



ELSEVIER



Journal of Statistical Planning and Inference ■■■ (■■■■) ■■■-■■■

journal of  
statistical planning  
and inference[www.elsevier.com/locate/jspi](http://www.elsevier.com/locate/jspi)

# Covariance changes detection in multivariate time series

Pedro Galeano\*, Daniel Peña

*Departamento de Estadística, Universidad Carlos III de Madrid, c/ Madrid 126, 28903 Getafe, Madrid, Spain*

Received 16 August 2004; accepted 16 September 2005

## Abstract

This paper studies the detection of step changes in the variances and in the correlation structure of the components of a vector of time series. Two procedures based on the likelihood ratio test (LRT) statistic and on a cumulative sums (cusum) statistic are considered and compared in a simulation study. We conclude that for a single covariance change the cusum procedure is more powerful in small and medium samples, whereas the likelihood ratio test is more powerful in large samples. However, for several covariance changes the cusum procedure works clearly better. The procedures are illustrated in two real data examples.

© 2005 Elsevier B.V. All rights reserved.

*Keywords:* Cusum statistic; Heteroskedasticity; Likelihood ratio test statistic; Step changes; VARMA models

## 1. Introduction

The problem of detection of a sudden change in the marginal variance of a univariate time series has been extensively studied. For instance, [Wichern et al. \(1976\)](#) considered a detection procedure for a variance change at an unknown position in a first order autoregressive model. [Abraham and Wei \(1984\)](#) studied inference in the parameters of a time series under the presence of a variance change at an unknown point. [Baufays and Rasson \(1985\)](#) proposed an iterative algorithm for changes in autoregressive models by means of a maximum likelihood procedure. [Tsay \(1988\)](#) studied outliers, level shifts and variance changes in ARIMA models. [Inclán \(1993\)](#) and [McCulloch and Tsay \(1993\)](#) proposed Bayesian procedures for autoregressive processes with variance changes. [Inclán and Tiao \(1994\)](#) proposed an iterative procedure based on a cumulative sum of squares statistic for the detection of several variance changes in Gaussian independent observations. The cusum statistic they proposed for variance changes was related to similar statistics studied for level shifts by [Andrews \(1993\)](#), [Bai \(1994a\)](#), [Bai et al. \(1998\)](#) and [Bai and Perron \(1998\)](#). [Chen and Gupta \(1997\)](#) considered an information theoretic approach based on the Bayesian Information Criteria (BIC) for detection of step changes. [Wang and Zivot \(2001\)](#) proposed a model for time series with changing variance. [Park et al. \(2000\)](#) and [Lee and Park \(2001\)](#) extended the Inclán and Tiao approach to autoregressive and moving average models, respectively.

The case of multivariate sequences has, to the best of our knowledge, not been considered yet. The multivariate case is more challenging than the univariate case because then we can consider changes in both the variances of the components and in the correlation structure among them. The effect of a change in the covariance structure depends on the magnitude of the change and on the parameters of the model. In fact, if we look for variance changes in each

\* Corresponding author. Tel.: +34 91 624 95 53; fax: +34 91 624 98 49.

E-mail address: [pedro.galeano@uc3m.es](mailto:pedro.galeano@uc3m.es) (P. Galeano).

of the components of the series, we can find that several of these variance changes can be explained by a single covariance change. In this article we study the detection of step changes in the variance and in the correlation structure of the components of a vector autoregressive moving average (VARMA) model. Two approaches are introduced and compared. The first one is a likelihood ratio approach and the second is a cusum approach.

The rest of this article is organized as follows. In Section 2 we present the model for covariance changes and two statistics that can be used for testing for the change. In Section 3 we consider the particular case of changes only in the variances of the components, and present two statistics for testing for such a change. In Section 4 we study two different procedures for detection and estimation of these changes. In Section 5 the two procedures are compared in a Monte Carlo experiment for different models, sample sizes, number of changes and location of the change-points. Finally, in Section 6 we illustrate the procedures by means of two real data examples. We conclude that the procedure based on the cusum statistic has an overall better performance than the one based on the likelihood ratio test, whereas for a single change and large sample size the likelihood ratio test appears to be more powerful.

**2. Testing for covariance changes in multivariate time series**

Let  $x_t = (x_{1t}, \dots, x_{kt})'$ ,  $t = 1, \dots, n$ , be a  $k$ -dimensional vector of time series following a vector ARIMA model, given by

$$\Phi(B)x_t = c + \Theta(B)a_t, \tag{1}$$

where  $B$  is the backshift operator,  $Bx_t = x_{t-1}$ ,  $\Phi(B) = I - \Phi_1 B - \dots - \Phi_p B^p$  and  $\Theta(B) = I - \Theta_1 B - \dots - \Theta_q B^q$ , are  $k \times k$  matrix polynomials of finite degrees  $p$  and  $q$ ,  $c$  is a  $k$ -dimensional constant vector, and  $a_t = (a_{1t}, \dots, a_{kt})'$  is a sequence of independent and identically distributed (iid) Gaussian random vectors with zero mean and positive-definite covariance matrix,  $\Sigma$ . We assume that  $\Phi(B)$  and  $\Theta(B)$  are left coprime and that all the zeros of the determinants  $|\Phi(B)|$  are on or outside the unit circle and those of  $|\Theta(B)|$  outside the unit circle. The series  $x_t$  is stationary if  $|\Phi(z)| \neq 0$  for all  $|z| = 1$  and is unit-root nonstationary if  $|\Phi(1)| = 0$ . The autoregressive representation of model (1) is given by  $\Pi(B)x_t = c_{\Pi} + a_t$ , where  $\Pi(B) = \Theta(B)^{-1}\Phi(B) = I - \sum_{i=1}^{\infty} \Pi_i B^i$ , and  $c_{\Pi} = \Theta(1)^{-1}c$  is a vector of constants. We also have the moving-average representation,  $x_t = c_{\Psi} + \Psi(B)a_t$ , where  $\Psi(B) = I + \sum_{i=1}^{\infty} \Psi_i B^i$  is defined by  $\Phi(B)\Psi(B) = \Theta(B)$  and  $c_{\Psi}$  is a constant if the process is stationary and a deterministic function of time otherwise.

Suppose that instead of observing  $x_t$  we observe a time series  $y_t = (y_{1t}, \dots, y_{kt})'$ , defined as follows. Let  $S_t^{(h)}$  be a step function such that  $S_t^{(h)} = 0$ ,  $t < h$ , and  $S_t^{(h)} = 1$ ,  $t \geq h$ . Let  $W$  be a lower triangular matrix of size  $k \times k$  denoting the magnitude of the covariance change. We assume that the innovations affecting the series,  $e_t$ , is not a sequence of iid  $N_k(0, \Sigma)$  variables, due to a change in the covariance of the components at  $t = h$ , given by  $e_t = a_t + WS_t^{(h)}a_t$ . Therefore, the observed vector time series  $y_t = (y_{1t}, \dots, y_{kt})'$  can be written as

$$\Phi(B)y_t = c + \Theta(B)(a_t + WS_t^{(h)}a_t). \tag{2}$$

The relation between the observed series,  $y_t$ , and the unobserved vector ARMA time series,  $x_t$ , is given by

$$y_t = x_t + \Psi(B)WS_t^{(h)}a_t. \tag{3}$$

The covariance of  $e_t$  changes from  $\Sigma$  to  $\Omega = (I + W)\Sigma(I + W)'$  at the time point  $t = h$ . Without loss of generality, it is assumed that  $(I + W)$  is a positive defined matrix so that  $W$  is well identified. For that, let  $\Sigma = L_{\Sigma}L'_{\Sigma}$  and  $\Omega = L_{\Omega}L'_{\Omega}$  be the Cholesky decompositions of  $\Sigma$  and  $\Omega$ , respectively. Then, by taking,

$$W = L_{\Omega}L'_{\Sigma}^{-1} - I, \tag{4}$$

we obtain  $\Omega = (I + W)\Sigma(I + W)'$ , and the matrix  $W$  is unique.

To test the significance of a covariance change at  $t = h$ , suppose that the parameters of the vector ARMA model are known and by using them we compute the innovations

$$e_t = y_t - \sum_{i=1}^p \Phi_i y_{t-i} - c + \sum_{j=1}^q \Theta_j e_{t-j}. \tag{5}$$

We want to test the hypothesis that these innovations are iid homoskedastic, versus the alternative hypothesis that they are heteroskedastic because of a covariance change at a known location  $t = h$ . Thus, in Eq. (2) we consider the null hypothesis  $H_0 : W = 0$  versus the alternative hypothesis  $H_1 : W \neq 0$ . The most usual method for testing the homogeneity of the covariance matrices of two Gaussian populations is the likelihood ratio test (LRT) statistic, which is asymptotically the most powerful test. Let us define the three matrices  $S = \sum_{t=1}^n (e_t e_t')/n$ ,  $S_1 = \sum_{t=1}^h (e_t e_t')/h$  and  $S_2 = \sum_{t=h+1}^n (e_t e_t')/(n - h)$ . The LRT statistic for testing for a covariance change at the time point  $m = h + 1$  of the innovations in (5) is given by

$$LR_h = n \log \frac{|S|}{|S_1|^v |S_2|^{1-v}}, \tag{6}$$

where  $v = h/n$ . Statistic (6), for a fixed time point  $h$ , under the null hypothesis of no covariance change and assuming that the model is known, has an asymptotic chi-squared distribution with  $k(k + 1)/2$  degrees of freedom.

An alternative cusum test statistic can be built as follows. Let  $A_h = \sum_{t=1}^h e_t' \Sigma^{-1} e_t$  be the multivariate cumulative sum of squares of  $\{e_1, \dots, e_h\}$  for a fixed time point  $h$  with  $1 \leq h \leq n$ . Let

$$C_h = \frac{h}{\sqrt{2kn}} \left( \frac{A_h}{h} - \frac{A_n}{n} \right) \tag{7}$$

be the centered and standardized cumulative sum of squares of the sequence  $\{e_1, \dots, e_h\}$ . We study the asymptotic behavior of statistic (7) under the hypothesis of homoskedasticity. Let  $M_v$  be a standard Brownian motion on  $[0, 1]$ . Let  $M_v^0$  denote a Brownian bridge given by  $M_v^0 = M_v - vM_1$ . The asymptotic distribution of the statistic  $C_h$  is obtained in the following theorem, which is proved in the appendix.

