|\mathcal{F}_k(t)]|^{2\iota} = O(\rho^k) \ \textit{and}$$
$$E\sup_\Theta |h_t(\vartheta) - E[h_t(\vartheta)|\mathcal{F}_k(t)]|^\iota = O(\rho^k),$$

for any $\iota \in (0,1)$. Furthermore, we can show that, for small enough $\iota > 0$,

$$E|w_t - E[w_t|\mathcal{F}_k(t)]|^\iota = O(k^{-\nu}) \text{ and } E\sup_\Theta \left|\frac{\varepsilon_t^2(\vartheta)}{h_t(\vartheta)} - E\left[\frac{\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}\middle|\mathcal{F}_k(t)\right]\right|^{2\iota} = o(\rho^k).$$

Note that w_t and h_t^{-1} is bounded and $E|\varepsilon_t(\vartheta)|^{8\iota} < \infty$ as $\iota < 1/4$. By the previous equations, we have

$$
\begin{aligned}
&E\sup_{\Theta}\left|\frac{w_t\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}-E\left[\frac{w_t\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}\middle|\mathcal{F}_k(t)\right]\right|^{2\iota}\\
&\quad\le O(1)E\sup_{\Theta}\left|\frac{w_t\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}-E\left[w_t\middle|\mathcal{F}_k(t)\right]E\left[\frac{\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}\middle|\mathcal{F}_k(t)\right]\right|^{2\iota}\\
&\quad\le O(1)E\left\{w_t\sup_{\Theta}\left|\frac{\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}-E\left[\frac{\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}\middle|\mathcal{F}_k(t)\right]\right|\right\}^{2\iota}\\
&\qquad+O(1)E\sup_{\Theta}\left\{\frac{\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}\middle|w_t-E[w_t|\mathcal{F}_k(t)]\right|\right\}^{2\iota}\\
&\quad\le O(1)E\sup_{\Theta}\left|\frac{\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}-E\left[\frac{\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}\middle|\mathcal{F}_k(t)\right]\right|^{2\iota}\\
&\qquad+O(1)\left(E\sup_{\Theta}\left\{\frac{\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}\right\}^{4\iota}E\left|w_t-E[w_t|\mathcal{F}_k(t)]\right|^{4\iota}\right)^{1/2}\\
&\quad=O(k^{-\nu}).
\end{aligned}
$$

By Holder's inequality and the previous inequality, we have

$$
\begin{aligned}
&E\sup_{\Theta}\left|\frac{w_t\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}-E\left[\frac{w_t\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}\middle|\mathcal{F}_k(t)\right]\right|^{1+\iota}\\
&\quad\le\left[E\sup_{\Theta}\left|\frac{w_t\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}-E\left[\frac{w_t\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}\middle|\mathcal{F}_k(t)\right]\right|^{2}\right]^{1/2}\\
&\qquad\times\left[E\sup_{\Theta}\left|\frac{w_t\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}-E\left[\frac{w_t\varepsilon_t^2(\vartheta)}{h_t(\vartheta)}\middle|\mathcal{F}_k(t)\right]\right|^{2\iota}\right]^{1/2}\\
&\quad\le\sqrt{E(w_t\xi_{\rho t}^2)^2}\,O(k^{-\nu})=O(k^{-\nu}).
\end{aligned}
$$

Similarly, we can show that

$$
E\sup_{\Theta}\left|w_t\log h_t(\vartheta)-E\left[w_t\log h_t(\vartheta)|\mathcal{F}_k(t)\right]\right|^{1+\iota}=O(k^{-\nu}).
$$

By the previous two equations, we can see that (7.2) holds.

