Estimating a change point in the long memory parameter

Keiko Yamaguchi

Department of Management Science, Faculty of Engineering, Tokyo University of Science

May 13, 2010

Abstract

We propose an estimator of change point in the long memory parameter d of an ARFIMA(p, d, q) process using the sup Wald test. We derive the consistency and the rate of convergence of the parameter. The convergence rate of our change point estimator depends on the magnitude of a shift. Furthermore, we obtain the limiting distribution of our change point estimator without depending on the distribution of the process. Therefore, we can construct the confidence interval of the change point. Simulations show the validity of the asymptotic theory of our estimator if the sample size is large enough. We apply our change point estimator to the yearly Nile river minimum time series.

Key Words : Break in persistence; long memory; change point

JEL classification : C22

1   Introduction

Long memory processes have been observed in many areas, such as hydrology, telecommunications, economics, and finance. The ARFIMA(p, d, q) model is one of the most widely used to model such dependence (Granger and Joyeux (1980); Hosking (1981)). This is a natural extension of the ARIMA (p, d, q) model and allows the parameter d to be a real number. The d parameter governs the amount of persistence in the process and is called a stationary long memory parameter when $d \in (0, 1/2)$. Because of its properties, the long memory model is often applied to long data series. In such long series, the chance of structural change will be greater with a change in the physical mechanism that generates the data or with a change in the way that observations are collected over time.

Modeling with taking changes in d into account is important, because even small changes in d may have a strong impact on statistical inferences such as the convergence rate of confidence

intervals for constants and long-term prediction. Beran and Terrin (1996) point out the existence of a change in d for some time series. To model changes with unknown change points, which is often the case in practice, we often use a two-step procedure. First, we test for the constancy of parameters. Then we estimate a change point, if the null hypothesis is rejected. This procedure is considered in Bai (1997) in the context of multiple regression models. Concerning the test for the constancy of the long memory parameter, there are several studies in the ARFIMA(p, d, q) model with an unknown change point. For example, Beran and Terrin (1996), Horvath and Shao (1999), and Ling (2007) consider the sup Wald test, which is one of the most common tests to deal with an unknown change point; however, they only consider the null limiting distribution. Horvath (2001) derives the consistency of the sup Wald test.

In contract, the literature contains little research about the estimation of change point in -changes in d-. One exception is Ray and Tsay (2002), who adopt the Bayesian approach. However, the estimation of the change point is considered important in order to build an accurate model and forecast properly. Furthermore, this enables us to uncover the underlying factors that fostered the changes.

In this paper, we propose an estimator of change point in d of the ARFIMA process using the sup Wald test. We derive the consistency and the rate of convergence of the parameter, including our change point estimator. The convergence rate of our change point estimator depends on the magnitude of a shift. This coincides with our intuition. Moreover, we obtain the limiting distribution of our change point estimator without depending on the distribution of the process. The limiting distribution is the same in previous research, and its density function is known. Therefore, we can construct the confidence interval (CI) of the change point.

The rest of the paper is organized as follows. After introducing the model and estimation method in Section 2, Section 3 describes the break fraction estimator and derives the asymptotic properties of our estimator. Section 4 contains the simulation results, and an empirical example using the same dataset as was used in Beran and Terrin (1996). The proofs of the results stated in the text are presented in Sections 5 and 6.

2 Model and estimation method

Consider the following ARFIMA(p, d, q) model with a change point k_0:

$$
y_t = \begin{cases} y_t^{(1)} = (1-L)^{-d_{10}}\phi^{-1}(L)\psi(L)\varepsilon_t, & t = 1, \ldots, k_0 \\ y_t^{(2)} = (1-L)^{-d_{20}}\phi^{-1}(L)\psi(L)\varepsilon_t, & t = k_0 + 1, \ldots, T \end{cases} \tag{1}
$$

where $\{\varepsilon_t\} \sim IID(0, \sigma^2)$, $\phi(L) = 1 - \sum_{i=1}^{p}\phi_i L^i$, and $\psi(L) = 1 - \sum_{i=1}^{q}\psi_i L^i$; L is the Lag operator, $0 < d < 0.5$, $\phi(z) \neq 0$ and $\psi(z) \neq 0$ for all z such that $|z| \leq 1$, $\phi_p \neq 0$, $\psi_q \neq 0$, and $\psi(z)$ and $\psi(z)$ have no common root. Denote $\eta = (\phi_1, \ldots, \phi_p, \psi_1, \ldots, \psi_q)'$, $\theta_{10} = (d_{10}, \eta')'$, $\theta_{20} = (d_{20}, \eta')'$, and $m = p + q + 1$. Assume θ_{10} and θ_{20} are unknown. The parameter space Θ is a compact subset of R^m, as in Ling (2007). We let $\tau_0 = k_0/T$ be the true break fraction, which is unknown, and assume $0 < \underline{\tau} \leq \tau_0 \leq \bar{\tau} < 1$.

For any given $k \in [\underline{\tau}T, \bar{\tau}T]$, the conditional-sum-of-squares (CSS) estimators of θ_{10} and θ_{20} are given by

$$
\hat{\theta}_1(k/T) = \arg\min_{\Theta} \frac{1}{T}\sum_{t=1}^{k}\left[(1-L)^{d_1}\phi(L)\psi^{-1}(L)(y_t I\{t > 0\})\right]^2 = \arg\min_{\Theta} \frac{1}{T}\sum_{t=1}^{k}e_t^2(\theta_1),
$$
$$
\hat{\theta}_2(k/T) = \arg\min_{\Theta} \frac{1}{T}\sum_{t=k+1}^{T}\left[(1-L)^{d_2}\phi(L)\psi^{-1}(L)(y_t I\{t > 0\})\right]^2 = \arg\min_{\Theta} \frac{1}{T}\sum_{t=k+1}^{T}e_t^2(\theta_2),
$$

$$(2)$$

where $I\{\cdot\}$ is an indicator function. Thus, the CSS estimation is a nonlinear least square estimation substituting zero(=mean) into y_t for $t < 0$.

Let denote $\tilde{D}_t(\theta) = \frac{\partial}{\partial\theta}e_t^2(\theta)$, $\tilde{P}_t(\theta) = -\frac{\partial^2}{\partial\theta\partial\theta'}e_t^2(\theta)$, $\Sigma_{1\tau}(\theta) = \sum_{t=1}^{k}\tilde{P}_t(\theta)$, $\Sigma_{2\tau}(\theta) = \sum_{t=k+1}^{T}\tilde{P}_t(\theta)$ and

$$
\hat{\Sigma}_T(k) = \Sigma_{1\tau}(\hat{\theta}_1(k/T)) + \Sigma_{2\tau}(\hat{\theta}_2(k/T)),
$$

$$
\hat{\Omega}_T(k) = \sum_{t=1}^{k}\tilde{D}_t(\hat{\theta}_1(k/T))\tilde{D}_t'(\hat{\theta}_1(k/T)) + \sum_{t=k+1}^{T}\tilde{D}_t(\hat{\theta}_2(k/T))\tilde{D}_t'(\hat{\theta}_2(k/T)).
$$

Let $z_1(\theta) = (1-L)^{d_1}\phi(L)\psi^{-1}(L)y_1$, which is similar to $e_1(\theta)$, but without the indicator function. Let $D_1(\theta) = \frac{\partial}{\partial\theta}z_1^2(\theta)$, $P_1(\theta) = -\frac{\partial^2}{\partial\theta\partial\theta'}z_1^2(\theta)$, $\Sigma = E[P_1(\theta_{10})]$, and $\Omega = E[D_1(\theta_{10})D_1'(\theta_{10})]$. Since Ling (2007) considers more general models, he distinguishes Ω from Σ. In this paper, however, we let $\Omega = \Sigma$ and be equivalent to the information matrix, because the CSS estimator is an approximate Maximum Likelihood (ML) estimator (see Robinson (2006)).