**Theorem 1.** *Let  $\{e_1, \dots, e_n\}$  be a sequence of independent, identically distributed Gaussian, random variables with zero mean and common covariance matrix  $\Sigma$ . For  $v = h/n$ , the statistic  $C_h$  in (7) converges weakly to  $M_v^0$ , a standard Brownian bridge on  $[0, 1]$ .*

A conclusion from Theorem 1 is that we can test for the presence of a covariance change at the time point  $m = h + 1$  in the innovations in (5), by obtaining the value of the statistic  $C_h$  in (7) and comparing it with the percentiles of the distribution of a Brownian bridge at the time point  $v = h/n$ , which is a normal distribution with zero mean and variance  $v(1 - v)$ . We note that the Brownian bridge limit is operational only to test for a break of known location.

Finally, the magnitude of a covariance change is estimated from (4) by

$$\widehat{W} = L_{S_2} L_{S_1}^{-1} - I. \tag{8}$$

### 3. The case of variance changes

When  $W$  is a diagonal matrix, the innovations have a change only in their variances but their correlations remain constant. The observed vector time series  $y_t = (y_{1t}, \dots, y_{kt})'$  can be written as in (3) with  $W$  diagonal and the covariance matrix of  $e_t$  at the time point  $t = h$  changes from  $\Sigma$  to  $\Omega = (I + W)\Sigma(I + W)$ . We assume that  $(I + W)$  is a positive defined matrix, so that the matrix  $W$  is well identified. To show this, let  $\Sigma = D_\Sigma R_\Sigma D_\Sigma$  and  $\Omega = D_\Omega R_\Omega D_\Omega$  be the spectral decompositions of the matrices  $\Sigma$  and  $\Omega$ , respectively, where  $R_\Sigma$  and  $R_\Omega$  are the correlation matrices of  $\Sigma$  and  $\Omega$ , which are assumed to be equal, and  $D_\Sigma$  and  $D_\Omega$  are diagonal matrices whose elements are the standard deviations of each component. By taking  $W = D_\Omega D_\Sigma^{-1} - I$ , with  $W$  diagonal, we obtain  $\Omega = (I + W)\Sigma(I + W)$ , and the matrix  $W$  is unique because the Cholesky decomposition of a matrix is unique. We note that this change may affect one or several components and the elements different from 0 of  $W$  indicate the components with changing variance.

As in the previous case, to test the significance of a change at  $m = h + 1$ , suppose that the parameters of the vector ARMA model are known and by using them we compute the innovations as in (5). We consider the null hypothesis  $H_0 : W = 0$  versus the alternative hypothesis  $H_1 : W \neq 0$  in Eq. (2) with  $W$  diagonal. Let us define the variances  $s^2(i) = \sum_{t=1}^n (e_{it}^2)/n$ ,  $s_1^2(i) = \sum_{t=1}^h (e_{it}^2)/h$  and  $s_2^2(i) = \sum_{t=h+1}^n (e_{it}^2)/(n - h)$ , for  $i = 1, \dots, k$ . The LRT statistic of

the innovations in (5) for a variance change at the time point  $m = h + 1$  is given by

$$LR_h = n \log \frac{s^2(1) \cdots s^2(k)}{(s_1^2(1) \cdots s_1^2(k))^v (s_2^2(1) \cdots s_2^2(k))^{1-v}}, \tag{9}$$

where  $v = h/n$ . Statistic (9), for a fixed time point  $h$ , under the null hypothesis of no variance change and assuming that the model is known, has an asymptotic chi-squared distribution with  $k$  degrees of freedom.

A cumulative sums statistic can be built as follows. Under the null hypothesis of homoskedasticity, define  $b_t = D_\Sigma^{-1} e_t$  with  $Cov(b_t) = R_\Sigma$ . The principal components of the series  $b_t$  are given by  $c_t = U_\Sigma b_t$ , where  $U_\Sigma$  is the matrix whose columns are the eigenvectors of the matrix  $R_\Sigma$ , and  $Cov(c_t) = \Lambda$ , which is a diagonal matrix. The components of  $c_t$  are uncorrelated with variances equal to the elements of the matrix  $\Lambda$ . Let  $A_h = \sum_{t=1}^h c_t' c_t$  be the multivariate cumulative sums of squares of the sequence  $\{c_1, \dots, c_h\}$  for a fixed  $h$  with  $1 \leq h \leq n$ . Let

$$C_h = \frac{h}{\sqrt{2tr(\Lambda^2)n}} \left( \frac{A_h}{h} - \frac{A_n}{n} \right) \tag{10}$$

be the centered and standardized cumulative sums of squares of the sequence  $\{e_1, \dots, e_h\}$ . The asymptotic distribution of statistic (10) under the hypothesis of homoscedasticity can be obtained similarly to the case of changing covariance, as it is shown in the following theorem, which is proved in the appendix.

**Theorem 2.** *Let  $\{e_1, \dots, e_n\}$  be a sequence of independent identically distributed Gaussian random variables with zero mean and common covariance matrix  $\Sigma$ . Let  $b_t = D_\Sigma^{-1} e_t$  with  $Cov(b_t) = \Lambda$ , where  $D_\Sigma$  is a diagonal matrix with elements the standard deviations of the variances of the components of  $e_t$  and let  $c_t$  be the principal components of the series  $b_t$ . Therefore, the statistic  $C_h$  in (10) converges weakly to  $M_v^0$ , a standard Brownian bridge on  $[0, 1]$ .*

Thus, to test for the presence of a variance change at the time point  $m = h + 1$  of the innovations in (5), we obtain the value of the statistic  $C_h$  in (10) and compare it with the percentiles of a  $N(0, v(1 - v))$ , the distribution of a Brownian bridge at the time point  $v = h/n$ . As in the case of covariance changes, we note that the Brownian bridge limit is operational only to test for a break of known location.

The magnitude of a variance change is estimated as follows:

$$\widehat{W} = \text{diag}(S_1)^{-1} \text{diag}(S_2) - I, \quad i = 1, \dots, k. \tag{11}$$

Under the null hypothesis of no variance change,  $(1 + \widehat{W}(i, i))^2$  is distributed as an  $F$  distribution with  $(n - h + 1, h - 1)$  degrees of freedom, so that we can test the null hypothesis of  $W(i, i) = 0$  by using this distribution. A confidence interval for  $W(i, i)$ , at significant level  $\alpha$ , is given by

$$1 - \alpha = P \left( \frac{1 + \widehat{W}(i, i)}{\sqrt{F_{(n-h+1, h-1)}^{1-\alpha/2}}} - 1 \leq W(i, i) \leq \frac{1 + \widehat{W}(i, i)}{\sqrt{F_{(n-h+1, h-1)}^{\alpha/2}}} - 1 \right),$$

where  $F_{(n-h+1, h-1)}^{\alpha/2}$  and  $F_{(n-h+1, h-1)}^{1-\alpha/2}$  are the critical values of the  $F$  distribution with  $n - h + 1$  and  $h - 1$  degrees of freedom, for significance levels  $\alpha/2$  and  $1 - \alpha/2$ , respectively.

#### 4. Procedures for covariance change detection

A series can be affected by several covariance or variance changes. In this case, we observe a time series  $y_t = (y_{1t}, \dots, y_{kt})'$  defined as follows:

$$y_t = x_t + \Psi(B)(I + W_r S_t^{(h_r)}) \cdots (I + W_1 S_t^{(h_1)}) a_t,$$

where  $\{h_1, \dots, h_r\}$  are the time of  $r$  changepoints and  $W_1, \dots, W_r$  are  $k \times k$  lower triangular or diagonal matrices denoting the impact of the  $r$  changes. Assuming that the parameters are known, the filtered series of innovations is

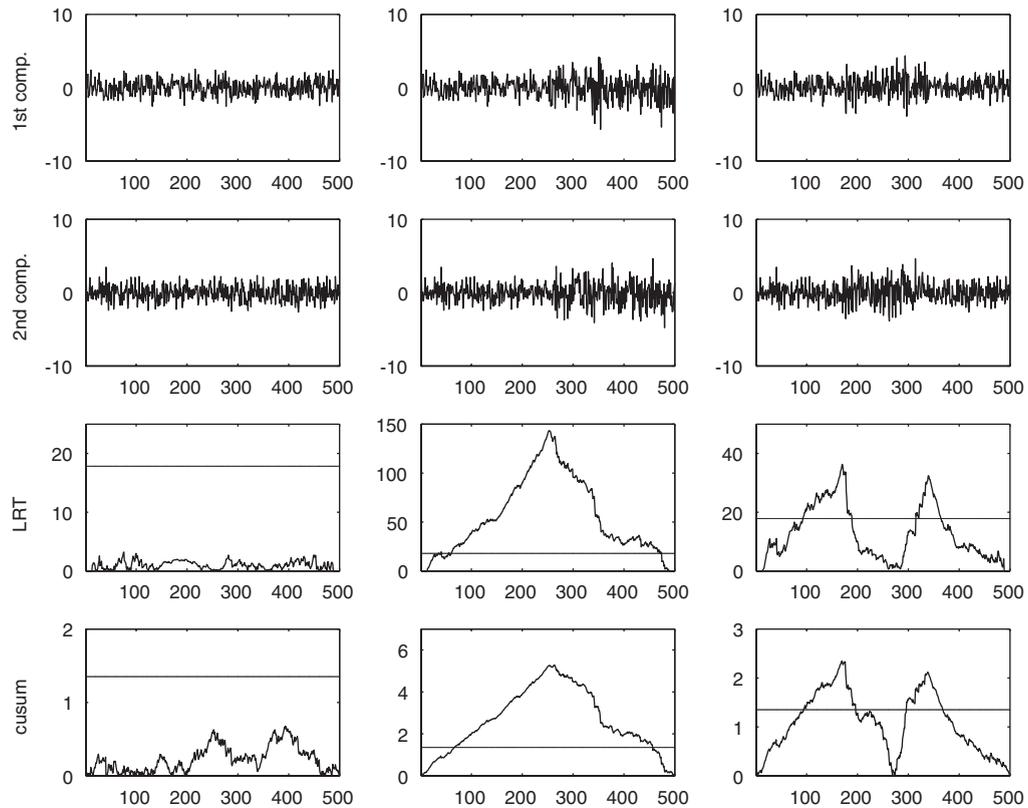


Fig. 1. Three bivariate series and two statistics for covariance change detection. The first column is the no changes case, the second column is the one change case at  $t = 250$  and the third is the two changes case at  $t = 166$  and  $t = 333$ .

given by

$$e_t = (I + W_r S_t^{(h_r)}) \cdots (I + W_1 S_t^{(h_1)}) a_t,$$

and the innovational covariance matrix of  $e_t$  changes from  $\Sigma$  to  $(I + W_1)\Sigma(I + W_1)'$  at  $t = h_1$ , and to  $(I + W_2)(I + W_1)\Sigma(I + W_1)'(I + W_2)'$  at  $t = h_2$ , and so on.

