

HOW BIG IS BIG ENOUGH?  
JUSTIFYING RESULTS  
OF THE IID TEST BASED  
ON THE CORRELATION INTEGRAL  
IN THE NON-NORMAL WORLD

L'uboř Briatka

**CERGE-EI**

Charles University  
Center for Economic Research and Graduate Education  
Academy of Sciences of the Czech Republic  
Economics Institute

**Working Paper Series**  
(ISSN 1211-3298)

**308**

**How Big is Big Enough?  
Justifying Results of the iid Test  
Based on the Correlation Integral  
in the Non-Normal World**

Luboš Briatka

CERGE-EI  
Prague, September 2006

**ISBN 80-7343-104-1 (Univerzita Karlova. Centrum pro ekonomický výzkum a doktorské studium)**  
**ISBN 80-7344-093-8 (Akademie věd České republiky. Národohospodářský ústav)**

# How Big is Big Enough? Justifying Results of the iid Test Based on the Correlation Integral in the Non-Normal World \*

Luboš Briatka <sup>†</sup>

CERGE-EI <sup>‡</sup>

September 2006

## Abstract

Kočenda (2001) introduced the test for nonlinear dependencies in time series data based on the correlation integral. The idea of the test is to estimate the correlation dimension by integrating over a range of proximity parameter  $\varepsilon$ . However, there is an unexplored avenue if one wants to use the test to identify nonlinear structure in non-normal data. Using the Monte Carlo studies, we show that non-normality leads to an over-rejection of the null hypothesis due to two reasons: First, the data are not iid, and second, the data are non-normal. It is shown that even a very small deviation from normality could lead to a rejection of the null hypothesis and hence a wrong conclusion. Therefore, the bootstrap method is introduced and it is shown that it helps to avoid the over-rejection problem; moreover the power of the test increases by a significant amount. These findings help us to extend the use of the test into many other fields that deal with nonlinear data that are not necessarily normal, e. g. financial economics, stock price volatility, stock market efficiency, stock exchange, behavior of equity indices, nonlinear dynamics in foreign exchange rates, or interest rates.

*Key Words:* chaos, nonlinear dynamics, correlation integral, Monte Carlo, power tests, high-frequency economic and financial data

*JEL Classification:* C14, C15, C52, C87, F31, G12.

---

\* I would like to thank Jan Hanousek, Jan Kmenta, Evžen Kočenda, and Petr Zemčík for their valuable comments. My thanks belong also to Marian Baranec, Martin Kubačka, Evžen Kočenda, the CERGE-EI computer department and many others that helped me with the extensive simulations. The work was supported by a grant from the Grant Agency of the Charles University No. 346/2005.

<sup>†</sup> E-mail: lubos.briatka@cerge-ei.cz, url: <http://home.cerge-ei.cz/briatka>

<sup>‡</sup> CERGE-EI (a joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute of the Academy of Sciences of the Czech Republic), P.O.Box 882, Politických vězňů 7, 111 21 Prague, Czech Republic

## Abstrakt

Kočenda (2001) představil proceduru na testování závislosti časových řad nelineárních procesů. Tato procedura vychází z korelačního integrálu a její idea spočívá v odhadnutí korelační dimenze pomocí integrace přes interval parametru  $\varepsilon$ . Použitím Monte Carlo metody v článku poukazujeme na to, že nenormalita testovaných dat vede v této proceduře k nadměrnému zamítání nulové hypotézy. To je způsobeno dvěma příčinami. Za prvé, testovaná data nejsou nezávislá a rovnoměrně rozdělená (iid), a za druhé, testovaná data nejsou z normálního rozdělení. V článku ukazujeme, že i velice jemná odchylka od normality může vést k okamžitému zamítání nulové hypotézy, a tím k nesprávnému závěru. Proto představujeme bootstrap metodu, která eliminuje problém nadměrného zamítání a zvyšuje sílu testu. Tyto závěry umožňují aplikovat test na mnohé oblasti, které běžně pracují s nenormálními daty, jako jsou analýzy burzovních trhů, modelování volatility výměnných kurzů a úrokových sazeb, nebo chování se akciových indexů.

## 1. INTRODUCTION

The end of the 20<sup>th</sup> century witnessed a renaissance of nonlinear modeling in econometrics. This huge movement in nonlinear modeling was motivated by early theoretical work in nonlinear estimation of a high-frequency time series (see Engle (1982) and Bollerslev (1986)) and later enormous empirical studies in finance (see Brock, Hsieh and LeBaron (1993) as reference study). The research of dozens of economists in this promising field was cumulated in 2003 with a Nobel Prize for one of the fathers of nonlinear modeling, Robert Engle: "for methods of analyzing economic time series with time-varying volatility".

A large expansion of theories based on Engle's findings in the eighties and nineties have been supported with empirical evidence from many fields. Many studies from the areas of financial economics, stock price volatility, stock market efficiency, stock exchange, behavior of equity indices, nonlinear dynamics in foreign exchange rates, and interest rates frequently brought new empirical evidence that there is a need to model nonlinearity. In many cases, not including a nonlinear component in a model caused vast quantitative and qualitative mistakes. Therefore since the eighties, the testing for nonlinear patterns in a time series became the norm.

As an alternative to the very popular BDS-test<sup>1</sup>, Kočenda (2001) introduced a test for nonlinear dependencies in a time series based on the correlation integral. The idea of the test is to estimate the correlation dimension by integrating over a range of proximity parameter  $\varepsilon$ . Using the Monte Carlo studies, Kočenda and Briatka (2005) showed that this test is very powerful over many nonlinear models and has an ability to reveal much more of the hidden nonlinear dependencies than other well-known nonlinear tests.

However, there is an "unexplored avenue" if one wants to use the Kočenda's test for the identification of nonlinear structure in data that do not meet the normality assumption. If this is the case, the test is biased towards rejecting the null hypothesis of independently and identically distributed data (see Kočenda, Briatka (2005) for some evidence). In this case, the rejection power of the test is deformed<sup>2</sup> by two components: The first component is the natural power of the test to reject non-iid data. The second one is derived from the fact that data are non-normal. In other words, the

---

<sup>1</sup> The BDS-test is one of the most common nonlinear tests and is exhaustively described in Section 2.2.

natural power of the test is strengthened with a disturbing effect of inappropriate non-normality in data. The intensity of this disturbing effect and its impact on the results of the test depend on the extent of the data's deviation from normality. The aim of this paper is to measure this effect and evaluate the ratio and strength of nonlinear and non-normal processes in the rejection power of the Kočenda's test. This justification fills in the gap in the methodology of the test and allows us to apply the testing procedure to a broader set of non-normal data.

The paper is organized as follows: In Section 2, we briefly present the concept of nonlinearity and basic methods used for testing nonlinearity in data; in Section 3, we introduce normality and discuss the connection between normality and nonlinearity. In Section 4, we provide Monte Carlo studies and the bootstrap testing of our procedure; in Section 5, we comment on the results and discuss possible implications and conclude in Section 6.

## 2. METHODOLOGY AND LITERATURE REVIEW

The idea of testing for nonlinearities in data arose in the early eighties when Engle (1982) was doing pioneer research in studying the volatility of high-frequency data. As the theory of modeling high-frequency data was growing and becoming very popular among econometricians due to its extensive applications (e.g. ARCH, GARCH or NLMA models), there was a need to test for nonlinearities. One of the largest groups of tests consists of testing procedures based on the correlation integral.

