

## SOME GOODNESS OF FIT TESTS FOR RANDOM SEQUENCES

Yuriy Kozachenko<sup>1</sup>, Tetiana Ianevych<sup>2</sup>

Taras Shevchenko National University of Kyiv, Mechanics and Mathematics Faculty  
Address: Volodymyrska str., 64/13, Kyiv, 01601, Ukraine  
E-mail: <sup>1</sup>ykoz@ukr.net, <sup>2</sup>yata452@univ.kiev.ua

Received: August 2013      Revised: September 2013      Published: November 2013

**Abstract.** In this paper we had made an attempt to incorporate the results from the theory of square Gaussian random variables in order to construct the goodness of fits test for random sequences (time series). We considered two versions of such tests. The first one was designed for testing the adequacy of the hypotheses on expectation and covariance function of univariate non-centered sequence, the other one was constructed for testing the hypotheses on covariance of the multivariate centered sequence. The simulation results illustrate the behavior of these tests in some particular cases.

**Keywords:** goodness of fit test, multivariate random sequence, time series, square Gaussian random variable.

### 1. Introduction

The task of investigating the properties of a random sequence is very important for its application. Very often some phenomena can be observed only at certain points in time. In some cases the values of continuous quantities, such as temperature and voltage, can be written only at discrete moments of time. And even if the observations can be recorded continuously we can only use discrete data for computational purposes. That is why we usually deal with random sequences or time series in practice. The latter term is used more frequently but we prefer to use the former one designating the connection to random processes.

There is much literature devoted to this topic, in particular the classic books on the statistical analysis of time series written by Anderson [1], Box and Jenkins [2], and Brockwell and Davis [4].

To date, many goodness-of-fit tests in time series are residual-based. For example, the classic portmanteau test of Box and Pierce [3] and its improvement by Ljung and Box [15] are based on the sample autocorrelations of the residuals. In the context of goodness of fit of nonlinear time series models, the McLeod and Li [16] test is based on the sample autocorrelations of the squared residuals. Based on a spectral approach of the residuals, Chen and Deo [6] proposed some new diagnostic tests. More recently, perhaps influenced by the empirical distribution function approach in the goodness-of-fit test for independent observations, substantial developments for time series data have taken place in the form of tests based on empirical processes marked by certain residuals, see for instance Chen and Härdle [7] and Escanciano [8]. For more information and details see the references within the literature proposed.

In the model-based approach to time series analysis, estimated residuals are computed once a fitted model has been obtained from the data, and these are then tested for “whiteness”, i. e. it is determined whether they behave like white noise. Tests for residual whiteness generally postulate whiteness of the residuals as the Null Hypothesis, so that significant rejections indicate model inadequacy. These tests require the computation of residuals from the fitted model, which can be quite tedious when the model does not have a finite order autoregressive representation. Also, in such cases, the residuals are not uniquely defined.

In this paper we use another approach gained from the theory of square Gaussian random variables. This theory was developed in works by Kozachenko et al. [11], [12], [13], for the investigation of stochastic processes. In the book by Buldygin and Kozachenko [5] the properties of the space of square Gaussian random variables were studied and the connection with Orlicz spaces of random variables was established. We use this theory for the construction goodness of fit tests on the expectation and covariance function for the *non*-centered univariate stationary Gaussian sequence and covariance function for the centered but multivariate stationary random sequence. Our tests do not require the computation

of residuals and can be applied to infinite order representations. This paper is the continuation of the work started in [14], which was devoted to testing the centered univariate random sequence.

The paper consists of 5 sections and 2 annexes. The second section is devoted to the theory of square Gaussian random variables and contains the main definitions and results. In particular, we found the estimate for the distribution of the maximum of the quadratic form of the square Gaussian random variables. Sections 3 and 4 apply the estimator obtained in section 2 to construct different aggregate tests.

The test in section 3 was constructed for testing the aggregated hypothesis on the expectation and covariance function of the *non*-centered stationary Gaussian sequence. It is based on the approach used in stochastic process  $L^2$  theory. Within this inference the process identification is made on the basis of the two main characteristics: mathematical expectation and covariance function.

In the section 4 we consider multivariate sequences. The approach analyzing the residuals dominates in the multivariate case too. See, for example, papers by Hosking [9], [10], Mahdi and McLeod [17] and the references therein. The goodness of fit test we have constructed for the centered Gaussian multivariate stationary sequence is based on fitting the covariance function.

The power properties of our tests were studied through simulations. Section 5 draws some conclusions. Some necessary mathematical calculations are relegated to the annexes at the end.

## 2. Square Gaussian random variables

Let  $\Xi = \{\gamma_i, i \in I\}$  be a family of joint Gaussian random variables for which  $E\gamma_i = 0$  for all  $i \in I$ .