In practice, the number, location and sizes of the covariance changes are unknown. To motivate the proposed procedures, let us consider a bivariate series generated by a first order vector autoregressive model. We consider three different situations which are illustrated in the matrix of plots in Fig. 1. The first column corresponds to the case of no covariance changes. The second column corresponds to the case of a single change at  $t = 250$ , where the variances and the covariances of the components goes to 3 and 1, respectively. The third column corresponds to the case of two changes at  $t = 166$ , where the innovational covariance matrix changes as in the previous case, and at  $t = 333$ , where the innovational covariance matrix goes back to be  $I$ . The first (second) rows in Fig. 1 shows a sample of 500 observations of the first (second) component of this bivariate series, and the third and fourth rows show the LRT statistic (6) and the absolute value of the cusum statistic (7), respectively. In the first column in Fig. 1, no covariance change case, the two statistics plotted in the third and fourth rows are always under the two straight lines corresponding to the 95% critical values of the distributions of the maximums of statistic (6) and the absolute value of statistic (7), computed as we will explain next. In the second column the maximum of both statistics is around  $t = 250$ , the time of the single covariance change, and we obtain values larger than the corresponding critical values, so that the hypothesis of no change is rejected in both cases. In the third column, these changes appear as two significant extremes around the times of the changes,  $t = 166$  and  $t = 333$ . These results suggest to look for the maximum of statistics (6) and (7) to discover the location of a change point.

We define the statistics,

$$A_{\max} = \max_{1 \leq t \leq n} LR_t, \quad \Gamma_{\max} = \max_{1 \leq t \leq n} |C_t|, \tag{12}$$

and let  $h_{\max}^{LR}$  and  $h_{\max}^C$  be the time indices in which the maximums are attained, respectively. The time indices  $h_{\max}^{LR} + 1$  and  $h_{\max}^C + 1$  are the estimates of the time of the change using the LRT and cusum statistics, respectively. The distribution of  $A_{\max}$  in (12) is intractable, and critical values are obtained by simulation. The distribution of  $\Gamma_{\max}$  in (12) is asymptotically the distribution of  $\sup\{|M_v^0| : 0 \leq v \leq 1\}$ , which is given by (see, Billingsley, 1968, p. 85)

$$P\{\sup |M_v^0| \leq a : 0 \leq v \leq 1\} = 1 + 2 \sum_{i=1}^{\infty} (-1)^i \exp(-2i^2 a^2),$$

and critical values can be obtained from this distribution. Note that the statistics in (12) can be obtained for covariance changes as well as for variance changes. In the latter case, statistics (6) and (7) are replaced by statistics (9) and (10).

It is important to note that, in practice, the model should be identified and their parameters estimated. The effects of model selection and parameter estimation in the statistics in (12) are very complex. In the case of the cusum statistics, Bai (1993, 1994b) and Ling (1998) showed that the asymptotic distribution of the cusum statistics for detecting a change in the level of stationary and non-stationary time series, respectively, are not affected by using estimated residuals. Park et al. (2000) showed similar results for a change in the variance of autoregressive models fitted by least squares. The case of covariance change detection in multivariate time series is much more difficult than in the previous cases because the statistic  $\Gamma_{\max}$  depends on estimated covariance matrices. We carried out Monte Carlo experiments that suggested that the statistic  $\Gamma_{\max}$  using the true model and parameters and using identified models and estimated parameters have the same asymptotic distribution. The critical values of the LRT statistics are obtained via simulation using estimated residuals.

In practice, we will compute the statistics in (12) for  $d + 1 \leq t \leq n - d$ , where  $d$  is a positive integer denoting the minimum number of residuals needed to estimate the parameters of the model. We have taken  $d$  as the number of parameters to estimate plus one in each case,

$$d = k(p + q + 1) + \frac{k(k + 1)}{2} + 1, \quad d = k(p + q + 1) + k + 1,$$

for covariance and variance changes, respectively. Usually, the limiting properties of the maximum test statistics may differ if one trims the top and bottom  $\alpha\%$  of the sample (see, for instance, Gombay and Horvath, 1990). In our case we choose  $d$  independent of the sample size and as  $d/n \rightarrow 0$ , the limiting distribution does not change. The empirical critical values of the LRT statistics are computed taking this proposed trimming into account.

If several changes have occurred in the series, we propose iterative procedures based on the statistics  $A_{\max}$  and  $\Gamma_{\max}$  to detect them and estimate their impacts. We present the procedures for the case of covariance changes, while the case of variance changes is considered later.

#### 4.1. LRT procedure

The following procedure is a generalization to the one in Tsay (1988) for univariate time series. The algorithm is based on cleaning the series after finding a change and proceeds as follows:

1. Assuming no covariance changes, specify a vector ARIMA model for the observed series  $y_t$ . Obtain the maximum likelihood estimates of the model and the filtered series of residuals, which we denote by  $\hat{e}_t$ . Define  $y_t^* = y_t$ .
2. Compute the statistics  $LR_h$ ,  $h = d + 1, \dots, n - d$ , by using the residuals obtained in Step 1. Obtain the statistic  $A_{\max}$  in (12) and denote by  $h_{\max}^{LR}$  the time index in which  $A_{\max}$  is attained. Compare  $A_{\max}$  with a specified critical value  $C$  for a given critical level. If  $A_{\max} < C$ , assume that there is not a covariance change and the procedure stops. If  $A_{\max} \geq C$ , assume that there is a possible covariance change at time  $t = h_{\max}^{LR} + 1$ .
3. Estimate the matrix  $W$  with (8), and compute a modified residual series as follows:

$$e_t^* = \begin{cases} \hat{e}_t, & t < h_{\max}^{LR} + 1, \\ (I + \hat{W})^{-1} \hat{e}_t, & t \geq h_{\max}^{LR} + 1, \end{cases}$$

and the corrected time series  $y_t$  by

$$y_t = \begin{cases} y_t^*, & t < h_{\max}^{LR} + 1, \\ \hat{c} + \hat{\Phi}_1 y_{t-1}^* + \dots + \hat{\Phi}_p y_{t-p}^* + e_t^* - \hat{\Theta}_1 e_{t-1}^* - \dots - \hat{\Theta}_q e_{t-q}^*, & t \geq h_{\max}^{LR} + 1, \end{cases}$$

where the polynomials  $\hat{\Phi}(B)$  and  $\hat{\Theta}(B)$  are the maximum likelihood estimates of the parameters. Then, go back to Step 1 considering  $y_t$  as the observed process.

4. When no more covariance changes are detected, estimate the parameters of the series and all the covariance changes found in the previous steps by using the model

$$\Phi(B)y_t = c + \Theta(B)(I + W_r S_t^{(h_r)}) \dots (I + W_1 S_t^{(h_1)})a_t. \tag{13}$$

This joint estimation is carried out in two steps. First, estimate the parameters assuming no covariance changes and then estimate the matrices  $W_i$ . After that, correct the series, and repeat these two steps until convergence.

#### 4.2. Cusum procedure

The following procedure is a generalization to the one proposed by [Inclán and Tiao \(1994\)](#) for univariate time series. The algorithm is based on successive divisions of the series into two pieces when a change is detected and proceeds as follows:

1. Assuming no covariance changes, specify a vector ARIMA model for the observed series  $y_t$ . Obtain the maximum likelihood estimates of the model and the filtered series of residuals,  $\hat{e}_t$ . Let  $t_1 = 1$ .
2. Compute the statistics  $C_h$ ,  $h = d + 1, \dots, n - d$ , using the residuals obtained in Step 1. Obtain  $\Gamma_{\max}$  in (12) and denote by  $h_{\max}^C$  the time index in which  $\Gamma_{\max}$  is attained. If  $\Gamma_{\max} < C$ , assume that there is not a covariance change and the procedure stops. If  $\Gamma_{\max} \geq C$ , where  $C$  is a specified critical value for a given critical level, go to Step 3.
3. Step 3 has three substeps:
  - (a) Obtain  $\Gamma_{\max}$  for  $t = 1, \dots, t_2$ , where  $t_2 = h_{\max}^C$ . If  $\Gamma_{\max} > C$ , redefine  $t_2 = h_{\max}^C$  and repeat Step 3(a) until  $\Gamma_{\max} < C$ . Define  $h_{\text{first}} = t_2$ , where  $t_2$  is the last value such that  $\Gamma_{\max} > C$ .
  - (b) Obtain  $\Gamma_{\max}$  for  $t = t_1, \dots, n$ , where  $t_1 = h_{\max}^C + 1$  in Step 2. If  $\Gamma_{\max} > C$ , redefine  $t_1 = h_{\max}^C$  and repeat Step 3(b) until  $\Gamma_{\max} < C$ . Define  $h_{\text{last}} = t_1$ , where  $t_1$  is the last value such that  $\Gamma_{\max} > C$ .
  - (c) If  $|h_{\text{last}} - h_{\text{first}}| < d$ , there is just one change point and the algorithm stops here. Otherwise, keep both values as possible change points and repeat Steps 2 and 3 for  $t = h_{\text{first}}, \dots, h_{\text{last}}$ , until no more possible changes are detected. Then, go to Step 4.
4. Define a vector  $\ell = (\ell_1, \dots, \ell_s)$  where  $\ell_1 = 1$ ,  $\ell_s = n$  and  $\ell_2, \dots, \ell_{s-1}$  are the points detected in Steps 2 and 3 in increasing order. Obtain the statistic  $\Gamma_{\max}$  in each interval  $(\ell_i, \ell_{i+2})$  and check if it is significant. If not, eliminate the corresponding point. Repeat Step 4 until the number of possible change points does not change, and the points found in previous iterations do not differ from those in the last one. The vector  $(\ell_2 + 1, \dots, \ell_{s-1} + 1)$  are the points of covariance changes.
5. Finally, estimate the parameters of the series and the magnitudes of the covariance changes detected in the previous steps jointly by using (13).

