OPTIMAL CHANGE-POINT ESTIMATION IN TIME SERIES

BY NGAI HANG CHAN∗, WAI LEONG NG†, CHUN YIP YAU∗ AND HAIHAN YU‡

The Chinese University of Hong Kong∗, The Hang Seng University of Hong Kong† and Iowa State University‡

This paper establishes asymptotic theory for optimal estimation of change-points in general time series models under α-mixing conditions. We show that the Bayes-type estimator is asymptotically minimax for change-point estimation under squared error loss. Two bootstrap procedures are developed to construct confidence intervals for the change-points. An approximate limiting distribution of the change-point estimator under small change is also derived. Simulations and real data applications are presented to investigate the finite sample performance of the Bayes-type estimator and the bootstrap procedures.

1. Introduction. Change-point analysis in time series models has received considerable attention in engineering, financial econometrics, genetics and environmetrics. One major goal of change-point analysis is to detect the existence of change-points. In particular, [16] considered a likelihood ratio test for change-points in autoregressive (AR) models. A Wald-type test for change-points in nonlinear time series models is studied in [29], and a self-normalization approach in testing for change-points in time series is developed in [41]. Another goal of change-point analysis in time series is to identify the locations of change-points. For example, [30] considered change-point estimation in time series using quasi-maximum likelihood estimator (QMLE). Methods for estimating the number and locations of change-points are investigated in [17] using the minimum description length principle, and in [23] using a multiscale approach. An accurate estimation of change-points has profound applications in neuroimaging and economic forecasting; see, e.g., [22]. Estimating the locations of change-points constitutes the primary focus of this paper.

Consider first the estimation of one change-point. Extensions to multiple change-point estimation will be discussed in Section 5.3. Let $x_1, x_2, \ldots, x_n$ be a series of observations whose distribution changes at some time point

AMS 2000 subject classifications: Primary 62M10; Secondary 60F25

Keywords and phrases: Bayes-type estimator, confidence interval, double-sided random process, piecewise stationary time series, structural break
\( k_0 \), where \( k_0 \) is unknown and called the change-point. Denote \( E_k \) as the expectation operator evaluated under a probability measure \( P_k \) such that an abrupt change occurs at time \( k \). To estimate \( k_0 \), an estimator \( \hat{k}_n \) is called an asymptotic minimax estimator with respect to the squared error loss if its maximal quadratic risk \( \sup_{k_0} E_{k_0} (\hat{k}_n - k_0)^2 \) attains the asymptotic minimax lower bound asymptotically, i.e.,

\[
\lim \inf_{n \to \infty} \inf_{T_n} \sup_{k_0} E_{k_0} (T_n - k_0)^2 \geq \lim \sup_{n \to \infty} \sup_{k_0} E_{k_0} (\hat{k}_n - k_0)^2,
\]

where \( \inf_{T_n} \) denotes the infimum over all estimators \( T_n = T_n(x_1, \ldots, x_n) \) of \( k_0 \). In other words, \( \hat{k}_n \) is optimal in the sense that the maximal quadratic risk of \( \hat{k}_n \) does not exceed that of any other estimator asymptotically. Since \( k_0 \in \mathbb{Z} \) and the mean squared error is not a smooth function with respect to \( k_0 \), the standard theory on the asymptotic optimality of maximum likelihood estimator (MLE) does not apply. In particular, \([39]\) proved that MLE is minimax under the 0-1 loss function for independent normal observations with a change-in-mean. However, MLE is not optimal under the squared error loss function: \([37]\) studied a Bayes-type estimator and established its asymptotic efficiency under squared error loss for a univariate parameter change in i.i.d. sequences. \([7]\) established the asymptotic minimaxity of the Bayes-type estimator under mild assumptions. \([1]\) and \([2]\) studied Bayes-type change-point estimators and their asymptotic distributions. All of the aforementioned works, however, focus on the independent case; optimality under dependence remains largely unexplored. Moreover, optimal inference for change-points is difficult in practice as a closed-form expression for the asymptotic distribution of the Bayes-type estimator of a change-point does not exist. On the other hand, bootstrap inference provides a reasonable approximation to the asymptotic distribution. \([20]\) studied some nonparametric change-point estimators and bootstrap methods to construct confidence regions. \([1]\) and \([2]\) also proposed bootstrap methods for constructing change-point confidence intervals. However, the aforementioned works mainly focus on the independent case. Although \([26]\) studied a circular moving block bootstrap to obtain confidence intervals for change-points in a time series setting, only simple location models were explored.

In this paper, we study the Bayes-type estimator of change-points for general time series models under \( \alpha \)-mixing conditions. The Bayes-type estimator is shown to be asymptotically minimax with respect to the squared error loss. An approximate asymptotic distribution of the estimator under small change is also derived. One challenging feature of the aforementioned asymptotic analysis is that when a structural change occurs in time series, the post-change segment is non-stationary. To address this, we develop some
Lipschitz-type conditions to technically control the aggregate effect from the dependence between the pre-change and post-change segments. Moreover, we propose two bootstrap procedures to approximate the asymptotic distribution of the estimator for inference of the change-point. The validities of the bootstrap procedures are established analytically. We also extend the procedure to conduct simultaneous inference for multiple change-points.

The paper is organized as follows. In Section 2, we introduce the Bayes-type estimator and establish its asymptotic minimaxity. In Section 3, an approximate asymptotic distribution under small change is derived. In Section 4, two bootstrap procedures are proposed to construct confidence intervals for the change-points. In Section 5, simulation studies are conducted to demonstrate the efficiency gains of the Bayes-type estimator compared to the MLE, and performances of the bootstrap procedures are compared. Real data applications are provided in Section 6. Technical assumptions and lemmas are discussed in Section 7. Section 8 concludes. All technical proofs are provided in the supplementary material [13].

2. Bayes-type Estimators of Change-points for General Time Series Models. Consider a stationary time series \( \{x_t\}_{t \in \mathbb{Z}} \) generated by

\[
x_t = f(\theta, X_{t-1}, \eta_t),
\]

where \( X_{t-1} = (x_{t-1}, x_{t-2}, \ldots) \), \( \theta \) is an unknown parameter vector from a compact parameter space \( \Theta \subset \mathbb{R}^d \), \( \{\eta_t\}_{t \in \mathbb{Z}} \) is a sequence of independent and identically distributed random variables with distribution function \( F_0 \), and \( f \) is a continuous function which is assumed to be known with its functional form depending solely on a class of pre-specified parametric models. The model (2.1) includes many time series models in the literature, such as ARMA, GARCH and random coefficient AR models. Asymptotic theory and change-point analysis related to model (2.1) are studied by [43], [44], [29] and [30]. Also, the independence of \( \{\eta_t\}_{t \in \mathbb{Z}} \) can be relaxed as long as \( \{x_t\}_{t \in \mathbb{Z}} \) satisfies the mixing condition in Assumption 1 in Section 7. Assume that \( x_t \) is \( \mathcal{F}_t \) measurable, where \( \mathcal{F}_t \) is the \( \sigma \)-algebra generated by \( \{\eta_t, \eta_{t-1}, \ldots\} \). Denote \( \{x_t\} \sim M(\theta) \) if \( \{x_t\} \) follows model (2.1).

We consider a change-point model, denoted as \( M_{k_0}(\theta_{10}, \theta_{20}) \), in which

\[
x_t = \begin{cases} f(\theta_{10}, X_{t-1}, \eta_t), & t \leq k_0, \\ f(\theta_{20}, X_{t-1}, \eta_t), & t > k_0, \end{cases}
\]

where \( \theta_{10} \neq \theta_{20} \) are interior points in \( \Theta \), and \( k_0 \) is an unknown integer representing a change-point. The limiting behavior under fixed change, i.e., \( \|\theta_{20} - \theta_{10}\| \) is a constant, is considered in this section, and the approximate
limiting behavior under small change is studied in Section 3. Given a set of innovations \( \{ \eta_t \} \), the following three different processes can be generated:

(i) stationary process \( \{ x_t^{(1)} \} \sim M(\theta_{10}) \) following (2.1) with \( \theta = \theta_{10} \),
(ii) stationary process \( \{ x_t^{(2)} \} \sim M(\theta_{20}) \) following (2.1) with \( \theta = \theta_{20} \),
(iii) change-point process \( \{ x_t \} \sim M_{k_0}(\theta_{10}, \theta_{20}) \) following (2.2).

Note that the first two processes are stationary and satisfy some weak dependence conditions described in Assumption 1 in Section 7. Also, \( \{ x_t \} \leq k_0 = \{ x_t^{(1)} \} \leq k_0 \) since they are generated using the same parameter and inputs. However, the post-change data \( \{ x_t \} > k_0 \) is non-stationary due to its dependence on the pre-change data which has a different dependence structure.

To establish asymptotic theory, it is necessary that the number of observations before and after the change-point increases to infinity as the horizon of observations grows. Thus, we assume that \( k_0 = [\tau_0 n] \), where \([x]\) is the greatest integer that is less than or equal to \( x \) and \( \tau_0 \in (\lambda, 1-\lambda) \) is known as the relative change-point and \( \lambda \) is a sufficiently small positive constant. Hence, the support of the change-point is \( \Gamma(n) = \{ k \in \mathbb{N}^+ : \lambda n \leq k \leq (1-\lambda)n \} \).