By (7.2), $\{w_tl_t(\vartheta)\}$ is an $L^{1+\iota}(v)$-NED sequence. By Theorem 2.1 of Ling (2007a), for each $\vartheta\in\Theta$,

$$
\frac{1}{n}\sum_{t=-n}^{-1}w_tl_t(\vartheta)=E(w_tl_t(\vartheta))+o(1), \tag{7.3}
$$

as $n\to\infty$. Denote $V_\delta=\{\vartheta^*:\|\vartheta-\vartheta^*\|\le\delta\}\subset\Theta$. Let

$$
\xi_t=\sup_{\theta^*\in V_\delta}|wl_t(\vartheta^*)-wl_t(\vartheta)|\ \ and\ \ \tilde{\xi}_t=\sup_{\theta^*\in V_\delta}|E[wl_t(\vartheta^*)|\mathcal{F}_k(t)]-E[wl_t(\vartheta)|\mathcal{F}_k(t)]|.
$$

Then, by (7.2),

$$E|\xi_t - \tilde{\xi}_t| \le E \sup_{\theta^* \in V_\delta} \Big| wl_t(\vartheta^*) - E[wl_t(\vartheta^*)|\mathcal{F}_k(t)] - wl_t(\vartheta) + E[wl_t(\vartheta)|\mathcal{F}_k(t)] \Big|$$
$$\le 2E \sup_{\theta^* \in V_\delta} |wl_t(\vartheta^*) - E[wl_t(\vartheta^*)|\mathcal{F}_k(t)]| = O(k^{-\nu}).$$

By the previous inequality, we have

$$E|\xi_t - E[\xi_t|\mathcal{F}_k(t)]|^{1+\iota} \le E|\xi_t - \tilde{\xi}_t + E[\xi_t - \tilde{\xi}_t|\mathcal{F}_k(t)]|^{1+\iota}$$
$$\le CE|\xi_t - \tilde{\xi}_t|^{1+\iota} = O(k^{-\nu}).$$

Thus, $\{\xi_t\}$ is an $L^{1+\iota}(\upsilon)$-NED sequence. By Theorem 2.1 of Ling (2007a), we have

$$\frac{1}{n}\sum_{t=-n}^{-1} \xi_t = E\xi_t + o(1), \tag{7.4}$$

as $n \to \infty$. Using (7.3) and (7.4) and partitioning Θ into finite balls with a radius δ small enough, we can show that Assumption 2.2 holds. By Theorem 2.3, we complete the proof. ■

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

$$\frac{1}{n}\sum_{t=1}^{k_0} \Big\| P_t(\hat{\theta}_{1n}) - P_t(\theta_{10}) \Big\| = o_p(1). \tag{7.5}$$

Note that $\hat{k} - k_0$ is bounded in probability. When $\hat{k}_n < k_0$, we can show that

$$\frac{1}{n}\sum_{t=\hat{k}_n+1}^{k_0} \Big\| P_t(\hat{\theta}_{1n}) \Big\| \le \frac{1}{n}\sum_{t=\hat{k}_n+1}^{k_0} \Big\| P_t(\hat{\theta}_{1n}) - P_t(\theta_{10}) \Big\|$$
$$+ \frac{1}{n}\sum_{t=\hat{k}_n+1}^{k_0} \|P_t(\theta_{10})\| = o_p(1). \tag{7.6}$$

Thus, by the previous two inequalities, we have

$$\frac{1}{n}\sum_{t=1}^{\hat{k}_n} P_t(\hat{\theta}_{1n}) = \frac{1}{n}\sum_{t=1}^{k_0} P_t(\hat{\theta}_{1n}) - \frac{1}{n}\sum_{t=\hat{k}_n+1}^{k_0} P_t(\hat{\theta}_{1n}) = \tau_0 \Sigma_1 + o_p(1). \tag{7.7}$$

By Taylor's expansion, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{\hat{k}_n} D_t(\hat{\theta}_{1n}) = \frac{1}{\sqrt{n}}\sum_{t=1}^{\hat{k}_n} D_t(\theta_{10}) + \left[\frac{1}{n}\sum_{t=1}^{\hat{k}_n} P_t(\tilde{\xi}_{1n})\right][\sqrt{n}(\hat{\theta}_{1n} - \theta_{10})], \tag{7.8}$$

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

$$\tilde{\theta}_{1n} = \theta_{10} + \frac{\Sigma_1^{-1}}{\tau_0 n} \sum_{t=1}^{k_0} D_t(\theta_{10}) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{7.9}$$

A similar expansion holds for $\tilde{\theta}_{2n}$. Thus, by the central limiting theorem, we see that (a) holds when $\hat{k}_n < k_0$. Similarly, we can show that (a) holds when $\hat{k}_n > k_0$.