Some comments with regard to these algorithms are in order. First, the procedures for variance changes are similar but with statistics (9) and (10), and the magnitudes of the changes are estimated by using (11). Second, the critical values for the LRT statistics have to be computed by simulation as we will study in Section 5, while those of the cusum procedure are the asymptotic critical values of the maximum of the absolute value of a Brownian Bridge. Third, in both algorithms we require a minimum distance between changes, larger than  $d$ . If several changes are found in an interval smaller than  $d$ , these changes will be considered as outliers and estimated by the procedure proposed in [Tsay et al. \(2000\)](#). Third, the last step in the LRT procedure is needed for avoiding bias in the size of the estimated covariance changes. Note that in Step 4, when a covariance change is detected, its magnitude is estimated and a new series is defined by using this estimation. Thus, for instance, if there are two covariance changes, the magnitude of the first change is estimated without taking into account the second one. Then, a new series is obtained based on the biased estimated covariance matrix, which introduces bias in the estimation of the magnitude of the second change. In conclusion, a joint estimation is needed to avoid biases in the estimation of the magnitudes of the changes.

Table 1  
Models for the simulation study

$k = 2$		$k = 3$			
$\Phi$	$\Sigma$	$\Phi$	$\Sigma$		
$\begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 0.4 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.6 & 0.2 & 0 \\ 0.2 & 0.4 & 0 \\ 0.6 & 0.2 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$		

Table 2  
Empirical quantiles of the  $A_{\max}$  and  $\Gamma_{\max}$  statistics based on 10 000 realizations

Probability-LRT						Probability-cusum				
<i>W diagonal</i>										
$k = 2$	50%	90%	95%	97.5%	99%	50%	90%	95%	97.5%	99%
$n = 100$	9.71	15.09	16.94	18.62	21.03	0.75	1.12	1.27	1.38	1.49
$n = 200$	10.42	15.49	17.13	18.89	21.65	0.77	1.14	1.28	1.39	1.51
$n = 500$	11.26	16.46	18.09	19.87	22.32	0.79	1.19	1.33	1.44	1.56
$n = 1000$	11.72	17.07	19.00	21.02	23.31	0.80	1.21	1.34	1.47	1.60
$n = \infty$	–	–	–	–	–	0.82	1.22	1.35	1.48	1.62
$k = 3$	50%	90%	95%	97.5%	99%	50%	90%	95%	97.5%	99%
$n = 100$	11.98	17.70	20.00	21.95	23.93	0.75	1.13	1.26	1.37	1.49
$n = 200$	13.17	18.88	20.99	22.96	25.45	0.76	1.16	1.29	1.41	1.55
$n = 500$	13.99	19.73	21.48	23.57	25.74	0.78	1.17	1.31	1.43	1.56
$n = 1000$	14.81	20.75	22.70	24.44	27.52	0.79	1.20	1.34	1.44	1.61
$n = \infty$	–	–	–	–	–	0.82	1.22	1.35	1.48	1.62
<i>W lower triangular</i>										
$k = 2$	50%	90%	95%	97.5%	99%	50%	90%	95%	97.5%	99%
$n = 100$	12.06	17.83	20.17	22.15	25.01	0.75	1.13	1.28	1.40	1.53
$n = 200$	12.61	18.30	20.30	22.40	25.30	0.78	1.17	1.29	1.43	1.55
$n = 500$	13.47	19.14	21.15	22.97	25.53	0.79	1.18	1.32	1.44	1.58
$n = 1000$	13.70	19.79	21.76	23.73	25.75	0.80	1.20	1.33	1.45	1.59
$n = \infty$	–	–	–	–	–	0.82	1.22	1.35	1.48	1.62
$k = 3$	50%	90%	95%	97.5%	99%	50%	90%	95%	97.5%	99%
$n = 100$	17.29	24.44	26.75	28.63	31.94	0.76	1.13	1.27	1.41	1.54
$n = 200$	18.29	25.25	27.70	29.48	32.20	0.78	1.15	1.28	1.43	1.56
$n = 500$	19.31	26.09	28.55	30.77	33.03	0.79	1.17	1.29	1.44	1.57
$n = 1000$	19.94	26.60	28.99	30.97	33.80	0.81	1.21	1.34	1.46	1.60
$n = \infty$	–	–	–	–	–	0.82	1.22	1.35	1.48	1.62

### 5. Monte Carlo results

The Monte Carlo results in this section and the analysis of real data examples in the next one have been carried out by means of various routines written by the authors in MATLAB (developed by The MathWorks, Inc). We first obtain critical values for the statistic  $A_{\max}$  in (12) for  $W$  diagonal and  $W$  lower triangular by simulating from the vector AR(1) models in Table 1, where  $k = 2, 3$  and sample sizes  $n = 100, 200, 500$  and  $1000$ . For each model and sample size, we generate 10 000 realizations, estimate a vector AR(1) model, obtain the residuals,  $\hat{\epsilon}_t$ , and compute the statistic  $A_{\max}$ . Table 2 provides some quantiles of the distribution of  $A_{\max}$  for both models and different sample sizes under the null hypothesis of no covariance change in the sample. Note that the quantiles depend on the time series dimension. The asymptotic distribution of the statistic  $\Gamma_{\max}$  is known but we also study the finite sample behavior of the quantiles of this statistic, see Table 2. As we can see, the finite sample quantiles are always smaller than the asymptotic ones, implying

Table 3  
Covariance matrices for the changes

$\Omega_1$	$\Omega_2$	$\Omega_3$	$\Omega_4$
$\begin{vmatrix} 2 & 0.5 \\ 0.5 & 2 \end{vmatrix}$	$\begin{vmatrix} 0.5 & -0.125 \\ -0.125 & 0.5 \end{vmatrix}$	$\begin{vmatrix} 2 & 0.5 & 0.5 \\ 0.5 & 2 & 0.5 \\ 0.5 & 0.5 & 2 \end{vmatrix}$	$\begin{vmatrix} 0.5 & -0.125 & -0.125 \\ -0.125 & 0.5 & -0.125 \\ -0.125 & -0.125 & 0.5 \end{vmatrix}$

Table 4  
Results for model 1 and one covariance change

$\Omega$	$n$	$h$	LRT procedure					Cusum procedure				
			Frequency			$\hat{h}$		Frequency			$\hat{h}$	
			0	1	$\geq 2$	Med	Mad	0	1	$\geq 2$	Med	Mad
$I$	100	–	95.6	4.4	0.0	–	–	97.2	2.8	0.0	–	–
$I$	200	–	95.4	4.6	0.0	–	–	97.1	2.9	0.0	–	–
$I$	500	–	95.2	4.5	0.3	–	–	97.0	2.6	0.4	–	–
$\Omega_1$	100	25	69.2	30.3	0.5	26	6	60.4	39.4	0.2	39	10
		50	46.6	53.4	0.0	50	5	18.2	80.6	1.2	52	3
		75	60.4	39.2	0.4	75	3	30.6	69.1	0.3	75	3
$\Omega_1$	200	50	29.1	69.3	1.6	51	5	18.4	79.3	2.3	61	10
		100	8.8	88.1	3.1	101	4	1.4	96.0	2.6	101	3
		150	16.8	81.8	1.4	150	5	4.8	92.0	3.2	148	5
$\Omega_1$	500	125	0.0	96.2	3.8	126	4	0.0	95.4	4.6	137	11
		250	0.0	97.4	2.6	251	3	0.0	93.8	6.2	252	4
		375	0.2	95.8	4.0	376	4	0.2	95.8	4.0	375	5
$\Omega_3$	100	25	63.6	35.9	0.5	25	3	46.6	51.8	1.6	27	4
		50	52.0	47.0	1.0	49	5	25.6	71.4	3.0	48	3
		75	65.0	34.6	0.4	75	4	64.2	35.1	0.7	65	8
$\Omega_3$	200	50	17.2	80.4	2.4	50	3	5.4	90.6	4.0	51	4
		100	4.8	91.7	3.5	99	3	0.6	94.8	4.6	99	3
		150	19.8	77.8	2.4	149	4	14.4	81.5	4.1	141	8
$\Omega_3$	500	125	0.0	97.6	2.4	124	2	0.2	94.0	5.8	125	3
		250	0.0	97.5	2.5	249	3	0.0	93.8	6.2	249	3
		375	0.0	97.0	3.0	374	3	0.0	95.7	4.3	367	8

Med is the median and Mad, the median of the absolute deviations from the Med.

that the use of the asymptotic quantile is a conservative decision and, therefore, the type I error will not increase. Note also that the quantiles do not depend on  $k$ .