3 Estimation of the break fraction

3.1 Estimation of the break fraction

In this subsection, we propose a change point estimator based on the Wald test for constancy of d. By treating each $k \in [\underline{\tau}T, \bar{\tau}T]$ as a hypothetical change point, the corresponding Wald statistics is defined as:

$$W_T(k/T) = \frac{k(T-k)}{T}\hat{\rho}^{-1}[\hat{d}_1(k/T) - \hat{d}_2(k/T)]^2,$$

where $\hat{\rho}$ is the first element in the first row of $(\hat{\Sigma}_T(k)/T)^{-1}$. We then define the change point $\hat{k}$ as a global maximizer of Wald statistics, i.e.:

$$\hat{k} = \operatorname*{Arg\,max}_{k} W_T(k/T),$$

and denote the break fraction estimator as $\hat{\tau} = \hat{k}/T$. In addition, we define $\hat{\theta}_1(\hat{k}/T)$ and $\hat{\theta}_2(\hat{k}/T)$ as the estimators of θ_{10} and θ_{20} as corresponding to $\hat{k}$.

Now, more specifically, we can see that our change point estimator is related to the sup Wald test for the null hypothesis of no structural change versus the alternative hypothesis that a change occurs at some unknown date. Indeed, the sup Wald test statistics proposed in Beran and Terrin (1996), and Horvath and Shao (1999) is identical to sup $W_T(k/T)$ $(= W_T(\hat{k}/T))$, except that $\hat{d}$ is estimated by minimizing the Whittle likelihood. Note that it is less important to consider the difference between the Whittle and CSS estimations here, because both are approximate ML estimations.

The idea of estimating change point using sup Wald statistics is suggested in Bai (1997) in the context of multiple regression models. As mentioned in Bai (1997), it is useful that a break point estimator is obtained automatically, combining the testing and estimation in a single step.

3.2 Asymptotic properties

We introduce an assumption about the magnitude of shift. We consider a small shift because we can detect a big change by plotting the data.

Assumption 3.1. $|d_{10} - d_{20}| = \delta\nu_T$, where $\nu_T > 0$, $\nu_T \to 0$ and $T^{1/2-\alpha}\nu_T \to \infty$ for some $\alpha \in (0, 1/2)$ and $\delta \neq 0$.

This assumption is parallel to the "-shrinking shift-" in Bai (1997) and Bai and Perron (1998). In the above assumption, the magnitude of the shift shrinks to zero as the sample size increases, but the shift cannot be too small to identify the break fraction. This assumption allows us to obtain the limiting distribution of $\hat{\tau}$ invariant to $\{\varepsilon_t\}$, as in Theorem 3.2, whereas the limiting distribution based on the assumption of a fixed magnitude of shift depends on the exact distribution of $\{\varepsilon_t\}$ (e.g., Hinkley (1970)).

We obtain the consistency and the convergence rate of $\hat{\tau}$ under Assumption 3.1.

Theorem 3.1. *Under Assumption 3.1, we have: (i) $\hat{\tau}_T \to_p \tau_0$; and (ii) for every $\eta > 0$ there exists a $C < \infty$ such that for all large T, $P(|T\nu_T^2(\hat{\tau}_T - \tau_0)| > C) < \eta$.*

Theorem 3.1 shows the convergence rate of $\hat{\tau}$ depends on the shrinking rate ν_T. Although $\hat{\tau}$ is consistent, $|\hat{k} - k|$ diverges. This enables us to apply Functional Central Limit Theorem and obtain a limit distribution not depending on $\{\varepsilon_t\}$ provided in Theorem 3.2 below.

Now define $B_i(r)$, $i = 1, 2$ as two independent standard Brownian motions defined on R_+. Let define $B^*(r) = B_1(-r)$ for $r < 0$ and $B^*(r) = B_2(r)$ for $r \geq 0$.

Theorem 3.2. *Under Assumption 3.1, we have*

$$\rho^{-1}T(d_{20} - d_{10})^2(\hat{\tau} - \tau_0) = \rho^{-1}(d_{20} - d_{10})^2(\hat{k} - k_0) \to^d Arg\ max\left(B^*(r) - \frac{1}{2}|r|\right). \qquad (3)$$

Similar results are obtained for iid models (e.g. Yao (1987)), multi regression models (e.g. Bai (1997)) and so on. The density function of the limiting distribution is given by

$$3/2e^{|x|}\Phi(-3\sqrt{|x|}/2) - 1/2\Phi(-\sqrt{|x|}/2),$$

where $\Phi(\cdot)$ is standard normal distribution (see e.g. Bai (1997)). For example, the 95% and 97.5% quantiles are 7.7 and 11.0. Note that the distribution is symmetric about the origin.

Then, all we need to construct the CI of $\hat{k}$ are the asymptotic properties of $\hat{L} = \hat{\rho}^{-1}(\hat{d}_2 - \hat{d}_1)^2$. The following corollary gives the consistency of the remaining parameters using $\hat{\tau}$.

Corollary 3.1. *Under Assumption 3.1, $\hat{\theta}_1(\hat{\tau}) - \theta_{10} = O_p(T^{-1/2})$, $\hat{\theta}_2(\hat{\tau}) - \theta_{20} = \max\{O_p(T^{-1/2}), O_p(T^{-d_{10}}\nu_T)\}$, and $\frac{1}{T}\hat{\Sigma}_T(\hat{k})$ is a consistent estimate of the information matrix.*

Corollary 3.1 implies that $(\hat{L} - L)(\hat{k} - k_0) \to^p 0$. Therefore, we can construct the CI for the change point.

