Mathematical Modelling and Analysis
Volume 17 Number 3, June 2012, 383–395
http://dx.doi.org/10.3846/13926292.2012.685959

Publisher: Taylor&Francis and VGTU
http://www.tandfonline.com/TMMA
Print ISSN: 1392-6292
Online ISSN: 1648-3510

Bickel–Rosenblatt Test for Weakly Dependent Data

Janis Valeinis and Audris Locmelis

University of Latvia
Zellu 8, 1002 Riga, Latvia
E-mail(*corresp.*): valeinis@lu.lv
E-mail: locmelis.audris@gmail.com

Received October 4, 2011; revised February 22, 2012; published online June 1, 2012

Abstract. The aim of this paper is to analyze the Bickel–Rosenblatt test for simple hypothesis in case of weakly dependent data. Although the test has nice theoretical properties, it is not clear how to implement it in practice. Choosing different bandwidth sequences first we analyze percentage rejections of the test statistic under H_0 by some empirical simulation analysis. This can serve as an approximate rule for choosing the bandwidth in case of simple hypothesis for practical implementation of the test. In the recent paper [12] a version of Neyman goodness-of-fit test was established for weakly dependent data in the case of simple hypotheses. In this paper we also aim to compare and discuss the applicability of these tests for both independent and dependent observations.

Keywords: goodness-of-fit, Bickel–Rosenblatt test, nonparametric density estimation, weak dependence, Neyman's smooth test.

AMS Subject Classification: 62G10; 65C05.

1 Introduction

Let $(X_t)_{t\in\mathbb{Z}}$ be a real-valued strictly stationary process defined on a probability space $(\Omega, \mathcal{F}, P)$, where X_t has some distribution function F. Considering the classical goodness-of-fit problem assume that we wish to test the following simple hypothesis:

$$H_0\colon \quad F = U[0,1] \quad \text{versus} \quad H_1\colon \quad F \neq U[0,1], \tag{1.1}$$

where $U[0,1]$ denotes the uniform distribution on the interval $[0,1]$. Note that testing H_0: $F = F_0$ for some general continuous distribution F_0 can be reduced to this situation by transforming the data to $F_0(X_t)$, $t \in \mathbb{Z}$.

For independent observations there exist many famous classical goodness-of-fit tests dealing with the hypothesis (1.1) such as Kolmogorov–Smirnov, chi-square, Cramer–von Mises tests. However, usually in practical econometric data problems observations are correlated. For example, consider the well known class of autoregressive moving average (ARMA) processes, which are widely used for forecasting and analysis of time series data. These processes are weak dependent processes with autocorrelations decaying exponentially fast. The weak dependence in general between observations can be described by notion of mixing sequences or processes. After Rosenblatt [15] has introduced the notion of strong mixing processes already in 1956, different other mixing concepts have been defined afterwards (see, for example, [3]). It appears that well known ARMA, GARCH, bilinear and different other nonlinear processes have some mixing property. The purpose of this paper is to analyze some goodness-of-fit tests for dependent observations.

There exist many papers about statistical methods for general mixing sequences. However, the goodness-of-fit problem has been analyzed less extensively. According to our knowledge the first result is due to Neumann and Paparoditis [13], where they consider tests based on the Bickel–Rosenblatt statistic under an exponential β-mixing rate. Next, for simple hypothesis (1.1) the version of Neyman smooth test statistic was introduced in [12] for general α-mixing sequences.

Our initial goal was to compare recently established Neyman test with the Bickel–Rosenblatt test for weakly dependent observations using some empirical power analysis. Although introduced already in 1973 (see [2]), the Bickel–Rosenblatt statistic is problematic to use in practice. The main reason is that the test statistic contains the smoothing bandwidth parameter, which has to be chosen by some data-driven selection rules. This problem differs from usual kernel density bandwidth selection problems and should be treated differently.

Although the Bickel–Rosenblatt test is less known and less used in comparison to the classical goodness-of-fit tests, there are many papers published in recognizable statistical journals analyzing its theoretical properties (for example, [1, 4, 5, 7, 11, 13]). It appears that the Bickel–Rosenblatt test statistic has outstanding property: it has the same asymptotic behaviour in the case of 1) the simple and composite hypotheses; 2) the independent and weak dependent cases. Also it is interesting to mention that unlike in common nonparametric problems, the uniform Kernel appears to be theoretically the best for the Bickel–Rosenblatt test (see [7]).

The main goal of our paper is to analyze the Bickel–Rosenblatt test statistic by empirical simulation study both for independent and dependent observations. More specifically, firstly we aim to find an empirical rule for bandwidth selection for the Bickel–Rosenblatt test in case of the simple hypothesis (1.1). Secondly, we perform an empirical power analysis for the Bickel–Rosenblatt test similarly to the analysis done in [8] and [12]. We also compare the empirical power of both Neyman and Bickel–Rosenblatt tests.

Recently the authors have become aware of the paper [6], where some test statistics based on kernel density estimators in nonparametric regression problems have been analyzed. The closed-form expressions were obtained to explicitly represent the leading terms of both the size and power functions depending on the bandwidth parameter. In this paper it was shown how to control the significance level simultaneously by maximizing the power of the test. It would be interesting to apply these ideas to the Bickel–Rosenblatt test in both the independent and dependent cases, but it is postponed to the future work.