Now, we consider the case of covariance changes and make a Monte Carlo analysis in order to study the size and power of both procedures. For that, we consider the models in Table 1 and sample sizes,  $n = 100, 200$  and  $500$ . For the case of one covariance change, for each  $n$ , we consider three locations of the change point,  $h = [0.25n], [0.50n]$  and  $[0.75n]$ . The changes are introduced by transforming the original covariance matrix,  $\Sigma = I$ , into one of the matrices  $\Omega_i, i = 1, 2, 3, 4$ , in Table 3. For each case, we generate 5000 realizations. Then, we apply the two procedures with the 95% critical values from Table 2. The results are shown in Tables 4 and 5, where columns 4–6 and 9–11 report the number of covariance changes detected by the algorithms and columns 7, 8, 12 and 13 show the median and the mean absolute deviation of the estimates of the change points for each case. The cases with  $\Omega = I$  indicate the type I error of the procedures, which is around 5% in all the sample sizes considered. From these two tables we conclude that when  $n = 100$  the cusum procedure appears to work better than the LRT procedure. For  $n = 200$ , the cusum procedure is slightly better than the LRT one, but for a larger change  $\Omega_2$  and  $\Omega_4$ , the LRT seems to be slightly more powerful. In

Table 5  
Results for model 2 and one covariance change

$\Omega$	$n$	$h$	LRT procedure					Cusum procedure				
			Frequency			$\hat{h}$		Frequency			$\hat{h}$	
			0	1	$\geq 2$	Med	Mad	0	1	$\geq 2$	Med	Mad
$I$	100	–	94.4	5.4	0.2	–	–	97.0	3.0	0.0	–	–
$I$	200	–	94.4	5.3	0.3	–	–	96.0	3.8	0.2	–	–
$I$	500	–	95.2	4.8	0.0	–	–	95.5	4.5	0.0	–	–
$\Omega_2$	100	25	58.3	40.7	1.0	26	5	49.1	49.7	1.2	34	7
		50	34.5	64.9	0.6	50	3	10.4	88.0	1.6	51	3
		75	47.5	51.5	1.0	75	5	18.4	81.6	0.0	74	3
$\Omega_2$	200	50	13.2	85.3	1.5	50	4	5.4	92.4	2.2	57	7
		100	3.0	95.6	1.4	101	3	0.2	97.5	2.3	101	3
		150	9.6	88.5	1.9	151	3	0.4	97.4	2.2	150	3
$\Omega_2$	500	125	0.0	97.6	2.4	125	3	0.0	91.3	8.7	133	8
		250	0.0	96.5	3.5	250	3	0.0	94.4	5.6	252	3
		375	0.0	96.8	3.2	375	2	0.0	94.2	5.8	375	3
$\Omega_4$	100	25	31.5	68.1	0.4	24	2	21.2	77.6	1.2	26	3
		50	14.2	84.3	1.5	50	2	8.4	89.5	2.1	49	2
		75	36.9	60.5	2.6	74	2	40.1	57.7	2.2	70	5
$\Omega_4$	200	50	1.8	95.2	3.0	50	2	0.0	97.0	3.0	51	3
		100	0.0	97.1	2.9	100	2	0.0	94.3	5.7	99	2
		150	3.4	94.4	2.2	150	2	4.2	89.8	6.0	145	5
$\Omega_4$	500	125	0.0	97.2	2.8	125	2	0.0	94.7	5.3	125	2
		250	0.0	97.6	2.4	250	1	0.0	92.5	7.5	249	2
		375	0.0	97.4	2.6	375	2	0.0	93.6	6.4	369	6

Med is the median and Mad, the median of the absolute deviations from the Med.

almost all the cases, the time of the changes is better estimated with the LRT procedure. Also for  $n = 500$ , the LRT procedure is slightly better than the cusum one.

For two change points, we consider the same sample sizes and changes at  $h_1 = [0.33n]$  and  $h_2 = [0.66n]$ . Each change point is associated with two matrices  $\Omega_i$ ,  $i = 1, 2, 3, 4$ , which give the residual covariance matrices after each change. Four combinations are considered. For each case, we generate 5000 realizations with the corresponding changes. Then, we apply the two procedures with the 95% critical values from Table 2. The results are shown in Tables 6 and 7. Columns 6–9 in these tables are the number of covariance changes detected by the algorithms, and columns 10–13 show the median and the mean absolute deviation of the estimates of the changepoints. For two changepoints, the advantage of the cusum procedure over the LRT one is clearer. Note that, first, the detection frequency of two change points is larger for the cusum procedure in almost all the cases, and, second, the LRT procedure suffers of an overestimation of the number of changes in some situations. When  $k = 2$ , the sample size is small ( $n = 100$ ) and small changes ( $\Omega_2$  and  $I$ ), the detection frequency of the cusum procedure is low: 9.6% and 17.4% for  $k = 2$  and  $k = 3$ , respectively. In the rest of the cases, the cusum procedure works quite well, with several cases over the 90% of detection frequency. As in the previous case, as the sample size increases, the change is larger and the number of components increases, the procedure works better. It also appears that the estimate of the second changepoint has smallest Mad, suggesting that the procedure detects more precisely the change at the end of the series. The medians of the estimates are quite closed to the real changepoints, except with the smallest sample size and the smallest changes.

Now, we study the case of changes in the variances. We make a simulation in order to study the power of the proposed procedures for a single change. We consider the same models and sample sizes for  $k = 2$  as in the previous case, and the changes are introduced by transforming the original covariance matrix,  $\Sigma = I$ , into one of the matrices,  $\Omega_5 = 2 \times I$  and  $\Omega_6 = 0.5 \times I$ . For each case, we generate 5000 realizations. Then, we apply the two procedures with the 5% critical values from Table 2. The results are shown in Table 8, with the same design as before. The case with  $\Omega = I$  shows the

Table 6  
Results for model 1 and two covariance changes

$\Omega$	$\Omega$	$n$	$h_1$	$h_2$	Frequency				$\hat{h}_1$		$\hat{h}_2$	
					0	1	2	$\geq 3$	Med	Mad	Med	Mad
<i>LRT procedure</i>												
$\Omega_1$	$I$	100	33	66	83.6	12.4	4.0	0.0	29	4	66	2
		200	66	133	61.1	20.0	18.8	0.0	66	4	133	4
		500	166	333	9.6	3.6	84.3	2.5	165	4	335	2
$\Omega_2$	$I$	100	33	66	90.2	9.2	0.5	0.1	30	4	66	2
		200	66	133	84.0	13.4	2.2	0.4	62	8	134	2
		500	166	333	44.5	6.8	42.4	6.3	161	5	340	7
$\Omega_1$	$\Omega_2$	100	33	66	7.1	73.3	19.2	0.4	31	5	66	1
		200	66	133	0.0	46.3	53.5	0.2	66	5	133	1
		500	166	333	0.0	0.2	91.4	6.6	166	4	333	1
$\Omega_2$	$\Omega_1$	100	33	66	15.4	82.6	2.0	0.0	28	18	67	1
		200	66	133	0.4	81.6	16.8	1.2	60	6	134	1
		500	166	333	0.0	4.2	89.0	6.8	162	5	334	1
<i>Cusum procedure</i>												
$\Omega_1$	$I$	100	33	66	75.4	9.2	15.0	0.4	36	3	65	2
		200	66	133	39.3	6.8	52.6	1.3	68	3	131	3
		500	166	333	0.6	0.0	94.0	5.4	168	4	331	3
$\Omega_2$	$I$	100	33	66	72.0	17.8	9.6	0.6	31	2	69	3
		200	66	133	46.7	8.4	41.3	3.6	65	2	135	2
		500	166	333	1.6	0.0	91.3	7.1	164	3	335	3
$\Omega_1$	$\Omega_2$	100	33	66	3.8	33.9	62.1	0.2	34	3	65	1
		200	66	133	0.2	6.6	90.4	2.8	68	4	132	1
		500	166	333	0.0	0.0	91.9	8.1	168	4	332	1
$\Omega_2$	$\Omega_1$	100	33	66	53.5	18.8	26.1	1.6	32	2	66	1
		200	66	133	0.4	18.3	74.7	6.6	65	3	134	1
		500	166	333	0.0	0.0	93.1	6.9	164	3	334	1

Med is the median and Mad, the median of the absolute deviations from the Med.

type I error of the procedures, which is around the 5% in all the sample sizes considered. When  $n = 100$ , the cusum procedure appears to work better than the LRT procedure. For  $n = 200$  the cusum procedure is slightly better than the LRT one, but when  $n = 500$ , the detection frequency of one changepoint is larger than 90% and there is a small increase in the detection of two or more changes by the cusum procedure. The time of the changes is slightly better estimated with the LRT procedure.

Finally, we study the power of these statistics when there is also a change in the parameter matrices, which will be called a structural change. Let  $y_t$  be a series generated by the following model:

$$\begin{cases} \Phi_1(B)y_t = c_1 + \Theta_1(B)a_t, & t < h, \\ \Phi_2(B)y_t = c_2 + \Theta_2(B)(a_t + WS_t^{(h)}a_t), & t \geq h, \end{cases}$$

such that the covariance matrix as well as the polynomials of the model changes at time  $t = h$ . The polynomials  $\Phi_1(B)$ ,  $\Phi_2(B)$ ,  $\Theta_1(B)$  and  $\Theta_2(B)$  are assumed to verify the conditions for stationarity and invertibility in Section 2. If the procedures had good power properties for detecting a covariance change under a structural change, they could be a useful tool for detecting both covariance changes and structural changes. We consider the case of  $k = 2$  and sample sizes  $n = 100, 200$  and  $500$ . The changes are introduced by transforming the original covariance matrix,  $\Sigma = I$ , into  $\Omega_1$  and  $\Omega_2$ , and the autoregressive polynomial  $\Pi_1$  into  $\Pi_2$ , where

$$\Pi_1 = \begin{pmatrix} 0.6 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}, \quad \Pi_2 = \begin{pmatrix} 0.3 & 0.4 \\ 0.4 & 0.7 \end{pmatrix}.$$