**Definition 1.** [13] The space  $SG_{\Xi}(\Omega)$  is the space of square Gaussian random variables if any element  $\xi \in SG_{\Xi}(\Omega)$  can be presented as

$$\xi = \vec{\gamma}^T \mathbf{A} \vec{\gamma} - E \vec{\gamma}^T \mathbf{A} \vec{\gamma}, \quad (1)$$

where  $\vec{\gamma}^T = (\gamma_1, \dots, \gamma_r)$ ,  $\gamma_i \in \Xi$ ,  $i = \overline{1, r}$ ,  $\mathbf{A}$  is a real-valued matrix; or the element  $\xi \in SG_{\Xi}(\Omega)$  is the square mean limit of the sequence  $\{\xi_n, n \geq 1\}$  of the form (1)

$$\xi = \text{l.i.m.}_{n \rightarrow \infty} \xi_n.$$

It was proved by Buldygin & Kozachenko in [5] that  $SG_{\Xi}(\Omega)$  is a linear space.

For the square Gaussian random variables the following results hold true.

**Theorem 1.** [11] Let  $\vec{\xi}^T = (\xi_1, \xi_2, \dots, \xi_d)$  be a random vector such that  $\xi_i \in SG_{\Xi}(\Omega)$  and  $\mathbf{B}$  be a symmetric semi-definite matrix. Then for all  $0 < s < \frac{1}{\sqrt{2}}$  the following inequality is true

$$E \cosh \left( \sqrt{s^2 \frac{\vec{\xi}^T \mathbf{B} \vec{\xi}}{E \vec{\xi}^T \mathbf{B} \vec{\xi}}} \right) \leq R(\sqrt{2}s), \quad (2)$$

where  $R(y) = \frac{1}{\sqrt{1-y}} \exp\{-\frac{y}{2}\}$ ,  $0 < y < 1$ .

**Theorem 2.** Let  $\{\eta_m, 1 \leq m \leq M\}$  be a sequence of random variables that can be presented as quadratic form of square Gaussian random variables (that is,  $\eta = \vec{\xi}^T \mathbf{B} \vec{\xi}$ , where  $\mathbf{B}$  is a symmetric semi-definite matrix). Then, for any  $x \geq 0$

$$P \left\{ \max_{1 \leq m \leq M} \frac{\eta_m}{E \eta_m} > x \right\} \leq M \cdot W(x), \quad (3)$$

where  $W(x) = \frac{R\left(\frac{\sqrt{2x}}{1+\sqrt{2x}}\right)}{\cosh\left(\frac{x}{1+\sqrt{2x}}\right)}$  and the function  $R$  is defined in Theorem 1.

*Proof.* For all  $0 < s < 1/\sqrt{2}$  and  $x > 0$

$$\begin{aligned} P \left\{ \max_{1 \leq m \leq M} \frac{\eta_m}{E \eta_m} > x \right\} &\leq M \max_{1 \leq m \leq M} P \left\{ \frac{\eta_m}{E \eta_m} > x \right\} \leq \\ &\leq M \max_{1 \leq m \leq M} \frac{E \cosh \sqrt{s^2 \frac{\eta_m}{E \eta_m}}}{\cosh \sqrt{s^2 x}} \leq M \frac{R(s\sqrt{2})}{\cosh(s\sqrt{x})}. \end{aligned}$$

Putting  $s = \frac{\sqrt{x}}{1+\sqrt{2x}}$ , which is approximately the minimal point, we obtain (3) and

$$\begin{aligned} \mathbb{P} \left\{ \max_{1 \leq m \leq M} \frac{\eta_m}{\mathbb{E} \eta_m} > x \right\} &\leq M \frac{(1 - \sqrt{2s})^{-1/2} \exp\{-s/\sqrt{2}\}}{\cosh(s\sqrt{x})} \leq \\ &\leq M e^{-1/2} \frac{\exp\left\{\frac{1}{2(1+\sqrt{2x})}\right\} \sqrt{1 + \sqrt{2x}}}{\cosh\left(\frac{x}{1+\sqrt{2x}}\right)} = \\ &= O(x^{1/4} e^{-\sqrt{x}}) \quad \text{as } x \rightarrow \infty. \end{aligned}$$

The theorem is proved.  $\square$

### 3. Testing a hypotheses on the expectation and covariance function of a univariate sequence

Using the inequality (3) it is possible to test a hypothesis on the expectation and covariance function of the non-centered univariate stationary Gaussian sequence. Hereinafter we will consider stationarity in a strict sense.