4 Simulation and empirical example

4.1 Simulation

We investigate the finite-sample properties of the change point estimator $\hat{k}$ (and hence $\hat{\tau}$). We consider the following data generating process:

$$(1 - L)^{d_1} y_t = \varepsilon_t \quad t = 1, \ldots, k_0,$$
$$(1 - L)^{d_2} y_t = \varepsilon_t \quad t = k_0 + 1, \ldots, T, \tag{4}$$

where $\{\varepsilon_t\} \sim NID(0, 1)$. We set $d_1 = 0.1$ and $d_2 = 0.2, 0.3, 0.4$, $k_0 = \tau_0 T$, $\tau_0 = 0.5$, and $T = 400, 800$. The number of replications is $1,000$ throughout the simulations. We search $\hat{k}$ among $\{0.15T, \ldots, 0.85T\}$.
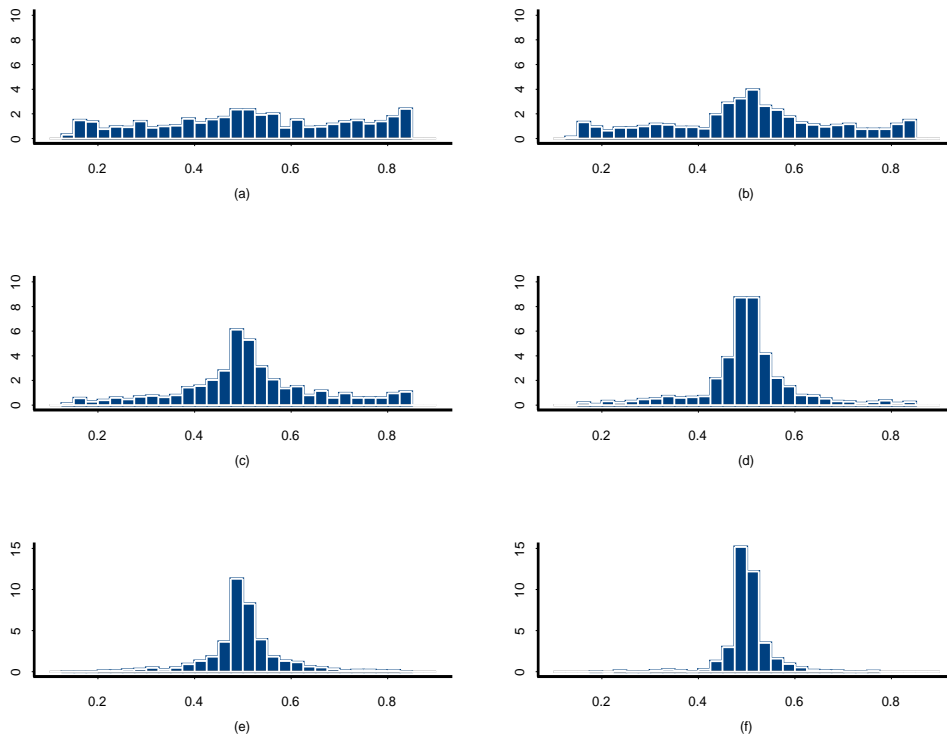

Figure 1: Histogram of $\hat{\tau}$ when $\tau_0 = 0.5$ and $d_1 = 0.1$ for selected values d_1, d_2 and T: (a) $d_2 = 0.2, T = 400$; (b) $d_2 = 0.2, T = 800$; (c) $d_2 = 0.3, T = 400$; (d) $d_2 = 0.3, T = 800$; (e) $d_2 = 0.4, T = 400$; (f) $d_2 = 0.4, T = 800$.

Figure 1 presents histograms of $\hat{\tau}$ in all cases. We can see that $\hat{\tau}$ is more concentrated at the true break fraction τ_0 as T. This confirms the consistency. The precisions of $\hat{\tau}$ are worse for the DGP with $d_1 = 0.1$, $d_2 = 0.2$ than for other cases. This is because the rate of convergence is propotional to the square of the magnitude of the shift, as in Theorem 3.2. Moreover, in most cases, histograms seem symmetric as predicted by Theorem 3.2.
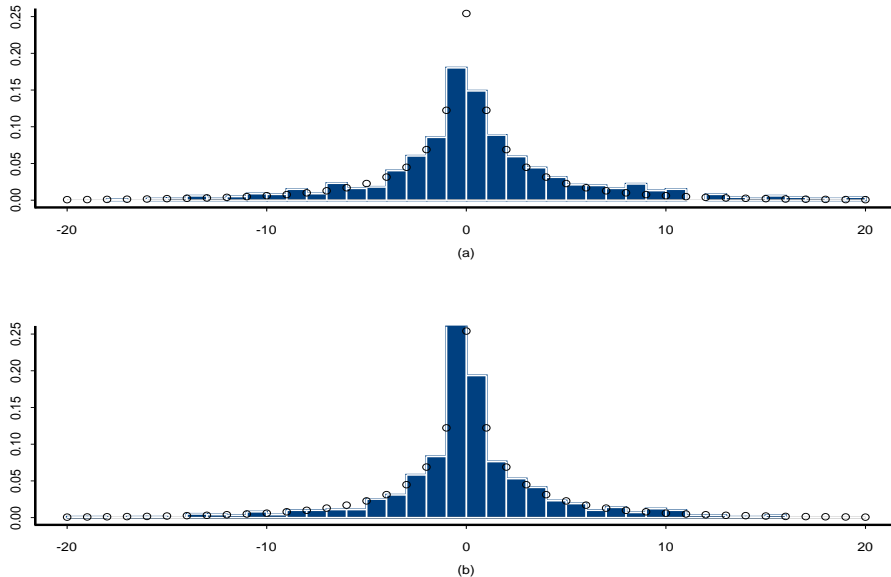

Figure 2: Distribution of $\rho^{-1}(\hat{d}_{20} - \hat{d}_{10})^2 T(\hat{\tau} - \tau_0)$ when $\tau_0 = 0.5$, $T = 400$: (a) $d_1 = 0.1$, $d_2 = 0.3$; (b) $d_1 = 0.1$, $d_2 = 0.4$. The circle plot represents the asymptotic distribution.

Figure 2 shows histograms of the normalized break fraction estimator $\hat{\rho}^{-1} T(\hat{d}_2 - \hat{d}_1)^2(\hat{\tau} - \tau_0)$ for $d_2 = 0.3, 0.4$ with a sample size of $T = 400$. In addition, Figure 2 plots the corresponding limiting distribution given by Theorem 3.2. It seems that the sample size $T = 400$ is sufficient to apply the asymptotic theory, as long as the magnitude of the shift is not too small.

4.2   Empirical example

We apply the proposed change point estimator to the yearly Nile river minimum time series, which is known to follow a long memory process (e.g., Beran (1994)). This Nile river data spans from 622 AD to 1284 AD, implying 663 observations. Beran and Terrin (1996) find a change in d in this data using their sup Wald test. In addition, they compute Wald statistics at intervals of 20 years, and find a change around 722 AD. Ray and Tsay (2002) find a change

in d at year 722 AD using a Bayesian method. They also point out that this change may be caused by the introduction of a new type of device for measurements around the year 715 AD. Using our method, we can construct a CI to consider inference about a change point from a classical viewpoint. We adopt the ARFIMA$(0, d, 0)$ model, the same as in Beran and Terrin (1996). We calculate Wald statistics $W_T(k/T)$ for every k, not at intervals of 20 as in Beran and Terrin (1996). The maximum of $W_T(k/T)$ is attained at 732 AD. The estimates $\hat{d}(\hat{\tau})$ before and after this change point are 0.0088 and 0.4631. Then, the estimated scale factor $\hat{L} = \pi^2/6(0.4631 - 0.0088)^2 = 0.3395$. Thus, the upper boundary of the 90% confidence interval will be 732AD + [11.0/0.3395] +1 = 755AD. Based on symmetry, the confidence interval is [709AD, 755AD]. It contains 715AD. This CI contains the year of the introduction of a new measurement device, corroborating the conjecture of Ray and Tsay (2002).