For (b), using (7.9) and by Taylor's expansion, we have

$$\begin{aligned} &L_n(\tilde{k}, \tilde{\theta}_{1n}, \tilde{\theta}_{2n}) - L_n(k_0, \theta_{10}, \theta_{20}) \\ &\quad = I\{\tilde{k}_n \le k_0 \sum_{t=\hat{k}_n+1}^{k_0} [l_t(\theta_{20}) - l_t(\theta_{10})] + I\{\tilde{k}_n > k_0 \sum_{t=k_0+1}^{\tilde{k}_n} [l_t(\theta_{20}) - l_t(\theta_{10})] \\ &\qquad + \frac{1}{2}\left(\|\Sigma_1^{1/2} u_1^*\|^2 + \|\Sigma_2^{1/2} u_2^*\|^2\right) + o_p(1), \end{aligned}$$

where $u_1^* = \sqrt{n\tau_0}(\tilde{\theta}_{1n} - \theta_{10})$ and $u_2^* = \sqrt{n(1-\tau_0)}(\tilde{\theta}_{2n} - \theta_{20})$. Thus, by the strict stationarity of $\{y_t\}$,

$$\begin{aligned} \tilde{k}_n - k_0 = \operatorname{argmax}_k \Bigg\{ & I\{k \le k_0\} \sum_{t=k+1}^{k_0} [l_t(\theta_{20}) - l_t(\theta_{10})] \\ & + I\{k > k_0\} \sum_{t=k_0+1}^{k} [l_t(\theta_{10}) - l_t(\theta_{20})] \Bigg\} \\ & - k_0 + o_p(1) \longrightarrow_{\mathcal{L}} \operatorname{argmax}_k W(k, \theta_{10}, \theta_{20}), \end{aligned}$$

as $n \to \infty$. This completes the proof. ■

We now give one more lemma. It is for the NED property of $D_t(\vartheta_0)$ and $P_t(\vartheta_0)$ when $\{y_t\} \in M(\vartheta_0)$.

LEMMA 7.2. *If $\{y_t\} \in M(\vartheta_0)$ and Assumptions 4.1–4.2 hold, then there exists a constant $\iota \in (0,1)$ such that*

(a) $E\left\| \frac{1}{\sqrt{h_t}} \frac{\partial \varepsilon_t(\vartheta_0)}{\partial \gamma} - E\left[\frac{1}{\sqrt{h_t}} \frac{\partial \varepsilon_t(\vartheta_0)}{\partial \gamma} \middle| \mathcal{F}_k(t)\right] \right\|^{2+\iota} = O(\rho^k)$,

(b) $E\left\| \frac{1}{h_t} \frac{\partial h_t(\vartheta_0)}{\partial \vartheta} - E\left[\frac{1}{h_t} \frac{\partial h_t(\vartheta_0)}{\partial \vartheta} \middle| \mathcal{F}_k(t)\right] \right\|^{2+\iota} = O(\rho^k)$,

where ρ is a constant with $\rho \in (0,1)$.

To make notation clear in the proof of Theorem 4.3, when $y_t \in M(\theta_{i0})$, y_t is denoted by y_{it}, $\varepsilon_t(\theta)$ by $\varepsilon_{it}(\theta)$, $h_t(\theta)$ by $h_{it}(\theta)$, $l_t(\theta)$ by $l_{it}(\theta)$, and similarly for $D_{it}(\theta)$, $P_{it}(\theta)$, $R_{i1t}(\theta)$, $R_{i2t}(\theta)$, and $U_{it}(\theta)$, $i = 1, 2$, etc. We further give one lemma as follows. The proofs of Lemmas 7.2 and 7.3 are in Ling (2014).

