12. E. Kant, Understanding and automatic algorithm design, *IEEE Trans. Software Eng.* SE-11(11):1361–1374 (1985).
13. J. R. Low, Automatic data structure selection: An example and overview, *Commun. ACM* 21(5):376–385 (1978).
14. C. V. Ramamoorthy, S. B. Ho, and W. T. Chen, On the automated generation of program test data, *IEEE Trans. Software Eng.* SE-2(4):293–300, 1976.
15. M. Shishibori, K. H. Park, and J. Aoe, A method of determining key search algorithms using classification knowledge, in *Proc. of the 45th Conf. on Information Processing of Japan*, 1992, pp. D-363–D-366.
16. J. Iwasawa and K. Nakamori, The algorithm animation for algorithm research in environment, *Information Algorithm Research Committee of Japan* 32(10):73–80 (1993).
17. D. M. Steier and E. Kant, The roles of execution and analysis in algorithm design, *IEEE Trans. Software Eng.* SE-11(11):1375–1386 (1985).
18. X. Tokunaka, X. Okumura, and X. Tanaka, The induction of viewpoint to conceptual label, *Information Research Committee of Japan* 30(8):970–975 (1989).

Received 6 March 1994; revised 24 July 1994, 30 November 1994

NORTH-HOLLAND

Time Series Segmentation: A Sliding Window Approach

CHIA-SHANG JAMES CHU

Department of Economics, University of Southern California, Los Angeles, California 90089

ABSTRACT

The aim of this paper is to present two on-line, sliding window segmentation algorithms. Detection nonstationarity is based on parameter fluctuations and change point localization of the Akaike information criterion. Asymptotic properties of the proposed algorithms are analyzed. Specifically, the limiting distributions are derived and the asymptotic threshold values are tabulated for future reference. Finite sample simulations are performed to illustrate the usefulness of these algorithms.

1. INTRODUCTION

The problem of segmentation of piecewise stationary time series frequently arises in econometric modeling. The parameters in an econometric model are often subject to shift due to policy changes. For example, the exchange rate volatility may depend on U.S. monetary policy regimes, that is, nonstationarity of the variance in exchange rate time series. Such parameter shifts, if unaccounted for, will bias parameter estimates and forecasts. Similar problems arise in many applied sciences such as speech recognition system analysis and biomedical signal processing (e.g., electroencephalograms). Engineering and statistics literature in this field are quite extensive; see [6, 29, 31], *Handbook of Statistics* (edited by Krishniah and Sen), and references therein.

Segmentation consists of two interdependent steps: detecting nonstationarity and localizing change points. Generally, there are two types of segmentation methods, methods based on one model and methods based on two models. The one-model method (perhaps better known as the residual-based approach) bases segmentation on significant deviations of the residuals from its behavior-should-be under the null model. Segmentation of this kind only requires estimation of the null model. Due to its relative simplicity in computation, the one-model segmentation procedures have

INFORMATION SCIENCES 85, 147–173 (1995)

655 Avenue of the Americas, New York, NY 10010

0020-0255/95/$9.50
SSDI 0020-0255(95)00021-G

been widely used in practice. Examples include the CUSUM segmentation of Brown *et al.* [11] and its variants such as the MOSUM (moving sum of recursive residuals) of Baul and Hackl [8], later resumed by Chu *et al.* [12] and the OLS-CUSUM (cumulative sum of ordinary least squares residuals) of Ploberger and Kramer [24]. The main drawback of these segmentation procedures is the power deficiency against alternatives in which the parameter jump is orthogonal to the mean regressors [25, 12].

On the other hand, the two-models approach bases segmentation on some convenient measures of differences between two models, which we refer to as the reference model (hypothetical model before the change) and test model (the model after the change). Two questions underlying the two-models segmentation arise: first, how to identify the reference and test models; second, how to measure the differences between the two models. Several statistics have been suggested to measure the difference between the two models, such as the maximal generalized likelihood ratio (max-GLR) of Appel and Brandt [4] and Quandt [27], Chernoff's distance as used in Basseville and Benveniste [6], maximal Wald statistic (max-W) of Hawkins [18], maximal Lagrange multiplier statistic (max-LM) of Andrews [2], quadratic mean of the difference between the two spectra of Bodenstein and Praetorius [10], and the fluctuation statistic of Sen [28] and Ploberger *et al.* [23].

The first question arises because the model before and after change is typically unknown, and has to be identified from the reference and test window. There are many possible selections concerning the location of the reference and test window in the literature; most of them can be classified into the following four identification schemes (see Figure 1):

(1a) A fixed reference window and sliding test windows, e.g., Bodenstein and Praetorius [10], in which the reference model is identified in a fixed window of size, say h, and a sequence of test models is identified from sliding test windows of the same size.

(1b) Growing reference windows and sliding test windows, e.g., Appel and Brandt [4] and Basseville and Benveniste [6], in which reference models are updated in such a way that the reference window is always adjacent to the test window of size h.

(1c) A fixed (global) reference window and growing test windows, e.g., Ploberger *et al.* [23], in which the reference model is estimated from the whole sample and a sequence of test models is identified from test windows of increasing size.

(1d) Growing reference windows and shrinking test windows, e.g., Andrews [2] and Deshayes and Picard [16], in which a sample split point k is used to determine the location of the reference window (containing

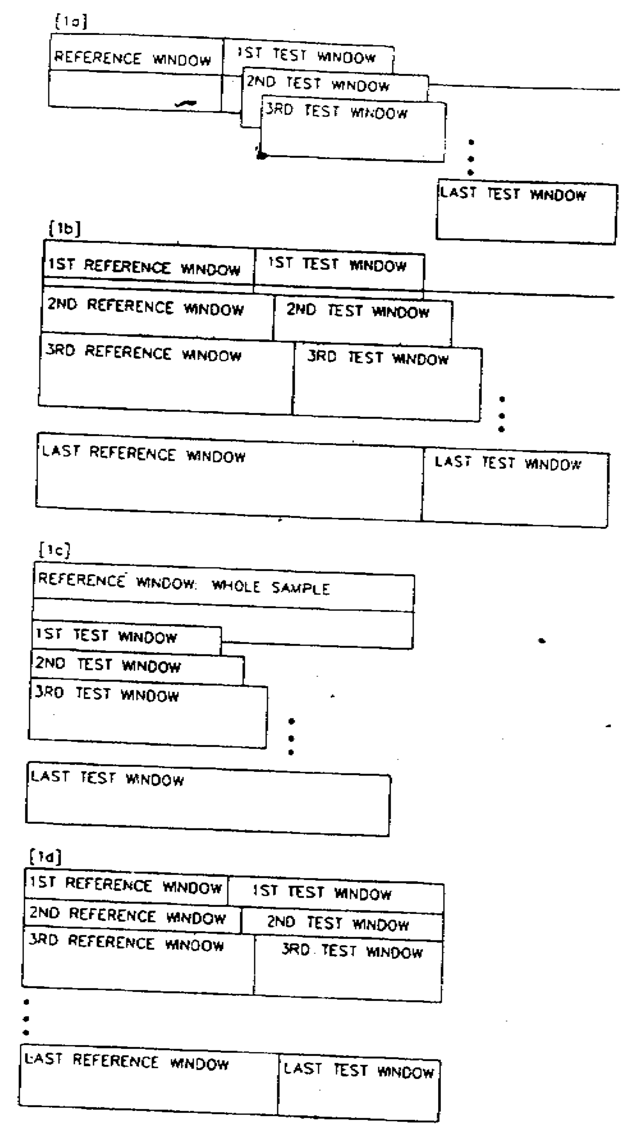

Fig. 1. Identification schemes.

all the pre-k observations) and the test window (containing all the post-k observations) such that the size of reference and test windows always sums up to the size of the whole sample.

The fundamental difference is that (1a) and (1b) permit on-line segmentation, while (1c) and (1d) do not. Since the change point localization depends very much on the structure of the nonstationarity detector, the choice concerning on-line or off-line is crucial. Specifically, in what way should we implement a detection procedure—the first step in segmentation?, on-line or off-line? In light of the results in the sequential tests literature, off-line detection is more powerful than the on-line detection essentially because they use all of the data that will (ever) be collected. However, there is a tradeoff, namely, the off-line detector offers a less appealing procedure in change point localization.

Off-line detection procedures based on identification scheme (1c) and (1d) are well known in the econometrics and statistics literature; the max-GLR, max-W, and max-LM statistic use (1d), while the fluctuation statistic uses (1c). Identification scheme (1d) is well motivated by the alternative hypothesis of a one-time parameter shift at an unknown point of time. This alternative apparently excludes the possibility of multiple change points; hence, it may appear questionable in practice to implement the max-GLR, max-W, or max-LM detector because there can be more than one change point. When multiple change points are present, these tests are potentially insensitive to multiple changes because multiple changes could be "absorbed" in the estimates of the variance of the process. However, from a detection point of view alone, the one-time shift alternative is harmless since it is proved by Andrews [2] that the max-GLR, max-W, and max-LM detectors all have nontrivial asymptotic power against quite general alternatives, including a multiple change points alternative. The same conclusion also holds for the fluctuation test. These celebrated results together with certain optimal properties of off-line detectors [16, 3, 7] seem overwhelmingly favorable to an off-line approach.