Theoretically, an arbitrary small but fixed \( \lambda \) is sufficient to ensure that the pre-change and post-change segments have \( O(n) \) observations to accurately estimate the specified parameter values, see [16], [17], [41], [30], [21] and [24] for similar settings. In practice, the searching space of the change-point estimates can be \( \{ 1 \leq k \leq n \} \) or \( \{ d \leq k \leq n-d \} \), where \( d \) is the dimension of the parameter space so that the model parameters are identifiable; see [30], [21] and [24]. In Section 4.5 of [17], the estimation performance remains promising when the change-point occurs close to the beginning of the series.

For estimation purpose, consider the objective function

\[
L_n(k, \theta_1, \theta_2) = \sum_{t=1}^{k} l_t(\theta_1) + \sum_{t=k+1}^{n} l_t(\theta_2),
\]

where \( l_t(\theta) = l(x_t | \theta, X_{t-1}) \) is a measurable objective function for parameter estimation of the time series model in a stationary segment. This setting includes many common estimators, such as least squares estimator (LSE), MLE, QMLE, and \( M \)-estimators.

An estimator of the change-point can be defined by

\[
\hat{k}_n = \arg \max_{k \in \Gamma(n)} L_n(k, \hat{\theta}_1(k), \hat{\theta}_2(k)),
\]

where \( \hat{\theta}_1(k) = \arg \max_{\theta_1 \in \Theta} \sum_{t=1}^{k} l_t(\theta_1) \) and \( \hat{\theta}_2(k) = \arg \max_{\theta_2 \in \Theta} \sum_{t=k+1}^{n} l_t(\theta_2) \) are the model parameter estimates given a change-point at \( k \). Once the estimated change-point \( \hat{k}_n \) is found, the model parameters are estimated by
\[ \hat{\theta}_1, \hat{\theta}_2 \equiv (\hat{\theta}_1(\hat{k}_n), \hat{\theta}_2(\hat{k}_n)) = \arg \max_{(\hat{\theta}_1, \hat{\theta}_2) \in \Theta^2} L_n(\hat{k}_n, \theta_1, \theta_2). \]

In this paper, we study the Bayes-type estimator defined by
\[ \tilde{k}_n = \frac{\sum_{k=1}^n k \exp L_n(k, \hat{\theta}_1, \hat{\theta}_2)}{\sum_{k=1}^n \exp L_n(k, \hat{\theta}_1, \hat{\theta}_2)}, \tag{2.6} \]
where \((\hat{\theta}_1, \hat{\theta}_2)\) is defined in (2.5). Under the independent setting, the asymptotic efficiency and the asymptotic minimaxity of the Bayes-type estimator in (2.6) have been established in [37] and [7], respectively.

Remark 2.1. In practice, \(X_{t-1} = (x_{t-1}, x_{t-2}, \ldots)\) is not available for computing \(l_t(\theta)\) since it involves data from the infinite past. One approach is to use 0 or some constant sequence as initial values to replace the unobserved \(\{x_t\}_{t \leq 0}\), i.e., \(l_t(\theta) = l(x_t|\theta, \{x_{t-1}, x_{t-2}, \ldots, x_1, 0, 0, \ldots\})\). Moreover, for change-point models, one may use 0 as the pre-change sequence in the post-change \(l_t\), i.e., \(l_t(\theta) = l(x_t|\theta, \{x_{t-1}, x_{t-2}, \ldots, x_k, 0, 0, \ldots\}) =: l_{t,k}(\theta)\), where \(k\) is the change-point and \(t > k\); see, e.g., [6], [29], [4], [31], and [30].

If the main goal is point estimation, then, under Assumption 4, Lemma 7.2 in Section 7 guarantees that the effect of the above treatment on the second sum in (2.3) is negligible. Hence, both \(l_{t,k}(\theta)\) and \(l(x_t|\theta, X_{t-1})\) give consistent estimators in the sense that \(\hat{k}_n/n - \tau_0 = O_p(n^{-1})\). On the other hand, if the main goal is to make inference, then only \(l_t(\theta) = l(x_t|\theta, X_{t-1})\) is appropriate. The reason is that, under (2.2), the generation of the post-change sequence involves the pre-change sequence. As the distribution of the change-point estimator depends heavily on the observations in the neighborhood of the change-point, \(l_t(\theta) = l(x_t|\theta, X_{t-1})\) has to be employed to capture the dependence on the pre-change sequence; see Section 4 for details.

Next we establish the main results of the paper, with the technical assumptions and lemmas deferred to Section 7. Define the double-sided random process
\[ W(k) = \begin{cases} \sum_{t=1}^k [l_t(\theta_{10}) - l_t(\theta_{20})], & k > 0, \\ 0, & k = 0, \\ \sum_{t=-k}^{-1} [l_t(\theta_{20}) - l_t(\theta_{10})], & k < 0, \end{cases} \tag{2.7} \]
where \( \{x_t\}_{t \in \mathbb{Z}} \) satisfies (2.2) with \( k_0 = 0 \). The following theorem describes the asymptotic distribution of the Bayes-type estimator \( \hat{k}_n \).

**Theorem 2.1.** If Assumptions 1 to 4 in Section 7 hold, then
\[
(\hat{\theta}_1, \hat{\theta}_2) \to (\theta_{10}, \theta_{20}), \quad \mathbb{P}_{k_0} \text{-a.s.}
\]
Also, uniformly for all \( k_0 \in \Gamma(n) \), under the probability measure \( \mathbb{P}_{k_0} \),
\[
\tilde{k}_n - k_0 \overset{L}{\to} \frac{\sum_{u \in \mathbb{Z}} u \exp W(u)}{\sum_{u \in \mathbb{Z}} \exp W(u)} =: \tilde{u},
\]
where “\( L \to \)” denotes weak convergence, and
\[
\lim_{n \to \infty} \mathbb{E}_{k_0} (\tilde{k}_n - k_0)^2 = \mathbb{E}(\tilde{u}^2) < \infty.
\]

Note that (2.8) implies that \( n(\tilde{\tau}_n - \tau_0) \overset{L}{\to} \tilde{u} \), where \( \tau_0 = k_0/n \) is the relative change-point and \( \tilde{\tau}_n = \tilde{k}_n/n \) is the corresponding estimator. The convergence rate of order \( \mathcal{O}_p(n^{-1}) \) for \( \tilde{\tau}_n \) is common in change-point estimators, see [30]. Before presenting the asymptotic optimality results, we define some terminology. First, with respect to the quadratic loss function, define the Bayes risk with respect to uniform prior over \( \Gamma(n) \) as
\[
\mathcal{R}_B(T_n) = \frac{1}{|\Gamma(n)|} \sum_{k_0 \in \Gamma(n)} \mathbb{E}_{k_0} [(T_n - k_0)^2],
\]
and define the maximal risk of a change-point estimator \( T_n \) as
\[
\mathcal{R}_M(T_n) = \sup_{k_0 \in \Gamma(n)} \mathbb{E}_{k_0} [(T_n - k_0)^2].
\]
The following theorem asserts that, when \( l_\cdot(\theta) \) is the log-likelihood function, the corresponding Bayes-type estimator, denoted as \( \tilde{k}_n^* \), is an asymptotic minimax estimator of \( k_0 \) with respect to the quadratic loss.

**Theorem 2.2.** If Assumptions 1-4 in Section 7 hold, then
\[
\lim_{n \to \infty} \inf_{T_n} \mathcal{R}_M(T_n) \geq \lim_{n \to \infty} \inf_{T_n} \mathcal{R}_B(T_n) = \lim_{n \to \infty} r_B(\tilde{k}_n^*) = \mathbb{E}[(\tilde{u}^*)^2] = \lim_{n \to \infty} r_M(\tilde{k}_n^*),
\]
where \( \inf_{T_n} \) denotes the infimum over all estimators of \( k_0 \), \( \tilde{k}_n^* \) and \( \tilde{u}^* \) are defined in (2.6) and (2.8) respectively, with \( l_\cdot(\cdot) \) being taken as the log-likelihood function. Specifically, the Bayes-type estimator \( \tilde{k}_n^* \) asymptotically minimizes the Bayes risk under quadratic loss, and together with (2.9) that the quadratic risk of \( \tilde{k}_n^* \) is asymptotically constant as a function of \( k_0 \in \Gamma(n) \), we have the asymptotic minimaxity of \( k_n^* \) with respect to the quadratic loss.
Remark 2.2. Testing for a change-point in time series models defined in (2.1) can be found in [29], which considers a max-type Wald test statistic
\[
\max_{k \in [1,n]} \frac{k(n-k)}{n^2} [\hat{\theta}_1(k) - \hat{\theta}_2(k)]' [\hat{\Sigma}_n(k) \hat{\Omega}_n^{-1}(k) \hat{\Sigma}_n(k)] [\hat{\theta}_1(k) - \hat{\theta}_2(k)] ,
\]
where \(\hat{\Sigma}_n(k)\) and \(\hat{\Omega}_n(k)\) are some variance estimators. The asymptotic theories for change-point estimation and change-point testing are very different.

For change-point estimation, argmax-type statistics are used and the corresponding asymptotic distributions involve the argmax of a double-sided random process, see [3] and [30]. For change-point testing, max-type test statistics are used. Under some regularity and weak dependence conditions, the change-point test statistics have a Darling-Erdős-type limit, see [29].