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

(a) $E\left\| \frac{1}{\sqrt{h_{1t}}} \frac{\partial \varepsilon_{1t}(\theta_{10})}{\partial \gamma} - \frac{1}{\sqrt{h_{2t}}} \frac{\partial \varepsilon_{2t}(\theta_{20})}{\partial \gamma} \right\|^2 = o(1)$,

(b) $E\left\| \frac{1}{h_{1t}} \frac{\partial h_{1t}(\theta_{10})}{\partial \vartheta} - \frac{1}{h_{2t}} \frac{\partial h_{2t}(\theta_{20})}{\partial \vartheta} \right\|^2 = o(1)$.

Proof of Theorem 4.3. Assumption 2.3(i) holds by Taylor's expansion and Lemma A.7 of Ling (2007b). Assumption 2.3(ii) and (iii) hold by Theorem 4.1 of Ling (2007b). Taking Taylor expansion at θ_{10}, we have

$$\sum_{t=-[mz]}^{-1} [l_{1t}(\theta_{20}) - l_{1t}(\theta_{10})]$$
$$= -\sum_{t=-[mz]}^{-1} d' D_{1t}(\theta_{10}) - \frac{1}{2} d' \sum_{t=-[mz]}^{-1} P_{1t}(\xi^*) d, \tag{7.10}$$

where ξ^* lies between θ_{10} and θ_{20} and $D_{1t}(\theta_{10}) = U_{1t}(\theta_{10})\xi_t$, where $\xi_t = [\eta_t, (1-\eta_t^2)/\sqrt{2}]'$. Since $\xi^* = \theta_{10} + O(1/\sqrt{m})$, by Lemma A.7 of Ling (2007b), we have

$$\frac{1}{m} \sum_{t=-[mz]}^{-1} P_{1t}(\xi^*) = \frac{1}{m} \sum_{t=-[mz]}^{-1} P_{1t}(\theta_{10}) + o_p(1), \tag{7.11}$$

where $P_{1t}(\theta_{10}) = U_{1t}(\theta_{10}) U_{1t}'(\theta_{10}) + \eta_t R_{11t}(\theta_{10}) + (\eta_t^2 - 1) R_{12t}(\theta_{10})$. Using Lemma 7.3, we can show that

$$\lim_{\|d\| \to 0} E\|U_{1t}(\theta_{10}) - U_{2t}(\theta_{20})\|^2 = 0 \text{ and } \lim_{\|d\| \to 0} \|P_{1t}(\theta_{10}) - P_{2t}(\theta_{20})\| = 0, \tag{7.12}$$

where $P_{2t}(\theta_{20}) = U_{2t}(\theta_{20}) U_{2t}'(\theta_{20}) + \eta_t R_{21t}(\theta_{20}) + (\eta_t^2 - 1) R_{22t}(\theta_{20})$. By (7.10)–(7.12), we can show that

$$\sum_{t=-[mz]}^{-1} [l_{1t}(\theta_{20}) - l_{1t}(\theta_{10})]$$
$$= -\sum_{t=-[mz]}^{-1} d' U_{2t}(\theta_{20}) \xi_t - \frac{1}{2} d' \sum_{t=-[mz]}^{-1} P_{2t}(\theta_{20}) d + o_p(1). \tag{7.13}$$

Using Lemma 7.2, we can show that $\|U_{2t}(\theta_{20})\|^2$ is $L^{1+\iota}(\upsilon)$-NED. Similarly, we can show that $\|R_{i2t}(\theta_{20})\|$, $i = 1, 2$, are also $L^{1+\iota}(\upsilon)$-NED. By Theorem 2.1 of Ling (2007a), we have

$$\frac{1}{m} \sum_{t=-[mz]}^{-1} P_{2t}(\theta_{20}) = \Sigma_{\theta_{20}} z + o_p(1).$$

By the previous three equations, we can see that Assumption 3.1 holds. By Lemma 7.2, $U_{2t}(\theta_{20})$ is $L^{2+\iota}(\upsilon)$-NED and hence so is $D_{2t}(\theta_{20})$, i.e., Assumption 3.2 holds. This completes the proof. ■