The structure of the maximal type detectors only permits a maximum likelihood estimator of a single change point; they can be rather misleading if multiple change points are present.[1] Motivated by the intuition that on-line identification such as (1a) or (1b) appears more promising in change point localization, applied researchers have been using some segmentation algorithms that basically borrow the structure of the off-line detector and implement it in an on-line fashion known as the sliding window approach. Although some success is documented, to the best of the author's knowledge, there is no formal analysis of the statistical properties of segmentation based on a sliding test windows approach.

[1]In the econometrics and statistics literature, most of the previous work on change point localization assumes single change point alternative, cf. [5, 16, 19, 21].

This paper investigates asymptotic properties of segmentation that use fluctuation statistic coupled with sliding test windows, i.e., identification schemes (1a) and (1b). The fluctuation statistic is studied because it leads to a relatively simple asymptotic theory. The difficult question of optimality is not addressed here, in part because we encounter an inherent tradeoff between powerful detection and accurate change point localization. This paper is organized as follows. In Sections 2.1 and 2.2, I derive the asymptotic distributions of the detection procedures based on the sliding fluctuation test for a simple autoregressive model. Extensions to pth order autoregressive models are considered in Section 3.1. Based on a minimized Akaike information criterion, a procedure of localizing change points is suggested in Section 3.2. Section 4 contains various simulation results, and Section 5 ends the paper with concluding remarks. All proofs are contained in the Appendix.

2. AUTOREGRESSIVE PROCESS OF ORDER ONE

2.1. *A FIXED REFERENCE WINDOW AND SLIDING TEST WINDOWS*

We consider first the fluctuation detector with identification scheme (1a) for a simple AR(1) process. We shall refer to the segmentation procedure based on a fixed reference window and sliding test windows as an FSW procedure. In the next section, we take on identification scheme (1b), and use the GSW procedure to denote segmentation using growing reference windows and sliding test windows. Let $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $t = 1, 2, \ldots, T$, where $\{\varepsilon_t\}$ is a white noise process. The prototype of the change detection problem can be formulated as $H_0\colon \alpha_t = \alpha_0$, for all t; $|\alpha_0| < 1$. A standard one-time change alternative hypothesis is that $H_1\colon \alpha_t = \alpha_1$ for $t = 1, 2, \ldots, r$, $|\alpha_1| < 1$; and $\alpha_t = \alpha_2$ for $t = r+1, \ldots, T$; $|\alpha_2| < 1$, where r is an unknown change point.

One can perform the likelihood ratio test of H_0 against H_1. Typically, the parameters before and after change are unknown, and are estimated using, say, the maximum likelihood principle. It can be shown [2, 16] that the likelihood ratio process indexed by the sample split point (i.e., identification scheme (1d)) has a well-defined limiting distribution in the space of continuous functions $C[0,1]$, known as the tied down Bessel process. Moreover, it is well known that max-GLR, max-W, and max-LM are asymptotically equivalent. These asymptotic results permit a formal off-line detector. However, the detector considered in this paper is based on the fluctuation statistic first proposed by Sen [28] and later extended by Ploberger *et al.* [23] and Chu and White [14]. The fluctuation statistic assumes no particular alternative, and is computationally easier since recursive algorithms for parameter estimates in linear models are often available.

We assume that the stochastic sequence $\{y_{t-1}\varepsilon_t/\sigma_v\}$ obeys a functional central limit theorem (FCLT). Formally, let $\{v_t \equiv y_{t-1}\varepsilon_t, t = 2, \ldots, T\}, v_1 = 0$, and define S_T to be the piecewise constant interpolation of

$$S_T\left(\frac{k-1}{T-1}\right) = \frac{1}{\sqrt{T-1}}\sigma_v^{-1}\sum_{t=2}^{k} v_t, \qquad k = 2, 3, \ldots, T, \tag{1}$$

so that

$$S_T(\tau) = \frac{1}{\sqrt{T-1}}\sigma_v^{-1}\sum_{t=2}^{[(T-1)\tau]+1} v_t, \qquad \tau \in [0,1].$$

Throughout this paper, we use the notation $[x]$ and "$\Rightarrow$" to denote the integer part of x and "weakly converges to," respectively. $S_T(\tau)$ is a random element in $D[0,1]$, the space of CADLAG (rcll) functions on interval $[0,1]$. The FCLT assumption is that

$$S_T(\tau) \Rightarrow W(\tau), \qquad \tau \in [0,1], \tag{2}$$

where $W(\tau)$ is the standard Wiener process and

$$\sigma_v^2 = \lim_{T\nearrow\infty}\frac{1}{T-1}E\left[\left(\sum_{t=2}^{T} v_t\right)^2\right] < \infty.$$

The FCLT requirement is weak and applies to a wide class of sequences $\{y_{t-1}\varepsilon_t/\sigma_v\}$. For more details on the type of weak convergence, we refer to [9, 26, 30].

The σ_v^2 can be simplified under stronger conditions, such as ε_t is Gaussian (so that ε_t and y_{t-1} are independent) or a martingale difference sequence, i.e., $E(\varepsilon_t \mid F_{t-1}) = 0$ and $E(\varepsilon_t^2 \mid F_{t-1}) = \sigma_\varepsilon^2$, where F_{t-1} is the σ-algebra generated by $\{y_{t-1}, y_{t-2}, \ldots\}$. These conditions imply that $E(v_t v_{t-k}) = 0$, for $k \neq 0$. It follows that $E[(\sum_{t=2}^T v_t)^2] = E[\sum_{t=2}^T v_t^2] = \sum_{t=2}^T \sigma_\varepsilon^2\, E(y_{t-1}^2) = (1/T-1)\,\sigma_\varepsilon^4(1-\alpha_0^2)^{-1}$. Hence, $\sigma_v^2 = \sigma_\varepsilon^4(1-\alpha_0^2)^{-1}$.

The reference model of the FSW procedure is identified from the first window of size h, a subsample consisting of observations at $t = 2, \ldots, [(T-1)h]+1$, $h \in (0, 1/2]$. Let the parameter estimate of the reference model be

$$\widehat{\alpha}_h = \left[\sum_{t=2}^{[(T-1)h]+1} y_{t-1}^2\right]^{-1}\left[\sum_{t=2}^{[(T-1)h]+1} y_t y_{t-1}\right]; \qquad h \in (0, 1/2].$$

The kth test model is identified from a window of the same size h, containing observations at $t = [(T-1)h]+k+1, \ldots, 2[(T-1)h]+k$. As we vary k, we slide the window of size h forward to obtain a sequence of test models indexed by k, $k = 1, 2, \ldots, T - 2[(T-1)h]$. Note that the first test window is adjacent, not overlapping with the reference window. For $k = 1, 2, \ldots, T - 2[(T-1)h]$, define the parameter estimated from the kth test model as

$$\widetilde{\alpha}_{k,h} = \left[\sum_{t=[(T-1)h]+k+1}^{2[(T-1)h]+k} y_{t-1}^2\right]^{-1}\left[\sum_{t=[(T-1)h]+k+1}^{2[(T-1)h]+k} y_t y_{t-1}\right]. \tag{3}$$

Following the spirit of the fluctuation test, the deviations of test models from the reference model, measured by the $|\widetilde{\alpha}_{k,h} - \widehat{\alpha}_h|$, should be insignificant under the null hypothesis. Straightforward algebra shows that

$$\begin{aligned}(\widetilde{\alpha}_{k,h} - \widehat{\alpha}_h) &= \left[\sum_{t=[(T-1)h]+k+1}^{2[(T-1)h]+k} y_{t-1}^2\right]^{-1}\left[\sum_{t=[(T-1)h]+k+1}^{2[(T-1)h]+k} v_t\right] - \left[\sum_{t=2}^{[(T-1)h]+1} y_{t-1}^2\right]^{-1}\left[\sum_{t=2}^{[(T-1)h]+1} v_t\right]\\ &= \left[\sum_{t=[(T-1)h]+k+1}^{2[(T-1)h]+k} y_{t-1}^2\right]^{-1}\left[\sum_{t=2}^{2[(T-1)h]+k} v_t - \sum_{t=2}^{[(T-1)h]+k} v_t\right]\\ &\quad - \left[\sum_{t=2}^{[(T-1)h]+1} y_{t-1}^2\right]^{-1}\left[\sum_{t=2}^{[(T-1)h]+1} v_t\right].\end{aligned}$$

To analyze the asymptotic properties of a discrete sequence $\{(\widetilde{\alpha}_{k,h} - \widehat{\alpha}_h)\}_k$, we consider the constant interpolation of normalized $(\widetilde{\alpha}_{k,h} - \widehat{\alpha}_h)$ and investigate its limiting behavior in space $C[0,1]$. For this, we introduce a random function $\mathrm{FSW}_{T,h} \equiv ([(T-1)h]/\sqrt{T-1})(\sigma_y^2/\sigma_v)(\widetilde{\alpha}_{k,h} - \widehat{\alpha}_h)$. In terms of $S_T(\cdot)$, we have

$$\begin{aligned}\mathrm{FSW}_{T,h} &= \left[\sum_{t=[(T-1)h]+k+1}^{2[(T-1)h]+k} y_{t-1}^2/[(T-1)h]\right]^{-1} \sigma_y^2 S_T\left(\frac{k-1}{T-1}+2\frac{[(T-1)h]}{T-1}\right)\\ &\quad-\left[\sum_{t=[(T-1)h]+k+1}^{2[(T-1)h]+k} y_{t-1}^2/[(T-1)h]\right]^{-1} \sigma_y^2 S_T\left(\frac{k-1}{T-1}+\frac{[(T-1)h]}{T-1}\right)\\ &\quad-\left[\sum_{t=2}^{[(T-1)h]+1} y_{t-1}^2/[(T-1)h]\right]^{-1} \sigma_y^2 S_T\left(\frac{[(T-1)h]}{T-1}\right). \qquad (4)\end{aligned}$$

The limiting behavior of $\mathrm{FSW}_{T,h}$ in (4) can be established by using the weak law of large numbers and the FCLT, as the following theorem shows.