3. Approximate Asymptotic Distributions of Bayes-type Estimators Under Small Change. The distribution of \(\tilde{u}\) in (2.8) does not have a closed form expression, and is dependent on unknown parameters \(\theta_{10}\) and \(\theta_{20}\). However, when the difference between two parameters approaches 0, we have a pivotal approximation to \(\tilde{u}\). Specifically, define
\[
(3.1) \quad \Delta = \theta_{20} - \theta_{10}.
\]
Denote \(\dot{l}_t(\theta) := \partial l_t(\theta)/\partial \theta\) and \(\ddot{l}_t(\theta) := \partial^2 l_t(\theta)/\partial \theta \partial \theta'\). Let \(E_{\theta_0}(\cdot)\) be the expectation under \(\{x_t\} \sim M(\theta_0)\). Under the moment conditions in Assumption 3 stated in Section 7, we define \(E_{\theta_0}(\dot{l}_t(\theta_0)\dot{l}_t(\theta_0)') = \Omega_0\) and \(E_{\theta_0}(\ddot{l}_t(\theta_0)) = -\Sigma_0\), where \(\Omega_0\) and \(\Sigma_0\) are positive-definite matrices. The following theorem gives the approximate asymptotic distribution of the Bayes-type estimator for the case \(\Delta \to 0\).

Theorem 3.1. Let \(s = (\Delta' \Sigma_{10} \Delta)^{-2} (\Delta' \Omega_{10} \Delta)\) or \((\Delta' \Sigma_{20} \Delta)^{-2} (\Delta' \Omega_{20} \Delta)\). If Assumptions 1-4 in Section 7 hold, then the asymptotic distribution of the Bayes-type estimator \(\tilde{u}\) in (2.8) satisfies
\[
(3.2) \quad \frac{1}{s} \tilde{u} = \frac{1}{s} \frac{\sum_{j=-\infty}^{\infty} j \exp W(j)}{\sum_{j=-\infty}^{\infty} \exp W(j)} \frac{\zeta}{\int_{-\infty}^{\infty} \exp \{\kappa (B(z) - |z|/2)\} dz} ,
\]
as \(\Delta \to 0\), where \(B(z)\) is the standard Brownian motion and
\[
\kappa = \lim_{\Delta \to 0} \frac{\Delta' \Omega_{10} \Delta}{\Delta' \Sigma_{10} \Delta} = \lim_{\Delta \to 0} \frac{\Delta' \Omega_{20} \Delta}{\Delta' \Sigma_{20} \Delta}.
\]

When \(l_t(\cdot)\) is the log-likelihood function, it is standard to check that \(\Omega_{i0} = \Sigma_{i0}, i = 1, 2\). Hence, \(\kappa = 1\) and we have the following corollary.
Corollary 3.2. If \( l_t(\cdot) \) is the log-likelihood function, then with \( s = (\Delta' \Sigma_{10} \Delta)^{-1} \) or \((\Delta' \Sigma_{20} \Delta)^{-1}\), the asymptotic distribution of the Bayes-type estimator \( \tilde{u}^* \) in (2.10) satisfies, as \( \Delta \to 0 \),

\[
\frac{1}{s} \tilde{u}^* = \frac{1}{s} \sum_{j=-\infty}^{\infty} j \exp W(j) \left( \int_{-\infty}^{\infty} z \exp \{ B(z) - |z| / 2 \} \, dz \right) / \left( \int_{-\infty}^{\infty} \exp \{ B(z) - |z| / 2 \} \, dz \right).
\]

The pivotal asymptotic distribution in Corollary 3.2 allows one to construct confidence intervals of change-points. Table 1 provides the percentiles of the asymptotic distribution, which are determined by Monte Carlo simulations of 100,000 paths. For each path, the range is from -10,000 to 10,000 with a step size of 0.01.

| Quantiles of the pivotal asymptotic distribution in Corollary 3.2. |
|-----------------|---|---|---|---|---|---|
| Critical Values | 2.5% | 5% | 10% | 90% | 95% | 97.5% |
|                 | -9.16 | -6.68 | -4.45 | 4.43 | 6.66 | 9.15 |

Remark 3.1. Similar to the framework described in Section 3 of [30], the true parameter \((\theta_{10}, \theta_{20})\) in Theorems 2.1 and 2.2 is fixed, and hence \( \Delta \) is fixed and does not depend on \( n \). On the other hand, Theorem 3.1 and Corollary 3.2 take \( \Delta \to 0 \) on the asymptotic distribution \( \tilde{u} \) in Theorem 2.2. Thus, the rate of \( \Delta \to 0 \) does not depend on \( n \), and the results should be interpreted as a reasonable pivotal approximation to \( \tilde{u} \) when \( \Delta \) is small.

4. Bootstrap Inference on Change-points. From Theorem 2.1 and (2.8), the asymptotic distribution of the Bayes-type estimator \( \hat{k}_n \) depends heavily on the models of the segments before and after the change-point. Also, the pivotal distribution in Theorem 3.1 requires that the change \( \Delta \) is small. To construct confidence intervals for the change-point in general, we propose two bootstrap procedures to approximate the distribution of \( \tilde{u} \).

4.1. Bootstrap Procedures. The basic idea of the bootstrap procedures is to obtain replicates of \( \tilde{u} \) in (2.8) using replicates of \( W(u) \). Since the summations in (2.8) involve all integers, a truncation parameter \( m \) has to be introduced for approximation. The choice of \( m \) is investigated in Section 4.2. The bootstrap procedures for a fixed \( m \) are described as follows.

- **Parametric Bootstrap (PB)**
  - Step 1: Set \( m \in Z^+ \). For \( i = 1, \ldots, B \), repeat Step 2 to Step 4.
  - Step 2: Simulate \( \{y_{t(i)}\}_{t=-m,...,m} \) according to model (2.2) with \( k_0 = 0 \), and \( \theta_{10} \) and \( \theta_{20} \) replaced by \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \), respectively.
Step 3: Calculate the objective functions \( \{ l_t^{(i)}(\hat{\theta}_1), l_t^{(i)}(\hat{\theta}_2) \}_{t=-m,...,m} \). Then, compute the double-sided random process

\[
\tilde{W}^{(i)}(j) = \begin{cases} 
\sum_{t=1}^{j}[l_t^{(i)}(\hat{\theta}_1) - l_t^{(i)}(\hat{\theta}_2)], & j > 0, \\
0, & j = 0, \\
\sum_{t=j+1}^{m}[l_t^{(i)}(\hat{\theta}_2) - l_t^{(i)}(\hat{\theta}_1)], & j < 0,
\end{cases}
\]

Step 4: Compute \( u^{(i)} = \left( \frac{\sum_{j=-m}^{m} j \exp \tilde{W}^{(i)}(j)}{\sum_{j=-m}^{m} \exp \tilde{W}^{(i)}(j)} \right) \).

Step 5: Define \( q^*_u \) and \( q^*_n \) as the \( \alpha/2 \) and \( 1 - \alpha/2 \) quantiles of the sample \( \{ u^{(1)}, u^{(2)}, \ldots, u^{(B)} \} \), respectively. The bootstrap confidence interval is given by \( CI_p = (\tilde{k}_n - q^*_u, \tilde{k}_n - q^*_n) \).

- **Hybrid Bootstrap (HB)**

Step 1: Calculate residuals \( \{ \tilde{y}_t \}_{t=\tilde{k}_n-[\log n]+1,...,n} \) based on estimates \( (\hat{\theta}_1, \hat{\theta}_2, \tilde{k}_n) \).

Step 2: Set \( m \in \mathbb{Z}^+ \). For \( i = 1, \ldots, B \), repeat Step 3 to Step 6.

Step 3: Sample a block of length \( m+1 \), \( \{ \tilde{y}_t^{(i)} \}_{t=-m,...,0} \) from \( \{ x_t \}_{t=1,...,\tilde{k}_n-[\log n]} \).

Step 4: Conditional on \( \{ \tilde{y}_t^{(i)} \}_{t=-m,...,0} \), simulate \( \{ \tilde{y}_t^{(i)} \}_{t=1,...,m} \) using \( \tilde{y}_t^{(i)} = f(\hat{\theta}_2, \tilde{Y}_{t-1}^{(i)}, \eta_t^*) \), where \( \tilde{Y}_{t-1}^{(i)} = \{ \tilde{y}_t^{(i)} \}_{t=-m,...,-1,0,1,...,t-1} \) and \( \eta_t^* \) are independently resampled from the residuals \( \{ \tilde{y}_t \}_{t=\tilde{k}_n+[\log n]+1,...,n} \).

Step 5: Calculate the objective functions \( \{ l_t^{(i)}(\hat{\theta}_1), l_t^{(i)}(\hat{\theta}_2) \}_{t=-m,...,m} \). Then, compute the double-sided random process

\[
\tilde{W}^{(i)}(j) = \begin{cases} 
\sum_{t=1}^{j}[\tilde{l}_t^{(i)}(\hat{\theta}_1) - \tilde{l}_t^{(i)}(\hat{\theta}_2)], & j > 0, \\
0, & j = 0, \\
\sum_{t=j+1}^{m}[\tilde{l}_t^{(i)}(\hat{\theta}_2) - \tilde{l}_t^{(i)}(\hat{\theta}_1)], & j < 0,
\end{cases}
\]

Step 6: Compute \( u^{(i)} = \left( \frac{\sum_{j=-m}^{m} j \exp \tilde{W}^{(i)}(j)}{\sum_{j=-m}^{m} \exp \tilde{W}^{(i)}(j)} \right) \).

Step 7: Define \( \tilde{q}_l \) and \( \tilde{q}_u \) as the \( \alpha/2 \) and \( 1 - \alpha/2 \) quantiles of the sample \( \{ u^{(1)}, u^{(2)}, \ldots, u^{(B)} \} \), respectively. The bootstrap confidence interval is given by \( CI_h = (\tilde{k}_n - \tilde{q}_u, \tilde{k}_n - \tilde{q}_l) \).