8. CONCLUDING REMARKS

This paper has established an asymptotic theory for change-points in linear and nonlinear time series models under some regular conditions. It was shown that the estimated change-point converges weakly to the location of the maxima of a double-sided random walk. When the magnitude of changed parameters is small, this limiting distribution can be further approximated by a known distribution to obtain its approximating quantiles. For the structure-changed ARMA-GARCH/IGARCH model, the self-weighted QMLE and the local QMLE were studied and the corresponding limiting distributions of the estimated change-point are derived, respectively. Our framework includes many other models as a special case, such as long memory FARIMA model, exponential GARCH model, random coefficient AR model, ARCH-type model, and smooth threshold AR model, among others. It can be readily extended to include the stationary multivariate time series models with exogenous variables. However, it cannot be applied for the threshold AR model with unknown thresholds for which some additional techniques are needed. Furthermore, it is still a challenging issue on the estimation of change-points in the unit root/cointegrated time series models if the parameters are fixed before and after the change-point. Some projects on these are ongoing.

REFERENCES

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–855.

Andrews, D.W.K. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62(6), 1383–1414.

Aue, A., L. Hörmann, L. Horváth, & M. Reimherr (2009) Break detection in the covariance structure of multivariate time series models. *Annals of Statistics* 37, 4046–4087.

Bai, J. (1994) Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis* 15, 453–472.

Bai, J. (1995) Least absolute deviation estimation of a shift. *Econometric Theory* 11, 403–436.

Bai, J., R.L. Lumsdaine, & J.H. Stock (1998) Testing for and dating common breaks in multivariate time series. *Review of Economic Studies* 65, 395–432.

Bai, J. & P. Perron (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47–78.

Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.

Breitung, J. & S. Eickmeier (2011) Testing for structural breaks in dynamic factor models. *Journal of Econometrics* 163, 71–84.

Chong, T.T.L. (2001) Structural change in AR(1) models. *Econometric Theory* 17, 87–155.

Chow, G.C. (1960) Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28, 591–604.

Chow, Y.S. & H. Teicher (1978) *Probability Theory. Independence, Interchangeability, Martingales*. Springer-Verlag.

Davis, R.A., D.W., Huang, & Y.C. Yao (1995) Testing for a change in the parameter values and order of an autoregressive model. *Annals of Statistics* 23, 282–303.

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

Dümbgen, L. (1991) The asymptotic behavior of some nonparametric change-point estimators. *Annals of Statistics* 19, 1471–1495.

Francq, C. & J.M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.

Haavelmo, T. (1944) The probability approach in econometrics. *Econometrica* 12, 1–118.

Han, X. & A. Inoue (2014) Tests for parameter instability in dynamic factor models. *Econometric Theory*, accepted.

Hansen, B.E. (1992) Tests for parameters instability in regression with $I(1)$ processes. *Journal of Business and Economic Statistics* 10, 321–335.

Hansen, B.E. (2001) The new econometrics of structural change: Dating changes in U.S. labor productivity. *Journal of Economic Perspectives* 15, 117–128.

Hansen, B.E. (2003) Tests for parameter instability in regressions with I(1) processes. *Journal of Business & Economic Statistics* 20, 45–59. Twentieth anniversary commemorative issue.

Hansen, B.E. (2009) Averaging estimators for regressions with a possible structural break. *Econometric Theory* 25, 1498–1514.

Hendry, D.F. & S. Johansen (2014) Model discovery and Trygve Haavelmo's legacy. *Econometric Theory*, to appear.

Hidalgo, J. & M.H. Seo (2013) Testing for structural stability in the whole sample. *Journal of Econometrics* 175, 84–93.

Hinkley, D.V. (1970) Inference about the change-point in a sequence of random variables. *Biometrika* 57, 1–17.

Hinkley, D.V. & E.A. Hinkley (1970) Inference about the change-point in a sequence of binomial variables. *Biometrika* 57, 477–488.

Horváth, L. (1993) The maximum likelihood method for testing changes in the parameters of normal observations. *Annals of Statistics* 21, 671–680.