THEOREM 1. *Suppose that*

(a) $y_t = \alpha_0 y_{t-1} + \varepsilon_t, |\alpha_0| < 1$, *where* $\{\varepsilon_t\}$ *is a white noise process with finite second moment,*
(b) *the sequence* $\{y_{t-1}\varepsilon_t/\sigma_v\}$ *obeys the FCLT with finite* $\sigma_v^2 = \lim_{T\nearrow\infty} (1/T-1)E[(\sum_{t=2}^{T} y_{t-1}\varepsilon_t)^2]$,
(c) $\lim_{T\nearrow\infty}(1/[(T-1)h])\sum_{t=[(T-1)h]+[N\tau]+1}^{[N\tau]+2[(T-1)h]} y_{t-1}^2 \to \sigma_y^2$ *uniformly in* $0 \le \tau \le 1-2h$ *for a given* $h \in (0,1/2]$, *where* $N = \{(T-1) - 2[(T-1)h]\}/(1-2h)$. *Then*

$$\operatorname*{Max}_{k=1,\dots,T-2[(T-1)h]} |\mathrm{FSW}_{T,h}| \Rightarrow \sup_{\tau\in[0,1-2h]} |\mathrm{FW}_h(\tau)|,$$

where $\mathrm{FW}_h(\tau) = W(\tau+2h) - W(\tau+h) - W(h)$.

Theorem 1 permits construction of an asymptotic test:

$$\operatorname*{Max}_{k=1,2,\dots,T-2[(T-1)h]} \frac{[(T-1)h]}{\sqrt{T-1}} \frac{\widehat{\sigma}_y^2}{\widehat{\sigma}_v} |\widetilde{\alpha}_{k,h} - \widehat{\alpha}_h|. \qquad (5)$$

Note that we have replaced the unknown σ_y^2 and σ_v in (5) with their consistent estimates $\widehat{\sigma}_y^2$ and $\widehat{\sigma}_v$, estimated from the reference window or the entire sample. This does not alter the conclusion of Theorem 1. Consistent

estimation for σ_v is straightforward under the assumption that $\{\varepsilon_t\}$ is a Gaussian or martingale difference sequence. In practice, possible misspecification can cause $\{\varepsilon_t\}$, and hence $\{v_t\}$, to be autocorrelated, which makes the consistent estimation of σ_v a challenging task. In this situation, some sort of heteroskedasticity and autocorrelation consistent estimation is called for, such as the Andrews' automatic bandwidth procedure [1].

The asymptotic threshold value can be determined from the probability of $\mathrm{FM}_h(\tau)$ crossing a pair of constant boundaries since

$$\lim_{T\to\infty} P\left\{\max_k |\mathrm{FSW}_{T,h}| > \beta\right\} = P\left\{\max_{\tau\in[0,1-2h]} |\mathrm{FW}_h(\tau)| > \beta\right\}.$$

Given that the window size is h and that the false alarm rate is controlled at $\alpha\%$, we can solve for the asymptotic threshold value β from $P\{\max_{\tau\in[0,1-2h]} |\mathrm{FW}_h(\tau)| > \beta\} = \alpha\%$. The process $\mathrm{FW}_h(\tau)$ is clearly Gaussian with a continuous path. It is easy to see that the covariance function of $\mathrm{FW}_h(\tau)$ is $C(s,\tau) = s + 2h - \min(\tau, s+h)$. It follows that the variance of $\mathrm{FW}_h(\tau)$ is $2h$. As the variance is constant, it is legitimate to consider the probability of the $\mathrm{FW}_h(\tau)$ process crossing a constant boundary. However, this covariance function is nonstandard, which makes the problem of solving boundary crossing probability difficult. Nevertheless, Theorem 2 below provides partial analytical solutions for the asymptotic threshold value when h is constrained between 1/3 and 1/2, in which case it is easier to deal with the covariance function. The theorem is not only interesting in its own right, but useful in assessing the accuracy of simulations. The case of $h < 1/3$ is obviously of interest. Although we do not have analytical solutions in this case, we can simulate the $\mathrm{FW}_h(\tau)$ to obtain asymptotic threshold values.

THEOREM 2. *Let* $h \in [1/3, 1/2]$ *and* $h_0 = 1 - 2h/6h - 1$.

$$P\left\{\max_{\tau\in[0,1-2h]} |\mathrm{FW}_h(\tau)| > \beta\right\} = \frac{1}{\sqrt{4\pi h}} \int_{|m|<\beta} p_1(\beta,m) e^{-m^2/4h}\,dm + \frac{1}{\sqrt{4\pi h}} \int_{|m|<\beta} p_2(\beta,m) e^{-m^2/4h}\,dm,$$

where

$$\begin{aligned}p_1(\beta,m) &= 1 - \Phi[(8hh_0)^{-1/2}((\beta-m)+(m+\beta)h_0)]\\ &\quad+\exp[-(\beta^2-m^2)/4h]\Phi[(8hh_0)^{-1/2}((m-\beta)+(m+\beta)h_0)],\end{aligned}$$

and

$$p_2(\beta, m) = 1 - \Phi[(8hh_0)^{-1/2}((m+\beta) - (m-\beta)h_0)] + \exp[-(\beta^2 - m^2)/4h]\Phi[-(8hh_0)^{-1/2}((\beta+m) + (m-\beta)h_0)].$$

2.2. *GROWING REFERENCE WINDOWS AND SLIDING TEST WINDOWS*

Detection procedures based on the fluctuation statistic and identification scheme (1b) can be analyzed similarly, although they yield quite different asymptotics. Define the parameter estimated from reference windows to be

$$\widehat{\alpha}_{k,h} = \left[\sum_{t=2}^{[(T-1)h]+k} y_{t-1}^2\right]^{-1} \left[\sum_{t=2}^{[(T-1)h]+k} y_t y_{t-1}\right].$$

The parameter estimated from sliding test windows is $\widetilde{\alpha}_{k,h}$, defined earlier in [3]. The parameter fluctuation is then

$$(\widetilde{\alpha}_{k,h} - \widehat{\alpha}_{k,h})$$
$$= \left[\sum_{t=[(T-1)h]+k+1}^{2[(T-1)h]+k} y_{t-1}^2\right]^{-1} \left[\sum_{t=2}^{2[(T-1)h]+k} \varepsilon_t y_{t-1} - \sum_{t=2}^{[(T-1)h]+k} \varepsilon_t y_{t-1}\right] - \left[\sum_{t=2}^{[(T-1)h]+k} y_{t-1}^2\right]^{-1} \left[\sum_{t=2}^{[(T-1)h]+k} \varepsilon_t y_{t-1}\right].$$

Suppose that we use the random function as in (4); a proof similar to Theorem 1 yields $\{[(T-1)h]/\sqrt{T-1}\}(\sigma_y^2/\sigma_v)(\widetilde{\alpha}_{k,h} - \widehat{\alpha}_h) \Rightarrow \mathrm{GW}_h(\tau) \equiv W(\tau+2h) - (1 + (h/\tau+h))W(\tau+h)$. Routine computation shows that for $s < t$, the covariance function of $\mathrm{GW}_h(\tau)$ is given by $(1 + (h/\tau+h))[(s+h) - \min(\tau, s+h)]$, and hence the variance $(1 + (h/\tau+h))h$. The variance of $\mathrm{GM}_h(\tau)$ is not constant, and is always less than that of $\mathrm{FM}_h(\tau)$ for all $\tau > 0$. This is owing to the device that more information is used to update the reference model, and the accuracy of the reference parameter estimate continues to increase until a change is detected. In contrast, the FSW detector by construction discards the observations between the test and reference window.

Unlike $\mathrm{FM}_h(\tau)$, it is not appropriate to consider a constant boundary function to monitor the behavior of the $\mathrm{GW}_h(\tau)$ process, which means that it is not appropriate for a test procedure based on the $\mathrm{GM}_h(\tau)$ process to use constant threshold values. However, the normalized process $[(\tau+h)/(\tau+2h)]^{1/2}\mathrm{GW}_h(\tau)$ does have a constant variance, h. The normalization factor $[(\tau+h)/(\tau+2h)]$ is precisely the ratio of the reference window size to the total observations used to identify both reference and test models. The proof of the following theorem is quite similar to that of Theorem 1; we omit the treatment here.

THEOREM 3. *Given the same conditions as in Theorem 1, let*

$$\mathrm{GSW}_{T,h} \equiv \left[\frac{[(T-1)h] + (k-1)}{2[(T-1)h] + (k-1)}\right]^{1/2} \frac{[(T-1)h]}{\sqrt{T-1}} \frac{\sigma_y^2}{\sigma_v} (\widetilde{\alpha}_{k,h} - \widehat{\alpha}_{k,h}).$$

Then

$$\underset{k=1,2,\ldots,T-2[(T-1)h]}{\mathrm{Max}} |\mathrm{GSW}_{T,h}| \Rightarrow \sup_{\tau\in[0,1-2h]} \sqrt{\frac{\tau+h}{\tau+2h}} |\mathrm{GW}_h(\tau)|,$$

where $\mathrm{GW}_h(\tau) = W(\tau+2h) - (1 + (h/\tau+h))W(\tau+h), h \in (0, 1/2]$.