### 2.1. Correlation Integral, Correlation Dimension, and The BDS-test

The tests based on the correlation integral are built on the idea of a correlation integral, a popular method of estimating the dimension of an attractor. Formally, let  $\{x_t\}$  be a time series of the size  $T$  generated randomly according to a density function  $f$ . Form  $m$ -dimensional vectors, called  $m$ -histories,  $x_t^m = (x_t, x_{t+1}, \dots, x_{t+m-1})$ . Then the correlation integral at embedding dimension  $m$  is defined as

$$C_m(\varepsilon) = \lim_{T \rightarrow \infty} C_{m,T}(\varepsilon), \quad (2.1)$$

---

<sup>2</sup> By deforming we mean incorrectly strengthening up.

where

$$C_{m,T}(\varepsilon) = 2 \sum_{t=1}^{T_m-1} \sum_{s=t+1}^{T_m} H(\varepsilon - \|x_t^m - x_s^m\|) / (T_m(T_m - 1)), \quad (2.2)$$

is the sample correlation integral at embedding dimension  $m$ . Here,  $\varepsilon$  is a sufficiently small number,  $T_m = T - m + 1$ ;  $H(z)$  is the Heaviside function that maps positive argument into 1 and nonpositive into 0; and  $\|\cdot\|$  is the distance induced by the selected form. Thus, the sample correlation integral measures the fraction of pairs that lie within the tolerance distance  $\varepsilon$  for the particular embedding dimension  $m$ . If  $C_{m,T}(\varepsilon)$  is large (close to 1), even for a very small  $\varepsilon$ , then the data are correlated.

Grassberger and Procaccia (1983) showed that for a very small  $\varepsilon$ ,  $C_{m,T}(\varepsilon)$  grows exponentially, i.e.  $C_{m,T}(\varepsilon) = \eta e^D$ , where  $\eta$  is a constant, and  $D$  is so called the correlation dimension. If the increase in  $C_{m,T}(\varepsilon)$  concerns increasing  $\varepsilon$ , then most data points are near to each other, and the data are well correlated. Therefore, the higher the correlation dimension, the less correlated the data are, and the system is random (stochastic). Formally, the correlation dimension is defined as

$$D = \lim_{\varepsilon \rightarrow 0} \lim_{T \rightarrow \infty} \frac{\ln C_{m,T}(\varepsilon)}{\ln \varepsilon}. \quad (2.3)$$

The correlation integral and the correlation dimension originate in Grassberger and Procaccia (1983) and appears to be a robust method of detecting nonlinear components in data.

Among many others tests<sup>3</sup>, a well-known BDS-test was invented by Brock, Dechert, and Scheinkman (1987) as a non-parametric method of testing for nonlinear patterns in a time series. Using the correlation integral, they constructed the test and provided its theoretical background. The null hypothesis of the test is that the data in a time series are independently and identically distributed and the alternative hypothesis is specified very widely. The alternative hypothesis is just opposite, i.e. that the null hypothesis is not true<sup>4</sup>. The test statistic, which measures the fraction of the data that lies within a given relative (and small enough) distance  $\varepsilon$  for embedding dimension  $m$  is formally described by the formula

$$\text{BDS}_{m,T}(\varepsilon) = T^{\frac{1}{2}} [C_{m,T}(\varepsilon) - C_{1,T}(\varepsilon)^m] / \sigma_{m,T}(\varepsilon), \quad (2.4)$$

---

<sup>3</sup> Other tests will be discussed in Section 2.4.



where  $\sigma_{m,T}(\varepsilon)$  is an estimate of the asymptotic standard deviation of  $C_{m,T}(\varepsilon) - C_{1,T}(\varepsilon)^m$ . The BDS statistic is asymptotically standard normal under the whiteness null hypothesis and can be estimated  $\sqrt{n}$ -consistently.

The intuition behind the BDS-statistic comes from the fact that the sample correlation integral  $C_{m,T}(\varepsilon)$  is an estimate of the probability that the distance between two  $m$ -histories,  $x_t^m$  and  $x_s^m$ , should be less than  $\varepsilon$ . If they were independent, then for  $t \neq s$  the probability of joint event is equal to the product of the individual probabilities. Moreover, if  $x_t^m$  were also identically distributed, all of the  $m$  probabilities under the product sign would be the same. This logic is equivalent to the null hypothesis of white noise.

In the early nineties the BDS-test turned out to be widely used. Several studies were published in order to provide strong statistical and empirical properties of the BDS-test. Hsieh (1991) and Hsieh and LeBaron (1993) ran a large number of power tests that showed the great ability of the BDS-test to detect nonlinearities in ARCH, GARCH, NLMA, and TAR models. Later, Brock et al. (1996) extended their original framework and prepared software to be run in the MATLAB and C++ environment. Further, Barnett et al. (1997) compared the power of the BDS-test with the power of the most common nonlinear tests in ARCH, GARCH, NLMA, and ARMA models and found the BDS-statistic to be one of the most appropriate. Finally, the most recent study by Kanzler (1999) proved that the BDS-statistic is badly sized on small samples and thus yields misleading conclusions. Fortunately, he provided asymptotically corrected critical values for the small sample time series.

Although these studies were mostly in favor of the BDS-test, there were several substantial difficulties with this test that appeared to be very discouraging. First, to compute the BDS-statistic, one needs to set two parameters *ex ante*: the embedding dimension  $m$  and the correlation distance (the proximity parameter)  $\varepsilon$ . The problem is that for various pairs of the parameters, it is possible to get qualitatively different values of the statistic and therefore possibly contradictory results. The right choice of the parameters is widely debated in the statistical literature but no general suggestion

---

<sup>4</sup> LeBaron (1994) argues that “this feature can be viewed as both a cost and a benefit. On the one hand, it can detect many types of nonlinear dependence that might be missed by other tests. On the other hand, a rejection using this test is not very informative.”

has been made.<sup>5</sup> Second, even though Monte Carlo simulation studies provide strong evidence for using the BDS-test in particular types of distributions, there are still many distributions the test does not fit very well.

## 2. 2. Kočenda's Test

To solve the problem that different parameters can get contradictory results, a new test for detecting hidden nonlinear patterns in data was proposed by Kočenda (2001). He suggested an alternative test using integration across the correlation integral

$$\hat{\beta}_m = \frac{\sum_{\varepsilon} (\ln(\varepsilon) - \overline{\ln(\varepsilon)}) \cdot (\ln(C_m(\varepsilon)) - \overline{\ln(C_m(\varepsilon))})}{\sum_{\varepsilon} (\ln(\varepsilon) - \overline{\ln(\varepsilon)})^2}, \quad (2.5)$$

where  $\varepsilon$  runs through a range of standard deviation  $\sigma$  of the time series. This test eliminates the arbitrary selection of the proximity parameter  $\varepsilon$  and reduces the choice of parameters to only one by integrating the correlation integral across the proximity parameter  $\varepsilon$ . The null hypothesis of the test is that data are iid, and the intuition behind this is to estimate the slope coefficient from the least squares regression of log of sample correlation integral to the log of proximity parameter  $\varepsilon$ . The regression can be formally described as

$$\ln(C_{m,T}(\varepsilon_i)) = \alpha_m + \beta_m \ln(\varepsilon_i) + u_i, \quad i = 1, \dots, n. \quad (2.6)$$

Kočenda (2001) suggested estimating the slope  $\beta_m$  using the integration of the correlation integral through the range  $(0.25\sigma; 1.00\sigma)$  within the formula (2.5). An empirical comparison of this new test with the BDS-test showed that Kočenda's test is able to detect the hidden patterns with much higher precision.

Belaire-Franch (2003) considered the Kočenda's test by using the bootstrapping method (the random shuffle) in order to avoid asymptotically biased critical values<sup>6</sup>. He recommended using two different ranges for the correlation distance  $\varepsilon$ . However, the new ranges were constructed only as an additive extension of the original range, and no theoretical or practical arguments were provided to support the choice.

---

<sup>5</sup> See Brock, Hsieh and LeBaron (1993) as a reference textbook for choice of parameters related issues.

<sup>6</sup> If data are non-normal, and one uses the tabulated critical values, it leads to overrejection of the null hypothesis.