We organize our paper as follows. In Section 2 we introduce both the Neyman and Bickel–Rosenblatt test statistics. Section 3 deals with some simulation study. More specifically, we simulate 1) the asymptotic behaviour of the Bickel–Rosenblatt test statistic under H_0; 2) the percentage rejections of the true H_0; 3) the power against some smooth alternatives discussed in literature.

2 Test Statistics

Bickel–Rosenblatt test. We will use the setup of [13], where they consider the Bickel–Rosenblatt test for dependent observations characterized by absolutely regular or β-mixing processes.

Let $(X_t)_{t\in\mathbb{Z}}$ be a real-valued strictly stationary process defined on a probability space $(\Omega,\mathcal{F},P)$. For any two σ-fields $\mathcal{A}$ and $\mathcal{B}\subset\mathcal{F}$ define the following measure of dependence

$$\beta(\mathcal{A},\mathcal{B}) := \sup\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\bigl|P(A_i\cap B_j)-P(A_i)P(B_j)\bigr|,$$

where this latter supremum is taken over all $\{A_1,\dots,A_I\}$ and $\{B_1,\dots,B_J\}$ from Ω such that $A_i\in\mathcal{A}$ for all i and $B_j\in\mathcal{B}$ for all j. Define $F_J^L := \sigma(X_k, J\le k\le L)$, when $-\infty\le J\le L\le\infty$. $(X_t)_{t\in\mathbb{Z}}$ is called absolutely regular or β-mixing if $\beta(n) := \sup_{J\in\mathbb{Z}}\beta(F_{-\infty}^J, F_{J+n}^\infty)\to 0$ when $n\to\infty$.

Under certain (usually mild) conditions ARMA, GARCH, nonlinear time series models, Markov chains are β-mixing processes (see [3]).

The Bickel–Rosenblatt test statistic has the following form

$$T_n = nh^{1/2}\int\bigl(\hat f_n(x)-(K_h*f)(x)\bigr)^2\,dx,$$

where f is the density under the null hypothesis,

$$\hat f_n(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)$$

is the nonparametric density estimator, where h is a smoothing parameter and finally define

$$(K_h*g)(\cdot)=\int h^{-1}K\left(\frac{\cdot-z}{h}\right)g(z)\,dz$$

for some kernel function K.

Assumptions. For a strictly stationary β-mixing process $(X_t)_{t\in\mathbb{Z}}$ assume the following:

(A′) $\beta(k)\le C\exp(-Ck)$.

(B′) K is bounded and compactly supported.

(C′) f is continuous, furthermore

$$\sup_{x_1,\dots,x_m} \{f_{X_{i_1},\dots,X_{i_m}}(x_1,\dots,x_m)\} < \infty \quad \forall m \text{ and } i_1 < \dots < i_m.$$

(D′) $h = o([\ln(n)]^{-3})$ and $h^{-1} = o(n)$.

Theorem 1. [13] *Assume that conditions* (A′)–(D′) *hold. Then under* H_0 *we have*

$$T_n - h^{-1/2}\int K^2(u)\,du \to_d N\big(0,\sigma^2\big),$$

where

$$\sigma^2 = 2\int f^2(x)\,dx \int\Big(\int K(u)K(u+v)\,du\Big)^2\,dv.$$

H_0 is rejected for large values of the test statistic.

For completeness we will make a small survey of literature regarding the bandwidth parameter selection problem and Bickel–Rosenblatt test. Fan [5] analyzed the Bickel–Rosenblatt test and its extensions in details concluding that for consistency of the test data should be undersmoothed in sense that $nh^{d+2m} \to 0$, where m is the order of the kernel function, d denotes the dimension of the data vector and h denotes the smoothing bandwidth parameter. This also agrees with the previous results in [2] and [16]. For practical simulations Fan [5] analyzes the composite hypothesis of normality and suggests to choose the smoothing parameter in the form of $h = h_0\hat{\sigma}n^{-\delta}$, where $\hat{\sigma}$ denotes the estimator of the standard deviation, δ and h_0 are some positive constants with $0 < \delta < 1$. He chooses $\delta = 2/8, 2/7$ and by empirical percentage rejections of the true H_0 of normality suggests that, for example, for $\delta = 1/4$, the value of $h_0 = 1.90$ gives a good approximation of both 5% and 10% levels for all sample sizes considered in the simulation study. This is maybe the best attempt in the statistical literature to find the right smoothing parameter for testing normality by simulations.

Chebana [4] also argues that one cannot choose the smoothing parameter $h = cn^{-1/5}$ which minimizes the integrated square error and fits for estimation but not for tests. Therefore for simulations she also suggests to follow [5] and chooses $h = h_0n^{-\delta}$, where $\delta = 0.25$ and $\delta = 0.30$ concluding that $h_0 = 0.05$ might be the best choice for the simulations. It is interesting to note that in [11] and [1] the Bickel–Rosenblatt test has been established for the first-order autoregressive time series models. For simulation study they use $h = 1/3n^{-1/5}$ and $h = (\hat{\sigma}^2/n)^{1/5}$, respectively. Finally, we also would like to mention that in the case of weakly dependent absolutely regular or β-mixing processes in [13] a fixed bandwidth $h = 0.03$ has been chosen in the simulation study.