To initialize either the FSW or the GSW detection procedure, one must determine the window size h. It is advisable that the window size be chosen such that the reference window is not contaminated, i.e., it does not contain a change point. This can be accomplished by applying off-line tests for the constancy of the reference parameter. The consequences of different h values on the segmentation performance will be discussed via simulations in Section 4.

3. EXTENSIONS AND CHANGE POINT LOCALIZATION

3.1. *EXTENSIONS TO AR(p) MODEL*

The previous analysis can be easily extended to autoregressive models of higher order. Consider a stationary and causal AR(p) model, $y_t = Y_t\phi + \varepsilon_t$, $t = 1, 2, \ldots, T$, where $Y_t = (y_{t-1}, y_{t-2}, \ldots, y_{t-p})$, $\phi = (\phi_1, \phi_2, \ldots, \phi_p)'$ and ε_t is a white noise with finite variance σ^2. Again we assume that $\{Y_t'\varepsilon_t\}$ obeys the multivariate FCLT [22], i.e.,

$$\underline{S}_T(\tau) = \frac{1}{\sqrt{T-p}} \Omega^{-1/2} \sum_{t=p+1}^{[(T-p)\tau]+p} Y_t'\varepsilon_t \Rightarrow \underline{W}(\tau), \qquad \tau \in [0,1], \quad (6)$$

where

$$\Omega = \lim_{T \nearrow \infty} \frac{1}{T-p} E\left[\left(\sum_{t=p+1}^{T} Y_t'\varepsilon_t\right)\left(\sum_{t=p+1}^{T} Y_t'\varepsilon_t\right)'\right],$$

and $\underline{W}(\tau)$ is a p-dimensional independent standard Wiener process. Let the parameters estimated from a fixed reference window and from a sequence of growing reference windows be

$$\widehat{\phi}_h = \left[\sum_{t=p+1}^{[(T-p)h]+p} Y_t'Y_t\right]^{-1}\left[\sum_{t=p+1}^{[(T-p)h]+p} Y_t'y_t\right]$$

and

$$\widehat{\phi}_{k,h} = \left[\sum_{t=p+1}^{[(T-p)h]+k+p-1} Y_t'Y_t\right]^{-1}\left[\sum_{t=p+1}^{[(T-p)h]+k+p-1} Y_t'y_t\right],$$

$k = 1, 2, \ldots, (T-p+1)-2[(T-1)h]$. Also denote the parameters estimated from sliding test windows by

$$\widetilde{\phi}_{k,h} = \left[\sum_{t=[(T-p)h]+k+p}^{2[(T-p)h]+k+p-1} Y_t'Y_t\right]^{-1}\left[\sum_{t=[(T-p)h]+k+p}^{2[(T-p)h]+k+p-1} Y_t'y_t\right].$$

For a p-dimensional vector V, let $\|V\| = \max_{j=1,2,\ldots,p}|V_j|$ be the norm of V. We have the following theorem.

THEOREM 4. *Suppose that*

(a) $y_t = Y_t\phi_0 + \varepsilon_t$ *is a stationary and causal process, where* $\{\varepsilon_t\}$ *is a white noise process with finite variance, and*
(b) *the sequence* $\{Y_t'\varepsilon_t\}$ *obeys a multivariate FCLT,*
(c) $\lim_{T\nearrow\infty}(1/[(T-p)h])\sum_{i=[(T-p)h]+[N\tau]+p}^{[N\tau]+2[(T-p)h]+p-1} Y_t'Y_t \to \Gamma_p$, *uniformly in* $\tau \in [0, 1-2h]$ *for a given* h, *where* $N = \{(T-p)-2[(T-p)h]\}/(1-2h)$.

Let

$$\underline{\mathrm{FSW}}_{T,h} \equiv \frac{[(T-p)h]}{\sqrt{T-p}} D_T^{-1/2}(\widetilde{\phi}_{k,h} - \widehat{\phi}_h)$$

and

$$\underline{\mathrm{GSW}}_{T,h} \equiv \left[\frac{[(T-p)h]+(k+1)}{2[(T-p)h]+(k-1)}\right]^{1/2}\frac{[(T-p)h]}{\sqrt{T-p}} D_T^{-1/2}(\widetilde{\phi}_{k,h} - \widehat{\phi}_{k,h}),$$

$$h \in (0, 1/2].$$

Then under the null hypothesis,

$$\max_{k=1,2,\ldots,(T-p+1)-2[(T-p)h]} \|\underline{\mathrm{FSW}}_{T,h}\| \Rightarrow \sup_{\tau\in[0,1-2h]} \|\underline{\mathrm{FW}}_h(\tau)\|$$

and

$$\max_{k=1,2,\ldots,(T-p+1)-2[(T-p)h]} \|\underline{\mathrm{GSW}}_{T,h}\| \Rightarrow \sup_{\tau\in[0,1-2h]} \sqrt{\frac{\tau+h}{\tau+2h}}\|\underline{\mathrm{GW}}_h(\tau)\|,$$

where $D_T = \Gamma_p^{-1}\Omega\Gamma_p^{-1}$, $\underline{\mathrm{FW}}_h(\tau) = \underline{W}(\tau+2h) - \underline{W}(\tau+h) - \underline{W}(h)$ and $\underline{\mathrm{GW}}_h(\tau) = \underline{W}(\tau+2h) - (1+(h/\tau+h))\underline{W}(\tau+h)$.

Several remarks are in order. First, it is necessary to estimate the unknown matrix D_T to implement the FSW or GSW test. Under stronger assumptions, e.g., Gaussian or martingale difference, Ω can be simplified to the familiar autocovariance matrix Γ_p, a Toeplitz form of $[\lambda(i-j)]_{i,j=1}^{p}$. Hence, D_T reduces to Γ_p^{-1} and consistent estimation of Γ_p is well known. Second, asymptotic threshold values of the $\underline{\mathrm{FSW}}$ or $\underline{\mathrm{GSW}}$ proeedure can be computed easily given the results in Sections 2.1 and 2.2 since

$$\begin{aligned} P&\left\{\sup_{t\in[0,1-2h]} \|\underline{\mathrm{FW}}_h(\tau)\| > \beta\right\} \\ &= 1 - P\{\|\underline{\mathrm{FW}}_h(\tau)\| \le \beta, \quad \text{for all } \tau \in [0, 1-2h]\} \\ &= 1 - \{P\{|\underline{\mathrm{FW}}_h(\tau)| \le \beta, \quad \text{for all } \tau \in [0, 1-2h]\}\}^p. \end{aligned} \tag{7}$$

Hence, only asymptotic threshold values for univariate $\mathrm{FW}_h(\tau)$ need to be tabulated.

3.2. *LOCALIZING THE CHANGE POINTS*

When the FSW or GSW test signals nonstationarity, the next step is to localize the change points. If the parameter estimates from the kth sliding window are sufficiently different from the reference parameters for the first time, the test rejects the stationarity and the detection procedure stops. Heuristically, the right boundary point of the kth sliding window, i.e., the last observation in the kth sliding window, is not an ideal estimate of the change point since there is almost certainly a positive alarm delay. However, the boundary point can be used to initialize the procedure of localizing change points.

Let t_b be the time point when the detection procedure stops. To search a change point from $t = p+1$ to $t = t_b$, the sample needs segmentation (from $t = 1$ to $t = t_b$), and is then divided into two portions by the split point s, $s = p+1, \ldots, t_b$. Change point localization can be based on classical test statistics such as the maximal likelihood ratio statistic [4], max-LM or max-W. In this paper, the Akaike information criterion is used, which is equivalent to minimized error sum of squares. Specifically, We are minimizing the sum of $\mathrm{ESS}_{p+1,s} = \sum_{t=p+1}^{s}(y_t - Y_t\widehat{\phi}_{p+1,s})^2$ and $\mathrm{ESS}_{s+1,t_b} = \sum_{t=s+1}^{t_b}(y_t - Y_t\widehat{\phi}_{s+1,t_b})^2$, where $\widehat{\phi}_{p+1,s}$ and $\widehat{\phi}_{s+1,t_b}$ are the parameter estimates from pre-s and post-s observations respectively. Segmentation is done when $s^* = \mathrm{argmin}(\mathrm{ESS}_{p+1,s} + \mathrm{ESS}_{s+1,t_b})$ is achieved. The detection procedure is resumed after the first change point is localized. The reference window now contains observation $t = s^*+1, \ldots, t_b$ of size $[(t_b - s^*)/(T - s^*)]$, with reference parameters $\widehat{\phi}_{s+1,t_b}$, already identified in the localization stage.

4. SIMULATIONS

We first simulate the univariate $\mathrm{FW}_h(\tau)$ and $[(\tau+h)/(\tau+2h)]^{1/2}\,\mathrm{GW}_h(\tau)$ process to approximate asymptotic threshold values. For this, we partition the [0, 1] interval into 3000 and 6000 subintervals of equal length, and discretize the standard Wiener process using the GAUSS normal random number generator. Simulated thresholds using finer 6000 partitions are reported in the first row of Table 1. Although results using 3000 partitions are not reported, the simulated critical values change very little from 3000 to 6000 partitions. Moreover, when $h = 0.33$, the simulated critical values with 6000 partitions are fairly close to the asymptotic values solved for analytically using Theorem 2. The same simulation technique is used to obtain critical values of the GSW procedure. For future reference, we also tabulate critical values of the FSW and GSW procedure for the higher order autoregressive process in Tables 1 and 2 up to tenth order.