Kočenda and Briatka (2005) extended Kočenda's original framework. They provided broad power test studies to declare the strong ability of Kočenda's test in finding nonlinear dependencies. As a result of these studies, two new ranges for correlation distance  $\varepsilon$ ,  $(0.50\sigma; 1.50\sigma)$  and  $(0.25\sigma; 2.00\sigma)$ , were offered, and new critical values computed.<sup>7</sup> Moreover, they ran sensitivity analysis within these intervals. By contracting and expanding these intervals, they find the optimal  $\varepsilon$ -range  $(0.60\sigma; 1.90\sigma)$  that was suggested to be used as a template option.<sup>8</sup> It was shown that using new  $\varepsilon$ -ranges, the test can be even more powerful than the BDS-test. The comparison with the existing results of the controlled competition of Barnett et al. (1997) as well as power tests on various nonlinear data were provided, and the result of the comparison was strongly in favor of this robust procedure.<sup>9</sup> Finally, new user-friendly software that is capable of running Kočenda's test for all embedding dimensions was built.<sup>10</sup>

### 2. 3. Other Tests for Nonlinear Dependencies

Until now, there have been dozens of diagnostic tests for nonlinearity available, so we will only mention those most used. Several authors used the test developed by Tsay (1986). Also, the bispectral test developed by Hinich (1982) is applied to several macro economic series in Ashley and Patterson (1989) and Barnett and Hinich (1993). The null hypothesis of this test is that the skewness function is flat, and hence, there is a lack of third order nonlinear dependence. Nychka et al. (1992) proposed a nonparametric test for positivity of the maximum Lyapunov exponent, which is a direct test for chaos. Further, White (1989)<sup>11</sup> introduced a test for nonlinearity, a test of the null hypothesis of the linearity in the mean, and Kaplan (1994)<sup>12</sup> described a test for nonlinearity, which is a test of the null hypothesis of linearity of the dynamics found in the data. These procedures are all tests for some kind of nonlinear dependencies; however, they are very specific.

---

<sup>7</sup> The results of these power tests are not in the paper themselves. They can be found in a working paper by Kočenda and Briatka (2004).

<sup>8</sup> A deeper discussion of how and when this interval is used can be found in Section 3.

<sup>9</sup> The exact formulas of processes used in this study are in Section 4.4.

<sup>10</sup> It is known that if the technique is easy to understand and to implement, and particularly if a convenient computer program is available, then it is likely to be used.

<sup>11</sup> For further details about this test see Lee, White, and Granger (1993) and Jungeilges (1996).

<sup>12</sup> The test depends on the embedding dimension  $m$  and works comparing the distances between points in an  $m$ -dimensional reconstructed space with the distances between their images.

### 3. NON-NORMALITY AND NONLINEARITY

#### 3. 1. Introducing Concepts of Normality and Nonlinearity

In this paper, we distinguish between two types of time series – non-normal (but still random) time series and nonlinear (and hence determined) time series. To make it clear for the rest of the paper, we will discuss the differences between these two time series here in more detail.

We consider the non-normal times series to be randomly drawn from statistical distribution with well-known properties (student distribution,  $\chi^2$ -distribution, beta distribution,  $F$ -distribution, etc.). An important fact is that the future values of non-normal time series cannot be forecasted using past values. However, because of the known distribution, we can still compute the probabilities that the expected value will be from a given interval. The non-normal time series is random in nature, and our test should not reject the null hypothesis of iid-ness.

Contrary to the non-normal time series, the nonlinear time series are meant to be driven by some kind of deterministic process (autoregressive conditional heteroscedasticity, nonlinear moving averages, chaos, etc.). Therefore, future values of nonlinear time series are fully observable once we obtain the formula that is behind the nonlinear process. It means that the future value is known with the probability equal to one. The nonlinear time series is not random, and hence, it is very likely that our test will reject the null hypothesis.

The difference between non-normal and nonlinear time series is more evident if we incorporate Kočenda's test. The usual testing procedure is described in Figure 1 and could be summarized in the following way. In the pre-test phase (first step), a researcher prepares data and runs a regression<sup>13</sup>. In the second step, the residuals are subject to tests for normality<sup>14</sup>. These residuals create the input time series for our test.

If the tests do not reject the normality of the time series (Branch A in Figure 1), then in the third step one proceeds with computing the Kočenda's statistics and comparing the results with tabulated critical values. If the test does not reject the null

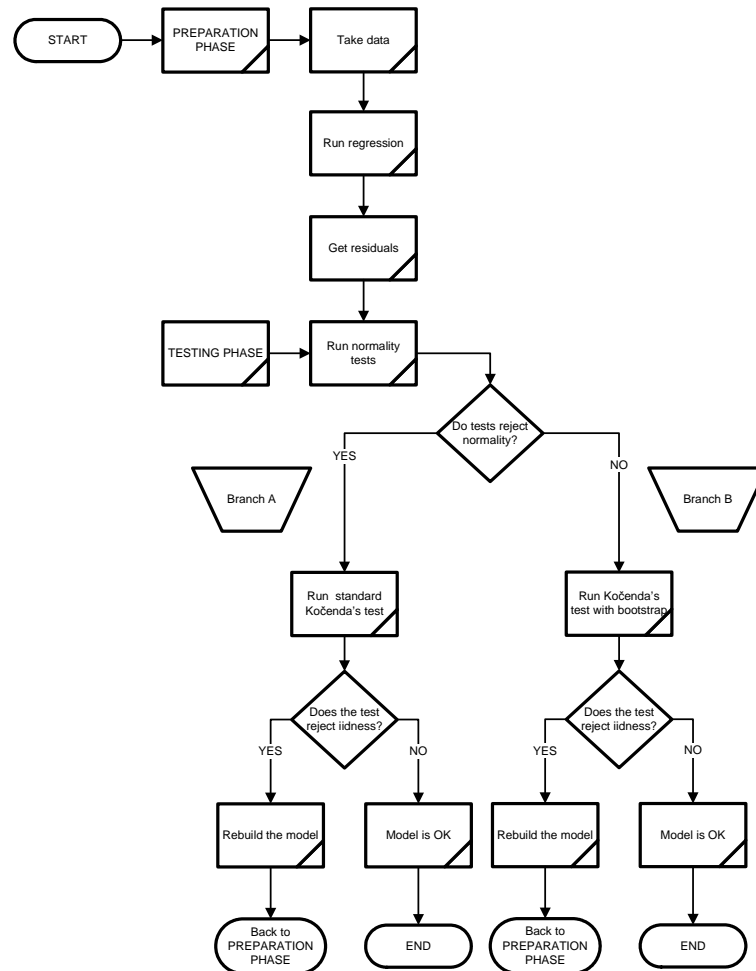
---

<sup>13</sup> We do not specify any kind of data or regression. It depends on the current research of a researcher. The parts relevant for our research start from the second step. We include the first step only in order to picture the complexity of the whole process. In general, the origin of the data could be unknown.

<sup>14</sup> We discuss the most used normality tests in the next section.

hypothesis, the residuals are iid and the specification of the model can be used in further analysis.

Figure 1  
Process of testing time series for nonlinearities



The second clearest result is if the test in the third step rejects the null hypothesis of iid-ness, then the data are not independent nor drawn from the random distribution. They contain some nonlinear structure, and therefore, one should try to rebuild the design of the model. After a new design is found, the previously mentioned analysis should be run again from the first step.

A different kind of logic applies in the case of tests in the second step rejecting the null hypothesis (branch B in Figure 1). It means that our residuals are not normal; hence, the result of the Kočenda's test when using a tabulated critical value is obscure, and one would not be able to distinguish whether the data are randomly drawn from known distribution or they follow a nonlinear process.

In this case, we have three options. The first option is to run the test against several types of one-sided hypotheses of exact distribution. This approach is very straightforward, but we are forced to compute a large amount of critical values. Moreover, in this kind of approach, the philosophy of the test is destroyed, and its strength (to reveal much more of the hidden nonlinear dependencies than other nonlinear tests) is broken down.

The second option is to proceed with the bootstrap. The bootstrap method gives us estimated critical values that can be easily used with the Kočenda's test without thoughts of normality. The problem of this approach is its exceptional demand for computational time<sup>15</sup> that is completely impractical if one wants to test hundreds of model specifications. Anyway, the results of this approach are very promising, and we shall discuss them in more details in Section 4.2.