5   Appendix 1: Proofs

5.1   Proof of Theorem 3.1 (i)

We shall show plim $\hat{\tau}_T = \tau_0$ using the contradiction argument such as in Bai and Perron (1998) and Chong (2001). Suppose $\hat{\tau}_T$ is not consistent. Then, with some positive probability, there exists $\epsilon > 0$, such that $\hat{\tau}_T \notin V_\epsilon$, $V_\epsilon = \{\tau; |\tau - \tau_0| \leq \epsilon\}$.

Let $\Upsilon_{1\epsilon} = \{\tau; \underline{\tau} \leq \tau < \tau_0 - \epsilon\}$ and $\Upsilon_{2\epsilon} = \{\tau; \tau_0 + \epsilon < \tau \leq \bar{\tau}\}$. Observe that

$$
\begin{aligned}
P(|\hat{\tau}_T - \tau_0| > \epsilon) &= P\left(\sup_{\tau \in \Upsilon_{1\epsilon} \cup \Upsilon_{2\epsilon}} W_T(\tau) > \sup_{\tau \in V_\epsilon} W_T(\tau) \right) \\
&\leq \sum_{i=1}^{2} P\left(\sup_{\tau \in \Upsilon_{i\epsilon}} [W_T(\tau) - W_T(\tau_0)] > 0 \right) \\
&= \sum_{i=1}^{2} P\left(\sup_{\tau \in \Upsilon_{i\epsilon}} \left[\frac{W_T(\tau) - W_T(\tau_0)}{T(d_{20} - d_{10})^2} \right] > 0 \right)
\end{aligned}
\tag{5}
$$

where the inequality follows from $\tau_0 \in V_\epsilon$ and the last equation follows from $(d_{20} - d_{10})^2 > 0$. If we show that

$$
\frac{W_T(\tau_0)}{T(d_{20} - d_{10})^2} - \tau_0(1 - \tau_0)\rho^{-1} = o_p(1),
\tag{6}
$$

$$
\sup_{\tau \in \Upsilon_{1\epsilon}} \left| \frac{W_T(\tau)}{T(d_{20} - d_{10})^2} - \tau \frac{(1 - \tau_0)^2}{(1 - \tau)} \rho^{-1} \right| = o_p(1),
\tag{7}
$$

and

$$\sup_{\tau \in \Upsilon_{2\epsilon}} \left| \frac{W_T(\tau)}{T(d_{20} - d_{10})^2} - \frac{\tau_0^2}{\tau}(1 - \tau)\rho^{-1} \right| = o_p(1), \tag{8}$$

then the right-hand side of (5) tends to 0 as $T \to \infty$, because

$$P\left(\sup_{\tau \in \Upsilon_{1\epsilon}} \left[\frac{W_T(\tau) - W_T(\tau_0)}{T(d_{20} - d_{10})^2} \right] > 0 \right) = P\left(\sup_{\tau \in \Upsilon_{1\epsilon}} \frac{(1 - \tau_0)}{(1 - \tau)}(\tau - \tau_0)\rho^{-1} + o_p(1) > 0 \right) \to 0,$$

and

$$P\left(\sup_{\tau \in \Upsilon_{2\epsilon}} \left[\frac{W_T(\tau) - W_T(\tau_0)}{T(d_{20} - d_{10})^2} \right] > 0 \right) = P\left(\sup_{\tau \in \Upsilon_{2\epsilon}} \frac{\tau_0}{\tau}(\tau_0 - \tau)\rho^{-1} + o_p(1) > 0 \right) \to 0.$$

Therefore, $\hat{\tau}_T$ is consistent.

We proceed to show (6)–(8). We will show the uniform convergence of $W_T(\tau)$ using the limiting theorems for near-epoch-dependent (NED) sequences, as in Ling (2007).

Define $Z(\tau) = T^{1/2}\tau(1 - \tau)\{\hat{d}_1(\tau) - \hat{d}_2(\tau)\} = T^{1/2}\tau(1 - \tau)\iota'\{\hat{\theta}_1(\tau) - \hat{\theta}_2(\tau)\}$, where ι is an $m \times 1$ vector whose first element is one and all other elements are zero. We can then rewrite $W_T(\tau)$ as

$$W_T(\tau) = \hat{\rho}^{-1}[(\tau(1 - \tau))^{-1/2}Z(\tau)']^2.$$

Thus, we need to examine the asymptotics of $\hat{\Sigma}_T(k)$, $Z(\tau_0)$, and $\sup_{\tau \in \Upsilon_{i\epsilon}} Z(\tau)$ for $i = 1, 2$. Note that $0 < \tau(1 - \tau) < \infty$, because we assume that $0 < \underline{\tau} \le \tau \le \bar{\tau} < 1$.

First, we consider $\sup_{\tau \in \Upsilon_{1\epsilon}} Z(\tau)$. Rewrite $Z(\tau)$ as

$$Z(\tau) = T^{1/2}\tau(1 - \tau)\iota' \left\{ \left(\hat{\theta}_1(\tau) - \theta_{10} \right) - \left(\hat{\theta}_2(\tau) - \theta_{20} \right) - (\theta_{20} - \theta_{10}) \right\}.$$

For $\hat{\theta}_1$, it is identical to $\hat{\lambda}$ in Ling (2007). Thus $\hat{\theta}_1(\tau)$ converges to θ_{10} uniformly in $\tau \in \Upsilon_{1\epsilon}$ and we can then expand it as following equation:

$$\hat{\theta}_1(\tau) - \theta_{10} = \left(\Sigma^{-1} + o_p(1) \right) \frac{1}{k} \left\{ \sum_{t=1}^{k} \frac{\partial e_t(\theta_{10})}{\partial \theta} e_t(\theta_{10}) \right\}. \tag{9}$$

For $\hat{\theta}_2(\tau)$, Lemma 6.3 provides that $\hat{\theta}_2(\tau)$ converges to θ_{20} uniformly in τ. Thus,

$$\hat{\theta}_2(\tau) - \theta_{20} = \left(\frac{1}{T - k}\hat{\Sigma}_{2\tau}(\tilde{\theta}) \right)^{-1} \frac{1}{T - k} \left\{ \sum_{t=k+1}^{k_0} \frac{\partial e_t(\theta_{20})}{\partial \theta} e_t(\theta_{20}) + \sum_{t=k_0+1}^{T} \frac{\partial e_t(\theta_{20})}{\partial \theta} e_t(\theta_{20}) \right\}, \tag{10}$$

where $|\tilde{\theta} - \theta_{20}| \le |\hat{\theta}_2(\tau) - \theta_{20}|$. From Lemma 6.6, we can rewrite

$$
\hat{\theta}_2(\tau) - \theta_{20} = \frac{\tau_0 - \tau}{1 - \tau} \left((\theta_{20} - \theta_{10}) + o_p(\nu_T) \right)
$$
$$
+ \left(\frac{1}{T-k} \hat{\Sigma}_{2\tau}(\tilde{\theta}) \right) \frac{1}{T-k} \left\{ \sum_{t=k+1}^{k_0} \frac{\partial e_t(\theta_{10})}{\partial \theta} e_t(\theta_{10}) + \sum_{t=k_0+1}^{T} \frac{\partial e_t(\theta_{20})}{\partial d} e_t(\theta_{20}) \right\}. \tag{11}
$$

From Lemma 6.5 and Lemma 6.2 in Ling (2007), $\hat{\Sigma}_{2\tau}/(T-k)$ converges to Σ uniformly in probability in $\tau \in \Upsilon_{1\epsilon}$. From Lemma 6.6, (11) holds uniformly in τ.

