Sequential Monitoring Algorithms for Time Series

Edit Gombay

Department of Mathematical and Statistical Sciences, University of Alberta,
425 CAB, Edmonton, AB
T2G 2G1, Canada
egombay@ualberta.ca

Abstract. In this presentation we will give a survey of some recently proposed algorithms that can be used in sequential testing or in sequential monitoring for the presence of a change in some parameters of the process. The underlying theory for weakly dependent time series will be referenced and some empirical results will show the power of these new procedures. Sequential tests under consideration are the continuous versions of the Pocock (1977) and O'Brien and Fleming (1979) type group sequential tests. The sequential monitoring strategy uses a version of Page's (1955) idea. These tests specify a maximal sample size, and are based on type I error considerations.

Keywords. approximations, Brownian motion, change in parameters, linear processes.

1 Introduction and results

Structural stability of observations over time is one of the most important topics in statistics, environmental studies, and econometrics. In this talk we shall define sequential tests and sequential change detection algorithms for various parameters of weakly dependent time series.

Retrospective change detection tests were defined in Gombay (2008) for autoregressive models, and in Gombay and Horváth (2009) for more general models. Sequential change detection algorithms for the autoregressive model were given in Gombay and Serban (2005, 2008), and in Gombay and Horváth (2009) for linear processes.

Sequential tests and sequential change detection algorithms are closely related. In sequential tests the change is at the first observation, so it is a special case of change detection tests, where change in an initial value can happen at any time. Hence, the underlying theory is similar for the two problems. It is described in the following theorem.

Let observations $\{Y_t, t \geq 1\}$ be defined as

$$Y_t = \sum_{i=0}^{\infty} \alpha_i \epsilon_{t-i},\ \sum_{i=1}^{\infty} i|\alpha_i| < \infty, \tag{1}$$

where $\{\epsilon_i\}$ is a sequence of i.i.d.r.v.'s, $\{\alpha_i\}$ is a sequence of constants. We need large sample approximations for the following partial sum sequences that arise in our tests:

$$\sum_{k=1}^{t} Y_k,\ t \geq 1, \tag{2}$$

$$\sum_{k=1}^{t} (Y_k Y_{k-s} - E(Y_k Y_{k-s})),\ t \geq 1,\ \text{for } s \geq 0 \text{ fixed.} \tag{3}$$

Theorem 1.1: Let $\{Y_k\}$ be a linear process as in (1), assume that for the i.i.d. sequence $\{\epsilon_k\}$ we have $E(\epsilon_k) = 0$, $Var(\epsilon_k) = \sigma^2$, $0 < \sigma^2 < \infty$, and $E|\epsilon_k|^\kappa < \infty$.

(i) If $\kappa > 2$, then there exists a Brownian motion $\{W(t),\ t \geq 0\}$, such that

$$|\sum_{k=1}^{[t]} Y_k - \sigma_1 W(t)| = o(t^{1/\nu}),\ a.s$$

for some $\sigma_1 > 0$ and some $\nu > 2$.

(ii) If $\kappa > 4$, then there exists a Brownian motion $\{W(t),\ t \geq 0\}$, such that

$$|\sum_{k=1}^{[t]}(Y_kY_{k-s} - E(Y_kY_{k-s})) - \sigma_2 W(t)| = o(t^{1/\nu}),\ a.s$$

for some $\sigma_2 > 0$ and some $\nu > 2$, where $0 \leq s$ is fixed.

Note, that the rates of approximation in the above theorem are the best possible, they are the same as in case of sums of independent identically distributed random variables. Based on this theorem we can define several sequential tests and sequential change detection algorithms. In both problems the initial value is given, the algorithms monitor the process up to a maximum sample size. The probability of type I error is under control, which is a new feature, as many classical sequential tests were designed so that the stopping time be minimized. The analysis of the new tests in case of i.i.d. observations was given in Gombay (2003), and their properties carry over to the dependent case by virtue of the invariance principle.

The following tests can be used in general for any situation where the test statistic can be approximated by a Brownian motion with the rate of error as in Theorem 1.1. For the practical implementation of these strategies one has to make sure, that the standardizing constant σ_i^2, $i = 1, 2$, can be estimated with precision that will not change the asymptotic distribution.

Let n denote the maximum sample size, or truncation point, and α the level of significance. Recall that observation $y_1, y_2, \ldots,$ are arriving one by one, so at stage k we have observed $y_1, \ldots, y_k$. First we list the sequential tests.