The third option is to use the basic Kočenda's test and use the tabulated critical values. However, to recognize the true process, one has to consider the extent of the deviation from normality, or in other words, one has to know how big is big enough. This point is crucial in using the test, and we shall come to it from a different perspective later in Section 5.

### 3. 2. Survey of Tests for Normality

In Section 2.3, we described the most common tests for nonlinearity. In this Section, we shall survey the basic tests for normality.

The basic testing for normality of the data is related to the moment statistics of a distribution. If we expected the time series to have zero mean (expectation of the linear deviation) and unit variance (expectation of the quadratic deviation), the next candidates for testing are skewness and kurtosis coefficients.<sup>16</sup> Skewness is a measure of the symmetricity of a distribution (expectation of the cubic deviation), while kurtosis is a measure of the thickness of the tails of a distribution (expectation of the quadruplicate deviation). The coefficients of skewness and kurtosis are formally defined as

---

<sup>15</sup> On a slow computer, it can take up to an hour to compute the Kočenda's statistics for a single time series using the bootstrap method.

<sup>16</sup> In generally the normal distribution with non zero mean or/and non unit variance is not a problem. The procedure used in Kočenda's test standardizes the time series to zero mean and unit variance itself.

$$\tau = \frac{\mu_3}{\sigma_3} = \frac{E[(x - \mu)^3]}{E[(x - \mu)^2]^{3/2}} \quad (3.1)$$

$$\kappa = \frac{\mu_4}{\sigma_4} = \frac{E[(x - \mu)^4]}{E[(x - \mu)^2]^2}, \quad (3.2)$$

where  $\{x_t\}$  is a time series with mean  $\mu$  and standard deviation  $\sigma$ . In practice, coefficients  $\tau$  and  $\kappa$  are estimated using the sample moments  $\hat{\mu}_r = T^{-1} \sum_{t=1}^T (x_t - \bar{x})^r$ . The normal distribution has skewness coefficient equal to 0 and kurtosis equal to 3. If  $\{x_t\}$  is normally distributed and iid, then  $\sqrt{T}\hat{\tau} \xrightarrow{D} N(0,6)$  and  $\sqrt{T}(\hat{\kappa} - 3) \xrightarrow{D} N(0,24)$ .

In order to test for normality very often, the Bera-Jarque test is employed. This standard procedure tests for the joint hypothesis of having  $\tau = 0$  and  $\kappa = 3$ . Bera and Jarque (1981) showed that

$$T \left( \frac{\hat{\tau}^2}{6} + \frac{(\hat{\kappa} - 3)^2}{24} \right) \xrightarrow{D} \chi_2^2. \quad (3.3)$$

In other words, the test is asymptotically distributed as chi-squared with two degrees of freedom.<sup>17</sup> Later in the paper, we shall use the skewness and kurtosis coefficients and the Bera-Jarque statistics as the basic description of connection to normality and measure of departure from normality.

Other tests for normality of the data are based on the empirical distribution function. The most known tests of this kind are the Kolmogorov-Smirnov (KS), the Anderson-Darling (AD) and the Cramer-von Mises (CM) tests. These tests are very often referred to as EDF tests. Because the idea of testing is the same for all cases and the only differences are the test statistics, we shall discuss just the Kolmogorov-Smirnov test in more detail.

The null hypothesis of the KS-test is that the data follows a specified distribution (in our case the normal one); the alternative is that the data do not follow this distribution. The Kolmogorov-Smirnov test statistic is defined as

$$D = \sup |F_n(x) - F(x)|, \quad (3.4)$$

where  $F$  is the theoretical cumulative distribution of the distribution being tested, which must be a continuous distribution, and it must be fully specified. The

---

<sup>17</sup> See also Bera and Jarque (1982) for more details.

hypothesis regarding the distributional form is rejected if the test statistic,  $D$ , is greater than the critical value obtained from a table.

There are several variations of these tables in the literature that use somewhat different scaling for the KS test statistic and critical regions. These alternative formulations should be equivalent, but it is necessary to ensure that the test statistic is calculated in a way that is consistent with how the critical values were tabulated.

As we shall see in the next Section, testing for normality is an important step in finding nonlinear dependencies in a time series. We undertook just a quick survey of the most common normality tests. For more details about testing normality issues, we referred to Lehman (1997).

#### **4. MONTE CARLO STUDIES AND BOOTSTRAP**

To test the impact of the non-normality vs. nonlinearity on the result of the test, we compare several models with low, middle, and heavy disturbances of non-normality in extensive Monte Carlo competition. The models were selected in order to deviate from the normality in all ways considering the deviation from Section 3.

The general procedure of generating a time series is summarized in the following setup. First, samples of 10 000 time series with 1 000 observations<sup>18</sup> were generated for the given process, and so-called “true” critical values were obtained from this generation. Second, new 1 000 time series were generated from the same process, and these generations were used as input for the testing procedure to get the value of the Kočenda’s statistics. Third, the Kočenda’s statistics computed in the previous step were compared with the critical values of the normal distribution and also with the “true” critical values. The differences between the results in the very last stage of the previous setup can be directly used as the measure of the error based on the usage of inappropriate critical values in the test.

After having a look at the error based on using inappropriate critical values, new 500 time series with 1 000 observations were generated and the bootstrap method employed to generate bootstrapping critical values for each time series.<sup>19</sup> The results

---

<sup>18</sup> This study was dealing only with a time series with 1 000 observations. A comparable study to the same extent, but with 2 500 observations in one file is not technically executable in the present time due to the extensive computer time that would be needed.

<sup>19</sup> We will come to this issue in more detail in Section 4.5.



based on the bootstrapping critical values and “true” critical values are now compared. The efficiency of the bootstrap method is discussed.

Before presenting the results, we introduce the basic distribution and processes used in our Monte Carlo study. These processes contain 4 time series drawn from student distribution (to evaluate the impact of improper kurtosis on results); 4 time series drawn from  $\chi^2$ -distribution (to evaluate the impact of improper skewness on results); 4 time series based on some nonlinear processes (ARCH, GARCH, NLMA, ARMA<sup>20</sup> – to evaluate the impact of bootstrap on the results); and 4 time series drawn from normal distribution with a different gauge of non-normal contamination (to evaluate the impact of unknown disturbances on the results).

The data were generated using an inversive congruential generator. The first 1 000 observations were discarded to avoid dependency on the initial condition. The generated data were randomly shuffled to reduce any hidden non-random dependencies in the data. The practical advantage of an inversive congruential generator (ICG) against a linear congruential generator (LCG) is that ICG guarantees the absence of a lattice structure. We have opted for the ICG for its superiority, despite the fact that it is significantly slower than LCG. Both generators are easy to implement, and there is abundant literature available with the portable code, parameters, and test results. For a concise survey of the performance of inversive random number generators in theoretical and empirical tests, as well as tables of parameters to implement inversive generators see Hellekalek (1995). For a survey of the latest concepts and results of random number generation, we recommend starting with L’Ecuyer (2004).

Finally, all computations are done using the  $(0.60\sigma; 1.90\sigma)$  interval for proximity parameter  $\varepsilon$ . This interval was found to be the optimal range for the Kočenda’s test. For more details see Kočenda and Briatka (2005).

---

<sup>20</sup> We know that the last process is linear; however it was used previously in Barnett et al. (1997), Kočenda and Briatka (2004), and Kočenda and Briatka (2005), and hence we hold it here for compactness of the study.

## 4. 1. Distributions to Consider for Simulations

### 4. 1. 1. Student Distribution

The student distribution is derived from the normal distribution - it is the distribution of the mean/standard deviation of a sample of normally distributed values with unknown variance. It is a characteristic due to its high kurtosis and consequently heavy tails. As a consequence, the data are expected to be concentrated in the centre of distribution more often than in the case of normal distribution. However, as we increase degree of freedom student distribution converges to normal distribution. Formally, the probability density function for student's distribution is given by the formula

$$f_n(x) = \frac{\Gamma(\frac{n}{2})}{\sqrt{n}\Gamma(\frac{n-1}{2})} (1+x^2)^{-n/2}, \quad (4.1)$$

where  $n$  is the number of freedom and  $\Gamma$  is the gamma function<sup>21</sup>.