Therefore, the following equation holds uniformly in $\tau \in \Upsilon_{1\epsilon}$:

$$
Z(\tau) = T^{1/2}\tau(1-\tau)\iota' \left\{ \left(\hat{\theta}_1(\tau) - \theta_{10} \right) - \left(\hat{\theta}_2(\tau) - \theta_{20} \right) - (\theta_{20} - \theta_{10}) \right\}
$$
$$
= -T^{1/2}\tau(1-\tau_0)(d_{20} - d_{10}) + T^{-1/2}\iota'\Sigma^{-1}\{A_1(\tau) + \tau A_0\} + o_p(T^{1/2}\nu_T), \tag{12}
$$

where

$$
A_0 = \left\{ \sum_{t=1}^{k_0} \frac{\partial e_t(\theta_{10})}{\partial \theta} e_t(\theta_{10}) + \sum_{t=k_0+1}^{T} \frac{\partial e_t(\theta_{20})}{\partial \theta} e_t(\theta_{20}) \right\},
$$

and

$$
A_1(\tau) = \sum_{t=1}^{k} \frac{\partial e_t(\theta_{10})}{\partial \theta} e_t(\theta_{10}).
$$

From Lemma 6.7, $A_1(\tau)$ is $O_p(T^{1/2})$ uniformly in τ. Because A_0 does not depend on τ, we can prove A_0 is $\max\{O_p(T^{1/2}), O_p(T^{1-d_{20}}(\log T)\nu_T)\}$ more easily from Lemma 6.4. Therefore, we establish (7).

Next, we consider (6). From Robinson (2006), $|\hat{d}_1(\tau_0) - d_{10}| = O_p(k_0^{-1/2})$ and from Robinson (2006) and Lemma 6.4 and 6.6, $|\hat{d}_2(\tau_0) - d_{20}|$ is $\max\{O_p((T - k_0)^{-1/2}), O_p(T^{-d_{10}}\nu_T)\}$. Hence,

$$
Z(\tau_0) = T^{1/2}\tau_0(1-\tau_0) \left\{ \left(\hat{d}_1(\tau_0) - d_{10} \right) - \left(\hat{d}_2(\tau_0) - d_{20} \right) - (d_{20} - d_{10}) \right\}
$$
$$
= O_p(1) + O_p(T^{1/2-d_{10}}\nu_T) + T^{1/2}\tau_0(1-\tau_0)(d_{20} - d_{10})
$$
$$
= T^{1/2}\tau_0(1-\tau_0)(d_{20} - d_{10}) + o_p(T^{1/2}\nu_T).
$$

Thus, from Assumption 3.1, we have (6). $\square$

5.2 Proof of Theorem 3.1 (ii)

We shall show $|\hat{\tau}_T - \tau_0| = O_p(\frac{1}{T\nu_T^2})$ using the contradiction argument, such as in Appendix C of Chong (2001). Remember that $\hat{\tau}_T \to_p \tau_0$. Suppose $\hat{\tau}$ is not $T\nu_T^2$-consistent, then there exists a sequence $M_T > 0$ such that $M_T \to \infty$ and $M_T/(T\nu_T^2) \to 0$ as $T \to \infty$, and

$$\liminf_{T\to\infty} P\left(|\hat{\tau}_T - \tau_0| > \frac{M_T}{T\nu_T^2}\right) = \alpha, \tag{13}$$

where α is a positive constant in $\alpha \in (0, 1]$. Note that

$$P\left(|\hat{\tau}_T - \tau_0| > \frac{M_T}{T\nu_T^2}\right) = P\left(\sup_{\tau\in\Upsilon_{1M_T}\cup\Upsilon_{2M_T}} W_T(\tau) > \sup_{\tau\in\Upsilon_{3M_T}} W_T(\tau)\right), \tag{14}$$

where $\Upsilon_{1M_T} = \{\tau; \underline{\tau} \le \tau < \tau_0 - M_T/(T\nu_T^2)\}$, $\Upsilon_{2M_T} = \{\tau; \tau_0 + M_T/(T\nu_T^2) < \tau \le \bar{\tau}\}$, and $\Upsilon_{3M_T} = \{\tau; |\tau - \tau_0| < M_T/(T\nu_T^2)\}$. Since $\tau_0 \in \Upsilon_{3M_T}$, the right-hand side of (14) is bounded by

$$P\left(\sup_{\tau\in\Upsilon_{1M_T}\cup\Upsilon_{2M_T}} W_T(\tau) > W_T(\tau_0)\right) \le \sum_{i=1}^{2} P\left(\sup_{\tau\in\Upsilon_{iM_T}} [W_T(\tau) - W_T(\tau_0)] > 0\right). \tag{15}$$

Recall that

$$W(\tau) = T\tau(1-\tau)\left[\iota'\left\{\left(\hat{\theta}_1(\tau) - \theta_{10}\right) - \left(\hat{\theta}_2(\tau) - \theta_{20}\right)\right\} - (d_{20} - d_{10})\right]^2, \tag{16}$$

and

$$W(\tau_0) = T\tau_0(1-\tau_0)\left[\iota'\left\{\left(\hat{\theta}_1(\tau_0) - \theta_{10}\right) - \left(\hat{\theta}_2(\tau_0) - \theta_{20}\right)\right\} - (d_{20} - d_{10})\right]^2. \tag{17}$$

The right-hand side of (16) is

$$T\tau\frac{(1-\tau_0)^2}{1-\tau}(d_{20} - d_{10})^2 - 2\frac{1-\tau_0}{1-\tau}\iota'\{A_1(\tau) + \tau A_0\}(d_{20} - d_{10})$$
$$+ T^{-1}\frac{1}{\tau(1-\tau)}(\iota'\{A_1(\tau) + \tau A_0\})^2 + o_p\left(\frac{\tau_0 - \tau}{1-\tau}T\nu_T^2\right),$$

and the right-hand side of (17) is

$$T\tau_0(1-\tau_0)(d_{20} - d_{10})^2 - 2\iota'\{A_1(\tau_0) + \tau_0 A_0\}(d_{20} - d_{10}) + T^{-1}\frac{1}{\tau_0(1-\tau_0)}(\iota'\{A_1(\tau_0) + \tau_0 A_0\})^2.$$