Note that in HB, the bootstrap procedures are conducted on \( \{ t = 1, \ldots, \tilde{k}_n-[\log n] \} \) and \( \{ t = \tilde{k}_n + 1 + [\log n], \ldots, n \} \) instead of \( \{ t = 1, \ldots, \tilde{k}_n \} \) and \( \{ t = \tilde{k}_n+1, \ldots, n \} \). Since \( \tilde{k}_n - k_0 = O_p(1) \), sampling outside \( [\tilde{k}_n-[\log n], \tilde{k}_n+[\log n]) \) guarantees asymptotically that the resamples are from a segment without a change-point.

The following theorem establishes the validity of the two bootstrap procedures. For notational simplicity, we omit the superscript \( (i) \) that indicates the \( i \)-th realization of the \( B \) resamples. We also denote by \( P^* \) the bootstrap probability distribution conditional on the observations \( \{ x_1, x_2, \ldots, x_n \} \).
Theorem 4.1. If Assumptions 1-4 in Section 7 hold and $m \to \infty$, $m/n \to 0$ as $n \to \infty$, then the distribution of the double-sided random process in PB satisfies
\[
\{W(j) : \{-m, \ldots, m\}\} \xrightarrow{L} \{W(j) : j \in \mathbb{Z}\} \text{ with probability one,}
\]
where $W(j)$ is defined in \((2.7)\). If additionally, the empirical distribution function of the residuals $\{\tilde{\eta}_t\}_{t=\hat{k}_n+\lceil \log n \rceil +1, \ldots, n}$ converges weakly to the distribution function of $\{\eta_t\}$, i.e., for all continuity point $x$ of $F_0$,
\[
(4.1) \quad \tilde{F}_{\tilde{\eta}}(x) = \frac{1}{n - \hat{k}_n - \lceil \log n \rceil} \sum_{i=\hat{k}_n + \lceil \log n \rceil +1}^{n} \mathbb{1}(\tilde{\eta}_i < x) \to F_0(x) \quad \text{a.s.,}
\]
then the distribution of the double-sided random process in HB satisfies
\[
\{\tilde{W}(j) : \{-m, \ldots, m\}\} \xrightarrow{L} \{W(j) : j \in \mathbb{Z}\} \text{ with probability one,}
\]
as $n \to \infty$. Hence, for any $u$ generated by PB and HB, as $n \to \infty$,
\[
\sup_{a \in \mathbb{R}} |P^*(u \leq a) - P(\tilde{u} \leq a)| \to 0 \text{ with probability one.}
\]

Remark 4.1. The buffer period $\lceil \log n \rceil$ used in HB can be replaced by any slowly increasing sequences of order $o(n)$. Since $\hat{k}_n - k_0 = O_p(1)$, the use of the buffer guarantees that the resampled block from the pre-change segment and the resampled residuals from the post-change segment are away from the true change-point. The results of the bootstrap procedure are not sensitive to the choice of the length of the buffer period when the sample size is sufficiently large. The choice of $\lceil \log n \rceil$ performs well in all simulations in Section 5. The additional assumption (4.1) for the HB bootstrap is not restrictive. Many commonly used time series models, including moving average models, autoregressive models, ARMA models, GARCH models and many others, fulfill these assumptions, see [8], [14], [15], and [27].

Remark 4.2. There are advantages and limitations of the two proposed bootstrap procedures. For PB, the distribution of the innovation terms $\{\eta_t\}$ is required to be known. Misspecification of this distribution will induce bias in the construction of the confidence intervals. On the other hand, HB is more robust under misspecification of the pre-change model and the distributional assumption of the innovations $\{\eta_t\}$, since it samples blocks of observations in the first segment and uses the residuals to generate bootstrap samples. However, HB can only apply to invertible time series models in...
the sense that the white noises can be estimated from the equation $x_t = f(\theta, X_{t-1}, \eta_t)$, see [25] and [42] for details on the invertibility. Nevertheless, if one assumes that the pre-change and post-change segments are independent, then nonparametric block bootstrap and objective function block bootstrap procedures are proposed to allow for valid confidence intervals under model misspecification in both segments, see the supplement [13] for details.

4.2. A Data-driven Choice of $m$. It is clear from the definition in (2.7) that the doubled-sided random process $W(k)$ tends to be larger for $k$ around 0 and tends to be more negative as $|k|$ increases. Therefore, when $m$ is sufficiently large, the error from the truncation is negligible. To be precise, the following theorem states the asymptotic behaviour for the error terms.

**Theorem 4.2.** If Assumptions 1-4 in Section 7 hold, then for any $u$ generated by PB or HB, and for all $a \in \mathbb{R}$ and $m \in \mathbb{Z}^+$, as $n \to \infty$,

$$
\mathbf{P}^*(u < a) \to \mathbf{P} \left( \tilde{u} < \frac{a + \epsilon_2(m)}{1 + \epsilon_1(m)} \right) \quad \text{with probability one,}
$$

where for $i = 1, 2$,

$$
\epsilon_i(m) = \left( \sum_{j=\infty}^{-m} j^{i-1} e^{W(j)} + \sum_{j=-m}^{\infty} j^{i-1} e^{W(j)} \right) / \sum_{j=-m}^{m} e^{W(j)} .
$$

Also, $\epsilon_1(m) \to 0$ and $\epsilon_2(m) \to 0$ a.s. when $m \to \infty$.

We now propose a data-driven choice of $m$. The goal is to bound $\epsilon_1(m)$ and $\epsilon_2(m)$ in (4.2). From (4.3), we have for any fixed $m$ that for $i = 1, 2$,

$$
|\epsilon_i(m)| < \left( \sum_{j=-\infty}^{-m} |j| e^{W(j)} + \sum_{j=-m}^{\infty} |j| e^{W(j)} \right) / \sum_{j=-m}^{m} e^{W(j)} .
$$

The following lemma helps to construct bounds for $\sum_{j=-\infty}^{-m} |j| e^{W(j)}$ and $\sum_{j=-m}^{\infty} |j| e^{W(j)}$.

**Lemma 4.3.** For any $m \in \mathbb{Z}^+$ and $G < 0$, define

$$
h(m|G) = \ln \left( \frac{m - (m - 1) e^G}{2m - 2 (m - e^G) e^G} \right) - mG .
$$

If $M$ is a positive even integer satisfying $h \left( \frac{M}{2} |G \right) > 0$, then

$$
\sum_{j=M}^{\infty} |j| e^{jG} < \sum_{j=M/2}^{M} |j| e^{jG} \quad \text{and} \quad \sum_{j=-\infty}^{-M} |j| e^{jG} < \sum_{j=-M}^{-M/2} |j| e^{jG} .
$$
Note that $E(W(j)) = |j| G_{\theta_1}$ when $j < 0$, and $E(W(j)) = |j| G_{\theta_2} + o(j)$ when $j > 0$, where

$$ (4.6) \quad G_{\theta_1} = E_{\theta_{10}}(l_t(\theta_{20}) - l_t(\theta_{10})) \quad \text{and} \quad G_{\theta_2} = E_{\theta_{20}}(l_t(\theta_{10}) - l_t(\theta_{20})), $$

are negative constants. Combining with Lemma 4.3 and (4.4), if $h(m/2 | G_{\theta_i}) > 0$, $i = 1, 2$, then $\epsilon_i(m), i = 1, 2$ can be approximately bounded by

$$ \left( \sum_{j=-m}^{-m/2} |j| e^{W(j)} + \sum_{j=m/2}^{m} |j| e^{W(j)} \right) / \sum_{j=-m}^{m} e^{W(j)} . $$

Note that $G_{\theta_1}$ and $G_{\theta_2}$ can be estimated by $\hat{W}(\tilde{k}_n)/|\tilde{k}_n|$ and $\hat{W}(1 - \tilde{k}_n)/|1 - \tilde{k}_n|$ respectively. This suggests a data-driven choice of $m$ for the bootstrap procedures, i.e., choosing the smallest $m$ such that

$$ \min(h(m/2 | \hat{W}(\tilde{k}_n))/|\tilde{k}_n|, h(m/2 | \hat{W}(1 - \tilde{k}_n)/|1 - \tilde{k}_n|)) > 0, $$

and

$$ (4.8) \quad e(m) := \left( \sum_{j=-m}^{-m/2} |j| e^{W(j)} + \sum_{j=m/2}^{m} |j| e^{W(j)} \right) / \sum_{j=-m}^{m} e^{W(j)} < \gamma, $$

for a pre-specified error tolerance level $\gamma > 0$.

5. Simulation Studies. This section examines the finite sample performance of the asymptotic results via Monte Carlo simulations. We consider the autoregressive (AR) model

$$ (5.1) \quad x_t = \psi_1 x_{t-1} \mathbb{I}_{\{t \leq k_0\}} + \psi_2 x_{t-1} \mathbb{I}_{\{t > k_0\}} + \epsilon_t, \quad \epsilon_t \overset{i.i.d.}{\sim} N(0, 1), $$

and generalized autoregressive conditional heteroskedastic (GARCH) model

$$ (5.2) \quad x_t = \epsilon_t \sqrt{h_t}, \quad \epsilon_t \overset{i.i.d.}{\sim} N(0, 1), $$

where $h_t = (\omega_1 + \alpha_1 x_{t-1}^2 + \beta_1 h_{t-1}) \mathbb{I}_{\{t \leq k_0\}} + (\omega_2 + \alpha_2 x_{t-1}^2 + \beta_2 h_{t-1}) \mathbb{I}_{\{t > k_0\}}$. Denote $\theta = (\omega, \alpha, \beta)$. The simulations are carried out in R, and the package fGarch is employed for the estimation of the GARCH models.