Neyman test. In [12] a modification of Neyman test has been introduced for strong mixing or α-mixing processes. For any two σ-fields $\mathcal{A}$ and $\mathcal{B} \subset \mathcal{F}$ define the following measure of dependence

$$\alpha(\mathcal{A},\mathcal{B}) := \sup\big|P(A\cap B) - P(A)P(B)\big|,$$

where this latter supremum is taken over all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. For $-\infty \leq J \leq L \leq \infty$ define $F_J^L := \sigma(X_k,\ J \leq k \leq L)$. $(X_t)_{t\in\mathbb{Z}}$ is called strongly mixing or α-mixing (see e.g. [3]) if $\alpha(n) := \alpha(F_{-\infty}^0, F_n^\infty) \to 0$ when $n \to \infty$.

It is well known that β-mixing property is stronger than α-mixing property (see [3]). That means that β-mixing processes are also α-mixing, but not vice-versa. A rescaled version of the Neyman statistic has the following form

$$N_k = \left(12\sigma^2\right)^{-1} R_k = \left(12\sigma^2\right)^{-1} \sum_{j=1}^{k} \left\{ n^{-\frac{1}{2}} \sum_{i=1}^{n} \phi_j(X_i) \right\}^2, \tag{2.1}$$

$$\sigma^2 = \sum_{t=-\infty}^{+\infty} \mathrm{Cov}(X_0, X_t), \tag{2.2}$$

where $\phi_0, \phi_1, \ldots,$ is an orthonormal system in $L_2[0,1]$ with $\phi_0(x) = 1$. Commonly one restricts to the Legendre polynomial system on the interval $[0,1]$. Note also that in the case of iid observations under H_0 the factor $(12\sigma^2)^{-1} = 1$ and N_k reduces to R_k, statistic initially proposed in [14].

Ledwina [10] has made the Neyman test appealing in comparison to other well-known tests choosing k by the Schwarz's [17] selection rule. We will use the following modifications of this rule

$$S_{mod} = \min\left\{k\colon 1 \leq k \leq d(n),\ R_k - k\log n \geq R_j - j\log n,\ j = 1, \ldots, d(n)\right\},$$
$$S_{mod2} = \min\left\{k\colon 1 \leq k \leq d(n),\ N_k - k\log n \geq N_j - j\log n,\ j = 1, \ldots, d(n)\right\}.$$

The first rule S_{mod} is based on the Neyman test statistic R_k in the independent case and is equivalent to the usual Schwarz's rule based on the penalized likelihood as shown in [9]. S_{mod2} proposed in [12] takes into account also the underlying dependence structure.

Finally, in [12] the consistency of the rescaled Neyman statistic $N_{S_{mod}}$ has been shown. It appears that under H_0 and rather mild conditions asymptotically the selection rules S_{mod} and S_{mod2} select the dimension 1. Moreover, $N_{S_{mod}}$ has the asymptotic chi-squared distribution with the degree of freedom 1 (see Theorem 2 and Corollaries 1 and 2). On the other hand under any sensible alternative (under H_1) it diverges to infinity.

Assumptions. For a strictly stationary α-mixing process $(X_t)_{t\in\mathbb{Z}}$ assume the following

(A) $\alpha(n) \leq a\rho^n$, for some $a > 0$, $0 < \rho < 1$.

(B) $E|X_t|^\gamma < +\infty$ for some $\gamma > 2$.

(C) $\sigma^2 = \sum_{t=-\infty}^{+\infty} \mathrm{Cov}(X_0, X_t) > 0$.

(D) $d(n) = o(\log n / \log\log n)$.

Let P_0 denote that the marginals of $(X_t)_{t\in\mathbb{Z}}$ are uniformly distributed on $[0,1]$.

Theorem 2. [12] *For a strictly stationary α-mixing process $(X_t)_{t\in\mathbb{Z}}$ assume* (A) *and* (D).

a) *Then* $\lim_{n\to\infty} P_0(S_{mod} = 1) = 1$.

b) *If we assume furthermore that* (B) *and* (C) *hold, then under* H_0*, as* $n \to \infty$*,* $N_{S_{mod}} \to_d \chi_1^2$.

Corollary 1. [12] Assume $\hat{\sigma}^2$ is a consistent estimator of σ^2 defined in (2.2) and further assume (A)–(D). Then

$$\left(12\hat{\sigma}^2\right)^{-1} R_{S_{mod}} \to_d \chi_1^2.$$

Now let us investigate the asymptotical behaviour of $N_{S_{mod}}$ under alternatives. Let X_t have a marginal distribution $P \neq P_0$ on $[0, 1]$. Suppose that

$$E_P\phi_1(X) = \cdots = E_P\phi_{K-1}(X) = 0, \quad E_P\phi_K(X) \neq 0 \tag{2.3}$$

for some $K = K(P) \geq 2$. Consistency of $N_{S_{mod}}$ will be proved for any alternative of the form (2.3). It will be assumed that $\lim_{n\to\infty} \inf d(n) \geq K$, which is certainly the case if $\lim_{n\to\infty} d(n) = \infty$, since K is fixed (see Assumption (D)).