Hence,

$$
\begin{aligned}
W_T(\tau) - W_T(\tau_0) =& T\,(d_{20} - d_{10})^2\,(\tau - \tau_0)\left(1 + \frac{\tau - \tau_0}{1 - \tau}\right) \\
& - 2\iota'\left[\left\{\left(1 + \frac{\tau - \tau_0}{1 - \tau}\right)A_1(\tau) - A_1(\tau_0)\right\} + (\tau - \tau_0)A_0\right](d_{20} - d_{10}) \\
& + T^{-1}\left[\frac{1}{\tau(1 - \tau)}\left(\iota'\{A_1(\tau) + \tau A_0\}\right)^2 - \frac{1}{\tau_0(1 - \tau_0)}\left(\iota'\{A_1(\tau_0) + \tau_0 A_0\}\right)^2\right] + o_p(M_T) \\
=& G_1(\tau) - 2G_2(\tau) + G_3(\tau).
\end{aligned}
$$

$$(18)$$

$G_2(\tau) = O_p(\sqrt{M_T})$ and $G_3(\tau) = O_p(1)$ uniformly in τ respectively, because

$$
A_1(\tau) = O_p(T^{1/2}), \quad A_1(\tau) - A_1(\tau_0) = O_p\left(\{T(\tau - \tau_0)\}^{1/2}\right) \quad \text{uniformly in } \tau, \qquad (19)
$$

and

$$
A_0 = \max\{O_p(T^{1/2}), O_p(T^{1-d_{20}}(\log T)\nu_T)\}. \qquad (20)
$$

Since $G_1(\tau) \sim M_T$, $\sup_{\tau \in \Upsilon_{1M_T}} W_T(\tau) - W_T(\tau_0)$ is dominated by $G_1(\tau)$. Then $G_1(\tau) \to -\infty$, because $-1 < \tau - \tau_0 < 0$ and $M_T \to \infty$. Thus,

$$
P\left(\sup_{\tau \in \Upsilon_{1M_T}}\,[W_T(\tau) - W_T(\tau_0)] > 0\right) \to 0. \qquad (21)
$$

Similarly, we can derive

$$
P\left(\sup_{\tau \in \Upsilon_{2M_T}}\,[W_T(\tau) - W_T(\tau_0)] > 0\right) \to 0. \qquad (22)
$$

From (21) and (22), we have (15) $\to 0$, but this contradicts $\alpha > 0$. Therefore, $|\hat{\tau}_T - \tau_0| = O_p(\frac{1}{T\nu_T^2})$. $\square$

5.3 Proof of Theorem 3.2

Without loss of generality, we assume $d_{20} - d_{10} > 0$. Because Theorem 3.1 (ii) holds,

$$
W_T(\tau) - W_T(\tau_0) = G_1(\tau) - 2G_2(\tau) + G_3(\tau) + o_p(1). \qquad (23)
$$

We will evaluate $G_i(\tau)$, $i = 1, 2, 3$ respectively in this case. From (19) and (20),

$$G_1(\tau) = -T(d_{20} - d_{10})^2(\tau_0 - \tau)(1 + o_p(1)) = -|s|\delta^2 + o_p(1),$$

$$G_2(\tau) = -2\iota'[((1 + o_p(1))A_1(\tau) - A_1(\tau_0)) + o_p(1)A_0](d_{20} - d_{10})$$

$$= 2\nu_T \iota' \sum_{t=k+1}^{k_0} \frac{\partial e_t(\theta_{10})}{\partial \theta} e_t(\theta_{10}) \, \delta + o_p(1)$$

$$\Longrightarrow 2B(-s),$$

and

$$
\begin{aligned}
G_3(\tau) = & T^{-1}(\tau(1 - \tau)\tau_0(1 - \tau_0))^{-1}(A_1(\tau)^2 - A_1(\tau_0)^2) \\
& + 2T^{-1}((1 - \tau)(1 - \tau_0))^{-1}A_0((1 - \tau_0)A_1(\tau) - (1 - \tau)A_1(\tau_0)) \\
& + T^{-1}((1 - \tau)(1 - \tau_0))^{-1}(\tau - \tau_0)A_0^2 \\
= & O_p(T^{-1/2}\nu_T^{-1}) = o_p(1).
\end{aligned}
\tag{24}
$$

(23) is dominated by G_1 and G_2. Thus, the result of Theorem 3.2 follows. $\qquad\square$

5.4   Proof of Corollary 3.1

Without loss of generality, we assume $\hat\tau < \tau_0$. Then we already show that $\frac{1}{T}\hat\Sigma_T(k)$ converges to Σ uniformly in k in the proof of Theorem 3.1 (i). From the proof of Theorem 3.1 (i), we know that $\hat\theta_1(\tau_0) - \theta_{10} = O_p(T^{-1/2})$, $\hat\theta_2(\tau_0) - \theta_{20} = \max\{O_p(T^{-1/2}), O_p(T^{-d_{10}}\nu_T)\}$. Thus it suffices to show that $\hat\theta_1(\hat\tau) - \hat\theta_1(\tau_0) = o_p(1/\sqrt{T})$, $\hat\theta_2(\tau_0) - \theta_{20} = O_p(T^{-1/2})$. From the definition of $A_1(\tau)$,

$$\hat\theta_1(\tau_0) - \theta_{10} = \frac{1}{T\tau_0}A_1(\tau_0),$$

and

$$\hat\theta_1(\hat\tau) - \theta_{10} = \frac{1}{T\hat\tau}A_1(\hat\tau) = \left(\frac{1}{T\tau_0} + \frac{1}{T\hat\tau} - \frac{1}{T\tau_0}\right)A_1(\hat\tau).$$

Then,

$$\hat\theta_1(\tau_0) - \hat\theta_1(\hat\tau) = \frac{1}{T\tau_0}\sum_{T\tau}^{T\tau_0} \frac{\partial e_t(\theta_{10})}{\partial d}e_t(\theta_{10}) - \frac{\tau_0 - \tau}{T\tau_0\hat\tau}\sum_{1}^{T\tau} \frac{\partial e_t(\theta_{10})}{\partial \theta}e_t(\theta_{10})$$

$$= O_p(T^{-1/2})O_p(T^{-1/2}\nu_T) + O_p(T^{-1/2})O_p(T^{-1}\nu_T^2) = o_p(T^{-1/2}).$$

We can obtain $\hat\theta_2(\tau_0) - \theta_{20}$ similarly. Therefore, we obtain the result. $\qquad\square$

6   Appendix 2: Auxiliary results

Lemma 6.1. *Let $w_t(d_{10}) = (1-L)^{-d_{10}}\phi^{-1}(L)\psi(L)\varepsilon_t$, and $w_t(d_{20}) = (1-L)^{-d_{20}}\phi^{-1}(L)\psi(L)\varepsilon_t$. Under Assumption 3.1, we have*

$$w_t(d_{10}) = w_t(d_{20}) + O_p(\nu_T).$$

Proof Define $\xi_t = \phi^{-1}(L)\psi(L)\varepsilon_t = \sum_{j=0}^{\infty}\pi_j\varepsilon_{t-j}$, and $(1-L)^d = \sum_{j=0}^{\infty}\kappa_j(d)L^j$. Note that

$$w_t(d_{10}) = w_t(d_{20}) + (d_{10} - d_{20})\sum_{k=0}^{\infty}\frac{1}{k}w_{t-k}(d_{20}) + (d_{10} - d_{20})^2\frac{\partial^2}{\partial d^2}w_t(d_2^*), \qquad (25)$$

where $|d_2^* - d_{20}| \le |d_{10} - d_{20}|$. Define the part of the third term in (25) as $L_t = \frac{\partial^2 w_t}{\partial d^2} = \sum_{h=1}^{\infty}\sum_{l=1}^{\infty}\frac{1}{h}\frac{1}{l}w_{t-l-h}$. We can then rewrite

$$L_t = \sum_{h=1}^{\infty}\sum_{j=1}^{h}\sum_{p=1}^{j-1}\frac{1}{p}\frac{1}{j-p}\kappa_{h-j}\xi_{t-h}.$$