2 Sequential Tests

Assume, that observations in (1) are including a constant term, that is,

$$Y_t(\mu) = \mu + \sum_{i=0}^{\infty} \alpha_i \epsilon_{t-i},\ \sum_{i=1}^{\infty} i|\alpha_i| < \infty. \tag{4}$$

To test sequentially if given value $\mu_0 = E(Y_t)$ is the true mean, define $Stat(k) = k^{-1/2}\sum_{i=1}^{k}(y_i - \mu_0)$, and define with a suitable standardizing estimator $\hat{\sigma}_{1,k}$ the following procedure.

TEST 1. *If for some $k > 1$*

$$\hat{\sigma}_{1,k}^{-1}\,|Stat(k)| \geq CV_1(\alpha, n),$$

then stop and reject the null hypothesis. If $\hat{\sigma}_{1,k}^{-1}\,|Stat(k)| < CV_1(\alpha, n)$ for all $k \leq n$, then no evidence against the null hypothesis has been found.

Critical value $CV_1(\alpha, n)$ can be approximated by the results of Vostrikova (1981) for each n. These approximations are based on the boundary crossing probabilities of Ornstein-Uhlenbeck processes and give less conservative critical values for finite n than the limiting double exponential distribution, as convergence for extremal values is known to be very slow. The one-dimensional approximation in Vostrikova (1981) with $T = \log n$ is

$$P\{\sup_{1<k\leq n} |\hat{\sigma}_{1,k}^{-1}\,|Stat(k)| > y\} \cong \frac{\exp(-y^2/2)y}{\sqrt{2\pi}}\left\{T(1-\frac{1}{y^2}) + \frac{4}{y^2} + O(\frac{1}{y^4})\right\};$$

therefore, $CV_1(\alpha, n)$ can be obtained from this equation.

If a sequential test for the value of $E(Y_kY_{k-s}) = \gamma_0(s)$, $s \geq 0$ is needed, then we replace $Stat(k)$ above by $Stat(k) = k^{-1/2}\sum_{i=1}^{k}(y_iy_{i-s} - \gamma_0(s))$, and replace $\hat{\sigma}_{1,k}$ with an appropriate estimator $\hat{\sigma}_{2,k}$, and perform the thus modified TEST 1. The proper choice of $\hat{\sigma}_{1,k}$ and $\hat{\sigma}_{2,k}$ is crucial for the performance of this test. Berkes et al. (2009) have proposed estimators with small enough errors rates.

TEST 1 is the continuous monitoring version of Pocock's (1977) group-sequential tests, hence it has its advantages of early stopping for large differences between the null-value and the true value of the parameter of interest. The continuous version of O'Brien and Fleming's (1979) group-sequential test requires different standardizing constants. Let $Stat^{(1)}(k) = n^{-1/2}\sum_{i=1}^{k}(y_i - \mu_0)$, or $Stat^{(2)}(k) = n^{-1/2}\sum_{i=1}^{k}(y_i y_{i-s} - \gamma_0(s))$, depending on the parameter of interest. Then for $i = 1, 2$ we have the following procedure.

TEST 2. *If for some k, $1 < k \leq n$,*

$$\hat{\sigma}_{ik}^{-1}|Stat^{(i)}(k)| \geq CV_2(\alpha, n),$$

then stop and reject the null hypothesis. Otherwise no evidence against the null hypothesis has been found.

Critical value $CV_2(\alpha, n) = CV_2(\alpha)$ is obtained from the well-known distribution of $\sup_{0<t<1}|W(t)|$. Note, that it does not depend on n. One-sided versions of these tests can be defined with the appropriate modification of the critical values.

TEST 2 has more sensitivity for small differences between the null and alternative values of the parameters. Hence, the choice between TEST 1 and TEST 2 are made on the same principles as the choice between the Pocock and O'Brien-Fleming tests. More details on the stopping characteristics of the above two testing strategies can be found in Gombay (2002).

Open ended tests, that is sequential tests without a maximal sample size n can be defined using boundary crossing probabilities for the Brownian motion. Such are readily available, see Gombay (2003) for some suggestions and references.

3 Sequential Change Detection

The following two tests use the CUSUM idea of Page (1955) for sequential change detection. When the interest is in the mean of the process we use the following test statistic,

$$Stat(k) = \hat{\sigma}_{1k}^{-1}\sqrt{\frac{k}{n}}\sum_{i=1}^{k}(Y_i - E_0(Y_i).$$

TEST 3. *Stop and conclude that the null hypothesis of no change is not supported by the data at the first k, $1 < k \leq n$, when*

$$\max_{1<j<k} n^{-1/2}\big(Stat(k) - Stat(j)\big) \geq CV_3(\alpha),$$

otherwise do not reject H_0.