Theorem 3. [12] *Let* $(X_t)_{t\in\mathbb{Z}}$ *be a strictly stationary* α*-mixing process. Assume that* (A) *holds. Then under any alternative* $P \neq P_0$ *with* K *defined in* (2.3)*, as* $n \to \infty$

$$\lim_{n\to\infty} P(S_{mod} \geq K) = 1 \quad \textit{and} \quad N_{S_{mod}} \xrightarrow{P} \infty.$$

Corollary 2. [12] Under the assumptions (A)–(D) Theorems 2 and 3, and Corollary 1 hold also for the selection rule S_{mod2} and the test statistic $N_{S_{mod2}}$.

To use the Neyman test in the dependent case additional to the iid case one needs to estimate σ^2 in (2.2). Define the autocovariance $\gamma(h) := \mathrm{Cov}(X_{t+h}, X_t)$ for all $t, h \in \mathbb{Z}$. For stationary processes and in particular for ARMA processes a common estimate for $\gamma(h)$ is given by

$$\hat{\gamma}(h) = (n-h)^{-1} \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X}) \quad \text{for } 0 \leq h \leq n-1.$$

Thus it naturally leads to the estimator

$$\hat{\sigma}^2 = \hat{\gamma}(0) + 2\sum_{j=1}^{q} \hat{\gamma}(j), \tag{2.4}$$

where q denotes the lag of the last autocovariance $\gamma(q)$, which has to be estimated. One possibility is simply to truncate the autocovariances rounding them to three decimal places in order to find appropriate q and estimate σ^2 as in (2.4) (see [12]). In case of the exponential decrease of mixing coefficients and positive dependence, such a truncation is plausible.

3 Simulation Study

In this section by simulation study we analyze the asymptotic behaviour of the test statistics, the empirical percentage rejections under H_0 and the empirical power against some smooth alternatives in the independent and dependent cases. In order to generate data from α-mixing process we will use the first order autoregressive process $\{X_t\}_{t\in Z}$ defined as

$$X_t - \theta X_{t-1} = Z_t,$$

where $\{Z_t\}_{t\in Z}$ is an innovation process which is weakly stationary with mean 0 and autocovariance $E(Z_t Z_{t+h}) = \sigma_Z^2 < \infty$ if $h = 0$ and 0 otherwise, and $|\theta| \le 1$ is the coefficient of the process. As described in [12] we generate mixing sequences from $AR(1)$ processes with uniform marginals first generating them with normally distributed innovations, then using simple transformation by the respective cumulative distribution function. For our simulation study we will generate data form $AR(1)$ processes with the coefficients $\theta = 0.3, 0.9$ (denoted by models M_1, M_2) and $\theta = -0.3, -0.9$ (denoted by models M_3, M_4) to represent both moderate and strong negative and positive dependence, respectively.

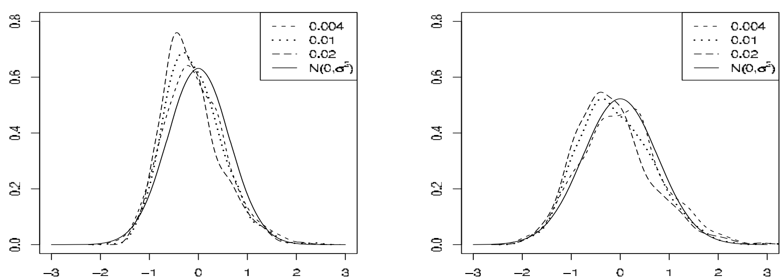

Figure 1. The true density under H_0 (solid line) and three kernel density estimators (with $h = \{0.004, 0.01, 0.02\}$) for the simulated statistic T_n under H_0 for the independent case with $n = 1000$ and 1000 replications. Kernels used in graphics are 1) $N(0,1)$ (left plot) and 2) $U[-1,1]$ (right plot), σ^2 is such as given in Theorem 1.

For illustration we have simulated the behaviour of the statistic T_n in order to see how it depends on the bandwidth sequences and the structure of underlying process. We used the nonparametric density approximations, based on the sample size $n = 1000$, different fixed bandwidth h values and 1000 replications. From the whole simulation study we found out that the results obtained using the uniform kernel only slightly differ from those using the standard normal kernel (see, for example, Figure 1).

In Figure 1 we see the behaviour of T_n in the independent case for both $N(0,1)$ and $U[-1,1]$ kernels. Figures 2 and 3 deal with the Models M_1, M_3 and M_2, M_4, respectively, with $N(0,1)$ kernel. For moderate dependence the optimal bandwidth is close to the independent case. However, for strongly negative and especially positive dependent observations first note that the limiting distribution is not well approximated when $n = 1000$. Moreover, from Figure 3 we can see that the bandwidth choice may strongly differ from the independent case.