Define $\eta_{h-j} = \kappa_{h-j}(d)\sum_{k=1}^{\infty}\pi_k$, then $\eta_m \sim m^{-1-d}$. $L_t = O_p(1)$, because

$$E|L_t|^2 = O(1)\sum_{h=1}^{\infty}\left\{\sum_{j=1}^{h}\sum_{p=1}^{j-1}\frac{1}{p}\frac{1}{j-p}\eta_{h-j}\right\}^2$$

$$= O(1)\sum_{h=1}^{\infty}\left\{\sum_{j=1}^{h}\frac{\log j}{j}\eta_{h-j}\right\}^2$$

$$= O(1)\sum_{h=1}^{\infty}(\log^3 h)h^{2d_2^*-2}$$

$$< \infty.$$

Therefore, the third term in (25) is $o_p(\nu_T)$, giving the stated result.   □

Lemma 6.2. *Define $(1 - L)^d\phi(L)\psi^{-1}(L) = \sum_{j=0}^{\infty}\mu_j(\theta)L^j$. For $t > k_0$, define $v_t(\theta) = \sum_{j=0}^{t-1}\mu_j(\theta)\{(1-L)^{-d_{20}}\xi_{t-j}\}$. Then, under Assumption 3.1, the following results hold:*

(a) $|e_t(\theta) - v_t(\theta)| = O_p((t - k_0)^{-d}\nu_T)$,

(b) $\left|\frac{\partial e_t(\theta)}{\partial\theta} - \frac{\partial v_t(\theta)}{\partial\theta}\right| = O_p((t - k_0)^{-d}\log(t - k_0)\nu_T)$,

(c) $\left|\frac{\partial^2 e_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 v_t(\theta)}{\partial\theta\partial\theta'}\right| = O_p((t - k_0)^{-d}\log^2(t - k_0)\nu_T)$.

Proof For (a), because DGP is (1), for $t > k_0$,

$$e_t(d) = \sum_{j=0}^{t-k_0-1} \mu_j(d) y_{t-j}^{(2)} + \sum_{j=t-k_0}^{t-1} \mu_j(d) y_{t-j}^{(1)}$$

$$= \sum_{j=0}^{t-k_0-1} \mu_j(d)\{(1-L)^{-d_{20}}\xi_{t-j}\} + \sum_{j=t-k_0}^{t-1} \mu_j(d)\{(1-L)^{-d_{10}}\xi_{t-j}\}.$$

From the fact that $\mu_j(d) \sim j^{-1-d}$ and Lemma 6.1,

$$\|e_t(d) - v_t(d)\|_2 \leq \sum_{j=t-k_0}^{t-1} |\mu_j(d)| \|w_t(d_{10}) - w_t(d_{20})\|_2 \tag{26}$$

$$= O_p((t-k_0)^{-d}\nu_T).$$

Next, consider (b) and (c). From the proof of Theorem 4.1 in Ling (2007), for $m = 1, 2$,

$$\frac{\partial^m v_t(d)}{\partial d^m} = \log^m(1-L)(1-L)^d w_t(d_{20}) I\{t > 0\},$$

where $\log^m(1-L)(1-L)^d = \sum_{i=1}^{\infty} a_{mi} L^i$ with $a_{mi}(d) = O(i^{-1-d}\log^m i)$. Since $\underline{d} < d$, proceeding in the same manner as in (a), we obtain the results. $\qquad\square$

Lemma 6.3. *Under Assumption 3.1, $\hat{\theta}_2(\tau)$ converges to θ_{20} in probability uniformly in $\tau \in \Upsilon_{1\epsilon}$.*

Proof Ling (2007) derives the uniform consistency of the CSS estimator, the objective function of which is $\frac{1}{T}\sum_{t=k+1}^{T} u_t^2(d_2)$ with $u_t(d_2) = (1-L)^{d_2}(w_t(d_{20})I\{t > 0\})$. So it suffices to show that

$$\sup_\tau \sup_{d_2} \left| \frac{1}{T}\sum_{t=k+1}^{T} e_t^2(d_2) - \frac{1}{T}\sum_{t=k+1}^{T} u_t^2(d_2) \right| = o_p(1). \tag{27}$$

Note that $u_t \equiv v_t$ for $t > k_0$ from the definition. Now

$$\left| \frac{1}{T}\sum_{t=k+1}^{T} e_t^2(d_2) - \frac{1}{T}\sum_{t=k+1}^{T} u_t^2(d_2) \right| \leq \frac{1}{T}\sum_{t=1}^{k_0} \left| \left[(1-L)^{d_2}(w_t(d_{10})I\{t > 0\})\right]^2 - \left[(1-L)^{d_2}(w_t(d_{20})I\{t > 0\})\right]^2 \right|$$

$$+ \frac{1}{T}\sum_{t=k_0+1}^{T} \left| e_t^2(d_2) - v_t^2(d_2) \right|. \tag{28}$$

For the first term,

$$\frac{1}{T}\sum_{t=1}^{k_0}\left|\left[(1-L)^{d_2}(w_t(d_{10})I\{t>0\})\right]^2 - \left[(1-L)^{d_2}(w_t(d_{20})I\{t>0\})\right]^2\right|$$

$$\leq \frac{1}{T}\sum_{t=1}^{k_0}\left|(1-L)^{d_2}(w_t(d_{10})-w_t(d_{20}))I\{t>0\}\right|\left|(1-L)^{d_2}(w_t(d_{10})+w_t(d_{20}))I\{t>0\}\right|.$$

(29)

From Lemma 6.1 and proceeding similarly to (26), we obtain $\sup_{d_2}\left|(1-L)^{d_2}(w_t(d_{10})-w_t(d_{20}))I\{t>0\}\right| = O_p(\nu_T)$. Therefore, (29) is $o_p(1)$ uniformly in τ. Similarly, we can prove the second term of (28) is $o_p(1)$ uniformly in d; neither does it depend on τ. Therefore, we show (27). $\square$

Lemma 6.4. *Define $v_t(\theta) = \sum_{j=0}^{t-1}\mu_j(\theta)\{(1-L)^{-d_{20}}\xi_{t-j}\}$, as in Lemma 6.2. Then,*

$$\frac{1}{T}\sum_{t=k_0+1}^{T}\frac{\partial e_t(\theta_{20})}{\partial\theta_2}e_t(\theta_{20}) = \frac{1}{T}\sum_{t=k_0+1}^{T}\frac{\partial v_t(\theta_{20})}{\partial\theta_2}v_t(\theta_{20}) + o_p(\nu_T).$$