The first four time series that were used in the Monte Carlo competition are time series randomly drawn from the student distribution with various degrees of freedom. Different degrees of freedom mean different departures from normality. The time series used in the competition are drawn from the student distribution with 3, 9, 25, and 36 degrees of freedom. All time series are normalized to have zero mean and unit variance. The skewness coefficients of these time series are close to zero, and kurtosis coefficients are approaching 3, it is 24.91 for student (3); 4.33 for student (9); 3.26 for student (25); and 3.21 for student (36).

It is obvious that the data from student distribution with 3 degrees of freedom deviate the most from the normal distribution, contrary to those from student distribution with 36 degrees of freedom that are very close to data drawn from normal distribution.

We used these data to model the behavior of the test in the non-normal world that are still independent. The results of the power tests of student distribution using tabulated critical values in Section 4.3 could provide the justification for using the

---

<sup>21</sup> The formula for the gamma function is  $\Gamma(a) = \int_0^{\infty} t^{a-1} e^{-t} dt$ .

tabulated critical values in the distribution with heavy tails. However, we show that this is not the case.

#### 4. 1. 2. $\chi^2$ -distribution

The second set of four time series was drawn from a  $\chi^2$ -distribution. The importance of the  $\chi^2$ -distribution stems from the fact that it describes the distribution of the variance of a sample taken from a normally distributed population. The  $\chi^2$ -distribution is derived from the normal distribution - it is the distribution of a sum of squared normally distributed variables, and therefore it is never less than zero. It is a typical nonsymmetrical distribution and is skewed to the right. Formally,

$$f(x) = \frac{e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)}, \text{ for } x \geq 0, \quad (4.2)$$

where  $\nu$  is the number of degrees of freedom  $\Gamma(\cdot)$  is the gamma function. Similarly to the student distribution, the standardized  $\chi^2$ -distribution also converges to normal distribution as degrees of freedom approaches infinity.

The time series used in the paper are drawn from the  $\chi^2$ -distribution with 3, 8, 18, and 48 degrees of freedom. The largest departure from the normality shows data from  $\chi^2$ -distribution with 3 degrees of freedom; on the other hand, distribution with 48 degrees of freedom is often approximated by normal distribution. All time series are normalized to have zero mean and unit variance. We chose these time series to measure the impact of skewness of a distribution on the results of the test. These coefficients should be close to zero; however, they are 1.62 for  $\chi^2(3)$ ; 1.08 for  $\chi^2(8)$ ; 0.69 for  $\chi^2(18)$ ; and 0.41 for  $\chi^2(48)$ .

We used the  $\chi^2$ -distribution to verify the behavior of the test in the non-normal world but still independent. The results of the power tests of  $\chi^2$ -distribution using tabulated critical values in Section 4.3 show the error one can make when using the tabulated critical values in skewed distributions.

### 4. 1. 3. Nonlinear Distributions

The third pack of time series are based upon four nonlinear processes. These processes are classically used for modeling the nonlinearity in time series econometrics. Moreover, these processes were previously used in tests' competition provided by Barnett et al. (1997) and extensive power tests competition provided by Kočenda and Briatka (2005), and therefore, we used them again as a proxy. These processes are GARCH, NLMA, ARCH, and ARMA. A more specific description follows:

1. A generalized autoregressive conditional heteroscedasticity model (GARCH) of the form:

$$y_t = h_t^{1/2} u_t, \quad (4.1)$$

where  $h_t$  is defined by  $h_t = 1 + 0.1y_{t-1}^2 + 0.8h_{t-1}$ , with  $h_0 = 1$  and  $y_0 = 0$ ;

2. A nonlinear moving average model (NLMA) of the form:

$$y_t = u_t + 0.8u_{t-1}u_{t-2}; \quad (4.2)$$

3. An autoregressive conditional heteroscedasticity model (ARCH) of the form:

$$y_t = (1 + 0.5y_{t-1}^2)^{1/2} u_t, \quad (4.3)$$

with the value of the initial observation set at  $y_0 = 0$ ;

4. An autoregressive moving average model (ARMA) of the form:

$$y_t = 0.8y_{t-1} + 0.15y_{t-2} + u_t + 0.3u_{t-1}, \quad (4.4)$$

with  $y_0 = 1$  and  $y_1 = 0.7$ .

These processes deviate more or less from the previously mentioned assumption of normality. All time series generated based on these processes are standardized to have zero mean and unit variance. Their skewness and kurtosis coefficients are  $-0.04$  and  $3.92$  for GARCH;  $-0.03$  and  $6.02$  for NLMA;  $0.02$  and  $3.31$  for ARCH; and  $0.01$  and  $3.08$  for the ARMA process.

From the first sign, it is clear that all processes are not skewed<sup>22</sup>. Moreover, their departures in kurtosis coefficients are comparable to those of student distribution with a high degree of freedom. The only large departure from the normality shows in

---

<sup>22</sup> However, we do not provide any statistical tests, but their skewness and kurtosis coefficient are much closer to those of Normal distribution than processes in Section 4.1.1 and 4.1.2.

NLMA data. Its statistical properties are comparable to a student distribution with 9 degrees of freedom.

We used these nonlinear processes to verify the behavior of the test in the non-normal and dependent world. The results of the power tests of these processes provide justification for the usage of the tabulated critical values in this kind of distribution. The comparison with the results based on bootstrap in Section 4.3 provides a comparison of the strength of the test.

## 4. 2. The Bootstrap Method<sup>23</sup>

Bootstrapping is a method for estimating the sampling distribution of an estimator by resampling with a replacement from the original sample. Because of an incredible increase of computer power, the bootstrap method has become very popular. For references, especially in econometrics, see Davidson and MacKinnon (2000) or Efron and Tibshirani (1998). Based on the original time series, the bootstrap method gives us virtual populations that can be considered as the true data population.

To proceed with bootstrapping, we first generate a sample of 500 time series. Every time series contains 1 000 observations and is based on one of the distributions or processes described in Section 4.1. Having chosen one time series from this sample, the bootstrap comes into play. For this single original time series, we generate 200 new time series. These 200 new time series are then used as a virtual population to compute so called bootstrapped critical values of Kočenda's test for the original time series. Then, the bootstrapped critical values are employed to evaluate the Kočenda's test on the original time series and bring a reject/no reject conclusion.

The bootstrap method here helps us to distinguish the power of the test between (1) a rejection that is made due to a usage of improper critical values, and (2) a rejection due to a nonlinear dependency in the data. If the test gives different results for the tabulated critical values and the bootstrapped ones, that would be a sign of inappropriate usage of the tabulated critical values. If the test offers the same results in both cases that would serve as a double check for usage of tabulated critical values.

---

<sup>23</sup> Bootstrapping alludes to a German legend about Baron Münchhausen, who was able to lift himself out of a swamp by pulling himself up by his own hair. In later versions, he was using his own boot straps to pull himself out of the sea, which gave rise to the term bootstrapping.

Bootstrapping is time consuming, and often it is not necessary to proceed with it to get clear evidence of the distribution of data. In the rest of the paper, we provide easy guidance based on the departures from normality that will show us when it is necessary to provide the bootstrap and when it would be satisfactory to use tabulated critical value.

### 4. 3. Results of the Monte Carlo studies

The Monte Carlo studies were performed to measure the size of a Type I and a Type II error of the test. There are two kinds of errors that can be made in significance testing: (1) a true null hypothesis can be incorrectly rejected, and (2) a false null hypothesis can fail to be rejected. The former error is called a Type I error, and the latter error is called a Type II error. The size of a Type I error is the probability that a Type I error is made and similarly the size of a Type II error is the probability that a Type II error is made. To describe the basic properties of the test means to describe the size of these errors.