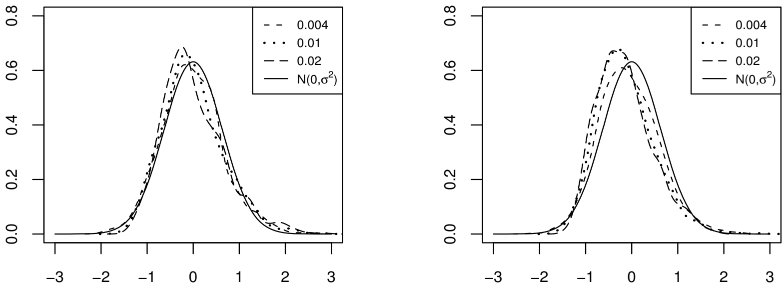

Figure 2. The true density under H_0 (solid line) and three kernel density estimators for the simulated statistic T_n under H_0 for models M_1 (left plot) and M_3 (right plot) with $n = 1000$ and 1000 replications. Kernel is $N(0,1)$ and σ^2 is such as given in Theorem 1.

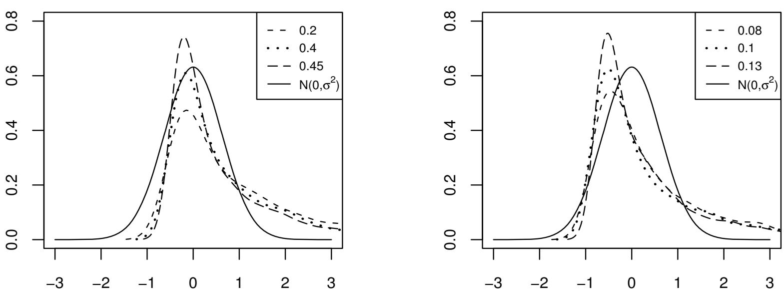

Figure 3. The true density under H_0 (solid line) and three kernel density estimators for the simulated statistic T_n under H_0 for models M_2 (left plot) and M_4 (right plot) with $n = 1000$ and 1000 replications. Kernel is $N(0,1)$ and σ^2 is such that in Theorem 1.

Following [5] and [4] throughout we will use the bandwidth sequence in the form $h = h_0 n^{-\delta}$, where $\delta = 1/4$ and $h_0 = (0.005, 0.01, 0.02, 0.03, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3)$. Other choices of δ lead to similar conclusions with a little different h_0 values.

In Table 1 simulated percentage rejections of the true H_0 are shown for the independent case and M_1, M_3 models. For the parameters h_0 and δ chosen correctly, we expect that the respective column in Table 1 will show approximately

Table 1. T_n percentage rejections of the true H_0 at 5% significance level with $n = 20, 50, 100$ for the *iid* case and models M_1 and M_3 made with 10,000 replications, $h = h_0 n^{-1/4}$; kernel $U[-1, 1]$.

	h_0											
n	0.005	0.01	0.02	0.03	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
$\{iid, 50\}$	6.90	6.61	6.03	5.59	**4.86**	4.34	3.93	3.49	2.98	2.65	2.39	2.25
$\{iid, 500\}$	6.06	6.21	6.35	5.60	5.67	**4.99**	4.51	4.45	4.21	3.88	3.70	3.45
$\{iid, 1000\}$	6.58	6.69	6.50	5.22	**4.99**	5.28	5.58	4.68	4.41	4.21	4.09	3.87
$\{M_1, 50\}$	7.15	7.23	7.35	7.31	7.46	7.98	7.79	7.64	7.50	7.11	6.80	6.32
$\{M_1, 500\}$	7.68	7.65	8.02	7.06	7.80	8.01	7.75	7.95	8.09	8.00	8.08	8.07
$\{M_1, 1000\}$	7.48	8.12	8.48	6.85	6.71	8.27	9.63	8.54	8.87	8.85	8.77	8.66
$\{M_2, 50\}$	37.47	46.68	56.37	62.36	68.14	74.47	75.82	76.36	75.81	75.00	74.01	72.62
$\{M_2, 500\}$	42.55	53.74	64.22	67.86	73.94	78.53	80.32	80.99	81.11	81.03	80.57	80.03
$\{M_2, 1000\}$	41.00	52.78	63.50	66.02	71.38	77.63	80.92	80.52	81.01	80.97	80.77	80.59
$\{M_3, 50\}$	6.26	5.97	**5.31**	**5.31**	4.59	3.59	2.91	2.48	2.07	1.63	1.49	1.29
$\{M_3, 500\}$	5.91	5.94	6.00	5.21	**5.20**	4.34	3.91	3.53	3.13	2.85	2.54	2.29
$\{M_3, 1000\}$	5.95	6.05	5.99	4.49	4.29	4.52	**4.88**	3.66	3.42	3.36	3.11	2.78
$\{M_4, 50\}$	15.91	18.77	21.61	23.77	25.90	28.22	28.21	27.59	26.87	25.88	24.54	23.19
$\{M_4, 500\}$	14.39	18.69	22.89	23.51	26.62	29.68	30.61	30.47	30.20	29.89	29.31	28.76
$\{M_4, 1000\}$	14.68	18.91	22.86	22.64	25.40	30.01	32.57	31.48	31.47	31.43	31.08	30.75

5% for all sample sizes large enough. We can see that for $\delta = -1/4$, $h_0 = 0.05$ could be a good choice for independent and also slight dependent cases (M_1 and M_3 models). Note also that for negatively correlated data the empirical percentage rejections are closer to 5% than for positively correlated data for the chosen set of bandwidth values. Moreover, the results for the independent case are more similar to the slight negative case characterized by model M_3. To conclude from Table 1 we see that for the dependent and the independent cases optimal h_0 choice shall differ already in the case of moderate dependence.