Let us consider the stationary sequence  $\{\gamma(n), n \geq 1\}$  for which  $\mathbb{E} \gamma(n) = a$  and  $\mathbb{E}(\gamma(n) - a)(\gamma(n+m) - a) = B(m)$ ,  $m \geq 0$  is its covariance function. We assume that we have  $N + M$  consecutive observations of this random sequence. Let us choose the estimators in the following way:

for the expectation

$$\hat{a}_m = \frac{1}{N} \sum_{n=1}^N \gamma(n+m), \quad 0 \leq m \leq M;$$

for the covariance function

$$\hat{B}(m) = \frac{1}{N} \sum_{n=1}^N (\gamma(n) - \hat{a}_0)(\gamma(n+m) - \hat{a}_m), \quad 0 \leq m \leq M.$$

We denote

$$\begin{aligned} E_a &:= \mathbb{E}(\hat{a}_m - a)^2 = \frac{1}{N^2} \sum_{n=1}^N \sum_{k=1}^N B(n-k) = \frac{1}{N} B(0) + \frac{2}{N^2} \sum_{i=1}^{N-1} (N-i) B(i); \\ E_B(m) &:= \mathbb{E} \hat{B}(m) = B(m) - \frac{1}{N^2} \sum_{n=1}^N \sum_{k=1}^N B(n-k-m) = \\ &= \left(1 - \frac{1}{N}\right) B(m) - \frac{2}{N^2} \sum_{i=1}^{N-1} (N-i) B(i-m), \quad 0 \leq m \leq M, \end{aligned}$$

and introduce the random variables:

$$\eta_a(m) := (\hat{a}_m - a)^2 - E_a, \quad 0 \leq m \leq M;$$

and

$$\eta_B(m) := \hat{B}(m) - E_B(m), \quad 0 \leq m \leq M.$$

It is easy to prove that these random variables are square Gaussian.

Let us define  $\vec{\eta}(m)^T = (\eta_a(m), \eta_B(m))$ ,  $0 \leq m \leq M$ . So, for any semi-definite matrix  $\mathbf{B}_m = (b_{ij}(m))_{i,j=1,2}$  the random variable  $\eta(m) := \vec{\eta}(m)^T \mathbf{B}_m \vec{\eta}(m)$  is actually the quadratic form of a square Gaussian random variable.

*Remark 1.* If  $b_{ij} = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases}$  (that is,  $\mathbf{B}$  is the identity matrix of order 2), then  $\eta(m) = \eta_a^2(m) + \eta_B^2(m)$  and  $\mathbb{E} \eta(m) = \mathbb{E} \eta_a^2(m) + \mathbb{E} \eta_B^2(m)$ . All the necessary calculations for the terms of  $\mathbb{E} \eta(m)$  are included to the Annex to this section.

*Remark 2.* If for every  $m$   $\mathbf{B}_m = \mathbf{C}^{-1}(m)$  is the inverse of matrix  $\mathbf{C}(m)$ , whose components are the covariances between the vector  $\vec{\eta}(m)$  items then  $\mathbb{E} \eta(m) = \text{const}$  for all  $m$ . But, in this case one should be careful since the matrices  $\mathbf{C}(m)$  have to be invertible.

**Criterion 1.** Let the null hypothesis  $H_0$  state that  $a$  and  $B(m)$ ,  $m \geq 0$  are the expectation and covariance function of the non-centered Gaussian stationary sequence  $\{\gamma(n), n \geq 1\}$  and the alternative  $H_a$  implies the opposite statement.

If for significance level  $\alpha$  and corresponding critical value  $\varepsilon_\alpha$  which can be found from the equation  $MW(\varepsilon_\alpha) = \alpha$

$$\max_{0 \leq m \leq M} \frac{\bar{\eta}^T(m) \mathbf{B}_m \bar{\eta}(m)}{E(\bar{\eta}^T(m) \mathbf{B}_m \bar{\eta}(m))} > \varepsilon_\alpha,$$

then the hypothesis  $H_0$  should be rejected and accepted otherwise.

*Remark 3.* The probability of a type I error for Criterion 1 is less than or equal to  $\alpha$ .

*Example 1.* Let us consider the *non*-centered Gaussian sequence  $\{\gamma(n), n \geq 1\}$  whose elements can be presented according to the expression

$$\gamma(n) = a + \sum_{j=0}^{\infty} \beta(j) \zeta_{n-j}, \quad n \geq 1 \quad (4)$$

where  $\beta(j) = e^{-\lambda j}$ ,  $j \geq 0$ ,  $\lambda > 0$  and  $\{\zeta_k, k \in \mathbb{Z}\}$  is a sequence of independent random variables such that for all  $k$   $E\zeta_k = 0$ ,  $E\zeta_k^2 = 1$ . In this case  $E\gamma(n) = a$  for all  $n$  and

$$B(m) = E(\gamma(n) - a)(\gamma(n+m) - a) = \sum_{j=0}^{\infty} \beta(j)\beta(j+m) = \sum_{j=0}^{\infty} e^{-\lambda j} e^{-\lambda(j+m)} = \frac{e^{-\lambda|m|}}{1 - e^{-2\lambda}}, m \in \mathbb{Z}. \quad (5)$$

Using the simulation study we investigated how the Criterion 1 works. We made 10 000 Monte Carlo simulations of the *non*-centered Gaussian stationary sequence  $\gamma(n)$ , with  $a = E\gamma(n)$  and covariance function defined by (5) with fixed parameter  $\lambda$ . For our needs we used the the simulation methods developed in paper [18].