The critical value $CV_3(\alpha) = CV_3(\alpha, n)$, all n, can be obtained from the well known distribution of $\sup_{0\leq s<t\leq 1}(W(t) - W(s))$ which is the same as the distribution of $\sup_{0\leq t\leq 1}|W(t)|$, where $W(\cdot)$ is a standard Brownian motion. For example, C(0.10)=1.96, C(0.05)=2.24 and C(0.01)=2.80.

When the interest is in change in the autocovariance at lag s, $s \geq 0$, then let

$$Stat(k) = \hat{\sigma}_{2k}^{-1}\sum_{i=1}^{k}(Y_i Y_{i-s} - E_0(Y_i Y_{i-s}))$$

and use test the following testing procedure.

TEST 4. *Stop and conclude that the null hypothesis of no change is not supported by the data at the first k, $1 < k \leq n$, when*

$$\max_{1<j<k} n^{-1/2}\big(Stat(k) - Stat(j)\big) \geq CV_3(\alpha),$$

otherwise do not reject H_0.

The greatest difficulty in implementing these tests is in the estimation of the standardizing constants σ_{1k} and σ_{2k}. Different aspects of these difficulties were discussed and solved in Berkes et al. (2009).

In Berkes et al. (2009), the use of moving average process approximations were advocated, and they were shown to work well in the retrospective change detection scenario. Earlier, in most works standardization based on approximating the process by autoregressive time series was suggested. It has been remarked in the econometric literature by several authors, that the moving average processes would be more natural, better suited to model many phenomena. However, mathematical convenience associated with an autoregressive model makes it widely applicable. This convenience is due to the existence of the likelihood function (or semi-likelihood, in case of non-normal errors), that leads, among other things, to explicit maximum likelihood estimators.

Moving average processes have m-dependent observations, and for those the strong law of large numbers (and the central limit theorem) can easily be proven. These theorems allow the definition of simple estimators for the standardizing constants. The estimator for σ_{1k} is the empirical variance of $\sum_{i=1}^{k}(Y_i - E_0(Y_i))$, and it turns out to be the same as the Bartlett estimator with uniform kernel. In the estimation of σ_{2k}, we extend this approach to higher moments. Also, by this method we can get an estimator for the covariance matrix, that is useful for simultaneous testing of several parameters.

Example: MA(1) process is $Y_t = \epsilon_t + \theta\epsilon_{t-1}$, so $q = 1$ and the covariance function of the process is $\gamma(0) = \sigma_\epsilon^2(1+\theta^2)$, $\gamma(1) = \theta\sigma_\epsilon^2$, and $\gamma(h) = 0$ for $|h| > 1$. The covariance matrix $\Gamma_{2\times 2}$ is estimated by $\hat{\Gamma}$ with components

$$\hat{\Gamma}_{00} = \sum_{j=1}^{k} Y_j^4 + 2\sum_{j=1}^{k} Y_j^2 Y_{j+1}^2 - 3\frac{1}{k}\left(\sum_{j=1}^{k} Y_j^2\right)^2,$$

$$\hat{\Gamma}_{11} = \sum_{j=1}^{k} Y_j^2 Y_{j-1}^2 +$$

$$+2\left\{\sum_{j=1}^{k} Y_j^2 Y_{j-1} Y_{j+1} + \sum_{j=1}^{k} Y_j Y_{j-1} Y_{j+1} Y_{j+2}\right\} - 5\frac{1}{k}\left(\sum_{j=1}^{k} Y_j Y_{j-1}\right)^2,$$

and

$$\hat{\Gamma}_{01} = \hat{\Gamma}_{10} = \sum_{j=1}^{k} Y_j^3 Y_{j-1} + \sum_{j=1}^{k} Y_j^3 Y_{j+1}$$

$$+\sum_{j=1}^{k} Y_j^2 Y_{j+1} Y_{j+2} + \sum_{j=1}^{k} Y_j^2 Y_{j-1} Y_{j-2} - \frac{4}{k}\left(\sum_{j=1}^{k} Y_j Y_{j-1}\right)\left(\sum_{j=1}^{k} Y_j^2\right).$$

When the process is AR(p), then we have closed form (quasi-)maximum likelihood estimators that converge to the unknown parameters at an optimum rate. This allows us to define the following tests.