Table 7  
Results for model 2 and two covariance changes

$\Omega$	$\Omega$	$n$	$h_1$	$h_2$	Frequency				$\hat{h}_1$		$\hat{h}_2$	
					0	1	2	$\geq 3$	Med	Mad	Med	Mad
<i>LRT procedure</i>												
$\Omega_3$	$I$	100	33	66	79.6	15.6	4.8	0.0	27	7	67	6
		200	66	133	47.7	19.6	32.3	0.4	66	4	133	4
		500	166	333	2.0	0.4	96.1	1.5	166	3	333	3
$\Omega_4$	$I$	100	33	66	88.8	11.0	0.2	0.0	29	1	82	1
		200	66	133	74.1	19.8	6.0	0.0	63	3	136	3
		500	166	333	18.2	0.2	59.6	21.9	163	3	336	3
$\Omega_3$	$\Omega_4$	100	33	66	0.6	75.9	23.2	0.3	30	4	66	1
		200	66	133	0.0	32.9	66.5	0.6	65	4	133	1
		500	166	333	1.8	0.2	95.9	2.1	166	3	333	3
$\Omega_4$	$\Omega_3$	100	33	66	4.6	94.0	1.4	0.0	28	9	70	2
		200	66	133	0.0	70.5	28.4	1.1	62	4	134	1
		500	166	333	0.0	0.0	77.2	22.8	163	3	333	0
<i>Cusum procedure</i>												
$\Omega_3$	$I$	100	33	66	68.5	12.6	18.8	0.0	34	2	65	2
		200	66	133	19.4	1.8	75.7	3.1	67	3	132	3
		500	166	333	0.0	0.0	91.4	8.6	167	3	331	3
$\Omega_4$	$I$	100	33	66	67.7	14.4	17.4	0.4	32	2	68	2
		200	66	133	26.7	1.0	68.5	3.8	65	2	134	2
		500	166	333	0.0	0.0	89.7	10.3	165	2	335	2
$\Omega_3$	$\Omega_4$	100	33	66	1.4	28.3	69.9	0.4	34	2	65	1
		200	66	133	0.2	1.4	95.1	3.3	67	3	133	1
		500	166	333	0.0	0.0	94.6	5.4	167	3	331	2
$\Omega_4$	$\Omega_3$	100	33	66	1.0	55.5	42.1	1.4	33	2	66	1
		200	66	133	0.2	3.6	87.4	8.8	65	2	134	1
		500	166	333	0.2	0.0	84.3	15.5	165	2	333	1

Med is the median and Mad, the median of the absolute deviations from the Med.

For each case, we generate 5000 realizations. Then, we apply the two procedures with the 5% critical values from Table 2. The results are shown in Table 9. We conclude that both procedures have a small decrease in power for small samples sizes, specially when  $h = [0.25n]$ , but they do not lose power for big sample sizes, here  $n = 500$ , except the LRT procedure when the change is  $h = [0.75n]$ . Note that we do not take into account the existence of the break in the parameter matrix. Therefore, the change in the parameter matrix affects the estimation of the covariance matrices and, thus, the value of the statistics for covariance changes. Comparing the results in Tables 4 and 9 we conclude that the presence of the parameter change decreases the power of both statistics for covariance change detection, but specially to the LRT one.

## 6. Illustrative examples

### 6.1. Example: flour data

We consider the trivariate series of the logarithms of monthly flour price indices from three US cities over the period August 1972 through November 1980. This vector series was analyzed by Tiao and Tsay (1989), Grubb (1992) and Lütkepohl and Poskitt (1996) and is shown in Fig. 2. Tiao and Tsay (1989) fitted a restricted vector ARMA(1,1) to the series, whereas Grubb (1992), by using the Akaike Information Criteria (AIC), chose a restricted VAR(2) model.

Table 8  
Results for model 1 with variances changes

$\Omega$	$n$	$h$	LRT procedure					Cusum procedure				
			Frequency			$\hat{h}$		Frequency			$\hat{h}$	
			0	1	$\geq 2$	Med	Mad	0	1	$\geq 2$	Med	Mad
$I$	100	–	94.2	5.8	0.0	–	–	96.8	3.2	0.0	–	–
$I$	200	–	95.6	4.2	0.2	–	–	95.6	4.2	0.2	–	–
$I$	500	–	96.4	3.6	0.0	–	–	96.0	3.8	0.2	–	–
$\Omega_5$	100	25	63.5	35.8	0.7	26	5	58.9	40.5	0.6	36	7
		50	42.9	56.7	0.4	50	5	14.2	85.0	0.8	52	3
		75	49.9	49.9	0.2	75	4	27.5	72.2	0.3	74	3
$\Omega_5$	200	50	22.2	76.2	1.6	51	5	11.2	85.6	3.2	60	9
		100	6.4	90.3	3.3	100	4	0.6	95.9	3.5	101	3
		150	16.8	80.2	3.0	150	5	3.2	95.2	1.6	150	4
$\Omega_5$	500	125	0.0	97.8	2.2	126	4	0.0	95.2	4.8	134	9
		250	0.0	96.2	3.8	250	3	0.0	93.6	6.4	251	3
		375	0.0	95.4	4.6	376	4	0.0	93.5	6.5	375	4
$\Omega_6$	100	25	66.7	33.1	0.2	25	3	54.1	45.7	0.2	26	3
		50	49.7	49.7	0.6	50	4	27.9	69.8	2.3	48	3
		75	64.9	33.7	1.4	74	6	66.5	31.9	1.6	65	8
$\Omega_6$	200	50	23.4	74.1	2.4	49	4	12.0	84.2	3.8	50	4
		100	11.8	84.3	3.9	99	4	1.2	96.0	2.8	98	3
		150	24.6	72.3	3.0	149	5	15.4	81.0	3.6	140	9
$\Omega_6$	500	125	0.2	97.0	2.8	124	4	0.0	91.8	8.2	125	4
		250	0.0	97.5	2.5	251	3	0.0	96.7	3.3	250	3
		375	0.2	97.0	2.8	374	4	0.2	94.3	5.5	365	10

Med is the median and Mad, the median of the absolute deviations from the Med.

Lütkepohl and Poskitt (1996) investigated cointegration in these series using Johansen’s test (see Johansen, 1991) in a VAR(2) model, rejecting the null hypothesis of cointegration. Then, they fitted a VAR(1) for the differenced series, which is showed in the second row in Table 10.

Using this model, we first apply the LRT and cusum procedures for variance changes. Table 11 summarizes the results. No variance changes are detected by both procedures (see the values for  $W$  diagonal). Then, we apply the LRT and cusum procedures for covariances changes, and both procedures detect one change point at  $t = 33$  (April, 1975). The estimation of  $W$  is done as in (8). We include in the table the values of the Akaike and Bayesian information criteria (AIC and BIC) for each model, given by  $-(2/n) \log(\text{maximized likelihood}) + (c/n)(\text{number of parameters})$ , where  $c = 2$  for AIC and  $c = \log(n)$  for BIC. Note that the values of both criteria are reduced when the covariance change is introduced. Both criteria indicate that the model with one covariance change at  $t = 33$  (April, 1975) appears to be the most appropriate for the data. The final estimated model is shown in the third row in Table 10.

Finally, we estimate a VAR(1) model to the subsamples 1–32 and 33–100. The two estimated models with their standard errors are given in the fourth and fifth rows in Table 10. As we can see, the parameters of the model and the sample residual covariance matrices are quite different in both models, so that we conclude that the series may have a structural change at  $t = 33$ . The model for the second part of the series is apparently a random walk and the sample residual covariance matrix has smaller values than those of the first part of the series.

### 6.2. Example: wheat data

We consider the series of the logarithms of the monthly wheat price indices from five provinces in Castilla, Spain, over the period July 1880 through December 1890. This vector series was analyzed in Peña and Box (1987) and is shown in Fig. 3. We investigate cointegration in these series using Johansen’s test and by using the BIC we chose a

Table 9  
Results for structural changes

$\Pi$	$\Omega$	$n$	$h$	LRT procedure					Cusum procedure				
				Frequency			$\hat{h}$		Frequency			$\hat{h}$	
				0	1	$\geq 2$	Med	Mad	0	1	$\geq 2$	Med	Mad
$\Pi_2$	$\Omega_1$	100	25	76.0	23.8	0.2	25	7	75.0	24.8	0.2	39	11
			50	58.5	40.8	0.7	50	6	27.5	72.2	0.3	52	4
			75	60.7	39.1	0.2	75	6	33.3	65.5	1.2	74	4
$\Pi_2$	$\Omega_1$	200	50	44.9	53.3	1.8	50	7	40.9	58.1	1.0	66	15
			100	15.8	81.2	3.0	101	7	2.6	94.2	3.2	102	6
			150	15.2	78.3	6.4	151	5	4.2	93.4	2.4	149	5
$\Pi_2$	$\Omega_1$	500	125	4.2	91.6	4.2	126	7	3.2	93.1	3.7	141	16
			250	0.0	91.0	9.0	251	5	0.2	95.6	4.2	253	5
			375	0.0	75.7	24.3	376	4	0.0	92.8	7.2	375	4
$\Pi_2$	$\Omega_2$	100	25	51.5	48.1	0.4	24	3	35.3	63.1	1.6	26	3
			50	41.9	56.7	1.4	50	4	23.2	73.6	3.1	49	3
			75	51.1	47.1	1.8	75	5	80.0	19.8	0.2	62	8
$\Pi_2$	$\Omega_2$	200	50	10.0	85.2	4.8	50	3	0.8	93.0	6.2	50	3
			100	5.0	89.3	5.7	100	4	0.2	92.9	6.9	99	3
			150	18.4	76.6	5.0	151	6	40.3	56.1	3.6	136	13
$\Pi_2$	$\Omega_2$	500	125	0.0	92.0	8.0	124	2	0.0	91.1	8.9	125	3
			250	0.0	87.9	12.1	250	4	0.0	91.9	8.1	248	4
			375	0.4	76.0	23.6	376	9	0.8	91.4	7.8	362	13

Med is the median and Mad, the median of the absolute deviations from the Med.