For power analysis we use some alternatives investigated in [10] and [8] which reflect different patterns of the density deviating from the null density. Following these authors we consider alternatives of the form

$$g_1(x) = c(\theta) \exp\left\{ \sum_{j=1}^{k} \theta_j \phi_j(x) \right\}, \tag{3.1}$$

where $\theta = (\theta_1, \ldots, \theta_k)^T \in \mathbb{R}^k$ is a parameter vector and θ^T denotes the transpose of a vector θ. Here $c(\theta)$ is a normalizing constant, such that g_1 integrates to one. We also analyze alternatives of the form

$$g_2(x) = 1 + \rho \cos(j\pi x). \tag{3.2}$$

For illustration see Figure 4 for alternatives g_1 and g_2 defined in (3.1) and (3.2) with different parameters used in simulations $\theta_1 = 0.3$; $\theta_2 = \{0, 0.4\}$; $\theta_3 = \{0.25, -0.35\}$ and $\{\rho, j\} = (\{0.4, 1\}, \{0.5, 2\}, \{0.7, 4\}, \{0.7, 5\}, \{0.7, 6\})$, respectively.

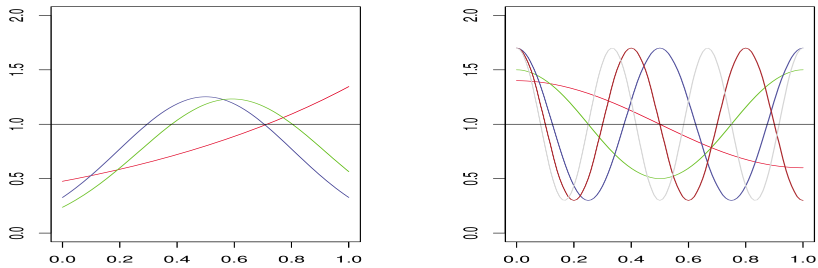

Figure 4. g_1 (left plot) and g_2 (right plot) alternatives chosen for simulation study.

Table 2. Estimated power (%) for g_1 and g_2 alternative defined in (3.1) and (3.2) of $R_{S_{mod}}$ (independent case), $N_{S_{mod2}}$ (M_1 model) when $n = 50$ and $d(n) = 10$ based on 10,000 samples in each case. For estimation of $\hat{\sigma}^2$ we use $q = 3$ in (2.4).

Statistic	θ			$\{\rho, j\}$		
	θ_1	θ_2	θ_3	$\{0.4; 1\}$	$\{0.5; 2\}$	$\{0.7; 4\}$
$R_{S_{mod}}$	38.02	58.16	57.62	34.23	58.16	52.73
$N_{S_{mod2}}$	33.95	56.18	23.95	31.82	30.35	14.94

In Table 2 the empirical power in percentage has been shown for the alternatives g_1 and g_2 for both: 1) the Neyman test $R_{S_{mod}}$ in the independent case; 2) the Neyman test $N_{S_{mod2}}$ in the positively dependent case (model M_1). For negatively correlated observations the Neyman test works much worse (see [12]), therefore we do not analyze this situation here. We base our simulations on 10,000 replications and fix $n = 50$ and $d(n) = 10$. As expected we conclude that in the independent case the Neyman test is more powerful than in the dependent case. For more detailed description and simulation study we refer to [12].

Next for the alternatives g_1 and g_2 we analyze the empirical power for the Bickel–Rosenblatt test statistic T_n. Again we consider only the case with $n = 50$ and base our simulations on 10,000 replications. We choose the kernel as $N(0, 1)$ density and $h = h_0 n^{-1/4}$ with $h_0 = (0.005, 0.01, 0.02, 0.03, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3)$. Tables 3–5 deal with the independent case and M_1, M_3 models, respectively. The maximal power is shown in black. In contrast to the Neyman test the power heavily depends on the smoothing parameter. Clearly if one wishes to control the "right" behaviour of the statistic T_n under H_0 choosing, say $h_0 = 0.01$ (see Table 1) then one may loose power of the test significantly.