For the symmetric semi-definite matrix  $\mathbf{B}_m$  we choose the identity matrix of order 2 –  $\mathbf{I}_2$ . Then  $\eta(m) = \bar{\eta}^T(m) \mathbf{B}_m \bar{\eta}(m) = \eta_a^2(m) + \eta_B^2(m)$ .

1. Let us check the null hypothesis ( $H_0$ ) that states that the stationary Gaussian sequence  $\gamma(n)$  has expectation  $a = 1$  and covariance function defined by the formula (5) with parameter  $\lambda = 1$  versus the alternative hypothesis ( $H_a$ ) implying that the stationary Gaussian sequence  $\gamma(n)$  has expectation  $a = 1$  and covariance function defined by the formula (5) with parameter  $\hat{\lambda} = 0.5$ .

We simulated 10 000 realizations of the sequences defined by (4) with parameters  $a = 1$  &  $\lambda = 1$  and  $a = 1$  &  $\hat{\lambda} = 0.5$ .

Let us define the necessary constants: the significance level  $\alpha = 0.1$ ,  $M = 10$ ,  $N = 1000$  ( $M + N = 1010$ ). In this case the critical value  $\varepsilon_\alpha = 87.82$ .

For the simulated sequences we obtained an estimate of the probability of a type I error  $\hat{\alpha} = 0$  and the estimate for the probability of a type II error  $\hat{\beta} = 0.23$ .

2. Let us now check the null hypothesis ( $H_0$ ) that states that the stationary Gaussian sequence  $\gamma(n)$  has expectation  $a = 1$  and covariance function defined by the formula (5) with parameter  $\lambda = 1$  versus the alternative hypothesis ( $H_a$ ) implying that the stationary Gaussian sequence  $\gamma(n)$  has expectation  $a = 0$  and covariance function defined by the formula (5) with parameter  $\lambda = 1$ .

We used again the simulated 10 000 realizations of the sequence defined by (4) with parameters  $a = 1$  &  $\lambda = 1$  and another sequence with parameters  $a = 0.25$  &  $\lambda = 1$ .

The required constants are the same as previously. In this case we obtained an estimate for the probability of a type I error  $\hat{\alpha} = 0$  and the estimate for the probability of a type II error  $\hat{\beta} = 0.0055$ .

*Remark 4.* It is evident that the more observations we have the more sensitive the criterion is. Finding the number  $N$  for which the null and alternative hypotheses can be distinguished is the subject of our continuing investigation.

#### 4. Testing a hypotheses on the covariance function of a centered multivariate sequence

The inequality (3) can also be useful for testing a hypothesis on the covariance function of a centered multivariate random sequence.

Let us assume that the components of the multivariate random sequence  $\vec{\gamma}(n)$ ,  $n \geq 1$  are jointly Gaussian, stationary (in the strict sense) sequences  $\{\gamma_k(n), n \geq 1, k = \overline{1, K}\}$  for which  $E\gamma_k(n) = 0$  and  $E\gamma_k(n)\gamma_l(n+m) = B_{kl}(m)$ ,  $m \geq 0$  is the covariance function of these sequences. It is worth mentioning that for  $k = l$   $B_{kk}$  is the ordinary autocovariance function of the  $k$ -th component and when  $k \neq l$   $B_{kl}$  are the joint covariances or sometimes called cross-covariances. Hereinafter we shall use the term *covariance function of the sequence*  $\vec{\gamma}(n)$ .

We suppose that the sequence  $\vec{\gamma}(n)$  is observed at points  $1, 2, \dots, N+M$  ( $N, M > 1$ ). As an estimator of the covariance function  $B_{kl}(m)$  we choose

$$\widehat{B}_{kl}(m) = \frac{1}{N} \sum_{n=1}^N \gamma_k(n)\gamma_l(n+m), \quad N \geq 1, \quad m = \overline{0, M}.$$

The estimator  $\widehat{B}_{kl}(m)$  is unbiased:

$$E\widehat{B}_{kl}(m) = \frac{1}{N} \sum_{n=1}^N E\gamma_k(n)\gamma_l(n+m) = B_{kl}(m).$$

The random variables  $\Delta_{kl}(m) = \widehat{B}_{kl}(m) - B_{kl}(m)$  are square Gaussian since  $\widehat{B}_{kl}(m)$  can be presented as  $(\gamma_k(1), \dots, \gamma_k(N))^T \mathbf{A} (\gamma_l(m+1), \dots, \gamma_l(N+m))$ , where the matrix

$$\mathbf{A} = \begin{pmatrix} \frac{1}{N} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{N} \end{pmatrix}.$$

Let  $\vec{\Delta}(m)$  be a vector with components  $\Delta_{kl}(m)$ .