Proof. From Lemma 6.2,

$$\left|\frac{\partial e_t(\theta)}{\partial\theta_2}e_t(\theta) - \frac{\partial v_t(\theta)}{\partial\theta_2}v_t(\theta)\right| \leq \left|\frac{\partial e_t(\theta)}{\partial\theta_2} - \frac{\partial v_t(\theta)}{\partial\theta_2}\right||e_t(\theta) - v_t(\theta)|$$

$$+ |v_t(\theta)|\left|\frac{\partial e_t(\theta)}{\partial\theta_2} - \frac{\partial v_t(\theta)}{\partial\theta_2}\right| + \left|\frac{\partial v_t(\theta)}{\partial\theta_2}\right||e_t(\theta) - v_t(\theta)|$$

$$= O_p((t-k_0)^{-d}\log(t-k_0)\nu_T).$$

Therefore,

$$\frac{1}{T}\sum_{t=k_0+1}^{T}\left|\frac{\partial e_t(\theta_{20})}{\partial\theta_2}e_t(\theta_{20}) - \frac{\partial v_t(\theta_{20})}{\partial\theta_2}v_t(\theta_{20})\right| = O_p(T^{-d_{20}}(\log T)\nu_T),$$

giving the stated result. $\square$

Lemma 6.5. *Under Assumption 3.1,*

$$\sup_{\tau}\sup_{\theta}\left|\frac{1}{T-k}\hat{\Sigma}_{2\tau}(\theta) - \frac{1}{T-k}\sum_{t=k+1}^{T}\frac{\partial^2}{\partial\theta\partial\theta'}u_t^2(\theta)\right| = o_p(1).$$

16

Proof. From triangular inequality,

$$
\left| \frac{1}{T-k} \hat{\Sigma}_{2\tau}(\theta) - \frac{1}{T-k} \sum_{t=k+1}^{T} \frac{\partial^2}{\partial\theta\partial\theta'} u_t^2(\theta) \right| \leq \frac{1}{T-k} \sum_{t=k+1}^{k_0} \left| \frac{\partial^2}{\partial\theta\partial\theta'} e_t^2(\theta) - \frac{\partial^2}{\partial d^2} u_t^2(\theta) \right|
$$

$$
+ \frac{1}{T-k} \sum_{t=k_0+1}^{T} \left| \frac{\partial^2}{\partial\theta\partial\theta'} e_t^2(\theta) - \frac{\partial^2}{\partial\theta\partial\theta'} v_t^2(\theta) \right|.
$$

For the second term, similarly to Lemma 6.4, from Lemma 6.2 and $\underline{d} < d$, we obtain

$$
\frac{1}{T-k} \sum_{t=k_0+1}^{T} \left| \frac{\partial^2}{\partial\theta\partial\theta'} e_t^2(\theta) - \frac{\partial^2}{\partial\theta\partial\theta'} v_t^2(\theta) \right| \leq \frac{1}{T-k} \sum_{t=k_0+1}^{T} \left| \frac{\partial^2 e_t(\theta)}{\partial\theta\partial\theta'} e_t(\theta) - \frac{\partial^2 v_t(\theta)}{\partial\theta\partial\theta'} v_t(\theta) \right|
$$

$$
= O_p(T^{-\underline{d}} (\log T)^2 \nu_T).
$$

Similarly, we can show that the first term is $O_p(\nu_T)$. □

Lemma 6.6. *For* $k < k_0$, *uniformly in* k,

$$
\frac{1}{k_0-k} \sum_{t=k+1}^{k_0} \frac{\partial e_t(\theta_{20})}{\partial\theta} e_t(\theta_{20}) = \frac{1}{k_0-k} \sum_{t=k+1}^{k_0} \frac{\partial e_t(\theta_{10})}{\partial\theta} e_t(\theta_{10}) - \Sigma(\theta_{20}-\theta_{10}) + o_p(\nu_T). \tag{30}
$$

Proof We can expand as follows;

$$
\frac{1}{k_0-k} \sum_{t=k+1}^{k_0} \frac{\partial e_t(\theta_{20})}{\partial\theta_2} e_t(\theta_{20}) = \frac{1}{k_0-k} \sum_{t=k+1}^{k_0} \frac{\partial e_t(\theta_{10})}{\partial\theta} e_t(\theta_{10}) + \frac{1}{k_0-k} \sum_{t=k+1}^{k_0} \frac{\partial^2 e_t(\theta_{10})}{\partial\theta\partial\theta'} (d_{20}-d_{10})
$$

$$
+ O_p\left(\nu_T(k_0-k)^{-1/2} \right) + O_p\left(\nu_T^2 \right). \tag{31}
$$

From Ling (2007), $\frac{\partial^2 e_t(\theta_{10})}{\partial\theta\partial\theta'}$ is NED and $\frac{1}{k_0-k} \sum_{t=k+1}^{k_0} \frac{\partial^2 e_t(\theta_{10})}{\partial\theta\partial\theta'}$ converges to Σ in probability uniformly in k. Therefore (30) follows. □

Lemma 6.7. *Define* $A_1(\tau) = \sum_{t=1}^{k} \frac{\partial e_t(\theta_{10})}{\partial\theta} e_t(\theta_{10})$. *Then* $A_1(\tau)$ *is* $O_p(T^{1/2})$ *uniformly in* τ.

Proof Ling (2007) shows that $A_1(\tau)$ can be approximated uniformly by martingale, the deference of which is NED. Furthermore, he shows that for some $\delta > 0$

$$
\max_k k^{\delta} \left| \frac{1}{\sqrt{k}} A_1(\tau) - \frac{1}{\sqrt{k}} \sum_{t=1}^{k} G_{1t} \right| = o_p(1), \tag{32}
$$

where $\{G_{1t} : t = 1, 2, \ldots\}$ is the sequence of *i.i.d.* $m \times 1$ normal vectors with mean zero and covariance matrix Ω. See Theorem 2.2 and Lemma 6.1 in Ling (2007) for details. □

ACKNOWLEDGMENTS

Notes

Corresponding author: Keiko Yamaguchi, Department of Management Science, Faculty of Engineering, Tokyo University of Science, Kagurazaka 1-3, Shinjuku-ku, Tokyo 162-8601, Japan. E-mail: kyamaguchi@ms.kagu.tus.ac.jp

References

BAI, J. (1997) Estimation of a Change Point in Multiple Regression Models. *Review of Economics and Statistics* 79, No. 4, 551–563.

BAI, J. AND PERRON, P. (1998) Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* 66, 47–78.

BERAN, J. (1994) *Statistics for Long Memory Processes*: Chapman & Hall.

BERAN, J. AND TERRIN, N. (1996) Testing for a change of the long-memory parameter. *Biometrika* 83, 627–638.

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

GRANGER, C. AND JOYEUX, R. (1980) An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–29.

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

HORVATH, L. AND SHAO, Q.-M. (1999) Limit theorems for quadratic forms with applications to Whittle's estimate. *Annals of Applied Probability* 9, 146–187.

HORVATH, L. (2001) Change-Point Detection in Long-Memory Processes. *Journal of Multivariate Analysis* 78, 218–234.

Hosking, J. (1981) Fractional differencing. *Biometrika* 68, 165–176.

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

Ray, B. K. and Tsay, R. S. (2002) Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis* 23, 687–705.

Robinson, P. (2006) Conditional-sum-of-squares estimation of models for stationary time series with long memory. *Lecture Notes-Monograph Series*, 130–137.

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.