To further assess the accuracy of the simulated asymptotic threshold values, we generate Gaussian AR(1) processes with an autoregressive parameter equal to 0.2, 0.5, and 0.8, respectively, and use the asymptotic critical values in Tables 1 and 2 to check empirical sizes. Results are summarized in Table 3. It is interesting to note that when the data are moderately correlated ($\alpha \leq 0.5$), both FSW and GSW procedures deliver satisfactory

TABLE 1
Critical Values of FSW Procedure for p-th Order Autoregressive Model

	$h = 0.10$		$h = 0.15$		$h = 0.20$		$h = 0.33$	
	10%	5%	10%	5%	10%	5%	10%	5%
$p =$ 1	1.306	1.426	1.505	1.652	1.666	1.844	1.899	2.138
= 2	1.425	1.523	1.652	1.779	1.842	2.007	2.135	2.353
= 3	1.480	1.583	1.723	1.857	1.938	2.092	2.251	2.485
= 4	1.521	1.617	1.777	1.905	2.002	2.158	2.342	2.547
= 5	1.558	1.645	1.807	1.951	2.049	2.191	2.418	2.615
= 6	1.578	1.671	1.849	1.978	2.087	2.210	2.477	2.659
= 7	1.597	1.693	1.881	2.001	2.113	2.245	2.510	2.701
= 8	1.616	1.705	1.900	2.022	2.152	2.284	2.544	2.743
= 9	1.633	1.722	1.922	2.036	2.177	2.305	2.584	2.763
= 10	1.643	1.738	1.937	2.052	2.185	2.326	2.607	2.796

Note: The order of an autoregressive process and the window size are denoted by p and h, respectively. Results are calculated from simulation of $\mathrm{FM}_h(t)$ process with 6000 partitions and equation (7) in the text.

TABLE 2
Critical Values of GSW Procedure

	$h = 0.10$		$h = 0.15$		$h = 0.20$		$h = 0.33$	
	10%	5%	10%	5%	10%	5%	10%	5%
$p =$ 1	1.030	1.105	1.147	1.242	1.308	1.418	1.444	1.575
= 2	1.104	1.176	1.241	1.328	1.417	1.513	1.574	1.703
= 3	1.146	1.220	1.292	1.375	1.472	1.573	1.647	1.774
= 4	1.173	1.250	1.325	1.412	1.509	1.608	1.696	1.815
= 5	1.202	1.271	1.348	1.446	1.544	1.636	1.739	1.867
= 6	1.219	1.284	1.369	1.471	1.569	1.664	1.769	1.897
= 7	1.235	1.298	1.393	1.483	1.590	1.689	1.797	1.916
= 8	1.249	1.309	1.406	1.497	1.606	1.702	1.813	1.941
= 9	1.261	1.326	1.426	1.512	1.619	1.712	1.840	1.952
= 10	1.267	1.336	1.440	1.523	1.631	1.722	1.863	1.961

Note: The order of an autoregressive process and the window size are denoted by p and h, respectively. Results are based on simulations of $\mathrm{GW}_h(t)$ process with 6000 partitions.

finite sample size results regardless of the choice of window size. However, when the observations are highly correlated, we observe significant finite sample size distortion (increasing false alarm rate) in the FSW and GSW procedures. Further inspection of Table 3 shows that the smaller the window size ($h = 0.1$ or 0.2), the greater the distortion. Heuristically, the parameter shift information contained in the sample is obscured by the autocorrelation, and the more seriously autocorrelated the observation, the more slowly the information is revealed. Moreover, since the variance term D_T in Theorem 4 is estimated from the reference window in the simulations, a slight estimation error for D_T will greatly bias the

TABLE 3
Empirical Sizes of FSW and GSW Procedure

	$\alpha = 0.2$ 10%	5%	1%	$\alpha = 0.5$ 10%	5%	1%	$\alpha = 0.8$ 10%	5%	1%
FSW:									
$h = 0.10$	9.50	4.90	0.80	10.50	5.80	1.30	16.80	12.10	5.00
$= 0.20$	11.00	4.80	1.00	10.30	5.20	1.00	12.50	7.00	1.90
$= 1/3$	9.90	4.50	0.80	9.90	5.10	1.10	10.00	4.90	1.20
GSW:									
$h = 0.10$	9.80	4.20	0.80	12.00	6.50	1.70	22.00	16.50	7.10
$= 0.20$	11.20	4.70	0.90	10.50	5.10	1.10	15.20	8.00	3.00
$= 0.30$	10.10	5.20	1.40	9.90	5.00	1.00	12.50	6.50	2.00

Note: 3000 observations are generated from an AR(1) process with AR coefficient α. The number of replications is 2500.

detection statistic. Intuitively, a larger window size in the case of highly correlated data improves the accuracy in estimating D_T. An unrealistic simulation has also been performed using the true variance instead of the estimated variance from the reference window. Results (not reported here) indicate that finite sample sizes of all choices of window size become acceptable.

An important message from Table 3 is that as far as controlling the false alarm rate is concerned, choose a larger window size for strongly dependent observations.

Turning to the finite sample detection power and change point localization performance of the FSW and GSW procedure, we generate a sample of size 3000 from two AR(1) processes: $y_t = 0.5y_{t-1} + \varepsilon_t$ for $t \leq [Tb]$ and $y_t = 0.8y_{t-1} + \varepsilon_t$, for $t \leq [Tb]$, where ε_t is normal iid(0,1) and $b \in (0, 1)$ is the location of the change point. We vary the window size h and the change point as follows. First, in Table 4, we set the change point located right after the first reference window ($b - h = 0$), cases most favorable to our segmentation procedures. Second, in Table 5, we let the break point fall inside the first reference window ($b - h < 0$), cases when the first reference window is contaminated. Lastly, we consider a change point outside the first reference window ($b - h > 0$) in Tables 6 and 7.

As shown in Table 4, although the FSW procedure exhibits perfect power, the performance of the FSW procedure in localizing the change point crucially depends on the choice of h. From the second to fourth column of Table 4, we observe the following:

(a) The mean delay (mean segmentation point minus $T \times b$) when $h =$ 0.1, 0.2, and 0.33 is 8.4, 8.6, and 0.3 periods, respectively, indicating that larger window size significantly lowers the bias.

TABLE 4
Noncontaminated Reference Models (Most Favorable Alternative)

	FSW Procedure $h = b = 0.1$	$h = b = 0.2$	$h = b = 1/3$	GSW Procedure $h = b = 0.1$	$h = b = 0.2$	$h = b = 0.3$
Detection:						
Power	1.0	1.0	1.0	1.0	1.0	1.0
Segmentation:						
Mean	309.4	609.6	1001.3	312.1	610.9	909.4
Std. deviation	46.9	40.8	32.4	47.1	36.1	36.6
1st quantile	298.0	599.0	992.0	299.0	600.0	899.0
Median	305.0	605.0	996.0	305.0	606.0	904.0
3rd quantile	320.0	619.0	1011.0	321.0	621.0	919.0
Maximal delay	401.0	291.0	209.0	541.0	394.0	224.0
Localization						
% within ± 10	46.4	49.2	53.0	48.3	48.7	50.1
% within ± 20	64.3	66.4	70.6	64.8	66.3	67.2
% within ± 50	84.6	86.9	90.0	86.1	87.9	88.0

Note: $T = 3000$ and the number of replications = 2500.

TABLE 5
Contaminated Reference Models

	FSW Procedure $b = 0.1$; $h = 0.15$	$b = 0.28$; $h = 1/3$	GSW Procedure $b = 0.1$; $h = 0.15$	$b = 0.25$; $h = 0.3$
Power	1.0	1.0	0.7	1.0
Mean segmentation	312.5	852.7	320.5	762.1
Std. deviation	40.6	33.6	67.0	32.1
1st quantile	300.0	839.0	300.0	750.0
Median	306.0	846.0	306.0	756.0
3rd quantile	323.0	861.0	324.0	769.0
Maximal delay	337.0	309.0	928.0	260.0
Localization				
% within ± 10	45.6	47.2	47.2	47.9
% within ± 20	63.4	67.2	63.6	67.9
% within ± 50	87.9	88.4	84.7	89.0

Note: $T = 3000$ and the number of replications = 1000.

(b) The localization accuracy is increased with respect to the window size in the sense of decreased standard deviation and a higher percentage of estimated change points within the interval of $Tb + q$ periods, $q = 10$, 20, and 50.

(c) The maximal delay (maximal estimated change point minus Tb) is 401, 291, and 209 for $h = 0.1$, 0.2, and 1/3, respectively.