Kejriwal, M. & P. Perron (2008) The limit distribution of the estimates in cointegrated regression models with multiple structural changes. *Journal of Econometrics* 146, 59–73.

Lai, T.L. (1995) Sequential changepoint detection in quality control and dynamical systems. With discussion and a reply by the author. *Journal of the Royal Statistical Society, Series B* 57, 613–658.

Ling, S. (2003) Testing and Estimation for Change-Points in GARCH and ARMA-GARCH Models. Working paper, HKUST.

Ling, S. (2007a) Testing for change-points in time series models and limiting theorems for NED sequences. *Annals of Statistics* 35, 1213–1237.

Ling, S. (2007b) Self-weighted and local quasi-maximum likelihood estimator for ARMA-GARCH/IGARCH models. *Journal of Econometrics* 140, 849–873.

Ling, S. (2014) Estimation of change-points in linear and nonlinear time series models—Supplementary Appendix: Proofs of Lemmas 7.1–7.3. Online address: http://www.math.ust.hk/maling/index-2.html.

Ling, S. & M. McAleer (2010) Strong consistency and asymptotic normality of estimates for time series models. *Statistica Neerlandica* 64, 97–111.

Perron, P. & Y. Yamamoto (2014) A note on estimating and testing for multiple structural changes in models with endogenous regressors via 2SLS. *Econometric Theory* 30, 491–507.

Picard, D. (1985) Testing and estimating change-points in time series. *Advances in Applied Probability* 17, 841–867.

Qu, Z. & P. Perron (2007) Estimating and testing structural changes in multivariate regressions. *Econometrica* 75, 459–502.
Quandt, R.E. (1960) Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association* 55, 324–330.
Ritov, Y. (1990) Asymptotic efficient estimation of the change point with unknown distributions. *Annals of Statistics* 18, 1829–1839.
Saikkonen, P., H. Lükepohl, & C. Trenkler (2006) Break date estimation for VAR processes with level shift with an application to cointegration testing. *Econometric Theory* 22, 15–68.
Shao, X. & X. Zhang (2010) Testing for change points in time series. *Journal of the American Statistical Association* 105, 1228–1240.
Stock, J.H. & M.W. Watson (1996) Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics* 14, 11–30.
Su, L. & H. White (2010) Testing structural change in partially linear models. *Econometric Theory* 26, 1761–1806.
Yao, Y.-C. (1987) Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *Annals of Statistics* 15, 1321–1328.
Zhu, K. (2010) On the LAD estimation and likelihood ratio test in time series. Ph.D. thesis, HKUST.

APPENDIX A: Consistency of $\hat{\tau}_n$

The following lemma shows that $\hat{\tau}_n$ is a consistent estimator of τ_0.

LEMMA A.1. *For any given $\tilde{\tau}, \tilde{\tau}_1 \in (0,1)$ with $\tilde{\tau} < \tilde{\tau}_1$, if Assumptions 2.1 and 2.2 hold and $\tau_0 \in (\tilde{\tau}, \tilde{\tau}_1)$, then it follows that*

$$\lim_{n\to\infty} P(\hat{\tau}_n \notin [\tilde{\tau}, \tilde{\tau}_1]) = 0.$$

Proof. We only consider the case with $\hat{\tau}_n < \tilde{\tau}$. By (6.1) and Lemma 1 in Chow and Teicher (1968, p. 31), for any $\epsilon > 0$, we have

$$\lim_{n\to\infty} P\left(\frac{1}{n}\max_{\log n\le k\le [n\tilde{\tau}]}\max_{\theta_1\in\Theta}\left|\sum_{t=1}^{k}[l_t(\theta_1) - El_t(\theta_1)]\right| > \epsilon\right) = 0.$$