As we mentioned before, the tests on iid sets of data were performed to show the size of a Type I error when using tabulated critical values. All tests were evaluated at a constant, 5% significance level. Results of the studies for student distribution are in Table 1.

Table 1  
Empirical power of Kočenda's test against Student distribution using tabulated critical values  
The numbers in the table represent the size of a Type I error of the test for the embedding dimension m

Process	m=2	m=3	m=4	m=5	m=6	m=7	m=8	m=9	m=10
student(3)	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
student(9)	92.20%	87.60%	81.80%	71.20%	62.40%	50.60%	42.00%	35.00%	32.00%
student(25)	25.00%	21.00%	17.80%	16.00%	13.00%	9.20%	8.00%	7.40%	9.20%
student(36)	13.40%	13.20%	10.60%	9.40%	8.00%	7.60%	8.20%	6.60%	6.00%

The results clearly state that using tabulated critical values in the case of fat-tailed distributions leads to over-rejection of the null hypothesis. This is evident from the results of student distributions with 3 and 9 degrees of freedom, where the smallest rejection rate is 32%, but the expected rate was about 5%. These results are far above the expected level of rejection and prove that the tabulated critical values are biased towards rejecting the null hypothesis for student distribution.

Because the kurtosis coefficient of student (9) distribution is 4.33 and of student (25) and student (36) is 3.26 and 3.21 respectively, we can conclude that the tabulated critical values should not be used for distribution with kurtosis greater than 4 and should be used very carefully in a time series with kurtosis close to 4. At any rate, if the kurtosis is less than 4, we should still be using the tabulated critical values carefully. We are technically not able to say the exact number for “safe-usage” of tabulated critical values<sup>24</sup>, but the results show that even the distribution with a very small kurtosis equal to 3.26 or 3.21 (as with student distributions with 25 and 36 degrees of freedom) can generate a rejection rate several times greater than the accepted one at 5%.

Similarly to Table 1, the results in Table 2 show the rejection rates (Type I error) for the  $\chi^2$ -distribution.

Table 2  
Empirical power of Kočenda's test against  $\chi^2$  distribution using tabulated critical values  
The numbers in the table represent the size of a Type I error of the test for the embedding dimension m

Process	m=2	m=3	m=4	m=5	m=6	m=7	m=8	m=9	m=10
chi(3)	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
chi(8)	100.00%	99.80%	99.40%	99.40%	98.40%	97.20%	93.00%	86.20%	82.20%
chi(18)	79.40%	72.80%	66.40%	58.20%	50.20%	39.80%	32.40%	24.20%	22.40%
chi(48)	24.00%	21.20%	17.60%	12.60%	12.00%	10.60%	9.60%	9.80%	8.60%

The results immediately illustrate that usage of tabulated critical values in the case of skewed distributions leads to over-rejection of the null hypothesis. The usage of improper critical values show the rejection rates between 8% and 100%, but the expected rate was about 5%. The rates in the table are too high, and therefore, the tabulated critical values are biased towards rejecting the null hypothesis for  $\chi^2$ -distribution.

Because the skewness coefficients (of used  $\chi^2$ -distribution) are between 0.41 and 1.62, we can realize that the tabulated critical values should not be used for distribution with skewness coefficient greater than 0.41 and should be used very carefully if skewness is less than 0.41 in absolute value. Table 2 implies that although the distributions with a very small skewness (i. e. in our case, small skewness equals 0.41) can give a several times greater rejection rate than it statistically should be.

<sup>24</sup> This is not possible because of the infinite number of nonlinear specifications of the model.

In order to show the ability of the test to correctly distinguish between truly random and random-like data we have performed a series of tests also on the nonlinear sets of processes. The results of these tests are in Table 3. The numbers in Table are now sizes of a Type II error.

Table 3  
Empirical power of Kočenda's test against non-linear processes using tabulated critical values  
The numbers in the table represent the size of a Type II error of the test for the embedding dimension  $m$

Process	m=2	m=3	m=4	m=5	m=6	m=7	m=8	m=9	m=10
GARCH	39.80%	26.20%	20.00%	15.40%	14.40%	15.80%	20.40%	27.60%	29.60%
NLMA	2.20%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.20%	0.20%
ARCH	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.00%	1.40%
ARMA	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

The results of the tests on the nonlinear data have different logic than in previous cases. In Table 1 and Table 2, the numbers represent Type I error (a true null hypothesis is incorrectly rejected). The data used to compute Table 3 were not iid, and therefore, the numbers in Table 3 represent Type II error (a false null hypothesis fails to be rejected). We can see that the results are mostly equal to or close to 0%, only the GARCH process brings results between 14.4% and 39.8%. Therefore, the results are as we expected, and the test successfully finds nonlinear dependencies. However, because of the high kurtosis (6.02) in the case of the NLMA process, the results of power studies for NLMA should be considered very wisely not as a rejection caused by the nonlinear dependencies but as a rejection caused by the high kurtosis itself (due to a large deviation from normality). In the case of such a large kurtosis, usage of Kočenda's test with the tabulated critical values brings void information.

On the other hand, GARCH, ARCH and the ARMA processes have skewness coefficients very close to 0 and a kurtosis coefficient smaller than 4. Therefore, we can use tabulated critical values and provide the valuable decision of the character of process. We do not go into deep details here because a similar discussion was provided in the study by Kočenda and Briatka (2005), and our power studies of nonlinear processes coincide with their results.



#### 4. 4. Results of the Bootstrap

The departure from normality due to a large kurtosis or skewness makes tabulated critical values of little use due to over-rejection of the null hypothesis<sup>25</sup>. Time series with high kurtosis or skewness coefficient should be analyzed with Kočenda's test only using the custom made (bootstrapped) critical values. In the previous chapter, we compared the results of the power tests based on the tabulated critical values with the true critical values and found that even a small deviation from normality could lead to over-rejection of the null hypothesis and hence to a wrong conclusion.

This chapter answers the question whether the bootstrap can be used as an alternative method that would deal with over-rejected difficulties of the tabulated critical values.

All power tests in this Section were evaluated at a 5% significance level; the numbers in Table 4 and 5 are the size of a Type I errors. Table 6 contains the size of a Type II error. Results of the power studies based on the bootstrapped critical values for student distribution are in Table 4.

Table 4  
Empirical power of Kočenda's test against Student distribution using the bootstrap  
The numbers in the table represent the size of a Type I error of the test for the embedding dimension m

Process	m=2	m=3	m=4	m=5	m=6	m=7	m=8	m=9	m=10
student(3)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.60%
student(9)	0.00%	0.00%	0.00%	0.60%	0.80%	1.40%	1.20%	2.00%	3.00%
student(25)	0.00%	0.00%	0.40%	0.40%	1.00%	3.00%	3.80%	3.40%	4.60%
student(36)	0.00%	0.00%	0.80%	0.60%	1.00%	2.20%	2.80%	1.80%	4.20%

At first sight, the results of the power tests based on the data from student distribution using bootstrap are different than those using tabulated critical values (Table 1). In the latter case, the values for the student distribution with 3 and 9 degrees of freedom were far above 30%, now the results are close to zero, not exceeding 3%. This directly indicates two things. The first, the bootstrap method works very well for student distribution. This means that heavy tails cause no problem in the Kočenda's test when using bootstrap. The second, the results of a power test based on the tabulated critical values together with the bootstrap imply that the bootstrap can be used as a double check for the Kočenda's test. In this case, different

<sup>25</sup> However, a time series with large kurtosis or skewness are not very common in real data. See Kočenda and Briatka (2004), Section 5.5 for more details and evidence.

results of the test based on bootstrapped and tabulated critical values are an attribute of the fact that an improper testing procedure was applied to iid data.

The results of tests based on bootstrap for data drawn from  $\chi^2$ -distribution are shown in Table 5.