5.1. Efficiency Gains of Bayes-type Estimators. To demonstrate the efficiency gains of the Bayes-type estimator compared to the MLE, we compare their mean and root-mean-squared error (RMSE). The standard errors of the RMSE, which are calculated by the delta method, are also provided to justify the significance of the difference between the RMSE of the two
Table 2
Mean, RMSE ($\times 10^{-2}$) and se(RMSE) ($\times 10^{-2}$) of the MLE and the Bayes-type estimator for AR models with Gaussian innovations of length $n = 500$.

<table>
<thead>
<tr>
<th>$\tau_0$</th>
<th>MLE</th>
<th>Bayes</th>
<th>MLE</th>
<th>Bayes</th>
<th>MLE</th>
<th>Bayes</th>
<th>MLE</th>
<th>Bayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.229</td>
<td>0.270</td>
<td>0.207</td>
<td>0.206</td>
<td>0.226</td>
<td>0.256</td>
<td>0.201</td>
<td>0.202</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.87</td>
<td>1.44</td>
<td>0.11</td>
<td>0.09</td>
<td>1.34</td>
<td>1.06</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>se(RMSE)</td>
<td>0.09</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.08</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Mean</td>
<td>0.499</td>
<td>0.500</td>
<td>0.505</td>
<td>0.500</td>
<td>0.505</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.78</td>
<td>0.37</td>
<td>0.08</td>
<td>0.05</td>
<td>0.54</td>
<td>0.28</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>se(RMSE)</td>
<td>0.04</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Mean</td>
<td>0.687</td>
<td>0.671</td>
<td>0.704</td>
<td>0.698</td>
<td>0.697</td>
<td>0.68</td>
<td>0.699</td>
<td>0.699</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.05</td>
<td>0.58</td>
<td>0.08</td>
<td>0.05</td>
<td>0.70</td>
<td>0.41</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>se(RMSE)</td>
<td>0.06</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
<td>0.06</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 3
Mean, RMSE ($\times 10^{-2}$) and se(RMSE) ($\times 10^{-2}$) of the MLE and the Bayes-type estimator for GARCH models of length $n = 500$.

<table>
<thead>
<tr>
<th>$\theta_1$=$\left(0.1, 0.1, 0.45\right)$</th>
<th>$\theta_2$=$\left(0.15, 0.15, 0.5\right)$</th>
<th>$\theta_2$=$\left(0.2, 0.2, 0.55\right)$</th>
<th>$\theta_2$=$\left(0.3, 0.3, 0.65\right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>Bayes</td>
<td>MLE</td>
<td>Bayes</td>
</tr>
<tr>
<td>Mean</td>
<td>0.286</td>
<td>0.278</td>
<td>0.221</td>
</tr>
<tr>
<td>RMSE</td>
<td>18.91</td>
<td>18.37</td>
<td>6.77</td>
</tr>
<tr>
<td>se(RMSE)</td>
<td>0.08</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td>Mean</td>
<td>0.517</td>
<td>0.509</td>
<td>0.510</td>
</tr>
<tr>
<td>RMSE</td>
<td>11.03</td>
<td>10.63</td>
<td>4.08</td>
</tr>
<tr>
<td>se(RMSE)</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>Mean</td>
<td>0.621</td>
<td>0.636</td>
<td>0.706</td>
</tr>
<tr>
<td>RMSE</td>
<td>23.32</td>
<td>19.83</td>
<td>4.13</td>
</tr>
<tr>
<td>se(RMSE)</td>
<td>0.07</td>
<td>0.07</td>
<td>0.04</td>
</tr>
</tbody>
</table>

estimators. In all simulations, the sample size is $n = 500$, and the number of replications is 100,000. The results are summarized in Tables 2 and 3. To explore the performance in time series with non-Gaussian innovations, simulations are conducted for AR models with $t_4$ innovations; see Table 4.

From Tables 2, 3 and 4, the RMSE of the Bayes-type estimator is smaller than that of the MLE in all cases. Also, the MLE is more heavy-tailed than the Bayes-type estimator. Note that the efficiency gain from the Bayes-type estimator differs from case to case since the signal-to-noise ratios and serial dependencies are different. Generally, a larger signal-to-noise ratio would lead to less efficiency gain. In the supplementary material [13], additional simulation results, including cases under sample size $n = 1000$ and AR models with centered $\chi_3^2$ innovations, are provided. It is seen that the efficiency
Table 4
Mean, RMSE ($\times 10^{-2}$) and se(RMSE) ($\times 10^{-2}$) of the MLE and the Bayes-type estimator for AR models with $t_4$ innovations of length $n = 500$.

<table>
<thead>
<tr>
<th>$\tau_0$</th>
<th>MLE</th>
<th>Bayes</th>
<th>MLE</th>
<th>Bayes</th>
<th>MLE</th>
<th>Bayes</th>
<th>MLE</th>
<th>Bayes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>Mean</td>
<td>0.230</td>
<td>0.243</td>
<td>0.208</td>
<td>0.206</td>
<td>0.229</td>
<td>0.235</td>
<td>0.202</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>13.30</td>
<td>11.17</td>
<td>3.62</td>
<td>3.12</td>
<td>11.65</td>
<td>9.74</td>
<td>2.31</td>
</tr>
<tr>
<td></td>
<td>se(RMSE)</td>
<td>0.09</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.08</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>0.5</td>
<td>Mean</td>
<td>0.500</td>
<td>0.500</td>
<td>0.506</td>
<td>0.502</td>
<td>0.506</td>
<td>0.503</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>8.53</td>
<td>6.66</td>
<td>2.71</td>
<td>2.27</td>
<td>7.20</td>
<td>5.79</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>se(RMSE)</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>0.7</td>
<td>Mean</td>
<td>0.686</td>
<td>0.682</td>
<td>0.705</td>
<td>0.701</td>
<td>0.698</td>
<td>0.691</td>
<td>0.700</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>10.61</td>
<td>8.25</td>
<td>2.82</td>
<td>2.37</td>
<td>8.52</td>
<td>6.86</td>
<td>2.16</td>
</tr>
<tr>
<td></td>
<td>se(RMSE)</td>
<td>0.07</td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
<td>0.06</td>
<td>0.05</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Gain of the Bayes-type estimator is higher under larger sample sizes.

5.2. Confidence Intervals and Approximate Distributions. This section examines the performance of asymptotic inference and the bootstrap procedures proposed in Sections 3 and 4 respectively. First, we investigate the performance of the bootstrap procedures under AR(1) and GARCH(1,1) models given in (5.1) and (5.2), respectively. For each of the PB and HB procedures, we calculate the coverage probabilities of the 90% confidence intervals with 1000 replications. Following the discussions in Section 4, for each of the bootstrap procedure, we try different values of $m$ and the data-driven choice of $m$ with tolerance level $\gamma = 10^{-5}$. Let $F^{(asym)}$ be the cumulative distribution function of the asymptotic distribution given in Corollary 3.2. Using Monte Carlo simulation, we obtain the 5% and 95% percentiles of the asymptotic distribution, $F^{(asym)}_{5\%}$ and $F^{(asym)}_{95\%}$, respectively. Then, a 90% confidence interval for the change-point is given by $[\hat{k}_n - \hat{s}F^{(asym)}_{95\%}, \hat{k}_n - \hat{s}F^{(asym)}_{5\%}]$, where $\hat{s} = (\hat{\Delta}'\hat{\Sigma}_{20}\hat{\Delta})^{-1}(\hat{\Delta}'\hat{\Omega}_{20}\hat{\Delta})$, $\hat{\Delta} = \hat{\theta}_2 - \hat{\theta}_1$, and $\hat{\Sigma}_{20}$ and $\hat{\Omega}_{20}$ are the corresponding sample estimators of $\Sigma_{20}$ and $\Omega_{20}$. For comparison, the confidence intervals of the MLE approach in [30] are also calculated. The results are summarized in Tables 5 and 6.

From Tables 5 and 6, bootstrap procedures have coverage probabilities close to the nominal level of 90% for large $n$. The bootstrap procedures with the data-driven choice of $m$ perform stably for most cases, especially for small sample sizes, and PB performs slightly better than HB. The asymptotic confidence intervals achieve reasonable performance when the change is small but perform poorly when the change is large. This is in line with Theorem 3.1 that the approximation only works for small change cases. Based
Table 5

Coverage accuracy for 90% confidence intervals of AR(1) models. “—” stands for the inapplicability of HB when the pre-change sequence is shorter than m.

<table>
<thead>
<tr>
<th>(ψ₁, ψ₂)</th>
<th>n</th>
<th>PB</th>
<th>HB</th>
<th>Asy</th>
<th>Ling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Adapted</td>
<td>91.0 %</td>
<td>84.7 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>m= 100</td>
<td>90.0 %</td>
<td>85.9 %</td>
<td>71.7 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 300</td>
<td>91.0 %</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>(0.2, 0.8)</td>
<td></td>
<td>1000</td>
<td>Adapted</td>
<td>90.7 %</td>
<td>89.2 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 100</td>
<td>90.8 %</td>
<td>88.6 %</td>
<td>66.3 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 300</td>
<td>90.6 %</td>
<td>86.8 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>Adapted</td>
<td>89.6 %</td>
<td>89.7 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 100</td>
<td>90.9 %</td>
<td>90.4 %</td>
<td>70.4 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 300</td>
<td>91.0 %</td>
<td>90.6 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adapted</td>
<td>91.2 %</td>
<td>88.2 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 100</td>
<td>91.5 %</td>
<td>88.0 %</td>
<td>84.1 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 300</td>
<td>90.5 %</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>(−0.4, 0.4)</td>
<td>1000</td>
<td>Adapted</td>
<td>92.7 %</td>
<td>90.5 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 100</td>
<td>90.4 %</td>
<td>88.1 %</td>
<td>85.4 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 300</td>
<td>91.1 %</td>
<td>88.3 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>Adapted</td>
<td>92.0 %</td>
<td>91.5 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 100</td>
<td>91.1 %</td>
<td>90.8 %</td>
<td>86.5 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 300</td>
<td>91.8 %</td>
<td>90.8 %</td>
<td></td>
</tr>
</tbody>
</table>

Table 6

Coverage accuracy for 90% confidence intervals of GARCH(1,1) models. “—” stands for the inapplicability of HB when the pre-change sequence is shorter than m.