TABLE 6
FSW with Noncontaminated Reference Models (Late Break Alternative)

	$b = 0.5$			$b = 0.7$		
	$h = 0.1$	$h = 0.2$	$h = 1/3$	$h = 0.1$	$h = 0.2$	$h = 1/3$
Power	1.0	1.0	1.0	1.0	1.0	1.0
Mean segmentation	1442.3	1485.7	1508.8	2018.7	2047.6	2065.0
Std. deviation	254.2	127.1	35.2	350.3	268.5	220.7
1st quantile	1493.0	1499.0	1499.0	2094.0	2098.0	2099.0
Median	1504.0	1504.0	1505.0	2104.0	2105.0	2104.0
3rd quantile	1519.0	1519.0	1520.0	2119.0	2121.0	2120.0
Maximal delay	313.0	168.0	238.0	219.0	180.0	211.0
Localization						
% within ± 10	44.0	47.5	50.5	42.2	42.7	46.6
% within ± 20	61.7	65.3	66.1	59.0	59.0	63.6
% within ± 50	80.7	86.1	88.2	80.0	77.7	84.3

$T = 3000$ and the number of replications = 1000.

TABLE 7
GSW with Noncontaminated Reference Models (Late Break Alternative)

	$b = 0.5$			$b = 0.7$		
	$h = 0.1$	$h = 0.2$	$h = 1/3$	$h = 0.1$	$h = 0.2$	$h = 0.3$
Power	1.0	1.0	1.0	1.0	1.0	1.0
Mean segmentation	1479.7	1488.5	1506.2	2025.0	2050.2	2069.7
Std. deviation	155.6	129.7	46.6	320.8	246.1	195.6
1st quantile	1498.0	1499.0	1499.0	2098.0	2094.0	2097.0
Median	1505.0	1505.0	1504.0	2105.0	2103.0	2105.0
3rd quantile	1521.0	1520.0	1517.0	2120.0	2118.0	2119.0
Maximal delay	187.0	287.0	349.0	174.0	170.0	238.0
Localization						
% within ± 10	44.0	46.2	50.7	44.3	45.9	44.6
% within ± 20	60.0	64.5	67.9	59.4	59.8	61.5
% within ± 50	81.9	84.5	87.6	78.5	78.7	81.4

Note: $T = 3000$ and the number of replications = 1000.

Similar results are observed in the GSW power simulations, i.e., the bias and maximal delay of the GSW procedure decreases with respect to h and the localization accuracy is improved with large window size. However, when compared to the FSW procedure, we see that the change point detection ability of the GSW is slightly less accurate than the FSW in terms of the maximal delay. Other statistics such as the estimated quantiles, mean, and standard deviations are rather similar. Note that the simulation environment here is set up in favor of the FSW procedure since the break point is located right after the first reference window. Obviously, the intuitive advantage of using growing reference windows to update reference parameters disappears.

As discussed in Section 3, an off-line test should be routinely applied to the reference model to ensure that the reference window does not include a change point. Since an off-line test is less powerful in detecting nonstationarity when the change point is located near the end of the test sample, it is quite possible in practice to mistakenly include a break point inside the first reference window ($b - h < 0$). It is of interest to simulate the segmentation performance when the reference models are misspecified. As evident from the second and fourth columns of Table 5 (in both cases, $b/h = 2/3$), the FSW outperforms GSW in detection power and localization accuracy. Specifically, FSW has smaller bias, lower standard deviation, and shorter maximal delay, although estimated quantile statistics are similar.

The fifth column reports the results of applying the GSW procedure with $h = 0.3$ while the change point is located at $t = 750(3000\times0.25)$. Compared to the previous GSW with $b = 0.1$ and $h = 0.15$ in the fourth column, we find that the detection power is improved, indicating that if the first reference model is misspecified, the closer the change point is to the right end of the reference window, the higher the detection power.[2] Moreover, when compared to the FSW with $h = 1/3$ and $b = 0.28$ (change point at $t = 840$), it is seen that the GSW and FSW perform similarly.[3] This is owing to the fact that the change point locations relative to the first reference window (the ratio of b to h in the third and fifth columns) differ very little. Results not reported here confirm that the farther away the change point is from the right end of the reference window (i.e., higher b/h), the lower the power. But we need not worry about this kind of misspecification since it can be effectively avoided by applying off-line detection statistics to a reference model.

Based on our simulations so far, it seems that the FSW procedure outperforms the GSW in terms of detection power and location accuracy provided one can choose a proper h in such a way that the change point is close to the first reference window. In practice, this cannot be guaranteed. It is therefore interesting to perform simulations in which the change point is far from close to the first reference window. We consider two scenarios: the change point is located in the middle ($b = 0.5$), and in the latter half of the sample ($b = 0.7$).

Table 6 summarizes simulation results for the FSW procedure. Given h, we observe higher localization accuracy when $b = 0.5$, confirming the

[2]However, when compared to the FSW with $h = b = 1/3$ in Table 4, with see that misspecification in the reference model results in a little deterioration in localization accuracy.

[3]Note that the location of break points in the third column and fifth columns differs by 90 periods. If we add 90 to the fifth column's mean and quantile statistics, the resulting figures resemble the third column.

intuition that it is more difficult to localize a change point close to the right end of the sample. On the other hand, given $b = 0.5$ or 0.7, the larger the window size, the smaller the bias and the lower is the standard deviation. Moreover, it is interesting to note the nonmonotone relations between the maximal delay and the window size. In particular, we see that $h = 0.2$ gives the shortest maximal delay. This seems to suggest a conjecture that an optimal window size may exist in the sense of minimized maximal delay, a well-accepted optimality criterion in the sequential test literature.

The results of GSW power simulations in Table 7 are similar to Table 6, except that no information was obtained on the potential optimal window size. Comparisons of Tables 6 and 7 show that neither procedure dominates the other. It appears that the GSW performs better than the FSW in the case of smaller windows. Specifically, when $b = 0.5$ or 0.7, the GSW with $h = 0.1$ yields smaller bias, lower standard deviation, and shorter maximal delay, although the localization accuracy is comparable. However, when $h = 0.2$, it is not clear which procedure performs better. This is somewhat inconsistent with the intuition that the advantage of updating the reference parameters in the GSW should lead to superior performance. One possibility could be that our simulation sample is so large that the mechanism of updating reference parameters actually gains very little. To confirm this conjecture, we performed the same simulations with a smaller sample size ($T = 600$ and $h = 0.2$) in favor of the GSW. Results not tabulated here support the conjecture.

5. CONCLUDING REMARKS

We have analyzed the asymptotic properties of sliding test window segmentation procedures (FSW and GSW) which base nonstationarity detection on fluctuation statistics and change point localization on the Akaike information criterion. Their limiting distributions are derived and asymptotic threshold values are tabulated for future reference. Although the discussions are limited to the AR models, techniques presented here can be applied to general ARIMA models. Finite but large sample simulations on the change point localization accuracy are performed to illustrate the usefulness of the proposed segmentation algorithms. Major findings are summarized as follows.

(A) When the time series exhibits less autocorrelation, the choice of the window size is not crucial in the sense of controlling the false alarm rate. On the other hand, if the time series is strongly autocorrelated, the FSW seems to have better size performance than the GSW.

(B) Segmentation with a larger window size performs better than with a smaller window size, provided that the reference window is not contaminated.

(C) FSW is more powerful and accurate than the GSW when the change point is located near the end of the first reference window.

(D) Localization accuracy deteriorates as the change point moves toward the end of the sample.

APPENDIX

I. PROOF OF THEOREM 1

LEMMA A (Chu et al. ([13]). *Let X_T be a sequence of a random process in $D[0,1]^m$ converging in distribution (with respect to the Skorohod topology) to a random process X in $C[0,1]^m$ (i.e., the limiting process has continuous path). Further, let $0 < h_T < 1$ be a sequence converging to $0 < h < 1$, and let $\kappa_T\colon [0,1-h] \to [0,1-h_T]$ be a sequence of maps such that $\sup_{\tau\in[0,1-h]} |\kappa_T(\tau) - \tau|$ tends to zero. Then if Z_T is the random process on $D[0,1-h]^m$ given by $Z_T = X_T(\kappa_T(\tau) + h_T)$, we have $Z_T \Rightarrow Z$, where $Z(\tau) = X(\tau + h)$, for $\tau \in [0,1\text{-}h]$.*

To prove Theorem 1, We first show that the first term on the right-hand side of (4), $S_T[(k-1/T-1) + 2([(T-1)h]/T-1)]$, weakly converges to $W(\tau+2h)$. Put $X_T(\tau) = S_T(\tau)$ and $X(\tau) = W(\tau)$; we have $X_T \Rightarrow X$ by condition (b). Define $h_T = [(T-1)h]/T-1$, $N = (T-1)-2[(T-1)h]/1-2h$, and $\kappa_T(\tau) = [N\tau]/T-1$. Observe that $S_T[([N\tau]/T-1)+2h_T]$ is a piecewise constant interpolation of $S_T[(k-1/T-1)+2([(T-1)h]/T-1)]$ on $[0, 1\text{-}2h]$ with interpolation nodes $k-1/N, k = 1,\ldots,2[(T-1)h]$. Since $\kappa_T(\tau) \to \tau$ and $h_T \to h$ as $T \to \infty$, it follows from Lemma A that $S_T[([N\tau]/T-1) + 2h_T] \Rightarrow W(\tau+2h)$.

Similarly, for the second and third terms on the right-hand side of (4), one can prove that $S_T[([N\tau]/T-1) + h_T] \Rightarrow W(\tau+h)$ and $S_T(h_T) \Rightarrow W(h)$. Since $[\sum_{t=[(T-1)h]+k+1}^{2[(T-1)h]+k} y_{t-1}^2/[(T-1)h]]^{-1}$ and $[\sum_{t=2}^{[(T-1)h]+1} y_{t-1}^2/[(T-1)h]]^{-1}$ converge to the same limit σ_y^2 by condition (c), the theorem follows from the continuous mapping theorem [9, p. 35].