Table 5  
Empirical power of Kočenda's test against  $\chi^2$  distribution using the bootstrap  
The numbers in the table represent the size of a Type I error of the test for the embedding dimension m

Process	m=2	m=3	m=4	m=5	m=6	m=7	m=8	m=9	m=10
chi(3)	0.20%	0.20%	0.20%	0.20%	0.40%	0.40%	0.60%	0.40%	0.60%
chi(8)	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.40%	1.20%	1.00%
chi(18)	0.00%	0.00%	0.00%	0.20%	1.40%	1.20%	1.80%	2.20%	2.80%
chi(48)	0.00%	0.00%	0.40%	1.20%	2.40%	2.40%	3.00%	5.20%	4.80%

The results are very similar as in the previous situation with the student distributions. The high rejection implied by tabulated critical values (see Table 2) was replaced with close-to-zero (actually none) rejection of bootstrapped critical values. Moreover, the tabulated critical values provide different results within themselves (it depends on the degree of deviation from normality), but the bootstrap results carry constant results over all four  $\chi^2$ -based distributions. The qualitative results of  $\chi^2$ -distributions coincide with the results of student distributions as describe earlier, i. e. the bootstrap method should be use as a surety check if the deviation from normality of time series is too high.

Table 6 shows the size of a Type II error based on the bootstrap for four nonlinear processes.

Table 6  
Empirical power of Kočenda's test against non-linear processes using the bootstrap  
The numbers in the table represent the size of a Type II error of the test for the embedding dimension m

Process	m=2	m=3	m=4	m=5	m=6	m=7	m=8	m=9	m=10
GARCH	98.60%	80.00%	55.20%	42.20%	35.60%	35.20%	41.00%	50.40%	50.20%
NLMA	11.20%	1.10%	0.00%	0.00%	0.00%	0.00%	0.00%	1.40%	0.20%
ARCH	37.00%	10.80%	7.60%	6.80%	8.60%	12.00%	18.60%	27.40%	39.80%
ARMA	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

The rejection rates are a little bit smaller (the numbers in the table are bigger) than those in Table 3; however, they are still high. The size of a Type II error for the NLMA and ARMA processes stand around 0%. The size of a Type II error in the case

of the ARCH process rises from 0% to 18% on average. In the case of the GARCH model, the average increase was 31%.

Following the previous partial results, we conclude that the bootstrap works fine even for nonlinear processes, and the power of the test has changed negligibly. More evidence will be provided in the discussion in the next Section.

## 5. DISCUSSION, COMMENTS, AND RECOMMENDATIONS

To conclude the results of the power tests from the previous section, we suggest adhering to the following strategy in order to handle a time series in the correct way and to minimize the error of rejecting the true null hypothesis.

Initially, perform the normality test for the time series. Now, the basic rules could be summarized in two points:

(1) If the skewness coefficient of a time series is in absolute value greater than 0.5 or kurtosis greater than 4.0, we have to use the bootstrap method to compute asymptotic critical values for Kočenda's test. The decision about the nonlinear behavior of the time series should be derived from the results of the bootstrap.

(2) If the skewness coefficient is in absolute values lower than 0.5 and kurtosis lower than 4.0, we can continue with the standard Kočenda's test without the need of bootstrap. The decision about the nonlinear behavior of the time series could be inferred from the results.

Several final thoughts and recommendations follow.

First, the results presented in this paper were computed using the  $(0.60\sigma; 1.90\sigma)$  interval for proximity parameter  $\varepsilon$ . We also run a very similar Monte Carlo study for different intervals, but the results are not materially different and are available upon request.

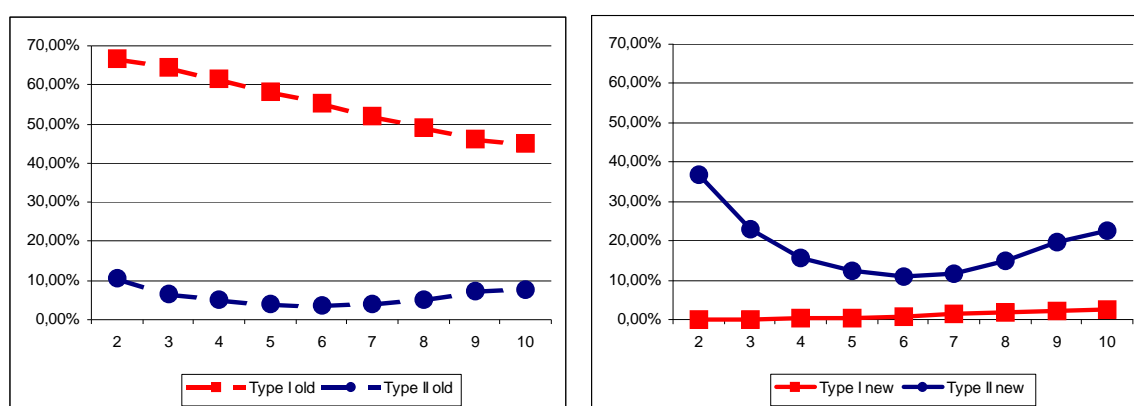
Second, in approaching the normal distribution, we found that the results of the standard and the bootstrap test match if  $|\tau| < 0.5$  and  $\kappa < 4$ . Therefore, this is the point that separates all time series into two groups. In the first group, there are time series that can be verified for nonlinear structure with the standard procedure. Time series from the second group have to be examined with the bootstrap method.

Third, following the basic implications of this paper, we can now run the testing procedure on much more extensive sets of data. Because of the bootstrap procedure,

we do not require the normality properties of the data. However, if the deviation from normality is too large<sup>26</sup>, we should be again careful in applying the test even with bootstrapped critical values.

Another important feature of the test is described in Figure 2. It is the comparison between the properties of the test when using standard methodology (tabulated critical values) and new methodology (based on the bootstrap).

Figure 2  
Comparison of Type I and Type II errors of the test for using tabulated critical values and the bootstrap



The Figure shows a comparison of sizes of Type I and Type II errors between standard methodology (left panel) and the bootstrap method (right panel). In the left panel there is a very low size of Type II error, which means that the standard methodology is very powerful in finding nonlinearities in a time series that are nonlinear in nature. This is the main finding in the paper by Kočenda and Briatka (2005). Yet, the power of the test suffers from a large Type I error, which means that many random time series could be incorrectly considered to be nonlinear.

The situation in the right panel is a little bit different. When using the bootstrap, the size of a Type I error is very low. It means that the probability of rejecting a true null hypothesis is very low. Further, the power of the test is still high enough (comparable to standard methodology) to reveal hidden dependencies in a nonlinear time series. This could be considered as the main finding of the paper.

Generally, to find the balance between the power of the test and the possible errors of the tests is a very hard task. We measured two alternatives for concepts of balance

<sup>26</sup> We do not measure the exact values here because in practice there is no need to consider such a

and as the best solution, we suggest to evaluate both concepts and make a decision based upon the results from both the standard and bootstrap method.

## 6. CONCLUSION

The basic idea of this paper is to identify two separate elements that constitute the rejection power of the Kočenda's test. The first element is nonlinear. It is derived from the fact that the test is designed to find nonlinear dependencies in data. The second element is non-normal, and it is derived from the fact that the test assumes normal distribution of input data. Both elements are jointly summed into the whole rejection power of the Kočenda's test. The main reason why one should distinguish between these two powers is to allow the testing procedure to be applied to a broader set of non-normal data.

Therefore, this paper builds on Kočenda (2001) and Kočenda and Briatka (2005) and sheds light on the properties of the test in the case of non-normally distributed data.

First, the bootstrap method is introduced, and it is shown that by using this method and the optimal interval suggested by Kočenda and Briatka (2005), the power of the test increases significantly for the data that do not look normally distributed although they still may be iid. Using Monte Carlo studies, we found that for a time series with the coefficient of skewness coefficient greater than 0.5 in absolute value and the coefficient of kurtosis greater than 4, bootstrapped critical values have to be employed for the Kočenda's test. Otherwise, the results could be biased towards rejecting the null hypothesis of iid-ness.