**Criterion 2.** Let the null hypothesis  $H_0$  state that  $B_{kl}(m)$ ,  $m \geq 0$  is the covariance function of the centered Gaussian stationary sequence  $\vec{\gamma}(n) = \{\gamma_k(n), n \geq 1\}_{k=\overline{1, K}}$  and the alternative  $H_a$  state the opposite.

If for significance level  $\alpha$  and corresponding critical value  $\varepsilon_\alpha$ , which can be found from the equation  $MW(\varepsilon_\alpha) = \alpha$ ,

$$\max_{0 \leq m \leq M} \frac{\vec{\Delta}^T(m) \mathbf{B}_m \vec{\Delta}(m)}{E(\vec{\Delta}^T(m) \mathbf{B}_m \vec{\Delta}(m))} > \varepsilon_\alpha,$$

then the hypothesis  $H_0$  should be rejected and accepted otherwise.

*Remark 5.* The probability of a type I error for the Criterion 2 is less or equal to  $\alpha$ .

*Remark 6.* The simplest way is to choose the matrix  $\mathbf{B}_m$  to be identical.

If for every  $m$  the matrix  $\mathbf{B}_m = \mathbf{C}^{-1}(m)$  is inverse to  $\mathbf{C}(m)$  which consists of the covariances of the vector  $\vec{\Delta}(m)$  components, then  $E(\vec{\Delta}^T(m) \mathbf{B}_m \vec{\Delta}(m)) = const$  for all  $m$ . But in this case we should pay attention to the invertibility of the matrices  $\mathbf{C}(m)$ .

Let's illustrate how this criterion works on the example.

*Example 2.* We consider the  $K = 2$  – component stationary centered Gaussian sequence  $\vec{\gamma}(n) = \{\gamma_k(n), n \geq 1, k = 1, 2\}$ . We assume that each component can be presented as moving average

$$\gamma_k(n) = \sum_{j=0}^{\infty} \beta_k(j) \xi_{n-j}, \quad n \geq 1,$$

with coefficients  $\beta_k(j) = e^{-\lambda_k j}$ ,  $\lambda_k > 0$ ,  $k = 1, 2$ ,  $j \geq 0$  and the random variables  $\xi_j$  are independent with zero mean and variance equal to 1.

If the components of  $\vec{\gamma}(n)$  are not independent then the covariance function of this sequence has a form:

$$B_{kl}(m) = \begin{cases} \frac{e^{-\lambda_l m}}{1 - e^{-(\lambda_k + \lambda_l)}}, & m \geq 0; \\ \frac{e^{-\lambda_k |m|}}{1 - e^{-(\lambda_k + \lambda_l)}}, & m < 0, \end{cases} \quad k, l = 1, 2. \quad (6)$$

In the case when  $k = l$  we obtain the covariance function of the  $k$ -th component:

$$B_{kk}(m) = \frac{e^{-\lambda_k|m|}}{1 - e^{-2\lambda_k}}, \quad m \in \mathbb{Z}, \quad k = 1, 2. \quad (7)$$

Let  $\vec{\Delta}^T(m) = (\Delta_{11}(m), \Delta_{12}(m), \Delta_{21}(m), \Delta_{22}(m))$  and matrix  $\mathbf{B} = \mathbf{I}_4$  is the identity matrix of 4-th order. Then

$$E(\vec{\Delta}^T(m)\mathbf{B}\vec{\Delta}(m)) = \sum_{k=1}^2 \sum_{l=1}^2 E\Delta_{kl}^2(m), \quad (8)$$

where

$$\begin{aligned} E\Delta_{kl}^2(m) &= \frac{1}{N} \left[ \frac{1}{(1 - e^{-2\lambda_k})(1 - e^{-2\lambda_l})} + \frac{e^{-2\lambda_l m}}{(1 - e^{-(\lambda_k + \lambda_l)})^2} \right] + \\ &+ \frac{2}{N} \sum_{t=1}^m \left(1 - \frac{t}{N}\right) \left[ \frac{e^{-(\lambda_k + \lambda_l)t}}{(1 - e^{-2\lambda_k})(1 - e^{-2\lambda_l})} + \frac{e^{-2\lambda_l m}}{(1 - e^{-(\lambda_k + \lambda_l)})^2} \right] + \\ &+ \frac{2}{N} \sum_{t=m+1}^{N-1} \left(1 - \frac{t}{N}\right) \left[ \frac{e^{-(\lambda_k + \lambda_l)t}}{(1 - e^{-2\lambda_k})(1 - e^{-2\lambda_l})} + \frac{e^{-(\lambda_k + \lambda_l)t - (\lambda_l - \lambda_k)m}}{(1 - e^{-(\lambda_k + \lambda_l)})^2} \right] = \\ &= O\left(\frac{1}{N}\right), \quad \text{as } N \rightarrow \infty. \end{aligned} \quad (9)$$