<table>
<thead>
<tr>
<th>(θ₁, θ₂)</th>
<th>n</th>
<th>PB</th>
<th>HB</th>
<th>Asy</th>
<th>Ling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Adapted</td>
<td>83.6 %</td>
<td>78.1 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>m= 100</td>
<td>85.2 %</td>
<td>78.0 %</td>
<td>60.2 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 300</td>
<td>85.5 %</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>(θ₁ = (0.1, 0.2, 0.4) , θ₂ = (0.05, 0.4, 0.1))</td>
<td>1000</td>
<td>Adapted</td>
<td>85.0 %</td>
<td>83.7 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 100</td>
<td>86.6 %</td>
<td>85.5 %</td>
<td>65.0 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 300</td>
<td>87.4 %</td>
<td>82.9 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>Adapted</td>
<td>88.1 %</td>
<td>88.1 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 100</td>
<td>85.8 %</td>
<td>87.1 %</td>
<td>64.7 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 300</td>
<td>88.0 %</td>
<td>87.2 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adapted</td>
<td>86.8 %</td>
<td>87.2 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 100</td>
<td>82.3 %</td>
<td>82.7 %</td>
<td>58.8 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 300</td>
<td>82.1 %</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>(θ₁ = (3, 0.1, 0.5) , θ₂ = (0.5, 0.1, 0.5))</td>
<td>1000</td>
<td>Adapted</td>
<td>87.1 %</td>
<td>87.7 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 100</td>
<td>89.5 %</td>
<td>88.8 %</td>
<td>55.1 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 300</td>
<td>89.1 %</td>
<td>88.1 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>Adapted</td>
<td>89.7 %</td>
<td>90.9 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 100</td>
<td>89.8 %</td>
<td>90.0 %</td>
<td>53.1 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>m= 300</td>
<td>89.9 %</td>
<td>90.1 %</td>
<td></td>
</tr>
</tbody>
</table>
on the results above, the data-driven procedures are recommended and will be employed in the following simulations.

In the supplementary material [13], three additional parameter combinations for AR and GARCH models are included. Simulations under model misspecifications are also provided. It is seen that PB is not robust against model misspecification. On the other hand, HB is robust to model misspecification in the pre-change model. Moreover, even though HB is not robust to the model misspecification in the post-change model, it still performs better than PB when both pre- and post-change models are misspecified.

5.3. Multiple Change-points by GLRSM. The bootstrap procedures for the Bayes-type estimator of a single change-point can be extended to multiple change-points through the generalized likelihood ratio scan method (GLRSM) proposed by [34]. The GLRSM, a generalization of the likelihood ratio scan method (LRSM) proposed by [45] from piecewise AR models to general short-memory time series models, is a computationally efficient procedure for multiple change-points inference. The GLRSM involves three steps: the first step detects potential change-points by scanning statistics. The second step obtains consistent change-point estimators by a model selection approach. The final step builds an “extended local window” around each estimated change-point. Each extended local window is shown to contain the true change-point in its middle half, so that the final change-point estimate and the approximate confidence interval can be obtained by maximum likelihood estimation over each extended local window. Here, we replace the last step by using the Bayes-type estimator and the bootstrap procedures proposed in Section 4 to provide improved final estimates and confidence intervals. Since GLRSM transforms multiple change-points problems into localized single change-point problems, the asymptotic optimality established in Theorem 2.2 can be applied to the change-point estimator with respect to each extended local window under the weak dependence condition between segments. Nevertheless, it remains an open issue if the proposed procedure is globally optimal, in the sense that the maximal risk accumulated over all change-points is minimized. The latter problem involves delicate analysis of the effect of the estimated number of change-points on the risk function, and will be investigated in future research.

We use the following models as in [34]: Piecewise stationary AR(1) model

$$X_t = \begin{cases} 0.4X_{t-1} + \epsilon_t, & \text{if } 1 \leq t \leq 400, \\ -0.6X_{t-1} + \epsilon_t, & \text{if } 401 \leq t \leq 700, \\ 0.5X_{t-1} + \epsilon_t, & \text{if } 701 \leq t \leq 1000, \end{cases}$$

and Piecewise stationary GARCH model
\[ X_t = \epsilon_t \sigma_t, \quad \sigma_t^2 = \begin{cases} 3 + 0.1X_{t-1}^2 + 0.5\sigma_{t-1}^2, & \text{if } 1 \leq t \leq 400, \\ 0.5 + 0.1X_{t-1}^2 + 0.5\sigma_{t-1}^2, & \text{if } 401 \leq t \leq 1600, \\ 0.8 + 0.1X_{t-1}^2 + 0.8\sigma_{t-1}^2, & \text{if } 1601 \leq t \leq 2000. \end{cases} \]

Table 7

<table>
<thead>
<tr>
<th>Model</th>
<th>( k_0 )</th>
<th>PB</th>
<th>HB</th>
<th>Ling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piecewise AR(1)</td>
<td>400</td>
<td>88.5%</td>
<td>86.0%</td>
<td>82.5%</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>90.8%</td>
<td>87.8%</td>
<td>84.8%</td>
</tr>
<tr>
<td>Piecewise GARCH(1,1)</td>
<td>400</td>
<td>87.6%</td>
<td>89.1%</td>
<td>79.0%</td>
</tr>
<tr>
<td></td>
<td>1600</td>
<td>91.6%</td>
<td>90.6%</td>
<td>77.5%</td>
</tr>
</tbody>
</table>

For confidence intervals, we also provide the MLE based confidence interval by [30]. It is found that GLRSM correctly identifies the number of change-points for all replications. The results of change-point estimates are summarized in Table 7. It can be seen from Table 7 that all bootstrap coverage accuracies are close to 90%, which is more accurate than the MLE approach proposed by [30].

6. Real Data Applications. In this section we analyze two real data examples using the proposed estimator and bootstrap procedures.

6.1. Array Comparative Genomic Hybridization (CGH) Data. We apply the Bayes-type estimator to array comparative genomic hybridization (CGH) data, which shows aberrations in genomic DNA. The observations consist of the log-ratios of normalized intensities from disease and control samples. The statistical problem is to identify regions on which the ratio differs significantly from 0, and can be regarded as a change-point problem in an i.i.d. Gaussian sequence. Independent Gaussian or quasi-Gaussian modeling has been adopted in [23] and [35]. The independence can be justified by the ACF plots provided in Figure 1 in the supplementary material [13].

We apply the proposed estimation procedure to a widely used data set GBM31 containing the array CGH profile of chromosome 13 (see also [23],
Allowing both mean and variance change, the Bayes-type change-point estimate is 548, and the 90% PB and HB confidence intervals are [534, 560] and [534, 557], respectively. The estimated locations are in accordance with the results in [23] and [28]. In terms of confidence intervals, PB and HB give similar results. Figure 1 depicts the GBM31 data and the estimated location with a 90% confidence interval constructed by PB.

6.2. Interhemispheric Gradient of $\Delta^{14}C$ Data. The $\Delta^{14}C$ of CO$_2$ (abbreviated as $\Delta^{14}C$) is a widely used tracer of past climate changes for both the ocean and the atmosphere. We analyze the interhemispheric gradient of $\Delta^{14}C$ over the period 850–1830 covering the Medieval Climate Anomaly ($\approx$950–1250) and the Little Ice Age ($\approx$1500–1800) using INTCAL04 [36] and SHCAL04 [32] tree-ring data. The data were recorded every 5 years.

Suggested by [5] and [38], AR(2) is employed to model both the pre-change and post-change sequences. Prior to the estimation procedure, the change-point test proposed by [16] is performed, giving a $p$-value less than $1 \times 10^{-4}$, which is a strong evidence for the existence of a change-point. Then, the Bayes-type change-point estimate is found to be 1450, and PB and HB generate the same 90% confidence interval [1420, 1485]. The estimated change-point is close to the result of [5], which is 1455, obtained from a SIC selection. However, a closer look at the ACF plots of the residuals in Figure 2 in [13] suggests that AR(2) modeling may not be appropriate. Moreover, the $p$-value of the Box-Ljung test using the first 20 autocorrelations on the pre-change residuals is 0.00013, indicating departures from independence. Under careful investigations, we propose to use AR(4) instead of AR(2).

An analysis is re-conducted under AR(4) modeling. A likelihood ratio test on the existence of change-points still outputs a $p$-value close to 0. The Bayes-type change-point estimate is 1545, and the 90% PB and HB confidence intervals are [1515, 1555] and [1515, 1570], respectively; see Figure 1. The residual ACF plots provided in Figure 3 in [13] suggest that AR(4) is adequate for modeling both the pre-change and post-change sequences. Adequacy is also validated by the results of the Box-Ljung test for the first 20 autocorrelations, in which the $p$-values for the pre-change and post-change residuals are 0.542 and 0.9807, respectively.