Second, the results of power tests on several iid data with different noises (from weak to strong) are provided, and the robustness of the test to not-Gaussian data is evaluated. It is found that when using bootstrap, the size of a Type I error is very low, and hence, the probability of rejecting a true null hypothesis is also low. Hopefully, the power of the test remains still high enough.

Finally, to answer the question from the title of this paper – How big is big enough? – we would say that every time series with  $|\tau| > 0.5$  or  $\kappa > 4$  is big enough. In our concept, it means that deviation from normality of such a time series is so big that

---

pathological case. However, to our knowledge, too large deviated time series should have  $|\tau| > 1$  or  $\kappa > 5$ .

it will influence the result of the Kočenda's test in a way that the test statistic will be biased toward rejecting the null hypothesis. In other words, we should use the bootstrap methodology to test this time series for nonlinearities.

These findings help us to extend the area of using the test into many other fields that deal with nonlinear data that are not necessarily normal, e. g. financial economics, stock price volatility, stock market efficiency, stock exchange, behavior of equity indices, nonlinear dynamics in foreign exchange rates, or interest rates.

## REFERENCES

Ashley, R. A. and Patterson, D. M., 1989. Linear versus nonlinear macroeconomics: A statistical test. *International Economic Review*, 30, 685-704.

Barnett, W. A. and Hinich, M., 1993. Has chaos been discovered with economic data? In: Day, R. H. and Chen, P. (Eds.), *Nonlinear Dynamics and Evolutionary Economics*, 254-265. Oxford University Press, Oxford.

Barnett, W. A., Gallant, A. R., Hinich, M. J., Jungeilges, J. A., Kaplan, D. T., Jensen, M. J., 1997. A single-blind controlled competition among tests for nonlinearity and chaos. *Journal of Econometrics*, 82, 157-192.

Belaire-Franch, J., 2003. A note on resampling the integration across the correlation integral with alternative ranges. *Econometric Reviews*, 22(4), 337-349.

Bera, A. K. and Jarque, C. M., 1981. Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economic Letters*, 7, 313-318.

Bera, A. K. and Jarque, C. M., 1982. Model specification tests. *Journal of Econometrics*, 20, 59-82.

Bollerslev T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.

Brock, W. A., Dechert, W. D., and Scheinkman, J. A., 1987. A test for independence based on the correlation dimension. *SSRI No. 8702, Department of Economics, University of Wisconsin, Madison*.

Brock, W. A., Dechert, W. D., Scheinkman, J. A., and LeBaron, B., 1996. A test for independence based on the correlation dimension. *Econometric Reviews*, 15, 197-235.

Brock, W. A., Hsieh, D. A., and LeBaron, B., 1993. *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*. Third edition, MIT Press, Cambridge, Massachusetts.

Cecen, A. A. and Erkal, C., 1996. Distinguishing between stochastic and deterministic behavior in high-frequency foreign exchange rate returns: Can nonlinear dynamics help forecasting? *International Journal of Forecasting*, 12, 465-473.

Davidson, R. and MacKinnon, J. G., 2000. Bootstrap tests: How many bootstraps? *Econometric Reviews*, 19, 2000, 55-68.

Efron, B. and Tibshirani, R. J., 1998. An introduction to the bootstrap. *Monographs on Statistics and Applied Probability*, 57. Chapman & Hall.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987-1006.

- Grassberger, P. and Procaccia, I., 1983. Measuring the strangeness of strange attractors. *Physica*, 9D, 189-208.
- Hellekalek, P., 1995. Inversible pseudorandom number generators: Concept, results, and links. *Proceedings of the 1995 Winter Simulation Conference*, 255-262.
- Hinich, M. J., 1982. Testing for Gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis*, 3(3), 169-176.
- Hsieh, D. A., 1991. Chaos and nonlinear dynamics: Application to financial markets. *Journal of Finance*, 46, 1839-1877.
- Hsieh, D. A., 1995. Nonlinear Dynamics in Financial Markets: Evidence and Implications. *Financial Analyst Journal*, 51, 55-62.
- Hsieh, D. A. and LeBaron, B., 1993. *Nonlinear Dynamics, Chaos, and Instability. Statistical Theory and Economic Evidence*. Third Edition, MIT Press, Cambridge, Massachusetts.
- Kanzler, L., 1999. Very Fast and Correctly Sized Estimation of the BDS Statistic, SSRN 151669, University of Oxford.
- Kaplan, D. T., 1994. Exceptional events as evidence for determinism. *Physica*, D73, 38-48.
- Kočenda, E., 2001. An alternative to the BDS-test: Integration across the correlation integral. *Econometric Reviews*, 20(3), 337-351.
- Kočenda, E. and Briatka, L., 2004. Advancing the iid test based on integration across the correlation integral: Ranges, competition, and power. *CERGE-EI Working Paper No. 235*, CERGE-EI.
- Kočenda, E. and Briatka, L., 2005. Optimal range for the iid test based on integration across the correlation integral. *Econometric Reviews*, 24(3), 265-296.
- L'Ecuyer, P., 2004. Random number generation. In: Gentle, J. E., Haerdle W., Mori Y. (Eds.), *Handbook of Computational Statistics*. Springer-Verlag.
- LeBaron, B., 1994. Chaos and nonlinear forecastability in economics and finance. *Philosophical Transactions of the Royal Society of London*, A348, 397-404.
- Lee, T. H., White, H., and Granger, C., 1993. Testing for neglected nonlinearities in time series models. *Journal of Econometrics*, 56, 269-290.
- Lehman, E. L., 1997. *Testing Statistical Hypotheses*. Springer-Verlag, New York.
- Nelson, D., B., 1990. ARCH models as diffusion approximations. *Journal of Econometrics*, 45, 7-38.



Nychka, D., Ellner, S., Gallant, R., and McCaffrey, D., 1992. Finding chaos in noisy systems. *Journal of the Royal Statistical Society*, B54, 399-426.

Tsay, R. S., 1986. Nonlinearity tests for time series. *Biometrika*, 73, 461-466.

White, H., 1989. An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. In: *Proceedings of the International Joint Conference on Neural Networks*, vol. 2, 451-455. IEEE Press. New York.

Individual researchers, as well as the on-line and printed versions of the CERGE-EI Working Papers (including their dissemination) were supported from the following institutional grants:

- Center of Advanced Political Economy Research [Centrum pro pokročilá politicko-ekonomická studia], No. LC542, (2005-2009),
- Economic Aspects of EU and EMU Entry [Ekonomické aspekty vstupu do Evropské unie a Evropské měnové unie], No. AVOZ70850503, (2005-2010);
- Economic Impact of European Integration on the Czech Republic [Ekonomické dopady evropské integrace na ČR], No. MSM0021620846, (2005-2011);

Specific research support and/or other grants the researchers/publications benefited from are acknowledged at the beginning of the Paper.

(c) Luboš Briatka, 2006

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical or photocopying, recording, or otherwise without the prior permission of the publisher.

Published by

Charles University in Prague, Center for Economic Research and Graduate Education (CERGE) and

Economics Institute (EI), Academy of Sciences of the Czech Republic

CERGE-EI, Politických vězňů 7, 111 21 Prague 1, tel.: +420 224 005 153, Czech Republic.

Printed by CERGE-EI, Prague

Subscription: CERGE-EI homepage: <http://www.cerge-ei.cz>

Editors: Directors of CERGE and EI

Managing editors: Deputy Directors for Research of CERGE and EI

ISSN 1211-3298

ISBN 80-7343-104-1 (Univerzita Karlova. Centrum pro ekonomický výzkum a doktorské studium)

ISBN 80-7344-093-8 (Akademie věd České republiky. Národohospodářský ústav)



CERGE-EI  
P.O.BOX 882  
Politických vězňů 7  
111 21 Praha 1  
Czech Republic  
<http://www.cerge-ei.cz>