Using simulations we investigated how the Criterion 2 works. We had made 10 000 Monte Carlo simulations of two sequences with covariance function defined by (6) and (7). For the simulations we used the methods described in the paper by Vasylyk et al. [18].

1. Let the null hypothesis ( $H_0$ ) state that the two components of the multivariate sequence  $\vec{\gamma}(n)$  are two jointly Gaussian, centered stationary sequences  $\gamma_1(n)$  and  $\gamma_2(n)$  with covariance function defined by (6) and (7) with parameters  $\lambda_1 = 1$  and  $\lambda_2 = 0.1$ , respectively, and the alternative hypothesis ( $H_a$ ) state that  $\lambda_1 = 0.3$  and  $\lambda_2 = 3$ .

We define the constants as  $\alpha = 0.1$ ,  $M = 10$ ,  $N = 1000$  ( $M + N = 1010$ ). Under these definitions the critical value  $\varepsilon_\alpha = 87.82$ .

We simulated 10 000 realizations of the two bivariate sequences with parameters  $\lambda_1 = 1$  &  $\lambda_2 = 0.1$  and  $\lambda_1 = 0.1$  &  $\lambda_2 = 1$ . We obtained an estimate for the probability of a type I error  $\hat{\alpha} = 0$  and an estimate for the probability of a type II error  $\hat{\beta} = 0.9616$ .

2. Let the null hypothesis ( $H_0$ ) state that the components of the multivariate sequence  $\vec{\gamma}(n)$  are two jointly Gaussian, centered stationary sequences  $\gamma_1(n)$  and  $\gamma_2(n)$  with covariance function defined by (6) and (7) with parameters  $\lambda_1 = 1$  and  $\lambda_2 = 0.1$ , respectively, and the alternative hypothesis ( $H_a$ ) state that  $\lambda_1 = 0.05$  and  $\lambda_2 = 5$ .

We used again simulated 10 000 realizations of the two bivariate sequences with parameters  $\lambda_1 = 1$  &  $\lambda_2 = 0.1$  and  $\lambda_1 = 0.05$  &  $\lambda_2 = 5$  and defined the constants as previously.

In this case we obtained an estimate for the probability of a type I error  $\hat{\alpha} = 0$  and the estimate for the probability of a type II error  $\hat{\beta} = 0.0104$ .

*Remark 7.* It is evident that the more observations we have the more sensitive the test is.

## 5. Conclusions

In this paper we estimated the distribution of the maximum of a random sequence which can be presented as a quadratic form of square Gaussian random variables. This result made it possible to build the criterion for testing a hypothesis on the expectation and covariance function of a *non*-centered univariate stationary Gaussian sequence and a hypothesis on the covariance function of a centered multivariate stationary Gaussian sequence. The simulation studies were also incorporated.

The inequality obtained in the section 2 can also be useful for testing the similar hypotheses for *non*-centered multivariate random sequences. Our test statistics are quite easy to compute and do not require the calculation of residuals from the fitted model. This is especially advantageous when the fitted model is not a finite order autoregressive model.

There is of course, a lot of room for improvement of the tests. Comparison with other tests and finding the number  $N$  for which the null and alternative hypotheses can be distinguishable are also very important issues for further investigation.

## 6. Acknowledgments

We are grateful to two anonymous referees for their insightful comments that have significantly improved the paper.

## 7. Annex to section 3

This Annex includes the requirements for the section 3 calculations for  $E\eta_a^2(m)$  and  $E\eta_B^2(m)$ .

$$E\eta_a^2(m) = E((\hat{a}_m - a)^2 - E_a)^2 = E(\hat{a}_m - a)^4 - (E_a)^2$$

Using the Isserlis' formula for the centered Gaussian random variables  $\tilde{\gamma}(n) := \gamma(n) - a$  we obtain

$$\begin{aligned} E(\hat{a}_m - a)^4 &= \frac{1}{N^4} \sum_{n=1}^N \sum_{k=1}^N \sum_{t=1}^N \sum_{s=1}^N E\tilde{\gamma}(n+m)\tilde{\gamma}(k+m)\tilde{\gamma}(t+m)\tilde{\gamma}(s+m) = \\ &= \frac{1}{N^4} \sum_{n=1}^N \sum_{k=1}^N \sum_{t=1}^N \sum_{s=1}^N [B(k-n)B(s-t) + B(t-n)B(s-k) + B(s-n)B(t-k)] = \\ &= \frac{3}{N^4} \left( \sum_{n=1}^N \sum_{k=1}^N B(k-n) \right)^2. \end{aligned} \quad (10)$$