II. PROOF OF THEOREM 2

Given that $h \in [1/3, 1/2]$, the process $\mathrm{FW}_h(t)$ is such that $E[\mathrm{FW}_h(t)] = 0, \forall t$ and for $s \le t, s,t \in [0,1-2h]$, $\mathrm{Cov}[\mathrm{FW}_h(s), \mathrm{Fw}_h(t)] = s+2h-t$. Conditional on $\mathrm{FW}_h(0) = m$, routine computations show that $E[\mathrm{FW}_h(t) | \mathrm{FW}_h(0) = m] = (1-t/2h)m$, $\mathrm{Cov}[\mathrm{FW}_h(s), \mathrm{FW}_h(t) | \mathrm{FW}_h(0) = m] = 2s-(st/2h)$. This conditional covariance function thus satisfies the Markov property, that is, $\mathrm{Cov}[\mathrm{FW}_h(s), \mathrm{FW}_h(t) | \mathrm{FW}_h(0) = m] = u(s)v(t)$, where $u(s) = s$ and $v(t) = 2-(t/2h)$. This suggests that we can rescale the time parameter [17] to characterize the conditional $\mathrm{FW}_h(t)$ process in terms of a Brownian motion, as shown in the following lemma.

Lemma B. *Let $h \in [1/3, 1/2]$. Conditional on $\mathrm{FW}_h(0) = m$,*

$$\mathrm{FW}_h(t) \stackrel{d}{=} \sqrt{8h}\frac{1}{\theta+1}W(\theta) + \left[\frac{2}{\theta+1} - 1\right]m, \qquad \theta \in [0,\ h_0],\ h_0 = \frac{1-2h}{6h-1}.$$

Proof. Let $a(t) = u(t)/v(t) = t/(2 - (t/2h))$, which is monotonically increasing with t. The zero-mean process $(2 - (t/2h))W(a(t))$ has covariance $2s - (st/2h)$, for $s < t$. Since this is identical to the covariance of conditional $\mathrm{FW}_h(t)$ process, $(2-(t/2h))W(a(t)) + (1 - (t/2h)m \stackrel{d}{=} \mathrm{FW}_h(t)|_{\mathrm{FM}_h(0)=m}$, where "$\stackrel{d}{=}$" signifies equality in distribution. Put $\lambda = (2 - (t/2h))$ so that $\lambda \in [3 - (1/2h), 2]$. We have

$$\begin{aligned}\left(2 - \frac{t}{2h}\right)W(a(t)) + \left(1 - \frac{t}{2h}\right)m &\stackrel{d}{=} \lambda W(\lambda^{-1}2h(2-\lambda)) + (\lambda - 1)m \\ &\stackrel{d}{=} \sqrt{2h}\lambda W(\lambda^{-1}(2-\lambda)) + (\lambda-1)m \\ &\stackrel{d}{=} \sqrt{2h}\frac{2}{\theta+1} \times W(\theta) + \left[\frac{2}{\theta+1} - 1\right]m \\ &\stackrel{d}{=} \sqrt{8h}\frac{1}{\theta+1}W(\theta) + \left[\frac{2}{\theta+1} - 1\right]m,\end{aligned}$$

where

$$\theta = \lambda^{-1}(2-\lambda) \text{ and } \theta \in \left[0,\ \frac{1-2h}{6h-1}\right].$$

Lemma B shows that the conditional probability of the $\mathrm{FW}_h(t)$ process crossing constant boundaries $\pm\beta$ is equal to the unconditional probability of a standard Wiener process crossing linear boundaries. That is,

$$\begin{aligned}p(\beta, m) &\equiv P\{|\mathrm{FW}_h(t)| > \beta, \text{ conditioning on } \mathrm{FW}_h(0) = m, \\ &\qquad \text{for some } t \in [0, 1-2h]\} \\ &= 1 - P\{-\beta \le \mathrm{FW}_h(t)|_{\mathrm{FW}_h(0)=m} \le \beta\} \\ &= 1 - P\left\{-\beta \le \sqrt{8h}\frac{1}{\theta+1}W(\theta) + \left[\frac{2}{\theta+1} - 1\right]m \le \beta\right\} \\ &= 1 - P\{(8h)^{-1/2}[-(m+\beta) + (m-\beta)\theta] \\ &\qquad \le W(\theta) \le (8h)^{-1/2}[(\beta - m) + (\beta + m)\theta]\} \\ &= P\{W(\theta) > (8h)^{-1/2}[(\beta - m) + (\beta + m)\theta]\} \\ &\qquad + P\{(8h)^{-1/2}[-(m+\beta) - (m-\beta)\theta] > W(\theta)\} \qquad \text{(A1)}\end{aligned}$$

since the probability that a Wiener process crosses both boundaries is negligible.

To evaluate the two probabilities in (A1), we borrow a result from Cuzik [15] that $P\{W(\tau) > a + c\tau, \text{ for some } \tau \le T\} = 1 - \Phi(T^{-1/2}(a + cT)) + \exp(-2ac)\Phi(T^{-1/2}(-a + cT))$. It follows that

$$\begin{aligned}p_1(\beta, m) &\equiv P\{W(\theta) > (8h)^{-1/2}[(\beta - m) + (\beta + m)\theta], \quad \text{for some } \theta \le h_0\} \\ &= 1 - \Phi[(8hh_0)^{-1/2}((\beta - m) + (m+\beta)h_0] \\ &\quad + \exp[-(\beta^2 - m^2)/4h] \\ &\quad \times \Phi[(8hh_0)^{-1/2}((m-\beta) + (m+\beta)h_0)],\end{aligned}$$

and

$$\begin{aligned}p_2(\beta, m) &\equiv P\{(8h)^{-1/2}[(m - \beta) - (\beta + m)\theta] > W(\theta), \quad \text{for some } \theta \le h_0\} \\ &= p\{W(\theta) > (8h)^{-1/2}[(\beta - m) + (\beta + m)\theta], \quad \text{for some } \theta \le h_0\} \\ &= 1 - \Phi[(8hh_0)^{-1/2}((m+\beta) - (m-\beta)h_0] \\ &\quad + \exp[-(\beta^2 - m^2)/4h] \\ &\quad \times \Phi[-(8hh_0)^{-1/2}((\beta + m) + (m-\beta)h_0)].\end{aligned}$$

The unconditional probability $P\{|\mathrm{FM}_h(t)| > \beta\}$ is given by

$$\frac{1}{\sqrt{4\pi h}}\int_{|m|<\beta} p(\beta, m)e^{-m^2/4h}dm = \frac{1}{\sqrt{4\pi h}}\int_{|m|<\beta} p_1(\beta, m)e^{-m^2/4h}dm + \frac{1}{\sqrt{4\pi h}}\int_{|m|<\beta} p_2(\beta, m)e^{-m^2/4h}dm.$$

□

III. PROOF OF THEOREM 4

Let $S_T(j - p/T - p) = (1/\sqrt{T-p})\Omega^{-1/2}\sum_{t=p+1}^{j} Y_t'\varepsilon_t, j = p+1, p+2, \ldots, T$. Since

$$(\tilde{\phi}_{k,h} - \hat{\phi}_h) = \left[\sum_{t=[(T-p)h]+k+p}^{2[(T-p)h]+k+p-1} Y_t'Y_t\right]^{-1}\left[\sum_{t=[(T-p)h]+k+p}^{2[(T-p)h]+k+p-1} Y_t'\varepsilon_t\right] - \left[\sum_{t=p+1}^{[(T-p)h]+p} Y_t'Y_t\right]^{-1}\left[\sum_{t=p+1}^{[(T-p)h]+p} Y_t'\varepsilon_t\right],$$

$$\frac{\lfloor (T-p)h \rfloor}{\sqrt{T-p}} D_T^{-1/2}(\widetilde{\phi}_{k,h} - \widehat{\phi}_h)$$

$$= \frac{\lfloor (T-p)h \rfloor}{\sqrt{T-p}} \Omega^{-1/2}\Gamma_p \left[\sum_{t=\lfloor (T-p)h \rfloor+k+p}^{2\lfloor (T-p)h \rfloor+k+p-1} Y_t'Y_t\right]^{-1} \times \left[\sum_{t=p+1}^{2\lfloor (T-p)h \rfloor+k+p-1} Y_t'\varepsilon_t - \sum_{t=p+1}^{\lfloor (T-p)h \rfloor+k+p-1} Y_t'\varepsilon_t\right]$$

$$- \frac{\lfloor (T-p)h \rfloor}{\sqrt{T-p}} \Omega^{-1/2}\Gamma_p \left[\sum_{t=p+1}^{\lfloor (T-p)h \rfloor+p} Y_t'Y_t\right]^{-1} \left[\sum_{t=p+1}^{\lfloor (T-p)h \rfloor+p} Y_t'\varepsilon_t\right]$$

$$= \frac{\lfloor (T-p)h \rfloor}{\sqrt{T-p}} \Omega^{-1/2}\Gamma_p \left[\sum_{t=\lfloor (T-p)h \rfloor+k+p}^{2\lfloor (T-p)h \rfloor+k+p-1} Y_t'Y_t\right]^{-1} \times \left[\underline{S}_T\left(\frac{k-1}{T-p} + 2\frac{\lfloor (T-p)h \rfloor}{T-p}\right) - \underline{S}_T\left(\frac{k-1}{T-p} + \frac{\lfloor (T-p)h \rfloor}{T-p}\right)\right] - \frac{\lfloor (T-p)h \rfloor}{\sqrt{T-p}} \Omega^{-1/2}\Gamma_p \times \left[\sum_{t=p+1}^{\lfloor (T-p)h \rfloor+p} Y_t'Y_t\right]^{-1} \underline{S}_T\left(\frac{\lfloor (T-p)h \rfloor}{T-p}\right).$$