3.1 Monitoring the mean μ.

We test the hypothesis of no change $H_0 : \mu_i = \mu_0,\ \ \sigma^2\ and\ \phi\ unknown, for\ all\ \ i \geq 1$, against the alternative

$$H_A : \mu_i = \mu_0,\ \ \sigma^2\ and\ \phi\ \ unknown\ \ for\ \ all\ \ 1 \leq i \leq \tau - 1$$
$$\mu_i = \mu_A,\ \ \sigma^2\ and\ \phi\ \ unknown,\ \ unchanged\ \ for\ \ all\ \ i \geq \tau.$$

The test statistic is based on $\max_{j<k}[W_k(\mu_0, \hat{\sigma}_k^2, \hat{\phi}_k) - W_j(\mu_0, \hat{\sigma}_k^2, \hat{\phi}_k)]$, where

$$W_k(\mu_0, \hat{\sigma}_k^2, \hat{\phi}_k) = \frac{1}{\hat{\sigma}_k} \sum_{i=1}^{k} \left[(Y_i - \mu_0) - \sum_{j=1}^{p} \hat{\phi}_{kj}(Y_{i-j} - \mu_0) \right].$$

Simulations results are shown in Figure 1. It contains eleven lines, each for a different size of change in mean denoted as $m = \mu_A - \mu_0$. Each of these lines shows the power achieved as the coefficient ϕ varies between -0.9 and 0.9. Note that, for fixed truncation point and fixed change-point, the power of the test decreases as the coefficient ϕ increases from -1 to 1. This can be seen in Figure 1 and it happens because the drift D is proportional to $(1 - \phi)$.

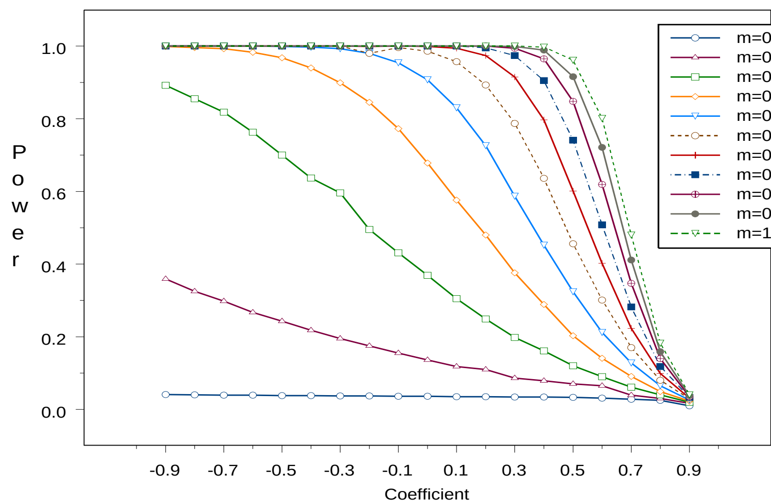

Fig. 1. Power vs Coefficient when testing change in the mean. The truncation point is $n_0 = 200$ and the change point is $\tau = 100$. The in-control value is $\mu_0 = 0$ and the change in mean is $m = \mu_A - \mu_0$.

3.2 Monitoring the variance σ^2.

When monitoring the variance σ^2, the mean μ and ϕ are nuisance parameters. The hypothesis of interest is $H_0 : \sigma_i^2 = \sigma_0^2,\ \mu \ and\ \phi\ unknown\ for\ all\ \ i \geq 1$, and the alternative

$$\begin{aligned} H_A : \sigma_i^2 &= \sigma_0^2, \ \ \mu \ \ and \ \ \phi \ \ unknown \ \ for \ \ all \ \ 1 \leq i \leq \tau - 1 \\ \sigma_i^2 &= \sigma_A^2, \ \ \mu \ \ and \ \ \phi \ \ unknown, \ \ unchanged \ \ for \ \ all \ \ i \geq \tau. \end{aligned}$$

The test statistic is based on $\max_{j<k}[W_k(\sigma_0^2, \hat{\mu}_k, \hat{\phi}_k) - W_j(\sigma_0^2, \hat{\mu}_k, \hat{\phi}_k)]$, where

$$W_k(\sigma_0^2, \hat{\mu}_k, \hat{\phi}_k) = 2^{-1/2}\sigma_0^{-2} \sum_{i=1}^{k} \left\{ \left[(Y_i - \hat{\mu}_k) - \sum_{j=1}^{p} \hat{\phi}_{kj}(Y_{i-j} - \hat{\mu}_k) \right]^2 - \sigma_0^2 \right\}.$$

From simulations studies unreported here one can see that for fixed truncation point and fixed change point, the power of the test remains almost the same for any coefficient ϕ between -1 and 1.

3.3 *Monitoring the coefficient* ϕ.