Then

$$E\eta_a^2(m) = \frac{3}{N^4} \left( \sum_{n=1}^N \sum_{k=1}^N B(k-n) \right)^2 - \left( \frac{1}{N^2} \sum_{n=1}^N \sum_{k=1}^N B(k-n) \right)^2 = \frac{2}{N^4} \left( \sum_{n=1}^N \sum_{k=1}^N B(k-n) \right)^2. \quad (11)$$

Let us make the required calculation for  $E\eta_B^2(m)$ .

$$E\eta_B^2(m) = E(\hat{B}(m) - E_B(m))^2 = E(\hat{B}(m))^2 - (E_B(m))^2.$$

$$E(\hat{B}(m))^2 = \frac{1}{N^2} \sum_{n=1}^N \sum_{k=1}^N E(\gamma(n) - \hat{a}_0)(\gamma(n+m) - \hat{a}_m)(\gamma(k) - \hat{a}_0)(\gamma(k+m) - \hat{a}_m).$$

Using the Isserlis' formula we obtain that

$$E\eta_B^2(m) = S_0 - 2S_1 + S_2 - (E_B(m))^2, \quad (12)$$

where

$$\begin{aligned} S_0 &= \frac{1}{N^2} \left( N^2 B^2(m) + \sum_{n=1}^N \sum_{k=1}^N [B^2(k-n) + B(k-n+m)B(k-n-m)] \right); \\ S_1 &= \frac{1}{N^2} B(m) \sum_{n=1}^N \sum_{k=1}^N B(k-n+m) + \\ &+ \frac{1}{N^3} \sum_{n=1}^N \left[ \left( \sum_{k=1}^N B(k-n) \right)^2 + \left( \sum_{k=1}^N B(k-n+m) \right) \left( \sum_{l=1}^N B(l-n-m) \right) \right]; \\ S_2 &= \frac{1}{N^4} \left[ \left( \sum_{n=1}^N \sum_{k=1}^N B(k-n) \right)^2 + 2 \left( \sum_{n=1}^N \sum_{k=1}^N B(k-n+m) \right)^2 \right]. \end{aligned}$$

## 8. Annex to section 4

In chapter 4 we need to find the expectation of  $\Delta_{kl}^2(m)$  in order to calculate the value of  $E(\vec{\Delta}^T(m)\mathbf{B}\vec{\Delta}(m))$  (see formula (8)). Let us do it.

$$\begin{aligned} E\Delta_{kl}^2(m) &= E(\hat{B}_{kl}(m) - B_{kl}(m))^2 = E \left[ \frac{1}{N} \sum_{n=1}^N (\gamma_k(n)\gamma_l(n+m) - B_{kl}(m)) \right]^2 = \\ &= \frac{1}{N^2} E \left[ \sum_{n=1}^N \sum_{i=1}^N (\gamma_k(n)\gamma_l(n+m) - B_{kl}(m))(\gamma_k(i)\gamma_l(i+m) - B_{kl}(m)) \right] = \\ &= \frac{1}{N^2} \sum_{n=1}^N \sum_{i=1}^N E \gamma_k(n)\gamma_l(n+m)\gamma_k(i)\gamma_l(i+m) - B_{kl}^2(m). \end{aligned}$$

Using again the Isserlis' formula for the centered Gaussian random variables we obtain

$$\begin{aligned} \mathbb{E} \gamma_k(n) \gamma_l(n+m) \gamma_k(i) \gamma_l(i+m) &= \mathbb{E} \gamma_k(n) \gamma_l(n+m) \mathbb{E} \gamma_k(i) \gamma_l(i+m) + \\ &+ \mathbb{E} \gamma_k(n) \gamma_k(i) \mathbb{E} \gamma_l(n+m) \gamma_l(i+m) + \\ &+ \mathbb{E} \gamma_k(n) \gamma_l(i+m) \mathbb{E} \gamma_l(n+m) \gamma_k(i) = \\ &= B_{kl}^2(m) + B_{kk}(i-n) B_{ll}(i-n) + B_{kl}(i-n+m) B_{lk}(i-n-m). \end{aligned}$$

Then

$$\begin{aligned} \mathbb{E} \Delta_{kl}^2(m) &= \frac{1}{N^2} \sum_{n=1}^N \sum_{i=1}^N [B_{kk}(i-n) B_{ll}(i-n) + B_{kl}(i-n+m) B_{lk}(i-n-m)] = \\ &= \frac{1}{N} [B_{kk}(0) B_{ll}(0) + B_{kl}^2(m)] + \\ &+ \frac{2}{N} \sum_{t=1}^{N-1} \left(1 - \frac{t}{N}\right) [B_{kk}(t) B_{ll}(t) + B_{kl}(t+m) B_{lk}(t-m)] \end{aligned}$$

Putting the covariance function defined by (6) into the last formula we get (9).