Thus, we have

$$\begin{aligned}
&\frac{1}{n}\max_{\log n\le k\le [n\tilde{\tau}]}\max_{\theta_1\in\Theta}\sum_{t=1}^{k}[l_t(\theta_1) - l_t(\theta_{10})] \\
&\quad\le \frac{2}{n}\max_{\log n\le k\le [n\tilde{\tau}]}\max_{\theta_1\in\Theta}\left|\sum_{t=1}^{k}[l_t(\theta_1) - El_t(\theta_1)]\right| \\
&\qquad + \frac{1}{n}\max_{\log n\le k\le [n\tilde{\tau}]} k\max_{\theta_1\in\Theta}[El_t(\theta_1) - El_t(\theta_{10})] \\
&\quad = o_p(1),
\end{aligned}$$

since $\max_{\theta_1\in\Theta}[El_t(\theta_1)-El_t(\theta_{10})]=0$ when $t\le k_0$. By Assumption 2.1 and the ergodic theorem, we have

$$\frac{1}{n}\max_{1\le k\le \log n}\max_{\theta_1\in\Theta}\sum_{t=1}^{k}[l_t(\theta_1)-l_t(\theta_{10})]\le\frac{2}{n}\sum_{t=1}^{\log n}\max_{\theta_1\in\Theta}|l_t(\theta_1)|=o_p(1).$$

By the previous two inequalities, we can claim that

$$\frac{1}{n}\max_{1\le k\le[n\tilde{\tau}]}\max_{\theta_1\in\Theta}\sum_{t=1}^{k}[l_t(\theta_1)-l_t(\theta_{10})]\le o_p(1). \tag{A.1}$$

By Assumption 2.2 and Lemma 1 in Chow and Teicher (1968), for any $\epsilon>0$, we have

$$\begin{aligned}
&P\left\{\frac{1}{n}\max_{1\le k\le[n\tilde{\tau}]}\max_{\theta_2\in\Theta}\left|\sum_{t=k+1}^{k_0}[l_t(\theta_2)-El_t(\theta_2)]\right|>\epsilon\right\}\\
&\quad=P\left\{\frac{1}{n}\max_{k_0-[n\tilde{\tau}]\le u\le k_0}\max_{\theta_2\in\Theta}\left|\sum_{t=-u}^{-1}[l_t(\theta_2)-El_t(\theta_2)]\right|>\epsilon\right\}\ (\textit{by stationarity})\\
&\quad\to 0.
\end{aligned} \tag{A.2}$$

Denote $\Theta_{2\delta}=\{\theta_2:\|\theta_2-\theta_{20}\|\le\delta\}$ and $C=El_t(\theta_{10})-El_t(\theta_{20})$. Then, $C>0$ when $t\le k_0$ by Assumption 2.1. Since $El_t(\theta_2)$ is a continuous function, we can take a small δ such that

$$\max_{\theta_2\in\Theta_{2\delta}}[El_t(\theta_2)-El_t(\theta_{10})]=-C+\max_{\theta_2\in\Theta_{2\delta}}[El_t(\theta_2)-El_t(\theta_{20})]\le -C/2.$$

Furthermore, by (8.2), we have

$$\begin{aligned}
&\frac{1}{n}\max_{1\le k\le[n\tilde{\tau}]}\max_{\theta_2\in\Theta_{2\delta}}\sum_{t=k+1}^{k_0}[l_t(\theta_2)-l_t(\theta_{10})]\\
&\quad\le\frac{2}{n}\max_{1\le k\le[n\tilde{\tau}]}\max_{\theta_2\in\Theta_{2\delta}}\left|\sum_{t=k+1}^{k_0}[l_t(\theta_2)-El_t(\theta_2)]\right|\\
&\qquad+\frac{1}{n}\max_{1\le k\le[n\tilde{\tau}]}(k_0-k)\max_{\theta_2\in\Theta_{2\delta}}[El_t(\theta_2)-El_t(\theta_{10})]\\
&\quad\le-\tilde{\tau}C/2+o_p(1).
\end{aligned} \tag{A.3}$$