Table 10  
AR parameter matrix ( $\hat{\Pi}$ ) and estimated covariance matrix ( $\hat{\Sigma}$ ) for three models fitted to the flour series

Model	$\hat{\Pi}$	$10^2 \hat{\Sigma}$
VAR(1)	$\begin{vmatrix} -0.86 & 1.01 & 0 \\ (0.17) & (0.18) & \\ -0.43 & 0.62 & 0 \\ (0.17) & (0.19) & \\ 0 & 0.25 & 0 \\ & (0.10) & \end{vmatrix}$	$\begin{vmatrix} 0.20 & 0.21 & 0.20 \\ 0.21 & 0.24 & 0.22 \\ 0.20 & 0.22 & 0.27 \end{vmatrix}$
VAR(1) + covariance change at $h = 33$	$\begin{vmatrix} -0.83 & 0.96 & 0 \\ (0.15) & (0.16) & \\ -0.48 & 0.64 & 0 \\ (0.13) & (0.14) & \\ 0 & 0.21 & 0 \\ & (0.11) & \end{vmatrix}$	$\begin{vmatrix} 0.20 & 0.21 & 0.20 \\ 0.21 & 0.23 & 0.22 \\ 0.20 & 0.22 & 0.26 \end{vmatrix}$
First part (1–32)	$\begin{vmatrix} -0.61 & 0.95 & 0 \\ (0.14) & (0.20) & \\ 0 & 0.40 & 0 \\ & (0.16) & \\ 0 & 0.43 & 0 \\ & (0.18) & \end{vmatrix}$	$\begin{vmatrix} 0.24 & 0.26 & 0.27 \\ 0.26 & 0.30 & 0.31 \\ 0.27 & 0.31 & 0.37 \end{vmatrix}$
Second part (33–100)	$\begin{vmatrix} -0.28 & 0.26 & 0 \\ (0.11) & (0.11) & \\ 0 & 0 & 0 \\ & & \\ 0 & 0 & 0 \end{vmatrix}$	$\begin{vmatrix} 0.19 & 0.26 & 0.17 \\ 0.20 & 0.21 & 0.19 \\ 0.17 & 0.19 & 0.21 \end{vmatrix}$

Standard errors of the coefficients are under parenthesis.

vector error correction model given by  $\nabla y_t = -\Pi y_{t-1} + e_t$ , with three cointegration relationships. This is in agreement with the two factors found by Peña and Box (1987).

First, we apply the LRT and cusum procedures for variance changes assuming the vector error correction model with three cointegration relationships shown in the second row in Table 12. Table 13 summarizes the results: the LRT

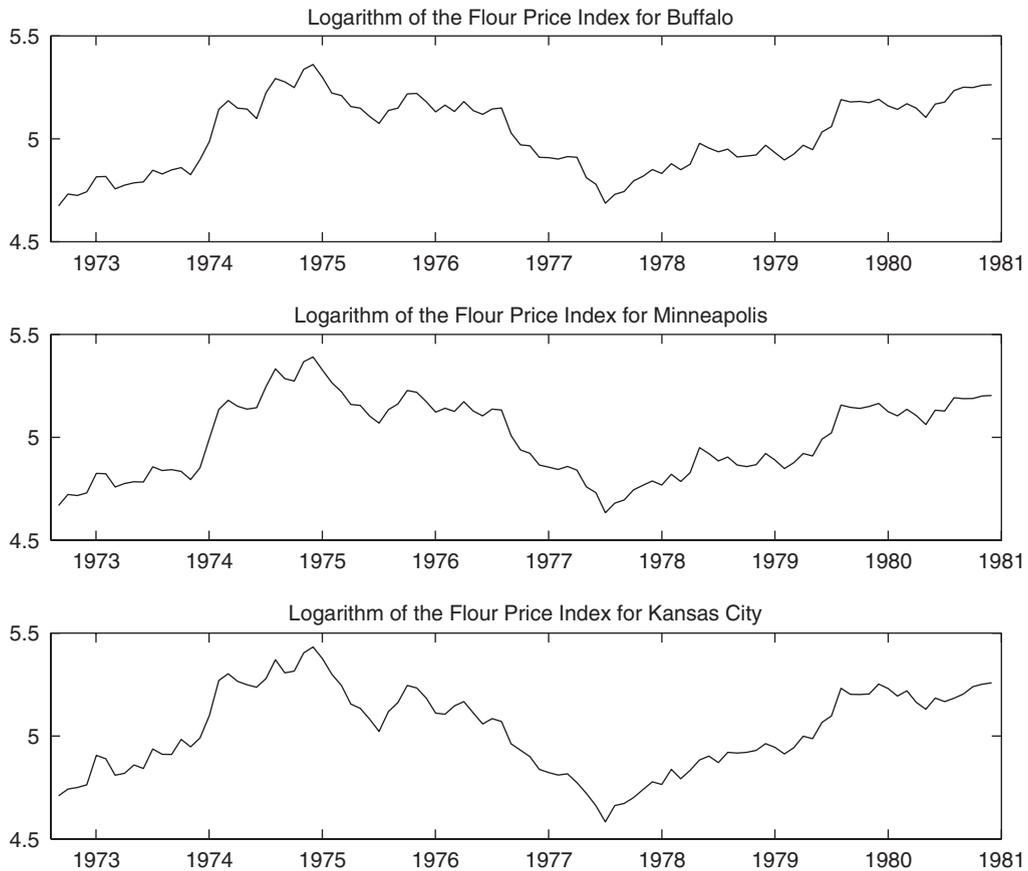


Fig. 2. Monthly flour price indices for three US cities.

Table 11  
Summary of the LRT and cusum procedures for the flour data

Method	VAR(1)	LRT $W_{diag}$	Cusum $W_{diag}$	LRT $W_{triang.}$	Cusum $W_{triang.}$
$h$	–	28	60	33	33
$A_{max}/\Gamma_{max}$	–	13.26	0.63	28.95	1.78
$\hat{W}$	–	–	–	$\begin{vmatrix} -0.14 & 0 & 0 \\ 0.40 & -0.53 & 0 \\ -0.02 & -0.17 & -0.09 \end{vmatrix}$	$\begin{vmatrix} -0.14 & 0 & 0 \\ 0.40 & -0.53 & 0 \\ -0.02 & -0.17 & -0.09 \end{vmatrix}$
AIC	–14.01	–14.01	–14.01	–27.39	–27.39
BIC	–13.77	–13.77	–13.77	–26.99	–26.99

procedure detects one change at  $h = 90$  (December, 1888), and the cusum procedure detects one change at  $h = 41$  (November, 1883). The estimation of the changes and their confidence intervals appear in Table 13. If one of the changes in one component is not significant, we represented it by 0. The minimum values of both the AIC and BIC correspond to the model proposed by the cusum with one change.

Then, we apply the LRT and cusum procedures for covariances changes, see Table 14. The LRT procedure detects now one change at  $h = 90$  (December, 1888), while the cusum procedure detects two changes: at  $h = 35$  (May, 1883)

Table 12  
AR parameter matrix ( $\hat{\Pi}$ ) and estimated covariance matrix ( $\hat{\Sigma}$ ) for two models fitted to the wheat series

Model	$\hat{\Pi}$	$10^3 \hat{\Sigma}$
ECM	$\begin{vmatrix} 0.04 & 0.13 & 0.02 & -0.37 & 0.17 \\ -0.15 & 0.67 & -0.05 & -0.41 & -0.03 \\ 0.08 & 0.31 & 0.56 & -0.94 & -0.00 \\ -0.10 & 0.23 & -0.17 & 0.07 & -0.03 \\ 0.13 & 0.04 & -0.10 & -0.56 & 0.48 \end{vmatrix}$	$\begin{vmatrix} 1.40 & 0.45 & 0.53 & 0.64 & 0.54 \\ 0.45 & 0.81 & 0.45 & 0.30 & 0.38 \\ 0.53 & 0.45 & 2.37 & 0.60 & 0.66 \\ 0.64 & 0.30 & 0.60 & 0.90 & 0.52 \\ 0.54 & 0.38 & 0.66 & 0.52 & 1.97 \end{vmatrix}$
ECM + 2 cov. changes	$\begin{vmatrix} 0.10 & 0.05 & -0.01 & -0.31 & 0.16 \\ 0.02 & 0.37 & -0.13 & -0.25 & -0.01 \\ 0.23 & -0.00 & 0.40 & -0.66 & 0.03 \\ -0.11 & 0.11 & -0.17 & 0.26 & -0.09 \\ 0.19 & -0.16 & -0.12 & -0.41 & 0.50 \end{vmatrix}$	$\begin{vmatrix} 2.21 & 0.88 & 1.05 & 1.25 & 0.99 \\ 0.88 & 1.43 & 1.18 & 0.54 & 0.83 \\ 1.05 & 1.18 & 4.62 & 1.10 & 1.26 \\ 1.25 & 0.54 & 1.10 & 1.66 & 0.94 \\ 0.99 & 0.83 & 1.26 & 0.94 & 2.47 \end{vmatrix}$