### References

1. Anderson, T. W., 1971: *The Statistical Analysis of Time Series*. New York: John Wiley & Sons, 704 p.
2. Box, G. E. P., Jenkins, G. M., Reinsel, G.C., 2011: *Time Series Analysis: Forecasting and Control*, 4th Edition. Wiley Series in Probability and Statistics, 784 p.
3. Box, G. E. P. and Pierce, D. A., 1970: Distribution of the Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models. *J. Amer. Statist. Assoc.*, 65, p. 1509–1526.
4. Brockwell, P. J., Davis, R. A., 2009: *Time Series: Theory and Methods*. New York: Springer Series in Statistics, Springer-Verlag, Second Edition, 586 p.
5. Buldygin, V. V., Kozachenko, Yu. V., 2000: *Metric Characterization of Random Variables and Random Processes*. Amer. Math. Soc., Providence, R I, 257 p.
6. Chen, W. W., Deo, R. S., 2004: A Generalized Portmanteau Goodness-of-Fit Test for Time Series Models. *Econometric Theory*, 20(2), p. 382–416.
7. Chen, S. X., Härdle, W., Li, M., 2003: An Empirical Likelihood Goodness-of-Fit Test for Time series. *J. R. Statist. Soc. B*, 65, Part 3, p. 663–678.
8. Escanciano, J. C., 2007: Model Checks using Residual Marked Empirical Processes. *Statist. Sinica*, 17, p. 115–138.
9. Hosking, J. R. M., 1980: The Multivariate Portmanteau Statistic. *Statist. Sinica*, 75, p. 602–608.
10. Hosking, J. R. M., 1981: Lagrange-Multiplier Tests of Multivariate Time-Series Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43(2), p. 219–230.
11. Kozachenko, Yu. V., Fedorynych, T. V., 2004: A Criterion for Testing Hypotheses about the Covariance Function of a Gaussian Stationary Process. *Theory of Probability and Mathematical Statistics*, 69, p. 85–94.
12. Kozachenko, Yu. V., Stadnik, A. I., 1991: Pre-Gaussian Processes and Convergence in  $C(T)$  of Estimators of Covariance Function *Theory of Probability and Mathematical Statistics*, 45, p. 51–57.
13. Kozachenko, Yu. V., Stus, O. V., 1998: Square-Gaussian Random Processes and Estimators of Covariance Functions. *Math. Communications*, 3(1), p. 83–94.
14. Kozachenko, Yu. V., Yakovenko, T. O., 2010: Criterion for Testing the Hypothesis about the Covariance Function of the Stationary Gaussian Random Sequence. *Bulletin of Uzhgorod University, Series: Mathematics & Informatics*, 20, p. 39–43. (In Ukrainian)
15. Ljung, G. M., Box, G. E. P., 1978: On a Measure on Lack of Fit in Time Series Models. *Biometrika*, 65(2), p. 297–303.
16. McLeod, A. I. and Li, W. K., 1983: Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations. *J. Time Series Anal.*, 4, p. 269–273.
17. Mahdi, E. and McLeod, A. I., 2012: Improved Multivariate Portmanteau Test. *J. Time Series Anal.*, 33(2), p. 211–222.
18. Vasylyk, O. I., Kozachenko, Yu. V., Yakovenko, T. O., 2009: Simulation of the Stationary Random Sequences. *Bulletin of the University of Kyiv, Series: Physics & Mathematics*, 1, p. 7–10. (In Ukrainian)



## KAI KURIŲ ATSTITKTINIŲ SEKŲ SUDERINAMUMO KRITERIJAI

Yuriy Kozachenko, Tetiana Ianevych

**Santrauka.** Kvadratinių Gauso atsitiktinių dydžių teorijos rezultatai pritaikyti atsitiktinių sekų (laiko eilučių) suderinamumo kriterijams sudaryti. Nagrinėjami du tokių kriterijų atvejai. Pirmasis kriterijus skirtas tikrinti hipotezei apie vienmatės necentruotos sekos vidurkį ir kovariacinę funkciją; antrasis kriterijus skirtas tikrinti hipotezei apie daugiamatės centruotos sekos kovariacinę funkciją. Modeliavimo rezultatai iliustruoja šių kriterijų elgesį kai kuriais atskirais atvejais.

**Reikšminiai žodžiai:** suderinamumo kriterijus, daugiamatė atsitiktinė seka, laiko eilutė, kvadratinis Gauso atsitiktinis dydis.