Finally, for dependent observations (models M_1, M_2, M_3 and M_4) we plot some power functions (see Figure 5) varying the sample size from $n = 10$ to $n = 500$ for the fixed bandwidth $h = h_0 n^{-1/4}$, where the constant $h_0 = 0.05$ has been selected from the Table 1 as the best one. For this analysis we use the Uniform kernel $U[-1, 1]$ and alternatives considered already in the Tables 3–5. We denote them in the same order shortly by $\{alt1; alt2; alt3; alt4; alt5\}$ from (3.2)

Table 3. *Independent case.* Simulated power of T_n for alternatives g_1 and g_2 with $n = 50$ and 10,000 replications; $h = h_0 n^{-1/4}$, kernel $U[-1, 1]$.

		h_0											
ρ	j	0.005	0.01	0.02	0.03	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
0.4	1	9.2	10.3	13.9	15.6	20.5	28.1	33.2	36.1	38.7	41.7	43.5	**45.4**
0.5	2	11.6	12.3	17.6	21.4	26.8	37.9	43.9	48.8	51.8	54.0	55.3	**55.5**
0.7	4	17.0	24.6	34.8	43.2	55.5	70.0	73.3	**73.6**	69.6	64.9	55.2	40.4
0.7	5	17.6	23.9	34.9	43.1	55.0	66.7	**69.2**	66.4	55.7	38.9	23.2	13.3
0.7	6	16.9	22.6	32.8	40.9	52.7	**62.4**	60.3	50.9	33.5	18.6	10.2	7.2
θ													
$(0, 3)$		25.9	23.5	24.4	26.0	28.4	32.8	36.4	40.1	42.5	44.6	46.6	**48.2**
$(0, -0.4)$		10.4	12.4	17.5	21.9	26.8	34.0	39.1	41.9	42.9	45.3	**46.0**	45.8
$(0.25, -0.35)$		10.0	12.0	17.6	21.0	27.2	33.7	38.2	42.5	45.2	48.5	49.6	**50.5**

Table 4. $AR(1)$ case with $\phi = 0.3$ (M_1 model). Simulated power for alternatives g_1 and g_2 with $n = 50$ and 10,000 replications; $h = h_0 n^{-1/4}$, kernel $U[-1, 1]$.

		h_0											
ρ	j	0.005	0.01	0.02	0.03	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
0.4	1	8.1	9.3	12.4	13.8	16.7	20.9	24.4	26.5	27.9	29.7	30.3	**30.8**
0.5	2	10.8	11.9	15.9	19.5	25.3	31.8	36.2	38.6	39.7	**40.5**	39.9	38.9
0.7	4	18.1	22.8	31.7	39.5	48.4	61.1	66.3	**64.0**	59.0	50.9	37.9	25.7
0.7	5	16.7	21.7	32.4	38.8	48.3	58.8	**60.3**	52.9	41.1	26.5	15.9	10.4
0.7	6	16.6	22.1	30.2	39.1	48.2	**56.7**	52.4	40.7	24.2	11.7	7.8	6.4
θ													
$(0, 3)$		27.6	23.3	22.5	24.0	26.1	28.4	32.5	33.3	33.6	35.1	36.5	**37.1**
$(0, -0.4)$		10.9	10.9	14.0	16.2	19.4	24.9	31.3	32.9	33.7	**34.1**	33.8	32.3
$(0.25, -0.35)$		9.5	10.6	14.3	16.5	20.0	26.9	32.9	34.9	37.1	38.2	**38.5**	38.0

Table 5. $AR(1)$ case with $\phi = -0.3$ (M_3 model). Simulated power for alternatives g_1 and g_2 with $n = 50$ and 10,000 replications; $h = h_0 n^{-1/4}$, kernel $U[-1, 1]$.

		h_0											
ρ	j	0.005	0.01	0.02	0.03	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
0.4	1	6.1	9.2	10.8	12.8	15.2	22.5	28.6	34.4	37.3	41.5	44.9	**48.1**
0.5	2	8.9	12.8	16.6	19.2	26.3	38.9	46.0	50.7	54.6	57.5	58.7	**60.7**
0.7	4	15.3	21.9	33.8	40.5	53.0	70.0	75.7	**77.9**	75.4	70.4	62.2	48.7
0.7	5	16.3	22.6	33.6	39.7	53.6	67.9	**70.8**	68.1	58.3	42.2	26.0	15.8
0.7	6	16.1	22.0	32.0	38.9	52.2	64.2	**64.8**	55.3	39.0	22.2	12.5	9.6
θ													
$(0, 3)$		25.1	25.3	25.2	24.5	27.6	34.6	40.1	45.3	49.8	53.7	57.4	**60.9**
$(0, -0.4)$		7.8	11.1	15.4	17.1	23.2	33.4	38.4	42.5	44.8	47.0	47.7	**48.3**
$(0.25, -0.35)$		9.3	11.7	15.5	17.4	23.7	33.8	40.2	45.7	48.5	51.6	53.3	**55.5**