The change-point estimate under AR(4) modeling is 1545, which is different from the results of 1455 in [5] and 1375 in [38]. Moreover, all of the confidence intervals fall outside of the usual recognized transition period ($\approx$1250–1500, see [5]). Hence, the question is raised concerning the original period identification between the Medieval Climate Anomaly and the Little Ice Age. Further scientific investigations are needed to clarify this issue.
7. Technical Assumptions and Lemmas. In this section, technical assumptions and lemmas are discussed in detail. Illustrative examples are presented in Appendix F of the supplementary material [13].

7.1. Notation. The i.i.d. innovations \( \{n_t\}_{t \in \mathbb{Z}} \) take values in a measurable space \((\mathbb{R}, \mathcal{B})\). The space \(\mathbb{R}^\infty\) is the subset of \(\mathbb{R}^N\) of finitely non-zero sequences \(\{u_k\}_{k>0}\) such that there exists \(N > 0\) with \(u_k = 0\) for \(k > N\). Let \(\mathbb{R}\) be endowed with its Borel \(\sigma\)-algebra \(\mathcal{B}\); then \(\mathbb{R}^\infty\) is considered together with its product \(\sigma\)-algebra \(\mathcal{B}^\otimes\). The function \(f\) in (2.1) is assumed to be a measurable function from \(\mathbb{R}^d \times \mathbb{R}^\infty \times \mathbb{R}\) with values in \(\mathbb{R}\). Let \(\| \cdot \|\) denote the Euclidean norm of \(\mathbb{R}\). Moreover, \(\| \cdot \|_p\) denotes the usual \(\mathbb{L}^p\)-norm, i.e., \(\|Y\|_p = E|Y|^p\) for \(p \geq 1\) for every \(\mathbb{R}\)-valued random variable \(Y\). Denote \(E_k\) as the expectation operator evaluated under a probability measure \(P_k\) under which \(\{x_t\} \sim M_k(\theta_{10}, \theta_{20})\) where an abrupt change occurs at time \(k\).

Recall that \(E_{\theta_0}(\cdot)\) is the expectation under which \(\{x_t\} \sim M(\theta_0)\). When the underlying distribution is clear in the content, the subscript \(\theta_0\) is omitted. Let \(\hat{l}_t(\theta) := \partial l_t(\theta)/\partial \theta\) and \(\hat{\partial}_l(\theta) := \partial^2 l_t(\theta)/\partial \theta^2\). For notational simplicity, we say that a result holds for \(h(x_t|\theta, X_{t-1})\) if the result holds for any of the functions \(l_t(\theta), \partial l_t(\theta)/\partial \theta_i\) or \(\partial^2 l_t(\theta)/\partial \theta_i \partial \theta_j\) for \(i,j \in \{1, \ldots, d\}\). Also, let \(h_q(x_t|\theta, X_{t-1}) = h(x_t|\theta, \{x_{t-1}, x_{t-2}, \ldots, x_{t-q}, 0, 0, \ldots\})\) be the truncated version of \(h(x_t|\theta, X_{t-1})\). The symbol 0 denotes a null vector.

7.2. Assumptions and Lemmas. The following weak dependence structure for the stationary processes \(\{x_t^{(i)}\}_{t \in \mathbb{Z}}\) for \(i = 1, 2\), and regularity assumption for the objective function \(l_t(\theta)\) are imposed for asymptotic theory.

Assumption 1. There exists some \(p^* > p + \nu\) with \(p > 3\) and \(\nu > 0\) such that both \(\{x_t^{(1)}\}_{t \in \mathbb{Z}} \sim M(\theta_{10})\) and \(\{x_t^{(2)}\}_{t \in \mathbb{Z}} \sim M(\theta_{20})\) are stationary \(\alpha\)-mixing sequences with mixing rate \(\alpha_x(n) = O\left(n^{-(p^*(p+\nu-1))/(p^*-(p+\nu))}\right)\).

The decay rate of \(\alpha\)-mixing coefficients stated in Assumption 1 satisfies the weak dependence condition for the convergence rate of the strong law of large numbers described in [40]. Specifically, it implies complete convergence of the sequences \(\{h(x_t|\theta, X_{t-1})\}\) in the sense that for both \(\{x_t\}_{t \in \mathbb{Z}} \sim M(\theta_{10})\) and \(\{x_t\}_{t \in \mathbb{Z}} \sim M(\theta_{20})\), there exists a \(p > 3\) such that

\[
\sum_{m=0}^{\infty} m^{p-2}P\left(\max_{j \leq m} \left\{ \sum_{t=1}^{j} [h(x_t|\theta, X_{t-1}) - E(h(x_t|\theta, X_{t-1}))]\right\} > Cm\right) < \infty,
\]

which follows from the maximal moment inequality for partial sums of sequences \(\{h(x_t|\theta, X_{t-1})\}\). It is possible that the NED weak dependence condition with suitable moment condition described in [29] or the weak dependence condition derived by the physical and predictive dependence measures
proposed in [44] imposed on sequences \( \{h(x_t|\theta, X_{t-1})\} \) can replace the weak
dependence condition in Assumption 1. Many commonly used time series
models fulfill this mixing condition. For example, Theorem 1 of [33] showed
that stationary ARMA models are geometric \( \alpha \)-mixing, i.e., the mixing co-

efficients \( \alpha(n) = O(a^n) \) for some \( 0 < a < 1 \), provided the innovations \( \{\eta_t\} \)
have absolutely continuous distribution with respect to Lebesgue measure,
see [18] for more details. Theorem 3.4.2 of [10] proved that GARCH models
are geometric \( \alpha \)-mixing, under the assumption that the innovations \( \{\eta_t\} \)
have absolutely continuous distribution with respect to Lebesgue measure, with
a positive density in a neighborhood of zero. See [11] for other conditions on
the innovations for the geometric \( \alpha \)-mixing of GARCH models.

**Assumption 2.** \( \mathbb{E}_{\theta_0}(\sup_{\theta \in \Theta} |l_t(\theta)|) < \infty \), and \( \mathbb{E}_{\theta_0}(l_t(\theta)) \) has a unique max-
imizer at \( \theta = \theta_0 \).

Note that in Assumption 2, \( \{x_t\} \sim M(\theta_0) \) is a stationary process and
hence \( \{l_t(\theta)\} \) is also a stationary process, therefore \( \mathbb{E}_{\theta_0}(l_t(\theta)) \) and its unique
maximizer do not depend on \( t \).

The objective function \( l_t(\theta) = l(x_t|\theta, \{x_{t-1}, x_{t-2}, \ldots\}) \), which involves
the infinite past \( \{x_t\}_{t < \tau} \), may not be strong-mixing even though the data
sequence \( \{x_t\} \) is strong mixing. On the other hand, its truncated counter-
part \( l_{t,q}(\theta) = l(x_t|\theta, \{x_{t-1}, x_{t-2}, \ldots, x_{t-q}, 0, 0, \ldots\}) \) is strong-mixing. The
following assumption provides some Lipschitz-type conditions on the ob-
jective function and its derivatives. These conditions state that the effects
of \( \{x_t\}_{t < \tau} \) on the function \( h(x_t|\theta, X_{t-1}) \) diminish when \( |t - t'| \to \infty \) at
a rate depending on the Lipschitz coefficients, and help establish closeness
between the objective functions and their truncated counterparts in Lemma
7.1. Hence, \( l_t(\theta) \) behaves similarly to a strong-mixing sequence.

**Assumption 3.** For all \( \theta \in \Theta \), \( \hat{l}_t(\theta) \) and \( \bar{l}_t(\theta) \) exist almost surely, and
there exist some \( r \in \mathbb{N} \), a sequence of positive numbers \( \{\alpha_j^{(h)}\}_{j \in \mathbb{N}} \) with \( \alpha_j^{(h)} = O(j^{-v}) \) for some \( v > 2 \), and some positive functions \( \{\psi_k^{(h)}\}_{k=1, \ldots, r} \), such that

\( \text{(i) } \sup_{\theta \in \Theta} |h(0|\theta, 0)| < \infty. \)
\( \text{(ii) For any } U_t = (u_t, u_{t-1}, \ldots) \text{ and } V_t = (v_t, v_{t-1}, \ldots) \in \mathbb{R}^\infty, \)

\[
\sup_{\theta \in \Theta} |h(u_t|\theta, U_{t-1}) - h(v_t|\theta, V_{t-1})| \leq \sum_{k=1}^r \left( \sum_{j=1}^\infty \alpha_j^{(h)} |u_{t-j+1} - v_{t-j+1}| \right)^k \psi_k^{(h)}(V_t),
\]

where \( U_{t-1} = (u_{t-1}, u_{t-2}, \ldots) \) and \( V_{t-1} = (v_{t-1}, v_{t-2}, \ldots) \).
(iii) For \( k = 1, \ldots, r \), \( \Psi_k^{(h)} \) satisfies \( \sup_{j \in \mathbb{N}} \mathbb{E}(\Psi_k^{(h)}(X_t^{(1)})^p | x_{t-j+1}^{(1)}|^{kp^*}) < \infty \) and \( \sup_{j \in \mathbb{N}} \mathbb{E}(\Psi_k^{(h)}(X_t^{(2)})^p | x_{t-j+1}^{(2)}|^{kp^*}) < \infty \) for \( i = 1, 2 \), where \( \{x_t^{(1)}\} \sim M(\theta_{10}) \) and \( \{x_t^{(2)}\} \sim M(\theta_{20}) \) are generated using the same set of noises \( \{\eta_t\}_{t \in \mathbb{Z}} \), \( X_t^{(1)} = (x_t^{(1)}, x_{t-1}^{(1)}), \) and \( p^* \) is defined in Assumption 1.