Similar to the proof of Theorem 1, let $h_T = \lfloor (T-p)h \rfloor / T-p$, $N = (T-p) - 2\lfloor (T-p)h \rfloor / 1 - 2h$ and $\kappa_T(\tau) = \lfloor N\tau \rfloor / T-p$. Note that $\Gamma_p[(1/\lfloor (T-p)h \rfloor)\sum_{t=\lfloor (T-p)h \rfloor+k+p}^{2\lfloor (T-p)h \rfloor+k+p-1} Y_t'Y_t]^{-1} - I_p \longrightarrow 0$ in probability by (c) and $\underline{S}_T((\lfloor N\tau \rfloor\ /T-p) + 2h_T)$ is a piecewise constant interpolation on $[0, 1-2h]$ of $\underline{S}_T((k-1/T-p) + 2(\lfloor (T-p)h \rfloor / T-p))$ with nodes $k-1/N, 1 \le k \le (T-p+1) - 2\lfloor (T-p)h \rfloor$; the proof is completed after applying the FCLT and the continuous mapping theorem.

Turn to the second part of the theorem. The random function

$$\frac{\lfloor (T-p)h \rfloor}{\sqrt{T-p}} D_T^{-1/2}(\widetilde{\phi}_{k,h} - \widehat{\phi}_h)$$

$$= \frac{\lfloor (T-p)h \rfloor}{\sqrt{T-p}} \Omega^{-1/2}\Gamma_p \left[\sum_{t=\lfloor (T-p)h \rfloor+k+p}^{2\lfloor (T-p)h \rfloor+k+p-1} Y_t'Y_t\right]^{-1} \times \left[\sum_{t=p+1}^{2\lfloor (T-p)h \rfloor+k+p-1} Y_t'\varepsilon_t - \sum_{t=p+1}^{\lfloor (T-p)h \rfloor+k+p-1} Y_t'\varepsilon_t\right] - \frac{\lfloor (T-p)h \rfloor}{\sqrt{T-p}} \Omega^{-1/2}\Gamma_p \left[\sum_{t=p+1}^{\lfloor (T-p)h \rfloor+k+p-1} Y_t'Y_t\right]^{-1} \times \left[\sum_{t=p+1}^{\lfloor (T-p)h \rfloor+k+p-1} Y_t'\varepsilon_t\right]$$

$$= \frac{1}{\sqrt{T-p}} \Omega^{-1/2}\Gamma_p \left[\frac{1}{\lfloor (T-p)h \rfloor} \sum_{t=\lfloor (T-p)h \rfloor+k+p}^{2\lfloor (T-p)h \rfloor+k+p-1} Y_t'Y_t\right]^{-1} \times \left[\underline{S}_T\left(\frac{k-1}{T-p} + 2\frac{\lfloor (T-p)h \rfloor}{T-p}\right) - \underline{S}_T\left(\frac{k-1}{T-p} + \frac{\lfloor (T-p)h \rfloor}{T-p}\right)\right] - \frac{1}{\sqrt{T-p}} \Omega^{-1/2}\Gamma_p \frac{\lfloor (T-p)h \rfloor}{\lfloor (T-p)h \rfloor + k - 1} \times \left[\frac{1}{\lfloor (T-p)h \rfloor + k - 1} \sum_{t=p+1}^{\lfloor (T-p)h \rfloor+k+p-1} Y_t'Y_t\right]^{-1} \times \underline{S}_T\left(\frac{k-1}{T-p} + \frac{\lfloor (T-p)h \rfloor}{T-p}\right).$$

Since

$$\Gamma_p \left[\frac{1}{\lfloor (T-p)h \rfloor} \sum_{t=\lfloor (T-p)h \rfloor+k+p}^{2\lfloor (T-p)h \rfloor+k+p-1} Y_t'Y_t\right]^{-1} \longrightarrow I_p, \Gamma_p \left[\frac{1}{\lfloor (T-p)h \rfloor + k - 1} \sum_{t=p+1}^{\lfloor (T-p)h \rfloor+k+p-1} Y_t'Y_t\right]^{-1} \longrightarrow I_p,$$

and

$$[(T-p)h]/[(T-p)h]+k-1 \longrightarrow h/\tau + h,$$

the result follows from the FCLT and the continuous mapping theorem.

This research was supported by Zumberg Research Innovation Fund, Grant Number 4060403000, University of Southern California. The author thanks Cheng Hsiao, Paul Wang, and two anonymous referees for helpful comments. All remaining errors are the author's.

REFERENCES

1. D. Andrews, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59:817–858 (1991).
2. D. Andrews, Tests for parameter instability and structural change with unknown change points, *Econometrica* 61:821–856 (1993).
3. D. Andrews and W. Ploberger, Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* (1994).
4. U. Appeal and A. Brandt, Adaptive sequential segmentation of piecewise stationary time series, *Information Sciences* 29:27–56 (1983).
5. J. Bai, Estimation of structural change in econometric models, Working Paper, Department of Economics, University of California, Berkeley, 1991.
6. M. Basseville and A. Benveniste, Sequential segmentation of nonstationary digital signals using spectral analysis. *Information Sciences* 29: 57–73 (1983).
7. M. Basseville and A. Benveniste, *Detection of Abrupt Changes in Signals and Dynamical Systems*, Springer-Verlag, New York, 1986.
8. P. Bauer, and P. Hackl, The use of MOSUMS for quality control, *Technometrics* 20:431–436 (1978).
9. P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.
10. G. Bodenstein and H. M. Praetorius, Feature extraction from encephalogram by adaptive segmentation, *Proceedings IEEE* 65:642–652 (1977).
11. R. Brown, J. Durbin, and J. Evans, Techniques for testing the constancy of regression relationships over time, *Journal of Royal Statistical Society* Series B37: 149–163 (1975).
12. C. S. Chu, K. Hornik, and C. K. Kuan, MOSUM test for parameter constancy, Working Paper, University of Southern California, 1994.
13. C. S. Chu, K. Hornik, and C. M. Kaun, The moving-estimates test for parameter stability, *Econometric Theory* (1995).
14. C. S. Chu and H. White, A direct test for changing trend, *Journal of Business and Econometric Statistics* 10:289–299 (1992).
15. J. Cuzik, Boundary crossing probabilities for stationary Gaussian processes and Brownian motion, *Transactions in American Mathematics Society* 263:469–492 (1981).
16. J. Deshayes and D. Picard, Off-line statistical analysis of change-point problems using nonparametric and likelihood methods, in *Detection of Abrupt Changes Signals and Dynamical Systems*, M. Basseville and A. Beneniste (Eds.), Springer-Verlag, New York, 1986.
17. J. Doob, Heuristic approach to Kologorov-Smirnov theorems, *Annals of Mathematical Statistics* 20:393–403 (1949).

18. D. Hawkins, A test for a change point in a parametric model based on a maximal Wald-type statistic, *Sankhya* 49:368–376 (1987).
19. D. Hinkley, Inference about the change point in a sequence of random variables, *Biometrika* 57:1-17 (1970).
20. W. Kramer, W. Ploberger, and R. Alt, Testing structural change in dynamic models, *Econometrica* 56:1335–1369 (1988).
21. L. Nunes, C.-M. Kuan, and P. Newbold, Spurious break, Working Paper, Department of Economics, University of Illinois at Urbana-Champaign, 1993.
22. P. Phillips and S. Durlauf, Multiple time series regression with integrated process, *Review of Economic Studies* 53:473–495 (1986).
23. W. Ploberger, W. Kramer, and K. Kontrus, A new test for structural stability in the linear regression model, *Journal of Econometrics* 40:307–318 (1989).
24. W. Ploberger and W. Kramer, The CUSUM test with OLS residuals, *Econometrica* 60:271–285 (1992).
25. W. Ploberger, W. Kramer, and A. Alt, Testing for structural change in dynamic models, *Econometrica* 60:271–285 (1988).
26. D. Pollard, *Convergence of Stochastic Processes*, Springer-Verlag, New York, 1984.
27. R. Quandt, Tests of the hypothesis that a linear regression system obeying two separate regimes, *Journal of the American Statistical Association* 55:324–330 (1960).
28. P. Sen, Asymptotic theory of some tests for a possible change in the regression slope occurring at unknown time point, *Zeitschrift Wahrscheinlichkeitstheorie und Verwandte Gebiete* 52:203–218 (1980).
29. A. S. Willsky, A survey of design methods for failure detection in dynamic systems, *Automatica* 12:601–611 (1976).
30. J. Wooldridge and H. White, Some invariance principles and central limit theorem for dependent heterogeneous processes, *Econometric Theory* 4:210–230 (1988).
31. S. Zacks, Survey of classical and Bayesian approaches to the change point problem: Fixed sample and sequential procedures for testing and estimation, in *Recent Advances in Statistics*, M. H. Rivzi *et al.* (Eds)., 1983.

Received 13 July 1994; revised 13 September 1994, 13 October 1994