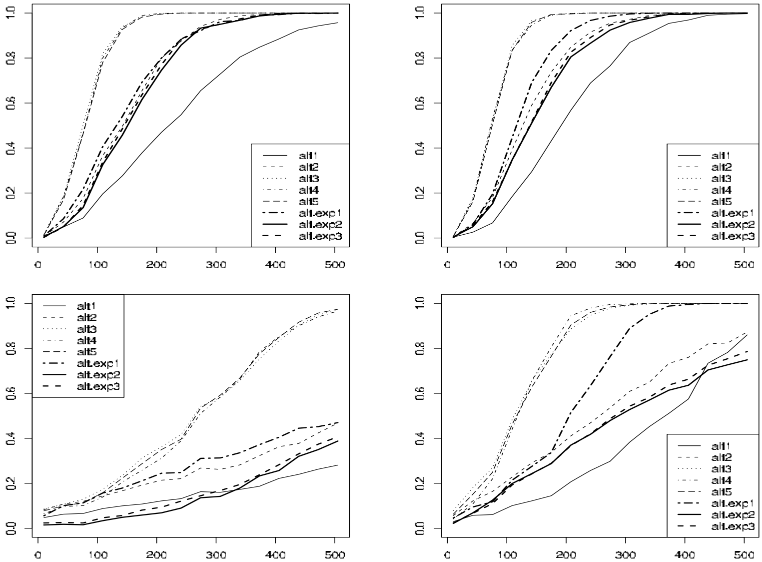

Figure 5. Empirical power plots for different alternatives (3.2) and (3.1) for models M_1 (top left plot), M_3 (top right plot), M_2 (bottom left plot) and M_4 (bottom right plot) with different sample sizes, based on 1000 replications. Kernel is $U[-1,1]$ and $h = h_0 n^{-1/4}$ with $h_0 = 0.05$.

and $\{alt.exp1, alt.exp2, alt.exp3\}$ from (3.1).

We conclude that the bandwidth choice heavily affects the power behaviour of the test. For different dependence structures the bandwidth parameter should be chosen differently under H_0. However, for slight dependences (positive or negative) the behaviour of the Bickel–Rosenblatt test is similar to the independent case and one could suggest to use for testing purposes, for example, the bandwidth $h = h_0 n^{-1/4}$ with $h_0 = 0.05$. As it has been mentioned in the Introduction it would be interesting to apply the ideas developed in [6] for the Bickel–Rosenblatt test both for independent and also for dependent observations.

Acknowledgements

Valeinis acknowledges a partial support of the project 2009/0223/1DP/ 1.1.1.2.0/09/APIA/VIAA/008 of the European Social Fund.

References

[1] D. Bachmann and H. Dette. A note on the Bickel–Rosenblatt test in autoregressive time series. *Statist. Probab. Lett.*, **74**(3):221–234, 2005. http://dx.doi.org/10.1016/j.spl.2005.04.003.

[2] P.J. Bickel and M. Rosenblatt. On some global measures of the deviations of density function estimates. *Ann. Statist.*, **1**(6):1071–1095, 1973. http://dx.doi.org/10.1214/aos/1176342558.

[3] R.C. Bradley. *Introduction to Strong Mixing Conditions.* Kendrick Press, Heber City, Utah, 2007.

[4] F. Chebana. On the optimization of the weighted Bickel–Rosenblatt test. *Statist. Probab. Lett.*, **68**(4):333–345, 2004. http://dx.doi.org/10.1016/j.spl.2004.03.007.

[5] Y. Fan. Testing the goodness of fit of a parametric density function by kernel method. *Econometric Theory*, **10**(2):316–356, 1994. http://dx.doi.org/10.1017/S0266466600008434.

[6] J. Gao and I. Gijbels. Bandwidth selection in nonparametric kernel testing. *J. Amer. Statist. Assoc.*, **103**(484):1584–1594, 2008. http://dx.doi.org/10.1198/016214508000000968.

[7] B.K. Ghosh and W.M. Huang. The power and optimal kernel of the Bickel–Rosenblatt test for goodness of fit. *Ann. Statist.*, **19**(2):999–1009, 1991. http://dx.doi.org/10.1214/aos/1176348133.

[8] W.C.M. Kallenberg and T. Ledwina. Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *Ann. Statist.*, **23**(5):1594–1608, 1995. http://dx.doi.org/10.1214/aos/1176324315.

[9] W.C.M. Kallenberg and T. Ledwina. Data-driven smooth tests when the hypothesis is composite. *J. Amer. Statist. Assoc.*, **92**(439):1094–1104, 1997. http://dx.doi.org/10.2307/2965574.

[10] T. Ledwina. Data-driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.*, **89**(427):1000–1005, 1994. http://dx.doi.org/10.2307/2290926.

[11] S. Lee and S. Na. On the Bickel–Rosenblatt test for first-order autoregressive models. *Statist. Probab. Lett.*, **56**(1):23–35, 2002. http://dx.doi.org/10.1016/S0167-7152(01)00143-2.

[12] A. Munk, J.P. Stockis, J. Valeinis and G. Giese. Neyman smooth goodness-of-fit tests for the marginal distribution of dependent data. *Ann. Inst. Statist. Math.*, **63**(5):939–959, 2011. http://dx.doi.org/10.1007/s10463-009-0260-2.

[13] M.H. Neumann and E. Paparoditis. On bootstrapping L2-type statistics in density testing. *Statist. Probab. Lett.*, **50**(2):137–147, 2000. http://dx.doi.org/10.1016/S0167-7152(00)00091-2.

[14] J. Neyman. Smooth test for goodness of fit. *Skand. Aktuarietidskr.*, **20**:150–199, 1937.

[15] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**(3):832–837, 1956. http://dx.doi.org/10.1214/aoms/1177728190.

[16] M. Rosenblatt. A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.*, **3**(1):1–14, 1975. http://dx.doi.org/10.1214/aos/1176342996.

[17] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, **6**(2):461–464, 1978. http://dx.doi.org/10.1214/aos/1176344136.