**Remark 7.1.** The Lipschitz-type conditions on \( l_t(\theta), \tilde{l}_t(\theta) \) and \( \tilde{\theta}_t(\theta) \) are not restrictive. They are fulfilled by many popular time series models, including ARMA \( (r = 2) \) and GARCH \( (r = 1) \) models; see Appendix F of the supplementary material [13]. Similar conditions are verified for a large class of causal processes satisfying Assumption 4 below in [4] to establish the strong consistency and asymptotic normality of the QMLE.

**Lemma 7.1.** Under Assumption 3, there exists an \( r \in \mathbb{N} \), a sequence of positive numbers \( \{\alpha_j^{(h)}\}_{j \in \mathbb{N}} \) with \( \alpha_j^{(h)} = O(j^{-v}) \) for some \( v > 2 \), and some positive functions \( \{\Psi_k^{(h)}\}_{k=1, \ldots, r} \) such that for any \( q \in \mathbb{N} \),

\[
\sup_{\theta \in \Theta} |h(x_t|\theta, X_{t-1}) - h_q(x_t|\theta, X_{t-1})| \leq \sum_{k=1}^{r} \left( \sum_{j=q+1}^{\infty} \alpha_j^{(h)} |x_{t-j}|^{kp^*} \right)^k \Psi_k^{(h)}(x_t, \ldots, x_{t-q}) \text{ a.s.}
\]

where \( \sup_{j \geq q+1} \mathbb{E}_{\theta_0}(\Psi_k^{(h)}(x_t, \ldots, x_{t-q})^p | x_{t-j}|^{kp^*}) < \infty \) for \( \{x_t\} \sim M(\theta_0) \), for \( k = 1, \ldots, r \), and \( p^* \) is defined in Assumption 1.

Recall that the post-change sequence \( \{x_t\}_{t > k_0} \) is not stationary since \( x_{t-1} = (x_{t-1}, \ldots, x_{k_0+1}, x_{k_0}, x_{k_0-1}, \ldots) \) involves \( \{x_{k_0}, x_{k_0-1}, \ldots\} \sim M(\theta_{10}) \). Note that \( X_{k_0} = \{x_t\}_{t \leq k_0} = X_{k_0}^{(1)} \sim M(\theta_{10}) \) and \( X_{k_0}^{(2)} = \{x_t^{(2)}\}_{t \leq k_0} \sim M(\theta_{20}) \) are stationary mixing sequences. For \( t > k_0 \), we can view \( x_t \) and \( x_t^{(2)} \) as

\[
x_t = g(\theta_{20}, \eta_{t}, \ldots, \eta_{k_0+1}, X_{k_0}^{(1)}), \quad x_t^{(2)} = g(\theta_{20}, \eta_{t}, \ldots, \eta_{k_0+1}, X_{k_0}^{(2)}).
\]

Hence, the difference between \( x_t \) and \( x_t^{(2)} \) is solely due to the initial values \( X_{k_0}^{(1)} \) and \( X_{k_0}^{(2)} \). To be precise, approximation of the non-stationary \( \{x_t\}_{t > k_0} \) by the stationary counterpart \( \{x_t^{(2)}\}_{t > k_0} \) can be viewed as a problem related to recursive approximation of the stationary measure studied in [9] and Proposition 3.1 in [19]. Let \( g_m : \mathbb{R}^{m-1} \rightarrow \mathbb{R} \) be the random function defined as \( u \mapsto f(\theta_{20}, (u, X_{k_0}^{(1)}), \eta_{k_0+m}) \) for each \( m \geq 2 \). Then the non-stationary
\{x_t\}_{t>k_0} can be expressed as \(x_{k_0+1} = g_1 = f(\theta_{20}, X_{k_0}^{(1)})\), and recursively for \(m \geq 2\), \(x_{k_0+m} = g_m(x_{k_0+m-1}, \ldots, x_{k_0+1})\).

The following Lipschitz-type condition on the function \(f\) in (2.1) is imposed so that we can quantify the difference between the non-stationary \(\{x_t\}\) and the stationary counterpart \(\{x_t^{(2)}\}\).

**Assumption 4.** There exists a sequence of positive numbers \(\{\beta_j\}_{j \in \mathbb{N}}\) with \(\beta_j = O(j^{-\mu})\) for some \(\mu > 2\) such that for all \(U_{t-1} = (u_{t-1}, u_{t-2}, \ldots)\) and \(V_{t-1} = (v_{t-1}, v_{t-2}, \ldots)\) in \(\mathbb{R}^\infty\) and \(\theta \in \{\theta_{10}, \theta_{20}\}\),

\[
\|f(\theta, U_{t-1}, \eta_t) - f(\theta, V_{t-1}, \eta_t)\|_{rp^*} \leq \sum_{j=1}^{\infty} \beta_j |u_{t-j} - v_{t-j}|,
\]

where \(r\) and \(p^*\) are defined in Assumption 3, innovations \(\eta_t\) have distribution function \(\mathbb{F}_0\) with \(|\eta_t|_{rp^*} < \infty\), \(\beta = \sum_{j=1}^{\infty} \beta_j < 1\), and \(\|f(\theta, 0, \eta_t)\|_{rp^*} < \infty\).

**Lemma 7.2.** Under Assumption 4, let \(\{x_t^{(2)}\} \sim M(\theta_{20})\) and \(\{x_t\} \sim M_{k_0}(\theta_{10}, \theta_{20})\) be generated using the same set of noises \(\{\eta_t\}_{t \in \mathbb{Z}}\). Then for any \(m \in \mathbb{N}\),

\[
\|x_{k_0+m}^{(2)} - x_{k_0+m}^{(2)}\|_{rp^*} \leq \frac{\|x_{k_0}^{(1)} - x_{k_0}^{(2)}\|_{rp^*}}{1 - \beta} \inf_{1 \leq q \leq m} \left( \beta^q + \frac{1}{1 - \beta} \sum_{i=q+1}^{\infty} \beta_i \right)
\leq C \left( \frac{\log m}{m} \right)^{\mu - 1}, \text{ for some } C > 0.
\]

Note that the sequence \(\{\beta_j\}_{j \in \mathbb{N}}\) only depends on the functional form of \(f\), and it does not depend on the true change-point \(k_0\). The Lipschitz-type condition in Assumption 4 is satisfied by a large class of time series models including ARMA and GARCH models, see Appendix F of the supplementary material [13]. The same condition can also be found in [19] to establish the existence of a strictly stationary solution and its weak dependence properties, and [4] to establish the asymptotic properties of the QMLE. As a consequence of the above assumptions, the following lemmas give bounds on the aggregate effect of the pre-change observations on the objective function and its derivatives, so that the non-stationary \(\{h(x_t|\theta, X_{t-1})\}_{t>k_0}\) can be approximated by its stationary counterpart \(\{h(x_t^{(2)}|\theta, X_{t-1}^{(2)})\}_{t>k_0}\).

**Lemma 7.3.** Under Assumptions 3 and 4, for any \(t > k_0\), \(\theta \in \Theta\), let \(R_t(\theta) = h(x_t|\theta, X_{t-1}) - h(x_t^{(2)}|\theta, X_{t-1}^{(2)})\),
where \(\{x^{(2)}_t, X^{(2)}_{t-1}\} \sim M(\theta_{20})\) and \(\{x_t, X_{t-1}\} \sim M_{k_0}(\theta_{10}, \theta_{20})\) are generated using the same set of noises \(\{\eta_t\}_{t \in \mathbb{Z}}\). Then, for all \(\epsilon > 0\) and \(\theta \in \Theta\),

\[
\sum_{m=1}^{\infty} m^{p-2} P_{k_0} \left( \max_{j \leq m} \left| \sum_{t=k_0+1}^{k_0+j} R_t(\theta) \right| > \epsilon m \right) < \infty,
\]

for \(h(x_t|\theta, X_{t-1})\) being \(l_t(\theta), \partial l_t(\theta)/\partial \theta_i\) and \(\partial^2 l_t(\theta)/\partial \theta_i \partial \theta_j, i, j \in \{1, \ldots, d\}\), where \(p\) is defined in Assumption 1.

**Lemma 7.4.** Under Assumptions 3 and 4, for all \(z \in \mathbb{R}\) and \(\theta \in \Theta\),

\[
\sum_{t=k_0+1}^{k_0+z/\Delta} R_t(\theta) \rightarrow 0 \text{ in } P_{k_0}-\text{probability as } \Delta \rightarrow 0.
\]

**8. Conclusion.** This paper establishes the asymptotic theory for the Bayes-type estimator of change-points in weak dependence (mixing) time series models. It is shown that if the objective function is taken as the log-likelihood function, then the Bayes-type estimator is an asymptotic minimax estimator with respect to the squared error loss. Furthermore, an approximate asymptotic distribution and two bootstrap procedures are developed for constructing confidence intervals for the change-point. Applications to multiple change-points problems are also demonstrated.

**Acknowledgements.** We would like to thank the co-Editor, an Associate Editor and the anonymous referees for their critical comments and thoughtful suggestions, which lead to a much improved version of this paper. Research supported in part by grants from HKSAR-RGC-GRF Nos 14308218 and 14325216 and HKSAR-RGC-TRF No. T32-101/15-R (Chan), and HKSAR-RGC-GRF Nos 14302719 and 14305517 (Yau).

**References.**