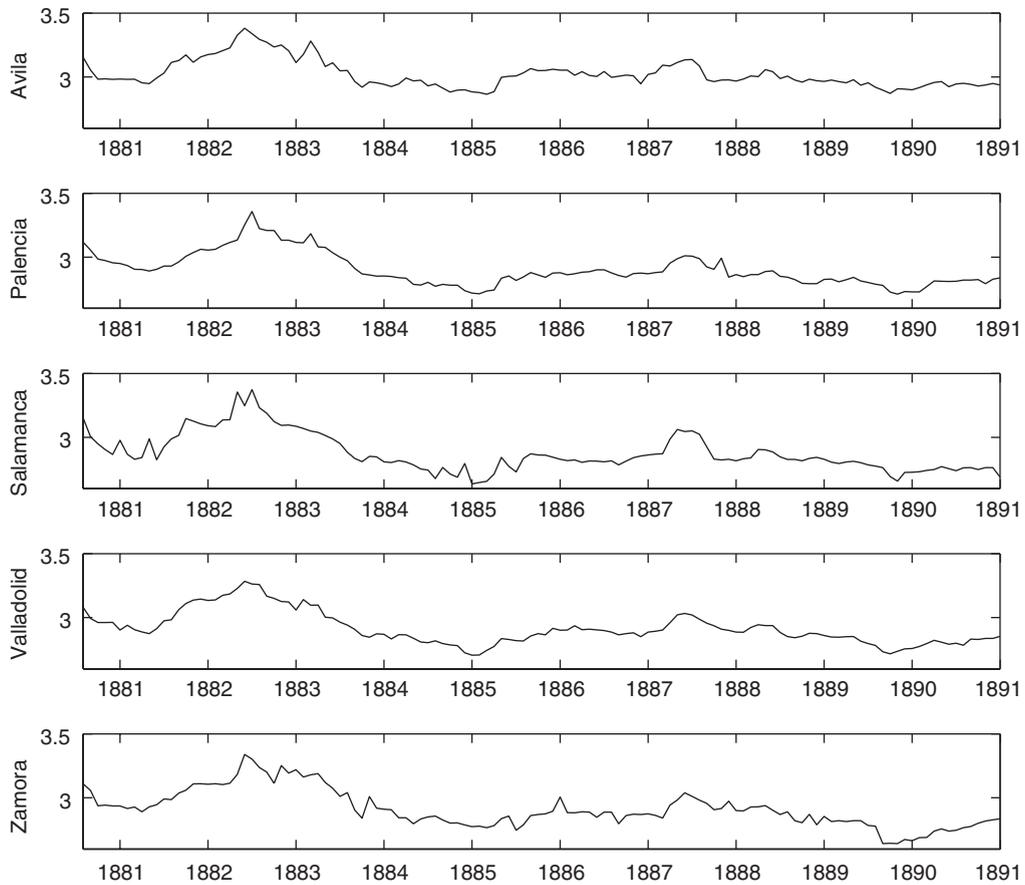


Fig. 3. Monthly wheat price indices for five provinces in Castilla, Spain.

and  $h = 90$  (December, 1888). The estimation of the changes appears in Table 14. The minimum of the values of both the AIC and BIC corresponds to the model proposed by the cusum with two changes.

The final model is selected by the BIC (although AIC gives the same result) and is the one obtained by the cusum procedure allowing covariance changes. The estimated parameters for this ECM are shown in the third row in Table 12.

Table 13  
Summary of the LRT and cusum procedures for variance changes for the wheat data

Method	VAR(1)	LRT	Cusum
$h$	–	90	41
$A_{\max}/\Gamma_{\max}$	–	72.58	2.25
$\widehat{W}(1, 1)$ <i>I.C.</i>	–	–0.47 (–0.56, –0.36)	–0.37 (–0.52, –0.50)
$\widehat{W}(2, 2)$ <i>I.C.</i>	–	–0.31 (–0.43, –0.17)	0
$\widehat{W}(3, 3)$ <i>I.C.</i>	–	–0.56 (–0.64, –0.47)	–0.30 (–0.47, –0.11)
$\widehat{W}(4, 4)$ <i>I.C.</i>	–	–0.28 (–0.41, –0.13)	–0.35 (–0.51, –0.18)
$\widehat{W}(5, 5)$ <i>I.C.</i>	–	–0.35 (–0.46, –0.21)	–0.29 (–0.46, –0.09)
AIC	–19.53	–28.05	–29.31
BIC	–18.97	–27.41	–28.64

Table 14  
Summary of the LRT and cusum procedures for covariance changes for the wheat data

Method	$h$	$A_{\max}/\Gamma_{\max}$	$\widehat{W}$	AIC	BIC
ECM	–	–	–	–19.53	–18.97
LRT	90	74.35	$\begin{pmatrix} -0.47 & 0 & 0 & 0 & 0 \\ -0.02 & -0.32 & 0 & 0 & 0 \\ -0.09 & -0.11 & -0.51 & 0 & 0 \\ -0.16 & 0.22 & -0.04 & -0.22 & 0 \\ -0.25 & 0.15 & -0.18 & 0.52 & -0.40 \end{pmatrix}$	–30.51	–29.04
	35	1.81	$\begin{pmatrix} -0.30 & 0 & 0 & 0 & 0 \\ -0.09 & -0.16 & 0 & 0 & 0 \\ 0.00 & 0.00 & -0.37 & 0 & 0 \\ -0.15 & 0.20 & 0.01 & -0.38 & 0 \\ -0.07 & -0.08 & 0.02 & -0.05 & -0.05 \end{pmatrix}$		
Cusum	90	2.03	$\begin{pmatrix} -0.37 & 0 & 0 & 0 & 0 \\ 0.00 & -0.37 & 0 & 0 & 0 \\ -0.01 & -0.27 & -0.35 & 0 & 0 \\ -0.04 & 0.06 & -0.10 & -0.02 & 0 \\ -0.03 & -0.09 & -0.09 & 0.48 & -0.24 \end{pmatrix}$	–30.97	–29.73

**Acknowledgement**

We acknowledge the financial support by MEC project SEJ2004-03303 and by Fundación BBVA, Spain. We also acknowledge an anonymous referee for very helpful comments.

**Appendix**

**Proof of Theorem 1.** We define the process  $\zeta_h = e'_h \Sigma^{-1} e_h - k$ , with  $E[\zeta_h] = 0$  and finite variance,

$$\sigma_{\zeta}^2 = E[\zeta_h^2] = E[(e'_h \Sigma^{-1} e_h - k)^2] = 2k.$$

The process  $\zeta_h$  satisfies the conditions of Theorem 3.4 in Phillips and Solo (1992). Therefore, for  $h = nv$ ,

$$\frac{1}{\sigma_{\zeta} \sqrt{n}} \sum_{t=1}^h \zeta_t \rightarrow_d M_v, \tag{14}$$

and

$$\frac{1}{\sigma_{\xi}\sqrt{n}} \sum_{t=1}^h \xi_t - \frac{h}{n} \frac{1}{\sigma_{\xi}\sqrt{n}} \sum_{t=1}^n \xi_t = \frac{h}{\sqrt{2kn}} \left( \frac{A_h}{h} - \frac{A_n}{n} \right) \rightarrow_d M_v^0,$$

which proves the stated result.  $\square$

**Proof of Theorem 2.** We define the process  $\xi_h = c'_h c_h - k$ , with  $E[\xi_h] = 0$  and finite variance,

$$\sigma_{\xi}^2 = E[\xi_h^2] = E[(c'_h c_h - k)^2] = 2tr(A^2).$$

The process  $\xi_h$  satisfies (14) for  $h = nv$  and consequently,

$$\frac{1}{\sigma_{\xi}\sqrt{n}} \sum_{t=1}^h \xi_t - \frac{h}{n} \frac{1}{\sigma_{\xi}\sqrt{n}} \sum_{t=1}^n \xi_t = \frac{h}{\sqrt{2tr(A^2)n}} \left( \frac{A_h}{h} - \frac{A_n}{n} \right) \rightarrow_d M_v^0,$$

which proves the stated result.  $\square$

## References

- Abraham, B., Wei, W.W.S., 1984. Inferences about the parameters of a time series model with changing variance. *Metrika* 31, 183–194.
- Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.
- Bai, J., 1993. On the partial sums of residuals in autoregressive and moving average models. *J. Time Ser. Anal.* 14, 247–260.
- Bai, J., 1994a. Least squares estimation of a shift in linear processes. *J. Time Ser. Anal.* 15, 453–472.
- Bai, J., 1994b. Weak-convergence of the sequential empirical processes of residuals in ARMA models. *Ann. Statist.* 22, 2051–2061.
- Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47–78.
- Bai, J., Lumsdaine, R.L., Stock, J.H., 1998. Testing for and dating common breaks in multivariate time series. *Rev. Econom. Stud.* 65, 395–432.
- Baufays, P., Rasson, J.P., 1985. Variance changes in autoregressive models. In: *Time Series. Theory and Methods*, second ed. Springer, New York, pp. 119–127.
- Billingsley, P., 1968. *Convergence of Probability Measures*. Wiley, New York.
- Chen, J., Gupta, A.K., 1997. Testing and locating variance changepoints with application to stock prices. *J. Amer. Statist. Assoc.* 92, 739–747.
- Gombay, E., Horvath, L., 1990. Asymptotic distributions of maximum likelihood tests for change in mean. *Biometrika* 77, 411–414.
- Grubb, H., 1992. A multivariate time series analysis of some flour price data. *Appl. Statist.* 26, 279–284.
- Inclán, C., 1993. Detection of multiple tests of variances using posterior odds. *J. Business Econom. Statist.* 11, 189–200.
- Inclán, C., Tiao, G.C., 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Amer. Statist. Assoc.* 89, 913–923.
- Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
- Lee, S., Park, S., 2001. The cusum of squares test for scale changes in infinite order moving average models. *Scand. J. Statist.* 28, 625–644.
- Ling, S.Q., 1998. Weak convergence of the sequential empirical processes of residuals in nonstationary autoregressive models. *Ann. Statist.* 26, 741–754.
- Lütkepohl, H., Poskitt, D.S., 1996. Specification of echelon-form VARMA models. *J. Business Econom. Statist.* 14, 69–79.
- McCulloch, R.E., Tsay, R.S., 1993. Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *J. Amer. Statist. Assoc.* 88, 968–978.
- Park, S., Lee, S., Jeon, J., 2000. The cusum of squares test for variance changes in infinite order autoregressive models. *J. Korean Statist. Soc.* 29, 351–361.
- Peña, D., Box, G.E.P., 1987. Identifying a simplifying structure in time series. *J. Amer. Statist. Assoc.* 82, 836–843.
- Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. *Ann. Statist.* 20, 971–1001.
- Tiao, G.C., Tsay, R.S., 1989. Model specification in multivariate time series (with discussion). *J. Roy. Statist. Soc. Ser. B* 51, 157–213.
- Tsay, R.S., 1988. Outliers, level shifts and variance changes in time series. *J. Forecasting* 7, 1–20.
- Tsay, R.S., Peña, D., Pankratz, A.E., 2000. Outliers in multivariate time series. *Biometrika* 87, 789–804.
- Wang, J., Zivot, E., 2001. A Bayesian time series model of multiple structural changes in level, trend, and variance. *J. Business Econom. Statist.* 18, 374–386.
- Wichern, D.W., Miller, R.B., Hsu, D.A., 1976. Changes of variance in first order autoregressive time series models—with an application. *Appl. Statist.* 25, 248–256.