Note that $\max_{\theta_2\in\Theta_{2\delta}}[El_t(\theta_2)-El_t(\theta_{20})]=0$ when $t>k_0$. By (6.1), we have

$$\begin{aligned}
&\frac{1}{n}\max_{\theta_2\in\Theta_{2\delta}}\sum_{t=k_0+1}^{n}[l_t(\theta_2)-l_t(\theta_{20})]\\
&\quad\le\frac{2}{n}\max_{\theta_2\in\Theta_{2\delta}}\left|\sum_{t=k_0+1}^{n}[l_t(\theta_2)-El_t(\theta_2)]\right|+\frac{n-k_0}{n}\max_{\theta_2\in\Theta_{2\delta}}[El_t(\theta_2)-El_t(\theta_{20})]\\
&\quad=o_p(1).
\end{aligned} \tag{A.4}$$

Let $\Theta_{2\delta}^c = \Theta - \Theta_{2\delta}$. Since $\theta_{10} \neq \theta_{20}$, we can take a small δ such that $\theta_{10} \in \Theta_{2\delta}^c$. Thus, $\max_{\theta_2 \in \Theta_{2\delta}^c}[El_t(\theta_2) - El_t(\theta_{10})] = 0$ when $t \leq k_0$. Thus, by (8.2), we have

$$\begin{aligned}
&\frac{1}{n} \max_{1 \leq k \leq [n\tilde{\tau}]} \max_{\theta_2 \in \Theta_{2\delta}^c} \sum_{t=k+1}^{k_0} [l_t(\theta_2) - l_t(\theta_{10})] \\
&\quad \leq \frac{2}{n} \max_{1 \leq k \leq [n\tilde{\tau}]} \max_{\theta_2 \in \Theta_{2\delta}^c} \left| \sum_{t=k+1}^{k_0} [l_t(\theta_2) - El_t(\theta_2)] \right| \\
&\qquad + \frac{1}{n} \max_{1 \leq k \leq [n\tilde{\tau}]} (k_0 - k) \max_{\theta_2 \in \Theta_{2\delta}^c} [El_t(\theta_2) - El_t(\theta_{10})] \\
&\quad = o_p(1). \qquad \textbf{(A.5)}
\end{aligned}$$

Denote $C_{1\delta} = -\max_{\theta_2 \in \Theta_{2\delta}^c}[El_t(\theta_2) - El_t(\theta_{20})]$. Then $C_{1\delta} > 0$ when $t > k_0$ by Assumption 2.1. By (6.1), we have

$$\begin{aligned}
&\frac{1}{n} \max_{\theta_2 \in \Theta_{2\delta}^c} \sum_{t=k_0+1}^{n} [l_t(\theta_2) - l_t(\theta_{20})] \\
&\quad \leq \frac{2}{n} \max_{\theta_2 \in \Theta_{2\delta}^c} \left| \sum_{t=k_0+1}^{n} [l_t(\theta_2) - El_t(\theta_2)] \right| + \frac{n - k_0}{n} \max_{\theta_2 \in \Theta_{2\delta}^c} [El_t(\theta_2) - El_t(\theta_{20})] \\
&\quad \leq -(1 - \tau_0) C_{1\delta} + o_p(1). \qquad \textbf{(A.6)}
\end{aligned}$$

By (6.2), (8.1), and (8.3)–(8.6), it follows that

$$\begin{aligned}
&P(\hat{\tau}_n < \tilde{\tau}) \\
&\quad \leq P\left(\frac{1}{n} \max_{1 \leq k \leq [n\tilde{\tau}]} \max_{(\theta_1, \theta_2) \in \Theta^2} \Delta_n(k, \theta_1, \theta_2) \geq 0 \right) \\
&\quad \leq P\left(\frac{1}{n} \max_{1 \leq k \leq [n\tilde{\tau}]} \max_{\theta_2 \in \Theta} \left\{ \sum_{t=k+1}^{k_0} [l_t(\theta_2) - l_t(\theta_{10})] + \sum_{t=k_0+1}^{n} [l_t(\theta_2) - l_t(\theta_{20})] \right\} \right. \\
&\qquad \left. + o_p(1) \geq 0 \right) \\
&\quad \leq P(\max\{-(1 - \tau_0) C_{1\delta}, -\tilde{\tau} C/2\} + o_p(1) \geq 0) \to 0,
\end{aligned}$$

as $n \to \infty$. This completes the proof. ■