In this case we test $H_0 : \phi_i = \phi_0\ ,\ \mu\ and\ \sigma^2\ \ unknown\ \ for\ all\ \ i \geq 1,$ against the alternative

$$\begin{aligned} H_A : & \phi_i = \phi_0\ ,\ \mu\ and\ \sigma^2\ \ unknown,\ \ for\ all\ 1 \leq i \leq \tau - 1 \\ & \phi_i = \phi_A\ ,\ \mu\ and\ \sigma^2\ \ \ unknown,\ \ unchanged\ \ for\ all\ i \geq \tau \end{aligned}$$

The test statistic is based on vector $\max_{j<k}[W_k(\phi_0, \hat{\mu}_k, \hat{\sigma}_k^2) - W_j(\phi_0, \hat{\mu}_k, \hat{\sigma}_k^2)]$,

$$W_k(\phi_0, \hat{\mu}_k, \hat{\sigma}_k^2) = \frac{1}{\hat{\sigma}_k} \Gamma^{-1/2}(\phi_0, \hat{\mu}_k \hat{\sigma}_k^2) \nabla_\phi l(\phi_0, \hat{\mu}_k, \hat{\sigma}_k^2),$$

where Γ is the information matrix, and $l(\phi_0, \hat{\mu}_k, \hat{\sigma})$ is the log-likelihood function with estimated nuisance parameters μ and σ^2. The five lines of Figure 2 present the empirical power of the test when the initial in-control value of ϕ_0 is $b =$ -0.9, -0.5, 0, 0.5, and 0.9, respectively. The value of ϕ after change is on the horizontal axis.

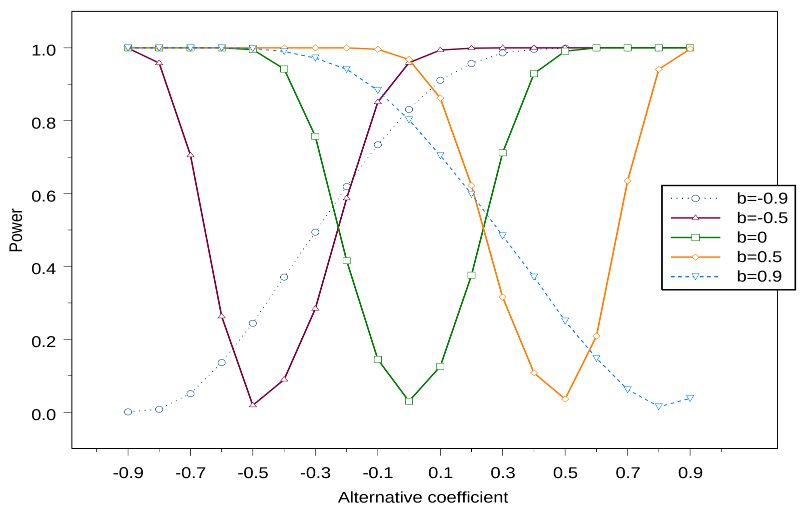

Fig. 2. Power curves when testing for change in the coefficient for different AR(1) models ($b = \phi_0$). The truncation point is $n_0 = 200$ and the change point is $\tau = 100$. The initial coefficient values are b =-0.9, -0.5, 0.0, 0.5 and 0.9.

References

Berkes, I., Gombay, E., and Horváth, L. (2009). Testing for changes in the covariance structure of linear processes. *J. of Statistical Planning and Inference*, **139** 2044–2063.

Gombay, E. (2002). Parametric sequential tests in the presence of nuisance parameters. *Theory Stoch. Processes*, **8** (24) No. 1-2 106-118.

Gombay, E. (2003). Sequential Change-Point Detection and Estimation, *Sequential Analysis*, **22** 203-222.

Gombay, E. (2008). Change detection in autoregressive time series. *J. of Multivariate Analysis*, **99** 451–464.

Gombay, E., and Horváth, L. (2009). Sequential tests and change detection in the covariance structure of weakly stationary time series. *S. Zacks Festschrift* of *Communications in Statistics - Theory & Methods*, to appear.

Gombay, E., and Serban, D. (2005). Monitoring Parameter Change in AR(p) Time Series Models. *Technical Report 05.04* Statistics Centre, University of Alberta, Edmonton, Canada.

Gombay, E., and Serban, D. (2009). Monitoring Parameter Change in AR(p) Time Series Models. *Journal of Multivariate Analysis*, **100** 715-725.

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials, *Biometrics*, **35** 549–556.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, **42** 523–527.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *J.* , *Biometrika*, **64** 191–199.

Vostrikova, L. J. (1981). Detection of a 'disorder' in a Wiener process, *Theory Probability and its Applications*, **26** 356–362.